Probabilidad

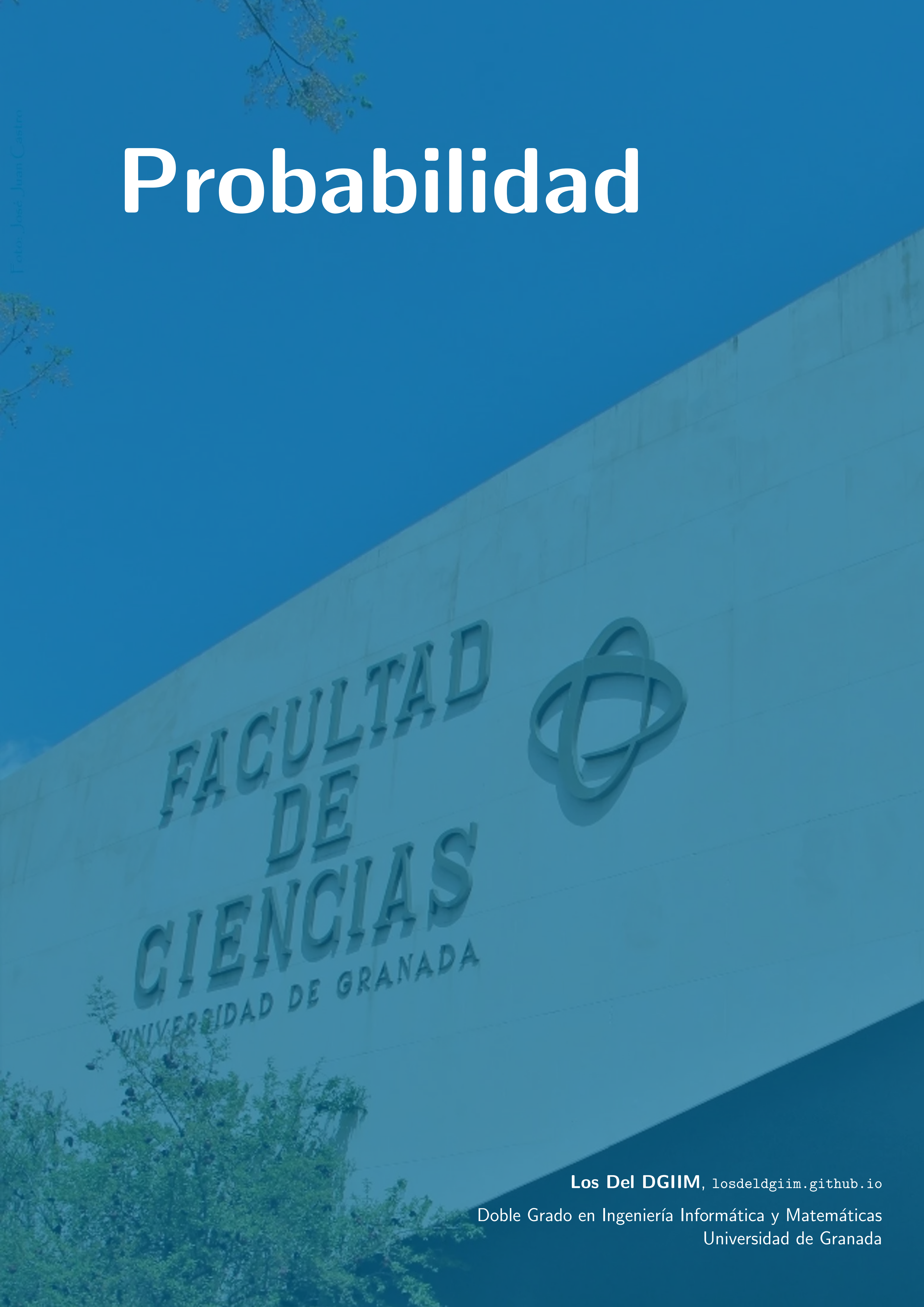


Foto: José Juan Castro

Los Del DGIIM, `losdeldgiim.github.io`

Doble Grado en Ingeniería Informática y Matemáticas
Universidad de Granada

Probabilidad

Los Del DGIIM, `losdeldgiim.github.io`

Arturo Olivares Martos

Granada, 2024-2025

Índice general

0. Introducción

En la asignatura de Estadística Descriptiva e Introducción a la Probabilidad se presentaron los conceptos más importantes de Probabilidad unidimensional, junto con diversos ejemplos de distribuciones de variables aleatorias discretas.

Este primer tema es un repaso de lo más importante de dicha asignatura, haciendo especial hincapié en las mencionadas distribuciones discretas. Se recomienda al lector por tanto que se dirija a los apuntes de dicha asignatura, revisando así estos conceptos.

1. Distribuciones de Probabilidad Continua

En el presente capítulo, se estudiarán las distribuciones de probabilidad continua más importantes. Al igual que en la asignatura de EDIP se vieron ejemplos para variables aleatorias discretas, en este tema se presentarán las distribuciones más relevantes para variables aleatorias continuas.

1.1. Distribución Uniforme Continua

Definición 1.1 (Distribución Uniforme Continua)**.** Una variable aleatoria continua X sigue una distribución uniforme en el intervalo $[a, b]$, con $a, b \in \mathbb{R}$, $a < b$, si su función de densidad toma un valor constante en dicho intervalo, siendo nula fuera de él. Lo denotaremos como $X \sim \mathcal{U}(a, b)$.

Proposición 1.1. *Sea $X \sim \mathcal{U}(a, b)$. Entonces, se tiene que:*

$$\begin{array}{rccl} f: & \mathbb{R} & \longrightarrow & [0,1] \\ & x & \longmapsto & \begin{cases} \dfrac{1}{b-a} & x \in [a,b] \\ 0 & x \notin [a,b] \end{cases} \end{array}$$

Demostración. Sea f la función de densidad de X. Para que sea una función de densidad, debe integrar 1 en todo $\mathbb{R}$. Como esta es nula fuera de $[a, b]$, tenemos que:

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

Como f es constante en $[a, b]$, sea entonces $f(x) = k$ para $x \in [a, b]$. Entonces:

$$\int_a^b f(x)dx = \int_a^b kdx = k\int_a^b dx = k(b-a) = 1 \Longrightarrow k = \frac{1}{b-a}$$

Por tanto, la función de densidad de X es:

$$f(x) = \frac{1}{b-a} \quad \forall x \in [a, b]$$

$\square$

Proposición 1.2. *Sea $X \sim \mathcal{U}(a,b)$, entonces su función de distribución es:*

$$\begin{array}{rccl} F_X: & \mathbb{R} & \longrightarrow & [0,1] \\ & x & \longmapsto & \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & x \in [a,b] \\ 1 & x > b \end{cases} \end{array}$$

Demostración. Tenemos que:

$$F_X(x) = \int_{-\infty}^{x} f(t)\ dt = \int_a^x \frac{1}{b-a}\ dt = \frac{1}{b-a}\int_a^x\ dt = \frac{x-a}{b-a}$$

□

Como consecuencia inmediata a la proposición anterior, vemos como definición alternativa que, si X es una variable aleatoria continua tal que $X \sim \mathcal{U}(a,b)$, entonces se tiene que la probabilidad de que X tome un valor en un intervalo $[c,d]$, con $a \leqslant c \leqslant d \leqslant b$, es directamente proporcional a la longitud del intervalo.

Proposición 1.3. *Sea X una variable aleatoria tal que $X \sim \mathcal{U}(a,b)$. Su función generatriz de momentos es:*

$$M_X(t) = \frac{e^{tb} - e^{ta}}{(b-a)t} \qquad t \neq 0$$

Para $t = 0$, $M_X(0) = 1$.

Demostración. Veamos en primer lugar el caso $t = 0$. Aunque ya esté demostrado en el temario de EDIP, esto es una propiedad general de las funciones generatrices de momentos, ya que:

$$M_X(0) = E\left[e^{0X}\right] = E[1] = 1$$

Para $t \neq 0$, tenemos que:

$$\begin{aligned} M_X(t) = E\left[e^{tX}\right] &= \int_a^b e^{tx}\frac{1}{b-a}\ dx = \frac{1}{b-a}\int_a^b e^{tx}\ dx = \\ &= \frac{1}{b-a}\left[\frac{e^{tx}}{t}\right]_a^b = \frac{e^{tb} - e^{ta}}{(b-a)t} \end{aligned}$$

□

Calculemos ahora los momentos de una variable aleatoria X tal que $X \sim \mathcal{U}(a,b)$.

Proposición 1.4. *Sea X una variable aleatoria tal que $X \sim \mathcal{U}(a,b)$. Entonces, su momento no centrado de orden $k \in \mathbb{N} \cup \{0\}$ es:*

$$m_k = E[X^k] = \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}$$

Demostración. Tenemos que:

$$m_k = E[X^k] = \int_a^b x^k \frac{1}{b-a}\, dx = \frac{1}{b-a}\int_a^b x^k\, dx =$$
$$= \frac{1}{b-a}\left[\frac{x^{k+1}}{k+1}\right]_a^b = \frac{b^{k+1}-a^{k+1}}{(k+1)(b-a)}$$

□

Como consecuencia, tenemos que:

$$m_1 = E[X] = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

Proposición 1.5. *Sea X una variable aleatoria tal que $X \sim \mathcal{U}(a,b)$. Entonces, su momento centrado de orden $k \in \mathbb{N}$ es:*

$$\mu_k = \begin{cases} 0 & k \text{ impar} \\ \dfrac{(b-a)^k}{(k+1)2^k} & k \text{ par} \end{cases}$$

Demostración. Tenemos que:

$$\mu_k = E[(X-m_1)^k] = E\left[\left(X - \frac{a+b}{2}\right)^k\right] = \int_a^b \left(x - \frac{a+b}{2}\right)^k \frac{1}{b-a}\, dx =$$
$$= \frac{1}{b-a}\int_a^b \left(x-\frac{a+b}{2}\right)^k dx = \frac{1}{b-a}\left[\frac{\left(x-\frac{a+b}{2}\right)^{k+1}}{k+1}\right]_a^b =$$
$$= \frac{\left(b-\frac{a+b}{2}\right)^{k+1} - \left(a-\frac{a+b}{2}\right)^{k+1}}{(k+1)(b-a)} = \frac{\left(\frac{b-a}{2}\right)^{k+1} - \left(-\frac{b-a}{2}\right)^{k+1}}{(k+1)(b-a)}$$

Distinguimos ahora en función de la paridad de k:

- Si k es impar, entonces $\mu_k = 0$.
- Si k es par, entonces $\mu_k = \dfrac{(b-a)^k}{(k+1)2^k}$.

□

Algunos ejemplos de su utilidad son los siguientes:

- La distribución uniforme proporciona una representación adecuada para redondear las diferencias que surgen al medir cantidades físicas entre los valores observados y los reales. Por ejemplo, si el peso de una persona se redondea al *kg* más cercano, entonces la diferencia entre el peso observado y el real será algún valor entre $-0{,}5$ y $0{,}5$ *kg*. Es común que el error de redondeo siga entonces una distribución $\mathcal{U}(-0{,}5, 0{,}5)$.
- La generación de números aleatorios en un intervalo $[a,b]$ debe seguir una distribución $\mathcal{U}(a,b)$.

1.2. Distribución Normal

Esta es la distribución de probabilidad más importante en la Teoría de la Probabilidad y la Estadística Matemática.

Definición 1.2 (Distribución Normal)**.** Una variable aleatoria continua X sigue una distribución normal con parámetros $\mu \in \mathbb{R}$ y $\sigma^2 \in \mathbb{R}^+$, si su función de densidad es:

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

donde $\sigma = +\sqrt{\sigma^2}$. Lo denotaremos como $X \sim \mathcal{N}(\mu, \sigma^2)$.

A pesar de darse como definición, hemos de demostrar que efectivamente es una función de densidad. Para ello, incluimos el siguiente Lema, cuya demostración no se incluye por su complejidad, al requerir de integración en varias variables con cambio a coordenadas polares.

Lema 1.6 (Integral de Gauss)**.** *Sea $a \in \mathbb{R}^+$, $b \in \mathbb{R}$. Entonces:*

$$\int_{-\infty}^{\infty} e^{-a(x+b)^2}\, dx = \sqrt{\frac{\pi}{a}}$$

Usando este lema, podemos demostrar que la función de densidad de la normal es efectivamente una función de densidad.

Demostración. La función f_X debe cumplir las siguientes propiedades:

- $f_X(x) \geqslant 0$ para todo $x \in \mathbb{R}$.

  Esto es directo puesto que el término exponencial siempre es positivo.

- f_X es integrable, por ser producto y composición de funciones integrables.

- $\displaystyle\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

$$\int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \overset{(*)}{=}$$
$$\overset{(*)}{=} \frac{1}{\sqrt{2\pi}\sigma} \cdot \sqrt{\frac{\pi}{\frac{1}{2\sigma^2}}} = 1$$

donde en $(*)$ hemos aplicado la Integral de Gauss con $a = \dfrac{1}{2\sigma^2}$ y $b = -\mu$.

□

Teorema 1.7 (Tipificación de la Normal)**.** *Sea $X \sim \mathcal{N}(\mu, \sigma^2)$. Entonces, la variable aleatoria $Z = \dfrac{X-\mu}{\sigma}$ se dice que es la variable aleatoria tipificada de X. Se cumple que:*

1. *$Z \sim \mathcal{N}(0, 1)$.*

2. $P[X \leqslant x] = P\left[Z \leqslant \dfrac{x-\mu}{\sigma}\right]$ *para todo* $x \in \mathbb{R}$.

Demostración. Demostramos cada uno de los puntos:

1. Para esto, hay que emplear el Teorema de Cambio de Variable de variable aleatoria continua a variable aleatoria continua. Definimos la siguiente función:

$$\begin{array}{rccc} g: & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & \dfrac{x-\mu}{\sigma} \end{array}$$

   Tenemos por tanto que $Z = g(X)$. La inversa de g es:

$$\begin{array}{rccc} g^{-1}: & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & z & \longmapsto & \sigma z + \mu \end{array}$$

   La función de densidad de Z es, por tanto:

$$\begin{aligned} f_Z(z) = f_X(g^{-1}(z)) \left|(g^{-1})(z)\right| = f_X(\sigma z + \mu) \cdot \sigma = \frac{1}{\sqrt{2\pi}\cancel{\sigma}} e^{-\dfrac{(\sigma z + \cancel{\mu} - \cancel{\mu})^2}{2\sigma^2}} \cdot \cancel{\sigma} = \\ = \frac{1}{\sqrt{2\pi}} e^{-\dfrac{z^2}{2}} \qquad \forall z \in \mathbb{R} \mid (\sigma z + \mu) \in \mathbb{R} = \\ = \frac{1}{\sqrt{2\pi}} e^{-\dfrac{z^2}{2}} \qquad \forall z \in \mathbb{R} \end{aligned}$$

   Por tanto, identificando términos, tenemos que $Z \sim \mathcal{N}(0, 1)$.

2. Este resultado se obtiene de forma directa por el Teorema de Cambio de Variable usando los dos Corolarios que lo acompañan vistos en EDIP.

□

Proposición 1.8. *Sea* $X \sim \mathcal{N}(\mu, \sigma^2)$*. Entonces, su función generatriz de momentos es:*

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Demostración. Demostraremos este resultado en dos pasos:

Caso particular Demostramos en primer lugar el caso para la variable $Z \sim \mathcal{N}(0, 1)$:

$$\begin{aligned} M_Z(t) = E\left[e^{tZ}\right] = \int_{\mathbb{R}} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{tz - \frac{z^2}{2}}\, dz = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{z^2 - 2tz + t^2 - t^2}{2}}\, dz = \\ = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(z-t)^2}{2}} e^{\frac{t^2}{2}}\, dz = \frac{e^{\frac{t^2}{2}}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(z-t)^2}{2}}\, dz \end{aligned}$$

Debido a que la integral de una función de densidad de una variable aleatoria con distribución $\mathcal{N}(t, 1)$ en todo $\mathbb{R}$ es 1, tenemos que:

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}}\, dz = 1$$

Por tanto:

$$M_Z(t) = e^{\frac{t^2}{2}}$$

Caso general Demostramos ahora el caso general para $X \sim \mathcal{N}(\mu, \sigma^2)$. Consideramos la variable aleatoria tipificada de X, $Z \sim \mathcal{N}(0,1)$:

$$Z = \frac{X-\mu}{\sigma}$$

Por tanto, tenemos que $X = \sigma Z + \mu$. Por tanto:

$$M_X(t) = E\left[e^{tX}\right] = E\left[e^{t(\sigma Z+\mu)}\right] = E\left[e^{t\sigma Z}e^{t\mu}\right]$$

Puesto que $e^{t\mu}$ es una constante, por la linealidad de la esperanza tenemos:

$$M_X(t) = E\left[e^{t\sigma Z}e^{t\mu}\right] = e^{t\mu}E\left[e^{(t\sigma)Z}\right] = e^{t\mu}M_Z(t\sigma) = e^{t\mu}e^{\frac{(t\sigma)^2}{2}} = e^{t\mu+\frac{\sigma^2t^2}{2}}$$

□

Una vez ya razonada la función generatriz de momentos, podemos entonces entender por qué los parámetros de la normal son μ y σ^2. Veamos en primer lugar que μ es la esperanza de la variable aleatoria ($E[X]$ se nota también como $\overline{X}$ o μ):

Proposición 1.9. *Sea $X \sim \mathcal{N}(\mu, \sigma^2)$. Entonces, su esperanza es:*

$$E[X] = \mu$$

Demostración. Tenemos que:

$$E[X] = \frac{d}{dt}M_X(t)\Big|_{t=0} = \frac{d}{dt}e^{t\mu+\frac{\sigma^2t^2}{2}}\Big|_{t=0} = e^{t\mu+\frac{\sigma^2t^2}{2}}\left(\mu+\sigma^2t\right)\Big|_{t=0} = e^0(\mu+0) = \mu$$

□

De igual forma, podemos ver que σ^2 es la varianza de la variable aleatoria (ya que $\mathrm{Var}[X]$ también se nota por σ^2):

Proposición 1.10. *Sea $X \sim \mathcal{N}(\mu, \sigma^2)$. Entonces, su varianza es:*

$$\mathrm{Var}[X] = \sigma^2$$

Demostración. Calculemos en primer lugar $E[X^2]$ con la función generatriz de momentos:

$$\begin{aligned}E[X^2] &= \frac{d^2}{dt^2}M_X(t)\Big|_{t=0} = \frac{d^2}{dt^2}e^{t\mu+\frac{\sigma^2t^2}{2}}\Big|_{t=0} = \frac{d}{dt}\left[e^{t\mu+\frac{\sigma^2t^2}{2}}(\mu+\sigma^2t)\right]\Big|_{t=0} = \\ &= \left(e^{t\mu+\frac{\sigma^2t^2}{2}}\left(\mu+\sigma^2t\right)^2 + e^{t\mu+\frac{\sigma^2t^2}{2}}\sigma^2\right)\Big|_{t=0} = e^0(\mu^2)+e^0\sigma^2 = \mu^2+\sigma^2\end{aligned}$$

Tenemos por tanto, usando que $E[X] = \mu$:

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = \mu^2+\sigma^2-\mu^2 = \sigma^2$$

□

Una de las propiedades más importantes de la distribución normal es que es simétrica respecto a su media.

Proposición 1.11. *Sea X una variable aleatoria tal que $X \sim \mathcal{N}(\mu, \sigma^2)$. Entonces, es simétrica respecto a su media, es decir:*

$$P[X \leqslant \mu - x] = P[X \geqslant \mu + x] \quad \forall x \in \mathbb{R}$$

Demostración. Para probar esto, probaremos que su función de densidad es simétrica respecto a su media. Tenemos que:

$$f_X(\mu - x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{((\mu - x) - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(-x)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{x^2}{2\sigma^2}} =$$
$$= \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{((\mu + x) - \mu)^2}{2\sigma^2}} = f_X(\mu + x)$$

Veamos ahora lo pedido. Como f_X es una función de densidad, tenemos que:

$$P[X \geqslant \mu + x] = 1 - P[X \leqslant \mu + x] = 1 - \int_{-\infty}^{\mu+x} f_X(t)\ dt =$$
$$= \int_{-\infty}^{+\infty} f_X(t)\ dt - \int_{-\infty}^{\mu+x} f_X(t)\ dt = \int_{\mu+x}^{+\infty} f_X(t)\ dt$$

donde podemos restar las integrales puesto que todas ellas son convergentes. Aplicamos ahora el cambio de variable $t = \mu + u$:

$$\left[\begin{array}{c} t = \mu + u \\ \dfrac{dt}{du} = 1 \end{array}\right] \qquad \left\{\begin{array}{l} \lim\limits_{u \to x} t = \mu + x \\ \lim\limits_{u \to \infty} t = +\infty \end{array}\right.$$

Por tanto, tenemos que:

$$P[X \geqslant \mu + x] = \int_{\mu+x}^{+\infty} f_X(t)\ dt = \int_{x}^{\infty} f_X(\mu + u)\ du = \int_{x}^{\infty} f_X(\mu - u)\ du$$

Notemos que este primer cambio de variable ha sido esencial para poder aplicar la simetría. Aplicamos ahora el cambio de variable $u = -v + \mu$:

$$\left[\begin{array}{c} u = -v + \mu \\ \dfrac{du}{dv} = -1 \end{array}\right] \qquad \left\{\begin{array}{l} \lim\limits_{v \to \mu - x} t = x \\ \lim\limits_{v \to -\infty} t = +\infty \end{array}\right.$$

Por tanto, tenemos que:

$$P[X \geqslant \mu + x] = \int_{x}^{\infty} f_X(\mu - u)\ du = -\int_{\mu-x}^{-\infty} f_X(v)\ dv = \int_{-\infty}^{\mu-x} f_X(v)\ dv =$$
$$= P[X \leqslant \mu - x]$$

Notemos que, de forma intuitiva, lo que hacemos con el primer cambio de variable es “llevarlo al eje de simetría”, y en ese eje aplicamos la simetría y “deshacemos” el cambio hecho anteriormente. □

Otra propiedad importante de la distribución normal es que la media, mediana y moda coincide.

Corolario 1.11.1. *Sea X una variable aleatoria tal que $X \sim \mathcal{N}(\mu, \sigma^2)$. Entonces:*

$$\mu = E[X] = Me[X] = Mo[X]$$

Demostración. Calculemos por separado ambos valores:

Cálculo de la Mediana Sabiendo que la distribución es simétrica respecto a su media, veamos que $P[X \leqslant \mu] = P[X \geqslant \mu] = 1/2$.

La primera igualdad es directa, puesto que $P[X \leqslant \mu] = P[X \geqslant \mu]$ por ser simétrica respecto de μ. Posteriormente, tenemos que:

$$P[X \geqslant \mu] = 1 - P[X \leqslant \mu] = 1 - P[X \geqslant \mu] \Longrightarrow P[X \geqslant \mu] = \frac{1}{2}$$

Por tanto, $\mu = Me[X]$.

Cálculo de la Moda Es el máximo absoluto de la función de densidad. Calculémoslo mediante el estudio de la primera derivada:

$$f'_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \left(-\frac{2(x-\mu)}{2\sigma^2} \right) = 0 \Longleftrightarrow x = \mu$$

Por tanto, vemos que el único candidato a extremo relativo es μ. Además, vemos que f'_X es creciente para $x < \mu$ y decreciente para $x > \mu$, de lo que deducimos que μ es un máximo absoluto. Por tanto, $\mu = Mo[X]$. □

Teorema 1.12 (Regla de la Probabilidad Normal)**.** *Sea X una variable aleatoria continua tal que $X \sim \mathcal{N}(\mu, \sigma^2)$. Entonces:*

1. $P[\mu - \sigma \leqslant X \leqslant \mu + \sigma] \approx 0{,}6826$
2. $P[\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma] \approx 0{,}9544$
3. $P[\mu - 3\sigma \leqslant X \leqslant \mu + 3\sigma] \approx 0{,}9974$

Demostración. Vamos a demostrar el primer apartado, siendo los demás análogos.

$$\begin{aligned} P[\mu - \sigma \leqslant X \leqslant \mu + \sigma] &= P\left[\frac{\mu - \sigma - \mu}{\sigma} \leqslant Z \leqslant \frac{\mu + \sigma - \mu}{\sigma}\right] = P[-1 \leqslant Z \leqslant 1] = \\ &= P[Z \leqslant 1] - P[Z \leqslant -1] = P[Z \leqslant 1] - P[Z \geqslant 1] = \\ &= 2P[Z \leqslant 1] - 1 \approx 2 \cdot 0{,}8413 - 1 \approx 0{,}6826 \end{aligned}$$

donde Z representa la variable aleatoria tipificada de X y, al terminar, hemos consultado los valores en la tabla de la distribución normal estándar. □

Su gráfica es ampliamente conocida y tiene forma de campana, como podemos ver en la Figura 1.1.

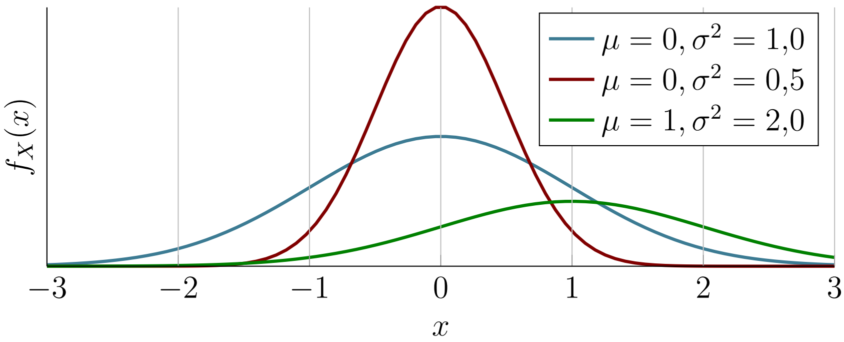


Figura 1.1: Función de densidad de una v. a. con distribución normal.

1.2.1. Aproximaciones

La distribución normal es una de las más importantes en la Estadística, y es común que se utilice para aproximar otras distribuciones. Esto se debe a que la distribución normal es una de las más sencillas de trabajar. En esta sección estudiaremos algunas de estas aproximaciones.

Observación. Notemos que estas aproximaciones solo tienen sentido hoy en día histórico o docente, ya que actualmente se dispone de herramientas computacionales que permiten trabajar con cualquier distribución sin necesidad de aproximarla. En el pasado, no obstante, estas aproximaciones eran muy útiles al no existir dichas herramientas. Igual ocurre en el ámbito docente actualmente.

Aproximación de la Binomial por la Normal

Proposición 1.13. *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias tal que $X_n \sim B(n,p)$. Entonces,*

$$\lim_{n\to\infty} P[X_n \leqslant x] = P\left[X \leqslant x\right]$$

donde $X \sim \mathcal{N}(np, np(1-p))$.

La demostración de este teorema se podrá entender en el último tema, por ser una consecuencia del Corolario 6.8.2. No obstante, se presenta aquí por su utilidad en la práctica, ya que nos permite aproximar la distribución binomial por la normal.

Si n es suficientemente grande, se tiene que la función de distribución de X_n se puede aproximar por la de una distribución normal con parámetros $\mu = np$ y $\sigma^2 = np(1-p)$. Informalmente, esto lo notamos como:

$$X_n \sim \mathcal{N}(np, np(1-p))$$

Empíricamente se ha comprobado que esta aproximación es buena si $n \geqslant 30$ y $0{,}1 \leqslant p \leqslant 0{,}9$.

Aproximación de la Poisson por la Normal

Proposición 1.14. *Para cada $\lambda \in \mathbb{R}^+$, sea $X_\lambda \sim \mathcal{P}(\lambda)$. Entonces, se tiene que:*

$$\lim_{\lambda\to\infty} P[X_\lambda \leqslant x] = P\left[X \leqslant x\right]$$

donde $X \sim \mathcal{N}(\lambda, \lambda)$.

De igual forma, la demostración de este teorema se podrá entender en el último tema, por ser una consecuencia del Teorema 6.8. No obstante, se presenta aquí por su utilidad en la práctica, ya que nos permite aproximar la distribución de Poisson por la normal.

Si λ es suficientemente grande, se tiene que la función de distribución de X_λ se puede aproximar por la de una distribución normal con parámetros $\mu = \sigma^2 = \lambda$. Informalmente, esto lo notamos como:

$$X_\lambda \sim \mathcal{N}(\lambda, \lambda)$$

Empíricamente se ha comprobado que esta aproximación es buena si $\lambda \geqslant 10$.

Corrección por Continuidad

Notemos que en muchos casos, como en las dos aproximaciones anteriores, se trata de aproximar una variable aleatoria discreta por una continua. En estos casos, es posible caer en el siguiente error. Supongamos X una variable aleatoria discreta que sigue una distribución binomial, y sea x_i un valor de la variable aleatoria con $P[X = x_i] > 0$. Al aproximarlo por una normal, se tendría que $P[X = x_i] = 0$, ya que la normal es continua. Esto es incoherente, por lo que se introduce una *corrección por continuidad*.

Esta corrección por continuidad siempre se realiza sumando o restando 0.5 (este valor se ha establecido así porque, empíricamente, se ha comprobado que mejora la aproximación) a los extremos de la desigualdad (según convenga). Lo que buscaremos es cubrir los valores de la variable aleatoria discreta en un intervalo de la variable aleatoria continua. Veamos algunos ejemplos:

- Para aproximar $P[X = x_i]$ en la binomial, se calculará con la expresión dada por $P[x_i - 0{,}5 \leqslant X \leqslant x_i + 0{,}5]$ en la normal.
- Para aproximar $P[X \leqslant x_i]$ en la binomial, se calculará $P[X \leqslant x_i + 0{,}5]$ en la normal.
- Para aproximar $P[X \geqslant x_i]$ en la binomial, se calculará $P[X \geqslant x_i - 0{,}5]$ en la normal.

1.3. Distribución Exponencial

Definición 1.3 (Distribución Exponencial)**.** Una variable aleatoria continua X sigue una distribución exponencial con parámetro $\lambda \in \mathbb{R}^+$, si su función de densidad es:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

Lo denotaremos como $X \sim \exp(\lambda)$.

Comprobemos ahora que efectivamente es una función de densidad.

Demostración. La función de densidad debe cumplir las siguientes propiedades:

- $f_X(x) \geqslant 0$ para todo $x \in \mathbb{R}$.

  Esto es directo puesto que el término exponencial siempre es positivo.

- $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,dx = 1.$

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^{\infty} \lambda e^{-\lambda x}\,dx = \lambda \int_0^{\infty} e^{-\lambda x}\,dx = \left[-e^{-\lambda x}\right]_0^{\infty} = 1$$

□

Veamos algunas aplicaciones de esta distribución:

- La distribución exponencial se utiliza para modelar el tiempo aleatorio entre dos fallos consecutivos en *fiabilidad* o entre dos llegadas consecutivas en *teoría de colas.*

- También se aplica en la modelización de tiempos aleatorios de supervivencia (*Análisis de Supervivencia*).

- En general, X suele representar un tiempo aleatorio transcurrido entre dos sucesos, que se producen de forma aleatoria y consecutiva en el tiempo. Dichos sucesos se contabilizan mediante un proceso de Poisson homogéneo. El parámetro λ representa la razón de ocurrencia de dichos sucesos, que en este caso es constante.

Proposición 1.15. *Sea X una variable aleatoria tal que $X \sim \exp(\lambda)$. Entonces, su función de distribución es:*

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

Demostración. Distinguimos dos casos:

- Si $x < 0$, entonces $F_X(x) = 0$.

- Si $x \geqslant 0$, entonces:

$$F_X(x) = \int_0^{x} \lambda e^{-\lambda t}\,dt = \left[-e^{-\lambda t}\right]_0^{x} = \left(-e^{-\lambda x} + 1\right) = 1 - e^{-\lambda x}$$

□

Proposición 1.16. *Sea X una variable aleatoria tal que $X \sim \exp(\lambda)$. Su función generatriz de momentos es:*

$$M_X(t) = \frac{\lambda}{\lambda - t} \quad si\ t < \lambda$$

Demostración. Tenemos que:

$$M_X(t) = E\left[e^{tX}\right] = \int_0^{\infty} e^{tx}\lambda e^{-\lambda x}\ dx = \int_0^{\infty} \lambda e^{(t-\lambda)x}\ dx = \left[\frac{\lambda e^{(t-\lambda)x}}{t-\lambda}\right]_0^{\infty}$$

Para que la integral sea convergente, necesitamos que $t-\lambda<0$, es decir, $t<\lambda$. Tenemos entonces:

$$M_X(t) = \left[\frac{\lambda e^{(t-\lambda)x}}{t-\lambda}\right]_0^{\infty} = 0 - \frac{\lambda e^0}{t-\lambda} = \frac{\lambda}{\lambda - t}$$

□

Proposición 1.17. *Sea X una variable aleatoria de forma que $X \sim \exp(\lambda)$. Entonces, sus momentos no centrados de orden $k \in \mathbb{N}$ son:*

$$E[X^k] = \frac{k!}{\lambda^k}$$

Demostración. Demostramos por inducción sobre k que:

$$\frac{d^k}{dt^k}M_X(t) = \frac{k!\cdot\lambda}{(\lambda-t)^{k+1}}$$

- <u>Caso base</u>: $k=1$.

$$\frac{d}{dt}M_X(t) = \frac{\lambda}{(\lambda-t)^2} = \frac{1!\cdot\lambda}{(\lambda-t)^{1+1}}$$

- <u>Supuesto cierto para k, demostramos para $k+1$</u>:

$$\begin{aligned}\frac{d^{k+1}}{dt^{k+1}}M_X(t) &= \frac{d}{dt}\left(\frac{d^k}{dt^k}M_X(t)\right) = \frac{d}{dt}\left(\frac{k!\cdot\lambda}{(\lambda-t)^{k+1}}\right) = \frac{k!\cdot\lambda}{(\lambda-t)^{2k+2}}\cdot(k+1)(\lambda-t)^k = \\ &= \frac{(k+1)k!\cdot\lambda}{(\lambda-t)^{k+2}} = \frac{(k+1)!\cdot\lambda}{(\lambda-t)^{k+2}}\end{aligned}$$

Por tanto, una vez demostrado este resultado, tenemos que:

$$E[X^k] = \frac{d^k}{dt^k}M_X(t)\Big|_{t=0} = \frac{k!\cdot\lambda}{(\lambda-0)^{k+1}} = \frac{k!}{\lambda^k}$$

□

Como consecuencia, tenemos que:

$$E[X] = \frac{1}{\lambda} \qquad \text{Var}[X] = E[X^2] - E[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Proposición 1.18 (Falta de memoria)**.** *Sea X una variable aleatoria tal que $X \sim \exp(\lambda)$. Entonces, se cumple la propiedad de falta de memoria:*

$$P(X \geqslant t+s \mid X \geqslant s) = P(X \geqslant t) \qquad \forall t, s \in \mathbb{R}^+$$

Demostración. Tenemos que:

$$P(X \geqslant t+s \mid X \geqslant s) = \frac{P(X \geqslant t+s, X \geqslant s)}{P(X \geqslant s)} = \frac{P(X \geqslant t+s)}{P(X \geqslant s)} \overset{(*)}{=}$$
$$\overset{(*)}{=} \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} \overset{(*)}{=} P(X \geqslant t)$$

donde en $(*)$ hemos usado que:

$$P(X \geqslant x) = 1 - P(X < x) = 1 - P(X \leqslant x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$$

□

La gráfica de la función de densidad de la exponencial es decreciente y asintótica al eje de abscisas, como podemos ver en la Figura 1.2.

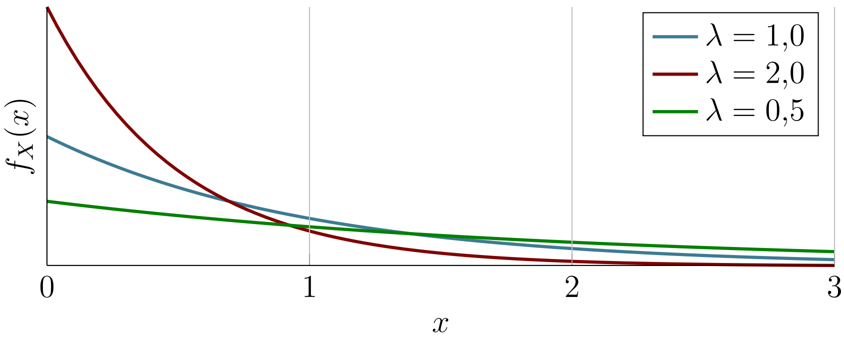


Figura 1.2: Función de densidad de una v. a. con distribución exponencial.

1.3.1. Relación con la Distribución Poisson

La distribución exponencial está estrechamente relacionada con la distribución de Poisson.

- Sea Y una variable aleatoria que indica el número de sucesos aleatorios que ocurren en un intervalo de tiempo de longitud t cuando la razón de ocurrencia de dichos sucesos es λ. Entonces, $Y \sim \mathcal{P}(\lambda t)$.

- Sea X una variable aleatoria que indica el tiempo que transcurre hasta que se produce el primer suceso aleatorio, o bien el tiempo que transcurre entre dos sucesos consecutivos, cuando la razón de ocurrencia de dichos sucesos es λ. Entonces, $X \sim \exp(\lambda)$.

Su relación es la siguiente:

$$P[Y = 0] = e^{-\lambda t} = P[X \geqslant t] = 1 - P[X < t] = 1 - 1 + e^{-\lambda t} = e^{-\lambda t}$$

Esto es coherente, ya que la probabilidad de que no se produzca ningún suceso en un intervalo de tiempo de longitud t ($P[Y = 0]$) es la misma que la probabilidad de que el tiempo que transcurra hasta que se produzca el primer suceso sea mayor que t ($P[X \geqslant t]$).

1.4. Distribución de Erlang

Definición 1.4 (Distribución de Erlang)**.** Una variable aleatoria continua X sigue una distribución de Erlang con parámetros $n \in \mathbb{N}$ y $\lambda \in \mathbb{R}^+$, si modela el tiempo que transcurre hasta que se producen n sucesos aleatorios consecutivos, cuando la razón de ocurrencia de dichos sucesos es λ. Lo denotaremos como $X \sim \mathcal{E}(n, \lambda)$.

Observación. La distribución de Erlang es una generalización de la distribución exponencial. En efecto, si $n = 1$, entonces la distribución de Erlang se reduce a la distribución exponencial.

Proposición 1.19. *Sea X una variable aleatoria tal que $X \sim \mathcal{E}(n, \lambda)$. Entonces, su función de densidad es:*

$$f_X(x) = \begin{cases} \dfrac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

donde $\Gamma(n) = (n-1)!$.

Demostración. La función de densidad debe cumplir las siguientes propiedades:

- $f_X(x) \geqslant 0$ para todo $x \in \mathbb{R}$.

  Esto es directo puesto que el término exponencial siempre es positivo.

- $\displaystyle\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

  $$\int_{-\infty}^{\infty} f_X(x)\, dx = \int_0^{\infty} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}\, dx = \frac{\lambda^n}{\Gamma(n)} \int_0^{\infty} x^{n-1} e^{-\lambda x}\, dx$$

  Para calcular la última integral, realizamos inducción sobre n para demostrar que:

  $$\int_0^{\infty} x^{n-1} e^{-\lambda x}\, dx = \frac{(n-1)!}{\lambda^n} \qquad \forall n \in \mathbb{N}$$

  - Caso base: $n = 1$.

    $$\int_0^{\infty} e^{-\lambda x}\, dx = \left[-\frac{e^{-\lambda x}}{\lambda}\right]_0^{\infty} = \frac{1}{\lambda}$$

  - Supuesto cierto para n, demostramos para $n+1$:

    $$\int_0^{\infty} x^n e^{-\lambda x}\, dx = \left[\begin{array}{cc} u(x) = x^n & v'(x) = e^{-\lambda x} \\ u'(x) = nx^{n-1} & v(x) = -\dfrac{e^{-\lambda x}}{\lambda} \end{array}\right] =$$
    $$= \left[-\frac{x^n e^{-\lambda x}}{\lambda}\right]_0^{\infty} + \frac{n}{\lambda} \int_0^{\infty} x^{n-1} e^{-\lambda x}\, dx \overset{(*)}{=} \frac{n}{\lambda} \frac{(n-1)!}{\lambda^n} = \frac{n!}{\lambda^{n+1}}$$

    donde en $(*)$ hemos usado la hipótesis de inducción.

Por tanto, tenemos que:

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \frac{\lambda^n}{\Gamma(n)} \cdot \frac{(n-1)!}{\lambda^n} = 1$$

$\square$

Proposición 1.20. *Sea X una variable aleatoria tal que $X \sim \mathcal{E}(n, \lambda)$. Entonces, su función generatriz de momentos es:*

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^n \qquad si\ t < \lambda$$

Demostración. Tenemos que:

$$M_X(t) = E\left[e^{tX}\right] = \int_0^{\infty} e^{tx} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}\ dx = \frac{\lambda^n}{\Gamma(n)} \int_0^{\infty} x^{n-1} e^{(t-\lambda)x}\ dx$$

Mediante una inducción análoga a la realizada en la demostración anterior, podemos demostrar (asumiendo que $t < \lambda$) que:

$$\int_0^{\infty} x^{n-1} e^{(t-\lambda)x}\ dx = \frac{\Gamma(n)}{(\lambda - t)^n} \qquad \forall n \in \mathbb{N}$$

Por tanto, tenemos que:

$$M_X(t) = \frac{\lambda^n}{\Gamma(n)} \cdot \frac{\Gamma(n)}{(\lambda - t)^n} = \left(\frac{\lambda}{\lambda - t}\right)^n$$

$\square$

1.5. Distribución Gamma

1.5.1. Función Gamma

Previamente al estudio de la distribución Gamma, vamos a estudiar la función Gamma, que es la función que da nombre a la distribución. Esta es:

$$\begin{array}{rccc} \Gamma : & \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\ & x & \longmapsto & \displaystyle\int_0^{\infty} t^{x-1} e^{-t}\ dt \end{array}$$

Algunas propiedades son:

1. $\Gamma(1) = 1$.

$$\Gamma(1) = \int_0^{\infty} e^{-t}\ dt = \left[-e^{-t}\right]_0^{\infty} = 1$$

2. $\Gamma(x+1) = x\Gamma(x), \qquad \forall x \in \mathbb{R}^+$.

Mediante integración por partes, tenemos que:

$$\Gamma(x+1) = \int_0^{\infty} t^x e^{-t}\ dt = \left[\begin{array}{cc} u(t) = t^x & v'(t) = e^{-t} \\ u'(t) = xt^{x-1} & v(t) = -e^{-t} \end{array}\right] =$$
$$= \left[-t^x e^{-t}\right]_0^{\infty} + x\int_0^{\infty} t^{x-1} e^{-t}\ dt = x\Gamma(x)$$

3. $\Gamma(n) = (n-1)!, \qquad \forall n \in \mathbb{N}$.

   Se deduce directamente de las dos propiedades anteriores.

4. Si $\lambda \in \mathbb{R}^+$, entonces $\displaystyle\int_0^\infty t^{x-1}e^{-\lambda t}\,dt = \frac{\Gamma(x)}{\lambda^x}$.

   Hacemos el cambio de variable $t = {}^u/_\lambda$:

$$\int_0^\infty t^{x-1}e^{-\lambda t}\,dt = \left[\begin{array}{c} t = {}^u/_\lambda \\ \frac{dt}{du} = {}^1/_\lambda \end{array}\right] = \int_0^\infty \left(\frac{u}{\lambda}\right)^{x-1} e^{-u}\,\frac{du}{\lambda} =$$
$$= \frac{1}{\lambda^x}\int_0^\infty u^{x-1}e^{-u}\,du = \frac{\Gamma(x)}{\lambda^x}$$

5. $\Gamma\left({}^1/_2\right) = \sqrt{\pi}$.

$$\Gamma\left({}^1/_2\right) = \int_0^\infty t^{-1/2}e^{-t}\,dt = \left[\begin{array}{c} t = {}^{u^2}/_2 \\ {}^{dt}/_{du} = u \end{array}\right] = \int_0^\infty \sqrt{2}\cdot u^{-1}\cdot e^{-u^2/2}\cdot u\,du =$$
$$= \sqrt{2}\int_0^\infty e^{-u^2/2}\,du$$

   Como la función $x \mapsto e^{-x^2}$ es par, tenemos que:

$$\Gamma\left({}^1/_2\right) = \sqrt{2}\int_0^\infty e^{-u^2/2}\,du = \frac{\sqrt{2}}{2}\int_{-\infty}^\infty e^{-u^2/2}\,du = \frac{1}{\sqrt{2}}\int_{-\infty}^\infty e^{-u^2/2}\,du =$$
$$= \frac{\sqrt{\pi}}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-u^2/2}\,du = \sqrt{\pi}\cdot\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,du \overset{(*)}{=} \sqrt{\pi}$$

   donde en $(*)$ hemos usado que el integrando es la función de densidad de una distribución $\mathcal{N}(0,1)$.

1.5.2. Distribución Gamma

Definición 1.5 (Distribución Gamma)**.** Una variable aleatoria continua X sigue una distribución Gamma con parámetros $u, \lambda \in \mathbb{R}^+$, si su función de densidad es:

$$f_X(x) = \begin{cases} \dfrac{\lambda^u}{\Gamma(u)}x^{u-1}e^{-\lambda x} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

Lo denotaremos como $X \sim \Gamma(u,\lambda)$.

- El parámetro u se conoce como *parámetro de forma.*
- El parámetro λ se conoce como *parámetro de escala.*

Observación. La distribución Gamma es una generalización de la distribución de Erlang. En efecto, si $u \in \mathbb{N}$, entonces la distribución Gamma se reduce a la distribución de Erlang.

Proposición 1.21. *Sea X una variable aleatoria tal que $X \sim \Gamma(u,\lambda)$. Entonces, su función de densidad cumple las condiciones de una función de densidad.*

Demostración. La función de densidad debe cumplir las siguientes propiedades:

- $f_X(x) \geqslant 0$ para todo $x \in \mathbb{R}$.

  Esto es directo puesto que el término exponencial siempre es positivo.

- $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,dx = 1.$

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^{\infty} \frac{\lambda^u}{\Gamma(u)} x^{u-1} e^{-\lambda x}\ dx = \frac{\lambda^u}{\Gamma(u)} \int_0^{\infty} x^{u-1} e^{-\lambda x}\ dx \overset{(*)}{=}$$
$$\overset{(*)}{=} \frac{\lambda^u}{\Gamma(u)} \cdot \frac{\Gamma(u)}{\lambda^u} = 1$$

donde en $(*)$ hemos usado la propiedad 4 de la función Gamma.

□

Proposición 1.22. *Sea X una variable aleatoria tal que $X \sim \Gamma(u, \lambda)$. Entonces, su función generatriz de momentos es:*

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^u \qquad \text{si } t < \lambda$$

Demostración. Tenemos que:

$$M_X(t) = E\left[e^{tX}\right] = \int_0^{\infty} e^{tx} \frac{\lambda^u}{\Gamma(u)} x^{u-1} e^{-\lambda x}\ dx = \frac{\lambda^u}{\Gamma(u)} \int_0^{\infty} x^{u-1} e^{(t-\lambda)x}\ dx$$

Usando de nuevo la propiedad 4 de la función Gamma, como $\lambda - t > 0$, tenemos:

$$M_X(t) = \frac{\lambda^u}{\Gamma(u)} \cdot \frac{\Gamma(u)}{(\lambda - t)^u} = \left(\frac{\lambda}{\lambda - t}\right)^u$$

□

Proposición 1.23. *Sea X una variable aleatoria tal que $X \sim \Gamma(u, \lambda)$. Entonces, sus momentos no centrados de orden $k \in \mathbb{N}$ son:*

$$E[X^k] = \frac{\Gamma(u + k)}{\lambda^k\ \Gamma(u)}$$

Demostración. Hay dos maneras de demostrar este resultado:

Opción 1 Usar la función generatriz de momentos.

Demostramos por inducción sobre k que:

$$\frac{d^k}{dt^k} M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^u \cdot \frac{\prod\limits_{i=0}^{k-1} (u + i)}{(\lambda - t)^k} \qquad \forall k \in \mathbb{N}$$

- Caso base: $k = 1$.

$$\frac{d}{dt} M_X(t) = u \left(\frac{\lambda}{\lambda - t}\right)^{u-1} \cdot \frac{\lambda}{(\lambda - t)^2} = \left(\frac{\lambda}{\lambda - t}\right)^u \cdot \frac{u}{\lambda - t}$$

- Supuesto cierto para k, demostramos para $k+1$:

$$\frac{d^{k+1}}{dt^{k+1}} M_X(t) = \frac{d}{dt}\left(\frac{d^k}{dt^k} M_X(t)\right) = \frac{d}{dt}\left(\left(\frac{\lambda}{\lambda - t}\right)^u \cdot \frac{\prod_{i=0}^{k-1}(u+i)}{(\lambda - t)^k}\right) =$$

$$= \left(\frac{\lambda}{\lambda - t}\right)^u \cdot \frac{u}{\lambda - t} \cdot \frac{\prod_{i=0}^{k-1}(u+i)}{(\lambda - t)^k} + \left(\frac{\lambda}{\lambda - t}\right)^u \cdot \frac{\prod_{i=0}^{k-1}(u+i)}{(\lambda - t)^{k+1}} \cdot k =$$

$$= \left(\frac{\lambda}{\lambda - t}\right)^u \cdot \frac{\prod_{i=0}^{k}(u+i)}{(\lambda - t)^{k+1}}$$

Por tanto, una vez demostrado este resultado, tenemos que:

$$E[X^k] = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0} = \left(\frac{\lambda}{\lambda - 0}\right)^u \cdot \frac{\prod_{i=0}^{k-1}(u+i)}{(\lambda - 0)^k} = \frac{\prod_{i=0}^{k-1}(u+i)}{\lambda^k} = \frac{\Gamma(u+k)}{\lambda^k\ \Gamma(u)}$$

Opción 2 Usar la definición de los momentos no centrados.

$$E[X^k] = \int_0^\infty x^k \frac{\lambda^u}{\Gamma(u)} x^{u-1} e^{-\lambda x}\ dx = \frac{\lambda^u}{\Gamma(u)} \int_0^\infty x^{u+k-1} e^{-\lambda x}\ dx \overset{(*)}{=}$$
$$\overset{(*)}{=} \frac{\Gamma(u+k)\lambda^u}{\lambda^{u+k}\ \Gamma(u)} = \frac{\Gamma(u+k)}{\lambda^k\ \Gamma(u)}$$

donde en $(*)$ hemos usado la propiedad 4 de la función Gamma.

$\square$

Como consecuencia, tenemos que:

$$E[X] = \frac{u}{\lambda} \qquad \text{Var}[X] = E[X^2] - E[X]^2 = \frac{u(u+1)}{\lambda^2} - \frac{u^2}{\lambda^2} = \frac{u}{\lambda^2}$$

Tiene muchas aplicaciones en experimentos o fenómenos aleatorios que tienen asociadas variables aleatorias que siempre son no negativas y cuyas distribuciones son sesgadas a la derecha.

1.6. Distribución Beta

Previamente al estudio de la distribución Beta, vamos a introducir la función Beta, que es la función que da nombre a la distribución.

1.6.1. Función Beta

Definición 1.6 (Función Beta)**.** La función Beta es una función definida como:

$$\begin{array}{rccl} \beta: & \mathbb{R}^+ \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\ & (p,q) & \longmapsto & \displaystyle\int_0^1 t^{p-1}(1-t)^{q-1}\ dt \end{array}$$

Algunas propiedades son:

1. Es simétrica. Es decir, $\forall p, q \in \mathbb{R}^+$, se cumple que $\beta(p,q) = \beta(q,p)$.

$$\begin{aligned} \beta(p,q) &= \int_0^1 t^{p-1}(1-t)^{q-1}\ dt = \left[\begin{array}{c} t = 1-u \\ dt = -du \end{array}\right] = -\int_1^0 (1-u)^{p-1}u^{q-1}\ du = \\ &= \int_0^1 (1-u)^{p-1}u^{q-1}\ du = \beta(q,p) \end{aligned}$$

2. $\beta(p,q) = \dfrac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$.

   Esta demostración no se incluye por ser requerir de integración en varias variables, siendo por tanto de mayor complejidad.

1.6.2. Distribución Beta

Definición 1.7 (Distribución Beta)**.** Una variable aleatoria continua X sigue una distribución Beta con parámetros $p, q \in \mathbb{R}^+$, si su función de densidad es:

$$f_X(x) = \begin{cases} \dfrac{1}{\beta(p,q)} x^{p-1}(1-x)^{q-1} & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases}$$

Lo denotaremos como $X \sim \beta(p,q)$.

Comprobemos que la función de densidad cumple las condiciones de una función de densidad.

Demostración. La función de densidad debe cumplir las siguientes propiedades:

- $f_X(x) \geqslant 0$ para todo $x \in \mathbb{R}$.

  Esto es directo puesto que los términos x^{p-1}, $(1-x)^{q-1}$ y $\beta(p,q)$ siempre son positivos.

- $\displaystyle\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x)\, dx &= \int_0^1 \frac{1}{\beta(p,q)} x^{p-1}(1-x)^{q-1}\ dx = \frac{1}{\beta(p,q)} \int_0^1 x^{p-1}(1-x)^{q-1}\ dx = \\ &= \frac{1}{\beta(p,q)} \cdot \beta(p,q) = 1 \end{aligned}$$

□

Proposición 1.24 (Simetría)**.** *Sea X una variable aleatoria tal que $X \sim \beta(p,q)$. Entonces, $1-X \sim \beta(q,p)$.*

Demostración. Calculemos la función de densidad de $Y = 1-X$ usando el Teorema de Cambio de Variable:

$$P[Y=y] = P[X=1-y] = \frac{1}{\beta(p,q)}(1-y)^{p-1}y^{q-1} = \frac{1}{\beta(q,p)}y^{q-1}(1-y)^{p-1}$$

Tenemos por tanto que $1-X \sim \beta(q,p)$. □

Proposición 1.25. *Sea X una variable aleatoria tal que $X \sim \beta(p,q)$. Entonces, sus momentos no centrados de orden $k \in \mathbb{N}$ son:*

$$E[X^k] = \frac{\beta(p+k,q)}{\beta(p,q)}$$

Demostración. Usamos la propiedad de la función Beta que relaciona la función Beta con la función Gamma:

$$\begin{aligned} E[X^k] &= \int_0^1 x^k \frac{1}{\beta(p,q)} x^{p-1}(1-x)^{q-1}\,dx = \frac{1}{\beta(p,q)}\int_0^1 x^{p+k-1}(1-x)^{q-1}\,dx = \\ &= \frac{\beta(p+k,q)}{\beta(p,q)} \end{aligned}$$

□

Como consecuencia, tenemos que:

$$E[X] = \frac{\beta(p+1,q)}{\beta(p,q)} = \frac{\Gamma(p+1)\Gamma(q)}{\Gamma(p+q+1)}\cdot\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} = \frac{\Gamma(p+1)}{\Gamma(p)}\cdot\frac{\Gamma(p+q)}{\Gamma(p+q+1)} = \frac{p}{p+q}$$

Para la varianza tenemos el siguiente resultado.

Proposición 1.26. *Sea X una variable aleatoria tal que $X \sim \beta(p,q)$. Entonces, su varianza es:*

$$\operatorname{Var}[X] = \frac{pq}{(p+q)^2(p+q+1)}$$

Demostración. Tenemos que:

$$\begin{aligned} E[X^2] &= \frac{\beta(p+2,q)}{\beta(p,q)} = \frac{\Gamma(p+2)\Gamma(q)}{\Gamma(p+q+2)}\cdot\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} = \frac{\Gamma(p+2)}{\Gamma(p)}\cdot\frac{\Gamma(p+q)}{\Gamma(p+q+2)} = \frac{p(p+1)}{(p+q)(p+q+1)} \\ \operatorname{Var}[X] &= E[X^2]-E[X]^2 = \frac{p(p+1)}{(p+q)(p+q+1)} - \frac{p^2}{(p+q)^2} = \frac{p(p+1)(p+q)-p^2(p+q+1)}{(p+q)^2(p+q+1)} = \\ &= \frac{p(p+q)\left[p+1-p\right]-p^2}{(p+q)^2(p+q+1)} = \frac{pq}{(p+q)^2(p+q+1)} \end{aligned}$$

□

Respecto a la forma que toma la función de densidad de la distribución Beta, podemos ver que toma formas muy variadas en función de los valores de los parámetros p y q. Esto es muy útil para modelar situaciones muy diversas. Tenemos los siguientes ejemplos:

- Si $p = q$, entonces la función de densidad es simétrica respecto a $x = 1/2$.
- Si $p = q = 1$, entonces $X \sim \mathcal{U}(0, 1)$.
- Si $p < q$, entonces la función de densidad es asimétrica a la derecha, y viceversa.
- Si $p < 1$ y $q \geqslant 1$, es decreciente y cóncava, mientras que si $p \geqslant 1$ y $q < 1$, es creciente y convexa.
- Si $p, q > 1$, entonces tiene solo un máximo.
- Si $p, q < 1$, entonces tiene solo un mínimo.

Por tanto, como hemos descrito, puede tomar formas muy diversas.

2. Vectores Aleatorios

Hasta ahora, hemos estudiado variable aleatoria unidimensional. En este capítulo, vamos a estudiar variables aleatorias multidimensionales, es decir, vectores aleatorios. Para ello, al igual que como hicimos con las variables aleatorias unidimensionales, hemos de definir en primer lugar la $\sigma-$álgebra de Borel en $\mathbb{R}^n$.

Definición 2.1. Sea $\mathbb{R}^n$ el espacio euclídeo de dimensión n. La $\sigma-$álgebra de Borel en $\mathbb{R}^n$, notada por $\mathcal{B}^n$, es la $\sigma-$álgebra generada por todos los intervalos de $\mathbb{R}^n$.

En particular, en Análisis Matemático II vimos que esta $\sigma-$álgebra está formada por los intervalos:

$$]-\infty, x] :=]-\infty, x_1] \times \cdots \times]-\infty, x_n], \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Además, en los presentes apuntes, usaremos la relación parcial de orden en $\mathbb{R}^n$ siguiente.

Notación. Dado $x, y \in \mathbb{R}^n$, notaremos:

$$x \leqslant y \iff x_i \leqslant y_i, \quad \forall i = 1, \ldots, n.$$

Esta es parcial, ya que no podemos comparar ciertos elementos, como $(1, 2)$ y $(2, 1)$.

Gráficamente, dados $x, x' \in \mathbb{R}^2$ tenemos que $x \leqslant x'$ si y solo si x está a la izquierda y por debajo de x'.

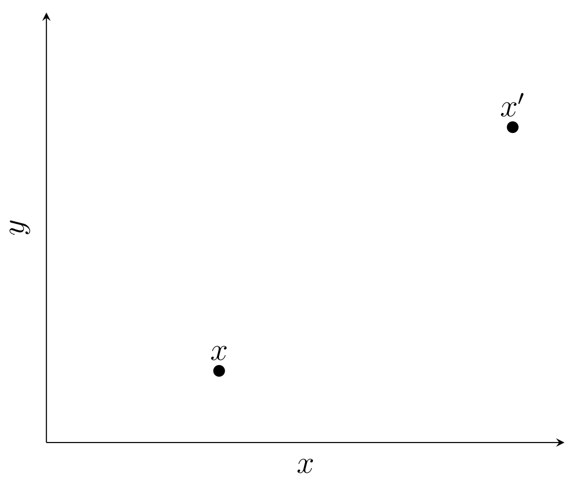


Figura 2.1: Relación parcial de orden en $\mathbb{R}^2$, donde $x \leqslant x'$.

Veamos ahora el equivalente a variable aleatoria en el caso multidimensional.

Definición 2.2 (Vector aleatorio)**.** Un vector aleatorio $X = (X_1, \dots, X_n)$ de un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ se define como una función medible:

$$X \ : \ (\Omega, \mathcal{A}, P) \to (\mathbb{R}^n, \mathcal{B}^n)$$

tal que se cumple que:

$$X^{-1}(B) \in \mathcal{A}, \quad \forall B \in \mathcal{B}^n.$$

Es decir:

$$X^{-1}(]-\infty, x]) = \{\omega \in \Omega \mid X_1(\omega) \leqslant x_1, \dots, X_n(\omega) \leqslant x_n\} \in \mathcal{A} \quad \forall x \in \mathbb{R}^n.$$

Además, considerando cada una de las componentes por separado, como cada componente de una función medible es medible, se tiene la siguiente caracterización de forma directa.

Teorema 2.1. *Sea* $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^n, \mathcal{B}^n)$*. Entonces:*

$$X \text{ es un vector aleatorio} \iff X_i \text{ es una variable aleatoria } \forall i = 1, \dots, n.$$

Introducimos ahora la distribución de probabilidad de un vector aleatorio, que será la función de densidad (o función masa de probabilidad) en el caso unidimensional.

Definición 2.3 (Distribución de probabilidad)**.** Sea X un vector aleatorio. La distribución de probabilidad de X es la medida de probabilidad en $\mathbb{R}^n$ definida por:

$$\begin{array}{rccl} P_X : & \mathcal{B}^n & \longrightarrow & [0,1] \\ & B & \longmapsto & P_X(B) = P(X^{-1}(B)), \quad \forall B \in \mathcal{B}^n. \end{array}$$

Notación. Al igual que en el caso unidimensional, dado $B \in \mathcal{B}^n$, tenemos:

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}$$

Por tanto, denotaremos $P_X(B)$ por $P[X \in B]$.

El gran uso que tiene P_X es que induce una medida de probabilidad en $\mathbb{R}^n$, y por tanto obtenemos el espacio probabilístico $(\mathbb{R}^n, \mathcal{B}^n, P_X)$. Podremos por tanto usar todas las propiedades vistas en EDIP sobre las probabilidades.

Proposición 2.2. *Sea X un vector aleatorio. Entonces, la distribución P_X es una medida de probabilidad sobre* $(\mathbb{R}^n, \mathcal{B}^n)$*.*

Demostración. Veamos que se cumplen las tres propiedades de la Axiomática de Kolmogorov:

1. <u>No negatividad</u>: $P_X(B) = P(X^{-1}(B)) \geqslant 0, \quad \forall B \in \mathcal{B}^n$ ya que P es una medida de probabilidad.

2. <u>Suceso seguro</u>: $P_X(\mathbb{R}^n) = P(X^{-1}(\mathbb{R}^n)) = P(\Omega) = 1$.

3. <u>$\sigma-$aditividad</u>: Sean $B_1, B_2, \ldots \in \mathcal{B}^n$ disjuntos dos a dos. Entonces, como $X^{-1}(B_1), X^{-1}(B_2), \ldots$ son disjuntos dos a dos, se tiene:

$$P_X\left(\bigcup_{i=1}^{\infty} B_i\right) = P\left(X^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right)\right) = P\left(\bigcup_{i=1}^{\infty} X^{-1}(B_i)\right) \stackrel{(*)}{=}$$
$$\stackrel{(*)}{=} \sum_{i=1}^{\infty} P(X^{-1}(B_i)) = \sum_{i=1}^{\infty} P_X(B_i).$$

donde en $(*)$ hemos usado la propiedad de $\sigma-$aditividad de la medida de probabilidad P. □

Así, tenemos que todo vector aleatorio X transforma el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ en el espacio de probabilidad $(\mathbb{R}^n, \mathcal{B}^n, P_X)$.

Al igual que en el caso unidimensional, definimos la función de distribución de un vector aleatorio a partir de la distribución de probabilidad.

Definición 2.4 (Función de distribución)**.** Sea X un vector aleatorio. La función de distribución de X es la función:

$$\begin{array}{rccl} F_X : & \mathbb{R}^n & \longrightarrow & [0,1] \\ & x & \longmapsto & F_X(x) = P_X(]-\infty, x]) \end{array}$$

Si $X = (X_1, \ldots, X_n)$, entonces denotaremos:

$$F_X(x) = P[X \leqslant x] = P[X_1 \leqslant x_1, \ldots, X_n \leqslant x_n].$$

Algunas de las propiedades que la función de distribución cumple son:

1. Es monótona no decreciente en cada una de sus componentes. Es decir, se cumple que $\forall\ i = 1, \ldots, n$ y $\forall\ x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \in \mathbb{R}$:

$$x_i \leqslant x_i' \Longrightarrow F_X(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) \leqslant F_X(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n).$$

Demostración. Como $x_i \leqslant x_i'$, tenemos que $]-\infty, x_i] \subset]-\infty, x_i']$. Por tanto, en $\mathbb{R}^n$ tenemos que:

$$]-\infty, x_1] \times \cdots \times]-\infty, x_{i-1}] \times]-\infty, x_i] \times]-\infty, x_{i+1}] \times \cdots \times]-\infty, x_n] \subset$$
$$\subset]-\infty, x_1] \times \cdots \times]-\infty, x_{i-1}] \times]-\infty, x_i'] \times]-\infty, x_{i+1}] \times \cdots \times]-\infty, x_n].$$

Por la monotonía de P_X por ser una medida de probabilidad, tenemos que:

$$P_X\left(]-\infty, x_1] \times \cdots \times]-\infty, x_{i-1}] \times]-\infty, x_i] \times]-\infty, x_{i+1}] \times \cdots \times]-\infty, x_n]\right) \leqslant$$
$$\leqslant P_X\left(]-\infty, x_1] \times \cdots \times]-\infty, x_{i-1}] \times]-\infty, x_i'] \times]-\infty, x_{i+1}] \times \cdots \times]-\infty, x_n]\right).$$

Por tanto, tenemos que:

$$F_X(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) \leqslant F_X(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n).$$

□

2. Es continua por la derecha en cada una de sus componentes. Es decir, $\forall\ i = 1, \dots, n$ y $\forall\ x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n \in \mathbb{R}$ se tiene que:

$$\forall x_i \in \mathbb{R}, \qquad F_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) = \lim_{x \to x_i^+} F_X(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n).$$

3. $\forall i = 1, \dots, n$ y $\forall x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n \in \mathbb{R}$ se tiene que:

$$\lim_{x \to -\infty} F_X(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) = 0.$$

4. Tenemos que:

$$\lim_{\substack{x_i \to +\infty \\ i=1,\dots,n}} F_X(x_1, \dots, x_n) = 1.$$

5. $\forall x_1, \dots, x_n \in \mathbb{R}$ y $\forall \varepsilon_1, \dots, \varepsilon_n \in \mathbb{R}^+$ se tiene que:

$$\begin{aligned}
&F_X(x_1 + \varepsilon_1, \dots, x_n + \varepsilon_n) - \\
&\quad - \sum_{i=1}^{n} F_X(x_1 + \varepsilon_1, \dots, x_{i-1} + \varepsilon_{i-1}, x_i, x_{i+1} + \varepsilon_{i+1}, \dots, x_n + \varepsilon_n) + \\
&\quad + \sum_{\substack{i,j=1 \\ i<j}}^{n} F_X(x_1 + \varepsilon_1, \dots, x_{i-1} + \varepsilon_{i-1}, x_i, x_{i+1} + \varepsilon_{i+1}, \dots \\
&\qquad\qquad \dots, x_{j-1} + \varepsilon_{j-1}, x_j, x_{j+1} + \varepsilon_{j+1}, \dots, x_n + \varepsilon_n) + \\
&\quad + \dots + (-1)^n F_X(x_1, \dots, x_n) \geqslant 0
\end{aligned}$$

Estas propiedades caracterizan la función de distribución de los vectores aleatorios. Es decir, dada una función que cumple estas propiedades, es la función de distribución de un vector aleatorio.

Veamos ahora una interpretación para la última propiedad, que puede ser bastante más compleja. Para el caso de dos variables, tenemos que fórmula queda:

$$F_X(x_1 + \varepsilon_1, x_2 + \varepsilon_2) - F_X(x_1, x_2 + \varepsilon_2) - F_X(x_1 + \varepsilon_1, x_2) + F_X(x_1, x_2) \geqslant 0.$$

Esa fórmula calcula el valor de la función de distribución en el siguiente rectángulo:

$$[x_1, x_1 + \varepsilon_1] \times [x_2, x_2 + \varepsilon_2].$$

Esto tiene sentido, ya que buscamos calcular probabilidades en subconjuntos de $\mathbb{R}^n$, y dichas probabilidades han de ser no negativas. De forma general, la fórmula calcula la probabilidad del siguiente rectángulo:

$$\prod_{i=1}^{n} [x_i, x_i + \varepsilon_i].$$

Al igual que ocurría con variables aleatorias unidimensionales, puesto que P_X es una medida de probabilidad, podemos calcular de forma sencilla la probabilidad de intervalos bidimensionales.

- $P[a < X_1 \leqslant b, X_2 \in I] = P[X_1 \leqslant b, X_2 \in I] - P[X_1 \leqslant a, X_2 \in I]$.
- $P[a < X_1 < b, X_2 \in I] = P[X_1 < b, X_2 \in I] - P[X_1 \leqslant a, X_2 \in I]$.
- $P[X_1 \leqslant b, c < X_2 \leqslant d] = P[X_1 \leqslant b, X_2 \leqslant d] - P[X_1 \leqslant b, X_2 \leqslant c]$.
- $P[X_1 \leqslant b, c \leqslant X_2 < d] = P[X_1 \leqslant b, X_2 < d] - P[X_1 \leqslant b, X_2 < c]$.

2.1. Clasificación de vectores aleatorios

Al igual que en el caso unidimensional, podemos clasificar los vectores aleatorios en discretos y continuos. Esto se hace en función de la naturaleza de los valores que toma el vector aleatorio.

Definición 2.5 (Recorrido de un vector aleatorio)**.** Sea $X = (X_1, \dots, X_n)$ un vector aleatorio. El recorrido de X es el conjunto de valores que toma el vector aleatorio:

$$E_X = \{x \in \mathbb{R}^n \mid \exists \omega \in \Omega \text{ tal que } X(\omega) = x\} = Img(X).$$

Usando el recorrido de una variable aleatoria unidimensional, tenemos que:

$$E_X \subset \prod_{i=1}^{n} E_{X_i}.$$

De forma directa, tenemos este resultado.

Proposición 2.3. *Sea X un vector aleatorio sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces:*

$$P[X \in E_X] = 1.$$

Demostración. Tenemos que:

$$P[X \in E_X] = P[X^{-1}(E_X)] = P[\Omega] = 1.$$

□

2.1.1. Vectores aleatorios discretos

Definición 2.6. Un vector aleatorio $X = (X_1, \dots, X_n)$ es discreto si $\exists A \subset \mathbb{R}^n$ numerable tal que

$$P[X \in A] = 1.$$

Veamos ahora la siguiente caracterización de vectores aleatorios discretos.

Teorema 2.4. *Un vector aleatorio $X = (X_1, \dots, X_n)$ es discreto si y solo si cada una de sus componentes X_i es discreta.*

Demostración. Demostramos por doble implicación.

$\Longrightarrow$) Supongamos que X es discreto con valores en E_X. Entonces, Para cada valor $i = 1, \dots, n$, consideramos la proyección de X en la componente i:

$$E_X^i = \{x_i \in siysolosi\mathbb{R} \mid \exists x \in E_X \text{ tal que } (x)_i = x_i\}.$$

Evidentemente, E_X^i es numerable por serlo E_X. Veamos ahora que se tiene $P[X_i \in E_X^i] = 1$. Tenemos la siguiente inclusión:

$$E_X \subset \mathbb{R} \times \cdots \times \mathbb{R} \times E_X^i \times \mathbb{R} \times \cdots \times \mathbb{R}.$$

Por tanto, como la probabilidad es una función creciente, tenemos que:

$$1 = P[X \in E_X] \leqslant P[X_1 \in \mathbb{R}, \dots, X_{i-1} \in \mathbb{R}, X_i \in E_X^i, X_{i+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}] = \\ = P[X_i \in E_X^i].$$

Por tanto, como toda probabilidad es menor o igual que 1, se tiene la igualdad $P[X_i \in E_X^i] = 1$, teniendo que X_i es discreta.

$\Longleftarrow$) Supongamos que cada una de las componentes de X es discreta. Es decir, para cada $i = 1, \dots, n$, tenemos que $\exists E_{X_i} \subset \mathbb{R}$ numerable tal que $P[X_i \in E_{X_i}] = 1$.

Consideramos ahora el conjunto $E_{X_1} \times \cdots \times E_{X_n} \subset \mathbb{R}^n$ numerable, y tenemos que:

$$P[X \in E_{X_1} \times \cdots \times E_{X_n}] = P\left[X^{-1}(E_{X_1} \times \cdots \times E_{X_n})\right] = P\left[\bigcap_{i=1}^n X_i^{-1}(E_{X_i})\right] \geqslant \\ \geqslant 1 - \sum_{i=1}^n P[X_i \notin E_{X_i}] = 1.$$

donde hemos usado la desigualdad de Boole y el lema anterior. Por tanto, como las probabilidades son menores o iguales que 1, se tiene el resultado de $P[X \in E_{X_1} \times \cdots \times E_{X_n}] = 1$, por lo que X es discreto.

□

Como en el caso de variables unidimensionales, los vectores de tipo discreto se manejan a partir de su función masa de probabilidad, y el tratamiento de este tipo de vectores es totalmente análogo al de las variables discretas.

Definición 2.7. Sea $X = (X_1, \dots, X_n)$ un vector aleatorio discreto. La función masa de probabilidad de X es la función:

$$\begin{array}{rccl} p_X: & E_X & \longrightarrow & [0,1] \\ & x & \longmapsto & p_X(x) = P[X = x] \end{array}$$

Esta satisface:

1. $p_X(x) \geqslant 0$.
2. $\sum\limits_{x \in E_X} p_X(x) = 1$.

La función de distribución de un vector aleatorio discreto se define por tanto como:

$$F_X(x) = P[X \leqslant x] = \sum_{\substack{t \in E_X \\ t \leqslant x}} P[X = t]$$

Ejemplo. Sea el experimento aleatorio de lanzar un dado, y sean las siguientes variables aleatorias:

$$X_1 = \begin{cases} -1 & \text{si sale impar} \\ 1 & \text{si sale par} \end{cases}$$

$$X_2 = \begin{cases} -2 & \text{si sale 1, 2, 3} \\ 0 & \text{si sale 4} \\ 3 & \text{si sale 5, 6} \end{cases}$$

Considerado el vector aleatorio $X = (X_1, X_2)$, se pide:

1. Calcular la función masa de probabilidad de X.

   Tenemos que:

   $$E_{X_1} = \{-1, 1\},$$
   $$E_{X_2} = \{-2, 0, 3\},$$

   Por tanto,

   $$E_X = E_{X_1} \times E_{X_2} = \{(-1,-2), (-1,0), (-1,3), (1,-2), (1,0), (1,3)\}.$$

   Tenemos por tanto que:

   - $P[X = (-1,-2)] = P[\text{sale impar y 1, 2, 3}] = {}^2\!/\!_6$.
   - $P[X = (-1,0)] = P[\text{sale impar y 4}] = {}^0\!/\!_6 = 0$.
   - $P[X = (-1,3)] = P[\text{sale impar y 5, 6}] = {}^1\!/\!_6$.
   - $P[X = (1,-2)] = P[\text{sale par y 1, 2, 3}] = {}^1\!/\!_6$.
   - $P[X = (1,0)] = P[\text{sale par y 4}] = \frac{1}{6}$.
   - $P[X = (1,3)] = P[\text{sale par y 5, 6}] = {}^1\!/\!_6$.

   Podemos resumir esta información como

$X_1 \backslash X_2$	-2	0	3
-1	${}^2\!/\!_6$	0	${}^1\!/\!_6$
1	${}^1\!/\!_6$	${}^1\!/\!_6$	${}^1\!/\!_6$

2. Calcular la función de distribución de X.

   Tenemos que:

   $$F_X(x_1, x_2) = \begin{cases} 0 & x_1 < -1 \text{ o } x_2 < -2 \\ {}^2\!/\!_6 & x_1 \in [-1, 1[\,, x_2 \in [-2, 3[\\ {}^3\!/\!_6 & x_1 \in [-1, 1]\,, x_2 \geqslant 3 \\ {}^3\!/\!_6 & x_1 \geqslant 1, x_2 \in [-2, 0[\\ {}^4\!/\!_6 & x_1 \geqslant 1, x_2 \in [0, 3[\\ 1 & x_1 \geqslant 1, x_2 \geqslant 3 \end{cases}$$

2.1.2. Vectores aleatorios continuos

Definición 2.8. Un vector aleatorio $X = (X_1, \ldots, X_n)$ es continuo si existe una función integrable no negativa $f_X : \mathbb{R}^n \to \mathbb{R}$ tal que su función de distribución es:

$$F_X(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(t_1, \ldots, t_n) dt_n \cdots dt_1, \quad \forall x_1, \ldots, x_n \in \mathbb{R}.$$

A la función f_X se le llama función de densidad de probabilidad de X.

Además, si f_X es continua en un punto $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, entonces aplicando el Teorema Fundamental del Cálculo en cada una de las coordenadas obtenemos que la función de distribución F_X es derivable en ese punto y se tiene que:

$$\frac{\partial^n F_X}{\partial x_1 \cdots \partial x_n}(x) = f_X(x).$$

Esta función f_X, por definición, cumple las siguientes propiedades:

1. $f_X(x) \geqslant 0$.
2. Es integrable en $\mathbb{R}^n$.
3. $\int_{\mathbb{R}^n} f_X(x)\ dx = 1$.

$$\int_{-\infty}^{+\infty} f_X(t)\ dt = \lim_{x \to +\infty} \int_{-\infty}^{x} f_X(t) = \lim_{x \to +\infty} F_X(x) \overset{(*)}{=} 1.$$

donde en $(*)$ hemos usado una de las propiedades de la función de distribución.

La función de densidad determina la función de distribución de un vector aleatorio continuo, y por tanto su distribución de probabilidad.

$$P_X(B) = P[X \in B] = \int_B f_X(x)\ dx, \quad \forall B \in \mathcal{B}^n.$$

Debido a lo estudiado en Análisis Matemático II, sabemos que si $E \subset \mathbb{R}^n$ es un conjunto numerable, entonces $P[X \in E] = 0$.

Al igual que en el caso de vectores discretos, tenemos el siguiente resultado (aunque en este caso no es una caracterización, ya que la implicación restante no será cierta en general).

Teorema 2.5. *Sea vector aleatorio $X = (X_1, \ldots, X_n)$ continuo. Entonces, cada una de sus componentes X_i es continua.*

Demostración. Fijemos $i \in \{1, \ldots, n\}$, y sea F_{X_i} la función de distribución de X_i. Entonces, tenemos que:

$$\begin{aligned}
F_{X_i}(x_i) &= P[X_i \leqslant x_i] = P[X_1 \in \mathbb{R}, \ldots, X_{i-1} \in \mathbb{R}, X_i \leqslant x_i, X_{i+1} \in \mathbb{R}, \ldots, X_n \in \mathbb{R}] = \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \int_{-\infty}^{x_i} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_X(t_1, \ldots, t_n) dt_n \cdots dt_{i+1} dt_i dt_{i-1} \cdots dt_1. = \\
&= \int_{-\infty}^{x_i} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_X(t_1, \ldots, t_n) dt_n \cdots dt_{i+1} dt_{i-1} \cdots dt_1 dt_i
\end{aligned}$$

Definimos por tanto:

$$f_{X_i}(x_i) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_X(t_1, \ldots, t_n) dt_n \cdots dt_{i+1} dt_{i-1} \cdots dt_1.$$

Esta función es integrable y no negativa, y cumple que:

$$F_{X_i}(x_i) = \int_{-\infty}^{x_i} f_{X_i}(t_i)\ dt_i.$$

Por tanto, X_i es una variable aleatoria continua cuya función de densidad es f_{X_i}. □

2.2. Distribuciones marginales

Dado un vector aleatorio $X = (X_1, \ldots, X_n)$, podemos estudiar las distribuciones de sus componentes por separado. Estas se conocen como distribuciones marginales.

Definición 2.9. Sea $X = (X_1, \ldots, X_n)$ un vector aleatorio. La distribución marginal de X_i es la distribución de probabilidad de la variable aleatoria X_i. A la distribución de probabilidad de X se le llama distribución conjunta de X.

Veamos cómo obtener la función de distribución de una variable aleatoria a partir de la función de distribución de un vector aleatorio.

Proposición 2.6. *Sea $X = (X_1, \ldots, X_n)$ un vector aleatorio con función de distribución F_X. Entonces, la función de distribución de X_i es:*

$$F_{X_i}(x_i) = \lim_{\substack{t_j \to +\infty, \\ j \neq i}} F_X(t_1, \ldots, t_{i-1}, x_i, t_{i+1}, \ldots, t_n).$$

Demostración. Esta propiedad se deduce de la continuidad de las funciones de probabilidad (Axiomática de Kolmogorov) y del hecho de que:

$$\mathbb{R} = \bigcup_{t \in \mathbb{R}^+}]-\infty, t].$$

Tenemos que:

$$\begin{aligned}
F_{X_i}(x_i) &= P[X_i \leqslant x_i] = P[X_1 \in \mathbb{R}, \ldots, X_{i-1} \in \mathbb{R}, X_i \leqslant x_i, X_{i+1} \in \mathbb{R}, \ldots, X_n \in \mathbb{R}] \overset{(*)}{=} \\
&\overset{(*)}{=} \lim_{\substack{t_j \to +\infty, \\ j \neq i}} P_X\left(]-\infty, t_1] \times \ldots \times]-\infty, t_{i-1}] \times]-\infty, x_i] \times]-\infty, t_{i+1}] \times \ldots \times]-\infty, t_n]\right) = \\
&= \lim_{\substack{t_j \to +\infty, \\ j \neq i}} F_X(t_1, \ldots, t_{i-1}, x_i, t_{i+1}, \ldots, t_n).
\end{aligned}$$

donde en $(*)$ hemos usado la propiedad de continuidad de la función de probabilidad. □

Caso discreto

En el caso de vectores aleatorios discretos, la función de masa de probabilidad de una variable aleatoria se obtiene a partir de la función de masa de probabilidad del vector aleatorio.

Proposición 2.7. *Sea $X = (X_1, \ldots, X_n)$ un vector aleatorio discreto con función de masa de probabilidad p_X. Entonces, la función de masa de probabilidad de X_i es:*

$$p_{X_i}(x_i) = \sum_{\substack{t=(t_1,\ldots,t_n)\in E_X \\ t_i = x_i}} p_X(t).$$

Caso continuo

En el caso de vectores aleatorios continuos, la función de densidad de probabilidad de una variable aleatoria se obtiene a partir de la función de densidad de probabilidad del vector aleatorio.

Proposición 2.8. *Sea $X = (X_1, \ldots, X_n)$ un vector aleatorio continuo con función de densidad de probabilidad f_X. Entonces, la función de densidad de probabilidad de X_i es:*

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{n-1}} f_X(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)\ dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

Notemos que esta demostración se ha hecho en el Teorema 2.5.

2.3. Distribuciones condicionadas

Dado un vector aleatorio $X = (X_1, \ldots, X_n)$, podemos estudiar la distribución de una de sus componentes condicionada a que otra de sus componentes tome un valor concreto.

Caso discreto

Definición 2.10. Sea $X = (X_1, \ldots, X_n)$ un vector aleatorio discreto, X_i una de sus componentes y $x_i^* \in \mathbb{R}$ tal que $P[X_i = x_i^*] > 0$. La distribución condicionada de $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ a $X_i = x_i^*$ es la distribución con función de masa de probabilidad:

$$\begin{gathered}
P[X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \ldots, X_n = x_n \mid X_i = x_i^*] = \\
= \frac{P[X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_i = x_i^*, X_{i+1} = x_{i+1}, \ldots, X_n = x_n]}{P[X_i = x_i^*]}, \\
\forall (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \mid (x_1, \ldots, x_{i-1}, x_i^*, x_{i+1}, \ldots, x_n) \in E_X.
\end{gathered}$$

Comprobemos que esta definición efectivamente es una función masa de probabilidad.

Demostración.

1. En primer lugar, toma valores no negativos, puesto que es el cociente de dos valores no negativos.

2. Veamos ahora que suma 1. Para ello, por la Proposición 2.7, tenemos que la siguiente sumatoria es la distribución marginal de X_i:

$$\sum_{\substack{(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)\mid\\(x_1,\dots,x_{i-1},x_i^*,x_{i+1},\dots,x_n)\in E_X}} P[X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i = x_i^*, X_{i+1} = x_{i+1}, \dots, X_n = x_n] = P[X_i = x_i^*].$$

   Por tanto, tenemos que:

$$\sum_{\substack{(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)\mid\\(x_1,\dots,x_{i-1},x_i^*,x_{i+1},\dots,x_n)\in E_X}} P[X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \dots, X_n = x_n \mid X_i = x_i^*] = 1$$

□

Caso continuo

Definición 2.11. Sea $X = (X_1, \dots, X_n)$ un vector aleatorio continuo, X_i una de sus componentes y $x_i^* \in \mathbb{R}$ tal que $f_{X_i}(x_i^*) > 0$. La distribución condicionada de $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ a $X_i = x_i^*$ es la distribución con función de densidad de probabilidad:

$$f_{X_1,\dots,X_{i-1},X_{i+1},\dots,X_n|X_i=x_i^*}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \frac{f_X(x_1, \dots, x_{i-1}, x_i^*, x_{i+1}, \dots, x_n)}{f_{X_i}(x_i^*)}.$$

Comprobemos que efectivamente dicha función se trata de una función de densidad.

Demostración.

1. $f_{X_1,\dots,X_{i-1},X_{i+1},\dots,X_n|X_i=x_i^*}$ es no negativa, por ser cociente de funciones de densidad.

2. Es integrable (se deja como ejercicio al lector).

3. Veamos que integra 1:

$$\int_{\mathbb{R}^{n-1}} \frac{f_X(x_1, \dots, x_{i-1}, x_i^*, x_{i+1}, \dots, x_n)}{f_{X_i}(x_i^*)} = \frac{\int_{\mathbb{R}^{n-1}} f_X(x_1, \dots, x_{i-1}, x_i^*, x_{i+1}, \dots, x_n)}{f_{X_i}(x_i^*)} \overset{(*)}{=}$$
$$\overset{(*)}{=} \frac{f_{X_i}(x_i^*)}{f_{X_i}(x_i^*)} = 1$$

   donde en $(*)$ hemos usado la Proposición 2.8, ya que es la función de densidad de la marginal de X_i.

□

2.4. Cambio de Variable

Al igual que ocurría en el caso unidimensional, tenemos un Teorema de Canvio de Variable. Dado un vector aleatorio $X = (X_1, \dots, X_n)$, podemos estudiar la distribución de otro vector aleatorio $Y = (Y_1, \dots, Y_m)$ que depende de X mediante una transformación.

Proposición 2.9. *Sea X un vector aleatorio $n-$dimensional, y una función g dada por $g : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^m, \mathcal{B}^m)$ una función medible. Entonces, $Y = g(X)$ es un vector aleatorio $m-$dimensional.*

Observación. Al igual que especificamos en EDIP, $Y = g(X)$ siendo formales representa:

$$Y = h \circ X$$

Demostración. Como la composición de funciones medibles es medible, tenemos que $Y = g(X)$ es medible por serlo X y g. Además, tenemos que:

$$Y^{-1}(B) = X^{-1}(g^{-1}(B)) \in \mathcal{A}, \quad \forall B \in \mathcal{B}^m.$$

por ser X un vector aleatorio y tener que $g^{-1}(B) \in \mathcal{B}^n$. □

Veamos ahora cómo se transforma la distribución de probabilidad de un vector aleatorio al aplicarle una transformación.

Proposición 2.10. *Sea $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^n, \mathcal{B}^n)$ un vector aleatorio $n-$dimensional y $g : \mathbb{R}^n \to \mathbb{R}^m$ una función medible. Entonces, la distribución de probabilidad de $Y = g(X)$ es:*

$$P_Y(B) = P_X(g^{-1}(B)), \quad \forall B \in \mathcal{B}^m.$$

Demostración. Como $Y = g \circ X$, tenemos que $Y^{-1} = X^{-1} \circ g^{-1}$. Por tanto, tenemos que:

$$P_Y(B) = P[Y^{-1}(B)] = P[X^{-1}(g^{-1}(B))] = P_X(g^{-1}(B)).$$

□

Como resultado inmediato, tenemos que la función de distribución de Y es:

$$F_Y(y) = P_Y(]-\infty, y]) = P_X(g^{-1}(]-\infty, y])) \qquad \forall y \in \mathbb{R}^m.$$

Veamos ahora algunos casos particulares de transformaciones de vectores aleatorios.

2.4.1. Discreto a Discreto

Como veremos en el siguiente teorema, al aplicarle una transformación a un vector aleatorio discreto, podemos afirmar que llegaremos a otro vector aleatorio discreto. Esto nos permite contemplar todos los casos, algo que no podremos hacer en el caso continuo.

Teorema 2.11. *Sea $X = (X_1, \dots, X_n)$ un vector aleatorio discreto con valores en E_X, y sea $g : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^m, \mathcal{B}^m)$ una función medible. Entonces, $Y = g(X)$ es un vector aleatorio discreto con valores en $E_Y = g(E_X)$, y su función masa de probabilidad es:*

$$P[Y = y] = \sum_{x \in E_X \cap g^{-1}(y)} P[X = x], \qquad \forall y \in E_Y.$$

Demostración. La demostración es totalmente análoga al caso unidimensional, disponible en el Temario de EDIP. □

Ejemplo. Sea $X = (X_1, X_2)$ un vector aleatorio con función de masa de probabilidad:

$X_2 \backslash X_1$	1	0	-1
-2	$1/6$	$1/12$	$1/6$
1	$1/6$	$1/12$	$1/6$
2	$1/12$	0	$1/12$

Calcular la función de masa de probabilidad de $Y = (|X_1|, X_2^2)$.

Tenemos que $E_Y = g(E_X)$, que es:

$$E_Y = \{(0, 1), (0, 4), (1, 1), (1, 4)\}.$$

Por tanto, tenemos que:

$$\begin{aligned}
P[Y = (0, 1)] &= P[X_1 = 0, X_2 = 1] = 1/12,\\
P[Y = (0, 4)] &= P[X_1 = 0, X_2 = 2] + P[X_1 = 0, X_2 = -2] = 0 + 1/12 = 1/12,\\
P[Y = (1, 1)] &= P[X_1 = 1, X_2 = 1] + P[X_1 = -1, X_2 = 1] = 1/6 + 1/6 = 1/3,\\
P[Y = (1, 4)] &= P[X_1 = 1, X_2 = 2] + P[X_1 = 1, X_2 = -2] + P[X_1 = -1, X_2 = 2]+\\
&\quad + P[X_1 = -1, X_2 = -2] = 1/12 + 1/6 + 1/12 + 1/6 = 1/2.
\end{aligned}$$

Por tanto, la función de masa de probabilidad de Y es:

$Y_2 \backslash Y_1$	0	1
1	$1/12$	$1/3$
4	$1/12$	$1/2$

2.4.2. Continuo a Discreto

En el caso de vectores aleatorios continuos, tras aplicarle una transformación no podemos asegurar que obtengamos un vector aleatorio continuo; ya que pueden suceder distintas combinaciones. Un caso particular es que obtengamos un vector aleatorio discreto, caso que contempla el siguiente teorema.

Teorema 2.12. *Sea $X = (X_1, \dots, X_n)$ un vector aleatorio continuo con función de densidad f_X, y sea $g : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^m, \mathcal{B}^m)$ una función medible tal que $Y = g(X)$ es un vector aleatorio discreto con valores en $E_Y \subset \mathbb{R}^m$. Su función masa de probabilidad se obtiene a partir de f_X como:*

$$P[Y = y] = \int_{g^{-1}(y)} f_X(x)\, dx, \qquad \forall y \in E_Y.$$

Demostración. La demostración es totalmente análoga al caso unidimensional, disponible en el Temario de EDIP. □

Ejemplo. Sea $X = (X_1, X_2)$ un vector aleatorio con función de densidad, para $\mu, \lambda \in \mathbb{R}^+$:

$$f_X(x_1, x_2) = \begin{cases} \lambda\mu e^{-\lambda x_1 - \mu x_2} & \text{si } x_1 > 0, x_2 > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

Calcular la función de masa de probabilidad de:

$$Y = \begin{cases} 1 & \text{si } X_1 > X_2 \\ 0 & \text{si } X_1 \leqslant X_2 \end{cases}$$

Tenemos que $E_Y = \{0, 1\}$, y calculamos cada una de las probabilidades:

$$\begin{aligned} P[Y=1] = P[X_1 > X_2] &= \int_0^\infty \int_{x_2}^\infty \lambda\mu e^{-\lambda x_1 - \mu x_2}\, dx_1 dx_2 = \int_0^\infty \mu e^{-\mu x_2} \int_{x_2}^\infty \lambda e^{-\lambda x_1}\, dx_1 dx_2 = \\ &= \int_0^\infty \mu e^{-\mu x_2} \left[-e^{-\lambda x_1}\right]_{x_2}^\infty dx_2 = \int_0^\infty \mu e^{-\mu x_2} e^{-\lambda x_2} dx_2 = \int_0^\infty \mu e^{-(\mu+\lambda)x_2} dx_2 = \\ &= \left[-\frac{\mu}{\mu+\lambda} e^{-(\mu+\lambda)x_2}\right]_0^\infty = \frac{\mu}{\mu+\lambda}. \end{aligned}$$

$$P[Y=0] = 1 - P[Y=1] = 1 - \frac{\mu}{\mu+\lambda} = \frac{\lambda}{\mu+\lambda}.$$

Por tanto, la función de masa de probabilidad de Y es:

$$P[Y=0] = \frac{\lambda}{\mu+\lambda}, \quad P[Y=1] = \frac{\mu}{\mu+\lambda}.$$

2.4.3. Continuo a Continuo

Otro caso especialmente relevante es que, tras aplicar la transformación, obtengamos un vector aleatorio continuo.

Teorema 2.13. *Sea $X = (X_1, \ldots, X_n)$ un vector aleatorio continuo con función de densidad f_X y con valores en $E_X \subset \mathbb{R}^n$. Sea $g = (g_1, \ldots, g_n) : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^n, \mathcal{B}^n)$ una función medible tal que:*

- $\exists g^{-1} = (g_1^*, \ldots, g_n^*)$.
- *La inversa es derivable en todas sus componentes:*

$$\exists \frac{\partial g_i^*}{\partial y_j}(y_1, \ldots, y_n), \quad \forall i, j \in \{1, \ldots, n\}, \forall y = (y_1, \ldots, y_n) \in g(E_X).$$

- *El jacobiano de la inversa no es nulo:*

$$\det(Jg^{-1}(y)) \neq 0 \quad \forall y \in g(E_X).$$

En estas condiciones, la transformación $Y = g(X)$ es un vector aleatorio de tipo continuo, y su función de densidad puede obtenerse a partir de f_X mediante la siguiente relación:

$$f_Y(y) = f_X(g^{-1}(y)) \left|\det(Jg^{-1}(y))\right|, \quad \forall y \in g(E_X).$$

2.4.4. Distribución del Máximo y del Mínimo

Definición 2.12. Dadas n funciones f_1, f_2, ... f_n definidas sobre un mismo conjunto A y con imagen en un mismo conjunto B, definimos las funciones

$$\text{mín}\{f_1, f_2, \dots, f_n\}, \text{máx}\{f_1, f_2, \dots, f_n\} : A \to B$$

dadas por, para cada $x \in A$:

$$\begin{aligned}\text{mín}\{f_1, f_2, \dots, f_n\}(x) &= \text{mín}\{f_1(x), f_2(x), \dots, f_n(x)\}\\ \text{máx}\{f_1, f_2, \dots, f_n\}(x) &= \text{máx}\{f_1(x), f_2(x), \dots, f_n(x)\}\end{aligned}$$

Dada una función $F : A^n \to B^m$ de n variables, podemos definir $\text{mín}\, F$ (análogamente $\text{máx}\, F$) como la función:

$$\text{mín}\, F = \text{mín}\{F_1, F_2, \dots, F_m\}$$

siendo $F_1, F_2, \dots, F_m$ las componentes de la función F.

Proposición 2.14. *Sea $g : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^m, \mathcal{B}^m)$ una función medible, entonces $\text{mín}\, g$ y $\text{máx}\, g$ son funciones medibles.*

Tenemos que dos cambios de variable frecuentes son los dados por las funciones máximo y mínimo, que sabemos que son medibles. Sea por tanto $X = (X_1, \dots, X_n)$ un vector aleatorio con función de distribución F_X, y buscamos hallar la función de distribución de $Z = \text{mín}\, X$ y $W = \text{máx}\, X$. Dado $x \in \mathbb{R}$, tenemos que:

- $\underline{Z = \text{mín} = \text{mín}\, X}$: $Z = \text{mín}\, X = \text{mín}\{X_1, \dots, X_n\}$.

  Dado $\omega \in \Omega$, tenemos que:

  $$Z(\omega) > x \Longleftrightarrow \text{mín}\{X_1(\omega), \dots, X_n(\omega)\} > x \Longleftrightarrow X_1(\omega) > x, \dots, X_n(\omega) > x.$$

  Por tanto, tenemos que:

  $$P[\text{mín} \leqslant x] = 1 - P[\text{mín} > x] = 1 - P[X_1 > x, \dots, X_n > x] \qquad \forall x \in \mathbb{R}.$$

- $\underline{W = \text{máx} = \text{máx}\, X}$: $W = \text{máx}\, X = \text{máx}\{X_1, \dots, X_n\}$.

  Dado $\omega \in \Omega$, tenemos que:

  $$W(\omega) < x \Longleftrightarrow \text{máx}\{X_1(\omega), \dots, X_n(\omega)\} < x \Longleftrightarrow X_1(\omega) < x, \dots, X_n(\omega) < x.$$

  Por tanto, tenemos que:

  $$P[\text{máx} \leqslant x] = P[X_1 \leqslant x, \dots, X_n \leqslant x] = P[X \leqslant (x, \dots, x)] \qquad \forall x \in \mathbb{R}.$$

Dado ahora $X = (X_1, \dots, X_n)$ un vector aleatorio, consideramos ahora la distribución conjunta:

$$(\text{máx}\, X, \text{mín}\, X)$$

Busquemos su función de distribución. Dados $(x, y) \in \mathbb{R}^2$, tenemos que:

$$P[(\text{máx}\, X, \text{mín}\, X) \leqslant (x, y)] = P[\text{máx}\, X \leqslant x, \text{mín}\, X \leqslant y]$$

- Si $x \leqslant y$, como siempre se tiene que $\min X \leqslant \max X$ y en este caso buscamos que $\max X \leqslant x$, tenemos que $\min X \leqslant \max X \leqslant x \leqslant y$, luego:

$$P[(\max X, \min X) \leqslant (x,y)] = P[\max X \leqslant x, \min X \leqslant y] = P[\max X \leqslant x] = F_X(x,\dots,x)$$

- Si $x > y$, entonces:

$$\begin{aligned} P[(\max X, \min X) \leqslant (x,y)] &= P[\max X \leqslant x, \min X \leqslant y] = \\ &= P[\max X \leqslant x] - P[\max X \leqslant x, \min X > y] = \\ &= F_X(x,\dots,x) - P[y < X_1 \leqslant x, \dots, y < X_n \leqslant x] \end{aligned}$$

Por tanto, la función de distribución de $(\max X, \min X)$ es:

$$F_{(\max X, \min X)}(x,y) = \begin{cases} F_X(x,\dots,x) & \text{si } x \leqslant y, \\ F_X(x,\dots,x) - P[y < X_1 \leqslant x, \dots, y < X_n \leqslant x] & \text{si } x > y. \end{cases}$$

2.5. Esperanza

Al igual que venimos haciendo en este tema, generalizamos el concepto de esperanza a vectores aleatorios.

Definición 2.13 (Esperanza)**.** Sea $X = (X_1, \dots, X_n)$ un vector aleatorio. Entonces:

- $\exists E[X] \iff \exists E[X_i] \quad \forall i \in \{1,\dots,n\}$.
- En tal caso, la esperanza de X es:

$$E[X] := (E[X_1], \dots, E[X_n]).$$

Al igual que ocurría en el caso unidimensional, tenemos el siguiente Teorema.

Teorema 2.15. *Sea $X = (X_1, \dots, X_n)$ un vector aleatorio y $g : \mathbb{R}^n \to \mathbb{R}$ una función medible. Entonces, suponiendo que dichas esperanzas existen:*

- *Si X es de tipo discreto con valores en E_X:*

$$E[g(X)] = \sum_{x \in E_X} g(x) P[X = x]$$

- *Si X es de tipo continuo con función de densidad f_X:*

$$E[g(X)] = \int_{\mathbb{R}^n} g(x) f_X(x) dx$$

Introducimos además el siguiente resultado, que no pudimos demostrar en la asignatura de EDIP.

Proposición 2.16 (Linealidad de la esperanza)**.** *Sean X e Y dos variables aleatorias definidas sobre el mismo espacio de probabilidad. Entonces, se verifica:*

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

Demostración. Distinguimos en función de si X e Y son discretas o continuas.

- Caso discreto: Supongamos que X e Y son discretas con valores en E_X y E_Y respectivamente. Entonces, tenemos que:

$$\begin{aligned}
E[X+Y] &= \sum_{(x,y)\in E_X\times E_Y} (x+y)P[X=x,Y=y] = \\
&= \sum_{x\in E_X}\sum_{y\in E_Y} xP[X=x,Y=y] + \sum_{x\in E_X}\sum_{y\in E_Y} yP[X=x,Y=y] = \\
&= \sum_{x\in E_X} x \sum_{y\in E_Y} P[X=x,Y=y] + \sum_{y\in E_Y} y \sum_{x\in E_X} P[X=x,Y=y] = \\
&= \sum_{x\in E_X} xP[X=x] + \sum_{y\in E_Y} yP[Y=y] = \\
&= E[X]+E[Y].
\end{aligned}$$

- Caso continuo: Supongamos que X e Y son continuas con funciones de densidad f_X y f_Y respectivamente. Entonces, tenemos que:

$$\begin{aligned}
E[X+Y] &= \int_{\mathbb{R}^2} (x+y)f_{X,Y}(x,y)\ dxdy = \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} (x+y)f_{X,Y}(x,y)\ dxdy = \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} xf_{X,Y}(x,y)\ dxdy + \int_{\mathbb{R}}\int_{\mathbb{R}} yf_{X,Y}(x,y)\ dxdy = \\
&= \int_{\mathbb{R}} x \int_{\mathbb{R}} f_{X,Y}(x,y)\ dydx + \int_{\mathbb{R}} y \int_{\mathbb{R}} f_{X,Y}(x,y)\ dxdy = \\
&= \int_{\mathbb{R}} xf_X(x)\ dx + \int_{\mathbb{R}} yf_Y(y)\ dy = \\
&= E[X]+E[Y].
\end{aligned}$$

□

Este resultado, ya conocido por el lector por usarse en la asignatura de EDIP, tiene importantes corolarios que recomendamos al lector recordar consultando dichos apuntes.

2.6. Momentos

2.6.1. Momentos No Centrados

Definición 2.14 (Momento no centrado)**.** Sea $X=(X_1,\dots,X_n)$ un vector aleatorio y $(k_1,\dots,k_n)\in(\mathbb{N}\cup\{0\})^n$. Entonces, el momento no centrado de orden $(k_1,\dots,k_n)$ de X es:

$$m_{k_1,\dots,k_n} := E[X_1^{k_1}\cdots X_n^{k_n}].$$

A estos momentos también se les llama momentos centrados en el origen.

Es importante notar que, fijado $i\in\{1,\dots,n\}$, si $k_j=0$ para todo $j\neq i$, entonces se obtienen los momentos de la variable aleatoria X_i. Estos se conocen como momentos marginales.

2.6.2. Momentos Centrados

Definición 2.15 (Momento centrado)**.** Sea $X = (X_1, \dots, X_n)$ un vector aleatorio y $(k_1, \dots, k_n) \in (\mathbb{N} \cup \{0\})^n$. Entonces, el momento centrado de orden $(k_1, \dots, k_n)$ de X es:

$$\mu_{k_1,\dots,k_n} := E[(X_1 - E[X_1])^{k_1} \cdots (X_n - E[X_n])^{k_n}].$$

A estos momentos también se les llama momentos centrados en la media.

De nuevo, es importante notar que, fijado $i \in \{1, \dots, n\}$, si $k_j = 0$ para todo $j \neq i$, entonces se obtienen los momentos centrados de la variable aleatoria X_i. Estos se conocen como momentos marginales centrados.

Al igual que ocurrió en Estadística Descriptiva Bidimensional, introducimos el concepto de covarianza:

Definición 2.16 (Covarianza)**.** Sea $X = (X_1, \dots, X_n)$ un vector aleatorio. Entonces, la covarianza de X_i y X_j, con $i \neq j$, es:

$$\mathrm{Cov}(X_i, X_j) = \mu_{0\dots0\overset{i)}{1}0\dots0\overset{j)}{1}0\dots0} = E[(X_i - E[X_i])(X_j - E[X_j])].$$

El siguiente lema técnico nos será de utilidad para el cálculo de la covarianza.

Lema 2.17. *Sean X, Y dos variables aleatorias. Entonces:*

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

Demostración. Tenemos que:

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] = E[XY - XE[Y] - E[X]Y + E[X]E[Y]] = \\
&= E[XY] - E[XE[Y]] - E[E[X]Y] + E[E[X]E[Y]] = \\
&= E[XY] - 2E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y].
\end{aligned}$$

donde hemos usado que $E[X], E[Y]$ son constantes. □

Como consecuencias inmediatas de este lema tenemos que, dadas X e Y dos variables aleatorias unidimensionales:

- $\mathrm{Cov}(X, X) = E[X^2] - E[X]^2 = \mathrm{Var}(X)$.
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.

De la definición de la covarianza, se deduce que esta mide la covariación conjunta de dos variables:

- Si es positiva nos dará la información de que a valores altos de una de las variable hay una mayor tendencia a encontrar valores altos de la otra variable y a valores bajos de una de las variable, correspondientemente se esperan valores bajos.
- Si la covarianza es negativa, la covariación de ambas variables será en sentido inverso: a valores altos le corresponderán bajos, y a valores bajos, altos.

- Si la covarianza es cero no hay una covariación clara en ninguno de los dos sentidos.

Este estudio se denomina correlación, y se entenderá a fondo en Regresión Multivariante, cuando se defina el coeficiente de correlación de Pearson.

Definición 2.17 (Matriz de Covarianzas)**.** Sea $X = (X_1, \ldots, X_n)$ un vector aleatorio. Entonces, la matriz de covarianzas de X, notada por Cov_X, es la matriz cuadrada de orden n cuyos elementos son las covarianzas de las variables aleatorias:

$$\mathrm{Cov}_X = (\mathrm{Cov}(X_i, X_j))_{i,j}$$

Como consecuencia de las dos propiedades anteriores, tenemos que Cov_X es simétrica, y sus elementos de la diagonal son las varianzas de las variables aleatorias. En el caso de que X sea un vector aleatorio bidimensional, la matriz de covarianzas se reduce a:

$$\mathrm{Cov}_X = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) \end{pmatrix}.$$

2.7. Función Generatriz de Momentos

Definición 2.18 (Función generatriz de momentos)**.** Sea $X = (X_1, \ldots, X_n)$ un vector aleatorio. Entonces, si $\exists E[e^{t_1X_1+\cdots+t_nX_n}]$ para todo $(t_1, \ldots, t_n) \in N$, siendo $N \subset \mathbb{R}^n$ un entorno del origen, la función generatriz de momentos de X es:

$$\begin{array}{rrcl} M_X: & N & \longrightarrow & \mathbb{R} \\ & (t_1, \ldots, t_n) & \longmapsto & E[e^{t_1X_1+\cdots+t_nX_n}] \end{array}$$

.

Notemos que, usando el producto escalar en $\mathbb{R}^n$ o la transpuesta de una matriz, podemos escribir la función generatriz de momentos como:

$$\begin{array}{rrcl} M_X: & N & \longrightarrow & \mathbb{R} \\ & t & \longmapsto & E[e^{\langle t, X\rangle}] = E[e^{tX^t}] \end{array}$$

.

Esta función, si existe, tiene propiedades análogas al caso unidimensional.

Teorema 2.18 (Unicidad)**.** *Si existe la función generatriz de momentos de un vector aleatorio, entonces determina la distribución de probabilidad de dicho vector aleatorio de forma unívoca.*

La relación con los momentos viene descrita en el siguiente teorema.

Teorema 2.19. *Sea $X = (X_1, \ldots, X_n)$ un vector aleatorio. Entonces, si existe la función generatriz de momentos de X:*

1. $\exists E[X_1^{k_1} \cdots X_n^{k_n}]$ *para todo* $(k_1, \ldots, k_n) \in (\mathbb{N} \cup \{0\})^n$.
2. M_X *es derivable y se tiene que:*

$$\left.\frac{\partial^{k_1+\cdots+k_n} M_X(t_1, \ldots, t_n)}{\partial t_1^{k_1} \cdots \partial t_n^{k_n}}\right|_{t_1=\cdots=t_n=0} = E[X_1^{k_1} \cdots X_n^{k_n}].$$

Por último, introducimos el concepto de función generatriz de momentos marginal. Dado $X = (X_1, \dots, X_n)$ un vector aleatorio con función generatriz de momentos M_X definida en N, existe la función generatriz de momentos de cada subvector de X $(X_{i_1}, \dots, X_{i_k})$. Esta se calcula como sigue:

$$\begin{aligned} M_{X_{i_1},\dots,X_{i_k}}(t_{i_1}, \dots, t_{i_k}) &= M_X(0, \dots, 0, t_{i_1}, 0, \dots, 0, t_{i_k}, 0, \dots, 0) \\ &\qquad \forall (t_{i_1}, \dots, t_{i_k}) \mid (0, \dots, 0, t_{i_1}, 0, \dots, 0, t_{i_k}, 0, \dots, 0) \in N. \end{aligned}$$

Observación. Notemos que esta notación no es del todo precisa, pero se emplea para no complicar la notación. La función generatriz de momentos de un subvector de X se obtiene evaluando la función generatriz de momentos de X en el punto cuyas componentes son todas nulas, excepto las correspondientes a las variables del subvector.

Esto es de demostración inmediata, ya que:

$$\begin{aligned} M_{X_{i_1},\dots,X_{i_k}}(t_{i_1}, \dots, t_{i_k}) &= E[e^{t_{i_1}X_{i_1}+\dots+t_{i_k}X_{i_k}}] = \\ &= E[e^{0X_1+\dots+0X_{i_1-1}+t_{i_1}X_{i_1}+0X_{i_1+1}+\dots+0X_{i_k-1}+t_{i_k}X_{i_k}+0X_{i_k+1}+\dots+0X_n}] = \\ &= M_X(0, \dots, 0, t_{i_1}, 0, \dots, 0, t_{i_k}, 0, \dots, 0). \end{aligned}$$

3. Independencia de variables Aleatorias

Estudiemos ahora la independencia de variables aleatorias, que puede recordarnos a indepencia de sucesos en estadística descriptiva.

Definición 3.1 (Independencia de variables aleatorias)**.** Sean $X_1, \ldots, X_n$ variables aleatorias definidas sobre el mismo espacio de probabilidad con funciones de distribución $F_{X_1}, \ldots, F_{X_n}$. Consideramos el vector aleatorio $X = (X_1, \ldots, X_n)$ con función de distribución conjunta F_X. Se dice que dichas variables son mutuamente independientes (o, simplemente, independientes) si:

$$F_X(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n), \quad \forall x_1, \ldots, x_n \in \mathbb{R}$$

3.1. Caracterizaciones de independencia para variables discretas

A continuación, veremos dos caracterizaciones de independencia para variables aleatorias discretas.

Proposición 3.1 (Caracterización mediante funciones de masa de probabilidad)**.** *Sean $X_1, \ldots, X_n$ variables aleatorias discretas definidas sobre el mismo espacio de probabilidad. Consideramos el vector aleatorio $X = (X_1, \ldots, X_n)$. Entonces, $X_1, \ldots, X_n$ son independientes si y solo si:*

$$P[X_1 = x_1, \ldots, X_n = x_n] = P[X_1 = x_1] \cdot \ldots \cdot P[X_n = x_n], \quad \forall x_1, \ldots, x_n \in \mathbb{R}$$

Ejemplo. Consideramos las variables aleatorias X_1 y X_2 con la siguiente función de masa de probabilidad conjunta:

$X_1 \backslash X2$	-2	1	
-1	$1/12$	$1/6$	$3/12 = 1/4$
0	$1/6$	$1/3$	$3/6 = 1/2$
1	$1/12$	$1/6$	$3/12 = 1/4$
	$4/12 = 1/3$	$4/6 = 2/3$	1

Comprobemos que X_1 y X_2 son independientes.

Notemos que hemos añadido la última fila y columna para facilitar los cálculos. Para comprobar la independencia, debemos comprobar que la función de masa de probabilidad conjunta es igual al producto de las funciones de masa de probabilidad marginales. Directamente con la tabla, vemos se tiene para todos los casos, aunque vamos a escribirlo explícitamente:

$$\begin{aligned}
P[X_1 = -1, X_2 = -2] &= 1/12 = P[X_1 = -1] \cdot P[X_2 = -2] = 1/3 \cdot 1/4 = 1/12 \\
P[X_1 = -1, X_2 = 1] &= 1/6 = P[X_1 = -1] \cdot P[X_2 = 1] = 2/3 \cdot 1/4 = 2/12 = 1/6 \\
P[X_1 = 0, X_2 = -2] &= 1/6 = P[X_1 = 0] \cdot P[X_2 = -2] = 1/2 \cdot 1/3 = 1/6 \\
P[X_1 = 0, X_2 = 1] &= 1/3 = P[X_1 = 0] \cdot P[X_2 = 1] = 1/2 \cdot 2/3 = 1/3 \\
P[X_1 = 1, X_2 = -2] &= 1/12 = P[X_1 = 1] \cdot P[X_2 = -2] = 1/4 \cdot 1/3 = 1/12 \\
P[X_1 = 1, X_2 = 1] &= 1/6 = P[X_1 = 1] \cdot P[X_2 = 1] = 1/4 \cdot 2/3 = 1/6
\end{aligned}$$

Por tanto, X_1 y X_2 son independientes.

Proposición 3.2 (Caracterización mediante factorización de la función masa de probabilidad conjunta)**.** *Sean $X_1, \ldots, X_n$ variables aleatorias discretas definidas sobre el mismo espacio de probabilidad. Entonces, $X_1, \ldots, X_n$ son independientes si y solo si:*

$$P[X_1 = x_1, \ldots, X_n = x_n] = h_1(x_1) \cdot \ldots \cdot h_n(x_n), \quad \forall x_1, \ldots, x_n \in \mathbb{R}$$

siendo $h_i : \mathbb{R} \to \mathbb{R}$ funciones arbitrarias para $i = 1, \ldots, n$.

Observación. Notemos que la segunda caracterización de independencia para variables discretas es más general que la primera, ya que las funciones h_i pueden ser arbitrarias. Por tanto, no es necesario el cálculo de las funciones masa de probabilidad marginales para comprobar la independencia.

Veamos un ejemplo de esta caracterización, donde la anterior observación entra en juego.

Ejemplo. Sea $X = (X_1, X_2)$ un vector aleatorio con función de masa de probabilidad conjunta:

$$P[X_1 = x_1, X_2 = x_2] = \frac{1}{2^{x_1+1}} \qquad \forall x_1 \in \mathbb{N},\ x_2 \in \{0, 1\}$$

Comprobemos que X_1 y X_2 son independientes.

Cálculo de las funciones masa de probabilidad marginales Calculamos ambas marginales:

$$P[X_1 = x_1] = \sum_{x_2 \in \{0,1\}} P[X_1 = x_1, X_2 = x_2] = \sum_{x_2 \in \{0,1\}} \frac{1}{2^{x_1+1}} = \frac{1}{2^{x_1+1}} \cdot 2 = \frac{1}{2^{x_1}}$$

$$P[X_2 = x_2] = \sum_{x_1 \in \mathbb{N}} P[X_1 = x_1, X_2 = x_2] = \sum_{x_1=1}^{\infty} \frac{1}{2^{x_1+1}} = \frac{1}{2} \sum_{x_1=1}^{\infty} \frac{1}{2^{x_1}} = \frac{1}{2} \left(-1 + \sum_{x_1=0}^{\infty} \frac{1}{2^{x_1}} \right) =$$
$$= \frac{1}{2} \left(-1 + \frac{1}{1 - 1/2} \right) = \frac{1}{2}(-1 + 2) = \frac{1}{2}$$

Por tanto, tenemos que:

$$P[X_1 = x_1] \cdot P[X_2 = x_2] = \frac{1}{2^{x_1}} \cdot \frac{1}{2} = \frac{1}{2^{x_1+1}} = P[X_1 = x_1, X_2 = x_2]$$

Por tanto, X_1 y X_2 son independientes.

Factorización de la función masa de probabilidad conjunta Consideramos las siguientes funciones auxiliares:

$$\begin{array}{rccl} h_1: & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & x_1 & \longmapsto & \begin{cases} \dfrac{1}{2^{x_1+1}} & \text{si } x_1 \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases} \end{array}$$

$$\begin{array}{rccl} h_2: & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & x_2 & \longmapsto & \begin{cases} 1 & \text{si } x_2 \in \{0,1\} \\ 0 & \text{en otro caso} \end{cases} \end{array}$$

Dado $x_1 \in \mathbb{N}$ y $x_2 \in \{0, 1\}$, se tiene:

$$P[X_1 = x_1, X_2 = x_2] = \frac{1}{2^{x_1+1}} = h_1(x_1) \cdot h_2(x_2)$$

Por tanto, X_1 y X_2 son independientes.

Como vemos, la segunda caracterización nos ha permitido comprobar la independencia de X_1 y X_2 sin necesidad de calcular las funciones masa de probabilidad marginales.

3.2. Caracterizaciones de independencia para variables continuas

A continuación, veremos dos caracterizaciones de independencia para variables aleatorias continuas; las cuales serán análogas a las de variables discretas.

Proposición 3.3 (Caracterización mediante funciones de densidad de probabilidad)**.** *Sean $X_1, \dots, X_n$ variables aleatorias continuas definidas sobre el mismo espacio de probabilidad. Entonces, $X_1, \dots, X_n$ son independientes si y solo si:*

$$\begin{cases} X = (X_1, \dots, X_n) \text{ es un vector aleatorio continuo} \\ \quad \wedge \\ f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n), \quad \forall x_1, \dots, x_n \in \mathbb{R} \end{cases}$$

Demostración. Demostraremos mediante doble implicación.

$\Longrightarrow$) Supongamos que $X_1, \dots, X_n$ son independientes. Entonces, la función de distribución conjunta es:

$$\begin{aligned} F_X(x_1, \dots, x_n) &= F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n) = \\ &= \left(\int_{-\infty}^{x_1} f_{X_1}(t_1)\, dt_1\right) \cdot \ldots \cdot \left(\int_{-\infty}^{x_n} f_{X_n}(t_n)\, dt_n\right) = \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f_{X_1}(t_1) \cdot \ldots \cdot f_{X_n}(t_n)\, dt_1 \ldots dt_n \end{aligned}$$

Por tanto, tenemos que X es de tipo continuo, con función de densidad:

$$f_X(x_1,\dots,x_n) = f_{X_1}(x_1)\cdot\ldots\cdot f_{X_n}(x_n) \qquad \forall x_1,\dots,x_n\in\mathbb{R}$$

Vemos, de hecho, que esta es integrable y no negativa.

$\Longleftarrow)$ Supongamos que $X_1,\dots,X_n$ son tales que X es un vector aleatorio continuo y:

$$f_X(x_1,\dots,x_n) = f_{X_1}(x_1)\cdot\ldots\cdot f_{X_n}(x_n), \quad \forall x_1,\dots,x_n\in\mathbb{R}$$

Comprobemos ahora que son independientes. Tenemos que:

$$\begin{aligned}F_X(x_1,\dots,x_n) &= \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} f_{X_1}(t_1)\cdot\ldots\cdot f_{X_n}(t_n)\ dt_1\ldots dt_n =\\ &= \left(\int_{-\infty}^{x_1} f_{X_1}(t_1)\ dt_1\right)\cdot\ldots\cdot\left(\int_{-\infty}^{x_n} f_{X_n}(t_n)\ dt_n\right) =\\ &= F_{X_1}(x_1)\cdot\ldots\cdot F_{X_n}(x_n)\end{aligned}$$

Por tanto, $X_1,\dots,X_n$ son independientes.

□

Proposición 3.4 (Caracterización mediante factorización de la función densidad de probabilidad conjunta)**.** *Sean $X_1,\dots,X_n$ variables aleatorias continuas definidas sobre el mismo espacio de probabilidad. Entonces, $X_1,\dots,X_n$ son independientes si y solo si:*

$$\begin{cases} X=(X_1,\dots,X_n) \text{ es un vector aleatorio continuo}\\ \quad\wedge\\ f_X(x_1,\dots,x_n) = h_1(x_1)\cdot\ldots\cdot h_n(x_n), \quad \forall x_1,\dots,x_n\in\mathbb{R}\end{cases}$$

siendo $h_i:\mathbb{R}\to\mathbb{R}$ funciones arbitrarias para $i=1,\dots,n$.

Observación. De nuevo, consideramos la misma observación que en el caso de variables discretas: la segunda caracterización de independencia para variables continuas es más general que la primera, ya que las funciones h_i pueden ser arbitrarias. Por tanto, no es necesario el cálculo de las funciones densidad de probabilidad marginales para comprobar la independencia.

3.3. Caracterización mediante conjuntos de Borel

Esta es la última caracterización de independencia que veremos, la cual es más general que las anteriores. Aunque esta no se usará en la práctica, es interesante verla puesto que nos relaciona directamente con la independencia de sucesos, explicada en Estadística Descriptiva.

Proposición 3.5 (Caracterización mediante conjuntos de Borel)**.** *Sean $X_1,\dots,X_n$ variables aleatorias definidas sobre el mismo espacio de probabilidad. Entonces, $X_1,\dots,X_n$ son independientes si y solo si:*

$$P[X_1\in B_1,\dots,X_n\in B_n] = P[X_1\in B_1]\cdot\ldots\cdot P[X_n\in B_n], \quad \forall B_1,\dots,B_n\in\mathcal{B}$$

Demostración. Demostramos mediante doble implicación:

$\Longrightarrow$) Supongamos que $X_1, \dots, X_n$ son independientes. Entonces, distinguimos en función de si son discretas o continuas:

- Si son discretas, para todo $B_1, \dots, B_n \in \mathcal{B}$, tenemos que:

$$\begin{aligned}
P[X_1 \in B_1, \dots, X_n \in B_n] &= \sum_{\substack{x_1 \in B_1 \\ \vdots \\ x_n \in B_n}} P[X_1 = x_1, \dots, X_n = x_n] \overset{(*)}{=} \\
&\overset{(*)}{=} \sum_{\substack{x_1 \in B_1 \\ \vdots \\ x_n \in B_n}} P[X_1 = x_1] \cdot \ldots \cdot P[X_n = x_n] = \\
&= \sum_{x_1 \in B_1} P[X_1 = x_1] \cdot \ldots \cdot \sum_{x_n \in B_n} P[X_n = x_n] = \\
&= P[X_1 \in B_1] \cdot \ldots \cdot P[X_n \in B_n]
\end{aligned}$$

  donde en $(*)$ hemos usado la independencia de las variables aleatorias discretas.

- Si son continuas, tenemos que:

$$\begin{aligned}
P[X_1 \in B_1, \dots, X_n \in B_n] &= \int_{B_1} \cdots \int_{B_n} f_X(x_1, \dots, x_n)\ dx_1 \dots dx_n \overset{(*)}{=} \\
&\overset{(*)}{=} \int_{B_1} \cdots \int_{B_n} f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)\ dx_1 \dots dx_n = \\
&= \int_{B_1} f_{X_1}(x_1)\ dx_1 \cdot \ldots \cdot \int_{B_n} f_{X_n}(x_n)\ dx_n = \\
&= P[X_1 \in B_1] \cdot \ldots \cdot P[X_n \in B_n]
\end{aligned}$$

  donde en $(*)$ hemos usado la independencia de las variables aleatorias continuas.

$\Longleftarrow$) Supongamos que $X_1, \dots, X_n$ son tales que:

$$P[X_1 \in B_1, \dots, X_n \in B_n] = P[X_1 \in B_1] \cdot \ldots \cdot P[X_n \in B_n], \quad \forall B_1, \dots, B_n \in \mathcal{B}$$

Tomando $B_i =]-\infty, x_i]$ para $i = 1, \dots, n$, siendo $x_i \in \mathbb{R}$, tenemos que:

$$\begin{aligned}
P[X_1 \leqslant x_1, \dots, X_n \leqslant x_n] &= P[X_1 \in B_1, \dots, X_n \in B_n] = \\
&= P[X_1 \in B_1] \cdot \ldots \cdot P[X_n \in B_n] = P[X_1 \leqslant x_1] \cdot \ldots \cdot P[X_n \leqslant x_n]
\end{aligned}$$

Por tanto, $X_1, \dots, X_n$ son independientes.

□

3.4. Propiedades de la independencia

Proposición 3.6. *Sea $X = c$ una variable aleatoria degenerada. Entonces X es independiente de cualquier otra variable aleatoria Y.*

Demostración. Sea $X = c$ una variable aleatoria degenerada, y sea Y otra variable aleatoria cualquiera. Entonces, X y Y son independientes, ya que:

$$P[X \leqslant x] = \begin{cases} 0 & \text{si } x < c \\ 1 & \text{si } x \geqslant c \end{cases} \qquad P[X \leqslant x, Y \leqslant y] = \begin{cases} 0 & \text{si } x < c \\ P[Y \leqslant y] & \text{si } x \geqslant c \end{cases}$$

Por tanto, $P[X \leqslant x, Y \leqslant y] = P[X \leqslant x]P[Y \leqslant y]$. □

Proposición 3.7. *Las variables de cualquier subconjunto de variables independientes son independientes. Es decir, si $X_1, \ldots, X_n$ son independientes, entonces $X_{i_1}, \ldots, X_{i_k}$ son independientes para cualquier subconjunto $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$.*

Proposición 3.8. *Sean $X_1, \ldots, X_n$ variables aleatorias sobre el mismo espacio de probabilidad. Estas son independientes si y solo si las distribuciones condicionadas de cualquier subvector a cualquier otro coinciden con la marginal del primero.*

Proposición 3.9. *Sean $X_1, \ldots, X_n$ variables aleatorias sobre el mismo espacio de probabilidad, y consideramos funciones medibles $g_i : \mathbb{R} \to \mathbb{R}$ para $i = 1, \ldots, n$. Entonces:*

$$X_1, \ldots, X_n \text{ son independientes} \Longrightarrow g_1(X_1), \ldots, g_n(X_n) \text{ son independientes}$$

Proposición 3.10 (Caracterización por funciones generatrices de momentos)**.** *Sean $X_1, \ldots, X_n$ variables aleatorias sobre el mismo espacio de probabilidad tales que $\exists M_{X_i}$ en un entorno de I_i de 0 para todo $i = 1, \ldots, n$ en un entorno de 0. Entonces, $X_1, \ldots, X_n$ son independientes si y solo si:*

$$\left\{ \begin{array}{l} \exists M_X(t_1, \ldots, t_n) \quad \forall (t_1, \ldots, t_n) \in I_1 \times \ldots \times I_n \\ M_X(t_1, \ldots, t_n) = M_{X_1}(t_1) \cdot \ldots \cdot M_{X_n}(t_n) \quad \forall t_i \in I_i \end{array} \right.$$

3.4.1. Teorema de la multiplicación de las esperanzas

Incluimos el siguiente teorema. Este es de tal importancia que se le ha dado una sección propia, ya que tiene gran variedad de resultados.

Teorema 3.11 (Teorema de la Multiplicación de las Esperanzas)**.** *Sean $X_1, \ldots, X_n$ variables aleatorias sobre el mismo espacio de probabilidad, donde suponemos que $\exists E[X_i]$ para todo $i = 1, \ldots, n$. Entonces:*

$$X_1, \ldots, X_n \text{ son independientes} \Longrightarrow \left\{ \begin{array}{l} \exists E[X_1 \cdot \ldots \cdot X_n] \\ \quad \wedge \\ E[X_1 \cdot \ldots \cdot X_n] = E[X_1] \cdot \ldots \cdot E[X_n] \end{array} \right.$$

Corolario 3.11.1. *Sean X, Y dos variables aleatorias sobre el mismo espacio de probabilidad independientes. Entonces, si $\exists \operatorname{Cov}[X, Y]$, se tiene:*

$$\operatorname{Cov}[X, Y] = 0$$

Demostración. Dado que X e Y son independientes, se tiene:

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0$$

□

Corolario 3.11.2. *Sean X e Y variables aleatorias independientes. Entonces, si $\exists E[X^2],\ E[Y^2]$, se tiene:*

$$\exists\,\text{Var}[X \pm Y] = \text{Var}[X] + \text{Var}[Y]$$

Demostración. Tenemos que:

$$\begin{aligned}\text{Var}[X \pm Y] &= E[(X \pm Y)^2] - E[X \pm Y]^2 = \\ &= E[X^2 + Y^2 \pm 2XY] - E[X]^2 - E[Y]^2 \mp 2E[X]E[Y] = \\ &= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 \pm 2E[XY] \mp 2E[X]E[Y] \overset{(*)}{=} \\ &\overset{(*)}{=} \text{Var}[X] + \text{Var}[Y]\end{aligned}$$

donde en $(*)$ hemos usado el teorema anterior. □

Corolario 3.11.3. *Sean $X_1, \ldots, X_n$ variables aleatorias sobre el mismo espacio de probabilidad independientes. Entonces, si $\exists E[X_i^2]$ para todo $i = 1, \ldots, n$, se tiene:*

$$\text{Var}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i^2\,\text{Var}[X_i] \quad \forall a_1, \ldots, a_n \in \mathbb{R}$$

Demostración.

$$\text{Var}\left[\sum_{i=1}^{n} a_i X_i\right] \overset{(*)}{=} \sum_{i=1}^{n} \text{Var}[a_i X_i] = \sum_{i=1}^{n} a_i^2\,\text{Var}[X_i]$$

donde en $(*)$ hemos usado el resultado anterior. □

Corolario 3.11.4. *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes con función generatriz de momentos $M_{X_i}(t)$ para todo $i = 1, \ldots, n$. Entonces, la función generatriz de momentos de la variable aleatoria $X = \sum\limits_{i=1}^{n} X_i$ es:*

$$M_X(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

Demostración.

$$M_X(t) = E[e^{tX}] = E\left[e^{t\sum\limits_{i=1}^{n} X_i}\right] = E\left[\prod_{i=1}^{n} e^{tX_i}\right] \overset{(*)}{=} \prod_{i=1}^{n} E[e^{tX_i}] = \prod_{i=1}^{n} M_{X_i}(t)$$

donde en $(*)$ hemos empleado el teorema de la multiplicación de las esperanzas. □

3.5. Distribuciones Reproductivas

Definición 3.2 (Reproductividad de distribuciones)**.** Una distribución de variables aleatorias se dice reproductiva si, dadas X_1, X_2, ..., X_n variables aleatorias que sigan una distribución de dicho tipo (aunque la distribución de cada variable tenga distintos parámetros), entonces la variable aleatoria dada por:

$$X = \sum_{i=1}^{n} X_i$$

sigue una distribución del mismo tipo.

Observación. Notemos que, para ver que una distribución es reproductiva, usaremos el Corolario 3.11.4 para, dadas n variables aleatorias $\{X_i\}_{i\in\{1,\dots,n\}}$ que sigan una distribución de dicho tipo, calcular la función generatriz de momentos de la variable aleatoria $X = \sum\limits_{i=1}^{n} X_i$ y, como $M_X(t)$ caracteriza la distribución de X, calculando $M_X(t)$ sabremos ya si X sigue una distribución del mismo tipo que las variables aleatorias X_i.

Veamos en primer lugar algunas distribuciones reproductivas para variables discretas.

Proposición 3.12 (Distribución Reproductiva - Binomial)**.** *Sean $X_1, \dots, X_n$ variables aleatorias independientes y $X_i \sim B(k_i, p)$ para $i = 1, \dots, n$. Entonces:*

$$\sum_{i=1}^{n} X_i \sim B\left(\sum_{i=1}^{n} k_i, p\right)$$

Demostración. Sabemos que la función generatriz de momentos de cada una de las variables aleatorias X_i es:

$$M_{X_i}(t) = (1 - p + pe^t)^{k_i} \qquad i = 1, \dots, n$$

Usando el Corolario 3.11.4, tenemos que:

$$M_{\sum\limits_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} (1 - p + pe^t)^{k_i} = (1 - p + pe^t)^{\sum\limits_{i=1}^{n} k_i}$$

Esta es la función generatriz de momentos de una variable aleatoria que sigue una distribución binomial con parámetros $\left(\sum\limits_{i=1}^{n} k_i, p\right)$. Por tanto, hemos demostrado que:

$$\sum_{i=1}^{n} X_i \sim B\left(\sum_{i=1}^{n} k_i, p\right)$$

□

Proposición 3.13 (Distribución Reproductiva - Poisson)**.** *Sean $X_1, \dots, X_n$ variables aleatorias independientes y $X_i \sim \mathcal{P}(\lambda_i)$ para $i = 1, \dots, n$. Entonces:*

$$\sum_{i=1}^{n} X_i \sim \mathcal{P}\left(\sum_{i=1}^{n} \lambda_i\right)$$

Demostración. Sabemos que la función generatriz de momentos de cada una de las variables aleatorias X_i es:

$$M_{X_i}(t) = e^{\lambda_i(e^t-1)} \qquad i = 1, \dots, n$$

Usando el Corolario 3.11.4, tenemos que:

$$M_{\sum\limits_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} e^{\lambda_i(e^t-1)} = e^{\left(\sum\limits_{i=1}^{n} \lambda_i\right)(e^t-1)}$$

Esta es la función generatriz de momentos de una variable aleatoria que sigue una distribución de Poisson con parámetro $\left(\sum\limits_{i=1}^{n} \lambda_i\right)$. Por tanto, hemos demostrado que:

$$\sum_{i=1}^{n} X_i \sim \mathcal{P}\left(\sum_{i=1}^{n} \lambda_i\right)$$

□

Proposición 3.14 (Distribución Reproductiva - Binomial Negativa)**.** *Sean $X_1, \dots, X_n$ variables aleatorias independientes y $X_i \sim BN(k_i, p)$ para $i = 1, \dots, n$. Entonces:*

$$\sum_{i=1}^{n} X_i \sim BN\left(\sum_{i=1}^{n} k_i, p\right)$$

Demostración. Sabemos que la función generatriz de momentos de cada una de las variables aleatorias X_i es:

$$M_{X_i}(t) = \left(\frac{p}{1-(1-p)e^t}\right)^{k_i} \qquad i = 1, \dots, n$$

Usando el Corolario 3.11.4, tenemos que:

$$M_{\sum\limits_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} \left(\frac{p}{1-(1-p)e^t}\right)^{k_i} = \left(\frac{p}{1-(1-p)e^t}\right)^{\sum\limits_{i=1}^{n} k_i}$$

Esta es la función generatriz de momentos de una variable aleatoria que sigue una distribución binomial negativa con parámetros $\left(\sum\limits_{i=1}^{n} k_i, p\right)$. Por tanto, hemos demostrado que:

$$\sum_{i=1}^{n} X_i \sim BN\left(\sum_{i=1}^{n} k_i, p\right)$$

□

Veamos el siguiente caso, último para distribuciones discretas. Aunque no es una familia de variables reproductivas, cumple la siguiente propiedad.

Proposición 3.15. *Sean $X_1, \dots, X_n$ variables aleatorias independientes y $X_i \sim G(p)$ para $i = 1, \dots, n$. Entonces:*

$$\sum_{i=1}^{n} X_i \sim BN(n, p)$$

Demostración. Sabemos que la función generatriz de momentos de cada una de las variables aleatorias X_i es:

$$M_{X_i}(t) = \frac{p}{1-(1-p)e^t} \qquad i = 1, \dots, n$$

Usando el Corolario 3.11.4, tenemos que:

$$M_{\sum\limits_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} \frac{p}{1-(1-p)e^t} = \frac{p^n}{(1-(1-p)e^t)^n}$$

Esta es la función generatriz de momentos de una variable aleatoria que sigue una distribución binomial negativa con parámetros $BN(n,p)$. Por tanto, hemos demostrado que:

$$\sum_{i=1}^{n} X_i \sim BN(n,p)$$

□

Veamos ahora algunas distribuciones reproductivas para variables continuas.

Proposición 3.16 (Distribución Reproductiva - Normal)**.** *Sean $X_1, \dots, X_n$ variables aleatorias independientes y $X_i \sim N(\mu_i, \sigma_i^2)$ para $i = 1, \dots, n$. Entonces:*

$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Demostración. Sabemos que la función generatriz de momentos de cada una de las variables aleatorias X_i es:

$$M_{X_i}(t) = e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}} \qquad i = 1, \dots, n$$

Usando el Corolario 3.11.4, tenemos que:

$$M_{\sum\limits_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}} = e^{\left(\sum\limits_{i=1}^{n} \mu_i\right)t + \frac{\left(\sum\limits_{i=1}^{n} \sigma_i^2\right)t^2}{2}}$$

Esta es la función generatriz de momentos de una variable aleatoria que sigue una distribución normal con parámetros $\left(\sum\limits_{i=1}^{n} \mu_i, \sum\limits_{i=1}^{n} \sigma_i^2\right)$. Por tanto, hemos demostrado que:

$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

□

Al igual que ocurría en el caso de la distribución geométrica, la distribución exponencial no es reproductiva. Sin embargo, cumple la siguiente propiedad.

Proposición 3.17. *Sean $X_1, \ldots, X_n$ variables aleatorias independientes tal que $X_i \sim \exp(\lambda)$ para $i = 1, \ldots, n$. Entonces:*

$$\sum_{i=1}^{n} X_i \sim \mathcal{E}\left(n, \lambda\right)$$

Demostración. Sabemos que la función generatriz de momentos de cada una de las variables aleatorias X_i es:

$$M_{X_i}(t) = \frac{\lambda}{\lambda - t} \qquad i = 1, \ldots, n$$

Usando el Corolario 3.11.4, tenemos que:

$$M_{\sum\limits_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} \frac{\lambda}{\lambda - t} = \frac{\lambda^n}{(\lambda - t)^n}$$

Esta es la función generatriz de momentos de una variable aleatoria que sigue una distribución de Erlang con parámetros n, λ. Por tanto, hemos demostrado que:

$$\sum_{i=1}^{n} X_i \sim \mathcal{E}\left(n, \lambda\right)$$

□

Proposición 3.18 (Distribución Reproductiva - Erlang)**.** *Sean $X_1, \ldots, X_n$ variables aleatorias independientes tal que $X_i \sim \mathcal{E}(k_i, \lambda)$ para $i = 1, \ldots, n$. Entonces:*

$$\sum_{i=1}^{n} X_i \sim \mathcal{E}\left(\sum_{i=1}^{n} k_i, \lambda\right)$$

Demostración. Sabemos que la función generatriz de momentos de cada una de las variables aleatorias X_i es:

$$M_{X_i}(t) = \frac{\lambda^{k_i}}{(\lambda - t)^{k_i}} \qquad i = 1, \ldots, n$$

Usando el Corolario 3.11.4, tenemos que:

$$M_{\sum\limits_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} \frac{\lambda^{k_i}}{(\lambda - t)^{k_i}} = \frac{\lambda^{\sum\limits_{i=1}^{n} k_i}}{(\lambda - t)^{\sum\limits_{i=1}^{n} k_i}}$$

Esta es la función generatriz de momentos de una variable aleatoria que sigue una distribución de Erlang con parámetros $\left(\sum\limits_{i=1}^{n} k_i, \lambda\right)$. Por tanto, hemos demostrado que:

$$\sum_{i=1}^{n} X_i \sim \mathcal{E}\left(\sum_{i=1}^{n} k_i, \lambda\right)$$

□

Proposición 3.19 (Distribución Reproductiva - Gamma)**.** *Sean $X_1, \ldots, X_n$ variables aleatorias independientes tal que $X_i \sim \Gamma(u_i, \lambda)$ para $i = 1, \ldots, n$. Entonces:*

$$\sum_{i=1}^{n} X_i \sim \Gamma\left(\sum_{i=1}^{n} u_i, \lambda\right)$$

Demostración. Sabemos que la función generatriz de momentos de cada una de las variables aleatorias X_i es:

$$M_{X_i}(t) = \frac{\lambda^{u_i}}{(\lambda - t)^{u_i}} \qquad i = 1, \ldots, n$$

Usando el Corolario 3.11.4, tenemos que:

$$M_{\sum\limits_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} \frac{\lambda^{u_i}}{(\lambda - t)^{u_i}} = \frac{\lambda^{\sum\limits_{i=1}^{n} u_i}}{(\lambda - t)^{\sum\limits_{i=1}^{n} u_i}}$$

Esta es la función generatriz de momentos de una variable aleatoria que sigue una distribución gamma con parámetros $\left(\sum\limits_{i=1}^{n} u_i, \lambda\right)$. Por tanto, hemos demostrado que:

$$\sum_{i=1}^{n} X_i \sim \Gamma\left(\sum_{i=1}^{n} u_i, \lambda\right)$$

□

3.6. Independencia para familias de variables aleatorias

Notemos que las definiciones de independencia que hemos dado hasta ahora son para un número finito de variables aleatorias. Sin embargo, podemos extender estas definiciones a familias de variables aleatorias infinitas numerables.

Definición 3.3 (Independencia mutua)**.** Sea $\{X_i\}_{i \in T}$ una familia de variables aleatorias, con T infinito numerable. Diremos que la familia $\{X_i\}_{i \in T}$, es independiente si, para todo subconjunto finito $T_0 \subset T$, las variables aleatorias $\{X_i\}_{i \in T_0}$ son independientes.

Definición 3.4 (Independencia dos a dos)**.** Sea $\{X_i\}_{i \in T}$ una familia de variables aleatorias, con T infinito numerable. Diremos que la familia $\{X_i\}_{i \in T}$, es independiente dos a dos si, para todo par de índices $i, j \in T$, $i \neq j$, las variables aleatorias X_i y X_j son independientes.

Es directo ver que la independencia mutua implica la independencia dos a dos. Sin embargo, la independencia dos a dos no implica la independencia mutua. Por tanto, la independencia mutua es una propiedad más fuerte que la independencia dos a dos.

3.7. Independencia para vectores aleatorios

Hasta ahora, hemos visto la independencia para variables aleatorias unidimensionales. Sin embargo, podemos extender estas definiciones a vectores aleatorios multidimensionales.

Definición 3.5 (Independencia para vectores aleatorios)**.** Sean $X^{(1)}, \dots, X^{(n)}$ vectores aleatorios de dimensión $d_1, \dots, d_n$ respectivamente. Sea $F_X^{(i)}$ la función de distribución de $X^{(i)}$ para $i = 1, \dots, n$, y F_X la función de distribución conjunta de $X = (X^{(1)}, \dots, X^{(n)})$. Diremos que los vectores aleatorios $X^{(1)}, \dots, X^{(n)}$ son independientes si:

$$F_X(x^{(1)}, \dots, x^{(n)}) = F_X^{(1)}(x^{(1)}) \cdot \dots \cdot F_X^{(n)}(x^{(n)}) \quad \forall x^{(1)} \in \mathbb{R}^{d_1}, \dots, x^{(n)} \in \mathbb{R}^{d_n}$$

4. Esperanza Condicionada

Cuando se considera un conjunto de variables aleatorias definidas en relación a un determinado experimento, es usual que existan relaciones entre ellas. El problema de regresión consiste en encontrar una función matemática que permita aproximar el valor de determinadas variables, conocidos los valores del resto.

Observación. Este concepto de regresión es el análogo al visto en la asignatura de EDIP. En este caso, en vez de establecer un carácter Y en función de un carácter X en función de sus frecuencias, buscamos establecer una relación entre dos variables aleatorias.

En el presente curso, por simplicidad, nos limitaremos a analizar el problema de regresión bidimensional; es decir, dadas dos variables aleatorias X e Y, queremos encontrar una función matemática que nos permita aproximar el valor de Y conocido el valor de X. Para encontrar esta función matemática usaremos el criterio de optimalidad denominado mínimos cuadrados, visto ya en EDIP.

4.1. Esperanza condicionada

Para encontrar esta función matemática, necesitamos introducir el concepto de esperanza condicionada.

Definición 4.1 (Esperanza condicionada)**.** Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se define la esperanza condicionada de X a Y, notada por $E[X \mid Y]$, como la variable aleatoria siguiente (en el caso de que exista):

$$E[X \mid Y](y) := E[X \mid Y = y]$$

Notemos que esta esperanza es una esperanza normal, solo que considerando la distribución condicionada de X a $Y = y$.

Veamos en función de si es discreta o continua:

- Si X e Y son discretas, entonces:

$$E[X \mid Y = y] = \sum_{x \in E_x} x \cdot P[X = x \mid Y = y] \qquad \text{con } P[Y = y] > 0$$

- Si X e Y son continuas, entonces:

$$E[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y=y}(x)\ dx \qquad \text{con } f_Y(y) > 0$$

Notemos que la definición de $E[X \mid Y = y]$ requiere que la distribución condicionada de X a $Y = y$ esté bien definida. Por este motivo, es preciso exigir que $P[Y = y] \neq 0$ en el caso discreto, y $f_Y(y) \neq 0$ en el caso continuo. Así, la variable aleatoria $E[X \mid Y]$ es una función de la variable Y, definida sobre el conjunto de sus valores, E_Y.

Veamos ahora qué hemos de imponer para que exista la esperanza condicionada de X a Y.

Proposición 4.1. *Sean dos variables aleatorias X e Y sobre el mismo espacio de probabilidad. Entonces:*

$$\exists E[X] \Longrightarrow \exists E[X \mid Y]$$

Demostración. Distinguimos en función de si X e Y son discretas o continuas:

- Si X e Y son discretas, entonces:

$$\begin{aligned}E[|X \mid Y = y|] &= \sum_{x\in E_x} |x| \cdot P[X = x \mid Y = y] = \sum_{x\in E_x} |x| \cdot \frac{P[X = x, Y = y]}{P[Y = y]} = \\ &= \frac{\sum\limits_{x\in E_x} |x| \cdot P[X = x, Y = y]}{P[Y = y]} \leqslant \frac{\sum\limits_{x\in E_x} |x| \cdot P[X = x]}{P[Y = y]} = \frac{E[|X|]}{P[Y = y]} < \infty\end{aligned}$$

  donde en la desigualdad hemos hecho uso de que

$$P[X = x, Y = y] \leqslant P[X = x] \qquad \forall x, y$$

  y, al final, hemos usado que $E[|X|] < \infty$ por tener $E[X]$ finita. Por tanto, $E[X \mid Y]$ existe.

- Si X e Y son continuas, entonces:

$$\begin{aligned}E[|X \mid Y = y|] &= \int_{-\infty}^{\infty} |x| \cdot f_{X|Y=y}(x)\ dx = \int_{-\infty}^{\infty} |x| \cdot \frac{f_{X,Y}(x, y)}{f_Y(y)}\ dx = \\ &= \frac{\int\limits_{-\infty}^{\infty} |x| \cdot f_{X,Y}(x, y)\ dx}{f_Y(y)}\end{aligned}$$

  Como $E[|X|] < \infty$, entonces:

$$\begin{aligned}E[|X|] &= \int_{-\infty}^{\infty} |x| \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy\, dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x| \cdot f_{X,Y}(x, y)\, dy\, dx < \infty \Longrightarrow \\ &\Longrightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x| \cdot f_{X,Y}(x, y)\, dx\, dy < \infty \Longrightarrow \int_{-\infty}^{\infty} |x| \cdot f_{X,Y}(x, y)\, dx < \infty \qquad \forall y \in \mathbb{R}\end{aligned}$$

  Por tanto, $E[X \mid Y]$ existe.

$\square$

Ejemplo. Sea (X, Y) un vector aleatorio bidimensional con función de densidad conjunta:

$$f_{X,Y}(x,y) = \begin{cases} 2 & \text{si } 0 < x < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Calcular $E[X \mid Y]$ y $E[Y \mid X]$.

Tenemos que:

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

Para calcular ambas funciones de densidad condicionada, necesitamos calcular las marginales de X e Y:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy = \int_x^1 2\ dy = 2(1-x)$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx = \int_0^y 2\ dx = 2y$$

Así, las funciones de densidad condicionada son:

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{2}{2y} = \frac{1}{y} \quad \text{si } 0 < x < y < 1$$
$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{2}{2(1-x)} = \frac{1}{1-x} \quad \text{si } 0 < x < y < 1$$

Por tanto, las esperanzas condicionadas son:

$$E[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y=y}(x)\ dx = \int_0^y x \cdot \frac{1}{y}\ dx = \frac{1}{y}\int_0^y x\ dx = \frac{1}{y}\left[\frac{x^2}{2}\right]_0^y = \frac{y}{2}$$
$$E[Y \mid X = x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X=x}(y)\ dy = \int_x^1 y \cdot \frac{1}{1-x}\ dy = \frac{1}{1-x}\int_x^1 y\ dy = \frac{1}{1-x}\left[\frac{y^2}{2}\right]_x^1 =$$
$$= \frac{1}{1-x}\left(\frac{1}{2} - \frac{x^2}{2}\right) = \frac{1}{2} \cdot \frac{1-x^2}{1-x} = \frac{1+x}{2}$$

Al igual que considerábamos $E[g(X)]$, consideramos ahora la esperanza condicionada de una función de X a Y. Formalmente, dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, y una función $g : \mathbb{R} \to \mathbb{R}$ medible, $g(X)$ es una variable aleatoria, y su esperanza condicionada a Y es:

$$E[g(X) \mid Y](y) := E[g(X) \mid Y = y]$$

Veamos en función de si son discretas o continuas:

- Si X e Y son discretas, entonces:

$$E[g(X) \mid Y = y] = \sum_{x \in E_x} g(x) \cdot P[X = x \mid Y = y] \qquad \text{con } P[Y = y] > 0$$

- Si X e Y son continuas, entonces:

$$E[g(X) \mid Y = y] = \int_{-\infty}^{\infty} g(x) \cdot f_{X|Y=y}(x)\, dx \qquad \text{con } f_Y(y) > 0$$

Al igual que ocurría con la esperanza condicionada de X a Y, para que exista basta con imponer que $E[g(X)] < \infty$.

Corolario 4.1.1. *Sean dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, y una función $g : \mathbb{R} \to \mathbb{R}$ medible. Entonces:*

$$\exists E[g(X)] \Longrightarrow \exists E[g(X) \mid Y]$$

Ejemplo. En el experimento aleatorio del lanzamiento de tres monedas se consideran las variables:

- X: Número de caras.
- Y: Diferencia, en valor absoluto, entre el número de caras y el número de cruces.

Calcular $E[X^2 \mid Y]$.

El espacio muestral es el siguiente:

$$\Omega = \{CCC, CC+, C+C, +CC, C++, +C+, ++C, +++\}$$

La función masa de probabilidad conjunta viene dada por:

$X \setminus Y$	1	3	$P[X = x_i]$
0	0	$1/2^3$	$1/2^3$
1	$3/2^3$	0	$3/2^3$
2	$3/2^3$	0	$3/2^3$
3	0	$1/2^3$	$1/2^3$
$P[Y = y_j]$	$3/2^2$	$1/2^2$	1

donde además hemos calculado las marginales de X e Y. Entonces, tenemos que:

$$\begin{aligned}
E[X^2 \mid Y = 1] &= \sum_{x \in E_x} x^2 \cdot P[X = x \mid Y = 1] = \\
&= 1^2 \cdot P[X = 1 \mid Y = 1] + 2^2 \cdot P[X = 2 \mid Y = 1] = \\
&= 1 \cdot \frac{P[X = 1, Y = 1]}{P[Y = 1]} + 4 \cdot \frac{P[X = 2, Y = 1]}{P[Y = 1]} = \\
&= 1 \cdot \frac{3/2^3}{3/2^2} + 4 \cdot \frac{3/2^3}{3/2^2} = \frac{5}{2} \\
E[X^2 \mid Y = 3] &= \sum_{x \in E_x} x^2 \cdot P[X = x \mid Y = 3] = \\
&= 0^2 \cdot P[X = 0 \mid Y = 3] + 3^2 \cdot P[X = 3 \mid Y = 3] = \\
&= 0 \cdot \frac{P[X = 0, Y = 3]}{P[Y = 3]} + 9 \cdot \frac{P[X = 3, Y = 3]}{P[Y = 3]} = \\
&= 9 \cdot \frac{1/2^3}{1/2^2} = \frac{9}{2}
\end{aligned}$$

Por tanto, la esperanza condicionada de X^2 a Y es:

$$E[X^2 \mid Y] = \begin{cases} 5/2 & \text{si } Y = 1 \\ 9/2 & \text{si } Y = 3 \end{cases}$$

Propiedades de la esperanza condicionada

Introducimos las siguientes propiedades, que se deducen de forma directa en la mayoría de los casos haciendo uso de que la esperanza condicionada de X a Y es la esperanza de X con respecto a la distribución condicionada de X a Y.

Proposición 4.2. *Sea Y una variable aleatoria discreta sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y $c \in \mathbb{R}$. Entonces:*

$$E[c \mid Y] = c$$

Demostración. Como Y es discreta, entonces:

$$E[c \mid Y = y] = \sum_{x \in E_x} x \cdot P[X = x \mid Y = y] = c \cdot P[X = c \mid Y = y] = c$$

Notemos que no consideramos Y continua, ya que c es una variable aleatoria constante (discreta) y solo hemos definido la esperanza condicionada para variables aleatorias del mismo tipo. □

Proposición 4.3 (Linealidad)**.** *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces:*

$$E[aX_1 + bX_2 \mid Y] = aE[X_1 \mid Y] + bE[X_2 \mid Y] \quad \forall a, b \in \mathbb{R}$$

Proposición 4.4. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Si $X \geqslant 0$, entonces:*

$$E[X \mid Y] \geqslant 0 \wedge E[X \mid Y] = 0 \Longleftrightarrow P[X = 0] = 1$$

Proposición 4.5 (Conservación del Orden)**.** *Sean X_1, X_2 e Y tres variables aleatorias sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Si $X_1 \leqslant X_2$, entonces:*

$$E[X_1 \mid Y] \leqslant E[X_2 \mid Y]$$

Proposición 4.6. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$, y consideramos una función $g : \mathbb{R} \to \mathbb{R}$ medible. Entonces:*

$$\text{Si } X, Y \text{ independientes} \Longrightarrow E[g(X) \mid Y] = E[g(X)]$$

En particular, tomando $g = Id$, tenemos:

$$\text{Si } X, Y \text{ independientes} \Longrightarrow E[X \mid Y] = E[X]$$

Demostración. Distinguimos en función de si X e Y son discretas o continuas:

- Si X e Y son discretas, como X e Y son independientes, entonces:

  $$P[X = x \mid Y = y] = P[X = x] \quad \forall x, y$$

  Por tanto, tenemos:

  $$E[g(X) \mid Y = y] = \sum_{x \in E_X} g(x) \cdot P[X = x \mid Y = y] = \sum_{x \in E_X} g(x) \cdot P[X = x] = E[g(X)]$$

- Si X e Y son continuas, como X e Y son independientes, entonces:

  $$f_{X|Y=y}(x) = f_X(x) \quad \forall x, y$$

  Por tanto, tenemos:

  $$E[g(X) \mid Y = y] = \int_{-\infty}^{\infty} g(x) \cdot f_{X|Y=y}(x)\ dx = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\ dx = E[g(X)]$$

$\square$

Proposición 4.7. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$, y consideramos una función $g : \mathbb{R} \to \mathbb{R}$ medible. Entonces:*

$$E[E[g(X) \mid Y]] = E[g(X)]$$

En particular, tomando $g = Id$, tenemos:

$$E[E[X \mid Y]] = E[X]$$

Demostración. Distinguimos en función de si X e Y son discretas o continuas:

- Si X e Y son discretas, entonces, como $E[g(X) \mid Y]$ es una variable aleatoria en función de Y, usando la esperanza de una función de una variable aleatoria, tenemos:

  $$\begin{aligned} E[E[g(X) \mid Y]] &= \sum_{y \in E_Y} E[g(X) \mid Y = y] \cdot P[Y = y] = \\ &= \sum_{y \in E_Y} \sum_{x \in E_X} g(x) \cdot P[X = x \mid Y = y] \cdot P[Y = y] = \\ &= \sum_{x \in E_X} g(x) \cdot \sum_{y \in E_Y} P[X = x, Y = y] = \\ &= \sum_{x \in E_X} g(x) \cdot P[X = x] = E[g(X)] \end{aligned}$$

- Si X e Y son continuas, entonces, como $E[g(X) \mid Y]$ es una variable aleatoria en función de Y, usando la esperanza de una función de una variable aleatoria, tenemos:

$$\begin{aligned}
E[E[g(X) \mid Y]] &= \int_{-\infty}^{\infty} E[g(X) \mid Y = y] \cdot f_Y(y)\ dy = \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) \cdot f_{X|Y=y}(x)\ dx \cdot f_Y(y)\ dy = \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) \cdot f_{X,Y}(x,y)\ dx\ dy = \\
&= \int_{-\infty}^{\infty} g(x) \cdot \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy\ dx = \\
&= \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\ dx = E[g(X)]
\end{aligned}$$

□

4.2. Momentos condicionados

Como una variable aleatoria que es, se pueden definir los momentos condicionados de una variable aleatoria a otra. Comenzamos con los momentos condicionados no centrados.

4.2.1. Momentos condicionados no centrados

Definición 4.2 (Momento condicionado)**.** Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad y $k \in \mathbb{N}$, se define el momento condicionado de orden k de X a Y como la variable aleatoria siguiente (en el caso de que exista):

$$E[X^k \mid Y]$$

Usando la esperanza condicionada de una función, tenemos que $\exists E[X^k]$ implica que exista $E[X^k \mid Y]$, y se calcula como:

- Si X e Y son discretas, entonces:

$$E[X^k \mid Y = y] = \sum_{x \in E_x} x^k \cdot P[X = x \mid Y = y] \qquad \text{con } P[Y = y] > 0$$

- Si X e Y son continuas, entonces:

$$E[X^k \mid Y = y] = \int_{-\infty}^{\infty} x^k \cdot f_{X|Y=y}(x)\ dx \qquad \text{con } f_Y(y) > 0$$

4.2.2. Momentos condicionados centrados

Respecto de los momentos condicionados centrados, tenemos la siguiente definición.

Definición 4.3 (Momento condicionado centrado)**.** Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad y $k \in \mathbb{N}$, se define el momento condicionado de orden k centrado de X a Y como la variable aleatoria siguiente (en el caso de que exista):

$$E[(X - E[X \mid Y])^k \mid Y]$$

En caso de existencia, los valores de estas variables se calculan teniendo en cuenta que al considerar el condicionamiento a un valor arbitrario, $Y = y$, la variable $E[X \mid Y]$ toma el valor $E[X \mid Y = y]$ y, por tanto:

$$E[(X - E[X \mid Y])^k \mid Y = y] = E[(X - E[X \mid Y = y])^k \mid Y = y] \quad \forall y \in E_Y$$

Entonces, ya que dado $Y = y$, $E[X \mid Y = y]$ es una constante, la variable $(X - E[X \mid Y = y])^k$ sólo depende de X, y aplicando de nuevo la expresión de la esperanza condicionada de una función de una variable aleatoria, se tiene:

- Si X e Y son discretas, entonces:

$$E[(X-E[X \mid Y])^k \mid Y = y] = \sum_{x \in E_x} (x-E[X \mid Y = y])^k \cdot P[X = x \mid Y = y] \qquad \text{con } P[Y = y] > 0$$

- Si X e Y son continuas, entonces:

$$E[(X-E[X \mid Y])^k \mid Y = y] = \int_{-\infty}^{\infty} (x-E[X \mid Y = y])^k \cdot f_{X|Y=y}(x)\, dx \qquad \text{con } f_Y(y) > 0$$

Las propiedades de los momentos condicionados son similares a las de los momentos sin condicionar. En particular, la existencia de los momentos condicionados no centrados equivale a la de los momentos condicionados centrados.

4.3. Varianza condicionada

La varianza condicionada es un caso particular de los momentos condicionados centrados, y es de especial relevancia en el estudio de la regresión, por lo que estudiaremos su definición y propiedades.

Definición 4.4. Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se define la varianza condicionada de X a Y como el momento condicionado centrado de orden 2 de X a Y:

$$\mathrm{Var}[X \mid Y] := E[(X - E[X \mid Y])^2 \mid Y]$$

Veamos ahora algunas de sus propiedades.

Proposición 4.8. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Si $\exists E[X^2]$, entonces:*

1. $\exists\, \mathrm{Var}[X \mid Y]$ *y* $\mathrm{Var}[X \mid Y] \geqslant 0$.

2. $\mathrm{Var}[X \mid Y] = E[X^2 \mid Y] - (E[X \mid Y])^2$.

Demostración. Demostramos cada uno de los apartados:

1. Dado que $\exists E[X^2]$, entonces $\exists E[X^2 \mid Y]$ (de hecho, son iguales). Como la existencia de momentos no centrados implica la de momentos centrados, entonces $\exists E[(X - E[X \mid Y])^2 \mid Y] = \text{Var}[X \mid Y]$.

   Por otro lado, como $E[(X - E[X \mid Y])^2 \mid Y]$, es la esperanza condicionada de una variable aleatoria no negativa, entonces $\text{Var}[X \mid Y] \geqslant 0$.

2. Partiendo de la definición de varianza condicionada, fijado $Y = y$, tenemos:

$$\begin{aligned}
&\text{Var}[X \mid Y = y] = E[(X - E[X \mid Y = y])^2 \mid Y = y] = \\
&= E[X^2 - 2XE[X \mid Y = y] + (E[X \mid Y = y])^2 \mid Y = y] \overset{(*)}{=} \\
&\overset{(*)}{=} E[X^2 \mid Y = y] - 2E[E[X \mid Y = y] \cdot X \mid Y = y] + E[(E[X \mid Y = y])^2 \mid Y = y] \overset{(**)}{=} \\
&\overset{(**)}{=} E[X^2 \mid Y = y] - 2E[X \mid Y = y] \cdot E[X \mid Y = y] + (E[X \mid Y = y])^2 = \\
&= E[X^2 \mid Y = y] - 2(E[X \mid Y = y])^2 + (E[X \mid Y = y])^2 = \\
&= E[X^2 \mid Y = y] - (E[X \mid Y = y])^2
\end{aligned}$$

   donde en $(*)$ hemos usado la linealidad de la esperanza condicionada, y en $(**)$ hemos usado que, como ya hemos condicionado a $Y = y$, $E[X \mid Y = y]$ es una constante.

□

Introducimos además esta última proposición, que cobrará gran importancia en la siguiente sección.

Proposición 4.9 (Descomposición de la varianza)**.** *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Si $\exists E[X^2]$, entonces se tiene que $\exists\,\text{Var}[E[X \mid Y]]$ y $\exists E[\text{Var}[X \mid Y]]$, y se cumple que:*

$$\text{Var}[X] = \text{Var}[E[X \mid Y]] + E[\text{Var}[X \mid Y]]$$

Demostración. En primer lugar, tenemos que $\exists E[X^2]$ implica $\exists E[E[X^2 \mid Y]]$. Por tanto, del segundo apartado de la proposición anterior, tenemos que:

$$0 \leqslant \text{Var}[X \mid Y] = E[X^2 \mid Y] - (E[X \mid Y])^2 \Longrightarrow 0 \leqslant (E[X \mid Y])^2 \leqslant E[X^2 \mid Y]$$

Por tanto, como $(E[X \mid Y])^2$ es una variable aleatoria acotada por variables aleatorias con esperanza, entonces $\exists E[(E[X \mid Y])^2]$. Tenemos por tanto, que:

- $\exists E[(E[X \mid Y])^2] \Longrightarrow \exists\,\text{Var}[E[X \mid Y]] = E[(E[X \mid Y])^2] - (E[E[X \mid Y]])^2$.
- $\exists E[(E[X \mid Y])^2]$ y $\exists E[E[X^2 \mid Y]]$, lo que implica que:

  $\exists E[E[X^2 \mid Y]] - E[(E[X \mid Y])^2] = E[E[X^2 \mid Y] - (E[X \mid Y])^2] = E[\text{Var}[X \mid Y]]$

Por tanto, hemos demostrado las dos existencias. Para demostrar la igualdad, tenemos que:

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

Como queríamos demostrar. □

4.4. Regresión Mínimo Cuadrática

Explicamos de nuevo el problema de regresión, que ya se introdujo para Estadística Descriptiva. Dadas dos variables aleatorias, X e Y, definidas sobre el mismo espacio de probabilidad, el problema de regresión de Y sobre X consiste en determinar una función de X que proporcione una representación, lo más precisa posible, de la variable Y. Esto es, se trata de aproximar la variable Y por una función de X, de manera que la aproximación sea óptima en algún sentido preestablecido.

$$Y \approx \varphi(X)$$

donde:

- Y es la variable dependiente, explicada o endógena.
- X es la variable independiente, explicativa o exógena.
- φ es la función de regresión.

El criterio de optimalidad más usual para abordar el problema de regresión es el basado en el principio de mínimos cuadrados, y consiste en encontrar la función que minimiza la media de las desviaciones cuadráticas de las aproximaciones respecto de los verdaderos valores de la variable aproximada. Esto es, se trata de encontrar una función, φ, que minimice $E[(Y-\varphi(X))^2]$.

Definición 4.5 (Error cuadrático medio)**.** Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, el error cuadrático medio asociado a la función de regresión φ de forma que $Y \approx \varphi(X)$ es:

$$\text{E.C.M.}(\varphi) = E[(Y-\varphi(X))^2]$$

Por tanto, el problema de regresión abordado bajo esta perspectiva consiste en encontrar una función, φ_{opt}, que minimice el E.C.M.

4.4.1. Búsqueda de la función de regresión óptima, φ_{opt}

En la presente subsección, suponemos que estamos buscando aproximar Y por una función de X, $\varphi(X)$, y que hemos decidido que la función de regresión óptima es aquella que minimiza el E.C.M. asociado a la aproximación.

Observación. Se deja para el lector generalizar el caso para aproximar X por una función de Y.

Así, el problema de regresión se reduce a encontrar la función de regresión óptima, φ_{opt}, que minimiza el E.C.M. El siguiente teorema no se demostrará por ser un resultado que excede los conocimientos de la asignatura, pero nos proporciona la solución al problema de regresión bajo el criterio de mínimos cuadrados.

Teorema 4.10. *Sea Y una variable aleatoria sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$, y X una variable aleatoria sobre el mismo espacio de probabilidad. Entonces, $E[Y \mid X]$ minimiza el E.C.M. Es decir:*

$$\varphi_{opt}(X) = E[Y \mid X]$$

Además, esta función de regresión óptima es única.

El E.C.M. asociado a la función de regresión óptima, que es el mínimo error cuadrático medio cometido al aproximar Y a partir de X, es:

$$\begin{aligned}\text{E.C.M.}(\varphi_{\text{opt}}) &= \text{E.C.M.}(E[Y \mid X]) = E[(Y - E[Y \mid X])^2] = \\ &= E[E[(Y - E[Y \mid X])^2 \mid X]] = E[\text{Var}[Y \mid X]]\end{aligned}$$

Considerando la descomposición de la varianza, tenemos:

$$\text{Var}[Y] = \text{Var}[E[Y \mid X]] + E[\text{Var}[Y \mid X]] = \text{Var}[\varphi_{\text{opt}}(X)] + \text{E.C.M.}(\varphi_{\text{opt}})$$

Definición 4.6 (Curva de regresión mínimo cuadrática)**.** Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, la curva de regresión mínimo cuadrática de Y sobre X es la curva obtenida empleando $\varphi_{\text{opt}}(X)$. Es decir:

- Curva de Regresión Mínimo Cuadrática de Y sobre X:

$$\widehat{Y}(x) = E[Y \mid X = x] \qquad \forall x \in E_X$$

- Curva de Regresión Mínimo Cuadrática de X sobre Y:

$$\widehat{X}(y) = E[X \mid Y = y] \qquad \forall y \in E_Y$$

Veamos algunos casos particulares, para los que antes debemos introducir las siguientes definiciones.

Definición 4.7 (Dependencia funcional)**.** Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se dice que Y depende funcionalmente de X si existe una función $f : E_X \to E_Y$ tal que:

$$Y = f(X) \qquad \forall x \in E_X$$

Dicha función f se denomina curva de dependencia.

Proposición 4.11. *Sean X, Y dos variables aleatorias sobre el mismo espacio de probabilidad. Equivalen:*

1. $\text{Var}[Y \mid X] = 0$.
2. $E[\text{Var}[Y \mid X]] = 0$.
3. *Y depende funcionalmente de X.*
4. *E.C.M.*$(\varphi_{opt}) = 0$.

Demostración. Demostramos mediante distintas equivalencias:

(1) $\Longleftrightarrow$ (2) Sabemos que $\text{Var}[Y \mid X] \geqslant 0$, por lo que la equivalencia se tiene de forma directa.

(3) $\Longrightarrow$ (4) Tenemos que $Y = \varphi_{\text{opt}}(X)$, por lo que:

$$\text{E.C.M.}(\varphi_{\text{opt}}) = E[(Y - \varphi_{\text{opt}}(X))^2] = E[(Y - Y)^2] = 0$$

(4) $\Longrightarrow$ (3) Aunque se ve intuitivamente, veámoslo formalmente.

Dado que E.C.M.$(\varphi_{\text{opt}}) = 0$, entonces:

$$E[(Y - \varphi_{\text{opt}}(X))^2] = 0$$

Esta es la esperanza de una variable aleatoria no negativa, por lo que:

$$Y - \varphi_{\text{opt}}(X) = 0 \Longrightarrow Y = \varphi_{\text{opt}}(X)$$

(1) $\Longrightarrow$ (3) Partimos de:

$$\text{Var}[Y \mid X = x] = 0 \qquad \forall x \in E_X$$

Por tanto, fijado $x \in E_X$, $\exists!\ c_x \in E_y$ tal que:

$$[Y \mid X = x] = c_x$$

Por tanto, definimos la función $f : E_X \to E_Y$ como:

$$f(x) = c_x \qquad \forall x \in E_X$$

Por tanto, Y depende funcionalmente de X.

(3) $\Longrightarrow$ (1) Partimos de que Y depende funcionalmente de X, es decir, $\exists f : E_X \to E_Y$ tal que:

$$Y = f(X)$$

Entonces, tenemos que:

$$\begin{aligned}\text{Var}[Y \mid X = x] &= \text{Var}[f(X) \mid X = x] = \\ &= \text{Var}[f(x) \mid X = x] = 0 \qquad \forall x \in E_X\end{aligned}$$

donde hemos usado que la varianza de una constante es 0.

$\square$

Definición 4.8 (Dependencia recíproca)**.** Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se dice que hay dependencia recíproca entre X e Y si Y depende funcionalmente de X y X depende funcionalmente de Y con la misma función. Es decir, existe una función $f : E_X \to E_Y$ tal que:

$$\begin{aligned} Y &= f(X) \qquad \forall x \in E_X \\ X &= f^{-1}(Y) \qquad \forall y \in E_Y \end{aligned}$$

Dicha función f se denomina curva de dependencia.

Corolario 4.11.1. *Sean X, Y dos variables aleatorias sobre el mismo espacio de probabilidad.*

$$\text{Var}[Y \mid X] = \text{Var}[X \mid Y] = 0 \Longleftrightarrow Y \textit{ y } X \textit{ dependen recíprocamente}$$

Demostración. Demostramos mediante doble implicación:

$\Longrightarrow$) Como corolario de la proposición anterior, sabemos que $\exists f : E_X \to E_Y$, $g : E_Y \to E_X$ tales que:

$$\begin{aligned} Y &= f(X) \qquad \forall x \in E_X \\ X &= g(Y) \qquad \forall y \in E_Y \end{aligned}$$

Sea $c_x \in E_y$, y consideramos su preimagen

$$f^{-1}(c_x) = \{x \in E_X \mid f(x) = c_x\} = \{x \in E_X \mid Y = c_x\}$$

No obstante, usando la demostración de la proposición anterior tenemos que dicho c_x es único. Por tanto, $f^{-1} = g$ y, por tanto, Y y X dependen recíprocamente.

$\Longleftarrow$) Se tiene de forma directa como corolario de la proposición anterior.

$\square$

Algunos casos particulares son:

- Si Y depende funcionalmente de X con curva de dependencia $Y = f(X)$, entonces la curva de regresión mínimo cuadrática de Y sobre X es la propia curva de dependencia:

  $$y = f(x) = E[Y \mid X = x] \qquad \forall x \in E_X$$

  Demostración. Dado que $Y = f(X)$, entonces:

  $$E[Y \mid X = x] = E[f(X) \mid X = x] = E[f(x) \mid X = x] = f(x) \qquad \forall x \in E_X$$

  $\square$

- Si hay dependencia recíproca entre X e Y, es decir, $Y = f(X)$ y $X = f^{-1}(Y)$, entonces ambas curvas de regresión mínimo cuadrática coinciden con las curvas de dependencia:

  $$\begin{aligned} y = f(x) &= E[Y \mid X = x] \qquad \forall x \in E_X \\ x = f^{-1}(y) &= E[X \mid Y = y] \qquad \forall y \in E_Y \end{aligned}$$

- Si X e Y son independientes, entonces la curva de regresión mínimo cuadrática de Y sobre X es la esperanza de Y, y la de X sobre Y es la esperanza de X:

  $$\begin{aligned} \widehat{Y}(x) &= E[Y] \qquad \forall x \in E_X \\ \widehat{X}(y) &= E[X] \qquad \forall y \in E_Y \end{aligned}$$

  Como vemos, estas curvas de regresión mínimo cuadrática son constantes y paralelas a los ejes, lo que muestra que no tiene sentido plantear un problema de regresión en este caso.

Demostración. Dado que X e Y son independientes, sabemos de forma directa que:

$$\begin{aligned} E[Y \mid X = x] &= E[Y] \qquad \forall x \in E_X \\ E[X \mid Y = y] &= E[X] \qquad \forall y \in E_Y \end{aligned}$$

□

Por tanto, a modo de resumen, en la Tabla 4.1 se recogen las predicciones de Y según lo que estemos observando.

	Sin observar X	Observando X	Para $X = x$
$\widehat{Y}$	$E[Y]$	$E[Y \mid X]$	$E[Y \mid X = x]$
E.C.M.	$\text{Var}[Y]$	$E[\text{Var}[Y \mid X]]$	$\text{Var}[Y \mid X = x]$

Tabla 4.1: Resumen de las predicciones de Y sobre en función de lo que se observe.

4.4.2. Razones de Correlación

Buscamos ahora estudiar el grado de bondad de la aproximación mínimo cuadrática de cada variable a partir de la otra. Partimos de la siguiente igualdad:

$$Y = E[Y \mid X] + (Y - E[Y \mid X])$$

Sabemos que $\widehat{Y} = E[Y \mid X]$ es la mejor aproximación que hemos obtenido. Vemos por tanto que el error es la variable aleatoria $(Y - E[Y \mid X])$ unidades, que denominaremos *residuo*.

Definición 4.9 (Residuo)**.** Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, el residuo asociado a la aproximación de Y a partir de X es la siguiente variable aleatoria:

$$R = Y - E[Y \mid X]$$

La aproximación será mejor cuanto menor sea $|R|$. Esta es una variable aleatoria, cuya comparación para estudiar la bondad no es sencilla. Usaremos por tanto valores numéricos que resuman la bondad de la aproximación.

Como primer intento, buscamos comparar mediante $E[R]$. Tenemos que:

$$E[R] = E[Y - E[Y \mid X]] = E[Y] - E[E[Y \mid X]] = E[Y] - E[Y] = 0$$

Por tanto, siempre será nulo, lo que no nos aporta información sobre la bondad de la aproximación. Buscamos por tanto comparar mediante la varianza de R, conocida como *varianza residual*.

Definición 4.10 (Varianza residual)**.** Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, la varianza residual asociada a la aproximación de Y a partir de X es varianza del residuo, es decir:

$$\text{Var}[R] = \text{Var}[Y - E[Y \mid X]]$$

Calculemos el valor de la varianza residual.

$$\begin{aligned}\mathrm{Var}[R] &= \mathrm{Var}[Y - E[Y \mid X]] = E[(Y - E[Y \mid X])^2] - (E[Y - E[Y \mid X])^2 = \\ &= E[(Y - E[Y \mid X])^2] - \cancel{E[R]^2} = E[(Y - E[Y \mid X])^2] = \\ &= E[E[(Y - E[Y \mid X])^2 \mid X]] = E[\mathrm{Var}[Y \mid X]] = \text{E.C.M.}(\varphi_{\text{opt}})\end{aligned}$$

Por tanto, usando la descomposición de la varianza, tenemos que:

$$\mathrm{Var}[Y] = \mathrm{Var}[E[Y \mid X]] + E[\mathrm{Var}[Y \mid X]] = \mathrm{Var}\left[\widehat{Y}\right] + \mathrm{Var}[R] = \mathrm{Var}\left[\widehat{Y}\right] + \text{E.C.M.}(\varphi_{\text{opt}}) \tag{4.1}$$

Vemos que el E.C.M.(φ_{opt}) será menor conforme mayor sea $\mathrm{Var}\left[\widehat{Y}\right]$, por lo que la aproximación será mejor conforme mayor sea $\mathrm{Var}\left[\widehat{Y}\right] = \mathrm{Var}[E[Y \mid X]]$. No obstante, usar este valor para comparar la bondad de la aproximación introduce los siguientes problemas:

- $\mathrm{Var}\left[E[Y \mid X]\right]$ no es adimensional, por lo que no es un valor que podamos comparar.
- No es invariante frente a cambios de escala, lo cual puede llevar a conclusiones engañosas. Por ejemplo, sean:
  - Y una variable aleatoria que mide la altura de una persona en metros.
  - Y' una variable aleatoria que mide la altura de una persona en centímetros.

  Entonces, es razonable pensar que cualquier variable X debe aproximar igual de bien a Y que a $Y' = 100$. No obstante, tenemos que:

  $$\mathrm{Var}\left[E[Y' \mid X]\right] = \mathrm{Var}\left[E[100Y \mid X]\right] = \mathrm{Var}\left[100E[Y \mid X]\right] = 100^2\,\mathrm{Var}\left[E[Y \mid X]\right]$$

  De esta forma, midiendo la bondad de la aproximación por su varianza sin tener en cuenta la unidad de medida de las variables aproximadas, podríamos concluir que la variable X aproxima mucho mejor a Y' que a Y.

Estos inconvenientes se salvan normalizando la varianza de la función de regresión y usando el siguiente coeficiente, que es el que buscábamos presentar en esta sección.

Definición 4.11 (Razón de correlación)**.** Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se define la razón de correlación de Y sobre X:

$$\eta^2_{Y/X} = \frac{\mathrm{Var}\left[E[Y \mid X]\right]}{\mathrm{Var}[Y]}$$

De forma análoga se define la razón de correlación de X sobre Y:

$$\eta^2_{X/Y} = \frac{\mathrm{Var}\left[E[X \mid Y]\right]}{\mathrm{Var}[X]}$$

Observación. En el caso de que, por ejemplo, Y sea constante, tenemos $\mathrm{Var}[Y] = 0$. Se define en este caso $\eta^2_{Y/X} = 1$. Análogamente, si X es constante, se define $\eta^2_{X/Y} = 1$.

Usando la Ecuación 4.1, tenemos:

$$\eta^2_{Y/X} = \frac{\operatorname{Var}[E[Y \mid X]]}{\operatorname{Var}[Y]} = \frac{\operatorname{Var}[Y] - E[\operatorname{Var}[Y \mid X]]}{\operatorname{Var}[Y]} = 1 - \frac{E[\operatorname{Var}[Y \mid X]]}{\operatorname{Var}[Y]}$$

Veamos que $\eta^2_{Y/X}$ solventa los dos problemas que presentábamos al usar $\operatorname{Var}[E[Y \mid X]]$ para medir la bondad de la aproximación.

- $\eta^2_{Y/X}$ es adimensional, ya que las unidades de medida de $\operatorname{Var}[E[Y \mid X]]$ y $\operatorname{Var}[Y]$ son las mismas.
- Veamos que es invariante frente a cambios de escala. Sean Y e Y' dos variables aleatorias, y $Y' = aY$ con $a \in \mathbb{R}$. Entonces:

$$\eta^2_{Y'/X} = \eta^2_{aY/X} = \frac{\operatorname{Var}[E[aY \mid X]]}{\operatorname{Var}[aY]} = \frac{\operatorname{Var}[aE[Y \mid X]]}{a^2 \operatorname{Var}[Y]} = \frac{a^2 \operatorname{Var}[E[Y \mid X]]}{a^2 \operatorname{Var}[Y]} = \eta^2_{Y/X}$$

Por tanto, la razón de correlación es un valor adimensional e invariante frente a cambios de escala que nos permite comparar la bondad de la aproximación de una variable a partir de la otra. En particular, $\eta^2_{Y/X}$ mide la proporción de la varianza de Y que queda explicada por la función de regresión $\widehat{Y} = E[Y \mid X]$. En este sentido, se puede interpretar como una medida de la bondad del ajuste de la distribución a la curva de regresión correspondiente, de forma que *a mayor valor del coeficiente, mejor será la aproximación.*

Proposición 4.12. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces:*

1. $0 \leqslant \eta^2_{Y/X}, \eta^2_{X/Y} \leqslant 1$.
2. $\eta^2_{Y/X} = 0 \iff$ *La curva de regresión de Y sobre X es $Y = E[Y]$.*
   $\eta^2_{X/Y} = 0 \iff$ *La curva de regresión de X sobre Y es $X = E[X]$.*
3. $\eta^2_{Y/X} = 1 \iff$ *Y depende funcionalmente de X.*
   $\eta^2_{X/Y} = 1 \iff$ *X depende funcionalmente de Y.*
4. $\eta^2_{Y/X} = \eta^2_{X/Y} = 1 \iff$ *X e Y tienen dependencia recíproca.*

Demostración. Demostraremos tan solo los resultados para $\eta^2_{Y/X}$, ya que los resultados para $\eta^2_{X/Y}$ son análogos.

1. Tenemos en primer lugar $\operatorname{Var}[E[Y \mid X]] \geqslant 0$ y $\operatorname{Var}[Y] \geqslant 0$ por ser varianzas. Por tanto lado, usamos que:

$$\eta^2_{Y/X} = 1 - \frac{E[\operatorname{Var}[Y \mid X]]}{\operatorname{Var}[Y]}$$

Tenemos que $\operatorname{Var}[Y \mid X], \operatorname{Var}[Y] \geqslant 0$ por ser varianzas. Además, como la esperanza de una variable aleatoria no negativa es no negativa, tenemos que $E[\operatorname{Var}[Y \mid X]] \geqslant 0$. Por tanto, $\eta^2_{Y/X} \leqslant 1$ y tenemos lo buscado.

2. Tenemos que:

$$\eta^2_{Y/X} = 0 \iff \text{Var}\left[E[Y \mid X]\right] = 0 \iff \exists c \in \mathbb{R} \text{ tal que } E[Y \mid X] = c \quad \forall x \in E_X$$

donde la última implicación se debe a las propiedades de la varianza. Por tanto, tenemos que:

$$c = E[c] = E[E[Y \mid X]] = E[Y]$$

Por tanto, tenemos que:

$$\eta^2_{Y/X} = 0 \iff E[Y \mid X] = E[Y]$$

3. Tenemos que:

$$\begin{aligned} \eta^2_{Y/X} = 1 \iff E[\text{Var}[Y \mid X]] = 0 \iff P[\text{Var}[Y \mid X] = 0] = 1 \iff \\ \iff \text{Var}[Y \mid X] = 0 \end{aligned}$$

Por tanto, tenemos que Y depende funcionalmente de X.

4. Por lo razonado anteriormente, tenemos que:

$$\eta^2_{Y/X} = \eta^2_{X/Y} = 1 \iff \text{Var}[Y \mid X] = 0 = \text{Var}[X \mid Y]$$

Por tanto, X e Y tienen dependencia recíproca.

□

En esta última proposición, vemos que nos falta estudiar el caso en el que se da $\eta^2_{Y/X} = \eta^2_{X/Y} = 0$. Este lo veremos más adelante, porque se trata de ejemplos de rectas de regresión.

4.5. Rectas de Regresión

Cuando el cálculo de la esperanza condicionada es complicado y se tiene relativa seguridad de que los puntos de la distribución teórica se ajustan a una determinada forma funcional (exponencial, parabólica, etc.) puede ser útil restringir la búsqueda de la función de regresión óptima a la clase de funciones de dicha forma.

Un caso particularmente importante es el de la regresión lineal. Las rectas que mejor se ajustan a los puntos de la distribución en el sentido de mínimos cuadrados (para aproximar cada una de las variables en términos de la otra) se denominan rectas de regresión.

Observación. Supongamos de nuevo que queremos obtener la recta de regresión de Y sobre X, es decir, $Y = \varphi(X)$.

En estos casos, tenemos que φ ha de ser una recta, por lo que:

$$\varphi(X) = aX + b \qquad a, b \in \mathbb{R}$$

Por tanto, razonando de forma idéntica a como hicimos en la sección anterior, tenemos que el problema de regresión se reduce a encontrar los valores de a y b que minimizan el E.C.M. asociado a la aproximación. Es decir, la recta de regresión de Y sobre X es:

$$\varphi_{\text{opt}}^L(X) = \min_{a,b\in\mathbb{R}} E[(Y - aX - b)^2]$$

El problema de regresión lineal se reduce por tanto a minimizar la siguiente función de dos variables:

$$L(a,b) = E[(Y - aX - b)^2] = E[Y^2] + a^2E[X^2] + b^2 - 2aE[XY] - 2bE[Y] + 2abE[X]$$

Para calcular el mínimo de L, derivamos parcialmente e igualamos a 0:

$$\begin{aligned}
\frac{\partial L}{\partial a} &= 2aE[X^2] - 2E[XY] + 2bE[X] = 0 \\
\frac{\partial L}{\partial b} &= 2b - 2E[Y] + 2aE[X] = 0 \Longrightarrow 2b = 2E[Y] - 2aE[X]
\end{aligned}$$

Por tanto, suponiendo $\text{Var}[X] > 0$ (caso que estudiaremos más adelante), tenemos que:

$$\begin{aligned}
2aE[X^2] - 2E[XY] + (2E[Y] - 2aE[X])E[X] = 0 \Longrightarrow \\
&\Longrightarrow a(E[X^2] - E[X]^2) = E[XY] - E[X]E[Y] \Longrightarrow \\
&\Longrightarrow a = \frac{\text{Cov}(X,Y)}{\text{Var}[X]} \Longrightarrow \\
&\Longrightarrow b = E[Y] - \frac{\text{Cov}(X,Y)}{\text{Var}[X]} \cdot E[X]
\end{aligned}$$

que, como se comprueba calculando la matriz de derivadas segundas[1], proporciona el mínimo de la función L:

$$\begin{vmatrix} \dfrac{\partial^2 L}{\partial a^2} & \dfrac{\partial^2 L}{\partial a \partial b} \\ \dfrac{\partial^2 L}{\partial b \partial a} & \dfrac{\partial^2 L}{\partial b^2} \end{vmatrix} = \begin{vmatrix} 2E[X^2] & 2E[X] \\ 2E[X] & 2 \end{vmatrix} = 4(E[X^2] - E[X]^2) = 4\,\text{Var}[X] > 0$$

Por tanto, tenemos que la recta de regresión de Y sobre X es:

$$\widehat{Y} = \varphi_{\text{opt}}^L(X) = E[Y] + \frac{\text{Cov}(X,Y)}{\text{Var}[X]}(X - E[X])$$

Este resultado refuerza el resultado que ya vimos del signo de la covarianza:

- Si $\text{Cov}(X,Y) > 0$, entonces las rectas de regresión son crecientes.
- Si $\text{Cov}(X,Y) < 0$, entonces las rectas de regresión son decrecientes.
- Si $\text{Cov}(X,Y) = 0$, entonces las rectas de regresión son constantes e iguales a la esperanza de la variable explicada.

[1]Minimización de funciones en varias variables es materia de Análisis Matemático II.

Calculemos ahora su error cuadrático medio:

$$\text{E.C.M.}(\varphi^L_{\text{opt}}) = E[(Y-\varphi^L_{\text{opt}}(X))^2] = E\left[\left(Y-E[Y]-\frac{\text{Cov}(X,Y)}{\text{Var}[X]}(X-E[X])\right)^2\right] =$$
$$= E\left[(Y-E[Y])^2 - 2\frac{\text{Cov}(X,Y)}{\text{Var}[X]}(Y-E[Y])(X-E[X]) + \frac{\text{Cov}^2(X,Y)}{\text{Var}^2[X]}(X-E[X])^2\right] =$$
$$= E[(Y-E[Y])^2] - 2\frac{\text{Cov}(X,Y)}{\text{Var}[X]}E[(Y-E[Y])(X-E[X])] + \frac{\text{Cov}^2(X,Y)}{\text{Var}^2[X]}E[(X-E[X])^2] =$$
$$= \text{Var}[Y] - 2\frac{\text{Cov}(X,Y)}{\text{Var}[X]}\cdot\text{Cov}(X,Y) + \frac{\text{Cov}^2(X,Y)}{\text{Var}^2[X]}\cdot\text{Var}[X] =$$
$$= \text{Var}[Y] - \frac{\text{Cov}^2(X,Y)}{\text{Var}[X]}$$

Veamos ahora el caso particular en el que $\text{Var}[X]=0$. Entonces, $X=k\in\mathbb{R}$ es una variable degenerada (constante). Por tanto, cualquier función lineal de X es constante, por lo que el problema de regresión lineal se reduce a encontrar la constante que minimiza $E[(Y-c)^2]$. Dicha constante es la esperanza de Y:

$$\varphi^L_{\text{opt}}(X) = E[Y]$$

De aquí en adelante, supondremos que $\text{Var}[X]>0$ cuando estemos trabajando con la regresión lineal de Y sobre X, y análogamente para la regresión lineal de X sobre Y.

4.5.1. Coeficiente de determinación lineal

Veamos ahora el equivalente a las razones de correlación en el caso de la regresión lineal. Para calcularlo, seguimos el mismo razonamiento que en la sección anterior, y este será:

$$\frac{\text{Var}[\varphi^L_{\text{opt}}(X)]}{\text{Var}[Y]}$$

Calculemos dicha varianza:

$$\text{Var}[\varphi^L_{\text{opt}}(X)] = \text{Var}\left[E[Y]+\frac{\text{Cov}(X,Y)}{\text{Var}[X]}(X-E[X])\right] =$$
$$= \frac{\text{Cov}^2(X,Y)}{\text{Var}^2[X]}\cdot\text{Var}[X] = \frac{\text{Cov}^2(X,Y)}{\text{Var}[X]}$$

Por tanto, tenemos que:

$$\frac{\text{Var}[\varphi^L_{\text{opt}}(X)]}{\text{Var}[Y]} = \frac{\text{Cov}^2(X,Y)}{\text{Var}[X]\cdot\text{Var}[Y]}$$

Introducimos entonces la siguiente definición:

Definición 4.12 (Coeficiente de determinación lineal)**.** Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se define el coeficiente de determinación lineal de Y sobre X, notado como $\rho^2_{X,Y}$, como:

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

Observación. Notemos que este coeficiente coincide con el producto de las pendientes de las rectas de regresión de Y sobre X y de X sobre Y. Esto nos será muy útil en la práctica.

Como primer resultado directo que se deduce de la definición, tenemos que:

$$\rho^2_{X,Y} = \rho^2_{Y,X}$$

Por tanto, tenemos que ambos coeficientes de correlación son iguales, lo que nos permite referirnos simplemente a $\rho^2_{X,Y}$ como coeficiente de determinación lineal.

Al igual que en el caso de las razones de correlación, el coeficiente de determinación lineal es un valor adimensional e invariante frente a cambios de escala que nos permite comparar la bondad de la aproximación de una variable a partir de la otra. En este sentido, se puede interpretar como una medida de la bondad del ajuste de la distribución a la recta de regresión correspondiente, de forma que *a mayor valor del coeficiente, mejor será la aproximación.*

Como resultados, incluimos las siguientes propiedades del coeficiente de determinación lineal:

Proposición 4.13. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces:*

1. $\rho^2_{aX+b,cY+d} = \rho^2_{X,Y}$ *para* $a, b, c, d \in \mathbb{R}$.
2. $0 \leqslant \rho^2_{X,Y} \leqslant 1$.
3. *Tenemos que:*

$$\begin{aligned} 0 = \rho^2_{X,Y} &\iff \textit{La recta de regresión de } Y \textit{ sobre } X \textit{ es } Y = E[Y] \\ &\iff \textit{La recta de regresión de } X \textit{ sobre } Y \textit{ es } X = E[X] \end{aligned}$$

4. $1 = \rho^2_{X,Y} \iff X$ *y* Y *tienen dependencia funcional lineal recíproca.*
5. $\rho^2_{X,Y} \leqslant \eta^2_{Y/X},\ \eta^2_{X/Y}$.
6. $\rho^2_{X,Y} = \eta^2_{Y/X} \iff$ *La curva de regresión de Y sobre X coincide con la recta de regresión de Y sobre X.*
   $\rho^2_{X,Y} = \eta^2_{X/Y} \iff$ *La curva de regresión de X sobre Y coincide con la recta de regresión de X sobre Y.*

4.5.2. Coeficiente de correlación lineal de Pearson

El coeficiente de determinación lineal es una medida de la bondad de la aproximación de una variable a partir de la otra, pero no nos da información sobre la dirección de la relación entre las variables. Para ello, introducimos el siguiente concepto:

Definición 4.13 (Coeficiente de correlación lineal de Pearson)**.** Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se define el coeficiente de correlación lineal de Pearson entre Y y X, notado como $\rho_{X,Y}$, como:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}[X] \cdot \mathrm{Var}[Y]}}$$

Como primer resultado, de la definición deducimos que $\rho_{X,Y} = \rho_{Y,X}$, por lo que nos referiremos a $\rho_{X,Y}$ como coeficiente de correlación lineal de Pearson. Notemos que $\rho_{X,Y}$ es la raíz cuadrada del coeficiente de determinación lineal empleado el signo de la covarianza. Debido a que el signo de $\rho_{X,Y}$ coincide con el signo de la covarianza, y este nos da la dirección de la relación entre las variables, el coeficiente de correlación lineal de Pearson nos da información sobre la dirección de la relación entre las variables.

- Si $\rho_{X,Y} > 0$, entonces las variables están positivamente correladas; es decir, si una de las variables aumenta, la otra también lo hace.
- Si $\rho_{X,Y} < 0$, entonces las variables están negativamente correladas; es decir, si una de las variables aumenta, la otra disminuye.
- Si $\rho_{X,Y} = 0$, entonces las rectas de regresión son constantes e iguales a las esperanzas de las variables dependientes. En este caso, se dice que son *incorreladas.*

Como resultado directo de esta última definición, tenemos la siguiente caracterización, que en muchos casos es la que se da como definición de incorrelación.

Proposición 4.14. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces:*

$$X, Y \text{ son incorreladas} \iff \operatorname{Cov}[X, Y] = 0$$

Llegados a este punto, podemos estudiar el caso que nos faltaba de las curvas de regresión, que es el caso en el que $\eta^2_{Y/X} = \eta^2_{X/Y} = 0$.

Corolario 4.14.1. *Sea X e Y dos variables aleatorias sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces:*

$$\eta^2_{Y/X} = \eta^2_{X/Y} = 0 \Longrightarrow X, Y \text{ son incorreladas.}$$

Demostración. Al ser ambas razones de correlación nulas, tenemos que las curvas de regresión son las esperanzas (constantes), luego son rectas y tenemos que:

$$\rho^2_{X,Y} = \eta^2_{Y/X} = \eta^2_{X/Y} = 0$$

Por tanto, tenemos que X e Y son incorreladas. □

El siguiente resultado se deduce de forma directa de la definición del coeficiente de correlación lineal de Pearson y de las propiedades del coeficiente de determinación lineal:

Proposición 4.15. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces:*

1. $-1 \leqslant \rho_{X,Y} \leqslant 1$.
2. *$\rho_{X,Y}$ es invariante frente a cambios de escala (salvo signo). Esto se suele notar como:*

$$\rho_{aX+b,cY+d} = \pm\rho_{X,Y} \qquad a, b, c, d \in \mathbb{R}$$

5. Algunos Modelos Multivariantes

En el presente Capítulo, y al igual que hemos hecho en el caso unidimensional, vamos a estudiar ahora ciertos modelos multivariantes que nos permitirán trabajar con más de una variable aleatoria. En concreto, vamos a generalizar la distribución binomial y la normal, ya estudiadas previamente.

5.1. Distribución Multinomial

Como ya hemos adelantado, esta distribución es una generalización de la distribución binomial. Fijamos $k \in \mathbb{N}$, y consideramos un experimento aleatorio con $k+1$ posibles resultados asociados a otros tantos sucesos $A_1, A_2, \ldots, A_{k+1}$ exhaustivos y mutuamente excluyentes, por lo tanto constituyen una partición del espacio muestral, con probabilidad de ocurrencia $p_i \in]0,1[$, $i = 1, \ldots, k+1$. Es decir:

$$\begin{aligned} \Omega &= \bigcup_{i=1}^{k+1} A_i, \\ A_i \cap A_j &= \emptyset, \quad i,j = 1, \ldots, k+1, \qquad i \neq j, \\ P(A_i) &= p_i, \quad i = 1, \ldots, k+1 \end{aligned}$$

Tenemos entonces que:

$$P(A_{k+1}) = 1 - P\left(\bigcup_{i=1}^{k} A_i\right) = 1 - \sum_{i=1}^{k} P(A_i) = 1 - \sum_{i=1}^{k} p_i$$

Supongamos que realizamos n repeticiones independientes del experimento en las mismas condiciones, por tanto las probabilidades p_i, $i = 1, 2, \ldots, k+1$ se mantienen constantes a lo largo de las repeticiones, y sean $X_1, \ldots, X_{k+1}$ variables aleatorias tales que X_i contabiliza el número de veces que ocurre el suceso A_i en las n repeticiones del experimento, $i = 1, 2, \ldots, k+1$. Tenemos que:

$$X_1 + X_2 + \ldots + X_{k+1} = n$$

En estas condiciones, podemos definir la distribución multinomial.

Definición 5.1 (Distribución Multinomial)**.** En las condiciones anteriores, se define la distribución multinomial $k-$dimensional con parámetros n y $p_1, \ldots, p_k$ como la distribución del vector aleatorio $X = (X_1, \ldots, X_k)$ en el cual cada componente X_i contabiliza el número de ocurrencias del suceso A_i, $i = 1, \ldots, k$. La notaremos por:

$$X \sim M_k(n, p_1, \ldots, p_k)$$

Observación. Notemos que la distribución binomial es un caso particular de la distribución multinomial, en concreto, la distribución multinomial con $k = 1$.

Razonemos ahora la función de masa de probabilidad de la distribución multinomial. Consideramos el vector $x = (x_1, \ldots, x_k)$ tal que $x_i \in \mathbb{N}_0$, $i = 1, \ldots, k$. Como se realizan n repeticiones del experimento, tenemos que:

$$\sum_{i=1}^{k} x_i \leqslant n$$

donde la igualdad solo se dará si no se produce el suceso A_{k+1} en ninguna de las repeticiones. Una de las posibles ordenaciones para que x tome dicho valor es que primero ocurran x_1 veces el suceso A_1, luego x_2 veces el suceso A_2, y así sucesivamente, hasta que ocurran x_k veces el suceso A_k y finalmente ocurran $n - \sum_{i=1}^{k} x_i$ veces el suceso A_{k+1}. Es decir, el suceso sería (notando $A_i^{x_i}$ al suceso A_i ocurrido x_i veces):

$$A_1^{x_1} \cap A_2^{x_2} \cap \ldots \cap A_k^{x_k} \cap A_{k+1}^{n - \sum\limits_{i=1}^{k} x_i}$$

La probabilidad de que ocurra este suceso, usando la independencia de las repeticiones del experimento, es:

$$P\left[\bigcap_{i=1}^{k} \left(\bigcap_{j=1}^{x_i} A_i\right) \cap \left(\bigcap_{j=1}^{n - \sum\limits_{i=1}^{k} x_i} A_{k+1}\right)\right] = \prod_{i=1}^{k} p_i^{x_i} \cdot p_{k+1}^{n - \sum\limits_{i=1}^{k} x_i} = \prod_{i=1}^{k} p_i^{x_i} \cdot \left(1 - \sum_{i=1}^{k} p_i\right)^{n - \sum\limits_{i=1}^{k} x_i}$$

Veamos ahora cuántas ordenaciones distintas nos pueden dar dicho vector x. Estas ordenaciones son permutaciones con repetición de n elementos clasificados en $k + 1$ grupos, habiendo x_i del tipo i−ésimo, $i = 1, \ldots, k$ y $n - \sum_{i=1}^{k} x_i$ del grupo $k + 1$−ésimo. Por tanto, el número de ordenaciones posibles es[1]:

$$PR_n^{x_1, \ldots, x_k, n - \sum_{i=1}^{k} x_i} = \frac{n!}{\left(\prod\limits_{i=1}^{k} x_i!\right) \cdot \left(n - \sum\limits_{i=1}^{k} x_i\right)!}$$

Por tanto, tenemos que:

$$P[X = (x_1, \ldots, x_k)] = \frac{n!}{\left(\prod\limits_{i=1}^{k} x_i!\right) \cdot \left(n - \sum\limits_{i=1}^{k} x_i\right)!} \cdot \prod_{i=1}^{k} p_i^{x_i} \cdot \left(1 - \sum_{i=1}^{k} p_i\right)^{n - \sum\limits_{i=1}^{k} x_i}$$

Comprobemos que, efectivamente, esta función de masa de probabilidad es tal. Para ello, es necesario introducir el siguiente lema técnico, descubierto por Leibniz.

[1]Ver sección de Combinatoria en los apuntes de EDIP.

Lema 5.1 (Fórmula de Leibniz). *Sean $p_1, \ldots, p_j$ números reales y $n \in \mathbb{N}$. Entonces, se cumple que:*

$$\left(\sum_{i=1}^{j} p_i\right)^n = \sum_{\sum_{i=1}^{j} x_i = n} \frac{n!}{\prod_{i=1}^{j} x_i!} \cdot \prod_{i=1}^{j} p_i^{x_i}$$

Proposición 5.2. *Sea $X \sim M_k(n, p_1, \ldots, p_k)$. Sean $x_1, \ldots, x_k$ tales que $x_i \in \{0, \ldots, n\}$, $i = 1, \ldots, k$ y $\sum_{i=1}^{k} x_i \leqslant n$. Entonces, la función de masa de probabilidad de X es:*

$$P[X = (x_1, \ldots, x_k)] = \frac{n!}{\left(\prod_{i=1}^{k} x_i!\right) \cdot \left(n - \sum_{i=1}^{k} x_i\right)!} \cdot \prod_{i=1}^{k} p_i^{x_i} \cdot \left(1 - \sum_{i=1}^{k} p_i\right)^{n - \sum_{i=1}^{k} x_i}$$

Demostración. Usaremos para ello la Fórmula de Leibniz identificando $k + 1 = j$ y definiendo $p_{k+1} = 1 - \sum_{i=1}^{k} p_i$ y $x_{k+1} = n - \sum_{i=1}^{k} x_i$:

$$\sum_{\sum_{i=1}^{k} x_i \leqslant n} \frac{n!}{\left(\prod_{i=1}^{k} x_i!\right) \cdot \left(n - \sum_{i=1}^{k} x_i\right)!} \cdot \prod_{i=1}^{k} p_i^{x_i} \cdot \left(1 - \sum_{i=1}^{k} p_i\right)^{n - \sum_{i=1}^{k} x_i} =$$

$$= \sum_{\sum_{i=1}^{k+1} x_i = n} \frac{n!}{\prod_{i=1}^{k+1} x_i!} \cdot \prod_{i=1}^{k+1} p_i^{x_i} = \left(\sum_{i=1}^{k+1} p_i\right)^n = \left(\sum_{i=1}^{k} p_i + 1 - \sum_{i=1}^{k} p_i\right)^n = 1^n = 1$$

Además, como todos los términos de la suma son no negativos, tenemos efectivamente que se trata de una función de masa de probabilidad. □

Calculemos ahora la función generatriz de momentos de la distribución multinomial.

Proposición 5.3. *Sea $X \sim M_k(n, p_1, \ldots, p_k)$. Entonces, la función generatriz de momentos de X es:*

$$M_X(t_1, \ldots, t_k) = \left[\left(\sum_{i=1}^{k} p_i e^{t_i}\right) + \left(1 - \sum_{i=1}^{k} p_i\right)\right]^n \qquad \forall t_1, \ldots, t_k \in \mathbb{R}$$

Demostración. Usaremos la definición de función generatriz de momentos:

$$M_X(t_1,\dots,t_k) = E[e^{t_1X_1+\ldots+t_kX_k}] =$$

$$= \sum_{\sum\limits_{i=1}^{k} x_i \leqslant n} e^{t_1x_1+\ldots+t_kx_k} \cdot \frac{n!}{\left(\prod\limits_{i=1}^{k} x_i!\right)\cdot\left(n-\sum\limits_{i=1}^{k} x_i\right)!} \cdot \prod_{i=1}^{k} p_i^{x_i} \cdot \left(1-\sum_{i=1}^{k} p_i\right)^{n-\sum\limits_{i=1}^{k} x_i} =$$

$$= \sum_{\sum\limits_{i=1}^{k} x_i \leqslant n} \prod_{i=1}^{k} (e^{t_i})^{x_i} \cdot \frac{n!}{\left(\prod\limits_{i=1}^{k} x_i!\right)\cdot\left(n-\sum\limits_{i=1}^{k} x_i\right)!} \cdot \prod_{i=1}^{k} p_i^{x_i} \cdot \left(1-\sum_{i=1}^{k} p_i\right)^{n-\sum\limits_{i=1}^{k} x_i} =$$

$$= \sum_{\sum\limits_{i=1}^{k} x_i \leqslant n} \frac{n!}{\left(\prod\limits_{i=1}^{k} x_i!\right)\cdot\left(n-\sum\limits_{i=1}^{k} x_i\right)!} \cdot \prod_{i=1}^{k} (p_ie^{t_i})^{x_i} \cdot \left(1-\sum_{i=1}^{k} p_i\right)^{n-\sum\limits_{i=1}^{k} x_i} \overset{(*)}{=}$$

$$\overset{(*)}{=} \left[\left(\sum_{i=1}^{k} p_ie^{t_i}\right) + \left(1-\sum_{i=1}^{k} p_i\right)\right]^n$$

donde en $(*)$ hemos usado la Fórmula de Leibniz. □

Calculamos ahora las marginales de la distribución multinomial.

Proposición 5.4. *Dado el vector aleatorio $X \sim M_k(n, p_1, \dots, p_k)$, consideramos un subvector de X $(X_{i_1}, \dots, X_{i_l})$, con $i_1, \dots, i_l \in \{1, \dots, k\}$, siendo $i_p \neq i_q$ si $p \neq q$. Entonces,*

$$(X_{i_1}, \dots, X_{i_l}) \sim M_l(n, p_{i_1}, \dots, p_{i_l}).$$

Observación. Notemos que esta notación no es del todo precisa, pero se emplea para no complicar la notación. La distribución marginal de un subvector aleatorio de l componentes sigue una distribución multinomial con los parámetros n y las probabilidades asociadas a las componentes del subvector.

Demostración. Esto se deduce directamente de la función generatiz de momentos de un subvector de X, que es:

$$\begin{aligned} M_{(X_{i_1},\dots,X_{i_l})}(t_{i_1},\dots,t_{i_l}) &= M_X(0,\dots,0,t_{i_1},0,\dots,0,t_{i_l},0,\dots,0) = \\ &= \left[\left(\sum_{i=1}^{k} p_ie^{t_i}\right) + \left(1-\sum_{i=1}^{k} p_i\right)\right]^n = \\ &= \left[\left(\sum_{j=1}^{l} p_{i_j}e^{t_{i_j}}\right) + \left(1-\sum_{j=1}^{l} p_{i_j}\right)\right]^n \end{aligned}$$

Como esta función generatriz de momentos coincide con la de una distribución multinomial con los parámetros n y las probabilidades asociadas a las componentes del subvector, se tiene que la distribución marginal de un subvector de l componentes de X sigue una distribución multinomial con los parámetros n y las probabilidades asociadas a las componentes del subvector. □

Tenemos que, en particular, para $j \in \{1, \ldots, k\}$, tenemos que:

$$X_j \sim M_1(n, p_j)$$

Su función generatriz de momentos entonces es:

$$M_{X_j}(t) = \left[p_j e^t + (1 - p_j)\right]^n$$

Por la unicidad de la función generatriz de momentos, tenemos que $X_j \sim B(n, p_j)$, como introducimos ya anteriormente.

Calculamos ahora las distribuciones condicionadas de la distribución multinomial.

Proposición 5.5. *Consideramos el vector aleatorio $X \sim M_k(n, p_1, \ldots, p_k)$. Sean los conjuntos* disjuntos $\{i_1, \ldots, i_q\}, \{j_1, \ldots, j_p\} \subset \{1, \ldots, k\}$, *con $i_p \neq i_q$ y $j_p \neq j_q$ si $p \neq q$. Entonces:*

$$(X_{i_1}, \ldots, X_{i_q}) \mid (X_{j_1} = x_{j_1}, \ldots, X_{j_p} = x_{j_p}) \sim M_q\left(n - \sum_{l=1}^{p} x_{j_l}, \frac{p_{i_1}}{1 - \sum\limits_{l=1}^{p} p_{j_l}}, \ldots, \frac{p_{i_q}}{1 - \sum\limits_{l=1}^{p} p_{j_l}}\right)$$

Observación. De nuevo, la notación no es del todo precisa, pero se emplea para no complicar la notación.

La distribución condicionada de un subvector de q componentes de X dado que otro subvector de p componentes de X toma unos valores concretos, sigue una distribución multinomial con los parámetros n menos la suma de los valores que toman las componentes del subvector condicionante, y las probabilidades asociadas a las componentes del subvector condicionado divididas por la probabilidad 1 menos la suma de las probabilidades asociadas a las componentes del subvector condicionante.

En el caso de que condicionemos una componente X_i de X a otra componente X_j de X $(i \neq j)$, tenemos que:

$$X_i \mid X_j = x_j \sim M_1\left(n - x_j, \frac{p_i}{1 - p_j}\right) = B\left(n - x_j, \frac{p_i}{1 - p_j}\right)$$

Veamos ahora que se trata de una distribución reproductiva.

Proposición 5.6 (Distribución Reproductiva - Multinomial)**.** *Fijado $p \in \mathbb{N}$, consideramos p vectores aleatorios $X_1, \ldots, X_p$* independientes *tales que $X_i \sim M_k(n_i, p_1, \ldots, p_k)$, $i = 1, \ldots, p$. Se tiene entonces que:*

$$\sum_{i=1}^{p} X_i \sim M_k\left(\sum_{i=1}^{p} n_i, p_1, \ldots, p_k\right)$$

Demostración. Por ser independientes, tenemos que:

$$M_{\sum\limits_{i=1}^{p} X_i}(t_1, \ldots, t_k) = \prod_{i=1}^{p} M_{X_i}(t_1, \ldots, t_k) = \prod_{i=1}^{p}\left[\left(\sum_{j=1}^{k} p_j e^{t_j}\right) + \left(1 - \sum_{j=1}^{k} p_j\right)\right]^{n_i} =$$

$$= \left[\left(\sum_{j=1}^{k} p_j e^{t_j}\right) + \left(1 - \sum_{j=1}^{k} p_j\right)\right]^{\sum\limits_{i=1}^{p} n_i}$$

Por tanto, la función generatriz de momentos de $\sum_{i=1}^{p} X_i$ coincide con la de una distribución multinomial con los parámetros $\sum_{i=1}^{p} n_i$ y las probabilidades asociadas a las componentes de X_i. □

Veamos ahora la esperanza de la distribución multinomial.

Proposición 5.7. *Sea $X \sim M_k(n, p_1, \dots, p_k)$. Entonces, la esperanza de X es:*

$$E[X] = n\,(p_1, \dots, p_k)$$

Demostración. De la definición de esperanza, si $X = (X_1, \dots, X_k)$, tenemos que:

$$E[X] = (E[X_1], \dots, E[X_k])$$

Como hemos visto anteriormente, para cada $j \in \{1, \dots, k\}$, tenemos que $X_j \sim B(n, p_j)$, por lo que $E[X_j] = n \cdot p_j$. Por tanto, tenemos que:

$$E[X] = (n \cdot p_1, \dots, n \cdot p_k) = n\,(p_1, \dots, p_k)$$

□

Veamos ahora la covarianza de cada par de componentes de la distribución multinomial.

Proposición 5.8. *Sea $X \sim M_k(n, p_1, \dots, p_k)$. Entonces, si $X = (X_1, \dots, X_k)$, se tiene que:*

$$\mathrm{Cov}(X_i, X_j) = -n \cdot p_i p_j \qquad \forall i, j \in \{1, \dots, k\}, i \neq j$$

Demostración. De la definición de covarianza, tenemos que:

$$\mathrm{Cov}(X_i, X_j) = E[X_i \cdot X_j] - E[X_i] \cdot E[X_j]$$

Para calcular $E[X_i \cdot X_j]$, usamos la función generatriz de momentos de X:

$$\begin{aligned}
E[X_i \cdot X_j] &= \left.\frac{\partial^2 M_X(t_1, \dots, t_k)}{\partial t_i \partial t_j}\right|_{t_1 = \ldots = t_k = 0} = \\
&= n(n-1) p_i p_j e^{t_i} e^{t_j} \left.\left[\left(\sum_{i=1}^{k} p_i e^{t_i}\right) + \left(1 - \sum_{i=1}^{k} p_i\right)\right]^{n-2}\right|_{t_1 = \ldots = t_k = 0} = \\
&= n(n-1) p_i p_j \left[\left(\sum_{i=1}^{k} p_i\right) + \left(1 - \sum_{i=1}^{k} p_i\right)\right]^{n-2} = n(n-1) p_i p_j
\end{aligned}$$

Por tanto, como $E[X_i] = n \cdot p_i$, $E[X_j] = n \cdot p_j$, tenemos que:

$$\mathrm{Cov}(X_i, X_j) = n(n-1) p_i p_j - n^2 p_i p_j = -n \cdot p_i \cdot p_j$$

□

5.2. Distribución Normal Bidimensional

Generalizamos ahora la distribución normal a dos dimensiones.

Definición 5.2 (Distribución Normal Bidimensional)**.** Consideramos un vector aleatorio bidimensional continuo $X = (X_1, X_2)$ y los parámetros $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 \in \mathbb{R}^+$ y $\rho \in]-1, 1[$. Consideramos además las siguientes matrices:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$
$$\mu = \begin{pmatrix} \mu_1 & \mu_2 \end{pmatrix}$$

donde $\sigma_i = \sqrt{\sigma_i^2}$, $i = 1, 2$. Diremos que X sigue una distribución normal bidimensional con parámetros $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ y ρ si su función de densidad es:

$$f_X(x) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{(x-\mu)\Sigma^{-1}(x-\mu)^t}{2}\right) \qquad x \in \mathbb{R}^2$$

La notaremos por las siguientes dos formas:

$$X \sim \mathcal{N}_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) \quad \text{o} \quad X \sim \mathcal{N}_2(\mu, \Sigma)$$

donde μ, Σ son las matrices definidas anteriormente.

Veamos que está bien definida. En primer lugar, hemos de calcular $|\Sigma|$:

$$|\Sigma| = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1-\rho^2) > 0$$

Por tanto, Σ es singular y, por tanto, $\exists \Sigma^{-1}$. Calculémosla:

$$\Sigma^{-1} = \frac{1}{|\Sigma|}\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

Calculamos ahora el producto de matrices descrito, considerando $x = (x_1, x_2)$:

$$\begin{aligned}
(x-\mu)\Sigma^{-1}(x-\mu)^t &= \frac{1}{|\Sigma|} \cdot \begin{pmatrix} x_1-\mu_1 & x_2-\mu_2 \end{pmatrix} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} = \\
&= \frac{1}{|\Sigma|} \cdot \begin{pmatrix} x_1-\mu_1 & x_2-\mu_2 \end{pmatrix} \begin{pmatrix} \sigma_2^2(x_1-\mu_1) - \rho\sigma_1\sigma_2(x_2-\mu_2) \\ -\rho\sigma_1\sigma_2(x_1-\mu_1) + \sigma_1^2(x_2-\mu_2) \end{pmatrix} = \\
&= \frac{1}{\sigma_1^2\sigma_2^2 \cdot (1-\rho^2)} \cdot \left[\sigma_2^2(x_1-\mu_1)^2 - 2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2) + \sigma_1^2(x_2-\mu_2)^2\right] = \\
&= \frac{1}{1-\rho^2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]
\end{aligned}$$

Por tanto, la función de densidad de la distribución normal bidimensional es:

$$f_X(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right)$$

Veamos ahora distintas expresiones alternativas para la función de densidad de la distribución normal bidimensional. Multiplicando el término de $\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2$ por $(1-\rho^2+\rho^2)$, el exponente queda:

$$
\begin{aligned}
&-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2(1-\rho^2+\rho^2)-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]=\\
&=-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}-\frac{1}{2(1-\rho^2)}\left[\rho^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]=\\
&=-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}-\frac{1}{2(1-\rho^2)}\left[-\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]^2=\\
&=-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}-\frac{1}{2\sigma_2^2(1-\rho^2)}\left[-\rho\cdot\frac{\sigma_2}{\sigma_1}\cdot(x_1-\mu_1)+(x_2-\mu_2)\right]^2
\end{aligned}
$$

Por tanto, podemos descomponer la función de densidad de la distribución normal bidimensional en dos funciones:

$$
\begin{aligned}
f_X(x_1,x_2)=&\overbrace{\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2\sigma_2^2(1-\rho^2)}\left[-\rho\cdot\frac{\sigma_2}{\sigma_1}\cdot(x_1-\mu_1)+(x_2-\mu_2)\right]^2\right)}^{h_{x_1}(x_2)}\cdot\\
&\cdot\underbrace{\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right)}_{g(x_1)} \qquad (5.1)
\end{aligned}
$$

Tenemos que $g(x_1)$ es la función de densidad de $\mathcal{N}(\mu_1,\sigma_1^2)$ y, para cada x_1, $h_{x_1}(x_2)$ es la función de densidad de $\mathcal{N}(\mu_2+\rho\cdot\frac{\sigma_2}{\sigma_1}\cdot(x_1-\mu_1),\sigma_2^2(1-\rho^2))$.
Esta descomposición es útil, y la emplearemos en diversas demostraciones en el futuro. Se podrá considerar también la descomposición análoga.

Para terminar de aceptar dicha definición, hemos de comprobar que, efectivamente, se trata de una función de densidad. Como $f_X(x)\geqslant 0\ \forall x\in\mathbb{R}^2$, tan solo nos falta por comprobar que:

$$
\begin{aligned}
\int_{\mathbb{R}^2}f_X(x)dx&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f_X(x_1,x_2)dx_2dx_1\overset{(*)}{=}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x_1)h_{x_1}(x_2)dx_2dx_1=\\
&=\int_{-\infty}^{+\infty}g(x_1)\int_{-\infty}^{+\infty}h_{x_1}(x_2)dx_2\ dx_1\overset{(**)}{=}\int_{-\infty}^{+\infty}g(x_1)dx_1\overset{(**)}{=}1
\end{aligned}
$$

donde en $(*)$ hemos empleado la descomposición de la Ecuación 5.1 y en $(**)$ hemos usado que $g(x_1)$ y $h_{x_1}(x_2)$ son funciones de densidad.

Veamos ahora las marginales de la distribución normal bidimensional.

Proposición 5.9. *Sean los parámetros $\mu_1,\mu_2\in\mathbb{R}$, $\sigma_1^2,\sigma_2^2\in\mathbb{R}^+$ y $\rho\in\,]-1,1[$, y sean sus matrices μ y Σ asociadas. Si $(X_1,X_2)\sim\mathcal{N}_2(\mu,\Sigma)$, entonces:*

$$X_1\sim\mathcal{N}(\mu_1,\sigma_1^2)\quad y\quad X_2\sim\mathcal{N}(\mu_2,\sigma_2^2)$$

Demostración. Calculamos la marginal de X_1:

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f_X(x_1, x_2)dx_2 \overset{(*)}{=} \int_{-\infty}^{+\infty} g(x_1)h_{x_1}(x_2)dx_2 = g(x_1)\int_{-\infty}^{+\infty} h_{x_1}(x_2)dx_2 \overset{(**)}{=} g(x_1)$$

donde en $(*)$ hemos empleado la descomposición de la Ecuación 5.1 y en $(**)$ hemos usado que $h_{x_1}(x_2)$ es una función de densidad. Por tanto, como vimos que $g(x_1)$ es la función de densidad de $\mathcal{N}(\mu_1, \sigma_1^2)$, tenemos que $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$. Usando la descomposición análoga, se demuestra que $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. □

Veamos ahora las condicionadas de la distribución normal bidimensional.

Proposición 5.10. *Sean los parámetros* $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 \in \mathbb{R}^+$ *y* $\rho \in]-1, 1[$, *y sean sus matrices* μ *y* Σ *asociadas. Si* $(X_1, X_2) \sim \mathcal{N}_2(\mu, \Sigma)$, *entonces, dado* x_1^*, $x_2^* \in \mathbb{R}$, *se tiene que:*

$$X_1 \mid X_2 = x_2^* \sim \mathcal{N}\left(\mu_1 + \rho \cdot \frac{\sigma_1}{\sigma_2} \cdot (x_2^* - \mu_2), \sigma_1^2(1 - \rho^2)\right)$$
$$X_2 \mid X_1 = x_1^* \sim \mathcal{N}\left(\mu_2 + \rho \cdot \frac{\sigma_2}{\sigma_1} \cdot (x_1^* - \mu_1), \sigma_2^2(1 - \rho^2)\right)$$

Demostración. Dado $x_1^* \in \mathbb{R}$, calculamos la condicionada de X_2:

$$f_{X_2|X_1=x_1^*}(x_2) = \frac{f_X(x_1^*, x_2)}{f_{X_1}(x_1^*)} \overset{(*)}{=} \frac{g(x_1^*)h_{x_1^*}(x_2)}{g(x_1^*)} = h_{x_1^*}(x_2)$$

donde en $(*)$ hemos empleado la descomposición de la Ecuación 5.1. Como $h_{x_1^*}(x_2)$ es la función de densidad de $\mathcal{N}\left(\mu_2 + \rho \cdot \frac{\sigma_2}{\sigma_1} \cdot (x_1^* - \mu_1), \sigma_2^2(1 - \rho^2)\right)$, tenemos que $X_2 \mid X_1 = x_1^* \sim \mathcal{N}\left(\mu_2 + \rho \cdot \frac{\sigma_2}{\sigma_1} \cdot (x_1^* - \mu_1), \sigma_2^2(1 - \rho^2)\right)$.

□

Observación. Una vez llegados aquí, podemos olvidarnos de la descomposición de la función de densidad de la distribución normal bidimensional descrita en la Ecuación 5.1. Esta descomposición en verdad era:

$$f_X(x_1, x_2) = f_{X_1}(x_1)f_{X_2|X_1=x_1}(x_2)$$

Esta descomposición ya es conocida por el lector.

Veamos ahora la función generatriz de momentos de la distribución normal bidimensional.

Proposición 5.11. *Sean los parámetros* $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 \in \mathbb{R}^+$ *y* $\rho \in]-1, 1[$, *y sean sus matrices* μ *y* Σ *asociadas. Si* $(X_1, X_2) \sim \mathcal{N}_2(\mu, \Sigma)$, *entonces, la función generatriz de momentos de* X_1 *y* X_2 *es:*

$$M_X(t) = \exp\left(t\mu^t + \frac{t\Sigma t^t}{2}\right) =$$
$$= \exp\left(t_1\mu_1 + t_2\mu_2 + \frac{t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2t_1t_2\rho\sigma_1\sigma_2}{2}\right) \qquad \forall t = (t_1, t_2) \in \mathbb{R}^2$$

Demostración. Sea $t = (t_1, t_2) \in \mathbb{R}^2$. Calculamos la función generatriz de momentos de X:

$$\begin{aligned}
M_X(t) = E\left[e^{t_1X_1+t_2X_2}\right] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{t_1x_1+t_2x_2} f_X(x_1,x_2)dx_1dx_2 = \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{t_1x_1}e^{t_2x_2} f_{X_2}(x_2) f_{X_1|X_2=x_2}(x_1)dx_1dx_2 = \\
&= \int_{-\infty}^{+\infty} e^{t_2x_2} f_{X_2}(x_2)\left[\int_{-\infty}^{+\infty} e^{t_1x_1} f_{X_1|X_2=x_2}(x_1)dx_1\right]dx_2
\end{aligned}$$

Como $X_1 \mid X_2 = x_2 \sim \mathcal{N}\left(\mu_1 + \rho\cdot\frac{\sigma_1}{\sigma_2}\cdot(x_2-\mu_2), \sigma_1^2(1-\rho^2)\right)$, usando la función generatriz de momentos de la distribución normal evaluada en t_1, tenemos que:

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{+\infty} e^{t_2x_2} f_{X_2}(x_2) M_{X_1|X_2=x_2}(t_1)dx_2 = \\
&= \int_{-\infty}^{+\infty} e^{t_2x_2} f_{X_2}(x_2)\exp\left[t_1\left(\mu_1+\rho\cdot\frac{\sigma_1}{\sigma_2}\cdot(x_2-\mu_2)\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right]dx_2 = \\
&= \exp\left[t_1\left(\mu_1-\rho\cdot\frac{\sigma_1}{\sigma_2}\cdot\mu_2\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right]\int_{-\infty}^{+\infty} e^{t_2x_2} f_{X_2}(x_2)\exp\left[t_1\rho\cdot\frac{\sigma_1}{\sigma_2}\cdot x_2\right]dx_2 = \\
&= \exp\left[t_1\left(\mu_1-\rho\cdot\frac{\sigma_1}{\sigma_2}\cdot\mu_2\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right]\int_{-\infty}^{+\infty} \exp\left[\left(t_2+t_1\rho\cdot\frac{\sigma_1}{\sigma_2}\right)x_2\right] f_{X_2}(x_2)dx_2
\end{aligned}$$

Como $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, usando la función generatriz de momentos de la distribución normal evaluada en $t_2 + t_1\rho\cdot\frac{\sigma_1}{\sigma_2}$, tenemos que:

$$\begin{aligned}
M_X(t) &= \exp\left[t_1\left(\mu_1-\rho\cdot\frac{\sigma_1}{\sigma_2}\cdot\mu_2\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right] M_{X_2}\left(t_2+t_1\rho\cdot\frac{\sigma_1}{\sigma_2}\right) \\
&= \exp\left[t_1\left(\mu_1-\rho\cdot\frac{\sigma_1}{\sigma_2}\cdot\mu_2\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right]\exp\left[\left(t_2+t_1\rho\cdot\frac{\sigma_1}{\sigma_2}\right)\mu_2+\frac{\left(t_2+t_1\rho\cdot\frac{\sigma_1}{\sigma_2}\right)^2\sigma_2^2}{2}\right] \\
&= \exp\left[t_1\mu_1 - \cancel{t_1\rho\cdot\frac{\sigma_1}{\sigma_2}\cdot\mu_2}+\frac{t_1^2\sigma_1^2}{2}-\cancel{\rho^2\cdot\frac{t_1^2\sigma_1^2}{2}}+t_2\mu_2+\cancel{t_1\rho\cdot\frac{\sigma_1}{\sigma_2}\cdot\mu_2}+\frac{t_2^2\sigma_2^2+\cancel{t_1^2\rho^2\sigma_1^2}+2t_1t_2\rho\sigma_1\sigma_2}{2}\right] \\
&= \exp\left[t_1\mu_1+t_2\mu_2+\frac{t_1^2\sigma_1^2+t_2^2\sigma_2^2+2t_1t_2\rho\sigma_1\sigma_2}{2}\right]
\end{aligned}$$

Calculemos ahora el producto de matrices $t\Sigma t^t$:

$$\begin{aligned}
t\Sigma t^t &= \begin{pmatrix} t_1 & t_2 \end{pmatrix}\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} t_1 & t_2 \end{pmatrix}\begin{pmatrix} \sigma_1^2 t_1 + \rho\sigma_1\sigma_2 t_2 \\ \rho\sigma_1\sigma_2 t_1 + \sigma_2^2 t_2 \end{pmatrix} = \\
&= t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2t_1t_2\rho\sigma_1\sigma_2
\end{aligned}$$

Llegando así a las dos expresiones de la función generatriz de momentos de la distribución normal bidimensional. □

Una vez estudiadas las marginales y la función generatriz de momentos, estamos en condiciones de ver por qué los parámetros de la distribución normal bidimensional se denominan $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ y ρ, y por qué se definen las matrices μ y Σ asociadas a estos parámetros. Veamos en primer lugar que, como era de esperar, μ es el vector esperanza de la distribución normal bidimensional.

Proposición 5.12. *Sean los parámetros $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 \in \mathbb{R}^+$ y $\rho \in \left]-1, 1\right[$, y sean sus matrices μ y Σ asociadas. Si $X = (X_1, X_2) \sim \mathcal{N}_2(\mu, \Sigma)$, entonces:*

$$E[X] = \mu = \begin{pmatrix} \mu_1 & \mu_2 \end{pmatrix}$$

Demostración. De la definición de esperanza, tenemos que:

$$E[X] = \begin{pmatrix} E[X_1] & E[X_2] \end{pmatrix}$$

Considerando las marginales, como $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ y $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, tenemos que $E[X_1] = \mu_1$ y $E[X_2] = \mu_2$. Por tanto, tenemos que:

$$E[X] = \begin{pmatrix} \mu_1 & \mu_2 \end{pmatrix} = \mu$$

□

Veamos ahora que la matriz Σ es la matriz de covarianzas de la distribución normal bidimensional.

Proposición 5.13. *Sean los parámetros $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 \in \mathbb{R}^+$ y $\rho \in \left]-1, 1\right[$, y sean sus matrices μ y Σ asociadas. Si $X = (X_1, X_2) \sim \mathcal{N}_2(\mu, \Sigma)$, entonces:*

$$\mathrm{Cov}_X = \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Demostración. De la definición de covarianza, tenemos que:

$$\mathrm{Cov}_X = \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) \end{pmatrix}$$

Como la covarianza de una variable consigo misma es su varianza y la covarianza es simétrica, tenemos que:

$$\mathrm{Cov}_X = \begin{pmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_1, X_2) & \mathrm{Var}[X_2] \end{pmatrix}$$

Considerando las marginales, como $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ y $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, tenemos que $\mathrm{Var}[X_1] = \sigma_1^2$ y $\mathrm{Var}[X_2] = \sigma_2^2$. Para calcular la covarianza de X_1 y X_2, consideramos la función generatriz de momentos de X:

$$\begin{aligned}
E[XY] &= \left.\frac{\partial^2 M_X(t)}{\partial t_1 \partial t_2}\right|_{t_1=t_2=0} = \left.\frac{\partial^2}{\partial t_1 \partial t_2} \exp\left(t_1\mu_1 + t_2\mu_2 + \frac{t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2t_1t_2\rho\sigma_1\sigma_2}{2}\right)\right|_{t_1=t_2=0} = \\
&= \left.\frac{\partial}{\partial t_2} \exp\left(t_1\mu_1 + t_2\mu_2 + \frac{t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2t_1t_2\rho\sigma_1\sigma_2}{2}\right)\left(\mu_1 + t_1\sigma_1^2 + \rho\sigma_1\sigma_2 t_2\right)\right|_{t_1=t_2=0} = \\
&= \exp\left(t_1\mu_1 + t_2\mu_2 + \frac{t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2t_1t_2\rho\sigma_1\sigma_2}{2}\right)\left(\mu_2 + t_2\sigma_2^2 + \rho\sigma_1\sigma_2 t_1\right)\left(\mu_1 + t_1\sigma_1^2 + \rho\sigma_1\sigma_2 t_2\right) + \\
&\quad + \left.\exp\left(t_1\mu_1 + t_2\mu_2 + \frac{t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2t_1t_2\rho\sigma_1\sigma_2}{2}\right)(\rho\sigma_1\sigma_2)\right|_{t_1=t_2=0} = \\
&= \exp(0)\,(\mu_2)(\mu_1) + \exp(0)\,(\rho\sigma_1\sigma_2) = \mu_1\mu_2 + \rho\sigma_1\sigma_2
\end{aligned}$$

Por tanto, tenemos que:

$$\text{Cov}(X_1, X_2) = E[X_1X_2] - E[X_1]E[X_2] = \mu_1\mu_2 + \rho\sigma_1\sigma_2 - \mu_1\mu_2 = \rho\sigma_1\sigma_2$$

Por tanto, hemos demostrado que $\text{Cov}_X = \Sigma$. □

Por último, interpretemos el parámetro ρ de la distribución normal bidimensional.

Proposición 5.14. *Sean los parámetros $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 \in \mathbb{R}^+$ y $\rho \in]-1, 1[$, y sean sus matrices μ y Σ asociadas. Si $X = (X_1, X_2) \sim \mathcal{N}_2(\mu, \Sigma)$, entonces ρ es el coeficiente de correlación de X_1 y X_2.*

$$\rho_{X_1,X_2} = \rho$$

Demostración. De la definición de Coeficiente de Correlación, tenemos que:

$$\rho_{X_1,X_2} = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}[X_1]\,\text{Var}[X_2]}}$$

Usando que Σ es la matriz de covarianzas de X, tenemos que:

$$\rho_{X_1,X_2} = \frac{\rho\sigma_1\sigma_2}{\sqrt{\sigma_1^2\sigma_2^2}} = \rho$$

□

Por tanto, ya hemos interpretado todos los parámetros de la distribución normal bidimensional. Por tanto, de aquí en adelante no es necesario prefijar $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ y ρ para definir la distribución normal bidimensional, sino que podemos definirla directamente con μ y Σ, siendo la primera un vector esperanza y la segunda una matriz de covarianzas. De ahí la notación de la distribución normal bidimensional $\mathcal{N}_2(\mu, \Sigma)$.

Estudiemos ahora las combinaciones lineales de variables aleatorias normales bidimensionales.

Proposición 5.15. *Fijado $q \in \{1, 2\}$, consideramos la matriz $A \in \mathcal{M}_{2\times q}(\mathbb{R})$. Sea también $X = (X_1, X_2) \sim \mathcal{N}_2(\mu, \Sigma)$. Entonces, tenemos que:*

$$Y := XA \sim \mathcal{N}_q(\mu A, A^t\Sigma A)$$

Demostración. Como $A \in \mathcal{M}_{2\times q}(\mathbb{R})$, tenemos que $XA \in \mathcal{M}_{1\times q}(\mathbb{R})$, por lo que Y es un vector aleatorio de dimensión q. Dado $t \in \mathbb{R}^q$, calculamos la función generatriz de momentos de Y:

$$M_Y(t) = E\left[e^{Yt^t}\right] = E\left[e^{XAt^t}\right] = E\left[e^{X(tA^t)^t}\right] = M_X(tA^t) = \exp\left(tA^t\mu^t + \frac{tA^t\Sigma(tA^t)^t}{2}\right) =$$
$$= \exp\left(t(\mu A)^t + \frac{tA^t\Sigma At^t}{2}\right)$$

Por tanto, tenemos que $Y \sim \mathcal{N}_q(\mu A, A^t\Sigma A)$ (Nótese que, si $q = 1$, también se trata de la función generatriz de momentos de la distribución normal unidimensional, donde usamos el producto escalar en $\mathbb{R}$). □

Por último, caractericemos la independencia en el caso de la distribución normal bidimensional.

Proposición 5.16. *Sean los parámetros $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 \in \mathbb{R}^+$ y $\rho \in]-1,1[$, y sean sus matrices μ y Σ asociadas. Si $X = (X_1, X_2) \sim \mathcal{N}_2(\mu, \Sigma)$, entonces:*

$$X_1 \text{ y } X_2 \text{ son independientes} \iff \rho = 0$$

Demostración. Demostramos mediante doble implicación:

$\Longrightarrow$) Supongamos que X_1 y X_2 son independientes. Entonces, vimos en su momento que $\text{Cov}(X_1, X_2) = 0$, por lo que $\rho_{X_1,X_2} = 0$. Además, como $\rho_{X_1,X_2} = \rho$, tenemos que $\rho = 0$.

$\Longleftarrow$) Supongamos que $\rho = 0$. Entonces, la función de densidad de X queda:

$$\begin{aligned} f_X(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right) = \\ &= \underbrace{\frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right)}_{g_1(x_1)} \underbrace{\exp\left(-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right)}_{g_2(x_2)} \end{aligned}$$

De esta forma, tenemos que $f_X(x_1, x_2) = g_1(x_1)g_2(x_2)$ para determinadas funciones g_1 y g_2. Por tanto, X_1 y X_2 son independientes.

□

Notemos que, como ya hemos visto en otros casos, esta equivalencia no es cierta en general para variables aleatorias normales bidimensionales.

6. Teorema Central del Límite

El objetivo de este tema es dar una solución aproximada al problema de encontrar la distribución de probabilidad de las sumas parciales de una sucesión de variables aleatorias. Es decir, dado $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$, se busca encontrar la distribución de probabilidad de la variable aleatoria:

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

Como se trata de sumas parciales (no infinitas), se trata de una transformación medible luego S_n efectivamente también es una variable aleatoria para todo $n \in \mathbb{N}$. Introducimos los siguientes tipos de convergencia.

Definición 6.1 (Convergencia Casi Segura)**.** Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Decimos que la sucesión converge casi seguramente a una variable aleatoria X, que notaremos por $\{X_n\} \xrightarrow{c.s.} X$, si:

$$P\left[\lim_{n\to\infty} X_n = X\right] = 1$$

Análogamente, usando la probabilidad en $(\Omega, \mathcal{A}, P)$, se tiene que:

$$P\left[\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\right] = 1$$

Notemos que, si $\{X_n\} \xrightarrow{c.s.} X$, entonces el conjunto de puntos de convergencia de X_n tiene probabilidad 1. Se tiene además el siguiente lema.

Lema 6.1. *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y X una variable aleatoria. Entonces:*

$$\{X_n\} \xrightarrow{c.s.} X \Longrightarrow \{X_n + c\} \xrightarrow{c.s} X + c \qquad \forall c \in \mathbb{R}$$

Demostración. Tenemos que:

$$\begin{aligned} 1 = P\left[\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\right] &= \\ = P\left[\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) + c = X(\omega) + c\right] &= \\ = P\left[\omega \in \Omega \mid \lim_{n\to\infty} (X_n + c)(\omega) = (X + c)(\omega)\right] & \end{aligned}$$

Por tanto, se tiene que $\{X_n + c\} \xrightarrow{c.s.} X + c$. $\square$

Definición 6.2 (Convergencia en Probabilidad)**.** Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Decimos que la sucesión converge en probabilidad a una variable aleatoria X, notado por $\{X_n\} \xrightarrow{P} X$, si:

$$\forall \varepsilon \in \mathbb{R}^+, \qquad \lim_{n\to\infty} P\left[|X_n - X| \leqslant \varepsilon\right] = 1$$

Notemos que, si $\{X_n\} \xrightarrow{P} X$, entonces la probabilidad de que la diferencia entre X_n y X sea menor que cualquier ε tiende a 1. Es decir, para cierto $n \in \mathbb{N}$ suficientemente grande, la probabilidad de que X_n esté cerca de X es muy alta. De la definición, se obtiene de forma directa el siguiente lema simplemente aplicando la definición de convergencia en probabilidad.

Lema 6.2. *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y X una variable aleatoria. Entonces:*

$$\{X_n\} \xrightarrow{P} X \Longrightarrow \{X_n + c\} \xrightarrow{P} X + c \qquad \forall c \in \mathbb{R}$$

Demostración. Fijado $\varepsilon \in \mathbb{R}^+$, se tiene que:

$$\begin{aligned} P\left[|X_n - X| \leqslant \varepsilon\right] &= P\left[|X_n + c - X - c| \leqslant \varepsilon\right] = \\ &= P\left[|(X_n + c) - (X + c)| \leqslant \varepsilon\right] \end{aligned}$$

Por tanto, si $\{X_n\} \xrightarrow{P} X$, entonces $\{X_n + c\} \xrightarrow{P} X + c$. □

Definición 6.3 (Convergencia en Ley o en Distribución)**.** Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Sean F_{X_n} la función de distribución de X_n para cada $n \in \mathbb{N}$. Decimos que la sucesión converge en ley a una variable aleatoria X con función de distribución F_X, notado por $\{X_n\} \xrightarrow{L} X$, si:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \qquad \forall x \in C(F_X)$$

donde $C(F_X)$ es el conjunto de puntos de continuidad de F_X.

Veamos ahora la relación que hay entre estas tres formas de convergencia.

Proposición 6.3. *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces:*

$$\{X_n\} \xrightarrow{c.s.} X \Longrightarrow \{X_n\} \xrightarrow{P} X \Longrightarrow \{X_n\} \xrightarrow{L} X$$

Introducimos ahora la siguiente proposición, que nos da una condición suficiente para la convergencia en Ley, que será necesaria más adelante.

Proposición 6.4. *Sean $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y X una variable aleatoria definida sobre el mismo espacio de probabilidad. Supongamos además que existen las funciones generatrices de momentos M_{X_n}, M_X de X_n, X respectivamente en un entorno N del origen. Se cumple que:*

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t) \qquad \forall t \in N \Longrightarrow \{X_n\} \xrightarrow{L} X$$

En las siguientes secciones, trataremos condiciones suficientes o necesarias relativas a:

- Convergencia en Ley (Teorema Central del Límite).
- Convergencia en Probabilidad (Ley Débil de los Grandes Números).
- Convergencia Casi Segura (Ley Fuerte de los Grandes Números).

6.1. Leyes de los Grandes Números

Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias *independientes* definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Adicionalmente, sea $\{S_n\}_{n\in\mathbb{N}}$ la variable aleatoria definida como las sumas parciales de la sucesión:

$$S_n = \sum_{i=1}^{n} X_i \qquad \forall n \in \mathbb{N}$$

Se consideran dos sucesiones de números reales $\{a_n\}_{n\in\mathbb{N}}, \{b_n\}_{n\in\mathbb{N}}$ tales que $\{b_n\} \nearrow +\infty$ (es decir, es creciente y diverge positivamente). Se definen entonces las siguientes Leyes.

Definición 6.4 (Ley Débil de los Grandes Números)**.** En las condiciones anteriores, se dice que $\{X_n\}$ satisface la Ley Débil de los Grandes Números respecto a $\{a_n\}, \{b_n\}$ si:

$$\left\{\frac{S_n - a_n}{b_n}\right\} \xrightarrow{P} 0$$

Definición 6.5 (Ley Fuerte de los Grandes Números)**.** En las condiciones anteriores, se dice que $\{X_n\}$ satisface la Ley Fuerte de los Grandes Números respecto a $\{a_n\}, \{b_n\}$ si:

$$\left\{\frac{S_n - a_n}{b_n}\right\} \xrightarrow{c.s.} 0$$

Notemos que, el nombre de convergencia *débil* y *fuerte* se debe a que, como la convergencia casi segura implica la convergencia en probabilidad, si se cumple la Ley Fuerte, entonces también se cumple la Ley Débil.

A continuación se formulan resultados que proporcionan condiciones suficientes o condiciones necesarias y suficientes para que una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ satisfaga las leyes de los grandes números, aunque antes se introducirá el siguiente resultado, muy útil por ser una caracterización de las hipótesis de la mayor parte de los resultados del tema.

Proposición 6.5. *Sean $X_1, \ldots, X_n$ variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces, se tiene que $X_1, \ldots, X_n$ son independientes e idénticamente distribuidas si y solo si:*

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_k}(x_i) \qquad \forall x_1, \ldots, x_n \in \mathbb{R}, k \in \{1, \ldots, n\}$$

Demostración. Demostraremos por doble implicación:

$\Longrightarrow$) Supongamos que $X_1, \ldots, X_n$ son independientes e idénticamente distribuidas. Entonces, por ser independientes, tenemos que:

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) \qquad \forall x_1, \ldots, x_n \in \mathbb{R}$$

Además, por ser idénticamente distribuidas, se tiene que:

$$F_{X_1}(x) = \cdots = F_{X_n}(x) \qquad \forall x \in \mathbb{R}$$

Por tanto, podemos escoger cualquier $k \in \{1, \ldots, n\}$ y se tiene que:

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_k}(x_i) \qquad \forall x_1, \ldots, x_n \in \mathbb{R}, k \in \{1, \ldots, n\}$$

$\Longleftarrow$) Supongamos que se cumple que:

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_k}(x_i) \qquad \forall x_1, \ldots, x_n \in \mathbb{R}, k \in \{1, \ldots, n\}$$

Fijado $x \in \mathbb{R}$, y tomando $x_1 = x_2 = \cdots = x_n = x$, se tiene que:

$$(F_{X_1}(x))^n = (F_{X_2}(x))^n = \cdots = (F_{X_n}(x))^n$$

Tomando la raíz $n-$ésima, como la función de distribución es no-negativa, se tiene que:

$$F_{X_1}(x) = F_{X_2}(x) = \cdots = F_{X_n}(x) \qquad \forall x \in \mathbb{R}$$

Por tanto, las variables aleatorias $X_1, \ldots, X_n$ son idénticamente distribuidas.

Sabiendo esto, tenemos que:

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_k}(x_i) = \prod_{i=1}^{n} F_{X_i}(x_i) \qquad \forall x_1, \ldots, x_n \in \mathbb{R}$$

Por tanto, también tenemos que las variables aleatorias $X_1, \ldots, X_n$ son independientes.

□

6.1.1. Leyes Débiles de los Grandes Números

Proposición 6.6 (Ley Débil de los Grandes Números de Khintchine, 1928)**.** *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas tal que $\exists \mu \in \mathbb{R}$ de forma que $E[X_n] = \mu$ $\forall n \in \mathbb{N}$. Entonces, se cumple que:*

$$\left\{ \frac{S_n - E[S_n]}{n} \right\} \xrightarrow{P} 0$$

De forma directa, por el Lema 6.2, se tiene el siguiente corolario.

Corolario 6.6.1. *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas tal que $\exists \mu \in \mathbb{R}$ de forma que $E[X_n] = \mu$ para algún $n \in \mathbb{N}$. Entonces, se cumple que:*

$$\left\{\frac{S_n}{n}\right\} \xrightarrow{P} \mu$$

Demostración. Tenemos que:

$$E[S_n] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \mu = n\mu$$

donde hemos usado que $E[X_i] = \mu\ \forall i \in \{1, \ldots, n\}$. Por tanto, aplicando el Lema 6.2, se tiene que:

$$\left\{\frac{S_n - E[S_n]}{n}\right\} \xrightarrow{P} 0 \Longrightarrow \left\{\frac{S_n - E[S_n]}{n} + \frac{E[S_n]}{n}\right\} \xrightarrow{P} 0 + \frac{E[S_n]}{n} = \frac{n\mu}{n} = \mu$$

□

Observación. Notemos que, como las variables aleatorias X_n son independientes e idénticamente distribuidas, se tiene que la condición de que $\exists \mu \in \mathbb{R}$ de forma que $E[X_n] = \mu\ \forall n \in \mathbb{N}$ es equivalente a que $\exists \mu \in \mathbb{R}$ de forma que $E[X_n] = \mu$ para algún $n \in \mathbb{N}$, ya que:

$$E[X_i] = E[X_j] \qquad \forall i, j \in \mathbb{N}$$

Históricamente, es relevante el caso en el que la distribución de las variables aleatorias X_n es de Bernoulli, ya que este se descubrió y demostró dos siglos antes. Se introduce entonces como corolario, pero notemos que es un caso particular de los anteriores usando que $X_n \sim B(p)$ y $E[X_n] = p$ para todo $n \in \mathbb{N}$.

Corolario 6.6.2 (Ley Débil de los Grandes Números de Bernoulli, 1713)**.** *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución de Bernoulli de parámetro $p \in\ ]0, 1[$. Es decir, $X_n \sim B(p)\ \forall n \in \mathbb{N}$. Entonces, se cumple que:*

$$\left\{\frac{S_n - E[S_n]}{n}\right\} \xrightarrow{P} 0$$

Corolario 6.6.3. *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución de Bernoulli de parámetro $p \in\ ]0, 1[$. Entonces, se cumple que:*

$$\left\{\frac{S_n}{n}\right\} \xrightarrow{P} p$$

6.1.2. Leyes Fuertes de los Grandes Números

En primer lugar, tenemos el siguiente resultado, cuya demostración no se incluye por exceder a los contenidos de este curso.

Proposición 6.7 (Ley fuerte de Kolmogorov). *Si $\{X_n\}_{n\in\mathbb{N}}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas. Se tiene:*

$$\left\{\frac{S_n - E[S_n]}{n}\right\} \xrightarrow{c.s.} 0 \Longleftrightarrow \exists E[X_n] \text{ para algún } n \in \mathbb{N}$$

Notemos que, como la convergencia casi segura implica la convergencia en probabilidad, tiene hipótesis más simples que la Ley Débil de los Grandes Números de Khintchine y resultados más fuertes (ya que se trata también de una condición necesaria). Además, se tiene el siguiente corolario, análogo al visto para el visto para la Ley Débil de los Grandes Números de Khintchine.

Corolario 6.7.1. *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas. Entonces, se cumple que:*

$$\left\{\frac{S_n}{n}\right\} \xrightarrow{c.s.} E[X_n] \qquad \forall n \in \mathbb{N} \Longleftrightarrow \exists E[X_n] \text{ para algún } n \in \mathbb{N}$$

Demostración. Fijado $n \in \mathbb{N}$, se tiene que:

$$E[S_n] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} E[X_n] = nE[X_n]$$

donde hemos usado que $E[X_i] = E[X_n]$ $\forall i \in \{1, \ldots, n\}$. Del Lema 6.1, se tiene:

$$\left\{\frac{S_n - E[S_n]}{n}\right\} \xrightarrow{c.s.} 0 \Longrightarrow \left\{\frac{S_n - E[S_n]}{n} + \frac{E[S_n]}{n}\right\} \xrightarrow{c.s.} 0 + \frac{E[S_n]}{n} = \frac{nE[X_n]}{n} = E[X_n]$$

□

Históricamente, es relevante el caso en el que la distribución de las variables aleatorias X_n es de Bernoulli, ya que este se descubrió antes. Se introduce entonces como corolario, pero notemos que es un caso particular de los anteriores usando que $X_n \sim B(p)$ y $E[X_n] = p$ para todo $n \in \mathbb{N}$.

Corolario 6.7.2 (Ley Fuerte de los Grandes Números de Borel). *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución de Bernoulli de parámetro $p \in]0,1[$. Es decir, $X_n \sim B(p)$ $\forall n \in \mathbb{N}$. Entonces, se cumple que:*

$$\left\{\frac{S_n - E[S_n]}{n}\right\} \xrightarrow{c.s.} 0$$

Corolario 6.7.3. *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución de Bernoulli de parámetro $p \in]0,1[$. Entonces, se cumple que:*

$$\left\{\frac{S_n}{n}\right\} \xrightarrow{c.s.} p$$

6.2. Problema Central Del Límite

En la presente sección buscamos dar soluciones al Problema Central del Límite. Dicho problema considera una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ independientes definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y su sucesión de sumas parciales $\{S_n\}_{n\in\mathbb{N}}$. Se trata de encontrar condiciones que garanticen las siguientes convergencias de las sumas centradas y tipificadas:

1. Convergencia en Ley a la Distribución Degenerada:
$$\exists E[X_n]\ \forall n \in \mathbb{N} \Longrightarrow \left\{\frac{S_n - E[S_n]}{n}\right\} \xrightarrow{L} 0$$
2. Convergencia en Ley a la Distribución Normal Tipificada:
$$\exists E[X_n^2]\ \forall n \in \mathbb{N} \Longrightarrow \left\{\frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right\} \xrightarrow{L} Z \sim \mathcal{N}(0,1)$$

La solución definitiva a este problema, que se obtuvo en el primer cuarto del siglo XX, establece condiciones necesarias y suficientes para la convergencia en ambos casos. Aquí nos limitamos a exponer una de las soluciones más populares, de enorme importancia por su gran aplicabilidad a diferentes problemas prácticos, el denominado Teorema de Lévy.

Teorema 6.8 (Teorema de Lévy)**.** *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas. Entonces, se cumple el Teorema Central del Límite, es decir:*

1. *Convergencia en Ley a la Distribución Degenerada:*
$$\exists E[X_n]\ \forall n \in \mathbb{N} \Longrightarrow \left\{\frac{S_n - E[S_n]}{n}\right\} \xrightarrow{L} 0$$
2. *Convergencia en Ley a la Distribución Normal Tipificada:*
$$\exists E[X_n^2]\ \forall n \in \mathbb{N} \Longrightarrow \left\{\frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right\} \xrightarrow{L} Z \sim \mathcal{N}(0,1)$$

Notemos que este es el resultado empleado en las aproximaciones que vimos en la Sección **??**. Históricamente, es relevante el caso en el que la distribución de las variables aleatorias X_n es de Bernoulli, ya que este resultado se descubrió antes. Se incluyen como corolarios, ya que es un caso particular del Teorema de Lévy. El caso de la convergencia en Ley a la Distribución Degenerada fue demostrado por Bernoulli, mientras que el caso de la convergencia en Ley a la Distribución Normal fue demostrado por De Moivre y Laplace.

Corolario 6.8.1 (Teorema De Bernouilli)**.** *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución de Bernoulli de parámetro $p \in]0,1[$. Es decir, $X_n \sim B(p)\ \forall n \in \mathbb{N}$. Entonces:*
$$\left\{\frac{S_n - E[S_n]}{n}\right\} \xrightarrow{L} 0$$

Corolario 6.8.2 (Teorema De Moivre-Laplace)**.** *Sea* $\{X_n\}_{n\in\mathbb{N}}$ *una sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución de Bernoulli de parámetro* $p \in]0,1[$*. Es decir,* $X_n \sim B(p)$ $\forall n \in \mathbb{N}$*. Entonces:*

$$\left\{\frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right\} \xrightarrow{L} Z \sim \mathcal{N}(0,1)$$

Buscamos por último entender el Teorema de De Moivre-Laplace. Si $X_n \sim B(p)$ $\forall n \in \mathbb{N}$, por la reproductividad de la binomial tenemos que $S_n \sim B(n,p)$ $\forall n \in \mathbb{N}$. Por tanto, se tiene que:

$$E[S_n] = np \qquad \mathrm{Var}[S_n] = np(1-p)$$

Por tanto, tenemos que:

$$\left\{\frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right\} = \left\{\frac{S_n - np}{\sqrt{np(1-p)}}\right\} \xrightarrow{L} Z \sim \mathcal{N}(0,1) \Longrightarrow S_n \xrightarrow{L} \sqrt{np(1-p)}Z + np$$

No obstante, tenemos que $\sqrt{np(1-p)}Z + np$ es una variable aleatoria normal de media np y varianza $np(1-p)$, es decir:

$$S_n \xrightarrow{L} X \sim \mathcal{N}(np, \sqrt{np(1-p)}Z + np)$$

Por tanto, dada una variable aleatoria $X_n \sim B(n,p)$ que represente el número de éxitos en n ensayos de Bernoulli, consideramos las variables aleatorias X_i que representan los éxitos en cada ensayo. Entonces, repitiendo el proceso, tenemos que:

$$X_n \xrightarrow{L} X \sim \mathcal{N}(np, \sqrt{np(1-p)}Z + np)$$

7. Relaciones de problemas

7.0. Introducción

Ejercicio 1. Se estudian las plantas de una determinada zona donde ha atacado un virus. La probabilidad de que cada planta esté contaminada es 0,35.

1. ¿Cuál es el número esperado de plantas contaminadas en 5 analizadas?
   Sea X la variable aleatoria que representa el número de plantas contaminadas en 5 análisis. Como la probabilidad de que una planta esté contaminada es 0,35, tenemos que sigue una distribución de probabilidad binomial con $n = 5$ y $p = 0{,}35$. Es decir:
   $$X \sim B(5, 0{,}35)$$
   En este caso, como nos piden el número esperado de plantas contaminadas, tenemos que calcular la esperanza:
   $$E[X] = n \cdot p = 5 \cdot 0{,}35 = 1{,}75$$
   Por tanto, y como el número de plantas contaminadas ha de ser un número entero, el número esperado de plantas contaminadas en 5 análisis es 2.

2. Calcular la probabilidad de encontrar entre 2 y 5 plantas contaminadas en 9 exámenes.

   Sea Y la variable aleatoria que representa el número de plantas contaminadas en 9 análisis. De igual forma que en el apartado anterior, sigue una distribución de probabilidad binomial con $n = 9$ y $p = 0{,}35$. Es decir:
   $$Y \sim B(9, 0{,}35)$$
   En este caso, nos piden calcular la probabilidad de encontrar entre 2 y 5 plantas contaminadas. Es decir:
   $$P[2 \leqslant Y \leqslant 5] = F_Y(5) - F_Y(1) = P[Y \leqslant 5] - P[Y \leqslant 1] \stackrel{(*)}{=} 0{,}9464 - 0{,}1211 = 0{,}8253$$
   donde en $(*)$ hemos utilizado la tabla de la distribución binomial.

3. Hallar la probabilidad de encontrar 4 plantas no contaminadas en 6 análisis.

   Sea Z la variable aleatoria que representa el número de plantas contaminadas en 6 análisis. De igual forma que en los apartados anteriores, sigue una distribución de probabilidad binomial con $n = 6$ y $p = 0{,}35$. Es decir:
   $$Z \sim B(6, 0{,}35)$$

En este caso, nos piden calcular la probabilidad de encontrar 4 plantas no contaminadas. Es decir, la probabilidad de que 2 plantas estén contaminadas. Es decir:
$$P[Z=2]=\binom{6}{2}\cdot 0{,}35^2\cdot 0{,}65^4=0{,}328$$

Ejercicio 2. Cada vez que una máquina dedicada a la fabricación de comprimidos produce uno, la probabilidad de que sea defectuoso es 0,01.

1. Si los comprimidos se colocan en tubos de 25, ¿cuál es la probabilidad de que en un tubo todos los comprimidos sean buenos?

   Sea X la variable aleatoria que representa el número de comprimidos correctos en un tubo de 25. Como la probabilidad de que un comprimido sea defectuoso es 0,01, tenemos que sigue una distribución de probabilidad binomial con $n=25$ y $p=1-0{,}01=0{,}99$. Es decir:
   $$X\sim B(25,0{,}99)$$

   En este caso, nos piden calcular la probabilidad de que en un tubo de 25 comprimidos todos sean buenos. Es decir:
   $$P[X=25]=\binom{25}{25}\cdot 0{,}99^{25}\cdot 0{,}01^0\approx 0{,}7778$$

2. Si los tubos se colocan en cajas de 10, ¿cuál es la probabilidad de que en una determinada caja haya exactamente 5 tubos con un comprimido defectuoso?

   En primer lugar, obtenemos cuál es la probabilidad de que un tubo tenga un comprimido defectuoso. Tenemos que:
   $$P[X=24]=\binom{25}{24}\cdot 0{,}99^{24}\cdot 0{,}01^1=\frac{25!}{24!1!}\cdot 0{,}99^{24}\cdot 0{,}01=25\cdot 0{,}99^{24}\cdot 0{,}01\approx 0{,}1964$$

   Sea Y la variable aleatoria que representa el número de tubos con un comprimido defectuoso en una caja de 10. En este caso, tenemos que sigue una distribución de probabilidad binomial con $n=10$ y $p=0{,}1964$. Es decir:
   $$Y\sim B(10,0{,}1964)$$

   En este caso, nos piden calcular la probabilidad de que en una determinada caja haya exactamente 5 tubos con un comprimido defectuoso. Es decir:
   $$P[Y=5]=\binom{10}{5}\cdot 0{,}1964^5\cdot(1-0{,}1964)^5\approx 0{,}02467$$

Ejercicio 3. Un pescador desea capturar un ejemplar de sardina que se encuentra siempre en una determinada zona del mar con probabilidad 0,15. Hallar la probabilidad de que tenga que pescar 10 peces de especies distintas de la deseada antes de:

1. Pescar la sardina buscada.

   Sea X la variable aleatoria que representa el número de peces de especies distintas de la deseada que el pescador tiene que pescar antes de pescar la sardina buscada. Como la probabilidad de que la sardina buscada se encuentre en la zona es 0,15, tenemos que sigue una distribución de probabilidad geométrica con $p = 0{,}15$. Es decir:
   $$X \sim G(0{,}15)$$

   En este caso, nos piden calcular la probabilidad de que tenga que pescar 10 peces de especies distintas de la deseada antes de pescar la sardina buscada. Es decir:

   $$\begin{aligned} P[X = 10] = P[X \leqslant 10] - P[X \leqslant 9] &= (1 - (1 - 0{,}15)^{11}) - (1 - (1 - 0{,}15)^{10}) = \\ &= (1 - 0{,}15)^{10} - (1 - 0{,}15)^{11} = 0{,}85^{10} \cdot (1 - 0{,}85) \approx 0{,}029 \end{aligned}$$

2. Pescar tres ejemplares de la sardina buscada.

   Sea Y la variable aleatoria que representa el número de peces de especies distintas de la deseada que el pescador tiene que pescar antes de pescar tres ejemplares de la sardina buscada. En este caso, tenemos que sigue una distribución de probabilidad negativa binomial con $k = 3$ y $p = 0{,}15$. Es decir:

   $$Y \sim BN(3, 0{,}15)$$

   En este caso, nos piden calcular la probabilidad de que tenga que pescar 10 peces de especies distintas de la deseada antes de pescar tres ejemplares de la sardina buscada. Es decir:

   $$\begin{aligned} P[Y = 10] &= \frac{(10 + 3 - 1)!}{10!(3 - 1)!} \cdot (1 - 0{,}15)^{10} \cdot 0{,}15^3 = \frac{12!}{10!2!} \cdot 0{,}85^{10} \cdot 0{,}15^3 = \\ &= \frac{12 \cdot 11}{2} \cdot 0{,}85^{10} \cdot 0{,}15^3 \approx 0{,}0438 \end{aligned}$$

Ejercicio 4. Un científico necesita 5 monos afectados por cierta enfermedad para realizar un experimento. La incidencia de la enfermedad en la población de monos es siempre del 30 %. El científico examinará uno a uno los monos de un gran colectivo, hasta encontrar 5 afectados por la enfermedad.

1. Calcular el número medio de exámenes requeridos.

   Consideraremos como éxito encontrar un mono afectado por la enfermedad. Sea X la variable aleatoria que representa el número de monos sanos que el científico tiene que examinar antes de encontrar 5 afectados. En este caso, tenemos que sigue una distribución de probabilidad negativa binomial con $k = 5$ y $p = 0{,}3$. Es decir:
   $$X \sim BN(5, 0{,}3)$$

   En este caso, nos piden calcular el número medio de exámenes requeridos. Es decir:
   $$5 + E[X] = 5 + \frac{k(1 - p)}{p} = 5 + \frac{35}{3} = 16.\overline{6}$$

Por tanto, se espera que se tengan que examinar 17 monos para encontrar 5 afectados.

2. Calcular la probabilidad de que encuentre 10 monos sanos antes de encontrar los 5 afectados.

   Tenemos que:

$$\begin{aligned}
P[X=10] &= \frac{(10+5-1)!}{10!(5-1)!}\cdot 0{,}3^5\cdot 0{,}7^{10} = \frac{14!}{10!45!}\cdot 0{,}3^5\cdot 0{,}7^{10} = \\
&= \frac{14\cdot 13\cdot 12\cdot 11}{4\cdot 3\cdot 2}\cdot 0{,}3^5\cdot 0{,}7^{10} \approx 0{,}0687
\end{aligned}$$

3. Calcular la probabilidad de que tenga que examinar por lo menos 20 monos.

   Como 5 de ellos serían afectados, se pide la probabilidad de que tenga que examinar al menos 15 monos sanos. Es decir:

$$P[X\geqslant 15] = 1 - P[X\leqslant 14] = 1 - 0{,}3^5\cdot\sum_{i=0}^{14}\binom{i+4}{4}0{,}7^i \approx 0{,}2822$$

Ejercicio 5. Se capturan 100 peces de un estanque que contiene 10000. Se les marca con una anilla y se devuelven al agua. Transcurridos unos días se capturan de nuevo 100 peces y se cuentan los anillados.

1. Calcular la probabilidad de que en la segunda captura se encuentre al menos un pez anillado.

   Sea X la variable aleatoria que representa el número de peces anillados en la segunda captura. En este caso, tenemos que sigue una distribución de probabilidad hipergeométrica con $N=10000$, $N_1=100$ y $n=100$. Es decir:

$$X\sim H(10000,100,100)$$

   La probabilidad de que en la segunda captura se encuentre al menos un pez anillado es:

$$\begin{aligned}
P[X\geqslant 1] &= 1-P[X=0] = 1-\frac{\binom{100}{0}\cdot\binom{9900}{100}}{\binom{10000}{100}} = \\
&= 1-\frac{1\cdot\dfrac{9900!}{100!9800!}}{\dfrac{10000!}{100!9900!}} = 1-\frac{9900!\cdot 100!\cdot 9900!}{10000!\cdot 100!\cdot 9800!}
\end{aligned}$$

   Como podemos ver, calcular dicha probabilidad de esta forma es complicado debido a la cantidad de factoriales que hay que calcular. Por ello, aproximaremos la distribución hipergeométrica a una binomial, tomando $n=100$ y $p=\dfrac{N_1}{N}=\dfrac{100}{10000}=0{,}01$. Es decir:

$$X\sim B(100,0{,}01)$$

Por lo que la probabilidad de que en la segunda captura se encuentre al menos un pez anillado es:

$$P[X \geqslant 1] = 1 - P[X = 0] = 1 - \binom{100}{0} \cdot 0{,}01^0 \cdot 0{,}99^{100} \approx 1 - 0{,}366 = 0{,}634$$

2. Calcular el número esperado de peces anillados en la segunda captura.

   El número esperado de peces anillados en la segunda captura es:

   $$E[X] = n \cdot \frac{N_1}{N} = 100 \cdot \frac{100}{10000} = 1$$

   Usando la aproximaxión a la binomial, tenemos que:

   $$E[X] = n \cdot p = 100 \cdot 0{,}01 = 1$$

   Efectivamente vemos que el resultado en ambos casos coincide.

Ejercicio 6. Cada página impresa de un libro tiene 40 líneas, y cada línea tiene 75 posiciones de impresión. Se supone que la probabilidad de que en cada posición haya error es $1/6000$.

1. ¿Cuál es la distribución del número de errores por página?

   Sea X la variable aleatoria que representa el número de errores por página. Tenemos que hay $n = 40 \cdot 75 = 3000$ posiciones de impresión por página; y la probabilidad de que en cada posición haya error es $p = 1/6000$. Por lo que sigue una distribución de probabilidad binomial con:

   $$X \sim B(3000, 1/6000)$$

   Tenemos además que X se puede aproximar a una distribución de Poisson con $\lambda = np = 3000 \cdot 1/6000 = 0{,}5$. Es decir:

   $$X \sim \mathcal{P}(0{,}5)$$

2. Calcular la probabilidad de que una página no contenga errores y de que contenga como mínimo 5 errores.

   En primer lugar, calculamos la probabilidad de que una página no contenga errores. Es decir:

   $$P[X = 0] = e^{-0{,}5} \cdot \frac{0{,}5^0}{0!} = e^{-0{,}5} \approx 0{,}6065$$

   Por otro lado, calculamos la probabilidad de que una página contenga como mínimo 5 errores. Es decir:

   $$P[X \geqslant 5] = 1 - P[X \leqslant 4] \stackrel{(*)}{=} 1 - 0{,}998 = 0{,}002$$

   donde en $(*)$ hemos utilizado la tabla de la distribución de Poisson.

3. ¿Cuál es la probabilidad de que un capítulo de 20 páginas no contenga errores?

   Sea Y la variable aleatoria que representa el número de páginas sin errores en un capítulo de 20 páginas. En este caso, tenemos que sigue una distribución de probabilidad binomial con $n = 20$ y $p = 0{,}6065$ (calculada en el apartado anterior). Es decir:
$$Y \sim B(20, 0{,}6065)$$

   En este caso, nos piden calcular la probabilidad de que un capítulo de 20 páginas no contenga errores. Es decir:

$$P[Y = 20] = \binom{20}{20} \cdot 0{,}6065^{20} \cdot (1 - 0{,}6065)^0 \approx 4{,}53 \cdot 10^{-5}$$

Ejercicio 7. En un departamento de control de calidad se inspeccionan las unidades terminadas que provienen de una línea de ensamble. La probabilidad de que cada unidad sea defectuosa es 0,05.

1. ¿Cuál es la probabilidad de que la vigésima unidad inspeccionada sea la segunda que se encuentra defectuosa?

   Sea X la variable aleatoria que representa el número de unidades correctas inspeccionadas hasta encontrar la segunda defectuosa. En este caso, tenemos que sigue una distribución de probabilidad binomial negativa con $k = 2$ y $p = 0{,}05$. Es decir:
$$X \sim BN(2, 0{,}05)$$

   Como buscamos la probabilidad de que la vigésima unidad inspeccionada sea la segunda que se encuentra defectuosa, hemos de calcular la probabilidad de encontrar 18 unidades no defectuosas antes de encontrar la segunda defectuosa. Es decir:

$$\begin{aligned} P[X = 18] &= \frac{(18 + 2 - 1)!}{18!(2 - 1)!} \cdot 0{,}05^2 \cdot 0{,}95^{18} = \frac{19!}{18!1!} \cdot 0{,}05^2 \cdot 0{,}95^{18} = \\ &= 19 \cdot 0{,}05^2 \cdot 0{,}95^{18} \approx 0{,}0188 \end{aligned}$$

2. ¿Cuántas unidades deben inspeccionarse por término medio hasta encontrar cuatro defectuosas?

   Sea Y la variable aleatoria que representa el número de unidades correctas inspeccionadas hasta encontrar cuatro defectuosas. En este caso, tenemos que sigue una distribución de probabilidad binomial negativa con $k = 4$ y $p = 0{,}05$. Es decir:
$$Y \sim BN(4, 0{,}05)$$

   En este caso, nos piden calcular el número medio de unidades que deben inspeccionarse hasta encontrar cuatro defectuosas. Es decir:

$$4 + E[Y] = 4 + \frac{k(1 - p)}{p} = 4 + \frac{4 \cdot 0{,}95}{0{,}05} = 4 + 76 = 80$$

3. Calcular la desviación típica del número de unidades inspeccionadas hasta encontrar cuatro defectuosas.

   Tenemos que:

$$\text{Var}[Y+4] \overset{(*)}{=} \text{Var}[Y] = \frac{k(1-p)}{p^2} = \frac{4\cdot 0{,}95}{0{,}05^2} = 1520$$

   donde en $(*)$ hemos usado que:

$$\text{Var}[aX+b] = a^2\,\text{Var}[X] \qquad \forall a,b\in\mathbb{R}$$

   Por tanto:

$$\sigma_{Y+4} = \sqrt{1520} \approx 38{,}98$$

Ejercicio 8. Los números $1,2,3,...,10$ se escriben en diez tarjetas y se colocan en una urna. Las tarjetas se extraen una a una y sin devolución. Calcular las probabilidades de los siguientes sucesos:

1. Hay exactamente tres números pares en cinco extracciones.

   Sea X la variable aleatoria que representa el número de números pares en cinco extracciones. En este caso, tenemos que sigue una distribución de probabilidad hipergeométrica con $N=10$, $N_1=5$ y $n=5$. Es decir:

$$X \sim H(10,5,5)$$

   La probabilidad de que haya exactamente tres números pares en cinco extracciones es:

$$\begin{aligned} P[X=3] &= \frac{\binom{5}{3}\cdot\binom{5}{2}}{\binom{10}{5}} = \frac{\frac{5!}{3!2!}\cdot\frac{5!}{2!3!}}{\frac{10!}{5!5!}} = \frac{(5!)^4}{10!(3!)^2(2!)^2} = \frac{5^4\cdot 4^4\cdot 3^4\cdot 2^4}{10!\cdot 3^2\cdot 2^4} = \\ &= \frac{5^4\cdot 3^4\cdot 2^{12}}{5^2\cdot 3^6\cdot 2^{12}\cdot 7} = \frac{5^2}{7\cdot 3^2} = \frac{25}{63} \approx 0{,}3968 \end{aligned}$$

2. Se necesitan cinco extracciones para obtener tres números pares.

   Definimos dos sucesos distintos:

   - A: Sacar 2 números pares en las primeras cuatro extracciones.
   - B: Sacar un número par en la quinta extracción.

   Nos piden calcular la probabilidad de que ocurra tanto A como B, es decir, $P(A\cap B)$. Tenemos que:

$$P(A\cap B) = P(A)\cdot P(B \mid A)$$

   Calculemos ambas probabilidades por separado:

- Para calcular $P(A)$, sea Y la variable aleatoria que representa el número de números pares en las primeras cuatro extracciones. En este caso, tenemos que sigue una distribución de probabilidad hipergeométrica con $N = 10$, $N_1 = 5$ y $n = 4$. Es decir:

$$Y \sim H(10, 5, 4)$$

  La probabilidad de que haya exactamente dos números pares en las primeras cuatro extracciones es:

$$P(A) = P[Y = 2] = \frac{\binom{5}{2} \cdot \binom{5}{2}}{\binom{10}{4}} = \approx 0{,}476$$

- Para calcular $P(B \mid A)$, usamos la Regla de Laplace. Como se habrán sacado 2 números pares en las primeras cuatro extracciones, en la quinta extracción habrá 3 números pares de un total de 6 candidatos, es decir:

$$P(B \mid A) = \frac{3}{6} = 0{,}5$$

Por tanto, la probabilidad de que se necesiten cinco extracciones para obtener tres números pares es:

$$P(A \cap B) = P(A) \cdot P(B \mid A) = 0{,}476 \cdot 0{,}5 = 0{,}238$$

3. Obtener el número 7 en la cuarta extracción.

Definimos los siguientes sucesos:

- C: No sacar un número 7 en las tres primeras extracciones.
- D: Sacar el número 7 en la cuarta extracción.

Nos piden calcular la probabilidad de que ocurra tanto C como D, es decir, $P(C \cap D)$. Tenemos que:

$$P(C \cap D) = P(C) \cdot P(D \mid C)$$

Calculemos ambas probabilidades por separado:

- Para calcular $P(C)$, sea Z la variable aleatoria que representa el número de extracciones distintas a 7 en las tres primeras extracciones. En este caso, tenemos que sigue una distribución de probabilidad hipergeométrica con $N = 10$, $N_1 = 9$ y $n = 3$. Es decir:

$$Z \sim H(10, 9, 3)$$

  La probabilidad de que no haya un número 7 en las tres primeras extracciones es:

$$P(C) = P[Z = 3] = \frac{\binom{9}{3} \cdot \binom{1}{0}}{\binom{10}{3}} = \frac{7}{10}$$

- Para calcular $P(D \mid C)$, usamos la Regla de Laplace. Como no se ha sacado un número 7 en las tres primeras extracciones, en la cuarta extracción habrá un número 7 de un total de 7 candidatos, es decir:

$$P(D \mid C) = \frac{1}{7}$$

Por tanto, la probabilidad de obtener el número 7 en la cuarta extracción es:

$$P(C \cap D) = P(C) \cdot P(D \mid C) = \frac{7}{10} \cdot \frac{1}{7} = \frac{1}{10} = 0{,}1$$

Ejercicio 9. Supongamos que el número de televisores vendidos en un comercio durante un mes se distribuye según una Poisson de parámetro 10, y que el beneficio neto por unidad es 30 euros.

1. ¿Cuál es la probabilidad de que el beneficio neto obtenido por un comerciante durante un mes sea al menos de 360 euros?

   Sea X la variable aleatoria que representa el número de televisores vendidos en un mes. En este caso, tenemos que sigue una distribución de probabilidad de Poisson con $\lambda = 10$. Es decir:

   $$X \sim \mathcal{P}(10)$$

   Sabemos que el beneficio neto por unidad es de 30 euros, por lo que para obtener un beneficio neto de al menos 360 euros, el comerciante debe vender al menos 12 televisores. Por lo que la probabilidad de que el beneficio neto obtenido por un comerciante durante un mes sea al menos de 360 euros es:

   $$P[X \geqslant 12] = 1 - P[X \leqslant 11] \overset{(*)}{=} 1 - 0{,}6968 = 0{,}3032$$

   donde en $(*)$ hemos utilizado la tabla de la distribución de Poisson.

2. ¿Cuántos televisores debe tener el comerciante a principio de mes para tener al menos probabilidad 0,95 de satisfacer toda la demanda?

   Se pide el menor valor de $\widehat{x} \in \mathbb{N}$ tal que:

   $$P[X \leqslant \widehat{x}] \geqslant 0{,}95$$

   Para resolverlo, buscamos el valor en la tabla de la distribución de Poisson que cumpla la condición. En este caso, el valor que cumple la condición es 15. Por tanto, el comerciante debe tener al menos 15 televisores a principio de mes para tener al menos probabilidad 0,95 de satisfacer toda la demanda.

Ejercicio 10. Sea el experimento de lanzar un dado de 6 caras. Obtener:

1. Función masa de probabilidad. Sea X la variable aleatoria que representa el número de la cara que sale en el dado. El espacio muestral del experimento es:

   $$\Omega = \{1, 2, 3, 4, 5, 6\}$$

   Entonces, la función masa de probabilidad de X, notada por $P_X : \Omega \to [0, 1]$, es:

   $$P_X(x) = P[X = x] = \frac{1}{6}, \quad x \in \Omega$$

2. Función de distribución.

   La función de distribución de X, notada por $F_X : \mathbb{R} \to [0,1]$, es:

$$F_X(x) = P[X \leqslant x] = \begin{cases} 0, & x < 1 \\ \frac{\lfloor x \rfloor}{6}, & 1 \leqslant x < 2 \\ 1, & x \geqslant 6 \end{cases}$$

3. Función generatriz de momentos.

   La función generatriz de momentos de X, notada por $M_X : \mathbb{R} \to \mathbb{R}$, es:

$$M_X(t) = E[e^{tX}] = \frac{1}{6}\sum_{x=1}^{6} e^{tx}$$

4. Valor esperado.

   El valor esperado de X, también conocido como la esperanza, es:

$$E[X] = \sum_{i=1}^{6} i \cdot P[X = i] = \frac{1}{6}\sum_{i=1}^{6} i = \frac{1}{6} \cdot \frac{6 \cdot 7}{2} = \frac{7}{2} = 3{,}5$$

   También se podría calcular como la primera derivada de la función generatriz de momentos evaluada en $t = 0$:

$$\begin{aligned} E[X] = M_X'(t)\Big|_{t=0} &= \frac{d}{dt}\left(\frac{1}{6}\sum_{x=1}^{6} e^{tx}\right)\Bigg|_{t=0} \\ &= \frac{1}{6}\sum_{x=1}^{6} \frac{d}{dt}\left(e^{tx}\right)\Big|_{t=0} = \frac{1}{6}\sum_{x=1}^{6} xe^{tx}\Big|_{t=0} = \frac{1}{6}\sum_{x=1}^{6} x = \frac{7}{2} = 3{,}5 \end{aligned}$$

   Como podemos ver, ambos métodos coinciden.

5. Varianza.

   La varianza de X es:

$$Var[X] = E[X^2] - E[X]^2 = \frac{1}{6}\sum_{i=1}^{6} i^2 - \left(\frac{7}{2}\right)^2 \approx 2{,}9167$$

6. La distribución de probabilidad que sigue el experimento.

   Tenemos que la variable aleatoria X sigue una distribución uniforme discreta.

$$X \sim \mathcal{U}(1, \ldots, 6)$$

Ejercicio 11. Consideramos la variable aleatoria X que representa el número de caras menos número de cruces obtenidas al lanzar tres monedas. Se pide:

1. Espacio muestral del experimento.

   El espacio muestral del experimento es, representando con C a una cara y X a una cruz:
$$\Omega = \{CCC, CCX, CXC, CXX, XCC, XCX, XXC, XXX\}$$
   Además, tenemos que:
$$X(\Omega) = \{-3, -1, 1, 3\}$$

2. Función masa de probabilidad.

   Usando la Regla de Laplace, y procesando cada una de las $8 = 2^3$ opciones, tenemos:
$$P[X = -3] = P[X = 3] = \frac{1}{2^3} = \frac{1}{8}$$
$$P[X = -1] = P[X = 1] = \frac{3}{2^3} = \frac{3}{8}$$

3. Valor esperado.

   El valor esperado de X es:
$$E[X] = \sum_{x\in\Omega} x \cdot P[X = x] = -3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} - 1 \cdot \frac{3}{8} + 1 \cdot \frac{3}{8} = 0$$

4. Varianza.

   Tenemos que:
$$\text{Var}[X] = E[X^2]-E[X]^2 = \sum_{x\in\Omega} x^2 \cdot P[X = x]-0 = 2\cdot 3^2\cdot\frac{1}{8}+2\cdot 1^2\cdot\frac{3}{8} = \frac{9}{4}+\frac{3}{4} = 3$$

5. Función de distribución.

   La función de distribución de X es:
$$F_X(x) = P[X \leqslant x] = \begin{cases} 0, & x < -3 \\ 1/8, & -3 \leqslant x < -1 \\ 1/2, & -1 \leqslant x < 1 \\ 7/8, & 1 \leqslant x < 3 \\ 1, & x \geqslant 3 \end{cases}$$

6. Probabilidad de que X sea positiva.

   Tenemos que:
$$P[X > 0] = 1 - P[X \leqslant 0] = 1 - 1/2 = 1/2$$

Ejercicio 12. Dado $k \in \mathbb{R}$, sea X una variable aleatoria con función de densidad dada por:
$$f_X(x) = \begin{cases} k/x^2 & 1 \leqslant x \leqslant 8 \\ 0, & \text{en otro caso} \end{cases}$$

Obtener:

1. El valor de k.

   Para obtener el valor de k, tenemos que:

   $$1 = \int_{-\infty}^{\infty} f_X(x)dx = \int_1^8 \frac{k}{x^2}dx = k \cdot \left[-\frac{1}{x}\right]_1^8 = k \cdot \left(-\frac{1}{8} + 1\right) = k \cdot \frac{7}{8}$$

   Por tanto, tenemos que:

   $$k = \frac{8}{7}$$

   Además, con dicho valor de k, tenemos que $f_X(x) \geqslant 0$ para todo $x \in \mathbb{R}$, por lo que la función de densidad es válida.

2. La función de distribución:

   La función de distribución de X es:

   $$F_X(x) = P[X \leqslant x] = \begin{cases} 0, & x < 1 \\ \displaystyle\int_1^x \frac{8}{7x^2}dx = \left[-\frac{8}{7x}\right]_1^x = -\frac{8}{7x} + \frac{8}{7}, & 1 \leqslant x < 8 \\ 1, & x \geqslant 8 \end{cases}$$

3. Valor esperado.

   El valor esperado de X es:

   $$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)dx = \int_1^8 x \cdot \frac{8}{7x^2}dx = \frac{8}{7}\int_1^8 \frac{1}{x}dx = \frac{8}{7}\left[\ln|x|\right]_1^8 = \frac{8}{7}\ln(8)$$

4. Varianza.

   En primer lugar, calculamos $E[X^2]$:

   $$E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)dx = \int_1^8 x^2 \cdot \frac{8}{7x^2}dx = \frac{8}{7}\int_1^8 dx = \frac{8}{7} \cdot 7 = 8$$

   Por tanto, la varianza de X es:

   $$\mathrm{Var}[X] = E[X^2] - E[X]^2 = 8 - \frac{64}{49} \cdot \ln(8)^2 \approx 2{,}352$$

Ejercicio 13. Una gasolinera vende una cantidad X (medida en miles) de litros de gasolina en un día. Supongamos que X tiene la siguiente función de densidad:

$$f_X(x) = \begin{cases} {}^3\!/\!_8 \cdot x^2 & 0 \leqslant x \leqslant 2 \\ 0, & \text{en otro caso} \end{cases}$$

Las ganancias de la gasolinera son 100 euros por cada mil litros de gasolina vendidos si la cantidad vendida es menor o igual a 1000 litros. Además, gana 40 euros extra (por cada 1000 litros) si vende por encima de dicha cantantidad. Calcule la ganancia esperada de la gasolinera en un día.

Tenemos que la función que mide la ganancia de la gasolinera en función de la cantidad de litros vendidos es:

$$G(x) = \begin{cases} 100 \cdot x, & 0 \leqslant x \leqslant 1 \\ 140 \cdot x & 1 < x \leqslant 2 \end{cases}$$

Calculamos el valor esperado de la ganancia de la gasolinera en un día:

$$\begin{aligned} E[G(X)] &= \int_{-\infty}^{\infty} G(x) \cdot f_X(x)dx = \int_0^2 G(x) \cdot \frac{3}{8}x^2 dx = \\ &= \int_0^1 100x \cdot \frac{3}{8}x^2 dx + \int_1^2 140x \cdot \frac{3}{8}x^2 dx = \\ &= \frac{3}{8}\int_0^1 100x^3 dx + \frac{3}{8}\int_1^2 140x^3 dx = \frac{3}{8}\left[25x^4\right]_0^1 + \frac{3}{8}\left[35x^4\right]_1^2 = \\ &= \frac{3}{8} \cdot 25 + \frac{3}{8} \cdot 35 \cdot 15 = \frac{825}{4} = 206{,}25 \end{aligned}$$

Observación. Notemos que piden el valor esperado de la ganancia, no la ganancia del valor esperado. Para calcular este último, habría que calcular el valor esperado de X y luego aplicar la función G.

Calculemos el valor esperado de X, que es la cantidad de litros vendidos en un día:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)dx = \int_0^2 x \cdot \frac{3}{8}x^2 dx = \frac{3}{8}\int_0^2 x^3 dx = \frac{3}{8} \cdot \left[\frac{1}{4}x^4\right]_0^2 = \frac{3}{8} \cdot \frac{1}{4} \cdot 2^4 = \frac{3}{2}$$

Tenemos por tanto que la ganancia del valor esperado es:

$$G(E[X]) = \frac{3}{2} \cdot 140 = 210$$

Notemos que ambos conceptos no son iguales.

7.1. Distribuciones de Probabilidad Continua

Ejercicio 7.1.1. La llegada de viajeros a una estación de tren se distribuye uniformemente en el tiempo. Cada 20 minutos se produce la salida del tren. Hallar:

1. La función de distribución de la variable aleatoria tiempo de espera, su media y su varianza.

   Sea X la variable aleatoria que representa el tiempo de espera de un viajero. Entonces, $X \sim \mathcal{U}(0, 20)$. La función de distribución de X es:

   $$F_X(x) = \begin{cases} 0 \text{ si } x < 0 \\ \dfrac{x}{20} \text{ si } 0 \leqslant x < 20 \\ 1 \text{ si } x \geqslant 20 \end{cases}$$

   La media de X es:

   $$E[X] = \int_0^{20} \frac{x}{20}\, dx = \left[\frac{x^2}{40}\right]_0^{20} = 10$$

   Para hallar la varianza de X, primero calculamos la esperanza de X^2:

   $$E[X^2] = \int_0^{20} \frac{x^2}{20}\, dx = \left[\frac{x^3}{60}\right]_0^{20} = \frac{400}{3}$$

   Por tanto, la varianza de X es:

   $$\mathrm{Var}(X) = E[X^2] - E[X]^2 = \frac{400}{3} - 100 = \frac{100}{3} \approx 33{,}33$$

2. La probabilidad de que un viajero espere al tren menos de 7 minutos.

   Queremos calcular entonces $P[X < 7]$. Tenemos que:

   $$P[X < 7] = F_X(7) = \frac{7}{20} = 0{,}35$$

Ejercicio 7.1.2. La temperatura media diaria en una región se distribuye según una normal con media 25 grados centígrados y desviación típica 10 grados centígrados.

1. Calcular la probabilidad de que en un día elegido al azar la temperatura media esté comprendida entre 20 y 32 grados centígrados.

   Sea X la variable aleatoria que representa la temperatura media diaria en una región. Entonces, $X \sim \mathcal{N}(25, 10^2)$. Queremos calcular entonces la siguiente probabilidad:

   $$\begin{aligned} P[20 \leqslant X \leqslant 32] &= P[X \leqslant 32] - P[X \leqslant 20] = \\ &= P\left[Z \leqslant \frac{32-25}{10}\right] - P\left[Z \leqslant \frac{20-25}{10}\right] = \\ &= P\left[Z \leqslant 0{,}7\right] - P\left[Z \leqslant -0{,}5\right] = P[Z \leqslant 0{,}7] - 1 + P[Z \leqslant 0{,}5] \approx \\ &\approx 0{,}75804 - 1 + 0{,}69146 \approx 0{,}4495 \end{aligned}$$

2. Calcular la probabilidad de que en un día elegido al azar la temperatura media difiera de la media de las temperaturas medias diarias más de 5 grados centígrados.

   Queremos calcular entonces $P[|X-\mu| > 5]$. Resolvamos primero dicha inecuación:

$$\begin{aligned} |X-\mu| > 5 &\Longleftrightarrow X-\mu > 5 \lor X-\mu < -5 \\ &\Longleftrightarrow X > 30 \lor X < 20 \end{aligned}$$

   Por tanto, tenemos que:

$$\begin{aligned} P[|X-\mu| > 5] &= P[X > 30] + P[X < 20] = \\ &= 1 - P[X \leqslant 30] + P[X \leqslant 20] = \\ &= 1 - P\left[Z \leqslant \frac{30-25}{10}\right] + P\left[Z \leqslant \frac{20-25}{10}\right] = \\ &= 1 - P[Z \leqslant 0{,}5] + P[Z \leqslant -0{,}5] = \\ &= 1 - P[Z \leqslant 0{,}5] + 1 - P[Z \leqslant 0{,}5] = \\ &= 2 \cdot (1 - P[Z \leqslant 0{,}5]) = \\ &= 2 \cdot (1 - 0{,}69146) = 2 \cdot 0{,}30854 = 0{,}61708 \end{aligned}$$

Ejercicio 7.1.3. De una variable aleatoria uniformemente distribuida se conoce su esperanza, μ, y su desviación típica, σ. Hallar el rango de valores de la variable, en función de μ y σ.

Sea $a, b \in \mathbb{R}$ tales que $a < b$. Entonces, la variable aleatoria X que se distribuye uniformemente en el intervalo $[a, b]$, notada por $X \sim \mathcal{U}(a, b)$, tiene esperanza y varianza:

$$\begin{aligned} E[X] &= \frac{a+b}{2} = \mu \\ \mathrm{Var}(X) &= \frac{(b-a)^2}{12} = \sigma^2 \end{aligned}$$

Por tanto, para hallar el rango de valores de la variable en función de μ y σ, despejamos a y b de las ecuaciones anteriores:

$$\begin{cases} a + b = 2\mu \\ b - a = \sqrt{12}\sigma \end{cases}$$

Sumándolas, vemos que:

$$\begin{aligned} b &= \mu + \sqrt{3}\sigma \\ a &= \mu - \sqrt{3}\sigma \end{aligned}$$

Ejercicio 7.1.4. Los precios de venta de un artículo se distribuyen según una ley normal. Se sabe que el 20 % son superiores a 1000 euros y que el 30 % no superan los 800 euros. Hallar la ganancia media y su desviación típica, si las ganancias (Y)

están relacionadas con los precios (X) según la expresión $Y = 350 + 0{,}15X$.

Sea X la variable aleatoria que representa el precio de venta de un artículo. Entonces, $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$. Calculemos μ_X y σ_X^2 a partir de los datos proporcionados.

Por un lado, tenemos que:

$$P[X > 1000] = 0{,}2 \Longrightarrow 1 - P[X \leqslant 1000] = 0{,}2 \Longrightarrow P[X \leqslant 1000] = 0{,}8 \Longrightarrow$$
$$\Longrightarrow P\left[Z \leqslant \frac{1000 - \mu_X}{\sigma_X}\right] = 0{,}8$$

Consultando la tabla, vemos que:

$$\left\{\begin{array}{l} P[Z \leqslant 0{,}84] = 0{,}79955 \\ P[Z \leqslant 0{,}85] = 0{,}80234 \end{array}\right\} \Longrightarrow 0{,}84 + \frac{0{,}85 - 0{,}84}{0{,}80234 - 0{,}79955} \cdot (0{,}8 - 0{,}79955) = 0{,}8416 \Longrightarrow$$
$$\Longrightarrow \frac{1000 - \mu_X}{\sigma_X} = 0{,}8416$$

Por otro lado, tenemos que:

$$P[X \leqslant 800] = 0{,}3 \Longrightarrow 1 - P[X \leqslant 800] = 0{,}3 \Longrightarrow P[X > 800] = 0{,}7 \Longrightarrow$$
$$\Longrightarrow P\left[Z \geqslant \frac{800 - \mu_X}{\sigma_X}\right] = 0{,}7 \Longrightarrow P\left[Z \leqslant \frac{\mu_X - 800}{\sigma_X}\right] = 0{,}7 \Longrightarrow$$
$$\Longrightarrow \frac{\mu_X - 800}{\sigma_X} = 0{,}53$$

Resolvemos por tanto el sistema de ecuaciones:

$$\left\{\begin{array}{l} 1000 - \mu_X = 0{,}8416\sigma_X \\ \mu_X - 800 = 0{,}53\sigma_X \end{array}\right.$$

Sumando, vemos que:

$$\sigma_X = \frac{1000 - 800}{0{,}8416 + 0{,}53} \approx 145{,}815 \mu_X \qquad = 800 + 0{,}53 \cdot 145{,}815 \approx 877{,}28$$

La ganancia media por tanto es:

$$E[Y] = E[350 + 0{,}15X] = 350 + 0{,}15 \cdot E[X] = 350 + 0{,}15 \cdot 877{,}28 \approx 481{,}59$$

Y la varianza de la ganancia es:

$$\mathrm{Var}[Y] = \mathrm{Var}[350 + 0{,}15X] = 0{,}15^2 \cdot \mathrm{Var}[X] = 0{,}15^2 \cdot 145{,}815^2$$

Por tanto, la desviación típica de la ganancia es:

$$\sigma_Y = \sqrt{0{,}15^2 \cdot 145{,}815^2} \approx 21{,}8722$$

Ejercicio 7.1.5. Se clasifican los cráneos en dolicocéfalos (si el índice cefálico, anchura/longitud, es menor que 75), mesocéfalos (si el índice está entre 75 y 80), y braquicéfalos (si el índice es superior a 80). Suponiendo que la distribución de los índices es normal, hallar la media y la desviación típica en una población en la que el 65 % de los individuos son dolicocéfalos, el 30 % mesocéfalos y el 5 % braquicéfalos.

Sea X la variable aleatoria que representa el índice cefálico de un individuo. Entonces, $X \sim \mathcal{N}(\mu, \sigma^2)$. Sabemos que:

$$\begin{aligned} P[X < 75] &= 0{,}65 \\ P[75 \leqslant X \leqslant 80] &= 0{,}30 \\ P[X > 80] &= 0{,}05 \end{aligned}$$

Por tanto, tenemos que:

$$P[X < 75] = P\left[Z < \frac{75-\mu}{\sigma}\right] = 0{,}65$$

$$P[75 \leqslant X \leqslant 80] = P[X \leqslant 80] - P[X < 75] = P[X \leqslant 80] - 0{,}65 = 0{,}30 \Longrightarrow$$

$$\Longrightarrow P[X \leqslant 80] = P\left[Z \leqslant \frac{80-\mu}{\sigma}\right] = 0{,}95$$

$$P[X > 80] = 1 - P[X \leqslant 80] = 1 - 0{,}95 = 0{,}05$$

Consultando la tabla, veamos qué valores de Z cumplen lo pedido.

- Buscamos $\hat{z}$ tal que $P[Z < \hat{z}] = 0{,}65$. Tras consultar la tabla, vemos:

  $$P[Z < 0{,}38] = 0{,}64803 \qquad P[Z < 0{,}39] = 0{,}65173$$

  Interpolamos por tanto de forma lineal entre amos valores:

  $$0{,}38 + \frac{0{,}39 - 0{,}38}{0{,}65173 - 0{,}64803} \cdot (0{,}65 - 0{,}64803) \approx 0{,}3853$$

  Por tanto, la ecuación que obtenemos es:

  $$\frac{75-\mu}{\sigma} = 0{,}3853$$

- Buscamos $\hat{z}$ tal que $P[Z \leqslant \hat{z}] = 0{,}95$. Tras consultar la tabla, vemos que $\hat{z} = 1{,}65$. Por tanto, tenemos que:

  $$\frac{80-\mu}{\sigma} = 1{,}65$$

Establecemos por tanto el sistema de ecuaciones:

$$\begin{cases} 0{,}3853\sigma + \mu = 75 \\ 1{,}65\sigma + \mu = 80 \end{cases}$$

Restando la primera ecuación a la segunda, obtenemos:

$$\sigma = \frac{5}{1{,}65 - 0{,}3853} \approx 3{,}95 \Longrightarrow \mu \approx 75 - 0{,}3853 \cdot 3{,}95 \approx 73{,}47$$

Por tanto, tenemos que:

$$\begin{aligned} \mu &\approx 73{,}47 \\ \sigma &\approx 3{,}95 \end{aligned}$$

Ejercicio 7.1.6. La probabilidad de contagio por unidad de tiempo viene dada por:

$$P[T \leqslant 1] = 1 - e^{-\lambda}, \qquad \lambda = 5$$

Calcular:

1. El número medio de nuevas infecciones, sobre la población de susceptibles, cuyo tamaño observado es de 50 individuos, e indicar la distribución aleatoria de la variable que contabiliza las nuevas infecciones en dicha población.

   Sea T la variable aleatoria que representa el tiempo entre dos contagios consecutivos. Entonces, $T \sim \exp(5)$. Sea X la variable aleatoria que representa el número de nuevas infecciones en una población de susceptibles de tamaño 50. La probabilidad de contagio la desconocemos, por lo que sea esta $p \in]0, 1[$. Tenemos que $X \sim B(50, p)$, por lo que:

   $$E[X] = 50 \cdot p$$

   Como podemos ver, este resultado no es concluyente, ya que no conocemos el valor de p, nos faltan datos en el enunciado.

2. Calcular la probabilidad de que se produzcan 10 contagios en un intervalo de tiempo de longitud 10 unidades temporales. Determinar la distribución de probabilidad de dicha variable aleatoria, así como El número medio de contagios en dicho intervalo temporal.

   Sea Y la variable aleatoria que representa el número de contagios en un intervalo de tiempo de longitud 10 unidades temporales. Entonces, $Y \sim \mathcal{P}(10 \cdot 5) = \mathcal{P}(50)$.

   La probabilidad de que se produzcan 10 contagios en un intervalo de tiempo de longitud 10 unidades temporales es:

   $$P[Y = 10] = \frac{e^{-50} \cdot 50^{10}}{10!} \approx 5{,}19 \cdot 10^{-12}$$

   Además, El número medio de contagios en dicho intervalo temporal es:

   $$E[Y] = 50$$

3. Calcular la probabilidad de que no se produzcan contagios en un intervalo de longitud 20 unidades temporales, así como el tiempo medio transcurrido entre contagios.

   Tenemos que la probabilidad pedida es:

   $$P[T \geqslant 20] = 1 - P[T \leqslant 20] = 1 - (1 - e^{-100}) = e^{-100} \approx 3{,}72 \cdot 10^{-44}$$

   Además, el tiempo medio transcurrido entre contagios es:

   $$E[T] = \frac{1}{\lambda} = \frac{1}{5} = 0{,}2$$

Ejercicio 7.1.7. La probabilidad de que un individuo sufra reacción al inyectarle un determinado suero es 0,1. Usando la aproximación normal adecuada, calcular la probabilidad de que al inyectar el suero a una muestra de 400 personas, sufran reacción entre 33 y 50.

Sea X la variable aleatoria que representa el número de personas que sufren reacción al inyectarles el suero. Entonces, $X \sim B(400, 0{,}1)$. Como $n = 400 > 30$ y $p \geqslant 0{,}1$, podemos aproximar la distribución binomial a una normal de media de valor $\mu = np = 400 \cdot 0{,}1 = 40$ y varianza $np(1-p) = 400 \cdot 0{,}1 \cdot 0{,}9 = 36$. Por tanto, $X \sim \mathcal{N}(40, 36)$. Tenemos que:

$$P[33 \leqslant X \leqslant 50] = P\left[X \leqslant 50\right] - P\left[X \leqslant 32\right]$$

Usando la corrección por continuidad de la aproximación normal a la binomial, tenemos que:

$$\begin{aligned} P\left[X \leqslant 50\right] &= P\left[X \leqslant 50{,}5\right] \\ P\left[X \leqslant 32\right] &= P\left[X \leqslant 32{,}5\right] \end{aligned}$$

Sea Z la variable aleatoria X tipificada, es decir, $Z = \dfrac{X-40}{6}$. Entonces, se tiene $Z \sim \mathcal{N}(0,1)$, y tenemos que:

$$\begin{aligned} P\left[X \leqslant 50{,}5\right] &= P\left[Z \leqslant \frac{50{,}5-40}{6}\right] = P\left[Z \leqslant 1{,}75\right] \approx 0{,}95994 \\ P\left[X \leqslant 32{,}5\right] &= P\left[Z \leqslant \frac{32{,}5-40}{6}\right] = P\left[Z \leqslant -1{,}25\right] = 1 - P\left[Z \leqslant 1{,}25\right] \approx \\ &\approx 1 - 0{,}89435 = 0{,}10565 \end{aligned}$$

Por tanto, $P[33 \leqslant X \leqslant 50] \approx 0{,}95994 - 0{,}10565 = 0{,}85429$.

Ejercicio 7.1.8. El tiempo de duración de una pieza de un cierto equipo, medido en horas, se distribuye según una ley exponencial de parámetro 0,2. Si el equipo deja de funcionar cuando fallan 3 piezas, determinar:

1. Probabilidad de que el equipo funcione más de 10 horas.

   Sea X la variable aleatoria que representa el tiempo de duración de una pieza del equipo. Entonces, $X \sim \mathcal{E}(3, 0{,}2)$. Queremos calcular entonces $P[X > 10]$. Tenemos que:

$$\begin{aligned} P[X > 10] = 1 - P[X \leqslant 10] &= 1 - \int_0^{10} \frac{0{,}2^3}{\Gamma(3)} x^{3-1} e^{-0{,}2x}\,dx = \\ &= 1 - \frac{0{,}2^3}{\Gamma(3)} \int_0^{10} x^2 e^{-0{,}2x}\,dx \end{aligned}$$

Calculemos ahora dicha integral:

$$\begin{aligned}
\int_0^{10} x^2 e^{-0,2x}\,dx &= \left[\begin{array}{cc} u(x) = x^2 & u'(x) = 2x \\ v'(x) = e^{-0,2x} & v(x) = -5e^{-0,2x} \end{array}\right] = \\
&= \left[-5x^2e^{-0,2x}\right]_0^{10} + 10\int_0^{10} xe^{-0,2x}\,dx = \\
&= \left[\begin{array}{cc} u(x) = x & u'(x) = 1 \\ v'(x) = e^{-0,2x} & v(x) = -5e^{-0,2x} \end{array}\right] = \\
&= -500e^{-2} + 10\left(\left[-5xe^{-0,2x}\right]_0^{10} + 5\int_0^{10} e^{-0,2x}\,dx\right) = \\
&= -500e^{-2} + 10\left(-50e^{-2} + 5\left[-5e^{-0,2x}\right]_0^{10}\right) = \\
&= -500e^{-2} - 500e^{-2} - 250e^{-2} + 250 = 250 - 1250e^{-2}
\end{aligned}$$

Por tanto, la probabilidad pedida es:

$$\begin{aligned}
P[X > 10] &= 1 - \frac{0,2^3}{\Gamma(3)}\cdot(250 - 1250e^{-2}) = 1 - \frac{0,2^3}{2}\cdot(250 - 1250e^{-2}) = \\
&= 1 - \frac{1}{250}\cdot(250 - 1250e^{-2}) = 1 - (1 - 5e^{-2}) = 5e^{-2} \approx 0,67667
\end{aligned}$$

2. Probabilidad de que el equipo funcione entre 10 y 15 horas.

   Queremos calcular entonces $P[10 \leqslant X \leqslant 15]$. Tenemos que:

$$\begin{aligned}
P[10 \leqslant X \leqslant 15] &= \int_{10}^{15} \frac{0,2^3}{\Gamma(3)} x^{3-1}e^{-0,2x}\,dx = \\
&= \frac{0,2^3}{2}\int_{10}^{15} x^2e^{-0,2x}\,dx
\end{aligned}$$

   Reutilizando el cálculo anterior, tenemos que:

$$\begin{aligned}
\int_{10}^{15} x^2e^{-0,2x}\,dx &= \left[-5x^2e^{-0,2x}\right]_{10}^{15} + 10\left(\left[-5xe^{-0,2x}\right]_{10}^{15} + 5\left[-5e^{-0,2x}\right]_{10}^{15}\right) = \\
&= -5(225e^{-3} - 100e^{-2}) + 10\left[-5(15e^{-3} - 10e^{-2}) - 25(e^{-3} - e^{-2})\right] = \\
&= e^-3(-1125 - 750 - 250) + e^{-2}(500 + 500 + 250) = \\
&= 1250e^{-2} - 2125e^{-3}
\end{aligned}$$

   Por tanto, la probabilidad pedida es:

$$\begin{aligned}
P[10 \leqslant X \leqslant 15] &= \frac{0,2^3}{2}\cdot(1250e^{-2} - 2125e^{-3}) = \frac{1}{250}\cdot(1250e^{-2} - 2125e^{-3}) = \\
&= 5e^{-2} - 8,5e^{-3} \approx 0,2535
\end{aligned}$$

Ejercicio 7.1.9. El número de piezas defectuosas diarias en un proceso de fabricación se distribuye según una Poisson. Sabiendo que El número medio de piezas defectuosas diarias es 25, calcular mediante la aproximación normal:

1. Probabilidad de que El número de defectuosas durante un día oscile entre 24 y 28.

   Sea X la variable aleatoria que representa el número de piezas defectuosas diarias. Entonces, $X \sim \mathcal{P}(25)$. Además, como $n = 25 > 10$, podemos aproximar la distribución de Poisson a una normal de media 25 y varianza 25. Por tanto, $X \sim \mathcal{N}(25, 25)$. Además, como se trata de una aproximación normal, podemos usar la corrección por continuidad. Por tanto, tenemos que:

$$\begin{aligned} P[24 \leqslant X \leqslant 28] &= P[X \leqslant 28] - P[X \leqslant 23] = \\ &= P[X \leqslant 28{,}5] - P[X \leqslant 23{,}5] = \\ &= P\left[Z \leqslant \frac{28{,}5 - 25}{\sqrt{25}}\right] - P\left[Z \leqslant \frac{23{,}5 - 25}{\sqrt{25}}\right] = \\ &= P\left[Z \leqslant 0{,}7\right] - P\left[Z \leqslant -0{,}3\right] = \\ &= P\left[Z \leqslant 0{,}7\right] - 1 + P\left[Z \leqslant 0{,}3\right] = \\ &= 0{,}75804 - 1 + 0{,}61791 = 0{,}37595 \end{aligned}$$

2. Número máximo de defectuosas que con probabilidad 0,97725 se fabrican al día.

   Queremos calcular entonces a tal que $P[X \leqslant a] = 0{,}97725$. Tenemos que:

$$P[X \leqslant a] = P\left[Z \leqslant \frac{a + 0{,}5 - 25}{5}\right] = 0{,}97725$$

   Tenemos por tanto que:

$$\frac{a + 0{,}5 - 25}{5} = 2 \Longrightarrow a = 34{,}5$$

   Por tanto, el número máximo de defectuosas que con probabilidad 0,97725 se fabrican al día es 34, ya que este debe ser un número entero.

3. Número mínimo de defectuosas que con probabilidad 0,15866 se fabrican al día.

   Queremos calcular entonces b tal que $P[X \geqslant b] = 0{,}15866$. Tenemos que:

$$\begin{aligned} P[X \geqslant b] = 1 - P[X \leqslant b - 1] \cong 1 - P\left[Z \leqslant \frac{b - 1 + 0{,}5 - 25}{5}\right] = 0{,}15866 \Longrightarrow \\ \Longrightarrow P\left[Z \leqslant \frac{b - 1 + 0{,}5 - 25}{5}\right] = 0{,}84134 \end{aligned}$$

   Tenemos por tanto que:

$$\frac{b - 1 + 0{,}5 - 25}{5} = 1 \Longrightarrow b = 30{,}5$$

   Por tanto, el número mínimo de defectuosas que con probabilidad 0,15866 se fabrican al día es 31, ya que este debe ser un número entero.

Ejercicio 7.1.10. Un grupo de investigadores ha determinado que el 3 % de los individuos afectados por cierto virus fallece. Determinar:

1. La probabilidad de que en una población de 10000 afectados fallezcan más de 100.

   Sea X la variable aleatoria que representa el número de individuos fallecidos en una población de 10^4 afectados. Entonces, $X \sim B(10^4, 0{,}03)$. Como $n = 10^4 > 30$ y $p = 0{,}03 \geqslant 0{,}01$, podemos aproximar la distribución binomial a una normal de media $\mu = np = 300$ y varianza $\sigma^2 = np(1-p) = 291$. Por tanto, $X \sim \mathcal{N}(300, 291)$. Además, como se trata de una aproximación normal, podemos usar la corrección por continuidad. Por tanto, tenemos que:

$$P[X > 100] = 1 - P[X \leqslant 100] = 1 - P[X \leqslant 100{,}5] = 1 - P\left[Z \leqslant \frac{100{,}5 - 300}{\sqrt{291}}\right] =$$
$$= 1 - P\left[Z \leqslant -11{,}7\right] = P[Z \leqslant 11{,}7] \approx 1$$

2. El número esperado de fallecidos en dicha población. Tenemos que:

$$E[X] = np = 10^4 \cdot 0{,}03 = 300$$

Ejercicio 7.1.11. La experiencia ha demostrado que las calificaciones obtenidas en un test de aptitud por los alumnos de un determinado centro siguen una distribución normal de media 400 y desviación típica 100. Si se realiza el test a un determinado grupo de alumnos, calcular:

1. El porcentaje de alumnos que obtendrán calificaciones comprendidas entre 300 y 500.

   Sea X la variable aleatoria que representa la calificación obtenida por un alumno. Entonces, $X \sim \mathcal{N}(400, 100^2)$. Queremos calcular entonces la probabilidad $P[300 \leqslant X \leqslant 500]$. Tenemos que:

$$\begin{aligned}
P[300 \leqslant X \leqslant 500] &= P\left[X \leqslant 500\right] - P\left[X \leqslant 300\right] = \\
&= P\left[Z \leqslant \frac{500-400}{100}\right] - P\left[Z \leqslant \frac{300-400}{100}\right] = \\
&= P\left[Z \leqslant 1\right] - P\left[Z \leqslant -1\right] = \\
&= P\left[Z \leqslant 1\right] - 1 + P\left[Z \leqslant 1\right] = \\
&= 2 \cdot P\left[Z \leqslant 1\right] - 1 \approx 2 \cdot 0{,}84134 - 1 = 0{,}68268
\end{aligned}$$

   Por tanto, el porcentaje de alumnos que obtendrán calificaciones comprendidas entre 300 y 500 es del 68,268 %.

2. La probabilidad de que, elegido un alumno al azar, su calificación difiera de la media en 150 puntos como máximo.

   Queremos calcular entonces $P[|X - 400| \leqslant 150]$. Veamos qué valores de X cumplen lo pedido.

$$|X - 400| \leqslant 150 \Longrightarrow 250 \leqslant X \leqslant 550$$

Por tanto, la probabilidad pedida es:

$$\begin{aligned} P[|X-400| \leqslant 150] &= P[250 \leqslant X \leqslant 550] = P[X \leqslant 550] - P[X \leqslant 250] = \\ &= P\left[Z \leqslant \frac{550-400}{100}\right] - P\left[Z \leqslant \frac{250-400}{100}\right] = \\ &= P[Z \leqslant 1{,}5] - P[Z \leqslant -1{,}5] = \\ &= P[Z \leqslant 1{,}5] - 1 + P[Z \leqslant 1{,}5] = \\ &= 2 \cdot P[Z \leqslant 1{,}5] - 1 \approx 2 \cdot 0{,}93319 - 1 = 0{,}86638 \end{aligned}$$

Ejercicio 7.1.12. En un parking público se ha observado que los coches llegan, aleatoria e independientemente, a razón de 360 coches por hora.

1. Utilizando la distribución exponencial, encontrar la probabilidad de que una vez que llega un coche, el próximo no llegue antes de medio minuto.

   Sea X la variable aleatoria que representa el tiempo que transcurre entre la llegada de dos coches consecutivos (en minutos). Entonces, como los coches llegan a razón de 360 coches por hora, tenemos que $\lambda = {}^{360}/_{6} = 6$ coches por minuto. Por tanto, $X \sim \exp(6)$. Queremos calcular entonces $P[X > 0{,}5]$. Tenemos que:

   $$P[X > 0{,}5] = 1 - P[X \leqslant 0{,}5] = 1 - (1 - e^{-6\cdot 0{,}5}) = e^{-3} \approx 0{,}04978$$

2. Utilizando la distribución de Poisson, obtener la misma probabilidad anterior.

   Sea Y la variable aleatoria que representa el número de coches que llegan en un intervalo de tiempo de longitud 0,5 minutos. Entonces, $Y \sim \mathcal{P}(\lambda \cdot 0{,}5)$, donde $\lambda = 6$ coches por minuto. Por tanto, $Y \sim \mathcal{P}(3)$. Queremos calcular entonces $P[Y = 0]$. Tenemos que:

   $$P[Y=0] = \frac{e^{-3} \cdot 3^0}{0!} = e^{-3}$$

   Por tanto, la probabilidad de que una vez que llega un coche, el próximo no llegue antes de medio minuto es e^{-3}, que coincide con el resultado obtenido en el apartado anterior.

Ejercicio 7.1.13. Cierta enfermedad puede ser producida por tres tipos de virus: A, B y C. En un laboratorio se tienen tres tubos con el virus A, dos tubos con el virus B y cinco con el virus C. La probabilidad de que el virus A produzca la enfermedad es $P(|X| < 4)$, siendo $X \sim \mathcal{N}(3, 25)$. La probabilidad de que el virus B produzca la enfermedad es $P(Y \geqslant 3)$, siendo $Y \sim B(5, 0{,}7)$. Por último, la probabilidad de que el virus C produzca la enfermedad es $P(Z \leqslant 5)$, siendo $Z \sim \mathcal{P}(4)$. Se elige un tubo al azar y al inocular el virus a un animal, contrae la enfermedad. Hallar la probabilidad de que el virus inoculado sea del tipo C.

Sean los siguientes sucesos:

- A: El virus inoculado es del tipo A.

- B: El virus inoculado es del tipo B.
- C: El virus inoculado es del tipo C.
- E: El animal contrae la enfermedad.

Tenemos que, según el enunciado:

$$P[E \mid A] = P[|X| < 4] \qquad P[E \mid B] = P[Y \geqslant 3] \qquad P[E \mid C] = P[Z \leqslant 5]$$

Calculemos cada una de estas probabilidades:

$$\begin{aligned}
P[E \mid A] = P[|X| < 4] = P[-4 < X < 4] = P\left[\frac{-4-3}{5} < Z < \frac{4-3}{5}\right] = P[-1{,}4 < Z < 0{,}2] = \\
&= P[Z < 0{,}2] - P[Z < -1{,}4] = P[Z < 0{,}2] - 1 + P[Z < 1{,}4] = \\
&= 0{,}57926 - 1 + 0{,}91924 = 0{,}4985
\end{aligned}$$

$$P[E \mid B] = P[Y \geqslant 3] = 1 - P[Y \leqslant 2] = 1 - \sum_{k=0}^{2} \binom{5}{k} 0{,}7^k 0{,}3^{5-k} = 0{,}83692$$

$$P[E \mid C] = P[Z \leqslant 5] = 0{,}7821$$

Calculemos ahora la probabilidad de que el virus inoculado sea de cada tipo mediante la Regla de Laplace. Sabiendo que hay 3 tubos con el virus A, 2 con el virus B y 5 con el virus C, tenemos que:

$$\begin{aligned}
P[A] &= \frac{3}{3+2+5} = \frac{3}{10} = 0{,}3 \\
P[B] &= \frac{2}{3+2+5} = \frac{2}{10} = 0{,}2 \\
P[C] &= \frac{5}{3+2+5} = \frac{5}{10} = 0{,}5
\end{aligned}$$

Nos piden por tanto calcular $P[C \mid E]$. Por la fórmula de Bayes, tenemos que:

$$\begin{aligned}
P[C \mid E] &= \frac{P[E \mid C] \cdot P[C]}{P[E]} = \frac{P[E \mid C] \cdot P[C]}{\sum\limits_{i=A,B,C} P[E \mid i] \cdot P[i]} = \\
&= \frac{0{,}7821 \cdot 0{,}5}{0{,}4985 \cdot 0{,}3 + 0{,}83692 \cdot 0{,}2 + 0{,}7821 \cdot 0{,}5} = \frac{0{,}39105}{0{,}14955 + 0{,}167384 + 0{,}39105} \approx \\
&\approx 0{,}5523
\end{aligned}$$

Ejercicio 7.1.14. Una máquina fabrica tornillos cuyas longitudes se distribuyen según una ley normal con media 20 mm y desviación típica 0,25 mm. Un tornillo se considera defectuoso si su longitud no está comprendida entre 19,5 y 20,5 mm. Los tornillos se fabrican de forma independiente.

1. ¿Cuál es la probabilidad de fabricar un tornillo defectuoso?

   Sea X la variable aleatoria que representa la longitud de un tornillo. Entonces, $X \sim \mathcal{N}(20, 0{,}25^2)$. Queremos calcular entonces $1 - P[19{,}5 \leqslant X \leqslant 20{,}5]$.

Tenemos que:

$$\begin{aligned}
P[19{,}5 \leqslant X \leqslant 20{,}5] &= P\left[X \leqslant 20{,}5\right] - P\left[X \leqslant 19{,}5\right] = \\
&= P\left[Z \leqslant \frac{20{,}5-20}{0{,}25}\right] - P\left[Z \leqslant \frac{19{,}5-20}{0{,}25}\right] = \\
&= P\left[Z \leqslant 2\right] - P\left[Z \leqslant -2\right] = \\
&= P\left[Z \leqslant 2\right] - 1 + P\left[Z \leqslant 2\right] = \\
&= 2 \cdot P\left[Z \leqslant 2\right] - 1
\end{aligned}$$

Por tanto, si A es el suceso de que un tornillo sea defectuoso, tenemos que:

$$P[A] = 1 - P[19{,}5 \leqslant X \leqslant 20{,}5] = 1 - 2 \cdot P\left[Z \leqslant 2\right] + 1 = 2 - 2 \cdot 0{,}97725 = 0{,}0455$$

2. Calcular la probabilidad de que en 10 tornillos fabricados no haya más de dos defectuosos.

   Sea Y la variable aleatoria que representa el número de tornillos defectuosos en 10 tornillos fabricados. Entonces, $Y \sim \mathcal{B}(10, 0{,}0455)$. Queremos calcular entonces $P[Y \leqslant 2]$. Tenemos que:

$$P[Y \leqslant 2] = \sum_{k=0}^{2} \binom{10}{k} 0{,}0455^k (1 - 0{,}0455)^{10-k} \approx 0{,}99112$$

3. Cuántos tornillos se fabricarán por término medio hasta obtener el primero defectuoso?

   Sea Z la variable aleatoria que representa el número de tornillos fabricados hasta obtener el primero defectuoso. Entonces, $Z \sim \mathcal{G}(0{,}0455)$. Queremos calcular entonces $E[Z]$. Tenemos que:

$$E[Z] = \frac{1-p}{p} = \frac{1-0{,}0455}{0{,}0455} \approx 20{,}98$$

Ejercicio 7.1.15. Si la proporción de personas que consumen una determinada marca de aceite de oliva sigue una distribución beta de parámetros 2 y 3, determinar la probabilidad de que dicha proporción esté comprendida entre el 0,1 y 0,5.

Sea X la variable aleatoria que representa la proporción de personas que consumen una determinada marca de aceite de oliva. Entonces, $X \sim \beta(2, 3)$.

Queremos calcular entonces $P[0{,}1 \leqslant X \leqslant 0{,}5]$. Tenemos que:

$$\begin{aligned}
P[0{,}1 \leqslant X \leqslant 0{,}5] &= \int_{0{,}1}^{0{,}5} \frac{\Gamma(2+3)}{\Gamma(2)\Gamma(3)} x^{2-1}(1-x)^{3-1}\,dx = \frac{4 \cdot 3 \cdot \Gamma(3)}{1 \cdot \Gamma(3)} \int_{0{,}1}^{0{,}5} x(1-x)^2\,dx = \\
&= 12 \int_{0{,}1}^{0{,}5} x(1-2x+x^2)\,dx = 12 \int_{0{,}1}^{0{,}5} x - 2x^2 + x^3\,dx = \\
&= 12 \left[\frac{x^2}{2} - \frac{2x^3}{3} + \frac{x^4}{4}\right]_{0{,}1}^{0{,}5} \approx 0{,}6352
\end{aligned}$$

Ejercicio 7.1.16. La proporción diaria de piezas defectuosas en determinada fábrica tiene distribución beta, y el segundo parámetro es 4. Sabiendo que la proporción media diaria es 0,2, calcular la probabilidad de que un día resulte una proporción de defectuosas superior a la media.

Sea X la variable aleatoria que representa la proporción diaria de piezas defectuosas. Entonces, $X \sim \beta(p, 4)$.

Para calcular el valor de p, usamos que $E[X] = 0{,}2$.

$$E[X] = \frac{p}{p+4} = 0{,}2 \Longrightarrow p = 0{,}2p + 0{,}8 \Longrightarrow 0{,}8p = 0{,}8 \Longrightarrow p = 1$$

Por tanto, $X \sim \beta(1, 4)$. Queremos calcular entonces $P[X > 0{,}2]$. Tenemos que:

$$\begin{aligned}
P[X > 0{,}2] &= 1 - P[X \leqslant 0{,}2] = 1 - \int_0^{0{,}2} \frac{\Gamma(1+4)}{\Gamma(1)\Gamma(4)} x^{1-1}(1-x)^{4-1}\,dx = \\
&= 1 - \frac{4 \cdot \Gamma(4)}{\Gamma(1)\Gamma(4)} \int_0^{0{,}2} (1-x)^3\,dx = 1 - 4\left[-\frac{(1-x)^4}{4}\right]_0^{0{,}2} = \\
&= 1 + \left[(1-x)^4\right]_0^{0{,}2} = (0{,}8)^4 \approx 0{,}4096
\end{aligned}$$

Ejercicio 7.1.17. El Instituto de Estadística de una determinada comunidad autónoma convoca unas pruebas selectivas para cubrir vacantes. La puntuación obtenida por cada candidato se calcula mediante el promedio de las calificaciones obtenidas en las pruebas realizadas, y se sabe, de experiencias previas, que dichas puntuaciones tienen media 100, se distribuyen de forma normal y que el 44,04 % de los aspirantes que realizan la prueba supera la puntuación 100,6.

1. La convocatoria de las pruebas establece una nota mínima de 105 puntos para superar la oposición. ¿Qué porcentaje de opositores consiguen una plaza?

   Sea X la variable aleatoria que representa la puntuación obtenida por un candidato. Entonces, $X \sim \mathcal{N}(100, \sigma^2)$. Sabemos que $P[X > 100{,}6] = 0{,}4404$. Por tanto, $P[X \leqslant 100{,}6] = 1 - 0{,}4404 = 0{,}5596$. Sea Z la variable aleatoria X tipificada, es decir, $Z = \dfrac{X - 100}{\sigma}$. Entonces, $Z \sim \mathcal{N}(0, 1)$, y tenemos que:

   $$P\left[Z \leqslant \frac{100{,}6 - 100}{\sigma}\right] = 0{,}5596 = P\left[Z \leqslant \frac{0{,}6}{\sigma}\right]$$

   Usando la tabla de la distribución normal estándar, buscamos el valor de z tal que $P[Z \leqslant z] = 0{,}5596$. Dicho valor, tras consultar la tabla, es $z = 0{,}15$. Por tanto,

   $$\frac{0{,}6}{\sigma} = z = 0{,}15 \Longrightarrow \sigma = 4 \Longrightarrow \sigma^2 = 16$$

   Tenemos por tanto que $X \sim \mathcal{N}(100, 16)$. Ahora, queremos calcular el valor de $P[X > 105]$. Tenemos que:

   $$\begin{aligned}
   P[X > 105] &= 1 - P[X \leqslant 105] = 1 - P\left[Z \leqslant \frac{105 - 100}{4}\right] = 1 - P\left[Z \leqslant 1{,}25\right] \approx \\
   &\approx 1 - 0{,}89435 = 0{,}10565
   \end{aligned}$$

2. No obstante, se sabe que en ocasiones el tribunal decide, dependiendo de las necesidades de personal, rebajar las condiciones para que un candidato sea admitido. ¿Cuál sería la nota mínima necesaria para que la probabilidad de superar la prueba de selección sea 0,33?

   Sea a la nota mínima necesaria para que la probabilidad de superar la prueba de selección sea 0,33; es decir, el valor buscado. Buscamos a tal que $P[X > a] = 0{,}33$. Tenemos que:

$$P[X > a] = 1 - P[X \leqslant a] = 1 - P\left[Z \leqslant \frac{a-100}{4}\right] = 0{,}33 \Longrightarrow$$
$$\Longrightarrow P\left[Z \leqslant \frac{a-100}{4}\right] = 0{,}67$$

   Usando la tabla de la distribución normal estándar, buscamos el valor de z tal que $P[Z \leqslant z] = 0{,}67$. Dicho valor, tras consultar la tabla, es $z = 0{,}44$. Por tanto,

$$\frac{a-100}{4} = z = 0{,}44 \Longrightarrow a = 100 + 4 \cdot 0{,}44 = 101{,}76$$

   Por tanto, la nota mínima necesaria para que la probabilidad de superar la prueba de selección sea 0,33 es 101,76.

3. El instituto decide crear una bolsa de interinos para cubrir temporalmente posibles eventualidades. A esa bolsa pertenecerán todos los candidatos cuyas puntuaciones estén entre la media de las puntuaciones y la nota establecida en el apartado anterior. ¿Qué porcentaje de candidatos estarán en dicha situación?

   En este caso, nos piden que calculemos $P[100 \leqslant X \leqslant 101{,}76]$. Tenemos que:

$$P[100 \leqslant X \leqslant 101{,}76] = P\left[0 \leqslant Z \leqslant \frac{101{,}76-100}{4}\right] = P\left[0 \leqslant Z \leqslant 0{,}44\right] =$$
$$= P\left[Z \leqslant 0{,}44\right] - P\left[Z \leqslant 0\right] \overset{(*)}{=} 0{,}67 - 0{,}5 = 0{,}17$$

   donde en $(*)$ hemos usado que $P[Z \leqslant 0] = 0{,}5$ por ser esta la distribución normal estándar y $P[Z \leqslant 0{,}44] = 0{,}67$ tras consultar la tabla de la distribución normal estándar.

7.2. Vectores Aleatorios. Parte 1

Ejercicio 7.2.1. Asociadas al experimento aleatorio de lanzar un dado y una moneda no cargados, se define la variable X como el valor del dado y la variable Y, que toma el valor 0 si sale cara en la moneda, y 1 si sale cruz. Calcular la función masa de probabilidad y la función de distribución del vector aleatorio (X, Y).

Calculemos los recorreidos de X e Y:

$$\begin{aligned} E_X &= \{1, 2, 3, 4, 5, 6\}, \\ E_Y &= \{0, 1\}. \end{aligned}$$

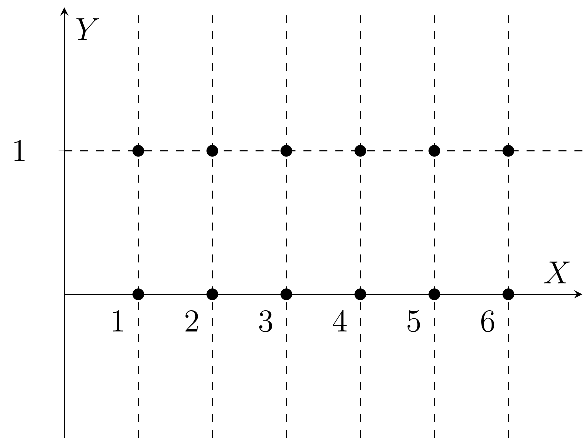
Por tanto, tenemos que:

$$E_{(X,Y)} = \{(1,0), (1,1), (2,0), (2,1), (3,0), (3,1), (4,0), (4,1), (5,0), (5,1), (6,0), (6,1)\}.$$

La función masa de probabilidad es:

$$\begin{array}{rccc} P_{(X,Y)}: & E_{(X,Y)} & \longrightarrow & [0,1] \\ & (x,y) & \longmapsto & 1/12 \end{array}$$

Para poder calcular la función de distribución, primero representamos los puntos del espacio muestral en el plano cartesiano:



La función de distribución es:

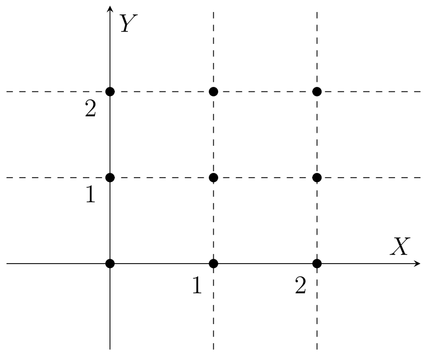
$$F_{(X,Y)}(x,y) = \begin{cases} 0, & x < 1 \text{ o } y < 0, \\ 1/12, & x \in [1,2[\text{ y } y \in [0,1[\,, \\ 2/12, & x \in [2,3[\text{ y } y \in [0,1[\quad \text{o} \quad x \in [1,2[\text{ y } y \geqslant 1, \\ 3/12, & x \in [3,4[\text{ y } y \in [0,1[\,, \\ 4/12, & x \in [4,5[\text{ y } y \in [0,1[\quad \text{o} \quad x \in [2,3[\text{ y } y \geqslant 1, \\ 5/12, & x \in [5,6[\text{ y } y \in [0,1[\quad \text{o} \quad x \in [3,4[\text{ y } y \geqslant 1, \\ 6/12, & x \geqslant 6 \text{ y } y \in [0,1[\quad \text{o} \quad x \in [3,4[\text{ y } y \geqslant 1, \\ 8/12, & x \in [4,5[\text{ y } y \geqslant 1, \\ 10/12, & x \in [5,6[\text{ y } y \geqslant 1, \\ 1, & x \geqslant 6 \text{ y } y \geqslant 1. \end{cases}$$

Ejercicio 7.2.2. El número de automóviles utilitarios, X, y el de automóviles de lujo, Y, que poseen las familias de una población se distribuye de acuerdo a las siguientes probabilidades:

$X\backslash Y$	0	1	2
0	$1/3$	$1/12$	$1/24$
1	$1/6$	$1/24$	$1/48$
2	$5/22$	$5/88$	$5/176$

Calcular la función de distribución del vector (X, Y) en los puntos $(0, 0)$; $(0, 2)$; $(1, 1)$ y $(2, 1)$, y la probabilidad de que una familia tenga tres o más automóviles.

Para calcular la función de distribución en los puntos $(0, 0)$; $(0, 2)$; $(1, 1)$ y $(2, 1)$, representamos antes los elementos de $E_{(X,Y)}$ en el plano cartesiano:



La función de distribución en los puntos $(0, 0)$; $(0, 2)$; $(1, 1)$ y $(2, 1)$ es:

$F_{(X,Y)}(0, 0) = P_{(X,Y)}(0, 0) = 1/3$,
$F_{(X,Y)}(0, 2) = P_{(X,Y)}(0, 0) + P_{(X,Y)}(0, 1) + P_{(X,Y)}(0, 2) = 1/3 + 1/12 + 1/24 = 11/24$,
$F_{(X,Y)}(1, 1) = P_{(X,Y)}(0, 0) + P_{(X,Y)}(0, 1) + P_{(X,Y)}(1, 0) + P_{(X,Y)}(1, 1) = 1/3 + 1/12 + 1/6 + 1/24 = 5/8$,
$F_{(X,Y)}(2, 1) = F_{(X,Y)}(1, 1) + P_{(X,Y)}(2, 0) + P_{(X,Y)}(2, 1) = 5/8 + 5/22 + 5/88 = 10/11$.

La probabilidad de que una familia tenga tres o más automóviles es:

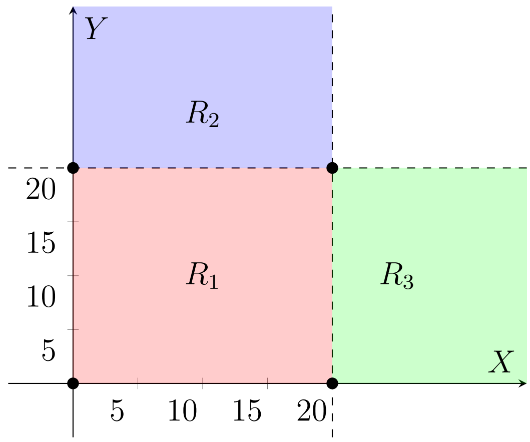
$$P[X + Y \geqslant 3] = P_{(X,Y)}(1, 2) + P_{(X,Y)}(2, 1) + P_{(X,Y)}(2, 2) = 1/48 + 5/88 + 5/176 = 7/66.$$

Ejercicio 7.2.3. La función de densidad del vector aleatorio (X, Y), donde X denota los Kg. de naranjas, e Y los Kg. de manzanas vendidos al día en una frutería está dada por:

$$f(x, y) = \frac{1}{400}; \quad 0 < x < 20; \quad 0 < y < 20.$$

siendo esta nula en caso contrario. Determinar la función de distribución de (X, Y) y la probabilidad de que en un día se vendan entre naranjas y manzanas, menos de 20 kilogramos.

Representamos los puntos de discontinuidad de la función de densidad en el plano cartesiano:



La función de distribución es:

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv$$

Estudiemos cada región por separado:

- Para $x \leqslant 0$ o $y \leqslant 0$:

  Tenemos que $f(u,v) = 0$ para $u \leqslant x$ o $v \leqslant y$, por lo que:

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{-\infty}^{x}\int_{-\infty}^{y} 0\,du\,dv = 0.$$

- Para $0 < x < 20$ y $0 < y < 20$ (región R_1):

  Tenemos que:

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{0}^{x}\int_{0}^{y} \frac{1}{400}\,du\,dv = \frac{xy}{400}.$$

- Para $0 < x < 20$ y $y \geqslant 20$ (región R_2):

  Tenemos que:

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{0}^{x}\int_{0}^{20} \frac{1}{400}\,du\,dv = \frac{20x}{400} = \frac{x}{20}.$$

- Para $x \geqslant 20$ y $0 < y < 20$ (región R_3):

  Tenemos que:

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{0}^{20}\int_{0}^{y} \frac{1}{400}\,du\,dv = \frac{20y}{400} = \frac{y}{20}.$$

- Para $x \geqslant 20$ y $y \geqslant 20$:

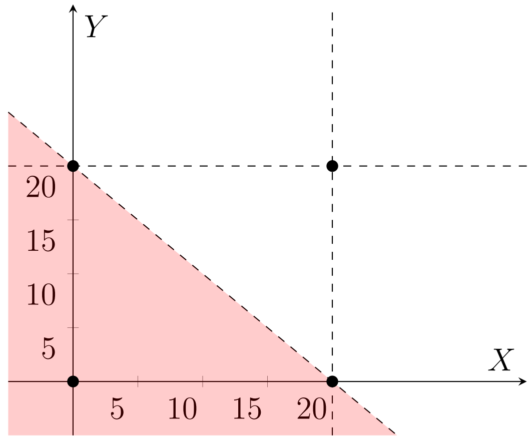
  Tenemos que:

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{0}^{20}\int_{0}^{20} \frac{1}{400}\,du\,dv = \frac{400}{400} = 1.$$

Por tanto, la función de distribución es:

$$F_{(X,Y)}(x,y) = \begin{cases} 0, & x \leqslant 0 \text{ o } y \leqslant 0, \\ \frac{xy}{400}, & 0 < x < 20 \text{ y } 0 < y < 20, \\ \frac{x}{20}, & 0 < x < 20 \text{ y } y \geqslant 20, \\ \frac{y}{20}, & x \geqslant 20 \text{ y } 0 < y < 20, \\ 1, & x \geqslant 20 \text{ y } y \geqslant 20. \end{cases}$$

Buscamos ahora la probabilidad de que en un día se vendan entre naranjas y manzanas, menos de 20 kilogramos. Para ello, buscamos la región del plano que cumple con la condición $X + Y < 20$. Tras representar la recta $y = 20 - x$, nos quedaremos con la región que queda por debajo de esta recta:



Por tanto, tenemos que $P[X + Y < 20]$ es la integral de la función de densidad en la región coloreada, $R = \{(x,y) \in \mathbb{R}^2; x + y < 20\}$:

$$P[X + Y < 20] = \int_{-\infty}^{+\infty} \int_{-\infty}^{20-x} f(x,y)\,dy\,dx = \int_0^{20} \int_0^{20-x} \frac{1}{400}\,dy\,dx =$$
$$= \int_0^{20} \frac{20-x}{400}\,dx = \frac{1}{400}\left[20x - \frac{x^2}{2}\right]_0^{20} = \frac{1}{400}[400 - 200] = \frac{1}{2}.$$

Ejercicio 7.2.4. La renta, X, y el consumo, Y, de los habitantes de una población, tienen por funciones de densidad

$$f_X(x) = 2 - 2x; \quad 0 < x < 1; \quad f_{Y|X}(y \mid x) = \frac{1}{x}; \quad 0 < y < x.$$

Determinar la función de densidad conjunta del vector aleatorio (X, Y) y la probabilidad de que el consumo sea inferior a la mitad de la renta.

Tenemos que, para $R = \{(x,y) \in \mathbb{R}^2; 0 < x < 1, 0 < y < x\}$, la función de densidad conjunta es:

$$f_{Y|X}(y \mid x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)} \Longrightarrow f_{(X,Y)}(x,y) = f_X(x) \cdot f_{Y|X}(y \mid x) = (2-2x) \cdot \frac{1}{x} = \frac{2-2x}{x}.$$

Tenemos ahora que:

$$P[Y < X/2] = \int_{-\infty}^{+\infty} \int_{-\infty}^{x/2} f_{(X,Y)}(x,y)\,dy\,dx = \int_0^1 \int_0^{x/2} \frac{2-2x}{x}\,dy\,dx =$$
$$= \int_0^1 \frac{2-2x}{x} \cdot \frac{x}{2}\,dx = \int_0^1 1 - x\,dx = \left[x - \frac{x^2}{2}\right]_0^1 = 1 - \frac{1}{2} = \frac{1}{2}.$$

Ejercicio 7.2.5. Una gasolinera tiene Y miles de litros en su depósito de gasóleo al comienzo de cada semana. A lo largo de una semana se venden X miles de litros del citado combustible, siendo la función de densidad conjunta de (X,Y) :

$$f(x,y) = \frac{1}{8}; \quad 0 < x < y < 4.$$

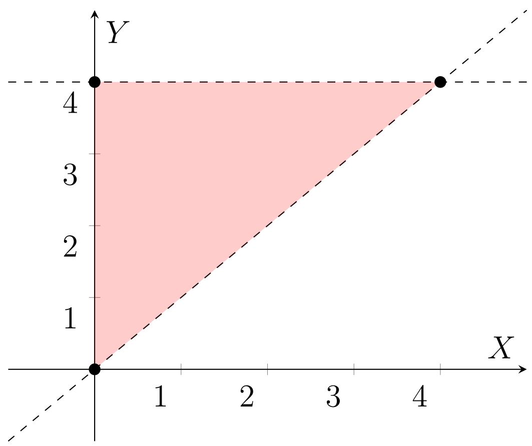
Se pide:

1. Probar que $f(x,y)$ es función de densidad y obtener la función de distribución.

   En primer lugar, vemos que es no negativa. Veamos si es integrable:

   $$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = \int_0^4 \int_x^4 \frac{1}{8}\,dy\,dx = \frac{1}{8}\int_0^4 4 - x\,dx =$$
   $$= \frac{1}{8}\left[4x - \frac{x^2}{2}\right]_0^4 = \frac{1}{8}[16-8] = 1.$$

   Tenemos por tanto que sí se trata de una función de densidad. Para obtener la función de distribución, representemos el conjunto en el que la función de densidad es no nula:

   Para obtener la función de distribución, distinguimos casos:

   - Para $x < 0$ o $y < 0$:

     $$F_{(X,Y)}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\,du\,dv = 0.$$

- Para $0 < x < y < 4$:

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dv\,du = \int_{0}^{x}\int_{u}^{y} \frac{1}{8}\,dv\,du = \frac{1}{8}\int_{0}^{x} y-u\,du =$$
$$= \frac{1}{8}\left[yu - \frac{u^2}{2}\right]_0^x = \frac{1}{8}\left[xy - \frac{x^2}{2}\right] = \frac{2xy-x^2}{16}.$$

- Para $0 < x < 4$ y $y \geqslant 4$:

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dv\,du = \int_{0}^{x}\int_{u}^{4} \frac{1}{8}\,dv\,du = \frac{1}{8}\int_{0}^{x} 4-u\,du =$$
$$= \frac{1}{8}\left[4u - \frac{u^2}{2}\right]_0^x = \frac{1}{8}\left[4x - \frac{x^2}{2}\right] = \frac{8x-x^2}{16}.$$

- Para $x \geqslant y$ y $0 \leqslant y < 4$:

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dv\,du = \int_{0}^{y}\int_{0}^{v} \frac{1}{8}\,du\,dv = \frac{1}{8}\int_{0}^{y} v\,dv =$$
$$= \frac{1}{8}\left[\frac{v^2}{2}\right]_0^y = \frac{1}{8}\cdot\frac{y^2}{2} = \frac{y^2}{16}.$$

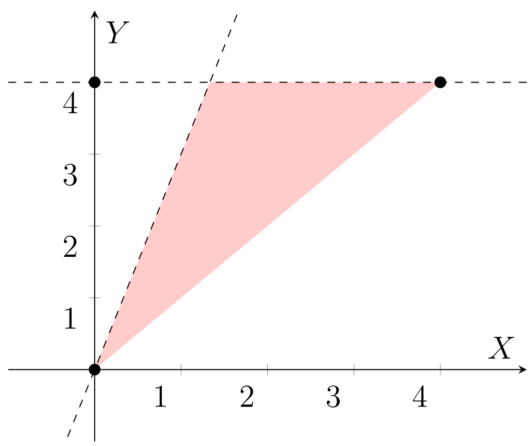
- Para $x \geqslant 4$ y $y \geqslant 4$:
Hemos visto anteriormente que $F_{(X,Y)}(x,y) = 1$.

Por tanto,

$$F_{(X,Y)}(x,y) = \begin{cases} 0, & x < 0 \text{ o } y < 0, \\ \dfrac{2xy-x^2}{16}, & 0 < x < y < 4, \\ \dfrac{8x-x^2}{16}, & 0 < x < 4 \text{ y } y \geqslant 4, \\ \dfrac{y^2}{16}, & x \geqslant y \text{ y } 0 \leqslant y < 4, \\ 1, & x \geqslant 4 \text{ y } y \geqslant 4. \end{cases}$$

2. Probabilidad de que en una semana se venda más de la tercera parte de los litros de que se dispone al comienzo de la misma.

En este caso, nos piden calcular $P[X > {}^{Y}\!/_{3}]$. Para ello, representamos la recta $y = 3x$ y nos quedamos con la región que queda por encima de esta recta:



Tenemos que integrar $f(x,y)$ en la región coloreada:

$$
\begin{aligned}
P[X > {}^{Y}\!/_{3}] &= \int_0^{4/3}\int_x^{3x}\frac{1}{8}\,dy\,dx + \int_{4/3}^{4}\int_x^{4}\frac{1}{8}\,dy\,dx = \\
&= \int_0^{4/3}\frac{3x-x}{8}\,dx + \int_{4/3}^{4}\frac{4-x}{8}\,dx = \frac{1}{8}\int_0^{4/3}2x\,dx + \frac{1}{8}\int_{4/3}^{4}4-x\,dx = \\
&= \frac{1}{8}\left[x^2\right]_0^{4/3} + \frac{1}{8}\left[4x-\frac{x^2}{2}\right]_{4/3}^{4} = \frac{1}{8}\left[\frac{16}{9}\right] + \frac{1}{8}\left[16-\frac{16}{2}-\frac{16}{3}+\frac{16}{18}\right] = \\
&= \frac{2}{9}+\frac{4}{9} = \frac{6}{9} = \frac{2}{3}.
\end{aligned}
$$

3. Si en una semana se han vendido 3,000 litros de gasóleo, ¿cuál es la probabilidad de que al comienzo de la semana hubiese entre 3,500 y 3,750 litros de combustible?

En este caso, nos piden:

$$P[3{,}5 < Y < 3{,}75 \mid X = 3] = \int_{3,5}^{3,75} f_{Y|X=3}(y)\,dy.$$

Veamos el valor de $f_{Y|X=3}(y)$:

$$f_{Y|X=3}(y) = \frac{f_{(X,Y)}(3,y)}{f_X(3)}$$

Calculemos por tanto la marginal de X:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy = \int_x^4 \frac{1}{8}\,dy = \frac{1}{8}\left[y\right]_x^4 = \frac{4-x}{8}$$

Por tanto, tenemos que:

$$f_{Y|X=3}(y) = \frac{f_{(X,Y)}(3,y)}{f_X(3)} = \frac{1/8}{1/8} = 1.$$

Por tanto, la probabilidad pedida es:

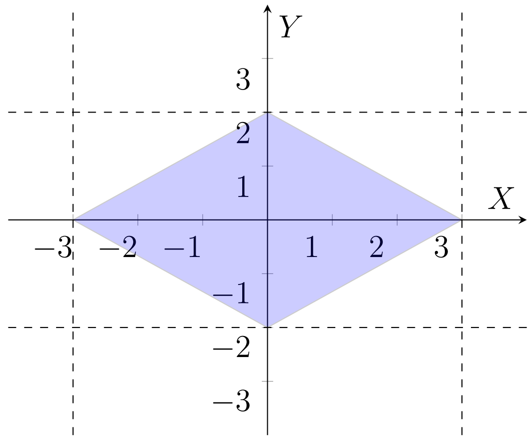
$$P[3{,}5 < Y < 3{,}75 \mid X = 3] = \int_{3,5}^{3,75} 1\,dy = \left[y\right]_{3,5}^{3,75} = 3{,}75 - 3{,}5 = 0{,}25.$$

Ejercicio 7.2.6. Sea (X,Y) un vector aleatorio continuo con función de densidad

$$f(x,y) = k, \quad (x,y) \in R,$$

siendo R el rombo de vértices $(3,0)$; $(0,2)$; $(-3,0)$; $(0,-2)$. Calcular k para que f sea una función de densidad. Hallar las distribuciones marginales y condicionadas.

Representamos en el plano cartesiano el rombo R:



Para que f sea una función de densidad, tenemos que:

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

Calculamos solo la integral en el primer cuadrante, ya que la función es simétrica. Para ello, calculamos la integral en el triángulo de vértices $(0,0)$; $(3,0)$; $(0,2)$. La recta que une los puntos $(3,0)$ y $(0,2)$ es $y = {}^{-2}/_{3}x + 2$. Por tanto, tenemos:

$$1 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = \int_R f(x,y)\,dx\,dy = 4\cdot\int_0^3\int_0^{-2/3\cdot x+2} k\,dy\,dx =$$
$$= 4k\cdot\int_0^3\left[-\frac{2}{3}\cdot x+2\right]dx = 4k\cdot\left[-\frac{1}{3}\cdot x^2+2x\right]_0^3 = 4k\cdot\left[-\frac{1}{3}\cdot 9+6\right] = 4k\cdot[-3+6] = 12k.$$

Por tanto, tenemos que $k = {}^{1}/_{12}$. En este caso, vemos que además $f_{(X,Y)}$ es no negativa e integrable.

Calculemos ahora la distribución marginal de X. Distinguimos:

- Si $x < -3$ o $x \geqslant 3$:
$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy = 0.$$
- Si $-3 \leqslant x < 0$:

  En este caso, tenemos que:
$$-\frac{2}{3}\cdot x - 2 \leqslant y \leqslant \frac{2}{3}\cdot x + 2.$$

  Por tanto, tenemos:
$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy = \int_{-2/3\cdot x-2}^{2/3\cdot x+2}\frac{1}{12}\,dy = \frac{1}{12}\cdot\left[{}^{2}/_{3}\cdot x+2-({}^{-2}/_{3}\cdot x-2)\right] =$$
$$= \frac{1}{12}\cdot\left[\frac{4}{3}\cdot x+4\right] = \frac{x}{9}+\frac{1}{3}.$$

- <u>Si $0 \leqslant x < 3$</u>:

  En este caso, tenemos que:

$$\frac{2}{3} \cdot x - 2 \leqslant y \leqslant -\frac{2}{3} \cdot x + 2.$$

  Por tanto, tenemos:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x,y)\,dy = \int_{2/3\cdot x-2}^{-2/3\cdot x+2} \frac{1}{12}\,dy = \frac{1}{12} \cdot [-2/3 \cdot x + 2 - (2/3 \cdot x - 2)] = \\ &= \frac{1}{12} \cdot \left[-\frac{4}{3} \cdot x + 4\right] = -\frac{x}{9} + \frac{1}{3}. \end{aligned}$$

Por tanto, tenemos que:

$$f_X(x) = \begin{cases} 0, & x < -3 \text{ o } x \geqslant 3, \\ \dfrac{x}{9} + \dfrac{1}{3}, & -3 \leqslant x < 0, \\ -\dfrac{x}{9} + \dfrac{1}{3}, & 0 \leqslant x < 3. \end{cases}$$

Calculemos ahora la distribución marginal de Y. Distinguimos:

- <u>Si $y < -2$ o $y \geqslant 2$</u>:

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx = 0.$$

- <u>Si $-2 \leqslant y < 0$</u>:

  En este caso, tenemos que:

$$-\frac{3}{2} \cdot y - 3 \leqslant x \leqslant \frac{3}{2} \cdot y + 3$$

  Por tanto, tenemos:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f(x,y)\,dx = \int_{-3/2\cdot y-3}^{3/2\cdot y+3} \frac{1}{12}\,dx = \frac{1}{12} \cdot [3/2 \cdot y + 3 - (-3/2 \cdot y - 3)] = \\ &= \frac{y/2 + 1}{2} = \frac{y}{4} + \frac{1}{2}. \end{aligned}$$

- <u>Si $0 \leqslant y < 2$</u>:

  En este caso, tenemos que:

$$\frac{3}{2} \cdot y - 3 \leqslant x \leqslant -\frac{3}{2} \cdot y + 3$$

  Por tanto, tenemos:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f(x,y)\,dx = \int_{3/2\cdot y-3}^{-3/2\cdot y+3} \frac{1}{12}\,dx = \frac{1}{12} \cdot [-3/2 \cdot y + 3 - (3/2 \cdot y - 3)] = \\ &= \frac{-y/2 + 1}{2} = -\frac{y}{4} + \frac{1}{2}. \end{aligned}$$

Por tanto, tenemos que:

$$f_Y(y) = \begin{cases} 0, & y < -2 \text{ o } y \geqslant 2, \\ \dfrac{y}{4} + \dfrac{1}{2}, & -2 \leqslant y < 0, \\ -\dfrac{y}{4} + \dfrac{1}{2}, & 0 \leqslant y < 2. \end{cases}$$

Calculemos ahora las distribuciones condicionadas. Dado $x^* \in]-3, 3[$, tenemos:

$$f_{Y|X=x^*}(y) = \frac{f_{(X,Y)}(x^*, y)}{f_X(x^*)} = \begin{cases} \dfrac{1}{12} \cdot \dfrac{1}{x/9 + 1/3}, & -3 < x^* < 0, \\ \dfrac{1}{12} \cdot \dfrac{1}{-x/9 + 1/3}, & 0 \leqslant x^* < 3. \end{cases}$$

Por otro lado, dado $y^* \in]-2, 2[$, tenemos que:

$$f_{X|Y=y^*}(x) = \frac{f_{(X,Y)}(x, y^*)}{f_Y(y^*)} = \begin{cases} \dfrac{1}{12} \cdot \dfrac{1}{y/4 + 1/2}, & -2 < y^* < 0, \\ \dfrac{1}{12} \cdot \dfrac{1}{-y/4 + 1/2}, & 0 \leqslant y^* < 2. \end{cases}$$

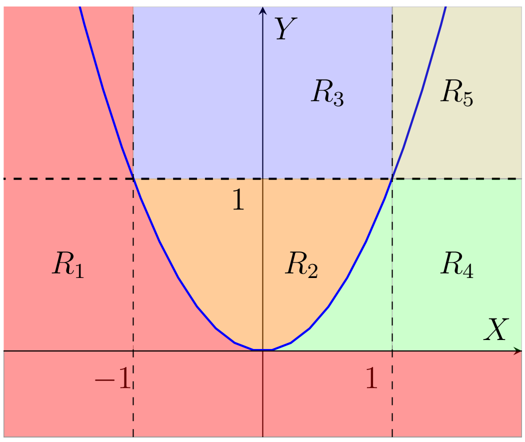
Ejercicio 7.2.7. Sea (X, Y) un vector aleatorio continuo con función de densidad

$$f(x, y) = k, \quad x^2 \leqslant y \leqslant 1,$$

anulándose fuera del recinto indicado.

1. Hallar la constante k para que f sea una función de densidad.

   Representamos en el plano cartesiano el recinto en el que la función de densidad es no nula:

   Para que f sea una función de densidad, tenemos que:

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

Tenemos por tanto que:

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = \int_{-1}^{1} \int_{x^2}^{1} k\,dy\,dx = k \cdot \int_{-1}^{1} [y]_{x^2}^{1}\,dx = k \cdot \int_{-1}^{1} 1 - x^2\,dx =$$
$$= k \cdot \left[x - \frac{x^3}{3}\right]_{-1}^{1} = k \cdot \left[1 - \frac{1}{3} - \left(-1 + \frac{1}{3}\right)\right] = k \cdot \left[2 - \frac{2}{3}\right] = \frac{4}{3}k$$

Por tanto, tenemos que $k = 3/4$. En este caso, vemos que además $f_{(X,Y)}$ es no negativa e integrable.

2. Calcular la función de distribución de probabilidad.

   Distinguimos casos:

   - Si $x \leqslant -1$ o $y \leqslant 0$ o $x \in \left]-1,0\right[$ y $y \leqslant x^2$ (zona R_1):

   $$F_{(X,Y)}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\,du\,dv = 0.$$

   - Si $x^2 \leqslant y \leqslant 1$ (zona R_2):

   $$F_{(X,Y)}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\,dv\,du = \int_{-\sqrt{y}}^{x} \int_{u^2}^{y} \frac{3}{4}\,dv\,du = \frac{3}{4} \int_{-\sqrt{y}}^{x} \left[y - u^2\right]\,du =$$
   $$= \frac{3}{4}\left[yu - \frac{u^3}{3}\right]_{-\sqrt{y}}^{x} = \frac{3}{4}\left[xy - \frac{x^3}{3} + y\sqrt{y} - \frac{y^{3/2}}{3}\right] =$$
   $$= \frac{3}{4}\left[xy - \frac{x^3}{3} + \frac{2}{3} \cdot y\sqrt{y}\right].$$

   - Si $y \geqslant 1$ y $x \in \left]-1,1\right[$ (zona R_3):

   $$F_{(X,Y)}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\,dv\,du = \int_{-1}^{x} \int_{u^2}^{1} \frac{3}{4}\,dv\,du = \frac{3}{4} \int_{-1}^{x} \left[1 - u^2\right]\,du =$$
   $$= \frac{3}{4}\left[u - \frac{u^3}{3}\right]_{-1}^{x} = \frac{3}{4}\left[x - \frac{x^3}{3} + 1 - \frac{1}{3}\right] = \frac{3}{4}\left[x - \frac{x^3}{3} + \frac{2}{3}\right].$$

   - Si $x \in \left]0,1\right[$ y $0 \leqslant y \leqslant x^2$ o $x \in \left]1,+\infty\right[$ y $y \in \left]0,1\right[$ (zona R_4):

   $$F_{(X,Y)}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\,dv\,du = \int_{-\sqrt{y}}^{\sqrt{y}} \int_{u^2}^{y} \frac{3}{4}\,dv\,du = \frac{3}{4} \int_{-\sqrt{y}}^{\sqrt{y}} \left[y - u^2\right]\,du =$$
   $$= \frac{3}{4}\left[yu - \frac{u^3}{3}\right]_{-\sqrt{y}}^{\sqrt{y}} = \frac{3}{4}\left[y\sqrt{y} - \frac{y^{3/2}}{3} + y\sqrt{y} - \frac{y^{3/2}}{3}\right] = y\sqrt{y}$$

   - Si $x \geqslant 1$ y $y \geqslant 1$ (zona R_5):

   $$F_{(X,Y)}(x,y) = 1.$$

Por tanto, tenemos que:

$$F_{(X,Y)}(x,y) = \begin{cases} 0, & x \leqslant -1 \text{ o } y \leqslant 0 \text{ o } x \in]-1,0[\text{ y } y \leqslant x^2, \\ \dfrac{3}{4}\left[xy - \dfrac{x^3}{3} + \dfrac{2}{3}\cdot y\sqrt{y}\right], & x^2 \leqslant y \leqslant 1, \\ \dfrac{3}{4}\left[x - \dfrac{x^3}{3} + \dfrac{2}{3}\right], & y \geqslant 1 \text{ y } x \in]-1,1[, \\ y\sqrt{y}, & x \in]0,1[\text{ y } 0 \leqslant y \leqslant x^2 \text{ o } x > 1 \text{ y } y \in]0,1[, \\ 1, & x \geqslant 1 \text{ y } y \geqslant 1. \end{cases}$$

3. Calcular $P(X \geqslant Y)$.

   Para calcular $P(X \geqslant Y)$, representamos la recta $y = x$ y nos quedamos con la región que queda por debajo de esta recta. Tenemos entonces que:

$$\begin{aligned} P(X \geqslant Y) &= \int_0^1 \int_{x^2}^{x} \frac{3}{4}\,dy\,dx = \frac{3}{4}\int_0^1 \left[x - x^2\right]\,dx = \frac{3}{4}\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \\ &= \frac{3}{4}\left[\frac{1}{2} - \frac{1}{3}\right] = \frac{3}{4}\left[\frac{3-2}{6}\right] = \frac{3}{4}\left[\frac{1}{6}\right] = \frac{1}{8}. \end{aligned}$$

4. Calcular las distribuciones marginales.

   Para $x \in]-1,1[$, tenemos que:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy = \int_{x^2}^{1} \frac{3}{4}\,dy = \frac{3}{4}\,[y]_{x^2}^{1} = \frac{3}{4}\left[1 - x^2\right].$$

   Para $y \in]0,1[$, tenemos que:

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{3}{4}\,dx = \frac{3}{4}\,[x]_{-\sqrt{y}}^{\sqrt{y}} = \frac{3}{4}\left[\sqrt{y} + \sqrt{y}\right] = \frac{3}{2}\sqrt{y}.$$

5. Calcular las distribuciones condicionadas.

   Dado $x^* \in]-1,1[$, tenemos que:

$$f_{Y|X=x^*}(y) = \frac{f_{(X,Y)}(x^*,y)}{f_X(x^*)} = \frac{3/4}{3/4\,[1-(x^*)^2]} = \frac{1}{1-(x^*)^2}.$$

   Dado $y^* \in]0,1[$, tenemos que:

$$f_{X|Y=y^*}(x) = \frac{f_{(X,Y)}(x,y^*)}{f_Y(y^*)} = \frac{3/4}{3/2\sqrt{y^*}} = \frac{1}{2\sqrt{y^*}}.$$

Ejercicio 7.2.8. Sea $k \in \mathbb{R}$. Consideramos la función de densidad de probabilidad

$$f(x,y) = \begin{cases} k\left[\frac{xy}{2} + 1\right], & 0 < x < 1, -1 < y < 1, \\ 0, & \text{en otro caso.} \end{cases}$$

Calcular:

1. La constante k para que f sea una función de densidad.

   Para que f sea una función de densidad, tenemos que:

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

   Tenemos por tanto que:

$$1 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = \int_{-1}^{1}\int_{0}^{1} k\left[\frac{xy}{2}+1\right]\,dx\,dy = k\int_{-1}^{1}\left[\frac{yx^2}{4}+x\right]_0^1 =$$
$$= k\int_{-1}^{1}\left[\frac{y}{4}+1\right]\,dy = k\left[\frac{y^2}{8}+y\right]_{-1}^{1} = k\left[\frac{1}{8}+1-\left(\frac{1}{8}-1\right)\right] = 2k \Longrightarrow k = \frac{1}{2}.$$

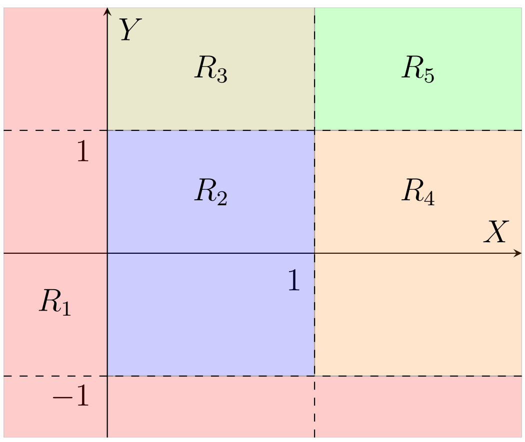
   Veamos ahora que, para dicho valor de k, f es no negativa. Para ello, tenemos que:

$$f(x,y) \geqslant 0 \Longleftrightarrow xy > -2 \Longleftrightarrow y > \frac{-2}{x}$$

   Esto último es cierto, ya que $x \in]0,1[$ e $y \in]-1,1[$. Por tanto, f es no negativa. Además, es integrable, por lo que f es una función de densidad.

2. La función de distribución de probabilidad.

   Representamos en el plano cartesiano la región en la que la función de densidad es no nula:



   Distinguimos casos:

   - Si $x \leqslant 0$ o $y \leqslant -1$ (zona R_1):

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = 0.$$

- Si $x \in]0,1[$ y $y \in]-1,1[$ (zona R_2):

$$\begin{aligned}
F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{0}^{x}\int_{-1}^{y} \frac{1}{2}\left[\frac{uv}{2}+1\right]\,dv\,du = \\
&= \int_0^x \left[\frac{uv^2}{8}+\frac{v}{2}\right]_{-1}^{y} du = \int_0^x \frac{uy^2}{8}+\frac{y}{2}-\frac{u}{8}+\frac{1}{2}\,du = \\
&= \left[\frac{u^2y^2}{16}+\frac{y}{2}u-\frac{u^2}{16}+\frac{u}{2}\right]_0^x = \frac{x^2y^2}{16}+\frac{xy}{2}-\frac{x^2}{16}+\frac{x}{2}.
\end{aligned}$$

- Si $x \in]0,1[$ y $y \geqslant 1$ (zona R_3):

$$\begin{aligned}
F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{0}^{x}\int_{-1}^{1} \frac{1}{2}\left[\frac{uv}{2}+1\right]\,dv\,du = \\
&= \int_0^x \left[\frac{uv^2}{8}+\frac{v}{2}\right]_{-1}^{1} du = \int_0^x \frac{u}{8}+\frac{1}{2}-\left(\frac{u}{8}-\frac{1}{2}\right)\,du = \int_0^x 1\,du = x.
\end{aligned}$$

- Si $y \in]-1,1[$ y $x \geqslant 1$ (zona R_4):

$$\begin{aligned}
F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{0}^{1}\int_{-1}^{y} \frac{1}{2}\left[\frac{uv}{2}+1\right]\,dv\,du = \\
&= \int_0^1 \left[\frac{uv^2}{8}+\frac{v}{2}\right]_{-1}^{y} du = \int_0^1 \frac{uy^2}{8}+\frac{y}{2}-\frac{u}{8}+\frac{1}{2}\,du = \\
&= \left[\frac{u^2y^2}{16}+\frac{y}{2}u-\frac{u^2}{16}+\frac{u}{2}\right]_0^1 = \frac{y^2}{16}+\frac{y}{2}+\frac{7}{16}
\end{aligned}$$

- Si $y \geqslant 1$ y $x \geqslant 1$ (zona R_5):

$$F_{(X,Y)}(x,y) = 1.$$

Por tanto, tenemos que:

$$F_{(X,Y)}(x,y) = \begin{cases}
0, & x \leqslant 0 \text{ o } y \leqslant -1, \\
\dfrac{x^2y^2}{16}+\dfrac{xy}{2}-\dfrac{x^2}{16}+\dfrac{x}{2}, & x \in]0,1[\text{ y } y \in]-1,1[, \\
x, & x \in]0,1[\text{ y } y \geqslant 1, \\
\dfrac{y^2}{16}+\dfrac{y}{2}+\dfrac{7}{16}, & y \in]-1,1[\text{ y } x \geqslant 1, \\
1, & y \geqslant 1 \text{ y } x \geqslant 1.
\end{cases}$$

3. Las distribuciones marginales.

   Para $x \in]0,1[$, tenemos que:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy = \int_{-1}^{1} \frac{1}{2}\left[\frac{xy}{2}+1\right]\,dy = \frac{1}{2}\left[\frac{xy^2}{4}+y\right]_{-1}^{1} = \frac{1}{2}\left[\frac{x}{4}+1-\frac{x}{4}+1\right] = 1$$

   Para $y \in]-1,1[$, tenemos que:

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx = \int_{0}^{1} \frac{1}{2}\left[\frac{xy}{2}+1\right]\,dx = \frac{1}{2}\left[\frac{x^2y}{4}+x\right]_{0}^{1} = \frac{1}{2}\left[\frac{y}{4}+1\right]$$

Ejercicio 7.2.9. Sea (X, Y) un vector aleatorio bidimensional continuo, con función de densidad de probabilidad

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

Responder a los siguientes apartados:

1. Hallar la constante k para que f sea una función de densidad de probabilidad.

   Para que f sea una función de densidad, tenemos que:

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

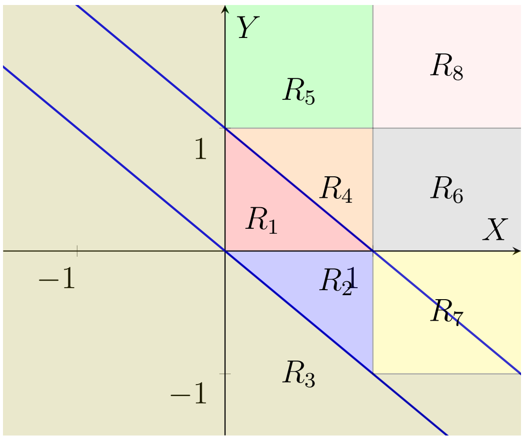
   Tenemos por tanto que:

$$1 = \int_0^1 \int_{-x}^{1-x} k\, dx\, dy = k \int_0^1 [y]_{-x}^{1-x}\, dx = k \int_0^1 1 - x + x\, dx = k \int_0^1 1\, dx = k$$

   Por tanto, tenemos que $k = 1$. En este caso, vemos que además $f_{(X,Y)}$ es no negativa e integrable.

2. Calcular la función de distribución de probabilidad.

   Dividimos el plano cartesiano en las distintas regiones:



   Distinguimos casos:

   - Si $x \leqslant 0$ o $x + y \leqslant 0$ (zona R_3):

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, du\, dv = 0.$$

   - Si $x \in]0,1[$ y $y \in]0,1[$ y $x + y \leqslant 1$ (zona R_1):

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, dv\, du = \int_0^x \int_{-u}^{y} 1\, dv\, du = \int_0^x [v]_{-u}^{y}\, du =$$
$$= \int_0^x y + u\, du = \left[yu + \frac{u^2}{2}\right]_0^x = xy + \frac{x^2}{2}.$$

- <u>Si $x \in]0,1[\quad$ y $\quad y \in]-1,0[\quad$ y $\quad x+y \geqslant 0$</u> (zona R_2):

$$\begin{aligned} F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{-y}^{x}\int_{-u}^{y} 1\,dv\,du = \int_{-y}^{x} [v]_{-u}^{y}\,du = \\ &= \int_{-y}^{x} y+u\,du = \left[yu+\frac{u^2}{2}\right]_{-y}^{x} = xy+\frac{x^2}{2}-y(-y)-\frac{y^2}{2} = \\ &= yx+\frac{x^2+y^2}{2} = \frac{(x+y)^2}{2}. \end{aligned}$$

- <u>Si $x \in]0,1[\quad$ y $\quad y \in]0,1[\quad$ y $\quad x+y \geqslant 1$</u> (zona R_4):

$$\begin{aligned} F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \\ &= \int_{0}^{1-y}\int_{0}^{y} 1\,dv\,du + \int_{1-y}^{x}\int_{0}^{1-u} 1\,dv\,du + \int_{0}^{x}\int_{-u}^{0} 1\,dv\,du = \\ &= \int_{0}^{1-y} y\,du + \int_{1-y}^{x} 1-u\,du + \int_{0}^{x} u\,du = \\ &= [yu]_0^{1-y} + \left[u-\frac{u^2}{2}\right]_{1-y}^{x} + \left[\frac{u^2}{2}\right]_0^{x} = \\ &= y(1-y) + \left(x-\cancel{\frac{x^2}{2}}\right) - \left(1-y-\frac{(1-y)^2}{2}\right) + \cancel{\frac{x^2}{2}} = \\ &= y-y^2+x-1+y+\frac{(1-y)^2}{2} = -(1-y)^2+x+\frac{(1-y)^2}{2} = \\ &= x-\frac{(1-y)^2}{2} \end{aligned}$$

- <u>Si $x \in]0,1[\quad$ y $\quad y \in]1,\infty[$</u> (zona R_5):

$$\begin{aligned} F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{0}^{x}\int_{-u}^{1-u} 1\,dv\,du = \\ &= \int_{0}^{x} [v]_{-u}^{1-u}\,du = \int_{0}^{x} 1-u+u\,du = \int_{0}^{x} 1\,du = x. \end{aligned}$$

- <u>Si $x \in]1,\infty[\quad$ y $\quad y \in]0,1[$</u> (zona R_6):

$$\begin{aligned} F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \\ &= \int_{0}^{1}\int_{-u}^{0} 1\,dv\,du + \int_{0}^{1-y}\int_{0}^{y} 1\,dv\,du + \int_{1-y}^{1}\int_{0}^{1-u} 1\,dv\,du = \\ &= \int_{0}^{1} u\,du + \int_{0}^{1-y} y\,du + \int_{1-y}^{1} 1-u\,du = \\ &= \left[\frac{u^2}{2}\right]_0^{1} + [yu]_0^{1-y} + \left[u-\frac{u^2}{2}\right]_{1-y}^{1} = \\ &= \cancel{\frac{1}{2}} + y(1-y) + 1 - \cancel{\frac{1}{2}} - \left[1-y-\frac{(1-y)^2}{2}\right] = \\ &= y-y^2+y+\frac{(1-y)^2}{2} = 1-\frac{(1-y)^2}{2} \end{aligned}$$

- Si $x \in]1, \infty[$ y $y \in]-1, 0[$ (zona R_7):

$$\begin{aligned}
F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \\
&= \int_{-y}^{1}\int_{-u}^{y} 1\,dv\,du = \int_{-y}^{1} [v]_{-u}^{y}\,du = \int_{-y}^{1} y+u\,du = \\
&= \left[yu + \frac{u^2}{2}\right]_{-y}^{1} = y + \frac{1}{2} + y^2 - \frac{y^2}{2} = y + \frac{y^2+1}{2}
\end{aligned}$$

- Si $x \in]1, \infty[$ y $y \in]1, \infty[$ (zona R_8):

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = 1$$

Por tanto, tenemos que:

$$F_{(X,Y)}(x,y) = \begin{cases}
0, & x \leqslant 0 \text{ o } x+y \leqslant 0,\ (R_3), \\
xy + \dfrac{x^2}{2}, & x \in]0,1[\text{ y } y \in]0,1[\text{ y } x+y \leqslant 1,\ (R_1), \\
\dfrac{(x+y)^2}{2}, & x \in]0,1[\text{ y } y \in]-1,0[\text{ y } x+y \geqslant 0,\ (R_2), \\
x - \dfrac{(1-y)^2}{2}, & x \in]0,1[\text{ y } y \in]0,1[\text{ y } x+y \geqslant 1,\ (R_4), \\
x, & x \in]0,1[\text{ y } y \geqslant 1,\ (R_5), \\
1 - \dfrac{(1-y)^2}{2}, & x \in]1,\infty[\text{ y } y \in]0,1[,\ (R_6), \\
y + \dfrac{y^2+1}{2}, & x \in]1,\infty[\text{ y } y \in]-1,0[,\ (R_7), \\
1, & x \in]1,\infty[\text{ y } y \geqslant 1,\ (R_8).
\end{cases}$$

3. Calcular las distribuciones marginales.

   Para $x \in]0, 1[$, ya que la función de densidad es constante, tenemos que:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy = \int_{-x}^{1-x} 1\,dy = [y]_{-x}^{1-x} = 1 - x + x = 1.$$

   Para $y \in]0, 1[$, ya que la función de densidad es constante, tenemos que:

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx = \int_{0}^{1-y} 1\,dx = [x]_{0}^{1-y} = 1 - y.$$

   Para $y \in]-1, 0[$, ya que la función de densidad es constante, tenemos que:

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx = \int_{-y}^{1} 1\,dx = [x]_{-y}^{1} = 1 + y.$$

   Por tanto, tenemos que, para $y \in]-1, 1[$:

$$f_Y(y) = 1 - |y|.$$

4. Calcular las distribuciones condicionadas.

   Dado $x^* \in]-1,1[$, tenemos para $y \in]0,1[$, $0 < x^* + y < 1$:

$$f_{Y|X=x^*}(y) = \frac{f_{(X,Y)}(x^*, y)}{f_X(x^*)} = \frac{1}{1} = 1.$$

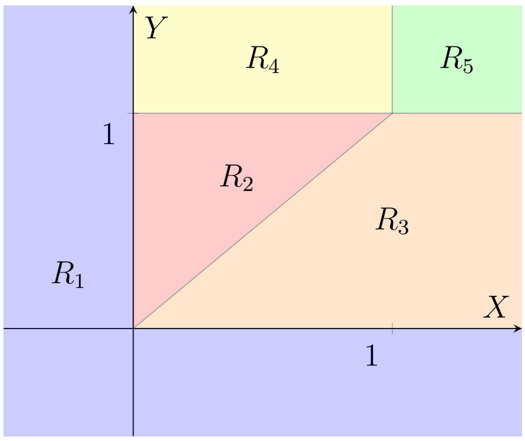
   Dado $y^* \in]-1,1[$, tenemos para $x \in]0,1[$, $0 < x + y^* < 1$:

$$f_{X|Y=y^*}(x) = \frac{f_{(X,Y)}(x, y^*)}{f_Y(y^*)} = \frac{1}{1-|y^*|}.$$

Ejercicio 7.2.10. Sea (X,Y) un vector aleatorio bidimensional continuo, con distribución de probabilidad uniforme sobre el triángulo de vértices $(0,0)$; $(0,1)$; $(1,1)$. Determinar:

1. La función de densidad de probabilidad.

   Veamos en primer lugar el triángulo en cuestión:



   La función de densidad de probabilidad es constante, por lo que:

$$f(x,y) = \begin{cases} k, & x \in [0,1], x \leqslant y \leqslant 1, \\ 0, & \text{en otro caso.} \end{cases}$$

   Para que f sea una función de densidad, tenemos que:

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = \int_0^1\int_x^1 k\,dy\,dx = k\int_0^1 [y]_x^1\,dx = \\ &= k\int_0^1 1-x\,dx = k\left[x - \frac{x^2}{2}\right]_0^1 = k\left(1-\frac{1}{2}\right) = \frac{k}{2} \Longrightarrow k = 2. \end{aligned}$$

2. La función de distribución de probabilidad.

   Distinguimos casos:

- Si $x \leqslant 0 \quad \text{o} \quad y \leqslant 0$ (zona R_1):

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = 0.$$

- Si $x \in]0,1[\quad \text{y} \quad x < y < 1$ (zona R_2):

$$\begin{aligned}F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{0}^{x}\int_{u}^{y} 2\,dv\,du = \int_{0}^{x} 2(y-u)\,du = \\ &= 2\left[yu - \frac{u^2}{2}\right]_0^x = 2\left(xy - \frac{x^2}{2}\right) = 2xy - x^2.\end{aligned}$$

- Si $y \in]0,1,\infty[\quad \text{y} \quad x > y$ (zona R_3):

$$\begin{aligned}F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{0}^{y}\int_{u}^{y} 2\,dv\,du = \int_{0}^{y} 2(y-u)\,du = \\ &= 2\left[yu - \frac{u^2}{2}\right]_0^y = 2\left(y^2 - \frac{y^2}{2}\right) = 2\cdot\frac{y^2}{2} = y^2.\end{aligned}$$

- Si $x \in]0,1[\quad \text{y} \quad y \geqslant 1$ (zona R_4):

$$\begin{aligned}F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = \int_{0}^{x}\int_{u}^{1} 2\,dv\,du = \int_{0}^{x} 2(1-u)\,du = \\ &= 2\left[u - \frac{u^2}{2}\right]_0^x = 2\left(x - \frac{x^2}{2}\right) = 2x - x^2.\end{aligned}$$

- Si $x, y \geqslant 1$ (zona R_5):

$$F_{(X,Y)}(x,y) = 1.$$

Por tanto, tenemos que:

$$F_{(X,Y)}(x,y) = \begin{cases} 0, & x \leqslant 0 \text{ o } y \leqslant 0, (R_1), \\ 2xy - x^2, & x \in]0,1[\text{ y } x < y < 1, (R_2), \\ y^2, & y \in]0,1[\text{ y } x > y, (R_3), \\ 2x - x^2, & x \in]0,1[\text{ y } y \geqslant 1, (R_4), \\ 1, & x, y \geqslant 1, (R_5). \end{cases}$$

3. Las distribuciones marginales.

   Para $x \in]0,1[$, ya que la función de densidad es constante, tenemos que:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy = \int_{x}^{1} 2\,dy = 2\,[y]_x^1 = 2(1-x).$$

   Para $y \in]0,1[$, ya que la función de densidad es constante, tenemos que:

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx = \int_{0}^{y} 2\,dx = 2\,[x]_0^y = 2y.$$

4. Las distribuciones condicionadas.

   Dado $x^* \in]0,1[$, tenemos para $y \in]0,1[$, $x^* < y < 1$:

$$f_{Y|X=x^*}(y) = \frac{f_{(X,Y)}(x^*, y)}{f_X(x^*)} = \frac{2}{2(1-x^*)} = \frac{1}{1-x^*}.$$

   Dado $y^* \in]0,1[$, tenemos para $x \in]0,1[$, $x < y^* < 1$:

$$f_{X|Y=y^*}(x) = \frac{f_{(X,Y)}(x, y^*)}{f_Y(y^*)} = \frac{2}{2y^*} = \frac{1}{y^*}.$$

Ejercicio 7.2.11. Sea (X,Y) una variable aleatoria bidimensional con distribución uniforme en el recinto

$$C = \{(x,y) \in \mathbb{R}^2; x^2 + y^2 < 1; x \geqslant 0, y \geqslant 0\}.$$

Calcular:

1. La función de densidad conjunta.

   La función de densidad conjunta es constante, por lo que:

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

   Para que f sea una función de densidad, tenemos que:

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

   Hay dos opciones:

   Integrando de la forma usual: Es necesario que:

$$1 = \int_0^1 \int_0^{\sqrt{1-x^2}} k\,dy\,dx = k \int_0^1 \sqrt{1-x^2}\,dx$$

   Haciendo el cambio de variable $x = \operatorname{sen}(t)$, tenemos que:

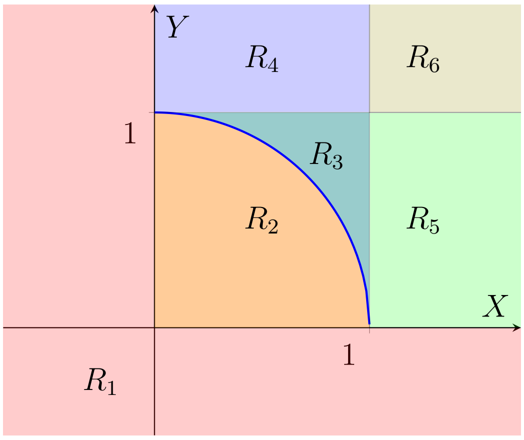
$$\begin{aligned} 1 = k \int_0^{\frac{\pi}{2}} \cos(t)\cos(t)\,dt = k \int_0^{\frac{\pi}{2}} \cos^2(t)\,dt = k \int_0^{\frac{\pi}{2}} \frac{1+\cos(2t)}{2}\,dt = \\ = k \left[\frac{t}{2} + \frac{\operatorname{sen}(2t)}{4}\right]_0^{\frac{\pi}{2}} = k \left[\frac{\pi}{4}\right] \Longrightarrow k = \frac{4}{\pi}. \end{aligned}$$

   Razonando la forma de C: Sabemos que C es un cuarto de círculo de radio 1, por lo que su área es $\pi/4$. Por tanto, tenemos que:

$$1 = \int_C f(x,y) = k \int_C 1 = k \cdot \lambda(C) = k \cdot \frac{\pi}{4} \Longrightarrow k = \frac{4}{\pi}.$$

2. La función de distribución conjunta.

   Dividimos el plano cartesiano en las distintas regiones:



   Distinguimos casos:

   - Si $x \leqslant 0 \quad$ o $\quad y \leqslant 0$ (zona R_1):

$$F_{(X,Y)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = 0.$$

   - Si $x \in]0,1[\quad$ y $\quad y \in \left]0, \sqrt{1-x^2}\right[$ (zona R_2):

$$\begin{aligned}F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dv\,du = \int_{0}^{x}\int_{0}^{y} \frac{4}{\pi}\,dv\,du = \int_{0}^{x} \frac{4}{\pi} y\,du = \\ &= \frac{4}{\pi}\left[yu\right]_0^x = \frac{4}{\pi}\cdot xy\end{aligned}$$

   - Si $x \in]0,1[\quad$ y $\quad y \in \left]\sqrt{1-x^2}, 1\right[$ (zona R_3):

$$\begin{aligned}F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dv\,du = \\ &= \int_{0}^{\sqrt{1-y^2}}\int_{0}^{y} \frac{4}{\pi}\,dv\,du + \int_{\sqrt{1-y^2}}^{x}\int_{0}^{\sqrt{1-u^2}} \frac{4}{\pi}\,dv\,du = \\ &= \int_{0}^{\sqrt{1-y^2}} \frac{4}{\pi} y\,du + \int_{\sqrt{1-y^2}}^{x} \frac{4}{\pi}\sqrt{1-u^2}\,du\end{aligned}$$

     Para resolver la segunda integral, hacemos el cambio de variable dado por $u = \operatorname{sen}(t)$, $du = \cos(t)dt$:

$$\begin{aligned}
F_{(X,Y)}(x,y) &= \frac{4}{\pi}y\sqrt{1-y^2} + \frac{4}{\pi}\int_{\operatorname{arc\,sen}(\sqrt{1-y^2})}^{\operatorname{arc\,sen}(x)} \cos^2(t)\,dt = \\
&= \frac{4}{\pi}y\sqrt{1-y^2} + \frac{4}{\pi}\int_{\operatorname{arc\,sen}(\sqrt{1-y^2})}^{\operatorname{arc\,sen}(x)} \frac{1+\cos(2t)}{2}\,dt = \\
&= \frac{4}{\pi}y\sqrt{1-y^2} + \frac{4}{\pi}\left[\frac{t}{2} + \frac{\operatorname{sen}(2t)}{4}\right]_{\operatorname{arc\,sen}(\sqrt{1-y^2})}^{\operatorname{arc\,sen}(x)} = \\
&= \frac{4}{\pi}y\sqrt{1-y^2} + \frac{4}{\pi}\left[\frac{\operatorname{arc\,sen}(x)}{2} + \frac{\operatorname{sen}(2\operatorname{arc\,sen}(x))}{4} - \right. \\
&\qquad \left. - \frac{\operatorname{arc\,sen}(\sqrt{1-y^2})}{2} - \frac{\operatorname{sen}(2\operatorname{arc\,sen}(\sqrt{1-y^2}))}{4}\right]
\end{aligned}$$

Veamos cuánto vale anteriormente $\operatorname{sen}(2\operatorname{arc\,sen}(x))$ para cierto $x \in \mathbb{R}$:

$$\operatorname{sen}(2\operatorname{arc\,sen}(x)) = 2\operatorname{sen}(\operatorname{arc\,sen}(x))\cos(\operatorname{arc\,sen}(x)) = 2x\sqrt{1-x^2}.$$

Por tanto, tenemos que:

$$\begin{aligned}
F_{(X,Y)}(x,y) &= \frac{4}{\pi}y\sqrt{1-y^2} + \frac{4}{\pi}\left[\frac{\operatorname{arc\,sen}(x)}{2} + \frac{2x\sqrt{1-x^2}}{4} - \right. \\
&\qquad \left. - \frac{\operatorname{arc\,sen}(\sqrt{1-y^2})}{2} - \frac{2\sqrt{1-y^2}\sqrt{y^2}}{4}\right] = \\
&= \frac{4}{\pi}y\sqrt{1-y^2} + \frac{2}{\pi}\operatorname{arc\,sen}(x) + \frac{2}{\pi}x\sqrt{1-x^2} - \\
&\qquad - \frac{2}{\pi}\operatorname{arc\,sen}(\sqrt{1-y^2}) - \frac{2}{\pi}y\sqrt{1-y^2}. = \\
&= \frac{2}{\pi}y\sqrt{1-y^2} + \frac{2}{\pi}\operatorname{arc\,sen}(x) + \frac{2}{\pi}x\sqrt{1-x^2} - \frac{2}{\pi}\operatorname{arc\,sen}(\sqrt{1-y^2})
\end{aligned}$$

- Si $x \in]0,1[$ y $y \geqslant 1$ (zona R_4):

$$\begin{aligned}
F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dv\,du = \int_0^x\int_0^{\sqrt{1-u^2}} \frac{4}{\pi}\,dv\,du = \\
&= \int_0^x \frac{4}{\pi}\sqrt{1-u^2}\,du
\end{aligned}$$

Para resolver la integral, de nuevo hacemos el cambio de variable dado por $u = \operatorname{sen}(t)$, $du = \cos(t)dt$:

$$\begin{aligned}
F_{(X,Y)}(x,y) &= \frac{4}{\pi}\int_0^{\operatorname{arc\,sen}(x)} \cos^2(t)\,dt = \frac{4}{\pi}\int_0^{\operatorname{arc\,sen}(x)} \frac{1+\cos(2t)}{2}\,dt = \\
&= \frac{4}{\pi}\left[\frac{t}{2} + \frac{\operatorname{sen}(2t)}{4}\right]_0^{\operatorname{arc\,sen}(x)} = \frac{4}{\pi}\left[\frac{\operatorname{arc\,sen}(x)}{2} + \frac{\operatorname{sen}(2\operatorname{arc\,sen}(x))}{4}\right] = \\
&= \frac{4}{\pi}\left[\frac{\operatorname{arc\,sen}(x)}{2} + \frac{2x\sqrt{1-x^2}}{4}\right] = \frac{2}{\pi}\operatorname{arc\,sen}(x) + \frac{2}{\pi}x\sqrt{1-x^2}.
\end{aligned}$$

- Si $y \in]0,1[$ y $x \geqslant 1$ (zona R_5):

$$\begin{aligned}
F_{(X,Y)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dv\,du = \\
&= \int_0^{\sqrt{1-y^2}}\int_0^y \frac{4}{\pi}\,dv\,du + \int_{\sqrt{1-y^2}}^{1}\int_0^{\sqrt{1-u^2}} \frac{4}{\pi}\,dv\,du = \\
&= \int_0^{\sqrt{1-y^2}} \frac{4}{\pi} y\,du + \int_{\sqrt{1-y^2}}^{1} \frac{4}{\pi}\sqrt{1-u^2}\,du
\end{aligned}$$

Para resolver la segunda integral, de nuevo hacemos el cambio de variable dado por $u = \operatorname{sen}(t)$, $du = \cos(t)dt$:

$$\begin{aligned}
F_{(X,Y)}(x,y) &= \frac{4}{\pi} y\,[u]_0^{\sqrt{1-y^2}} + \frac{4}{\pi}\int_{\operatorname{arc\,sen}(\sqrt{1-y^2})}^{\pi/2} \cos^2(t)\,dt = \\
&= \frac{4}{\pi} y\sqrt{1-y^2} + \frac{4}{\pi}\int_{\operatorname{arc\,sen}(\sqrt{1-y^2})}^{\pi/2} \frac{1+\cos(2t)}{2}\,dt = \\
&= \frac{4}{\pi} y\sqrt{1-y^2} + \frac{4}{\pi}\left[\frac{t}{2} + \frac{\operatorname{sen}(2t)}{4}\right]_{\operatorname{arc\,sen}(\sqrt{1-y^2})}^{\pi/2} = \\
&= \frac{4}{\pi} y\sqrt{1-y^2} + \frac{4}{\pi}\left[\frac{\pi/2}{2} + \frac{\operatorname{sen}(\pi)}{4} - \frac{\operatorname{arc\,sen}(\sqrt{1-y^2})}{2} - \right. \\
&\qquad \left. - \frac{\operatorname{sen}(2\operatorname{arc\,sen}(\sqrt{1-y^2}))}{4}\right] = \\
&= \frac{4}{\pi} y\sqrt{1-y^2} + 1 - \frac{2}{\pi}\operatorname{arc\,sen}(\sqrt{1-y^2}) - \frac{2}{\pi} y\sqrt{1-y^2} = \\
&= \frac{2}{\pi} y\sqrt{1-y^2} + 1 - \frac{2}{\pi}\operatorname{arc\,sen}(\sqrt{1-y^2})
\end{aligned}$$

- Si $x, y \geqslant 1$ (zona R_6):

$$F_{(X,Y)}(x,y) = 1.$$

Por tanto, tenemos que:

$$F_{(X,Y)}(x,y) = \begin{cases}
0, & (x,y) \in R_1, \\
4/\pi\, xy, & (x,y) \in R_2, \\
2/\pi\left[y\sqrt{1-y^2} + x\sqrt{1-x^2} - \operatorname{arc\,sen}(\sqrt{1-y^2})\right], & (x,y) \in R_3 \\
2/\pi\left[\operatorname{arc\,sen}(x) + x\sqrt{1-x^2}\right], & (x,y) \in R_4, \\
2/\pi\left[y\sqrt{1-y^2} + \pi/2 - \operatorname{arc\,sen}(\sqrt{1-y^2})\right], & (x,y) \in R_5, \\
1, & (x,y) \in R_6.
\end{cases}$$

3. Las funciones de densidad marginales.

Para $x \in [0,1]$, ya que la función de densidad es constante, tenemos que:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy = \int_0^{\sqrt{1-x^2}} \frac{4}{\pi}\,dy = \frac{4}{\pi}\,[y]_0^{\sqrt{1-x^2}} = \frac{4}{\pi}\cdot\sqrt{1-x^2}.$$

Para $y \in [0,1]$, ya que la función de densidad es constante, tenemos que:

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx = \int_0^{\sqrt{1-y^2}} \frac{4}{\pi}\,dx = \frac{4}{\pi}\,[x]_0^{\sqrt{1-y^2}} = \frac{4}{\pi}\cdot\sqrt{1-y^2}.$$

4. Las funciones de densidad condicionadas.
   Dado $x^* \in [0,1]$, tenemos para $y \in [0, \sqrt{1-(x^*)^2}]$:

$$f_{Y|X=x^*}(y) = \frac{f_{(X,Y)}(x^*,y)}{f_X(x^*)} = \frac{4/\pi}{\dfrac{4}{\pi}\cdot\sqrt{1-(x^*)^2}} = \frac{1}{\sqrt{1-(x^*)^2}}.$$

   Dado $y^* \in [0,1]$, tenemos para $x \in [0, \sqrt{1-(y^*)^2}]$:

$$f_{X|Y=y^*}(x) = \frac{f_{(X,Y)}(x,y^*)}{f_Y(y^*)} = \frac{4/\pi}{\dfrac{4}{\pi}\cdot\sqrt{1-(y^*)^2}} = \frac{1}{\sqrt{1-(y^*)^2}}.$$

7.3. Vectores Aleatorios. Parte 2

Ejercicio 7.3.1. Sea la función de densidad del vector (X, Y) dada por:

$$f(x,y) = \begin{cases} k\left[\dfrac{xy}{2}+1\right] & 0<x<1,\ -1<y<1, \\ 0 & \text{en otro caso.} \end{cases}$$

Se considera la transformación $Z = X - Y$, $T = X + 2Y$. Hallar la función de densidad de probabilidad conjunta de la variable transformada (Z, T).

Buscamos en primer lugar el valor de k:

$$\begin{aligned}
1 &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = \int_0^1\int_{-1}^1 k\left[\frac{xy}{2}+1\right]dy\,dx = \\
&= k\int_0^1 \left[\frac{xy^2}{4}+y\right]_{-1}^1 dx = k\int_0^1\left[\frac{x}{4}+1-\frac{x}{4}+1\right]dx = \\
&= k\int_0^1 2\,dx = k\,[2x]_0^1 = 2k \Longrightarrow k=\frac{1}{2}
\end{aligned}$$

Definimos ahora la función:

$$\begin{array}{rccc} g: & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & (X,Y) & \longmapsto & (Z,T) = (X-Y, X+2Y) \end{array}$$

Para obtener g^{-1}, buscamos obtener X, Y en función de Z, T:

$$\left\{\begin{array}{l} Z = X-Y, \\ T = X+2Y. \end{array}\right\} \Longrightarrow Z - T = -3Y \Longrightarrow Y = \frac{T-Z}{3} \Longrightarrow X = \frac{2Z+T}{3}.$$

Por tanto, tenemos que:

$$\begin{array}{rccc} g^{-1}: & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & (Z,T) & \longmapsto & (X,Y) = \left(\dfrac{2Z+T}{3}, \dfrac{T-Z}{3}\right) \end{array}$$

Tenemos que todas las componentes de g^{-1} son derivables:

$$\begin{array}{ll}
\dfrac{\partial X}{\partial Z}(T,Z) = 2/3, & \dfrac{\partial X}{\partial T}(Z,T) = 1/3, \\
\dfrac{\partial Y}{\partial Z}(Z,T) = -1/3, & \dfrac{\partial Y}{\partial T}(Z,T) = 1/3.
\end{array}$$

Además, tenemos que:

$$\det Jg^{-1}(z,t) = \begin{vmatrix} 2/3 & 1/3 \\ -1/3 & 1/3 \end{vmatrix} = \frac{1}{3^2}\begin{vmatrix} 2 & 1 \\ -1 & 1 \end{vmatrix} = \frac{1}{3^2}(2-(-1)) = \frac{3}{3^2} = \frac{1}{3} \neq 0 \qquad \forall (z,t)\in\mathbb{R}^2.$$

Por tanto, $(Z,T) = g(X,Y)$ es un vector aleatorio continuo. Buscamos ahora la función de densidad de probabilidad de (Z,T):

$$f_{(Z,T)}(z,t) = f_{(X,Y)}\left(\frac{2z+t}{3}, \frac{t-z}{3}\right)\cdot\left|\det Jg^{-1}(z,t)\right| =$$

$$= \begin{cases} \dfrac{1}{3}\cdot\dfrac{1}{2}\left[\dfrac{\frac{2z+t}{3}\cdot\frac{t-z}{3}}{2}+1\right] = \dfrac{1}{6}\left[\dfrac{(2z+t)(t-z)}{18}+1\right] & 0<\dfrac{2z+t}{3}<1,\ -1<\dfrac{t-z}{3}<1, \\ 0 & \text{en otro caso.} \end{cases}$$

Por tanto, la función de densidad de probabilidad de (Z,T) es:

$$f_{(Z,T)}(z,t) = \begin{cases} \dfrac{1}{6}\left[\dfrac{(2z+t)(t-z)}{18}+1\right] & 0<2z+t<3,\ -3<t-z<3, \\ 0 & \text{en otro caso.} \end{cases}$$

Ejercicio 7.3.2. Sea (X,Y) un vector aleatorio bidimensional con función de densidad de probabilidad

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

Calcular la función de densidad de probabilidad de la variable aleatoria $Z = X+2Y$, a partir del cálculo de la densidad de probabilidad conjunta de la variable aleatoria bidimensional transformada (Z,T), siendo $Z = X+2Y$, y $T = Y$.

Definimos la transformación:

$$\begin{array}{cccl} g: & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & (X,Y) & \longmapsto & (Z,T) = (X+2Y, Y) \end{array}$$

Para obtener g^{-1}, buscamos obtener X,Y en función de Z,T:

$$\left.\begin{cases} Z = X+2Y, \\ T = Y. \end{cases}\right\} \Longrightarrow \begin{cases} X = Z-2T, \\ Y = T. \end{cases}$$

Por tanto, tenemos que:

$$\begin{array}{cccl} g^{-1}: & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & (Z,T) & \longmapsto & (X,Y) = (Z-2T, T) \end{array}$$

Tenemos que todas las componentes de g^{-1} son derivables:

$$\begin{aligned} \frac{\partial X}{\partial Z}(Z,T) &= 1, & \frac{\partial X}{\partial T}(Z,T) &= -2, \\ \frac{\partial Y}{\partial Z}(Z,T) &= 0, & \frac{\partial Y}{\partial T}(Z,T) &= 1. \end{aligned}$$

Además, tenemos que:

$$\det Jg^{-1}(z,t) = \begin{vmatrix} 1 & -2 \\ 0 & 1 \end{vmatrix} = 1 \neq 0 \qquad \forall (z,t) \in \mathbb{R}^2.$$

Por tanto, $(Z,T) = g(X,Y)$ es un vector aleatorio continuo. Buscamos ahora la función de densidad de probabilidad de (Z,T):

$$\begin{aligned} f_{(Z,T)}(z,t) &= f_{(X,Y)}(z-2t,t) \cdot \left|\det Jg^{-1}(z,t)\right| = \\ &= \begin{cases} \exp(-(z-2t)-t) = \exp(-z+t) & z-2t>0,\ t>0, \\ 0 & \text{en otro caso.} \end{cases} \end{aligned}$$

Finalmente, hallamos la marginal de Z. Para todo $z > 0$:

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{+\infty} f_{(Z,T)}(z,t)\,dt = \int_0^{+\infty} \exp(-z+t)\,dt = \\
&= \int_0^{z/2} \exp(-z+t)\,dt = e^{-z}\left[\exp(t)\right]_0^{z/2} = e^{-z}\left[e^{z/2}-1\right]
\end{aligned}$$

Por tanto, la función de densidad de probabilidad de Z es:

$$f_Z(z) = \begin{cases} e^{-z}\left[e^{z/2}-1\right] & z > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

Ejercicio 7.3.3. Sea (X, Y) un vector aleatorio bidimensional con función de densidad de probabilidad

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

1. Calcular k para que f sea función de densidad de probabilidad de un vector aleatorio continuo (X, Y).

   Para que f sea función de densidad de probabilidad, necesitamos que:

   $$1 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = k\int_0^1\int_0^1 dy\,dx = k\int_0^1 1\,dx = k\left[x\right]_0^1 = k.$$

2. Calcular la función de densidad de probabilidad conjunta del vector bidimensional $(Z, T) = (X + Y, X - Y)$.

   Definimos la transformación:

   $$\begin{array}{rccl} g: & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & (X,Y) & \longmapsto & (Z,T) = (X+Y, X-Y) \end{array}$$

   Para obtener g^{-1}, buscamos obtener X, Y en función de Z, T:

   $$\left\{\begin{array}{l} Z = X + Y, \\ T = X - Y. \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} X = \dfrac{Z+T}{2}, \\ Y = \dfrac{Z-T}{2}. \end{array}\right.$$

   Por tanto, tenemos que:

   $$\begin{array}{rccl} g^{-1}: & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & (Z,T) & \longmapsto & (X,Y) = \left(\dfrac{Z+T}{2}, \dfrac{Z-T}{2}\right) \end{array}$$

   Tenemos que todas las componentes de g^{-1} son derivables:

   $$\begin{array}{ll} \dfrac{\partial X}{\partial Z}(Z,T) = 1/2, & \dfrac{\partial X}{\partial T}(Z,T) = 1/2, \\ \dfrac{\partial Y}{\partial Z}(Z,T) = 1/2, & \dfrac{\partial Y}{\partial T}(Z,T) = -1/2. \end{array}$$

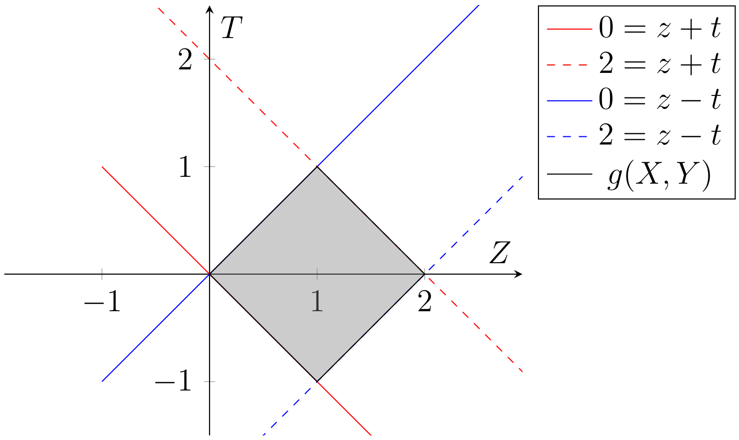
Además, tenemos que:

$$\det Jg^{-1}(z,t) = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = \frac{1}{2^2}\begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = \frac{1}{2^2}(-1-1) = -\frac{2}{2^2} = -\frac{1}{2} \neq 0 \qquad \forall (z,t) \in \mathbb{R}^2.$$

Por tanto, $(Z,T) = g(X,Y)$ es un vector aleatorio continuo. Veamos el valor de $g(X,Y)$ para $X,Y \in [0,1]$:

$$\begin{aligned} g(X,Y) &= \left\{(z,t) \in \mathbb{R}^2 \mid 0 < \frac{z+t}{2} < 1,\ 0 < \frac{z-t}{2} < 1\right\} = \\ &= \left\{(z,t) \in \mathbb{R}^2 \mid 0 < z+t < 2,\ 0 < z-t < 2\right\} \end{aligned}$$

Veámoslo gráficamente:



Buscamos ahora la función de densidad de probabilidad de (Z,T):

$$\begin{aligned} f_{(Z,T)}(z,t) &= f_{(X,Y)}\left(\frac{z+t}{2}, \frac{z-t}{2}\right) \cdot \left|\det Jg^{-1}(z,t)\right| = \\ &= \begin{cases} k \cdot \dfrac{1}{2} & (z,t) \in g(X,Y), \\ 0 & \text{en otro caso.} \end{cases} \end{aligned}$$

Por tanto, la función de densidad de probabilidad de (Z,T) es:

$$f_{(Z,T)}(z,t) = \begin{cases} \dfrac{1}{2} & 0 < z+t < 2,\ 0 < z-t < 2, \\ 0 & \text{en otro caso.} \end{cases}$$

3. Determinar las funciones de densidad de probabilidad marginales del vector transformado (Z,T).

   Para $z \in [0,2]$, tenemos que:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{(Z,T)}(z,t)\, dt$$

   Los límites de integración los vemos claros en la gráfica anterior. Distinguimos en función del valor de z:

- Si $z \in [0,1]$, entonces:

$$f_Z(z) = \int_{-z}^{z} \frac{1}{2}\, dt = \frac{1}{2}\, [t]_{-z}^{z} = \frac{1}{2}(z-(-z)) = z.$$

- Si $z \in [1,2]$, entonces:

$$f_Z(z) = \int_{z-2}^{2-z} \frac{1}{2}\, dt = \frac{1}{2}\, [t]_{z-2}^{2-z} = \frac{1}{2}(2-z-(z-2)) = \frac{1}{2}(4-2z) = 2-z.$$

Por tanto, la función de densidad de probabilidad de Z es:

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

Para $t \in [-1,1]$, tenemos que:

$$f_T(t) = \int_{-\infty}^{+\infty} f_{(Z,T)}(z,t)\, dz$$

Los límites de integración los vemos claros en la gráfica anterior. Distinguimos en función del valor de t:

- Si $t \in [-1,0]$, entonces:

$$f_T(t) = \int_{-t}^{2+t} \frac{1}{2}\, dz = \frac{1}{2}\, [z]_{-t}^{2+t} = \frac{1}{2}(2+t-(-t)) = 1+t.$$

- Si $t \in [0,1]$, entonces:

$$f_T(t) = \int_{t}^{2-t} \frac{1}{2}\, dz = \frac{1}{2}\, [z]_{t}^{2-t} = \frac{1}{2}(2-t-t) = 1-t.$$

Por tanto, la función de densidad de probabilidad de T es:

$$f_T(t) = \begin{cases} 1+t & -1 < t < 0, \\ 1-t & 0 < t < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

4. Determinar la función de distribución de probabilidad de ${}^X/_Y$ y XY.

   Definimos la transformación:

$$\begin{array}{rccc} g: & \mathbb{R}^+ \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \times \mathbb{R}^+ \\ & (X,Y) & \longmapsto & (Z,T) = ({}^X/_Y, XY) \end{array}$$

   Para obtener g^{-1}, buscamos obtener X, Y en función de Z, T:

$$\left\{\begin{array}{l} Z = {}^X/_Y, \\ T = XY. \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} X = ZY = \sqrt{ZT}, \\ Y = \sqrt{{}^T/_Z}. \end{array}\right.$$

Como $X, Y > 0$, entonces $Z, T > 0$, por lo que la inversa está bien definida.

$$\begin{array}{rccc} g^{-1}: & \mathbb{R}^+ \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \times \mathbb{R}^+ \\ & (Z,T) & \longmapsto & (X,Y) = (\sqrt{ZT}, \sqrt{T/Z}) \end{array}$$

Tenemos que todas las componentes de g^{-1} son derivables:

$$\begin{aligned} &\frac{\partial X}{\partial Z}(Z,T) = \frac{T}{2\sqrt{ZT}} = \frac{\sqrt{T}}{2\sqrt{Z}}, && \frac{\partial X}{\partial T}(Z,T) = \frac{\sqrt{Z}}{2\sqrt{T}}, \\ &\frac{\partial Y}{\partial Z}(Z,T) = \frac{-\frac{T}{Z^2}}{2\sqrt{T/Z}} = -\frac{\sqrt{T}}{2Z\sqrt{Z}}, && \frac{\partial Y}{\partial T}(Z,T) = \frac{1}{2Z\sqrt{T/Z}} = \frac{1}{2\sqrt{TZ}}. \end{aligned}$$

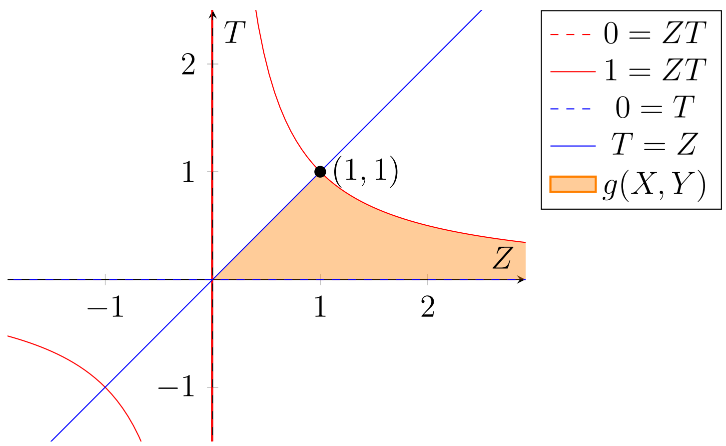
Además, tenemos que:

$$\det Jg^{-1}(z,t) = \begin{vmatrix} \frac{\sqrt{T}}{2\sqrt{Z}} & \frac{\sqrt{Z}}{2\sqrt{T}} \\ -\frac{\sqrt{T}}{2Z\sqrt{Z}} & \frac{1}{2\sqrt{TZ}} \end{vmatrix} = \frac{1}{4Z} + \frac{1}{4Z} = \frac{1}{2Z} > 0 \qquad \forall (z,t) \in \mathbb{R}^+ \times \mathbb{R}^+.$$

Por tanto, $(Z,T) = g(X,Y)$ es un vector aleatorio continuo. Estudiamos ahora el conjunto $g(X,Y)$:

$$\begin{aligned} g(X,Y) &= \left\{(z,t) \in \mathbb{R}^+ \times \mathbb{R}^+ \mid 0 < \sqrt{ZT} < 1,\ 0 < \sqrt{T/Z} < 1\right\} = \\ &= \left\{(z,t) \in \mathbb{R}^+ \times \mathbb{R}^+ \mid 0 < ZT < 1,\ 0 < T < Z\right\} \end{aligned}$$

Veamos el conjunto $g(X,Y)$ gráficamente:



Buscamos ahora la función de densidad de probabilidad de (Z,T):

$$\begin{aligned} f_{(Z,T)}(z,t) &= f_{(X,Y)}\left(\sqrt{ZT}, \sqrt{T/Z}\right) \cdot \left|\det Jg^{-1}(z,t)\right| = \\ &= \begin{cases} \dfrac{1}{2z} & (z,t) \in g(X,Y), \\ 0 & \text{en otro caso.} \end{cases} \end{aligned}$$

Por tanto, la función de densidad de probabilidad de (Z, T) es:

$$f_{(Z,T)}(z,t) = \begin{cases} \dfrac{1}{2z} & 0 < ZT < 1,\ 0 < T < Z, \\ 0 & \text{en otro caso.} \end{cases}$$

Para obtener la función de densidad de probabilidad de $Z = {}^X\!/_Y$, tenemos que:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{(Z,T)}(z,t)\,dt$$

Los límites de integración los vemos claros en la gráfica anterior. Distinguimos en función del valor de z:

- Si $z \in]0, 1]$, entonces:

$$f_Z(z) = \int_0^z \frac{1}{2z}\,dt = \frac{1}{2z}\,[t]_0^z = \frac{1}{2z}z = \frac{1}{2}.$$

- Si $z \in [1, +\infty[$, entonces:

$$f_Z(z) = \int_0^{1/z} \frac{1}{2z}\,dt = \frac{1}{2z}\,[t]_0^{1/z} = \frac{1}{2z^2}.$$

Por tanto, la función de densidad de probabilidad de Z es:

$$f_Z(z) = \begin{cases} \dfrac{1}{2} & 0 < z < 1, \\[2em] \dfrac{1}{2z^2} & 1 < z, \\ 0 & \text{en otro caso.} \end{cases}$$

Para obtener la función de densidad de probabilidad de $T = XY$, tenemos que:

$$f_T(t) = \int_{-\infty}^{+\infty} f_{(Z,T)}(z,w)\,dz$$

Los límites de integración los vemos claros en la gráfica anterior. Para $t \in]0, 1]$, tenemos que:

$$f_T(t) = \int_t^{1/t} \frac{1}{2z}\,dz = \frac{1}{2}\,[\ln(z)]_t^{1/t} = \frac{1}{2}\,[\ln({}^1\!/_t) - \ln(t)] = \frac{1}{2}\,[\ln(1) - 2\ln(t)] = -\ln(t)$$

Por tanto, la función de densidad de probabilidad de T es:

$$f_T(t) = \begin{cases} -\ln(t) & 0 < t < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Una vez tenemos ambas marginales, es fácil obtener la función de distribución de cada una. Respecto de Z, distinguimos en función del valor de z:

- Si $z \in]0,1]$, entonces:

$$F_Z(z) = \int_{-\infty}^{z} f_Z(z)\,dz = \int_0^z \frac{1}{2}\,dz = \frac{1}{2}\,[z]_0^z = \frac{z}{2}$$

- Si $z \in [1,+\infty[$, entonces:

$$\begin{aligned}F_Z(z) &= \int_{-\infty}^{z} f_Z(z)\,dz = \int_0^1 \frac{1}{2}\,dz + \int_1^z \frac{1}{2z^2}\,dz = \frac{1}{2} - \frac{1}{2}\left[\frac{1}{z}\right]_1^z = \\ &= \frac{1}{2} - \frac{1}{2}\left(\frac{1}{z} - 1\right) = 1 - \frac{1}{2z}\end{aligned}$$

Por tanto, la función de distribución de probabilidad de $Z = {}^X\!/_Y$ es:

$$F_Z(z) = \begin{cases} \dfrac{z}{2} & 0 < z < 1, \\ \\ 1 - \dfrac{1}{2z} & 1 < z, \\ 0 & \text{en otro caso.} \end{cases}$$

Respecto de T, para $t \in]0,1]$, tenemos que:

$$\begin{aligned}F_T(t) &= \int_{-\infty}^{t} f_T(t)\,dt = \int_0^t -\ln(t)\,dt = -\left[t\ln(t) - t\right]_0^t = \\ &= -t\ln(t) + t + \lim_{t\to 0} t\ln(t) = -t\ln(t) + t\end{aligned}$$

Por tanto, la función de distribución de probabilidad de $T = XY$ es:

$$F_T(t) = \begin{cases} 0 & t \leqslant 0, \\ -t\ln(t) + t & 0 < t < 1, \\ 1 & 1 \leqslant t, \end{cases}$$

5. Determinar la función de distribución de probabilidad de $\text{máx}(X,Y)$, y del $\text{mín}(X,Y)$.

   Dado $x \in \mathbb{R}$, tenemos que:

$$P[\text{máx}(X,Y) \leqslant x] = P[X \leqslant x, Y \leqslant x] = P[(X,Y) \leqslant (x,x)]$$

   Calculamos por tanto dicho valor sabiendo $f_{(X,Y)}$. Para $z \in [0,1]$, tenemos que:

$$P[(X,Y) \leqslant (z,z)] = \int_0^z \int_0^z k\,dy\,dx = k\int_0^z z\,dx = kz\,[x]_0^z = kz^2 = z^2$$

   Por tanto, la función de distribución de probabilidad de $\text{máx}(X,Y)$ es:

$$F_{\text{máx}(X,Y)}(z) = \begin{cases} 0 & z \leqslant 0, \\ z^2 & 0 < z < 1, \\ 1 & 1 \leqslant z. \end{cases}$$

Respecto del $\text{mín}(X, Y)$, dado $z \in \mathbb{R}$, tenemos que:

$$P[\text{mín}(X, Y) \leqslant z] = 1 - P[\text{mín}(X, Y) > z] = 1 - P[(X, Y) > z]$$

Calculamos dicho valor sabiendo $f_{(X,Y)}$. Distinguimos en función de z:

- Si $z \leqslant 0$, entonces $P[(X, Y) > z] = 1$.
- Si $0 < z < 1$, entonces:

$$P[(X, Y) > z] = \int_z^1 \int_z^1 k\,dy\,dx = k \int_z^1 (1 - z)\,dx = k(1 - z)\,[x]_z^1 = (1 - z)^2$$

- Si $z \geqslant 1$, entonces $P[(X, Y) > z] = 0$.

Por tanto, la función de distribución de probabilidad de $\text{mín}(X, Y)$ es:

$$F_{\text{mín}(X,Y)}(z) = 1 - P[(X, Y) > z] = \begin{cases} 0 & z \leqslant 0, \\ 1 - (1 - z)^2 & 0 < z < 1, \\ 1 & 1 \leqslant z. \end{cases}$$

6. Determinar la función de distribución de probabilidad conjunta del $\text{máx}(X, Y)$, y del $\text{mín}(X, Y)$.

   Dado $z, t \in \mathbb{R}$, distinguimos casos:

   - Si $z \leqslant t$, como $\text{mín}(X, Y) \leqslant \text{máx}(X, Y)$, entonces:

     $$P[\text{máx}(X, Y) \leqslant z, \text{mín}(X, Y) \leqslant t] = P[\text{máx}(X, Y) \leqslant z] = F_{\text{máx}(X,Y)}(z)$$

     Distinguimos en función de z:

     - Si $z \leqslant 0$, entonces $F_{\text{máx}(X,Y)}(z) = 0$.
     - Si $0 < z < 1$, entonces $F_{\text{máx}(X,Y)}(z) = z^2$.
     - Si $z \geqslant 1$, entonces $F_{\text{máx}(X,Y)}(z) = 1$.

   - Si $z > t$, entonces:

     $$\begin{aligned} &P[\text{máx}(X, Y) \leqslant z, \text{mín}(X, Y) \leqslant t] \\ &\qquad = P[\text{máx}(X, Y) \leqslant z] - P[\text{máx}(X, Y) \leqslant z, \text{mín}(X, Y) > t] = \\ &\qquad = P[\text{máx}(X, Y) \leqslant z] - P[t < X \leqslant z, t < Y \leqslant z] \end{aligned}$$

     Distinguimos en función de z:

     - Si $z \leqslant 0$, entonces $P[\text{máx}(X, Y) \leqslant z] = 0$. Además, $t < z \leqslant 0$. Por tanto:

       $$P[t < X \leqslant z, t < Y \leqslant z] = 0$$

       Por tanto, $P[\text{máx}(X, Y) \leqslant z, \text{mín}(X, Y) \leqslant t] = 0$.
     - Si $0 < z < 1$, entonces $P[\text{máx}(X, Y) \leqslant z] = z^2$. Además, sabemos que $t < z < 1$. Distinguimos en función de t:

- Si $t \leqslant 0$, entonces:

$$P[t < X \leqslant z, t < Y \leqslant z] = \int_0^z \int_0^z k\,dy\,dx = kz^2 = z^2$$

Por tanto, $P[\text{máx}(X,Y) \leqslant z, \text{mín}(X,Y) \leqslant t] = z^2 - z^2 = 0$.

- Si $0 < t < z$, entonces:

$$P[t < X \leqslant z, t < Y \leqslant z] = \int_t^z \int_t^z k\,dy\,dx = k(z-t)^2 = (z-t)^2$$

Por tanto, tenemos que:

$$P[\text{máx}(X,Y) \leqslant z, \text{mín}(X,Y) \leqslant t] = z^2-(z-t)^2 = z^2-z^2+2zt-t^2 = 2zt-t^2$$

- Si $z \geqslant 1$, entonces $P[\text{máx}(X,Y) \leqslant z] = 1$. Distinguimos en función de t:
  - Si $t \leqslant 0$, entonces:

$$P[t < X \leqslant z, t < Y \leqslant z] = \int_0^1 \int_0^1 k\,dy\,dx = k = 1$$

Por tanto, $P[\text{máx}(X,Y) \leqslant z, \text{mín}(X,Y) \leqslant t] = 1 - 1 = 0$.

  - Si $0 < t < 1$, entonces:

$$P[t < X \leqslant z, t < Y \leqslant z] = \int_t^1 \int_t^1 k\,dy\,dx = k(1-t)^2 = (1-t)^2$$

Por tanto,

$$P[\text{máx}(X,Y) \leqslant z, \text{mín}(X,Y) \leqslant t] = 1-(1-t)^2 = 1-1+2t-t^2 = 2t-t^2$$

  - Si $t \geqslant 1$, $t < z$, entonces:

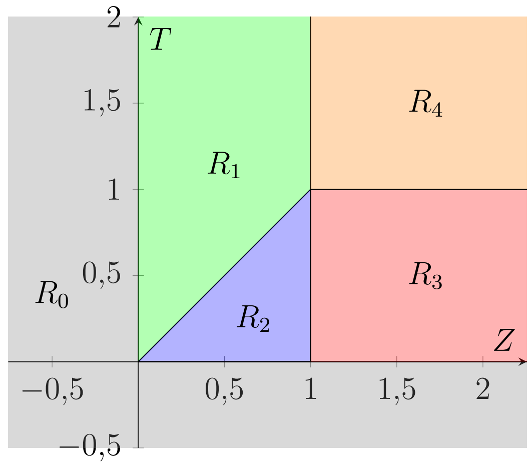
$$P[t < X \leqslant z, t < Y \leqslant z] = 0$$

Por tanto, $P[\text{máx}(X,Y) \leqslant z, \text{mín}(X,Y) \leqslant t] = 1 - 0 = 1$.

Por tanto, la función de distribución de probabilidad conjunta de $\text{máx}(X,Y)$ y $\text{mín}(X,Y)$ es:

$$F_{\text{máx}(X,Y),\text{mín}(X,Y)}(z,t) = \begin{cases} 0 & z \leqslant 0 \lor t \leqslant 0 \quad (R_0), \\ z^2 & z \leqslant t \land 0 < z < 1 \quad (R_1), \\ 2zt - t^2 & 0 < t < z < 1 \quad (R_2), \\ 2t - t^2 & 0 < t < 1 \leqslant z \quad (R_3), \\ 1 & 1 \leqslant t \land 1 \leqslant z \quad (R_4). \end{cases}$$

Veamos gráficamente cada una de las regiones:



Ejercicio 7.3.4. Sea (X, Y) un vector aleatorio bidimensional discreto, cuya función masa de probabilidad conjunta se calcula como el producto de las funciones masa de probabilidad marginales de X e Y. Las variables aleatorias X e Y se distribuyen según una Poisson con parámetro $\lambda > 0$. Calcular la función de distribución de probabilidad marginal del máximo y del mínimo, así como la distribución conjunta del máximo y del mínimo.

Tenemos que:

$$P[X = x, Y = y] = P[X = x] \cdot P[Y = y]$$

Como $X, Y \sim \mathcal{P}(\lambda)$, tenemos que:

$$\begin{aligned} P[X = x, Y = y] &= P[X = x] \cdot P[Y = y] = \\ &= \frac{e^{-\lambda}\lambda^x}{x!} \cdot \frac{e^{-\lambda}\lambda^y}{y!} = \\ &= \frac{e^{-2\lambda}\lambda^{x+y}}{x!y!} \end{aligned}$$

Calculemos la marginal del máximo. Para $n \in \mathbb{N}$, tenemos que:

$$\begin{aligned} P[\text{máx}(X, Y) \leqslant n] &= P[X \leqslant n, Y \leqslant n] = P[(X, Y) \leqslant (n, n)] = \\ &= \sum_{i=0}^{n} \sum_{j=0}^{n} P[X = i, Y = j] = \\ &= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{e^{-2\lambda}\lambda^{i+j}}{i!j!} = \\ &= e^{-2\lambda} \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{\lambda^{i+j}}{i!j!} = e^{-2\lambda} \sum_{i=0}^{n} \frac{\lambda^i}{i!} \sum_{j=0}^{n} \frac{\lambda^j}{j!} = \\ &= P[X \leqslant n] \cdot P[Y \leqslant n] \end{aligned}$$

Calculamos la marginal del mínimo. Para $n \in \mathbb{N}$, tenemos que:

$$\begin{aligned}
P[\text{mín}(X,Y) \leqslant n] &= 1 - P[\text{mín}(X,Y) > n] = \\
&= 1 - P[X > n, Y > n] = \\
&= 1 - P[(X,Y) > (n,n)] = \\
&= 1 - \sum_{i=n+1}^{+\infty} \sum_{j=n+1}^{+\infty} P[X=i, Y=j] = \\
&= 1 - \sum_{i=n+1}^{+\infty} \sum_{j=n+1}^{+\infty} \frac{e^{-2\lambda}\lambda^{i+j}}{i!j!} = \\
&= 1 - e^{-2\lambda} \sum_{i=n+1}^{+\infty} \sum_{j=n+1}^{+\infty} \frac{\lambda^{i+j}}{i!j!} = \\
&= 1 - e^{-2\lambda} \sum_{i=n+1}^{+\infty} \frac{\lambda^i}{i!} \sum_{j=n+1}^{+\infty} \frac{\lambda^j}{j!} = \\
&= 1 - P[X > n] \cdot P[Y > n]
\end{aligned}$$

Calculamos ahora la distribución conjunta del máximo y del mínimo. Para los valores $n, m \in \mathbb{N}$, tenemos que:

- Si $n \leqslant m$, entonces:

$$P[\text{máx}(X,Y) \leqslant n, \text{mín}(X,Y) \leqslant m] = P[\text{máx}(X,Y) \leqslant n]$$

- Si $m < n$, entonces:

$$\begin{aligned}
&P[\text{máx}(X,Y) \leqslant n, \text{mín}(X,Y) \leqslant m] = \\
&\qquad = P[\text{máx}(X,Y) \leqslant n] - P[\text{máx}(X,Y) \leqslant n, \text{mín}(X,Y) > m] = \\
&\qquad = P[\text{máx}(X,Y) \leqslant n] - P[m < X \leqslant n, m < Y \leqslant n] = \\
&\qquad = P[X \leqslant n] \cdot P[Y \leqslant n] - P[m+1 \leqslant X \leqslant n, m+1 < Y \leqslant n]
\end{aligned}$$

Calculamos la probabilidad $P[m+1 \leqslant X \leqslant n, m+1 < Y \leqslant n]$:

$$\begin{aligned}
P[m+1 \leqslant X \leqslant n, m+1 < Y \leqslant n] &= \sum_{i=m+1}^{n} \sum_{j=m+1}^{n} P[X=i, Y=j] = \\
&= \sum_{i=m+1}^{n} \sum_{j=m+1}^{n} \frac{e^{-2\lambda}\lambda^{i+j}}{i!j!} = \\
&= e^{-2\lambda} \sum_{i=m+1}^{n} \sum_{j=m+1}^{n} \frac{\lambda^{i+j}}{i!j!} = \\
&= \sum_{i=m+1}^{n} e^{-\lambda}\frac{\lambda^i}{i!} \sum_{j=m+1}^{n} e^{-\lambda}\frac{\lambda^j}{j!} = \\
&= \left(P[X \leqslant n] - P[X \leqslant m]\right)\left(P[Y \leqslant n] - P[Y \leqslant m]\right)
\end{aligned}$$

Por tanto, la distribución conjunta del máximo y del mínimo es:

$$P[\text{máx}(X,Y) \leqslant n, \text{mín}(X,Y) \leqslant m] =$$
$$= \begin{cases} P[X \leqslant n] \cdot P[Y \leqslant n] & n \leqslant m, \\ P[X \leqslant n] \cdot P[Y \leqslant n] - (P[X \leqslant n] - P[X \leqslant m])\,(P[Y \leqslant n] - P[Y \leqslant m]) & m < n. \end{cases}$$

Ejercicio 7.3.5. Sea (X,Y) un vector aleatorio con función de densidad de probabilidad
$$f(x,y) = \begin{cases} 2 & 0 < x < 1,\ 0 < y < x, \\ 0 & \text{en otro caso.} \end{cases}$$
Calcular la densidad de probabilidad de las variables $Z = aX + bY$, $T = X/Y$, a partir de la densidad de probabilidad conjunta de $(Z,T) = (aX + bY, X/Y)$, $a, b > 0$.

Definimos la transformación:

$$\begin{array}{rccl} g: & E_{(X,Y)} & \longrightarrow & \mathbb{R}^2 \\ & (X,Y) & \longmapsto & (Z,T) = (aX + bY, X/Y) \end{array}$$

Para obtener g^{-1}, buscamos obtener X,Y en función de Z,T:

$$\left\{\begin{array}{l} Z = aX + bY, \\ T = X/Y. \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} X = \dfrac{TZ}{aT+b}, \\ Y = \dfrac{Z}{aT+b}. \end{array}\right.$$

Notemos que $a, b > 0$, y como $X, Y > 0$, entonces $T > 0$. Por tanto, $aT + b > 0$, por lo que está bien definida la transformación.

Por tanto, tenemos que:

$$\begin{array}{rccl} g^{-1}: & g(E_{(X,Y)}) & \longrightarrow & E_{(X,Y)} \\ & (Z,T) & \longmapsto & (X,Y) = \left(\dfrac{TZ}{aT+b}, \dfrac{Z}{aT+b}\right) \end{array}$$

Tenemos que todas las componentes de g^{-1} son derivables:

$$\frac{\partial X}{\partial Z}(Z,T) = \frac{T}{aT+b}, \qquad \frac{\partial X}{\partial T}(Z,T) = \frac{bZ}{(aT+b)^2},$$
$$\frac{\partial Y}{\partial Z}(Z,T) = \frac{1}{aT+b}, \qquad \frac{\partial Y}{\partial T}(Z,T) = \frac{-aZ}{(aT+b)^2}.$$

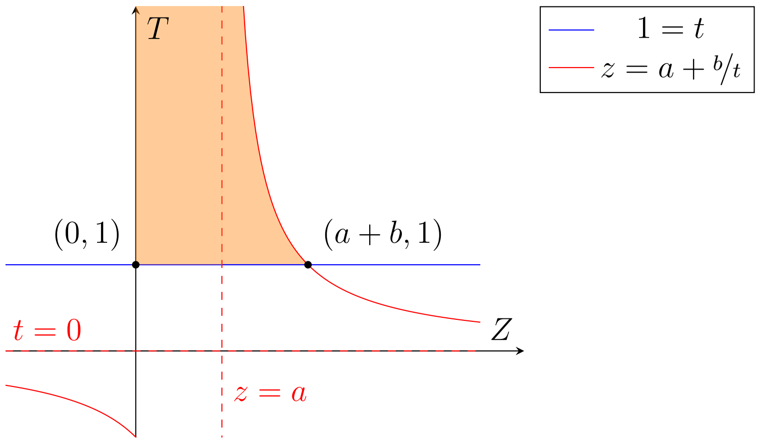
Además, tenemos que:

$$\det Jg^{-1}(z,t) = \begin{vmatrix} \dfrac{T}{aT+b} & \dfrac{bZ}{(aT+b)^2} \\ \dfrac{1}{aT+b} & \dfrac{-aZ}{(aT+b)^2} \end{vmatrix} = \frac{1}{(aT+b)^4} \begin{vmatrix} T(aT+b) & bZ \\ aT+b & -aZ \end{vmatrix} =$$
$$= \frac{1}{(aT+b)^4}(aT+b)(-TaZ - bZ) = -\frac{Z}{(aT+b)^2} \neq 0 \qquad \forall (z,t) \in g(E_{(X,Y)}).$$

Por tanto, $(Z,T) = g(X,Y)$ es un vector aleatorio continuo. Veamos ahora el valor de $g(X,Y)$ para $X \in [0,1]$, $Y \in [0,X]$:

$$\begin{aligned} g(X,Y) &= \left\{(z,t) \in \mathbb{R}^2 \mid 0 < \frac{tz}{at+b} < 1,\ 0 < \frac{z}{at+b} < \frac{tz}{at+b}\right\} = \\ &= \left\{(z,t) \in \mathbb{R}^2 \mid 0 < tz < at+b,\ 0 < 1 < t\right\} = \\ &= \left\{(z,t) \in \mathbb{R}^2 \mid 0 < z < a + \frac{b}{t},\ 1 < t\right\} \end{aligned}$$

Veamos este conjunto gráficamente:



La densidad de probabilidad de (Z,T) es:

$$f_{(Z,T)}(z,t) = f_{(X,Y)}\left(\frac{tz}{at+b}, \frac{z}{at+b}\right) \cdot \left|-\frac{Z}{(aT+b)^2}\right| =$$
$$= \begin{cases} \dfrac{2z}{(at+b)^2} & (z,t) \in g(X,Y), \\ 0 & \text{en otro caso.} \end{cases}$$

Por tanto, la densidad de probabilidad de (Z,T) es:

$$f_{(Z,T)}(z,t) = \begin{cases} \dfrac{2z}{(at+b)^2} & (z,t) \in g(X,Y), \\ 0 & \text{en otro caso.} \end{cases}$$

Calculemos ahora la densidad de probabilidad de $Z = aX + bY$. Distinguimos en función de z:

- Para $z \in [0,a]$, tenemos que:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_{(Z,T)}(z,t)\,dt = \int_1^{+\infty} \frac{2z}{(at+b)^2}\,dt = \\ &= -\frac{2z}{a}\left[\frac{1}{at+b}\right]_1^{+\infty} = -\frac{2z}{a}\left[-\frac{1}{a+b}\right] = \frac{2z}{a(a+b)} \end{aligned}$$

- Para $z \in]a, a+b[$, tenemos que:

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{+\infty} f_{(Z,T)}(z,t)\,dt = \int_1^{\frac{b}{z-a}} \frac{2z}{(at+b)^2}\,dt = \\
&= -\frac{2z}{a}\left[\frac{1}{at+b}\right]_1^{\frac{b}{z-a}} = -\frac{2z}{a}\left[\frac{1}{\frac{ba}{z-a}+b} - \frac{1}{a+b}\right] = -\frac{2z}{a}\left[\frac{z-a}{bz} - \frac{1}{a+b}\right] = \\
&= -\frac{2\cancel{z}}{a}\left[\frac{(z-a)(a+b)-bz}{b\cancel{z}(a+b)}\right] = -\frac{2}{\cancel{a}}\left[\frac{\cancel{a}z + \cancel{bz} - a^{\cancel{2}} - \cancel{a}b - \cancel{bz}}{b(a+b)}\right] = \\
&= 2\cdot\left[\frac{a+b-z}{b(a+b)}\right]
\end{aligned}
$$

Por tanto, la densidad de probabilidad de $Z = aX + bY$ es:

$$
f_Z(z) = \begin{cases} \dfrac{2z}{a(a+b)} & z \in [0,a]\,, \\ 2\cdot\left[\dfrac{a+b-z}{b(a+b)}\right] & z \in]a, a+b[\,, \\ 0 & \text{en otro caso.} \end{cases}
$$

Calculemos ahora la densidad de probabilidad de $T = {}^X\!/_Y$. Para $t > 1$, tenemos que:

$$
\begin{aligned}
f_T(t) &= \int_{-\infty}^{+\infty} f_{(Z,T)}(z,t)\,dz = \int_0^{a+\frac{b}{t}} \frac{2z}{(at+b)^2}\,dz = \\
&= \frac{1}{(at+b)^2}\left[z^2\right]_0^{a+\frac{b}{t}} = \frac{1}{(at+b)^2}\left(a+\frac{b}{t}\right)^2 = \left(\frac{a+\dfrac{b}{t}}{at+b}\right)^2 = \\
&= \left(\frac{at+b}{(at+b)t}\right)^2 = \\
&= \frac{1}{t^2}
\end{aligned}
$$

Por tanto, la densidad de probabilidad de $T = {}^X\!/_Y$ es:

$$
f_T(t) = \begin{cases} \dfrac{1}{t^2} & t > 1, \\ 0 & \text{en otro caso.} \end{cases}
$$

Ejercicio 7.3.6. Sea (X,Y) un vector aleatorio, cuya función de densidad de probabilidad conjunta se calcula como producto de las funciones de densidad de probabilidad marginales de X e Y, siendo $X \sim \exp(\lambda)$ e $Y \sim \exp(\mu)$. Calcular la función de distribución de probabilidad conjunta del vector aleatorio

$$
(Z,T) = (\text{mín}(X,Y), T), \qquad T = \begin{cases} 0 & Y < X \\ 1 & X < Y \end{cases}
$$

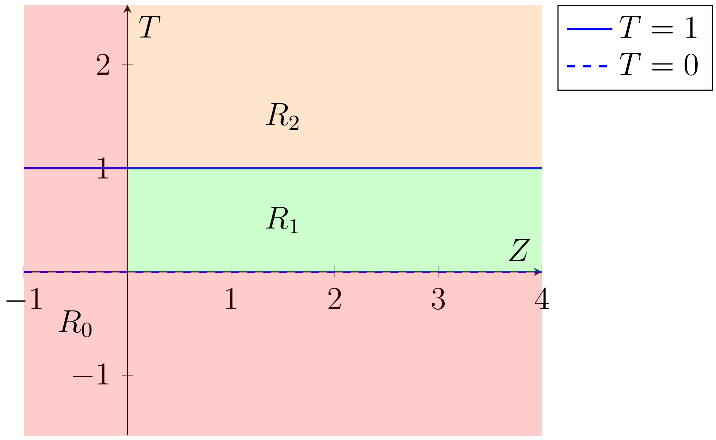
Como $X \sim \exp(\lambda)$ e $Y \sim \exp(\mu)$, tenemos que:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geqslant 0, \\ 0 & x < 0, \end{cases} \qquad f_Y(y) = \begin{cases} \mu e^{-\mu y} & y \geqslant 0, \\ 0 & y < 0. \end{cases}$$

Por tanto, la función de densidad de probabilidad conjunta de (X, Y) es:

$$f_{(X,Y)}(x, y) = f_X(x) \cdot f_Y(y) = \begin{cases} \lambda \mu e^{-\lambda x - \mu y} & x, y \geqslant 0, \\ 0 & \text{en otro caso.} \end{cases}$$

Tenemos que $E_X = E_Y = \mathbb{R}^+$, por lo que $E_Z = \mathbb{R}^+$. Además, $E_T = \{0, 1\}$. Tenemos por tanto la siguiente situación:



Calculamos la función de distribución de probabilidad conjunta de (Z, T):

$$P[Z \leqslant z, T \leqslant t] = P[\min(X, Y) \leqslant z, T \leqslant t]$$

Dados $z, t \in \mathbb{R}$, distinguimos casos:

- Si $z \leqslant 0$ o $t < 0$, (es decir, $(z, t) \in R_0$), entonces:

$$P[\min(X, Y) \leqslant z, T \leqslant t] = 0$$

- Si $z > 0$ y $t \in [0, 1[$, (es decir, $(z, t) \in R_1$), entonces:

$$\begin{aligned} P[\min(X, Y) \leqslant z, T \leqslant t] &= P[\min(X, Y) \leqslant z, T = 0] = P[\min(X, Y) \leqslant z, Y < X] = \\ &= P[Y \leqslant z, Y < X] = \int_0^z \int_y^{+\infty} f_{(X,Y)}(x, y)\, dx\, dy = \\ &= \int_0^z \int_y^{+\infty} \lambda \mu e^{-\lambda x - \mu y}\, dx\, dy = \int_0^z \mu e^{-\mu y} \int_y^{+\infty} \lambda e^{-\lambda x}\, dx\, dy = \\ &= \int_0^z \mu e^{-\mu y} \left[-e^{-\lambda x}\right]_y^{+\infty} dy = \int_0^z \mu e^{-\mu y} e^{-\lambda y}\, dy = \\ &= \mu \int_0^z e^{-(\lambda+\mu) y}\, dy = \mu \left[\frac{e^{-(\lambda+\mu)y}}{-(\lambda + \mu)}\right]_0^z = \\ &= -\frac{\mu}{\lambda + \mu} \left[\exp(-(\lambda + \mu) z) - 1\right] \end{aligned}$$

- Si $z > 0$ y $t \geqslant 1$, (es decir, $(z,t) \in R_2$), entonces:

  Tenemos dos opciones para calcular $P[\min(X,Y) \leqslant z, T \leqslant t]$:

  Opción 1) Como $E_T = \{0,1\}$, sabemos que:

  $$P[\min(X,Y) \leqslant z, T \leqslant t] = P[\min(X,Y) \leqslant z]$$

  Por tanto, calculamos $P[\min(X,Y) \leqslant z]$:

  $$\begin{aligned}
  P[\min(X,Y) \leqslant z] &= 1 - P[\min(X,Y) > z] = 1 - P[X > z, Y > z] = \\
  &= 1 - \int_z^{+\infty} \int_z^{+\infty} f_{(X,Y)}(x,y)\,dy\,dx = \\
  &= 1 - \int_z^{+\infty} \int_z^{+\infty} \lambda\mu e^{-\lambda x - \mu y}\,dy\,dx = \\
  &= 1 - \int_z^{+\infty} \lambda e^{-\lambda x}\,dx \int_z^{+\infty} \mu e^{-\mu y}\,dy = \\
  &= 1 - \left[-e^{-\lambda x}\right]_z^{+\infty} \left[-e^{-\mu y}\right]_z^{+\infty} = 1 - \left[e^{-\lambda z} - 0\right]\left[e^{-\mu z} - 0\right] = \\
  &= 1 - (e^{-\lambda z})(e^{-\mu z}) = 1 - \exp(-(\lambda+\mu)z)
  \end{aligned}$$

  Opción 2) Como $E_T = \{0,1\}$, sabemos que:

  $$P[\min(X,Y) \leqslant z, T \leqslant t] = P[\min(X,Y) \leqslant z, T = 0] + P[\min(X,Y) \leqslant z, T = 1]$$

  La primera ya la hemos obtenido anteriormente. por tanto, calculamos la segunda probabilidad:

  $$\begin{aligned}
  P[\min(X,Y) \leqslant z, T = 1] &= P[\min(X,Y) \leqslant z, X < Y] = P[X \leqslant z, X < Y] = \\
  &= \int_0^z \int_x^{+\infty} f_{(X,Y)}(x,y)\,dy\,dx = \int_0^z \int_x^{+\infty} \lambda\mu e^{-\lambda x - \mu y}\,dy\,dx = \\
  &= \int_0^z \lambda e^{-\lambda x} \int_x^{+\infty} \mu e^{-\mu y}\,dy\,dx = \int_0^z \lambda e^{-\lambda x} \left[-e^{-\mu y}\right]_x^{+\infty} dx = \\
  &= \int_0^z \lambda e^{-\lambda x} e^{-\mu x}\,dx = \lambda \int_0^z e^{-(\lambda+\mu)x}\,dx = \lambda \left[\frac{e^{-(\lambda+\mu)x}}{-(\lambda+\mu)}\right]_0^z = \\
  &= -\frac{\lambda}{\lambda+\mu}\left[\exp(-(\lambda+\mu)z) - 1\right]
  \end{aligned}$$

  Sumando ambos resultados, tenemos que:

  $$\begin{aligned}
  P[\min(X,Y) \leqslant z, T \leqslant t] &= \left[\exp(-(\lambda+\mu)z) - 1\right]\left(-\frac{\mu}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu}\right) = \\
  &= 1 - \exp(-(\lambda+\mu)z)
  \end{aligned}$$

En cualquier caso, la función de distribución de probabilidad conjunta (Z,T) es:

$$F_{(Z,T)}(z,t) = \begin{cases} 0 & z \leqslant 0 \text{ o } t < 0, \\ -\dfrac{\mu}{\lambda+\mu}\left[\exp(-(\lambda+\mu)z) - 1\right] & z > 0 \text{ y } t \in [0,1[\,, \\ 1 - \exp(-(\lambda+\mu)z) & z > 0 \text{ y } t \geqslant 1. \end{cases}$$

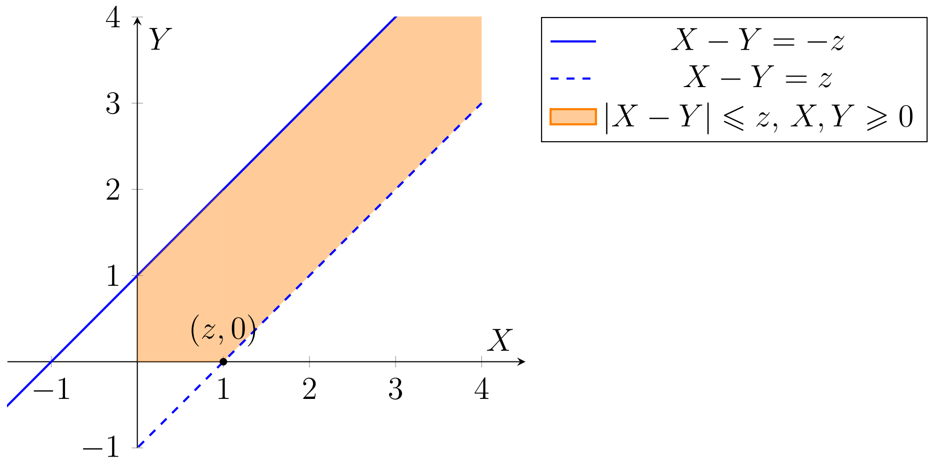


Figura 7.1: Región de integración para $P[|X-Y| \leqslant z]$

Ejercicio 7.3.7. Sea (X, Y) un vector aleatorio, cuya función de densidad de probabilidad conjunta se calcula como en el problema anterior, considerando $\lambda = \mu$. Calcular la distribución de probabilidad de:

1. $|X - Y|$,

   Del apartado anterior, tenemos que:

$$f_{(X,Y)}(x,y) = \begin{cases} \lambda^2 e^{-\lambda(x+y)} & x, y \geqslant 0, \\ 0 & \text{en otro caso.} \end{cases}$$

   Buscamos calcular $P[|X-Y| \leqslant z]$ para todo $z \in \mathbb{R}$. Si $z \leqslant 0$, tenemos que $P[|X-Y| \leqslant z] = 0$, por lo que sea $z \in \mathbb{R}^+$. Tenemos que:

$$P[|X-Y| \leqslant z] = P[-z \leqslant X - Y \leqslant z]$$

   Sabiendo que $X, Y \geqslant 0$, tenemos que la situación es la descrita en la Figura 7.1.

Por tanto, tenemos que:

$$\begin{aligned}
&P[|X-Y| \leqslant z] = P[-z \leqslant X-Y \leqslant z] = \\
&= \int_0^z \int_0^{x+z} f_{(X,Y)}(x,y)\,dy\,dx + \int_z^{+\infty} \int_{x-z}^{x+z} f_{(X,Y)}(x,y)\,dy\,dx = \\
&= \int_0^z \int_0^{x+z} \lambda^2 e^{-\lambda(x+y)}\,dy\,dx + \int_z^{+\infty} \int_{x-z}^{x+z} \lambda^2 e^{-\lambda(x+y)}\,dy\,dx = \\
&= \int_0^z \lambda e^{-\lambda x} \int_0^{x+z} \lambda e^{-\lambda y}\,dy\,dx + \int_z^{+\infty} \lambda e^{-\lambda x} \int_{x-z}^{x+z} \lambda e^{-\lambda y}\,dy\,dx = \\
&= \int_0^z \lambda e^{-\lambda x} \left[-e^{-\lambda y}\right]_0^{x+z} dx + \int_z^{+\infty} \lambda e^{-\lambda x} \left[-e^{-\lambda y}\right]_{x-z}^{x+z} dx = \\
&= \int_0^z \lambda e^{-\lambda x} \left[1 - e^{-\lambda(x+z)}\right] dx + \int_z^{+\infty} \lambda e^{-\lambda x} \left[e^{-\lambda(x-z)} - e^{-\lambda(x+z)}\right] dx = \\
&= \int_0^z \lambda(e^{-\lambda x} - e^{-\lambda(2x+z)})\,dx + \int_z^{+\infty} \lambda(e^{-\lambda(2x-z)} - e^{-\lambda(2x+z)})\,dx = \\
&= \left[-e^{-\lambda x} + \frac{1}{2}e^{-\lambda(2x+z)}\right]_0^z + \frac{1}{2}\left[-e^{-\lambda(2x-z)} + e^{-\lambda(2x+z)}\right]_z^{+\infty} = \\
&= -e^{-\lambda z} + \cancel{\frac{1}{2}e^{-\lambda(3z)}} + 1 - \cancel{\frac{1}{2}e^{-\lambda z}} + \frac{1}{2}\left[0 + \cancel{e^{-\lambda(z)}} - \cancel{e^{-\lambda(3z)}}\right] = 1 - e^{-\lambda z}
\end{aligned}$$

Por tanto, la distribución de probabilidad de $|X-Y|$ es:

$$P[|X-Y| \leqslant z] = \begin{cases} 0 & z \leqslant 0, \\ 1 - e^{-\lambda z} & z > 0. \end{cases}$$

2. $\max(X, Y^3)$,

Buscamos calcular $P[\max(X,Y^3) \leqslant z]$ para todo $z \in \mathbb{R}$.

$$P[\max(X,Y^3) \leqslant z] = P[X \leqslant z, Y^3 \leqslant z] = P[X \leqslant z, Y \leqslant \sqrt[3]{z}]$$

Para $z \leqslant 0$, como $X, Y \geqslant 0$, tenemos que $P[\max(X,Y^3) \leqslant z] = 0$. Por tanto, sea $z > 0$.

$$\begin{aligned}
P[\max(X,Y^3) \leqslant z] &= P[X \leqslant z, Y \leqslant \sqrt[3]{z}] = \int_0^z \int_0^{\sqrt[3]{z}} f_{(X,Y)}(x,y)\,dy\,dx = \\
&= \int_0^z \int_0^{\sqrt[3]{z}} \lambda^2 e^{-\lambda(x+y)}\,dy\,dx = \int_0^z \lambda e^{-\lambda x}\,dx \int_0^{\sqrt[3]{z}} \lambda e^{-\lambda y}\,dy = \\
&= \left[-e^{-\lambda x}\right]_0^z \left[-e^{-\lambda y}\right]_0^{\sqrt[3]{z}} = (1 - e^{-\lambda z})(1 - e^{-\lambda \sqrt[3]{z}})
\end{aligned}$$

Por tanto, la distribución de probabilidad de $\max(X,Y^3)$ es:

$$P[\max(X,Y^3) \leqslant z] = \begin{cases} 0 & z \leqslant 0, \\ (1 - e^{-\lambda z})(1 - e^{-\lambda \sqrt[3]{z}}) & z > 0. \end{cases}$$

3. $\min(X^5, Y)$.

   Buscamos calcular $P[\min(X^5, Y) \leqslant z]$ para todo $z \in \mathbb{R}$.

$$P[\min(X^5, Y) \leqslant z] = 1 - P[\min(X^5, Y) > z] = 1 - P[X^5 > z, Y > z] =$$
$$= 1 - P[X > \sqrt[5]{z}, Y > z]$$

   Para $z \leqslant 0$, como $X, Y \geqslant 0$, tenemos que $P[\min(X^5, Y) \leqslant z] = 0$. Por tanto, sea $z > 0$.

$$P[\min(X^5, Y) \leqslant z] = 1 - P[X > \sqrt[5]{z}, Y > z] = 1 - \int_{\sqrt[5]{z}}^{+\infty} \int_{z}^{+\infty} f_{(X,Y)}(x, y)\, dy\, dx =$$
$$= 1 - \int_{\sqrt[5]{z}}^{+\infty} \int_{z}^{+\infty} \lambda^2 e^{-\lambda(x+y)}\, dy\, dx = 1 - \int_{\sqrt[5]{z}}^{+\infty} \lambda e^{-\lambda x}\, dx \int_{z}^{+\infty} \lambda e^{-\lambda y}\, dy =$$
$$= 1 - \left[-e^{-\lambda x}\right]_{\sqrt[5]{z}}^{+\infty} \left[-e^{-\lambda y}\right]_{z}^{+\infty} = 1 - \left[0 - e^{-\lambda \sqrt[5]{z}}\right] \left[0 - e^{-\lambda z}\right] =$$
$$= 1 - (e^{-\lambda \sqrt[5]{z}})(e^{-\lambda z}) = 1 - \exp(-\lambda(\sqrt[5]{z} + z))$$

   Por tanto, la distribución de probabilidad de $\min(X^5, Y)$ es:

$$P[\min(X^5, Y) \leqslant z] = \begin{cases} 0 & z \leqslant 0, \\ 1 - \exp(-\lambda(\sqrt[5]{z} + z)) & z > 0. \end{cases}$$

Ejercicio 7.3.8. Sea (X, Y) un vector aleatorio discreto con función masa de probabilidad

$$P[X = x, Y = y] = \frac{k}{2^{x+y}}, \qquad x, y \in \mathbb{N}$$

Observación. Consideramos $\mathbb{N} = \mathbb{N} \cup \{0\}$.

1. Calcular el valor de k para que la ecuación anterior defina la función masa de probabilidad de un variable aleatoria bidimensional discreta.

   Tenemos que la suma de una serie geométrica de razón $r \in]-1, 1[$ es:

$$\sum_{n=0}^{+\infty} r^n = \frac{1}{1 - r}$$

   Además, la suma de sus n primeros términos es:

$$\sum_{k=0}^{n} r^k = \frac{1 - r^{n+1}}{1 - r}$$

   Para que la función masa de probabilidad sea válida, tenemos que:

$$1 = \sum_{x,y=0}^{+\infty} P[X = x, Y = y] = \sum_{x,y=0}^{+\infty} \frac{k}{2^{x+y}} = k \sum_{x=0}^{+\infty} \frac{1}{2^x} \sum_{y=0}^{+\infty} \frac{1}{2^y} =$$
$$= k \cdot \frac{1}{1 - 1/2} \cdot \frac{1}{1 - 1/2} = k \cdot 2 \cdot 2 = 4k \Longrightarrow k = \frac{1}{4}$$

2. Calcular las funciones masa de probabilidad marginales y condicionadas.

   La función masa de probabilidad marginal de X es:

$$P[X = x] = \sum_{y=0}^{+\infty} P[X = x, Y = y] = \sum_{y=0}^{+\infty} \frac{1}{4 \cdot 2^{x+y}} = \frac{1}{4} \sum_{y=0}^{+\infty} \frac{1}{2^{x+y}} =$$

$$= \frac{1}{4} \sum_{y=0}^{+\infty} \frac{1}{2^x} \frac{1}{2^y} = \frac{1}{4} \frac{1}{2^x} \sum_{y=0}^{+\infty} \frac{1}{2^y} = \frac{1}{4} \frac{1}{2^x} \frac{1}{1 - 1/2} = \frac{1}{4} \frac{1}{2^x} \frac{1}{1/2} = \frac{1}{2^{x+1}}$$

   De forma análoga, la función masa de probabilidad marginal de Y es:

$$P[Y = y] = \frac{1}{2^{y+1}}$$

   La función masa de probabilidad condicionada de X dado $Y = y^* \in \mathbb{N}$ es:

$$P[X = x \mid Y = y^*] = \frac{P[X = x, Y = y^*]}{P[Y = y^*]} = \frac{\frac{1}{4 \cdot 2^{x+y^*}}}{\frac{1}{2^{y^*+1}}} = \frac{1}{4 \cdot 2^{x+y^*}} \cdot \frac{2^{y^*+1}}{1} =$$

$$= \frac{2^{y^*-1}}{2^{x+y^*}} = 2^{-(x+1)} \qquad \forall x \in \mathbb{N}$$

   La función masa de probabilidad condicionada de Y dado $X = x^* \in \mathbb{N}$ es análoga, y es:

$$P[Y = y \mid X = x^*] = 2^{-(y+1)} \qquad \forall y \in \mathbb{N}$$

3. Calcular la función masa de probabilidad de $X + Y$.

   Notamos $Z = X + Y$. Definimos las transformación:

$$\begin{array}{rccc} g: & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & (X, Y) & \longmapsto & Z = X + Y \end{array}$$

   Como $Z = X + Y$, y $X, Y \in \mathbb{N}$, tenemos que $Z \in \mathbb{N}$. Busquemos la función masa de probabilidad de Z:

$$P[Z = z] = \sum_{\substack{x,y \in \mathbb{N} \\ x+y=z}} P[X = x, Y = y] = \sum_{\substack{x,y \in \mathbb{N} \\ x+y=z}} \frac{1}{4 \cdot 2^{x+y}} = \sum_{x=0}^{z} \frac{1}{4 \cdot 2^{x+z-x}} = \sum_{x=0}^{z} \frac{1}{4 \cdot 2^z} =$$

$$= \frac{z+1}{4 \cdot 2^z}$$

4. Calcular la función masa de probabilidad de $X - Y$.

   Notamos $T = X - Y$. Definimos las transformación:

$$\begin{array}{rccc} h: & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & (X, Y) & \longmapsto & T = X - Y \end{array}$$

Como $T = X - Y$, y $X, Y \in \mathbb{N}$, tenemos que $T \in \mathbb{Z}$. Busquemos la función masa de probabilidad de T:

$$P[T=t] = \sum_{\substack{x,y\in\mathbb{N}\\ x-y=t}} P[X=x, Y=y] = \sum_{\substack{x,y\in\mathbb{N}\\ x-y=t}} \frac{1}{4\cdot 2^{x+y}} \overset{y=x-t\geqslant 0}{=} \sum_{x=t}^{\infty} \frac{1}{4\cdot 2^{x+x-t}} =$$

$$= \frac{1}{4}\sum_{x=t}^{\infty} \frac{1}{\cdot 2^{2x-t}} = \frac{1}{4\cdot 2^{-t}} \sum_{x=t}^{\infty} \frac{1}{\cdot 4^x} = \frac{1}{4\cdot 2^{-t}} \left(\sum_{x=0}^{\infty} \frac{1}{\cdot 4^x} - \sum_{x=0}^{t-1} \frac{1}{\cdot 4^x} \right) =$$

$$= \frac{1}{4\cdot 2^{-t}} \left(\frac{1}{1-1/4} - \frac{1-(1/4)^t}{1-1/4} \right) = \frac{1}{4\cdot 2^{-t}} \left(\frac{4}{3} - \frac{4}{3}\frac{4^t-1}{4^t} \right) =$$

$$= \frac{1}{3\cdot 2^{-t}} \left(1 - \frac{4^t-1}{4^t} \right) = \frac{1}{3\cdot 2^{-t}\cdot 4^t} = \frac{1}{3\cdot 2^t}$$

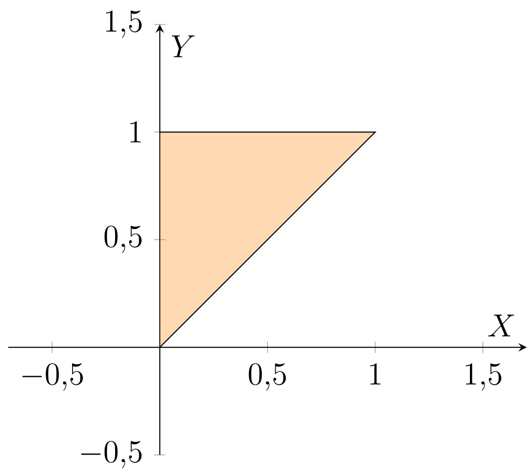
Ejercicio 7.3.9. El vector aleatorio (X, Y) se distribuye según una uniforme sobre el recinto

$$R_1 = \{(x,y); 0 < x < y < 1\}.$$

Calcular:

1. Su función generatriz de momentos conjunta.

   Veamos en primer lugar el conjunto R_1:



   Veamos ahora la función de densidad. Como se distribuye uniformemente, tenemos que:

$$f_{(X,Y)}(x,y) = \begin{cases} k & (x,y)\in R_1, \\ 0 & \text{en otro caso.} \end{cases}$$

   Para obtener el valor de k, tenemos que:

$$1 = \int_0^1 \int_x^1 k\,dy\,dx = k\int_0^1 (1-x)\,dx = k\left[x - \frac{x^2}{2}\right]_0^1 = k\left[1-\frac{1}{2}\right] = \frac{k}{2} \Longrightarrow k = 2$$

Por tanto, la función de densidad de probabilidad conjunta es:

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

La función generatriz de momentos conjunta es:

$$\begin{aligned}
M_{(X,Y)}(t_1,t_2) = E[e^{t_1X+t_2Y}] &= \int_0^1 \int_x^1 e^{t_1x+t_2y} f_{(X,Y)}(x,y)\,dy\,dx = \\
&= 2\int_0^1 \int_x^1 e^{t_1x+t_2y}\,dy\,dx = 2\int_0^1 e^{t_1x} \int_x^1 e^{t_2y}\,dy\,dx = \\
&= 2\int_0^1 e^{t_1x}\left[\frac{e^{t_2y}}{t_2}\right]_x^1 dx = 2\int_0^1 e^{t_1x}\left[\frac{e^{t_2}-e^{t_2x}}{t_2}\right] dx = \\
&= \frac{2}{t_2}\int_0^1 e^{t_1x+t_2} - e^{(t_1+t_2)x}\,dx = \frac{2}{t_2}\left[\frac{e^{t_1x+t_2}}{t_1} - \frac{e^{(t_1+t_2)x}}{t_1+t_2}\right]_0^1 = \\
&= \frac{2}{t_2}\left[\frac{e^{t_1+t_2}}{t_1} - \frac{e^{t_1+t_2}}{t_1+t_2} - \frac{e^{t_2}}{t_1} + \frac{1}{t_1+t_2}\right] = \frac{2}{t_2}\left[\frac{e^{t_1+t_2}-e^{t_2}}{t_1} - \frac{e^{t_1+t_2}-1}{t_1+t_2}\right]
\end{aligned}$$

Como no podemos evaluar en el origen, sería necesario ver si la función generatriz de momentos es continua en el origen. Para ello, es necesario calcular:

$$\lim_{(t_1,t_2)\to(0,0)} M_{(X,Y)}(t_1,t_2)$$

Este límite en varias variables es objetivo de estudio de Análisis, escapándose a los conocimientos de esta asignatura.

2. Las distribuciones generatrices de momentos marginales.

3. La covarianza de X e Y.

   La covarianza de X e Y es:

$$\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

Calculamos cada una de las esperanzas:

$$\begin{aligned}
E[X] &= \int_0^1 x f_X(x)\,dx = \int_0^1 x\int_x^1 2\,dy\,dx = 2\int_0^1 x(1-x)\,dx = 2\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \\
&= 2\left[\frac{1}{2} - \frac{1}{3}\right] = \frac{1}{3} \\
E[Y] &= \int_0^1 y f_Y(y)\,dy = \int_0^1 y\int_0^y 2\,dx\,dy = 2\int_0^1 y^2\,dy = 2\left[\frac{y^3}{3}\right]_0^1 = \frac{2}{3} \\
E[XY] &= \int_0^1\int_x^1 xy f_{(X,Y)}(x,y)\,dy\,dx = 2\int_0^1 x\int_x^1 y\,dy\,dx = 2\int_0^1 x\left[\frac{y^2}{2}\right]_x^1 dx = \\
&= 2\int_0^1 x\left[\frac{1}{2} - \frac{x^2}{2}\right] dx = 2\int_0^1 \left[\frac{x}{2} - \frac{x^3}{2}\right] dx = 2\left[\frac{x^2}{4} - \frac{x^4}{8}\right]_0^1 = \\
&= 2\left[\frac{1}{4} - \frac{1}{8}\right] = \frac{1}{4}
\end{aligned}$$

Por tanto, la covarianza de X e Y es:

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{1}{4} - \frac{1}{3} \cdot \frac{2}{3} = \frac{1}{4} - \frac{2}{9} = \frac{1}{36}$$

7.4. Independencia de Variables Aleatorias

Ejercicio 7.4.1. Sea un experimento aleatorio que consiste en lanzar un tetraedro regular cuyas caras están numeradas del 1 al 4, y se definen los sucesos:

$$A = \{1 \text{ ó } 2\}$$
$$B = \{2 \text{ ó } 3\}$$
$$C = \{2 \text{ ó } 4\}.$$

Responder a los siguientes apartados:

1. Calcular $P(A)$, $P(B)$, $P(C)$.

   Usando la Ley de Laplace, tenemos que:

   $$P(A) = \frac{2}{4} = \frac{1}{2}$$
   $$P(B) = \frac{2}{4} = \frac{1}{2}$$
   $$P(C) = \frac{2}{4} = \frac{1}{2}.$$

2. Calcular la probabilidad de obtener un dos.

   Usando de nuevo la Ley de Laplace, tenemos que:

   $$P(\{2\}) = \frac{1}{4}.$$

3. Indicar si los sucesos A, B y C son independientes dos a dos.

   Para comprobar si dos sucesos son independientes, debemos comprobar si se cumplen las siguientes igualdades:

   $$P(A \cap B) = P(A)P(B), \quad P(A \cap C) = P(A)P(C), \quad P(B \cap C) = P(B)P(C).$$

   Tenemos que:

   $$P(A \cap B) = P(\{2\}) = \frac{1}{4} = P(A)P(B) = \frac{1}{4}$$
   $$P(A \cap C) = P(\{2\}) = \frac{1}{4} = P(A)P(C) = \frac{1}{4}$$
   $$P(B \cap C) = P(\{2\}) = \frac{1}{4} = P(B)P(C) = \frac{1}{4}.$$

   Por lo tanto, los sucesos A, B y C son independientes dos a dos.

4. Indicar si los sucesos A, B y C son mutuamente independientes.

   Para comprobar si tres sucesos son mutuamente independientes, debemos comprobar si se cumple la siguiente igualdad:

   $$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Tenemos que:

$$P(A \cap B \cap C) = P(\{2\}) = \frac{1}{4} \neq P(A)P(B)P(C) = \frac{1}{8}.$$

Por lo tanto, los sucesos A, B y C no son mutuamente independientes.

Ejercicio 7.4.2. Sea (X_1, X_2, X_3) un vector aleatorio con función masa de probabilidad

$$P[(X_1, X_2, X_3) = (x_1, x_2, x_3)] = \frac{1}{4},$$

siendo $(x_1, x_2, x_3) = \{(1,0,0), (0,1,0), (0,0,1), (1,1,1)\}$.

1. Indicar si son X_1, X_2, X_3 independientes dos a dos.

   Calculamos en primer lugar las funciones masa de probabilidad marginales. Para el caso de X_1, tenemos que:

$$P[X_1 = 0] = P[(0,1,0)] + P[(0,0,1)] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$
$$P[X_1 = 1] = P[(1,0,0)] + P[(1,1,1)] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

   Para X_2 y X_3 se obtienen los mismos resultados. Por lo tanto, las funciones masas de probabilidad marginales para $i \in \{1,2,3\}$ son:

$$P[X_i = 0] = \frac{1}{2},$$
$$P[X_i = 1] = \frac{1}{2}.$$

   Calculemos ahora las funciones masa de probabilidad bidimensionales. Para el caso de X_1 y X_2, tenemos que:

$$P[(X_1, X_2) = (0,0)] = P[(0,1,0)] = \frac{1}{4}$$
$$P[(X_1, X_2) = (0,1)] = P[(0,0,1)] = \frac{1}{4}$$
$$P[(X_1, X_2) = (1,0)] = P[(1,0,0)] = \frac{1}{4}$$
$$P[(X_1, X_2) = (1,1)] = P[(1,1,1)] = \frac{1}{4}.$$

   Para (X_1, X_3) y (X_2, X_3) se obtienen los mismos resultados. Por lo tanto, las funciones masa de probabilidad bidimensionales para $i, j \in \{1,2,3\}$, $i \neq j$, son:

$$P[(X_i, X_j) = (0,0)] = \frac{1}{4},$$
$$P[(X_i, X_j) = (0,1)] = \frac{1}{4},$$
$$P[(X_i, X_j) = (1,0)] = \frac{1}{4},$$
$$P[(X_i, X_j) = (1,1)] = \frac{1}{4}.$$

Veamos ahora si son independientes dos a dos. Para $i, j \in \{1, 2, 3\}$, $i \neq j$ y $a, b \in \{0, 1\}$, tenemos que:

$$\frac{1}{4} = P[(X_i, X_j) = (a, b)] = P[X_i = a]P[X_j = b] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Por lo tanto, las variables X_1, X_2 y X_3 son independientes dos a dos.

2. Indicar si son X_1, X_2, X_3 mutuamente independientes.

   Tenemos que:

$$0 = P[(X_1, X_2, X_3) = (0, 0, 0)] \neq P[X_1 = 0]P[X_2 = 0]P[X_3 = 0] = \frac{1}{8}.$$

   Por lo tanto, las variables X_1, X_2 y X_3 no son mutuamente independientes.

3. Indicar si $X_1 + X_2$ y X_3 son independientes.

   Representamos en la siguiente tabla los valores de la suma $Z := X_1 + X_2$. Notemos que no es la función masa de probabilidad conjunta de (X_1, X_2), sino la tabla de los valores de Z.

$X_1 \backslash X_2$	0	1
0	0	1
1	1	2

   Notemos que Z toma los valores 0, 1 y 2, calculemos cada una de las probabilidades usando el Teorema de Cambio de Variable:

$$\begin{aligned}
&P[Z = 0] = P[(X_1, X_2) = (0, 0)] = 1/4 \\
&P[Z = 1] = P[(X_1, X_2) = (0, 1)] + P[(X_1, X_2) = (1, 0)] = 1/4 + 1/4 = 1/2 \\
&P[Z = 2] = P[(X_1, X_2) = (1, 1)] = 1/4.
\end{aligned}$$

   Tenemos que:

$$0 = P[(Z, X_3) = (0, 0)] = P[X_1 = 0, X_2 = 0, X_3 = 0] \neq P[Z = 0]P[X_3 = 0] = 1/4 \cdot 1/2.$$

   Por lo tanto, las variables $X_1 + X_2$ y X_3 no son independientes.

Ejercicio 7.4.3. Definimos sobre el experimento de lanzar diez veces una moneda las variables aleatorias X como el número de lanzamientos hasta que aparece la primera cara (si no aparece cara $X = 0$), e Y como el número de lanzamientos hasta que aparece la primera cruz (con $Y = 0$ si no aparece cruz). Indicar si X e Y son independientes.

La probabilidad de que haga falta 1 lanzamiento para obtener la primera cara, al igual que para obtener la primera cruz, usamos la Ley de Laplace:

$$P[X = 1] = P[Y = 1] = \frac{1}{2}$$

No obstante, por no poder darse en un mismo lanzamiento cara y cruz a la vez, tenemos que:

$$P[X = 1, Y = 1] = 0$$

Por tanto, como $P[X = 1]P[Y = 1] \neq P[X = 1, Y = 1]$, las variables X e Y no son independientes.

Ejercicio 7.4.4. El número de automóviles utilitarios, X, y el de automóviles de lujo, Y, que poseen las familias de una población se distribuye de acuerdo a las siguientes probabilidades:

$X\backslash Y$	0	1	2	
0	$1/3$	$1/12$	$1/24$	$11/24$
1	$1/6$	$1/24$	$1/48$	$11/48$
2	$5/22$	$5/88$	$5/176$	$5/16$
	$8/11$	$2/11$	$1/11$	

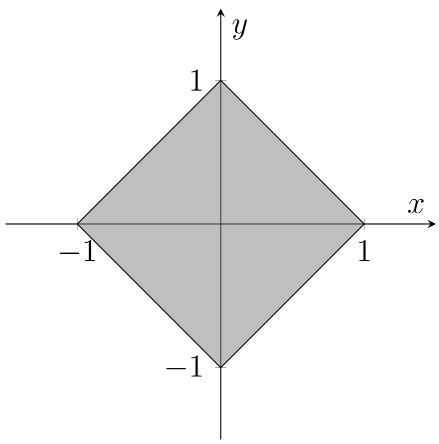
Comprobar que las variables X e Y son independientes.

Notemos que hemos incluido las funciones masa de probabilidad marginales en la última fila y columna de la tabla. Podemos comprobar así fácilmente que las variables X e Y son independientes.

Ejercicio 7.4.5. En los siguientes dos apartados, estudiar la independencia de las variables aleatorias X e Y, cuando su densidad de probabilidad conjunta se define como sigue:

1. $f(x, y) = 1/2$, si (x, y) pertenece al cuadrado de vértices $(1, 0); (0, 1); (-1, 0); (0, -1)$.

   El recinto descrito es:



   Calculamos cada una de las funciones de densidad marginales:

   - f_X si $x \in [-1, 0]$:

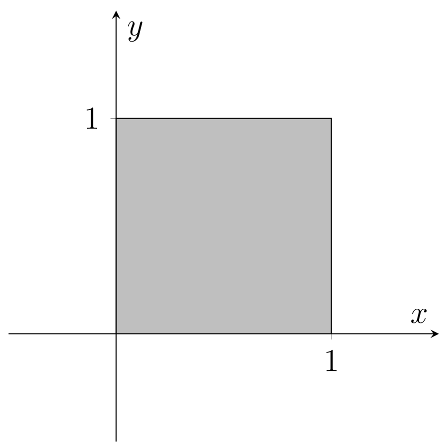
$$f_X(x) = \int_{-x-1}^{x+1} \frac{1}{2}\, dy = \left[\frac{1}{2}y\right]_{-x-1}^{x+1} = \frac{1}{2}(x + 1) - \frac{1}{2}(-x - 1) = x + 1.$$

- f_X si $x \in [0,1]$:
$$f_X(x) = \int_{x-1}^{-x+1} \frac{1}{2}\,dy = \left[\frac{1}{2}y\right]_{x-1}^{-x+1} = \frac{1}{2}(-x+1) - \frac{1}{2}(x-1) = 1 - x.$$
- f_Y si $y \in [-1,0]$:
$$f_Y(y) = \int_{-y-1}^{y+1} \frac{1}{2}\,dx = \left[\frac{1}{2}x\right]_{-y-1}^{y+1} = \frac{1}{2}(y+1) - \frac{1}{2}(-y-1) = y + 1.$$
- f_Y si $y \in [0,1]$:
$$f_Y(y) = \int_{y-1}^{-y+1} \frac{1}{2}\,dx = \left[\frac{1}{2}x\right]_{y-1}^{-y+1} = \frac{1}{2}(-y+1) - \frac{1}{2}(y-1) = 1 - y.$$

Por tanto, tenemos que las funciones de densidad marginales son:
$$f_X(x) = 1 - |x|, \quad f_Y(y) = 1 - |y|.$$
Por tanto, tenemos que:
$$f_X(x)f_Y(y) = (1-|x|)(1-|y|) = 1-|x|-|y|+|x||y| \qquad \forall x \in [-1,1], y \in [-x,x].$$
Tomando como ejemplo el origen, tenemos que:
$$\frac{1}{2} = f(0,0) \neq f_X(0)f_Y(0) = 1.$$

2. $f(x,y) = 1$, si (x,y) pertenece al cuadrado de vértices $(0,0)$; $(0,1)$; $(1,0)$; $(1,1)$.

   El recinto descrito es:



Definimos la siguiente función auxiliar:
$$\begin{array}{rccl} h_1 : & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & t & \longmapsto & \begin{cases} 1 & t \in [0,1] \\ 0 & t \notin [0,1] \end{cases} \end{array}$$
Tenemos que:
$$f(x,y) = h_1(x)h_1(y) \qquad \forall x,y \in \mathbb{R}.$$
Por tanto, las variables X e Y son independientes.

Ejercicio 7.4.6. Sean X_1 y X_2 variables aleatorias independientes con distribución Binomial con parámetros n_i, $i = 1, 2$, y $p = 1/2$. Calcular la distribución de $X_1 - X_2 + n_2$.

Sea $Z = X_1 - X_2 + n_2$. Calculamos generatriz de momentos de Z:

$$M_Z(t) = E[e^{tZ}] = E[e^{t(X_1 - X_2 + n_2)}] = E[e^{tX_1}e^{-tX_2}e^{tn_2}] = e^{tn_2}E[e^{tX_1}e^{-tX_2}].$$

Fijado $t \in \mathbb{R}$, como la transformación $u \mapsto e^{tu}$ es medible, tenemos que e^{tX_1} y e^{-tX_2} son independientes. Por tanto, usando el Teorema de la Multiplicación de las Esperanzas, tenemos que:

$$M_Z(t) = e^{tn_2}E[e^{tX_1}]E[e^{-tX_2}] = e^{tn_2}M_{X_1}(t)M_{X_2}(-t)$$

Usando la generatriz de momentos de la Binomial, tenemos que:

$$\begin{aligned}
M_Z(t) &= e^{tn_2}\left(1 + p(e^t - 1)\right)^{n_1}\left(1 + p(e^{-t} - 1)\right)^{n_2} = \\
&= (1 + p(e^t - 1))^{n_1}[e^t(1 + p(e^{-t} - 1))]^{n_2} = \\
&= (1 + p(e^t - 1))^{n_1}(1 + p(e^t - 1))^{n_2} \overset{(*)}{=} \\
&\overset{(*)}{=} (1 + p(e^t - 1))^{n_1 + n_2}.
\end{aligned}$$

donde en $(*)$ hemos usado que $p = 1/2$.

Por tanto, la variable Z sigue una distribución Binomial con parámetros $n_1 + n_2$ y $p = 1/2$. Es decir,

$$X_1 - X_2 + n_2 \sim B(n_1 + n_2, p).$$

Ejercicio 7.4.7. La demanda en miles de toneladas de un producto, X, y su precio por kilogramo en euros, Y, tienen por función de densidad conjunta

$$f(x, y) = kx^2(1 - x)^3y^3(1 - y)^2, \quad x, y \in]0, 1[.$$

Calcular la constante k para que f sea una función de densidad de probabilidad, y determinar si X e Y son independientes. Obtener la función de densidad de probabilidad del precio para una demanda fija.

Para que f sea una función de densidad de probabilidad, debe cumplir que:

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

Por tanto, calculamos la constante k:

$$\begin{aligned}
1 &= \int_0^1 \int_0^1 kx^2(1-x)^3y^3(1-y)^2\;dx\;dy = k\int_0^1 x^2(1-x)^3\;dx \int_0^1 y^3(1-y)^2\;dy = \\
&= k\int_0^1 x^2(1-3x+3x^2-x^3)\;dx \int_0^1 y^3(1-2y+y^2)\;dy = \\
&= k\int_0^1 (x^2-3x^3+3x^4-x^5)\;dx \int_0^1 (y^3-2y^4+y^5)\;dy = \\
&= k\left[\frac{1}{3}x^3-\frac{3}{4}x^4+\frac{3}{5}x^5-\frac{1}{6}x^6\right]_0^1 \left[\frac{1}{4}y^4-\frac{2}{5}y^5+\frac{1}{6}y^6\right]_0^1 = \\
&= k\left(\frac{1}{3}-\frac{3}{4}+\frac{3}{5}-\frac{1}{6}\right)\left(\frac{1}{4}-\frac{2}{5}+\frac{1}{6}\right) = \\
&= k\cdot\frac{1}{60}\cdot\frac{1}{60} = \frac{k}{3600} \Longrightarrow k = 3600.
\end{aligned}$$

Comprobamos ahora si X e Y son independientes. Para ello, definimos funciones auxiliares:

$$\begin{array}{rccl}
h_1: & \mathbb{R} & \longrightarrow & \mathbb{R} \\
& x & \longmapsto & \begin{cases} k\cdot x^2(1-x)^3 & x\in\,]0,1[\\ 0 & x\notin\,]0,1[\end{cases}
\end{array}$$

$$\begin{array}{rccl}
h_2: & \mathbb{R} & \longrightarrow & \mathbb{R} \\
& y & \longmapsto & \begin{cases} y^3(1-y)^2 & y\in\,]0,1[\\ 0 & y\notin\,]0,1[\end{cases}
\end{array}$$

Por tanto, tenemos que:

$$f(x,y) = h_1(x)h_2(y) \qquad \forall x,y\in\;\mathbb{R}$$

Por tanto, las variables X e Y son independientes. Buscamos ahora la función de densidad de probabilidad del precio para una demanda fija $D\in\,]0,1[$. Para ello, como X e Y son independientes, tenemos que:

$$\begin{aligned}
f_{Y|x=D}(y) = f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\;dx = ky^3(1-y)^2\int_0^1 x^2(1-x)^3\;dx = \\
&= 3600\cdot y^3(1-y)^2\cdot\frac{1}{60} = 60\cdot y^3(1-y)^2 \qquad \forall y\in\;]0,1[\,.
\end{aligned}$$

7.5. Esperanza Condicionada

Ejercicio 7.5.1. Sea X una variable aleatoria que se distribuye uniformemente en el intervalo $]0,1[$. Comprobar si las variables aleatorias X y $|1/2 - X|$ son incorreladas.

Como X es uniforme en $]0,1[$, tenemos que:

$$f_X(x) = \begin{cases} 1, & x \in]0,1[\\ 0, & \text{en otro caso} \end{cases}$$

Calculamos las siguientes esperanzas:

$$E[X] = \int_0^1 x f_X(x)\, dx = \int_0^1 x\, dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

$$E[|1/2 - X|] = \int_0^{1/2} (1/2 - x)\, dx + \int_{1/2}^1 (x - 1/2)\, dx = \left[\frac{x}{2} - \frac{x^2}{2}\right]_0^{1/2} + \left[\frac{x^2}{2} - \frac{x}{2}\right]_{1/2}^1 = \frac{1}{4}$$

$$E[X \cdot |1/2 - X|] = \int_0^{1/2} x(1/2 - x)\, dx + \int_{1/2}^1 x(x - 1/2)\, dx = \left[\frac{x^2}{4} - \frac{x^3}{3}\right]_0^{1/2} + \left[\frac{x^3}{3} - \frac{x^2}{4}\right]_{1/2}^1 = \frac{1}{8}$$

Calculamos ahora la covarianza:

$$\mathrm{Cov}[X, |1/2 - X|] = E[X \cdot |1/2 - X|] - E[X]E[|1/2 - X|] = \frac{1}{8} - \frac{1}{2} \cdot \frac{1}{4} = 0$$

Por tanto, tenemos que X y $|1/2 - X|$ son incorreladas.

Ejercicio 7.5.2. Calcular las curvas de regresión y las razones de correlación para las siguientes distribuciones, comentando los resultados.

1. Considerar las distribución conjunta (X,Y) con función de masa de probabilidad dada por:

$X \setminus Y$	10	15	20	
1	0	$2/6$	0	$2/6$
2	$1/6$	0	0	$1/6$
3	0	0	$3/6$	$3/6$
	$1/6$	$2/6$	$3/6$	1

Tras haber calculado las distribuciones marginales, calculamos ahora las distribuciones condicionadas. La siguiente tabla muestra la distribución condicionada de Y dado X, $P[Y = y \mid X = x]$:

$X \setminus Y$	10	15	20
1	0	1	0
2	1	0	0
3	0	0	1

La distribución condicionada de X dado Y, $P[X = x \mid Y = y]$ viene dada por la misma tabla, ya que en este caso tenemos que:

$$P[X = x \mid Y = y] = P[Y = y \mid X = x] \qquad \forall x, y$$

Calculemos ahora las curvas de regresión y las razones de correlación.

- Curva de regresión de Y sobre X:

$$\widehat{Y}(x) = E[Y \mid X = x] = \sum_y yP[Y = y \mid X = x] \qquad \forall x \in E_x$$

Por tanto, la curva de regresión de Y sobre X es:

$$\widehat{Y}(1) = 15 \qquad \widehat{Y}(2) = 10 \qquad \widehat{Y}(3) = 20$$

Para calcular la razón de correlación de Y sobre X, hay dos opciones:

Opción 1. Método rutinario.
Usamos la fórmula:

$$\eta^2_{Y/X} = \frac{\text{Var}(E[Y \mid X])}{\text{Var}(Y)}$$

- Calculemos $E[Y]$:

$$E[Y] = \sum_y yP[Y = y] \quad = 10 \cdot {}^1\!/_6 + 15 \cdot {}^2\!/_6 + 20 \cdot {}^3\!/_6 = \frac{50}{3}$$

- Calculemos ahora $E[Y^2]$:

$$E[Y^2] = \sum_y y^2 P[Y = y] = 10^2 \cdot {}^1\!/_6 + 15^2 \cdot {}^2\!/_6 + 20^2 \cdot {}^3\!/_6 = \frac{875}{3}$$

- Calculemos ahora $E[(E[Y \mid X])^2]$:

$$E[(E[Y \mid X])^2] = \sum_x (E[Y \mid X = x])^2 P[X = x] =$$
$$= 10^2 \cdot {}^1\!/_6 + 15^2 \cdot {}^2\!/_6 + 20^2 \cdot {}^3\!/_6 = \frac{875}{3} = E[Y^2]$$

- Calculemos ahora $E[E[Y \mid X]]$.

$$E[E[Y \mid X]] = E[Y]$$

Por tanto, usando lo anterior, tenemos:

$$\eta^2_{Y/X} = \frac{\text{Var}(E[Y \mid X])}{\text{Var}(Y)} = \frac{E[(E[Y \mid X])^2] - E[E[Y \mid X]]^2}{E[Y^2] - E[Y]^2}$$
$$= \frac{E[Y^2] - E[Y]^2}{E[Y^2] - E[Y]^2} = 1$$

Opción 2. Razonando por dependencia funcional.
En este caso, vemos que Y es función de X. Por tanto:

$$\eta^2_{Y/X} = 1$$

- Curva de regresión de X sobre Y:

$$\widehat{X}(y) = E[X \mid Y = y] = \sum_x xP[X = x \mid Y = y] \qquad \forall y \in E_y$$

Por tanto, la curva de regresión de X sobre Y es:

$$\widehat{X}(10) = 2 \qquad \widehat{X}(15) = 1 \qquad \widehat{X}(20) = 3$$

De nuevo, razonando ahora por dependencia funcional, tenemos que:

$$\eta^2_{X/Y} = 1$$

Como vemos, en este caso, hay dependencia recíproca entre X e Y. Por tanto, el ajuste es el idea, ya que $Y = f(X)$ y $X = g(Y)$. Cada una explica la totalidad de la variabilidad de la otra.

2. Considerar las distribución conjunta (X, Y) con función de masa de probabilidad dada por:

$X \setminus Y$	10	15	20	25	
1	0	$3/7$	0	$1/7$	$4/7$
2	0	0	$1/7$	0	$1/7$
3	$2/7$	0	0	0	$2/7$
	$2/7$	$3/7$	$1/7$	$1/7$	1

Tras haber calculado las distribuciones marginales, calculamos ahora las distribuciones condicionadas. La siguiente tabla muestra la distribución condicionada de Y dado X, $P[Y = y \mid X = x]$:

$X \setminus Y$	10	15	20	25
1	0	$3/4$	0	$1/4$
2	0	0	1	0
3	1	0	0	0

La distribución condicionada de X dado Y, $P[X = x \mid Y = y]$ viene dada por la siguiente tabla:

$X \setminus Y$	10	15	20	25
1	0	1	0	1
2	0	0	1	0
3	1	0	0	0

Calculemos ahora las curvas de regresión y las razones de correlación.

- Curva de regresión de Y sobre X:

$$\widehat{Y}(x) = E[Y \mid X = x] = \sum_y yP[Y = y \mid X = x] \qquad \forall x \in E_x$$

Por tanto, la curva de regresión de Y sobre X es:

$$\widehat{Y}(1) = 15 \cdot 3/4 + 25 \cdot 1/4 = 17{,}5$$
$$\widehat{Y}(2) = 20$$
$$\widehat{Y}(3) = 10$$

Para calcular la razón de correlación de Y sobre X, tenemos que:

$$\eta^2_{Y/X} = \frac{\mathrm{Var}(E[Y \mid X])}{\mathrm{Var}(Y)}$$

- Calculemos $E[Y]$:

$$E[Y] = \sum_y yP[Y = y] = 10 \cdot {}^2\!/_7 + 15 \cdot {}^3\!/_7 + 20 \cdot {}^1\!/_7 + 25 \cdot {}^1\!/_7 = \frac{110}{7}$$

- Calculemos ahora $E[Y^2]$:

$$\begin{aligned} E[Y^2] &= \sum_y y^2 P[Y = y] \\ &= 10^2 \cdot {}^2\!/_7 + 15^2 \cdot {}^3\!/_7 + 20^2 \cdot {}^1\!/_7 + 25^2 \cdot {}^1\!/_7 = \frac{1900}{7} \end{aligned}$$

- Calculemos ahora $E[(E[Y \mid X])^2]$:

$$\begin{aligned} E[(E[Y \mid X])^2] &= \sum_x (E[Y \mid X = x])^2 P[X = x] \\ &= 17{,}5^2 \cdot {}^4\!/_7 + 20^2 \cdot {}^1\!/_7 + 10^2 \cdot {}^2\!/_7 = \frac{1825}{7} \end{aligned}$$

- Calculemos ahora $E[E[Y \mid X]]$.

$$E[E[Y \mid X]] = E[Y]$$

Por tanto, usando lo anterior, tenemos:

$$\begin{aligned} \eta^2_{Y/X} &= \frac{\mathrm{Var}(E[Y \mid X])}{\mathrm{Var}(Y)} = \frac{E[(E[Y \mid X])^2] - E[E[Y \mid X]]^2}{E[Y^2] - E[Y]^2} \\ &= \frac{E[(E[Y \mid X])^2] - E[Y]^2}{E[Y^2] - E[Y]^2} \\ &= \frac{\frac{1825}{7} - \left(\frac{110}{7}\right)^2}{\frac{1900}{7} - \left(\frac{110}{7}\right)^2} = \frac{9}{16} \approx 0{,}5625 \end{aligned}$$

Por tanto, tenemos que X explica el 56,25 % de la variabilidad de Y. Tenemos entonces que no es un ajuste ideal.

- Curva de regresión de X sobre Y:
  En este caso, tenemos que $X = f(Y)$, por lo que:

$$\widehat{X}(10) = 3, \quad \widehat{X}(15) = 1, \quad \widehat{X}(20) = 2, \quad \widehat{X}(25) = 1$$

Por tanto, como X es función de Y, tenemos que:

$$\eta^2_{X/Y} = 1$$

Tenemos que Y explica la totalidad de la variabilidad de X, por lo que el ajuste es el ideal.

Ejercicio 7.5.3. Sea X el número de balanzas e Y el número de dependientes en los puntos de venta de un mercado. Determinar las rectas de regresión y el grado de ajuste a la distribución, si la función masa de probabilidad de (X, Y) viene dada por:

$X \setminus Y$	1	2	3	4	
1	$1/24$	$2/24$	0	0	$3/24$
2	$1/24$	$2/24$	$3/24$	$1/24$	$7/24$
3	0	$1/24$	$2/24$	$6/24$	$9/24$
4	0	0	$2/24$	$3/24$	$5/24$
	$2/24$	$5/24$	$7/24$	$10/24$	1

Tras calcular las distribuciones marginales, calculamos ahora las esperanzas necesarias para que el resto de cálculos posteriormente sean directos.

$$\begin{aligned}
E[X] &= 1 \cdot 3/24 + 2 \cdot 7/24 + 3 \cdot 9/24 + 4 \cdot 5/24 = 8/3 \\
E[Y] &= 1 \cdot 2/24 + 2 \cdot 5/24 + 3 \cdot 7/24 + 4 \cdot 10/24 = 73/24 \\
E[X^2] &= 1^2 \cdot 3/24 + 2^2 \cdot 7/24 + 3^2 \cdot 9/24 + 4^2 \cdot 5/24 = 8 \\
E[Y^2] &= 1^2 \cdot 2/24 + 2^2 \cdot 5/24 + 3^2 \cdot 7/24 + 4^2 \cdot 10/24 = 245/24 \\
E[XY] &= \sum_{\substack{x \in E_X \\ y \in E_y}} xyP[X = x, Y = y] = 209/24 \\
\mathrm{Var}[X] &= E[X^2] - E[X]^2 = 8/9 \\
\mathrm{Var}[Y] &= E[Y^2] - E[Y]^2 = 551/576 \\
\mathrm{Cov}[X, Y] &= E[XY] - E[X]E[Y] = 43/72
\end{aligned}$$

Calculamos ahora las rectas de regresión.

- Recta de regresión de Y sobre X:

$$\begin{aligned}
\widehat{Y}(x) &= E[Y] + \frac{\mathrm{Cov}[X, Y]}{\mathrm{Var}[X]}(X - E[X]) = \frac{73}{24} + \frac{43/72}{8/9}(X - 8/3) = \\
&= 73/24 + \frac{43}{64}(X - 8/3)
\end{aligned}$$

- Recta de regresión de X sobre Y:

$$\begin{aligned}
\widehat{X}(y) &= E[X] + \frac{\mathrm{Cov}[X, Y]}{\mathrm{Var}[Y]}(Y - E[Y]) = 8/3 + \frac{43/72}{551/576}(Y - 73/24) = \\
&= 8/3 + \frac{344}{551}(Y - 73/24)
\end{aligned}$$

Para determinar el grado de ajuste a la distribución, calculamos ahora el coeficiente de determinación lineal:

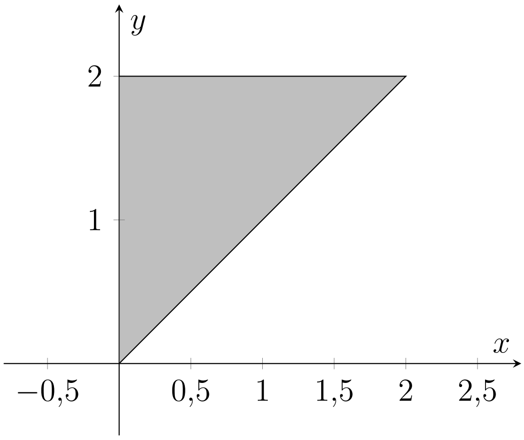
$$\rho^2_{X,Y} = \frac{\mathrm{Cov}[X, Y]^2}{\mathrm{Var}[X]\,\mathrm{Var}[Y]} = \frac{1849}{4408} \approx 0{,}419$$

Por tanto, tenemos que el 41,9 % de la variabilidad de Y queda explicada por la regresión lineal de Y sobre X. Vemos por tanto que el ajuste no es bueno.

Ejercicio 7.5.4. Sea (X, Y) un vector aleatorio con valores en el conjunto dado por $T = \{(x, y) \in \mathbb{R}^2 / 0 < x < y < 2\}$ y función de densidad constante. Calcular:

1. Función de densidad de probabilidad conjunta.

   Veamos en primer lugar el conjunto en el que se distribuye el vector aleatorio (X, Y):

   Como la función de densidad es constante, tenemos que:

   $$f_{(X,Y)}(x, y) = \begin{cases} k, & (x, y) \in T \\ 0, & \text{en otro caso} \end{cases}$$

   Para calcular k, tenemos que:

   $$1 = \int_T f_{(X,Y)} = \int_T k = k\lambda(T) = k \cdot 2 \Longrightarrow k = 1/2$$

2. Curvas y rectas de regresión de X sobre Y y de Y sobre X.

   Comenzamos calculando las curvas de regresión, ya que si estas son funciones lineales, las rectas de regresión coincidirán con las curvas de regresión. Para ello, calculamos las funciones de densidad marginales y condicionadas.

   - Función de densidad de X. Para $x \in]0, 2[$:

     $$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y)\, dy = \int_x^2 1/2\, dy = 1/2(2 - x) = 1 - \frac{x}{2}$$

   - Función de densidad de Y. Para $y \in]0, 2[$:

     $$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y)\, dx = \int_0^y 1/2\, dx = \frac{y}{2}$$

   - Función de densidad condicionada de X dado $Y = y^* \in]0, 2[$. Para $x \in]0, y^*[$:

     $$f_{X|Y=y^*}(x) = \frac{f_{(X,Y)}(x, y^*)}{f_Y(y^*)} = \frac{1/2}{y^*/2} = \frac{1}{y^*}$$

- Función de densidad condicionada de Y dado $X = x^* \in]0, 2[$. Para $y \in]x^*, 2[$:

$$f_{Y|X=x^*}(y) = \frac{f_{(X,Y)}(x^*, y)}{f_X(x^*)} = \frac{1/2}{1 - x^*/2} = \frac{1}{2 - x^*}$$

Calculamos ahora las curvas de regresión.

- Curva de regresión de X sobre Y:

$$\widehat{X}(y) = E[X \mid Y = y] = \int_0^y x f_{X|Y=y}(x)\, dx = \int_0^y x \cdot \frac{1}{y}\, dx = \frac{1}{y}\int_0^y x\, dx =$$
$$= \frac{1}{y}\left[\frac{x^2}{2}\right]_0^y = \frac{y}{2} \qquad \forall y \in]0, 2[$$

- Curva de regresión de Y sobre X:

$$\widehat{Y}(x) = E[Y \mid X = x] = \int_x^2 y f_{Y|X=x}(y)\, dy = \int_x^2 y \cdot \frac{1}{2 - x}\, dy = \frac{1}{2 - x}\int_x^2 y\, dy =$$
$$= \frac{1}{2 - x}\left[\frac{y^2}{2}\right]_x^2 = \frac{2^2 - x^2}{2(2 - x)} = \frac{4 - x^2}{2(2 - x)} = \frac{2 + x}{2} = 1 + \frac{x}{2} \qquad \forall x \in]0, 2[$$

Tenemos además que las curvas de regresión son funciones lineales, por lo que las rectas de regresión coinciden con las curvas de regresión.

3. Razones de correlación y coeficiente de correlación lineal.

Como las curvas de regresión son funciones lineales, tenemos que las razones de correlación coinciden y son iguales al coeficiente de determinación lineal.

$$\eta^2_{Y/X} = \eta^2_{X/Y} = \rho^2_{X,Y} = \frac{\text{Cov}[X, Y]^2}{\text{Var}[X]\,\text{Var}[Y]}$$

Como este es el producto de las pendientes de las rectas de regresión, tenemos que:

$$\rho^2_{X,Y} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Como la correlación es positiva por ser ambas rectas crecientes, tenemos que:

$$\rho_{X,Y} = +\sqrt{\rho^2_{X,Y}} = 1/2$$

4. Error cuadrático medio asociado a cada una de las funciones de regresión.

Por lo visto en teoría, sabemos que:

$$\text{ECM}(\widehat{X}) = \text{Var}[X] - \frac{\text{Cov}[X, Y]^2}{\text{Var}[Y]}$$
$$\text{ECM}(\widehat{Y}) = \text{Var}[Y] - \frac{\text{Cov}[X, Y]^2}{\text{Var}[X]}$$

Tenemos dos opciones:

Método Rutinario Calculamos por tanto los valores necesarios:

$$E[X] = \int_0^2 x f_X(x)\ dx = \int_0^2 x(1-x/2)\ dx = \left[\frac{x^2}{2} - \frac{x^3}{6}\right]_0^2 = \frac{4}{2} - \frac{8}{6} = \frac{2}{3}$$

$$E[Y] = \int_0^2 y f_Y(y)\ dy = \int_0^2 y \cdot y/2\ dy = \left[\frac{y^3}{6}\right]_0^2 = \frac{8}{6} = \frac{4}{3}$$

$$E[X^2] = \int_0^2 x^2 f_X(x)\ dx = \int_0^2 x^2(1-x/2)\ dx = \left[\frac{x^3}{3} - \frac{x^4}{8}\right]_0^2 = \frac{8}{3} - \frac{16}{8} = \frac{2}{3}$$

$$E[Y^2] = \int_0^2 y^2 f_Y(y)\ dy = \int_0^2 y^2 \cdot y/2\ dy = \left[\frac{y^4}{8}\right]_0^2 = \frac{16}{8} = 2$$

$$E[XY] = \int_0^2 \int_x^2 xy f_{(X,Y)}(x,y)\ dy\ dx = \int_0^2 \int_x^2 xy \cdot 1/2\ dy\ dx = \frac{1}{2}\int_0^2 x \left[\frac{y^2}{2}\right]_x^2 dx =$$

$$= \frac{1}{4}\int_0^2 x\left(4 - x^2\right)\ dx = \frac{1}{4}\left[2x^2 - \frac{x^4}{4}\right]_0^2 = \frac{1}{4}(8-4) = 1$$

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = \frac{2}{3} - \left(\frac{2}{3}\right)^2 = \frac{2}{9}$$

$$\mathrm{Var}[Y] = E[Y^2] - E[Y]^2 = 2 - \left(\frac{4}{3}\right)^2 = \frac{2}{9}$$

$$\mathrm{Cov}[X,Y] = E[XY] - E[X]E[Y] = 1 - \frac{2}{3}\cdot\frac{4}{3} = \frac{1}{9}$$

Por tanto, tenemos que:

$$\mathrm{ECM}(\widehat{X}) = \mathrm{Var}[X] - \frac{\mathrm{Cov}[X,Y]^2}{\mathrm{Var}[Y]} = \frac{2}{9} - \frac{1/9^2}{2/9} = \frac{2}{9} - \frac{1}{18} = \frac{3}{18} = \frac{1}{6}$$

$$\mathrm{ECM}(\widehat{Y}) = \mathrm{Var}[Y] - \frac{\mathrm{Cov}[X,Y]^2}{\mathrm{Var}[X]} = \frac{2}{9} - \frac{1/9^2}{2/9} = \frac{2}{9} - \frac{1}{18} = \frac{3}{18} = \frac{1}{6}$$

Razonar los datos que ya conocemos Tenemos que:

$$\mathrm{ECM}(\widehat{X}) = \mathrm{Var}[X] - \frac{\mathrm{Cov}[X,Y]^2}{\mathrm{Var}[Y]} =$$

$$= \mathrm{Var}[X] - \frac{\mathrm{Cov}[X,Y]^2}{\mathrm{Var}[Y]\,\mathrm{Var}[X]}\cdot \mathrm{Var}[X] = \mathrm{Var}[X] - \rho_{X,Y}^2\,\mathrm{Var}[X]$$

$$\mathrm{ECM}(\widehat{Y}) = \mathrm{Var}[Y] - \frac{\mathrm{Cov}[X,Y]^2}{\mathrm{Var}[X]}$$

Por tanto, por este método tan solo necesitaríamos calcular las varianzas, ahorrándonos el cálculo de $E[XY]$, que al ser una integral doble puede resultar complicada.

Ejercicio 7.5.5. Dada la función masa de probabilidad del vector aleatorio (X,Y)

$X \setminus Y$	0	1	2	3	
0	0,2	0,2	0,05	0	0,45
1	0,1	0,1	0,1	0,05	0,35
2	0	0,05	0,05	0,1	0,2
	0,3	0,35	0,2	0,15	1

1. Determinar la aproximación lineal mínimo cuadrática de Y para $X = 1$.

   Nos piden la recta de regresión de Y sobre X, y evaluarla en el punto $X = 1$. Calculamos las esperanzas necesarias para el cálculo de la recta de regresión.

$$\begin{aligned}
E[X] &= 0 \cdot 0{,}3 + 1 \cdot 0{,}35 + 2 \cdot 0{,}2 = 0{,}75 \\
E[Y] &= 0 \cdot 0{,}3 + 1 \cdot 0{,}35 + 2 \cdot 0{,}2 + 3 \cdot 0{,}15 = 1{,}2 \\
E[X^2] &= 0^2 \cdot 0{,}3 + 1^2 \cdot 0{,}35 + 2^2 \cdot 0{,}2 = 1{,}15 \\
E[XY] &= \sum_{\substack{x \in E_X \\ y \in E_Y}} xyP[X = x, Y = y] = 1{,}35 \\
\mathrm{Var}[X] &= E[X^2] - E[X]^2 = 1{,}15 - 0{,}75^2 = 0{,}5875 \\
\mathrm{Cov}[X, Y] &= E[XY] - E[X]E[Y] = 1{,}35 - 0{,}75 \cdot 1{,}2 = 0{,}45
\end{aligned}$$

   Calculamos ahora la recta de regresión de Y sobre X.

$$\widehat{Y}(x) = E[Y] + \frac{\mathrm{Cov}[X, Y]}{\mathrm{Var}[X]}(x - E[X]) = 1{,}2 + \frac{0{,}45}{0{,}5875}(x - 0{,}75) = 1{,}2 + \frac{36}{47}(x - 0{,}75)$$

   Evaluando en $X = 1$, tenemos que:

$$\widehat{Y}(1) = 1{,}2 + \frac{36}{47}(1 - 0{,}75) = 1{,}2 + \frac{36}{47} \cdot 0{,}25 = 1{,}2 + \frac{9}{47} \approx 1{,}391489$$

2. Determinar la aproximación mínimo cuadrática de Y para $X = 1$.

   En este caso, piden $E[Y \mid X = 1]$. Para ello, hemos de calcular las distribución condicionada de Y dado $X = 1$.

$$\begin{aligned}
P[Y = 0 \mid X = 1] &= \frac{P[X = 1, Y = 0]}{P[X = 1]} = \frac{0{,}1}{0{,}35} = \frac{2}{7} \\
P[Y = 1 \mid X = 1] &= \frac{P[X = 1, Y = 1]}{P[X = 1]} = \frac{0{,}1}{0{,}35} = \frac{2}{7} \\
P[Y = 2 \mid X = 1] &= \frac{P[X = 1, Y = 2]}{P[X = 1]} = \frac{0{,}1}{0{,}35} = \frac{2}{7} \\
P[Y = 3 \mid X = 1] &= \frac{P[X = 1, Y = 3]}{P[X = 1]} = \frac{0{,}05}{0{,}35} = \frac{1}{7}
\end{aligned}$$

   Por tanto, tenemos que:

$$\begin{aligned}
E[Y \mid X = 1] = \sum_{y \in E_Y} yP[Y = y \mid X = 1] = \\
&= 0 + 1 \cdot P[Y = 1 \mid X = 1] + 2 \cdot P[Y = 2 \mid X = 1] + 3 \cdot P[Y = 3 \mid X = 1] = \\
&= \frac{2}{7} + 2 \cdot \frac{2}{7} + 3 \cdot \frac{1}{7} = \frac{9}{7}
\end{aligned}$$

Ejercicio 7.5.6. Dadas las siguientes distribuciones, determinar qué variable, X ó X', aproxima mejor a la variable Y:

$X \setminus Y$	0	1	2	
0	$1/5$	0	0	$1/5$
2	0	$1/5$	0	$1/5$
3	$1/5$	0	$1/5$	$2/5$
4	0	0	$1/5$	$1/5$
	$2/5$	$1/5$	$2/5$	1

$X' \setminus Y$	0	1	2	
0	$1/5$	0	$1/5$	$2/5$
2	0	$1/5$	0	$1/5$
3	$1/5$	0	0	$1/5$
4	0	0	$1/5$	$1/5$
	$2/5$	$1/5$	$2/5$	1

Hemos de obtener $\eta^2_{Y/X}$ y $\eta^2_{Y/X'}$ para compararlas.

$$\eta^2_{Y/X} = \frac{\mathrm{Var}[E[Y \mid X]]}{\mathrm{Var}[Y]}, \qquad \eta^2_{Y/X'} = \frac{\mathrm{Var}[E[Y \mid X']]}{\mathrm{Var}[Y]}$$

Calculamos las esperanzas necesarias para el cálculo de las varianzas.

$$\begin{aligned}
E[Y] &= 0 \cdot 2/5 + 1 \cdot 1/5 + 2 \cdot 2/5 = 1 \\
E[Y^2] &= 0^2 \cdot 2/5 + 1^2 \cdot 1/5 + 2^2 \cdot 2/5 = 9/5 \\
E[E[Y \mid X]] &= E[Y] = 1 \\
E[E[Y \mid X']] &= E[Y] = 1 \\
E[(E[Y \mid X])^2] &= \sum_{x \in E_X} (E[Y \mid X = x])^2 P[X = x] \\
E[(E[Y \mid X'])^2] &= \sum_{x \in E_{X'}} (E[Y \mid X' = x])^2 P[X' = x]
\end{aligned}$$

Calculamos por tanto las esperanzas condicionadas.

$$\begin{aligned}
E[Y \mid X = 0] &= 0 \cdot 1 + 1 \cdot 0 + 2 \cdot 0 = 0 \\
E[Y \mid X = 2] &= 0 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 = 1 \\
E[Y \mid X = 3] &= 0 \cdot 1/2 + 1 \cdot 0 + 2 \cdot 1/2 = 1 \\
E[Y \mid X = 4] &= 0 \cdot 0 + 1 \cdot 0 + 2 \cdot 1 = 2 \\
E[Y \mid X' = 0] &= 0 \cdot 1/2 + 1 \cdot 0 + 2 \cdot 1/2 = 1 \\
E[Y \mid X' = 2] &= 0 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 = 1 \\
E[Y \mid X' = 3] &= 0 \cdot 1 + 1 \cdot 0 + 2 \cdot 0 = 0 \\
E[Y \mid X' = 4] &= 0 \cdot 0 + 1 \cdot 0 + 2 \cdot 1 = 2
\end{aligned}$$

Por tanto, tenemos que:

$$\begin{aligned}
E[(E[Y \mid X])^2] &= \sum_{x \in E_X} (E[Y \mid X = x])^2 P[X = x] = \\
&= 0^2 \cdot 1/5 + 1^2 \cdot 1/5 + 1^2 \cdot 2/5 + 2^2 \cdot 1/5 = 7/5 \\
E[(E[Y \mid X'])^2] &= \sum_{x \in E_{X'}} (E[Y \mid X' = x])^2 P[X' = x] = \\
&= 1^2 \cdot 2/5 + 1^2 \cdot 1/5 + 0^2 \cdot 1/5 + 2^2 \cdot 1/5 = 7/5 \\
\mathrm{Var}[E[Y \mid X]] &= E[(E[Y \mid X])^2] - E[E[Y \mid X]]^2 = 7/5 - 1 = 2/5 \\
\mathrm{Var}[E[Y \mid X']] &= E[(E[Y \mid X'])^2] - E[E[Y \mid X']]^2 = 7/5 - 1 = 2/5
\end{aligned}$$

Por tanto, tenemos que:

$$\eta^2_{Y/X} = \frac{\operatorname{Var}[E[Y \mid X]]}{\operatorname{Var}[Y]} = \frac{2/5}{4/5} = \frac{1}{2} \stackrel{(*)}{=} \eta^2_{Y/X'}$$

donde en $(*)$ hemos usado que $\operatorname{Var}[E[Y \mid X]] = \operatorname{Var}[E[Y \mid X']]$.

Por tanto, aun siendo ambos ajustes de igual calidad, tenemos que ninguno de los ajustes es bueno, ya que solo explican la mitad de la variabilidad de Y.

Ejercicio 7.5.7. Probar que las variables $X = U+V$ e $Y = U-V$ son incorreladas, pero no independientes, si U y V son variables aleatorias con función de densidad conjunta:

$$f_{U,V}(u,v) = \exp(-u-v), \quad u,v > 0.$$

Para estudiar si son incorreladas, calcularemos su covarianza. Previamente calculamos las marginales.

$$f_U(u) = \int_0^\infty f_{U,V}(u,v)\ dv = \int_0^\infty \exp(-u-v)\ dv = e^{-u} \int_0^\infty e^{-v}\ dv = e^{-u}\left[-e^{-v}\right]_0^\infty = e^{-u}$$
$$f_V(v) = \int_0^\infty f_{U,V}(u,v)\ du = \int_0^\infty \exp(-u-v)\ du = e^{-v} \int_0^\infty e^{-u}\ du = e^{-v}\left[-e^{-u}\right]_0^\infty = e^{-v}$$

Por tanto, tenemos que:

$$U,V \sim \exp(1) \Longrightarrow E[U^k] = E[V^k] = 1 \qquad \forall k \in \mathbb{N}$$

Por tanto, tenemos:

$$\begin{aligned}
E[X] &= E[U+V] = E[U] + E[V] = 1+1 = 2\\
E[Y] &= E[U-V] = E[U] - E[V] = 1-1 = 0\\
E[XY] &= E[(U+V)(U-V)] = E[U^2 - V^2] = E[U^2] - E[V^2] = 1-1 = 0
\end{aligned}$$

Tenemos por tanto que:

$$\operatorname{Cov}[X,Y] = E[XY] - E[X]E[Y] = 0 - 2\cdot 0 = 0$$

Por tanto, tenemos que X e Y son incorreladas.

Veamos que no son independientes. Para ello, calculamos en primer lugar $f_{X,Y}(x,y)$. Para ello, definimos la transformación:

$$\begin{array}{rrcl} g: & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & (U,V) & \longmapsto & (X,Y) = (U+V, U-V) \end{array}$$

Para obtener g^{-1}, buscamos obtener U,V en función de X,T:

$$\left\{\begin{array}{l} X = U+V \\ Y = U-V \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} U = {}^{X+Y}\!/_2 \\ V = {}^{X-Y}\!/_2 \end{array}\right\}$$

Por tanto, tenemos que $\exists g^{-1}$, con:

$$\begin{array}{rrcl} g^{-1}: & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & (X,Y) & \longmapsto & (U,V) = \left(\frac{X+Y}{2}, \frac{X-Y}{2}\right) \end{array}$$

Tenemos que todas las componentes de g^{-1} son derivables:

$$\frac{\partial U}{\partial X} = \frac{1}{2}, \qquad \frac{\partial U}{\partial Y} = \frac{1}{2}, \qquad \frac{\partial V}{\partial X} = \frac{1}{2}, \qquad \frac{\partial V}{\partial Y} = -\frac{1}{2}$$

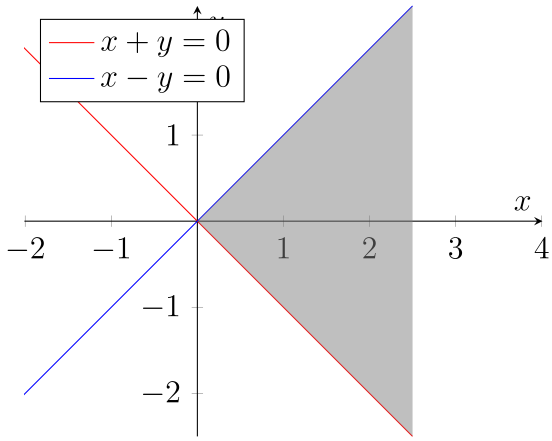
Además, veamos que el jacobianos no se anula:

$$\det Jg^{-1}(x,y) = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = \frac{1}{2^2} \cdot \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -\frac{1}{2} \neq 0 \qquad \forall (x,y) \in \mathbb{R}^2.$$

Por tanto, podemos aplicar el teorema del cambio de variable para obtener $f_{X,Y}(x,y)$:

$$\begin{aligned} f_{X,Y}(x,y) &= f_{U,V}(g^{-1}(x,y)) \cdot |Jg^{-1}(x,y)| = \frac{1}{2} \cdot f_{U,V}\left(\frac{x+y}{2}, \frac{x-y}{2}\right) \\ &= \frac{1}{2} \cdot \exp\left(-\frac{x+y}{2} - \frac{x-y}{2}\right) = \frac{1}{2} \cdot \exp\left(\frac{-x-y-x+y}{2}\right) = \\ &= \frac{1}{2} \cdot \exp(-x) \qquad \forall (x,y) \in \mathbb{R}^2 \text{ tal que } x+y, x-y > 0 \end{aligned}$$

Veamos el conjunto gráficamente:



Tenemos que el conjunto donde $f_{X,Y}(x,y)$ no se puede expresar como producto cartesiano, por lo que intuimos que X e Y no son independientes. Para comprobarlo, calculamos la función de densidad marginal de X y Y:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = \int_{-x}^{x} 1/2 \exp(-x)\, dy = 1/2 \exp(-x)\, [y]_{-x}^{x} = \\ &= 1/2 \exp(-x)(x+x) = xe^{-x} \qquad \forall x > 0 \end{aligned}$$

Para calcular la marginal de Y, distinguimos en función de y:

- Si $y > 0$:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx = \int_{y}^{\infty} 1/2 \exp(-x)\, dx = 1/2\, [-\exp(-x)]_y^{\infty} = \\ &= 1/2\, [0 - (-\exp(-y))] = \frac{e^{-y}}{2} \qquad \forall y > 0 \end{aligned}$$

- Si $y \leqslant 0$:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx = \int_{-y}^{\infty} 1/2 \exp(-x)\ dx = 1/2\left[-\exp(-x)\right]_{-y}^{\infty} =$$
$$= 1/2\left[0 - (-\exp(y))\right] = \frac{e^y}{2} \qquad \forall y \leqslant 0$$

Por tanto, tenemos que:

$$f_Y(y) = \begin{cases} \dfrac{e^{-y}}{2} & y > 0 \\ \\ \dfrac{e^{y}}{2} & y \leqslant 0 \end{cases}$$

Para el $(1,1)$, tenemos que:

$$f_X(1) = e^{-1}, \qquad f_{X,Y}(1,1) = \frac{1}{2}e^{-1}$$
$$f_Y(1) = \frac{1}{2}e^{-1}, \qquad f_X(1)f_Y(1) = \frac{1}{2}e^{-2} \neq f_{X,Y}(1,1)$$

Por tanto, no se cumple que $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ para todo $(x,y) \in \mathbb{R}^2$, por lo que X e Y no son independientes.

Ejercicio 7.5.8. Sea X una variable aleatoria con distribución uniforme en el intervalo $[0,1]$, y sea Y una variable aleatoria continua tal que

$$f_{Y|X=x}(y) = \begin{cases} 1/x^2 & y \in [0, x^2] \\ 0 & \text{en caso contrario} \end{cases}$$

1. Calcular la función de densidad de probabilidad conjunta de X e Y. Calcular la función de densidad de probabilidad marginal de Y.

   Para calcular $f_{X,Y}(x,y)$, usamos el siguiente resultado:

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

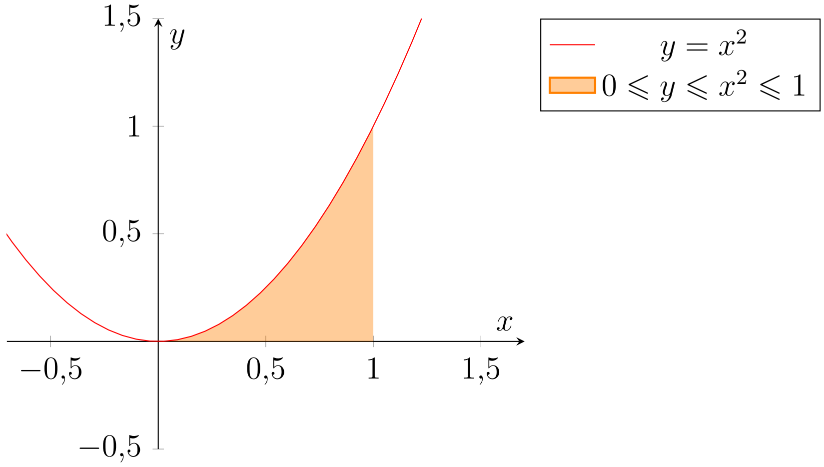
   Por tanto, calculamos la marginal de X. Como esta es uniforme en $[0,1]$, tenemos que:

$$1 = \int_0^1 f_X(x)\ dx = \int_0^1 k\ dx = k \Longrightarrow f_X(x) = 1 \qquad \forall x \in [0,1]$$

   Por tanto, tenemos que:

$$f_{X,Y}(x,y) = f_{Y|X=x}(y) \cdot f_X(x) = \begin{cases} 1/x^2 & 0 \leqslant y \leqslant x^2,\ 0 < x \leqslant 1 \\ 0 & \text{en caso contrario} \end{cases}$$

   Veamos gráficamente el conjunto donde $f_{X,Y}(x,y)$ no se anula:



Para calcular la marginal de Y, para cada $y \in]0,1]$, tenemos que:

$$f_Y(y) = \int_{\sqrt{y}}^{1} f_{X,Y}(x,y)\ dx = \int_{\sqrt{y}}^{1} {}^1/_{x^2}\ dx = \left[-\frac{1}{x}\right]_{\sqrt{y}}^{1} = -1 + \frac{1}{\sqrt{y}}$$

Por tanto, tenemos que:

$$f_Y(y) = \begin{cases} -1 + {}^1/_{\sqrt{y}} & 0 < y \leqslant 1 \\ 0 & \text{en caso contrario} \end{cases}$$

Observación. En este ejercicio, el lector podrá ver que es difícil comprender correctamente cuáles son los intervalos en los que se trabaja. Esto se debe a que no se especifica en el enunciado a qué valor de x estás condicionando, ya que $x = 0$ no es válido.

2. Calcular $E[X \mid Y = y]$ y $E[Y \mid X = x]$.

   Calculamos en primer lugar la distribución condicionada de X dado un valor $Y = y \in [0,1]$.

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{{}^1/_{x^2}}{-1 + {}^1/_{\sqrt{y}}} \qquad \forall x \in [0,1] \mid \sqrt{y} \leqslant x$$

   Calculamos ahora por tanto las esperanzas condicionadas:

$$\begin{aligned} E[X \mid Y = y] &= \int_{\sqrt{y}}^{1} x \cdot f_{X|Y=y}(x)\ dx = \int_{\sqrt{y}}^{1} x \cdot \frac{{}^1/_{x^2}}{-1 + {}^1/_{\sqrt{y}}}\ dx = \frac{1}{-1 + {}^1/_{\sqrt{y}}} \int_{\sqrt{y}}^{1} {}^1/_x\ dx = \\ &= \frac{1}{-1 + {}^1/_{\sqrt{y}}} [\ln(x)]_{\sqrt{y}}^{1} = \frac{1}{-1 + {}^1/_{\sqrt{y}}} [\ln(1) - \ln(\sqrt{y})] = \\ &= -\frac{\ln(\sqrt{y})}{-1 + {}^1/_{\sqrt{y}}} \qquad \forall y \in [0,1] \end{aligned}$$

$$\begin{aligned} E[Y \mid X = x] &= \int_0^{x^2} y \cdot f_{Y|X=x}(y)\ dy = \int_0^{x^2} y \cdot \frac{1}{x^2}\ dy = \frac{1}{x^2} \int_0^{x^2} y\ dy = \\ &= \frac{1}{x^2} \left[\frac{y^2}{2}\right]_0^{x^2} = \frac{1}{x^2} \cdot \frac{x^4}{2} = \frac{x^2}{2} \qquad \forall x \in [0,1] \end{aligned}$$

3. Para la misma densidad de probabilidad condicionada del apartado 1, considerando ahora que X es una variable aleatoria continua con función de densidad de probabilidad:
$$f_X(x) = \begin{cases} 3x^2 & x \in [0,1] \\ 0 & \text{en caso contrario} \end{cases}$$
Calcular de nuevo la función de densidad de probabilidad conjunta de X e Y, y la función de densidad de probabilidad marginal de Y, así como $E[X \mid Y = y]$ y $E[Y \mid X = x]$.

   Repetimos todo el proceso anterior, pero con la nueva función de densidad de X.
$$f_{X,Y}(x,y) = f_{Y|X=x}(y) \cdot f_X(x) = \begin{cases} 3 & 0 \leqslant y \leqslant x^2 \leqslant 1 \\ 0 & \text{en caso contrario} \end{cases}$$
   Para calcular la marginal de Y, para cada $y \in [0,1]$, tenemos que:
$$f_Y(y) = \int_{\sqrt{y}}^1 f_{X,Y}(x,y)\ dx = 3\int_{\sqrt{y}}^1 1\ dx = 3\,[x]_{\sqrt{y}}^1 = 3 - 3\sqrt{y} \qquad \forall y \in [0,1]$$
   Calculamos ahora la distribución condicionada de X dado un valor $Y = y \in [0,1]$.
$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{3}{3-3\sqrt{y}} = \frac{1}{1-\sqrt{y}} \qquad \forall x \in [0,1] \mid \sqrt{y} \leqslant x$$
   Calculamos ahora las esperanzas condicionadas:
$$\begin{aligned} E[X \mid Y = y] &= \int_{\sqrt{y}}^1 x \cdot f_{X|Y=y}(x)\ dx = \int_{\sqrt{y}}^1 x \cdot \frac{1}{1-\sqrt{y}}\ dx = \frac{1}{1-\sqrt{y}}\int_{\sqrt{y}}^1 x\ dx = \\ &= \frac{1}{1-\sqrt{y}}\left[\frac{x^2}{2}\right]_{\sqrt{y}}^1 = \frac{1}{1-\sqrt{y}}\left[\frac{1}{2} - \frac{y}{2}\right] = \frac{1}{2(1-\sqrt{y})}(1-y) = \\ &= \frac{1+\sqrt{y}}{2} \qquad \forall y \in [0,1] \\ E[Y \mid X = x] &= \int_0^{x^2} y \cdot f_{Y|X=x}(y)\ dy = \int_0^{x^2} y \cdot \frac{1}{x^2}\ dy = \frac{1}{x^2}\int_0^{x^2} y\ dy = \\ &= \frac{1}{x^2}\left[\frac{y^2}{2}\right]_0^{x^2} = \frac{1}{x^2} \cdot \frac{x^4}{2} = \frac{x^2}{2} \qquad \forall x \in [0,1] \end{aligned}$$

Ejercicio 7.5.9. Sean X e Y variables aleatorias con función de densidad conjunta:
$$f_{(X,Y)}(x,y) = \begin{cases} x+y & (x,y) \in [0,1] \times [0,1] \\ 0 & \text{en caso contrario} \end{cases}$$

1. Calcular la predicción mínimo cuadrática de Y a partir de X y el error cuadrático medio asociado.

Para calcular la predicción mínimo cuadrática de Y a partir de X, hemos de calcular $E[Y \mid X]$. Para ello, hemos de calcular la distribución marginal de X y la distribución condicionada de Y dado $X = x \in [0, 1]$. condicionada de Y dado $X = x \in [0, 1]$.

$$f_X(x) = \int_0^1 f_{(X,Y)}(x, y)\, dy = \int_0^1 x + y\, dy = \left[xy + \frac{y^2}{2}\right]_0^1 = x + \frac{1}{2} \qquad \forall x \in [0, 1]$$

$$f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x, y)}{f_X(x)} = \frac{x + y}{x + {}^1\!/\!_2} \qquad \forall y \in [0, 1]$$

Calculamos ahora la esperanza condicionada de Y dado $X = x$:

$$\begin{aligned}
E[Y \mid X = x] &= \int_0^1 y \cdot f_{Y|X=x}(y)\, dy = \int_0^1 y \cdot \frac{x + y}{x + {}^1\!/\!_2}\, dy = \frac{1}{x + {}^1\!/\!_2} \int_0^1 y \cdot (x + y)\, dy = \\
&= \frac{1}{x + {}^1\!/\!_2} \int_0^1 xy + y^2\, dy = \frac{1}{x + {}^1\!/\!_2} \left[\frac{xy^2}{2} + \frac{y^3}{3}\right]_0^1 = \frac{1}{x + {}^1\!/\!_2} \left(\frac{x}{2} + \frac{1}{3}\right) = \\
&= \frac{2}{2x + 1} \left(\frac{3x + 2}{6}\right) = \frac{3x + 2}{6x + 3} \qquad \forall x \in [0, 1]
\end{aligned}$$

Calculamos su error cuadrático medio:

$$\text{ECM}(E[Y \mid X]) = E[\text{Var}[Y \mid X]]$$

Calculamos por tanto en primer lugar $\text{Var}[Y \mid X] = E[Y^2 \mid X] - E[Y \mid X]^2$:

$$\begin{aligned}
E[Y^2 \mid X = x] &= \int_0^1 y^2 \cdot f_{Y|X=x}(y)\, dy = \int_0^1 y^2 \cdot \frac{x + y}{x + {}^1\!/\!_2}\, dy = \frac{1}{x + {}^1\!/\!_2} \int_0^1 y^2 \cdot (x + y)\, dy = \\
&= \frac{1}{x + {}^1\!/\!_2} \int_0^1 xy^2 + y^3\, dy = \frac{1}{x + {}^1\!/\!_2} \left[\frac{xy^3}{3} + \frac{y^4}{4}\right]_0^1 = \frac{1}{x + {}^1\!/\!_2} \left(\frac{x}{3} + \frac{1}{4}\right) = \\
&= \frac{4x + 3}{12x + 6} \qquad \forall x \in [0, 1]
\end{aligned}$$

$$\begin{aligned}
\text{Var}[Y \mid X = x] &= E[Y^2 \mid X = x] - E[Y \mid X = x]^2 = \frac{4x + 3}{12x + 6} - \left(\frac{3x + 2}{6x + 3}\right)^2 = \\
&= \frac{4x + 3}{6(2x + 1)} - \frac{(3x + 2)^2}{9(2x + 1)^2} = \frac{3(4x + 3)(2x + 1) - 2(3x + 2)^2}{18(2x + 1)^2} = \\
&= \frac{3(8x^2 + 10x + 3) - 2(9x^2 + 4 + 12x)}{18(2x + 1)^2} = \\
&= \frac{24x^2 + 30x + 9 - 18x^2 - 8 - 24x}{18(2x + 1)^2} = \\
&= \frac{6x^2 + 6x + 1}{18(2x + 1)^2} \qquad \forall x \in [0, 1]
\end{aligned}$$

Por tanto, tenemos que:

$$\begin{aligned}
\text{ECM}(E[Y \mid X]) = E[\text{Var}[Y \mid X]] &= \int_0^1 \frac{6x^2 + 6x + 1}{18(2x + 1)^2} \cdot f_X(x)\, dx = \\
&= \int_0^1 \frac{6x^2 + 6x + 1}{18(2x + 1)^2} \cdot (x + {}^1\!/\!_2)\, dx = \int_0^1 \frac{6x^2 + 6x + 1}{36(2x + 1)}\, dx
\end{aligned}$$

Para resolver dicha integral, realizamos la división de polinomios:

$$\begin{array}{rl} \big(\quad 6x^2 + 6x + 1\big) : \big(2x+1\big) = 3x + \frac{3}{2} + \dfrac{-\frac{1}{2}}{2x+1} \\ -\,6x^2 - 3x \qquad\qquad & \\ \hline 3x + 1 & \\ -\,3x - \frac{3}{2} & \\ \hline -\,\frac{1}{2} & \end{array}$$

Por tanto, tenemos que:

$$\begin{aligned} \mathrm{ECM}(E[Y \mid X]) &= \frac{1}{36}\int_0^1 3x + \frac{3}{2} - \frac{1/2}{2x+1}\,dx = \\ &= \frac{1}{36}\left[\frac{3x^2}{2} + \frac{3}{2}x - \frac{1}{4}\ln(2x+1)\right]_0^1 = \frac{1}{36}\left(\frac{3}{2} + \frac{3}{2} - \frac{1}{4}\ln(3)\right) = \\ &= \frac{1}{36}\left(3 - \frac{1}{4}\ln(3)\right) \approx 0{,}0757 \end{aligned}$$

2. Si se observa $X = 1/2$, ¿qué predicción de Y tiene menor error cuadrático medio? Calcular dicho error.

   En este caso, tenemos que la predicción de Y a partir de $X = 1/2$ es:

$$E[Y \mid X = 1/2] = \frac{3 \cdot 1/2 + 2}{6 \cdot 1/2 + 3} = \frac{7}{12}$$

   Calculamos su error cuadrático medio. Para esto, hay dos opciones:

   Usando el apartado anterior En este caso, tenemos que:

$$\begin{aligned} \mathrm{ECM}(E[Y \mid X = 1/2]) &= E[\mathrm{Var}[Y \mid X = 1/2]] = \\ &= E\left[\frac{6 \cdot 1/4 + 6 \cdot 1/2 + 1}{18(2 \cdot 1/2 + 1)^2}\right] = E\left[\frac{11}{144}\right] = \frac{11}{144} \approx 0{,}07639 \end{aligned}$$

   Usando la definición de ECM En este caso, tenemos que:

$$\begin{aligned} \mathrm{ECM}(E[Y \mid X = 1/2]) &= E[(Y - E[Y \mid X = 1/2])^2] = E\left[\left(Y - \frac{7}{12}\right)^2\right] = \\ &= E[Y^2] - \frac{7}{6}E[Y] + \frac{49}{144} \end{aligned}$$

   Para calcular las esperanzas, calculamos en primer lugar la marginal de Y. Para $y \in [0,1]$:

$$f_Y(y) = \int_0^1 f_{(X,Y)}(x,y)\,dx = \int_0^1 x + y\,dx = \left[\frac{x^2}{2} + xy\right]_0^1 = \frac{1}{2} + y \qquad \forall y \in [0,1]$$

Calculamos por tanto las esperanzas necesarias:

$$\begin{aligned}
E[Y] &= \int_0^1 y \cdot f_Y(y)\ dy = \int_0^1 y \cdot \left(\frac{1}{2} + y\right)\ dy = \int_0^1 \frac{y}{2} + y^2\ dy = \\
&= \left[\frac{y^2}{4} + \frac{y^3}{3}\right]_0^1 = \frac{1}{4} + \frac{1}{3} = \frac{7}{12} \\
E[Y^2] &= \int_0^1 y^2 \cdot f_Y(y)\ dy = \int_0^1 y^2 \cdot \left(\frac{1}{2} + y\right)\ dy = \int_0^1 \frac{y^2}{2} + y^3\ dy = \\
&= \left[\frac{y^3}{6} + \frac{y^4}{4}\right]_0^1 = \frac{1}{6} + \frac{1}{4} = \frac{5}{12}
\end{aligned}$$

Por tanto, tenemos que:

$$\mathrm{ECM}(E[Y \mid X = 1/2]) = E[Y^2] - \frac{7}{6}E[Y] + \frac{49}{144} = \frac{5}{12} - \frac{7}{6} \cdot \frac{7}{12} + \frac{49}{144} = \frac{11}{144}$$

3. Supóngase que una persona debe pagar una cantidad C por la oportunidad de observar el valor de X antes de predecir el valor de Y, o puede simplemente predecir el valor de Y sin observar X. Si la persona considera que su pérdida total es la suma de C y el error cuadrático medio de su predicción, qué valor máximo de C estaría dispuesta a pagar?

   Estudiemos cada una de las opciones por separado.

   Observar X antes de predecir Y: En este caso, la pérdida total es:

$$C + \mathrm{ECM}(E[Y \mid X]) = C + \frac{1}{36}\left(\frac{3}{2} + \frac{1}{8}\ln(3)\right)$$

   donde hemos hecho uso del resultado calculado en el primer apartado.

   Predecir Y sin observar X: En este caso, la pérdida total es (ya que no pagamos C):

$$\mathrm{ECM}(E[Y]) = E[(Y - E[Y])^2] = \mathrm{Var}[Y] = E[Y^2] - E[Y]^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}$$

   Por tanto, la persona elegirá la primera opción siempre que:

$$C + \mathrm{ECM}(E[Y \mid X]) \leqslant \mathrm{ECM}(E[Y]) \Longleftrightarrow C \leqslant \mathrm{ECM}(E[Y]) - \mathrm{ECM}(E[Y \mid X])$$

   Por tanto, la persona estará dispuesta a pagar un máximo de:

$$C = \frac{11}{144} - \frac{1}{36}\left(\frac{3}{2} + \frac{1}{8}\ln(3)\right) \approx 0{,}03090$$

Ejercicio 7.5.10. Sea (X, Y) un vector aleatorio con función de densidad:

$$f_{(X,Y)}(x, y) = e^{-y}, \quad 0 < x < y$$

Obtener y representar las rectas y curvas de regresión. Calcular el coeficiente de correlación lineal, las razones de correlación y el error cuadrático medio cometido al predecir cada variable según cada una de las funciones de regresión. Interpretar los resultados.

Obtenemos en primer lugar las marginales:

$$f_X(x) = \int_x^\infty e^{-y}\,dy = \left[-e^{-y}\right]_x^\infty = e^{-x} \qquad \forall x > 0$$
$$f_Y(y) = \int_0^y e^{-y}\,dx = ye^{-y} \qquad \forall y > 0$$

Obtenemos ahora las condicionadas. Dados $x^*, y^* \in \mathbb{R}^+$, tenemos que:

$$f_{Y|X=x^*}(y) = \frac{f_{(X,Y)}(x^*, y)}{f_X(x^*)} = \frac{e^{-y}}{e^{-x^*}} = e^{x^*-y} \qquad \forall y > x^*$$
$$f_{X|Y=y^*}(x) = \frac{f_{(X,Y)}(x, y^*)}{f_Y(y^*)} = \frac{e^{-y^*}}{y^*e^{-y^*}} = \frac{1}{y^*} \qquad \forall x \in]0, y^*[$$

Calculamos ahora las curvas de regresión. Para la de Y sobre X, tenemos que:

$$\widehat{Y}(x) = E[Y \mid X = x] = \int_x^\infty y \cdot e^{x-y}\,dy = e^x \int_x^\infty y \cdot e^{-y}\,dy$$

Para calcular la integral, realizamos integración por partes:

$$\left[\begin{array}{cc} u(y) = y & u'(y) = 1 \\ v'(y) = e^{-y} & v(y) = -e^{-y} \end{array}\right] \Longrightarrow \int_x^\infty y \cdot e^{-y}\,dy = \left[-ye^{-y}\right]_x^\infty - \int_x^\infty -e^{-y}\,dy =$$
$$= \left[-ye^{-y} - e^{-y}\right]_x^\infty = e^{-x}(x+1)$$

Por tanto, tenemos que la curva de regresión de Y sobre X es:

$$\widehat{Y}(x) = e^x e^{-x}(x+1) = x+1 \qquad \forall x > 0$$

Como vemos, la curva de regresión de Y sobre X es $y = x+1$, que es una recta y por tanto coincide con la recta de regresión.

Calculamos ahora la curva de regresión de X sobre Y. Para ello, tenemos que:

$$\widehat{X}(y) = E[X \mid Y = y] = \int_0^y x \cdot \frac{1}{y}\,dx = \frac{1}{y}\int_0^y x\,dx = \frac{1}{y}\left[\frac{x^2}{2}\right]_0^y = \frac{y}{2} \qquad \forall y > 0$$

Como vemos, la curva de regresión de X sobre Y es $x = y/2$, que es una recta y por tanto coincide con la recta de regresión.

Calculamos ahora el coeficiente de determinación lineal. Para ello, calculamos las varianzas y covarianzas:

$$\begin{aligned}
E[X] &= \int_0^\infty x \cdot e^{-x}\,dx = \left[-xe^{-x} - e^{-x}\right]_0^\infty = 1\\
E[Y] &= \int_0^\infty y^2 \cdot e^{-y}\,dy = \left[-y^2e^{-y}\right]_0^\infty + 2\int_0^\infty y \cdot e^{-y}\,dy = 0 + 2\cdot 1 = 2\\
E[X^2] &= \int_0^\infty x^2 \cdot e^{-x}\,dx = 2\\
E[Y^2] &= \int_0^\infty y^3 \cdot e^{-y}\,dy = \left[-y^3e^{-y}\right]_0^\infty + 3\int_0^\infty y^2 \cdot e^{-y}\,dy = 0 + 3\cdot 2 = 6\\
E[XY] &= \int_0^{+\infty}\int_0^y xyf_{(X,Y)}(x,y)\,dx\,dy = \int_0^{+\infty}\int_0^y xye^{-y}\,dx\,dy =\\
&= \int_0^{+\infty} ye^{-y}\cdot\frac{y^2}{2}\,dy = \frac{1}{2}\int_0^{+\infty} y^3e^{-y}\,dy = \frac{1}{2}\cdot 6 = 3\\
\operatorname{Var}[X] &= E[X^2] - E[X]^2 = 2 - 1 = 1\\
\operatorname{Var}[Y] &= E[Y^2] - E[Y]^2 = 6 - 4 = 2\\
\operatorname{Cov}[X,Y] &= E[XY] - E[X]E[Y] = 3 - 2\cdot 1 = 1
\end{aligned}$$

Por tanto, podemos calcular el coeficiente de determinación lineal. Además, como las rectas de regresión coinciden con las curvas de regresión, el coeficiente de determinación lineal coincide con las razones de correlación:

$$\eta^2_{X/Y} = \eta^2_{Y/X} = \rho^2_{X,Y} = \frac{\operatorname{Cov}[X,Y]^2}{\operatorname{Var}[X]\operatorname{Var}[Y]} = \frac{1^2}{1\cdot 2} = \frac{1}{2}$$

Como podemos ver, tenemos que la mitad de la varianza de X y Y se explica por la regresión, lo que nos indica que los ajustes no son muy buenos.

Calculamos ahora el coeficiente de correlación lineal. Como la covarianza es positiva, el coeficiente de correlación lineal también lo será:

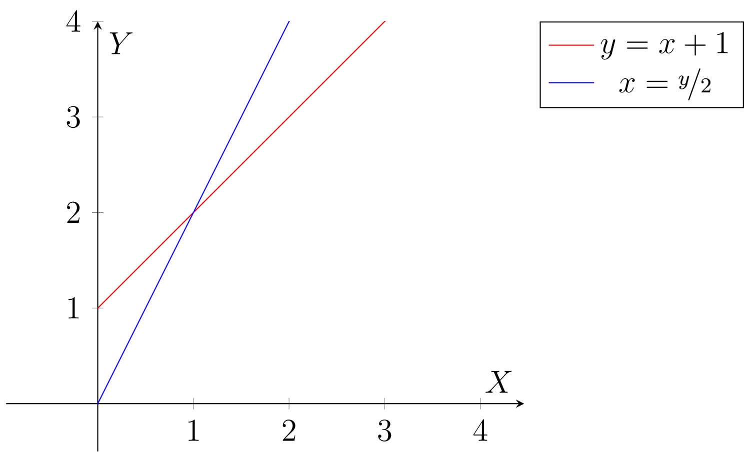
$$\rho_{X,Y} = +\sqrt{\rho^2_{X,Y}} = +\sqrt{\frac{1}{2}} = +\frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$$

Como podemos ver, el coeficiente de correlación lineal es positivo, lo que nos indica que las variables están correlacionadas positivamente.

Calculamos ahora el error cuadrático medio cometido al predecir cada variable según cada una de las funciones de regresión.

$$\begin{aligned}
\operatorname{ECM}(\widehat{Y}(X)) &= E[(Y-\widehat{Y}(X))^2] = E[(Y-X-1)^2] = E[Y^2+X^2+1-2Y+2X-2YX] =\\
&= E[Y^2]+E[X^2]+1-2E[Y]+2E[X]-2E[XY] = 6+2+1-4+2-6 = 1\\
\operatorname{ECM}(\widehat{X}(Y)) &= E[(X-\widehat{X}(Y))^2] = E[(X-{}^{Y}\!/_{2})^2] = E[X^2+{}^{Y^2}\!/_{4}-YX] =\\
&= E[X^2]+\frac{1}{4}E[Y^2]-E[XY] = 2+\frac{1}{4}\cdot 6-3 = \frac{1}{2}
\end{aligned}$$

Por último, representamos las rectas de regresión:



Ejercicio 7.5.11. Sea (X, Y) un vector aleatorio con función de densidad uniforme sobre el cuadrado unidad. Obtener y representar las rectas y curvas de regresión. Calcular el coeficiente de correlación lineal, las razones de correlación y el error cuadrático medio cometido al predecir cada variable según cada una de las funciones de regresión. Interpretar los resultados.

Dado que la función de densidad es uniforme en el cuadrado unidad, tenemos que (compruébese):

$$f_{(X,Y)}(x, y) = \begin{cases} 1, & (x, y) \in [0, 1] \times [0, 1] \\ 0, & \text{en caso contrario} \end{cases}$$

Definimos la siguiente función auxiliar:

$$\begin{aligned} g: \mathbb{R} &\longrightarrow \mathbb{R} \\ t &\longmapsto \begin{cases} 1, & t \in [0, 1] \\ 0, & \text{en caso contrario} \end{cases} \end{aligned}$$

Tenemos que $f_{(X,Y)}(x, y) = g(x)g(y)$ para todo $(x, y) \in \mathbb{R}^2$, por lo que X e Y son independientes. Por tanto, tenemos que las curvas de regresión son:

$$\widehat{Y}(x) = E[Y \mid X = x] = E[Y] = \int_0^1 y\, dy = \left[\frac{y^2}{2}\right]_0^1 = \frac{1}{2} \qquad \forall x \in [0, 1]$$

$$\widehat{X}(y) = E[X \mid Y = y] = E[X] = \int_0^1 x\, dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2} \qquad \forall y \in [0, 1]$$

Al ser independientes, en particular son incorreladas, por lo que:

$$\rho^2_{X,Y} = \eta^2_{X/Y} = \eta^2_{Y/X} = 0$$

Calculamos ahora el error cuadrático medio cometido al predecir cada variable según cada una de las funciones de regresión.

$$\mathrm{ECM}(\widehat{Y}(X)) = E[(Y - \widehat{Y}(X))^2] = E[(Y - E[Y])^2] = \mathrm{Var}[Y]$$

$$\mathrm{ECM}(\widehat{X}(Y)) = E[(X - \widehat{X}(Y))^2] = E[(X - E[X])^2] = \mathrm{Var}[X]$$

Calculamos por tanto ambas varianzas:

$$\begin{aligned}
E[X^2] &= \int_0^1 x^2\ dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3} \\
E[Y^2] &= \int_0^1 y^2\ dy = \left[\frac{y^3}{3}\right]_0^1 = \frac{1}{3} \\
\mathrm{Var}[X] &= E[X^2] - E[X]^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12} \\
\mathrm{Var}[Y] &= E[Y^2] - E[Y]^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}
\end{aligned}$$

Por tanto, tenemos que:

$$\begin{aligned}
\mathrm{ECM}(\widehat{Y}(X)) &= \mathrm{Var}[Y] = \frac{1}{12} \\
\mathrm{ECM}(\widehat{X}(Y)) &= \mathrm{Var}[X] = \frac{1}{12}
\end{aligned}$$

Ejercicio 7.5.12. Supongamos que (X, Y) tiene función de densidad de probabilidad conjunta dada por:

$$f_{(X,Y)}(x, y) = \begin{cases} 1, & |y| < x, x \in]0, 1[\\ 0, & \text{en otro caso} \end{cases}$$

Obtener y representar las rectas y curvas de regresión. Calcular el coeficiente de correlación lineal, las razones de correlación y el error cuadrático medio cometido al predecir cada variable según cada una de las funciones de regresión. Interpretar los resultados.

Calculamos las marginales:

$$\begin{aligned}
f_X(x) &= \int_{-x}^{x} 1\ dy = 2x \qquad \forall x \in]0, 1[\\
f_Y(y) &= \int_{|y|}^{1} 1\ dx = 1 - |y| \qquad \forall y \in]-1, 1[
\end{aligned}$$

Calculamos ahora las condicionadas. Dados $x^* \in]0, 1[$ e $y^* \in]-1, 1[$, tenemos que:

$$\begin{aligned}
f_{Y|X=x^*}(y) &= \frac{f_{(X,Y)}(x^*, y)}{f_X(x^*)} = \frac{1}{2x^*} \qquad \forall y \in]-x^*, x^*[\\
f_{X|Y=y^*}(x) &= \frac{f_{(X,Y)}(x, y^*)}{f_Y(y^*)} = \frac{1}{1 - |y^*|} \qquad \forall x \in]|y^*|, 1[
\end{aligned}$$

Calculamos ahora las curvas de regresión. Tenemos que:

$$\widehat{Y}(x) = E[Y \mid X = x] = \int_{-x}^{x} y \cdot \frac{1}{2x}\ dy = \frac{1}{2x} \int_{-x}^{x} y\ dy = \frac{1}{2x}\left(\frac{x^2}{2} - \frac{x^2}{2}\right) = 0 \qquad \forall x \in]0, 1[$$

$$\begin{aligned}
\widehat{X}(y) = E[X \mid Y = y] &= \int_{|y|}^{1} x \cdot \frac{1}{1 - |y|}\ dx = \frac{1}{1 - |y|} \int_{|y|}^{1} x\ dx = \frac{1}{1 - |y|}\left(\frac{1}{2} - \frac{y^2}{2}\right) = \\
&= \frac{1 - y^2}{2(1 - |y|)} = \frac{(1 - y^2)(1 + |y|)}{2(1 - |y|)(1 + |y|)} = \frac{1 + |y|}{2} \qquad \forall y \in]-1, 1[
\end{aligned}$$

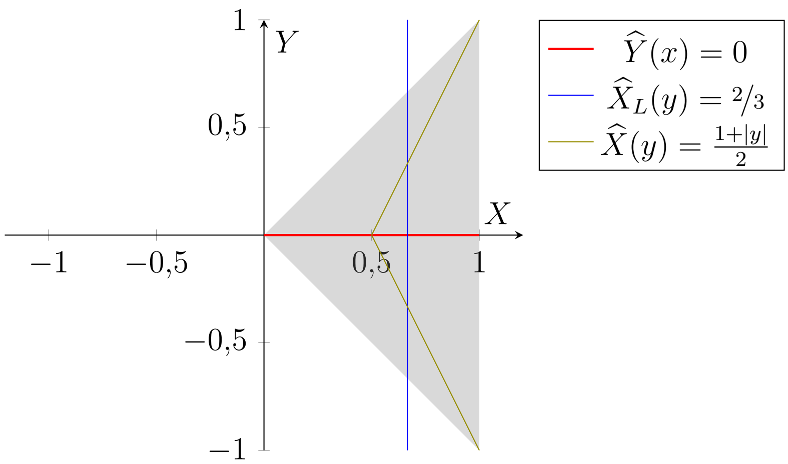
Como la curva de regresión de Y sobre X es constante, la recta de regresión de Y sobre X es la misma que la curva de regresión. Calculamos ahora la recta de regresión X sobre Y. Para ello, previamente calculamos:

$$\begin{aligned}
E[X] &= \int_0^1 x\cdot 2x\ dx = 2\int_0^1 x^2\ dx = 2\left[\frac{x^3}{3}\right]_0^1 = \frac{2}{3}\\
E[Y] &= \int_{-1}^1 y\cdot(1-|y|)\ dy = \int_{-1}^0 y(1+y)\ dy + \int_0^1 y(1-y)\ dy =\\
&= \left[\frac{y^2}{2}+\frac{y^3}{3}\right]_{-1}^0 + \left[\frac{y^2}{2}-\frac{y^3}{3}\right]_0^1 = -\frac{1}{2}+\frac{1}{3}+\frac{1}{2}-\frac{1}{3} = 0\\
E[X^2] &= \int_0^1 x^2\cdot 2x\ dx = 2\int_0^1 x^3\ dx = 2\left[\frac{x^4}{4}\right]_0^1 = \frac{1}{2}\\
E[Y^2] &= \int_{-1}^1 y^2\cdot(1-|y|)\ dy = \int_{-1}^0 y^2(1+y)\ dy + \int_0^1 y^2(1-y)\ dy =\\
&= \left[\frac{y^3}{3}+\frac{y^4}{4}\right]_{-1}^0 + \left[\frac{y^3}{3}-\frac{y^4}{4}\right]_0^1 = \frac{1}{3}-\frac{1}{4}+\frac{1}{3}-\frac{1}{4} = \frac{1}{6}\\
E[XY] &= \int_0^1\int_{-x}^x xy\ dy\ dx = \int_0^1 x\left[\frac{y^2}{2}\right]_{-x}^x\ dx = \int_0^1 x\left(\frac{x^2}{2}-\frac{x^2}{2}\right)\ dx = 0\\
\mathrm{Var}[X] &= E[X^2]-E[X]^2 = \frac{1}{2}-\left(\frac{2}{3}\right)^2 = \frac{1}{18}\\
\mathrm{Var}[Y] &= E[Y^2]-E[Y]^2 = \frac{1}{6}-0 = \frac{1}{6}\\
\mathrm{Cov}[X,Y] &= E[XY]-E[X]E[Y] = 0
\end{aligned}$$

Por tanto, tenemos que la recta de regresión de X sobre Y es:

$$\widehat{X}_L(y) = E[X] + \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[Y]}(y-E[Y]) = \frac{2}{3} \qquad \forall y\in\,]-1,1[$$

La representación gráfica de las rectas y curvas de regresión es:



El coeficiente de determinación lineal es:

$$\rho_{X,Y}^2 = \frac{\mathrm{Cov}[X,Y]^2}{\mathrm{Var}[X]\,\mathrm{Var}[Y]} = 0$$

Como vemos, el ajuste lineal es nulo, ya que no explica nada de la variabilidad de X ni de Y.

El coeficiente de correlación lineal es:

$$\rho_{X,Y} = 0$$

Como vemos, las variables son linealmente incorreladas, como era de esperar tras el cálculo del coeficiente de determinación lineal.

Calculamos las razones de correlación. Como la curva de regresión de Y sobre X coincide con la recta de regresión, tenemos que:

$$\eta^2_{Y/X} = \rho^2_{X,Y} = 0$$

Respecto de la curva de regresión de X sobre Y, tenemos que:

$$\eta^2_{X/Y} = \frac{\mathrm{Var}[\widehat{X}(Y)]}{\mathrm{Var}[X]} = \frac{\mathrm{Var}\left[\frac{1+|Y|}{2}\right]}{\mathrm{Var}[X]}$$

Calculemos la varianza del numerador:

$$\begin{aligned}
E[|Y|] &= \int_{-1}^{1} |y|\cdot(1-|y|)\ dy = \int_{-1}^{0} -y(1+y)\ dy + \int_{0}^{1} y(1-y)\ dy = \\
&= -\left[\frac{y^2}{2}+\frac{y^3}{3}\right]_{-1}^{0} + \left[\frac{y^2}{2}-\frac{y^3}{3}\right]_{0}^{1} = \frac{1}{2}-\frac{1}{3}+\frac{1}{2}-\frac{1}{3} = \frac{1}{3} \\
E\left[\frac{1+|Y|}{2}\right] &= \frac{1}{2}+\frac{1}{2}E[|Y|] = \frac{1}{2}+\frac{1}{2}\cdot\frac{1}{3} = \frac{1}{2}+\frac{1}{6} = \frac{2}{3} \\
E\left[\left(\frac{1+|Y|}{2}\right)^2\right] &= \frac{1}{4}\left[1+2E[|Y|]+E[Y^2]\right] = \frac{1}{4}\left[1+2\cdot\frac{1}{3}+\frac{1}{6}\right] = \frac{11}{24} \\
\mathrm{Var}\left[\frac{1+|Y|}{2}\right] &= E\left[\left(\frac{1+|Y|}{2}\right)^2\right] - E\left[\frac{1+|Y|}{2}\right]^2 = \frac{11}{24} - \left(\frac{2}{3}\right)^2 = \frac{1}{72}
\end{aligned}$$

Por tanto, tenemos que:

$$\eta^2_{X/Y} = \frac{\mathrm{Var}\left[\frac{1+|Y|}{2}\right]}{\mathrm{Var}[X]} = \frac{1/72}{1/18} = \frac{1}{4}$$

En este caso, tan solo se explica un cuarto de la variabilidad de X por la regresión, por lo que el ajuste sigue siendo malo.

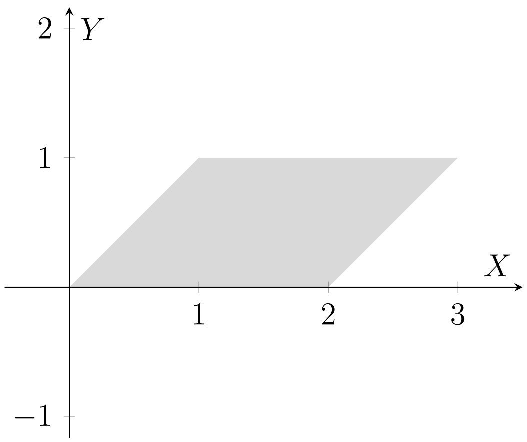
Calculamos ahora el error cuadrático medio de cada una de las funciones de regresión:

$$\begin{aligned}
\mathrm{ECM}(\widehat{Y}(X)) &= E[(Y-\widehat{Y}(X))^2] = E[Y^2] = \frac{1}{6}\\
\mathrm{ECM}(\widehat{X}(Y)) &= E[(X-\widehat{X}(Y))^2] = E\left[\left(X-\frac{1+|Y|}{2}\right)^2\right] =\\
&= E[X^2] + E\left[\left(\frac{1+|Y|}{2}\right)^2\right] - E[X(1+|Y|)] =\\
&= E[X^2] + E\left[\left(\frac{1+|Y|}{2}\right)^2\right] - E[X] - E[|Y|\cdot X] =\\
&= \frac{1}{2}+\frac{11}{24}-\frac{2}{3}-\frac{1}{4}=\frac{1}{24}\\
E[X\cdot|Y|] &= \int_0^1\int_{-x}^{x} x|y|\cdot 1\;dy\;dx = \int_0^1 2x\int_0^x y\;dy\;dx =\\
&= \int_0^1 2x\cdot\frac{x^2}{2}\;dx = \int_0^1 x^3\;dx = \left[\frac{x^4}{4}\right]_0^1 = \frac{1}{4}\\
\mathrm{ECM}(\widehat{X}_L(Y)) &= E\left[(X-E[X])^2\right] = \mathrm{Var}[X] = \frac{1}{18}
\end{aligned}$$

Aquí podemos ver que el error cuadrático medio cometido con la regresión lineal es mayor que el cometido con la curva de regresión, lo que indica que la regresión lineal es peor que la curva de regresión (como es de esperar).

Ejercicio 7.5.13. Sea (X,Y) un vector aleatorio distribuido uniformemente en el paralelogramo de vértices $(0,0)$; $(2,0)$; $(3,1)$ y $(1,1)$. Calcular el error cuadrático medio asociado a la predicción de X a partir de la variable Y y a la predicción de Y a partir de la variable aleatoria X. Determinar la predicción más fiable a la vista de los resultados obtenidos.

Veamos gráficamente el paralelogramo:



Notemos por P al paralelogramo. Tenemos que la función de densidad conjunta es:

$$f_{(X,Y)}(x,y) = \begin{cases} k, & (x,y)\in P\\ 0, & \text{en otro caso}\end{cases}$$

Para calcular k, tenemos que:

$$1 = \int_P f_{(X,Y)} = \int_P k = k\lambda(P) = k \cdot 2 \cdot 1 = 2k \Longrightarrow k = \frac{1}{2}$$

Calculamos en primer lugar las marginales. Para $x \in [0,3]$, tenemos que:

- Si $x \in [0,1]$, tenemos que:

$$f_X(x) = \int_0^x \frac{1}{2}\ dy = \frac{x}{2}$$

- Si $x \in [1,2]$, tenemos que:

$$f_X(x) = \int_0^1 \frac{1}{2}\ dy = \frac{1}{2}$$

- Si $x \in [2,3]$, tenemos que:

$$f_X(x) = \int_{x-2}^1 \frac{1}{2}\ dy = \frac{1}{2}(1-x+2) = \frac{3-x}{2}$$

Por tanto, tenemos que:

$$f_X(x) = \begin{cases} x/2, & x \in [0,1] \\ 1/2, & x \in [1,2] \\ \frac{3-x}{2}, & x \in [2,3] \\ 0, & \text{en otro caso} \end{cases}$$

Para $y \in [0,1]$, tenemos que:

$$f_Y(y) = \int_y^{y+2} \frac{1}{2}\ dx = \frac{1}{2}(y+2-y) = 1$$

Calculamos ahora las condicionadas. Dado $y^* \in [0,1]$, tenemos que:

$$f_{X|Y=y^*}(x) = \frac{f_{(X,Y)}(x,y^*)}{f_Y(y^*)} = \frac{1/2}{1} = \frac{1}{2} \qquad \forall x \in [0,3]$$

Respecto a la condicionada de Y dado $X = x^*$, distinguimos en función de x^*:

$$f_{Y|X=x^*}(y) = \frac{f_{(X,Y)}(x^*,y)}{f_X(x^*)} = \begin{cases} 1/x^*, & x^* \in [0,1],\ y \in [0,x^*] \\ 1, & x^* \in [1,2],\ y \in [0,1] \\ \frac{1}{3-x^*}, & x^* \in [2,3],\ y \in [x^*-2,1] \end{cases}$$

Tenemos que calcular los siguientes errores cuadráticos medios:

$$\mathrm{ECM}(\widehat{X}(Y)) = E[\mathrm{Var}[X \mid Y]]$$
$$\mathrm{ECM}(\widehat{Y}(X)) = E[\mathrm{Var}[Y \mid X]]$$

Para calcular $E[\mathrm{Var}[X \mid Y]]$, tenemos que:

$$\begin{aligned}
E[X \mid Y=y] &= \int_y^{y+2} x \cdot \frac{1}{2}\,dx = \frac{1}{2}\left[\frac{x^2}{2}\right]_y^{y+2} = \frac{1}{2}\left(\frac{(y+2)^2}{2} - \frac{y^2}{2}\right) = \frac{4y+4}{4} = y+1 \\
E[X^2 \mid Y=y] &= \int_y^{y+2} x^2 \cdot \frac{1}{2}\,dx = \frac{1}{2}\left[\frac{x^3}{3}\right]_y^{y+2} = \frac{1}{2}\left(\frac{(y+2)^3}{3} - \frac{y^3}{3}\right) = \frac{6y^2+12y+8}{6} = \\
&= y^2+2y+4/3 \\
\mathrm{Var}[X \mid Y=y] &= E[X^2 \mid Y=y] - E[X \mid Y=y]^2 = y^2+2y+\frac{4}{3} - (y+1)^2 = \\
&= y^2+2y+4/3-y^2-2y-1 = 1/3 \\
E[\mathrm{Var}[X \mid Y]] &= E\left[\frac{1}{3}\right] = \frac{1}{3}
\end{aligned}$$

Para calcular $E[\mathrm{Var}[Y \mid X]]$, calculamos previante lo necesario:

- Si $x \in [0,1]$, tenemos que:

$$\begin{aligned}
E[Y \mid X=x] &= \int_0^x y \cdot \frac{1}{x}\,dy = \frac{1}{x}\left[\frac{y^2}{2}\right]_0^x = \frac{x}{2} \\
E[Y^2 \mid X=x] &= \int_0^x y^2 \cdot \frac{1}{x}\,dy = \frac{1}{x}\left[\frac{y^3}{3}\right]_0^x = \frac{x^2}{3} \\
\mathrm{Var}[Y \mid X=x] &= E[Y^2 \mid X=x] - E[Y \mid X=x]^2 = \frac{x^2}{3} - \left(\frac{x}{2}\right)^2 = \frac{x^2}{12}
\end{aligned}$$

- Si $x \in [1,2]$, tenemos que:

$$\begin{aligned}
E[Y \mid X=x] &= \int_0^1 y \cdot 1\,dy = \left[\frac{y^2}{2}\right]_0^1 = \frac{1}{2} \\
E[Y^2 \mid X=x] &= \int_0^1 y^2 \cdot 1\,dy = \left[\frac{y^3}{3}\right]_0^1 = \frac{1}{3} \\
\mathrm{Var}[Y \mid X=x] &= E[Y^2 \mid X=x] - E[Y \mid X=x]^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}
\end{aligned}$$

- Si $x \in [2,3]$, tenemos que:

$$\begin{aligned}
E[Y \mid X=x] &= \int_{x-2}^1 y \cdot \frac{1}{3-x}\,dy = \frac{1}{3-x}\left[\frac{y^2}{2}\right]_{x-2}^1 = \frac{1}{3-x}\left(\frac{1}{2} - \frac{(x-2)^2}{2}\right) = \\
&= \frac{1}{3-x}\left(\frac{-x^2+4x-3}{2}\right) = \frac{-x^2+4x-3}{2(3-x)}
\end{aligned}$$

Realizamos la división:

$$\begin{array}{l}
\big(-x^2+4x-3\big) : \big(-x+3\big) = x-1 \\
\underline{\quad x^2-3x \quad} \\
\qquad\quad x-3 \\
\qquad \underline{-x+3} \\
\qquad\qquad 0
\end{array}$$

Por tanto, tenemos que:

$$E[Y \mid X = x] = \frac{(x-1)(3-x)}{2(3-x)} = \frac{x-1}{2}$$

$$E[Y^2 \mid X = x] = \int_{x-2}^{1} y^2 \cdot \frac{1}{3-x}\, dy = \frac{1}{3-x}\left[\frac{y^3}{3}\right]_{x-2}^{1} = \frac{1}{3-x}\left(\frac{1}{3} - \frac{(x-2)^3}{3}\right) =$$
$$= \frac{1}{3-x}\left(\frac{-x^3+6x^2-12x+9}{3}\right)$$

Realizamos la división:

$$\begin{array}{l}
\big(-x^3+6x^2-12x+9\big) : \big(-x+3\big) = x^2-3x+3 \\
\quad x^3 - 3x^2 \\
\hline
\qquad\quad 3x^2 - 12x \\
\qquad -3x^2 + 9x \\
\hline
\qquad\qquad -3x+9 \\
\qquad\qquad\ \ 3x-9 \\
\hline
\qquad\qquad\qquad 0
\end{array}$$

Por tanto, tenemos que:

$$E[Y \mid X = x] = \frac{x-1}{2}$$

$$E[Y^2 \mid X = x] = \frac{(x^2-3x+3)(3-x)}{3(3-x)} = \frac{x^2-3x+3}{3}$$

$$\mathrm{Var}[Y \mid X = x] = E[Y^2 \mid X = x] - E[Y \mid X = x]^2 = \frac{x^2-3x+3}{3} - \left(\frac{x-1}{2}\right)^2 =$$
$$= \frac{x^2-3x+3}{3} - \frac{x^2-2x+1}{4} = \frac{4x^2-12x+12-3x^2+6x-3}{12} =$$
$$= \frac{x^2-6x+9}{12}$$

Por tanto, tenemos que:

$$E[\mathrm{Var}[Y \mid X]] = \int_0^1 \frac{x^2}{12}\cdot\frac{x}{2}\, dx + \int_1^2 \frac{1}{12}\cdot\frac{1}{2}\, dx + \int_2^3 \frac{x^2-6x+9}{12}\cdot\frac{3-x}{2}\, dx =$$
$$= \int_0^1 \frac{x^3}{24}\, dx + \int_1^2 \frac{1}{24}\, dx - \int_2^3 \frac{(x-3)^3}{24}\, dx =$$
$$= \frac{1}{24}\left[\frac{x^4}{4}\right]_0^1 + \frac{1}{24}\,[x]_1^2 - \frac{1}{24}\left[\frac{(x-3)^4}{4}\right]_2^3 =$$
$$= \frac{1}{24}\cdot\frac{1}{4} + \frac{1}{24}\cdot 1 - \frac{1}{24}\cdot\left(-\frac{1}{4}\right) = \frac{1}{16}$$

En resumen, tenemos que:

$$\mathrm{ECM}(\widehat{X}(Y)) = \frac{1}{3}$$
$$\mathrm{ECM}(\widehat{Y}(X)) = \frac{1}{16}$$

Para comparar ambas regresiones, hemos de calcular las razones de correlación. Para ello, calculamos previamente las varianzas de X e Y:

$$\begin{aligned}
E[X] &= \int_0^1 x \cdot \frac{x}{2}\, dx + \int_1^2 x \cdot \frac{1}{2}\, dx + \int_2^3 x \cdot \frac{3-x}{2}\, dx = \\
&= \left[\frac{x^3}{6}\right]_0^1 + \left[\frac{x^2}{4}\right]_1^2 + \left[\frac{3x^2}{4} - \frac{x^3}{6}\right]_2^3 = \frac{1}{6} + \frac{4}{4} - \frac{1}{4} + \frac{27}{4} - \frac{27}{6} - \frac{12}{4} + \frac{8}{6} = \frac{3}{2} \\
E[X^2] &= \int_0^1 x^2 \cdot \frac{x}{2}\, dx + \int_1^2 x^2 \cdot \frac{1}{2}\, dx + \int_2^3 x^2 \cdot \frac{3-x}{2}\, dx = \\
&= \left[\frac{x^4}{8}\right]_0^1 + \left[\frac{x^3}{6}\right]_1^2 + \left[\frac{3x^3}{6} - \frac{x^4}{8}\right]_2^3 = \frac{1}{8} + \frac{8}{6} - \frac{1}{6} + \frac{81}{6} - \frac{81}{8} - \frac{24}{6} + \frac{16}{8} = \frac{8}{3} \\
\mathrm{Var}[X] &= E[X^2] - E[X]^2 = \frac{8}{3} - \left(\frac{3}{2}\right)^2 = \frac{5}{12} \\
E[Y] &= \int_0^1 y \cdot 1\, dy = \left[\frac{y^2}{2}\right]_0^1 = \frac{1}{2} \\
E[Y^2] &= \int_0^1 y^2 \cdot 1\, dy = \left[\frac{y^3}{3}\right]_0^1 = \frac{1}{3} \\
\mathrm{Var}[Y] &= E[Y^2] - E[Y]^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}
\end{aligned}$$

Por tanto, tenemos que:

$$\begin{aligned}
\eta^2_{X/Y} &= 1 - \frac{\mathrm{ECM}(\widehat{X}(Y))}{\mathrm{Var}[X]} = 1 - \frac{1/3}{5/12} = \frac{1}{5} \\
\eta^2_{Y/X} &= 1 - \frac{\mathrm{ECM}(\widehat{Y}(X))}{\mathrm{Var}[Y]} = 1 - \frac{1/16}{1/12} = \frac{1}{4}
\end{aligned}$$

Por tanto, la predicción más fiable es la de Y a partir de X, aunque no son buenas predicciones.

Ejercicio 7.5.14. Sea (X, Y) un vector aleatorio con rectas de regresión

$$x + 4y = 1 \qquad x + 5y = 2$$

1. ¿Cuál es la recta de regresión de Y sobre X?

   Suponemos que la recta de regresión de Y sobre X es $x + 4y = 1$. Por tanto, tenemos que:

   $$y = \widehat{Y}(x) = \frac{1}{4} - \frac{x}{4} = E[Y] + \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}\,(x - E[X]) \Longrightarrow \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]} = -\frac{1}{4}$$

   Por otro lado, tenemos que la recta de regresión de X sobre Y es $x + 5y = 2$. Por tanto, tenemos que:

   $$x = \widehat{X}(y) = 2 - 5y = E[X] + \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[Y]}\,(y - E[Y]) \Longrightarrow \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[Y]} = -5$$

Por tanto, el coeficiente de determinación lineal es:

$$\rho_{X,Y}^2 = \frac{\mathrm{Cov}[X,Y]^2}{\mathrm{Var}[X]\,\mathrm{Var}[Y]} = -\frac{1}{4} \cdot (-5) = \frac{5}{4} > 1$$

Esto es un absurdo, luego nuestra suposición era errónea. La recta de regresión de Y sobre X es $x + 5y = 2$ y la de X sobre Y es $x + 4y = 1$.

2. Calcular el coeficiente de correlación lineal y la proporción de varianza de cada variable que queda explicada por la regresión lineal.

   Como la recta de regresión de Y sobre X es $x + 5y = 2$, tenemos que:

$$y = \widehat{Y}(x) = \frac{2-x}{5} = E[Y] + \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]}(x - E[X]) \Longrightarrow \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]} = -\frac{1}{5}$$

   Por otro lado, como la recta de regresión de X sobre Y es $x+4y = 1$, tenemos que:

$$x = \widehat{X}(y) = 1 - 4y = E[X] + \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[Y]}(y - E[Y]) \Longrightarrow \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[Y]} = -4$$

   Por tanto, el coeficiente de determinación lineal es:

$$\rho_{X,Y}^2 = \frac{\mathrm{Cov}[X,Y]^2}{\mathrm{Var}[X]\,\mathrm{Var}[Y]} = -\frac{1}{5} \cdot (-4) = \frac{4}{5} = 0{,}8$$

   Por tanto, la proporción de varianza de cada variable que queda explicada por la regresión lineal es un 80 %. Además, como la covarianza es negativa, el coeficiente de correlación lineal es:

$$\rho_{X,Y} = -\sqrt{\rho_{X,Y}^2} = -\sqrt{\frac{4}{5}} = -\frac{2\sqrt{5}}{5}$$

3. Calcular las medias de ambas variables.

   De la forma de las rectas de regresión, planteamos el siguiente sistema:

$$\begin{cases} E[Y] + {}^1\!/_5 E[X] = {}^2\!/_5 \\ E[X] + 4E[Y] = 1 \end{cases}$$

   Resolviendo el sistema, llegamos a que:

$$E[X] = -3 \qquad E[Y] = 1$$

7.6. Algunos Modelos Multivariantes

Ejercicio 7.6.1. El 60 % de los clientes de un almacén paga con dinero, el 30 % con tarjeta y el 10 % con cheques. Calcular la probabilidad de que de 10 clientes, 5 paguen con dinero, 2 con tarjeta y 3 con cheques.

Dado $i \in \{\text{dinero}, \text{tarjeta}, \text{cheques}\}$, definimos X_i como la variable aleatoria que cuenta el número de clientes que pagan con el método i. Entonces, $(X_{\text{dinero}}, X_{\text{tarjeta}}, X_{\text{cheques}})$ sigue una distribución multinomial con parámetros $n = 10$ y probabilidades $(p_{\text{dinero}}, p_{\text{tarjeta}}, p_{\text{cheques}})$, donde:

$$p_{\text{dinero}} = 0{,}6, \qquad p_{\text{tarjeta}} = 0{,}3, \qquad p_{\text{cheques}} = 0{,}1.$$

Tenemos entonces que:

$$P[X_{\text{dinero}} = 5, X_{\text{tarjeta}} = 2, X_{\text{cheques}} = 3] = \frac{10!}{5!2!3!} \cdot p_{\text{dinero}}^5 \cdot p_{\text{tarjeta}}^2 \cdot p_{\text{cheques}}^3 =$$
$$= \frac{10!}{5!2!3!} \cdot 0{,}6^5 \cdot 0{,}3^2 \cdot 0{,}1^3 \approx 0{,}0176$$

Ejercicio 7.6.2. En un hotel hay tres salas de televisión. En un determinado instante, cada televisor puede sintonizar uno entre 6 canales distintos, A, B, C, D, E y F. Cada canal tiene probabilidad $1/36, 3/36, 5/36, 7/36, 9/36$ y $11/36$, respectivamente, de ser sintonizado, con independencia unas televisiones de otras. Calcular:

1. Probabilidad de que en un instante dado se sintonicen los canales B, D y E.

   Dado $i \in \{A, B, C, D, E, F\}$, definimos X_i como la variable aleatoria que cuenta el número de televisores sintonizando el canal i. Entonces, $(X_A, X_B, X_C, X_D, X_E)$ sigue una distribución multinomial con parámetros $n = 3$ y probabilidades $(p_A, p_B, p_C, p_D, p_E)$, donde

   $$p_A = \frac{1}{36}, \qquad p_B = \frac{3}{36},$$
   $$p_C = \frac{5}{36}, \qquad p_D = \frac{7}{36}, \qquad p_E = \frac{9}{36}.$$

   Consideramos ahora el vector aleatorio (X_B, X_D, X_E), que sabemos que:

   $$(X_B, X_D, X_E) \sim M_3\left(3, p_B, p_D, p_E\right).$$

   Tenemos que:

   $$P[X_B = 1, X_D = 1, X_E = 1] = \frac{3!}{1!1!1!0!} \cdot p_B \cdot p_D \cdot p_E = 6 \cdot \frac{3}{36} \cdot \frac{7}{36} \cdot \frac{9}{36} = \frac{7}{288} \approx 0{,}0243$$

2. Probabilidad de que en un instante dado haya un televisor sintonizando el canal B y otro el E.

   Consideramos ahora el vector aleatorio (X_B, X_E), que sabemos que:

   $$(X_B, X_E) \sim M_2\left(3, p_B, p_E\right).$$

   Tenemos que:

   $$P[X_B = 1, X_E = 1] = \frac{3!}{1!1!1!} \cdot p_B \cdot p_E \cdot (1 - p_B - p_E) = 6 \cdot \frac{3}{36} \cdot \frac{9}{36} \cdot \frac{24}{36} = \frac{1}{12} \approx 0{,}0833$$

3. Probabilidad de que en un instante dado los tres televisores sintonicen el canal F.

   Consideramos ahora el vector aleatorio (X_F), que sabemos que:

   $$(X_F) \sim M_1\left(3, p_F\right).$$

   Tenemos que:

   $$P[X_F = 3] = \frac{3!}{3!0!} \cdot p_F^3 = \left(\frac{11}{36}\right)^3 = \frac{1331}{46656} \approx 0{,}0285$$

4. Probabilidad de que en un instante dado no estén sintonizados A, B, C, D.

   Consideramos ahora el vector aleatorio (X_A, X_B, X_C, X_D), que sabemos que:

   $$(X_A, X_B, X_C, X_D) \sim M_4\left(3, p_A, p_B, p_C, p_D\right).$$

   Tenemos que:

   $$\begin{aligned} P[X_A = 0, &X_B = 0, X_C = 0, X_D = 0] = \frac{3!}{0!0!0!0!3!} \cdot (1 - p_A - p_B - p_C - p_D)^3 = \\ &= 1 \cdot \left(1 - \frac{1}{36} - \frac{3}{36} - \frac{5}{36} - \frac{7}{36}\right)^3 = \left(\frac{5}{9}\right)^3 = \frac{125}{729} \approx 0{,}1715 \end{aligned}$$

Ejercicio 7.6.3. Un ordenador genera números aleatorios del 0 al 9 con independencia e igual probabilidad para cada dígito. Si se generan 12 números aleatorios, calcular:

1. La probabilidad de que aparezca 6 veces el dígito 0, 4 veces el 1 y 2 veces el 2.

   Dado $i \in \{0, 1, 2, \ldots, 8\}$, definimos X_i como la variable aleatoria que cuenta el número de veces que aparece el dígito i. Entonces, $(X_0, X_1, X_2, \ldots, X_8)$ sigue una distribución multinomial con parámetros $n = 12$ y probabilidades $(p_0, p_1, p_2, \ldots, p_8)$, donde:

   $$p := p_0 = p_1 = p_2 = \ldots = p_8 = \frac{1}{10}.$$

   Consideramos ahora el vector aleatorio (X_0, X_1, X_2), que sabemos que:

   $$(X_0, X_1, X_2) \sim M_3\left(12, p_0, p_1, p_2\right).$$

   Tenemos que:

   $$P[X_0 = 6, X_1 = 4, X_2 = 2] = \frac{12!}{6!4!2!0!} \cdot p_0^6 \cdot p_1^4 \cdot p_2^2 = \frac{12!}{6!4!2!0!} p^{12} \approx 13{,}86 \cdot 10^{-9}$$

2. El número esperado de veces que aparece el 0.

   Tenemos que $X_0 \sim M_1\left(12, p_0\right)$, por lo que:

   $$E[X_0] = n \cdot p_0 = 12 \cdot \frac{1}{10} = \frac{6}{5} = 1{,}2$$

3. La probabilidad de que aparezca 4 veces el 1 y 3 veces el 6.

   Consideramos ahora el vector aleatorio (X_1, X_6), que sabemos que:

$$(X_1, X_6) \sim M_2 (12, p_1, p_6).$$

   Tenemos que:

$$P[X_1 = 4, X_6 = 3] = \frac{12!}{4!3!5!} \cdot p_1^4 \cdot p_6^3 \cdot (1 - p_1 - p_6)^5 = \frac{12!}{4!3!5!} p^7 (1 - 2p)^5 \approx 9{,}08 \cdot 10^{-4}$$

Ejercicio 7.6.4. Sea $F : \mathbb{R} \to \mathbb{R}$ una función de distribución continua, y sean α_1 y α_2 números reales tales que $F(\alpha_1) = 0{,}3$ y $F(\alpha_2) = 0{,}8$. Si se seleccionan al azar 25 observaciones independientes de la distribución cuya función de distribución es F. Calcular la probabilidad de que seis de los valores observados sean menores que α_1, diez de los valores observados estén entre α_1 y α_2 y 9 sean mayores que α_2.

Como F es una función de distribución, es creciente. Como $F(\alpha_1) = 0{,}3$ y $F(\alpha_2) = 0{,}8$, tenemos que $\alpha_1 \leqslant \alpha_2$. Definimos las siguientes variables aleatorias:

$$\begin{aligned} X_1 &:= \text{Número de observaciones menores que } \alpha_1, \\ X_2 &:= \text{Número de observaciones entre } \alpha_1 \text{ y } \alpha_2, \\ X_3 &:= \text{Número de observaciones mayores que } \alpha_2. \end{aligned}$$

Tenemos las siguientes probabilidades:

$$\begin{aligned} p_1 &= \text{Prob. de que una observación sea menor que } \alpha_1 = F(\alpha_1) = 0{,}3, \\ p_2 &= \text{Prob. de que una observación esté entre } \alpha_1 \text{ y } \alpha_2 = F(\alpha_2) - F(\alpha_1) = 0{,}8 - 0{,}3 = 0{,}5, \\ p_3 &= \text{Prob. de que una observación sea mayor que } \alpha_2 = 1 - F(\alpha_2) = 0{,}2 = 1 - p_1 - p_2. \end{aligned}$$

Tenemos que:

$$(X_1, X_2) \sim M_2 (25, p_1, p_2).$$

Por lo que la probabilidad pedida es:

$$P[X_1 = 6, X_2 = 10] = \frac{25!}{6!10!9!} \cdot p_1^6 \cdot p_2^{10} \cdot p_3^9 = \frac{25!}{6!10!9!} \cdot 0{,}3^6 \cdot 0{,}5^{10} \cdot 0{,}2^9 \approx 0{,}0059$$

Ejercicio 7.6.5. Si se lanzan 5 dados equilibrados de forma independiente, calcular la probabilidad de que los números 1 y 4 aparezcan el mismo número de veces.

Dado $i \in \{1, 2, 3, 4, 5, 6\}$, definimos X_i como la variable aleatoria que cuenta el número de veces que aparece el número i. Entonces, $(X_1, X_2, X_3, X_4, X_5)$ sigue una distribución multinomial con parámetros $n = 5$ y probabilidades $(p_1, p_2, p_3, p_4, p_5)$, donde:

$$p := p_1 = p_2 = p_3 = p_4 = p_5 = \frac{1}{6}.$$

Consideramos ahora el vector aleatorio (X_1, X_4), que sabemos que:

$$(X_1, X_4) \sim M_2 (5, p_1, p_4).$$

Tenemos que la probabilidad pedida es:

$$P[X_1 = X_2] = \sum_{i=0}^{2} P[X_1 = i, X_4 = i]$$

Calculamos cada uno de los sumandos:

$$P[X_1 = 0, X_4 = 0] = \frac{5!}{0!0!5!} \cdot p_1^0 \cdot p_4^0 \cdot (1 - p_1 - p_4)^5 = \frac{5!}{5!} \cdot \left(\frac{4}{6}\right)^5 = \frac{32}{243},$$
$$P[X_1 = 1, X_4 = 1] = \frac{5!}{1!1!3!} \cdot p_1^1 \cdot p_4^1 \cdot (1 - p_1 - p_4)^3 = 5 \cdot 4 \cdot \left(\frac{1}{6}\right) \cdot \left(\frac{1}{6}\right) \cdot \left(\frac{4}{6}\right)^3 = \frac{40}{243},$$
$$P[X_1 = 2, X_4 = 2] = \frac{5!}{2!2!1!} \cdot p_1^2 \cdot p_4^2 \cdot (1 - p_1 - p_4)^1 = \frac{5!}{2!2!} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{1}{6}\right)^2 \cdot \frac{4}{6} = \frac{5}{324}.$$

Por lo que la probabilidad pedida es:

$$P[X_1 = X_2] = \frac{32}{243} + \frac{40}{243} + \frac{5}{324} = \frac{101}{324} \approx 0{,}312$$

Ejercicio 7.6.6. En un determinado juego de azar existen tres posibles resultados A, B y C, que se dan con probabilidad 0,8,0,15 y 0,05, respectivamente. Una persona realiza 5 veces el juego de forma independiente, calcular la probabilidad de que no obtenga ninguna vez el resultado C ni más de una vez el resultado B.

Dado $i \in \{A, B, C\}$, sea p_i la probabilidad de obtener el resultado i. Definimos la variable aleatoria X_i como el número de veces que se obtiene el resultado i. Se pide:

$$P[X_B \leqslant 1, X_C = 0] = P[X_B = 0, X_C = 0] + P[X_B = 1, X_C = 0]$$

Para calcular cada una de las probabilidades, consideramos el vector aleatorio (X_B, X_C), que sabemos que:

$$(X_B, X_C) \sim M_2\left(5, p_B, p_C\right).$$

Tenemos que:

$$P[X_B = 0, X_C = 0] = \frac{5!}{0!0!5!} \cdot p_B^0 \cdot p_C^0 \cdot (1 - p_B - p_C)^5 = \frac{5!}{5!} \cdot 0{,}8^5 = \frac{1024}{3125},$$
$$P[X_B = 1, X_C = 0] = \frac{5!}{1!0!4!} \cdot p_B^1 \cdot p_C^0 \cdot (1 - p_B - p_C)^4 = 5 \cdot 0{,}15 \cdot 0{,}8^4 = \frac{192}{625}.$$

Por lo que la probabilidad pedida es:

$$P[X_B \leqslant 1, X_C = 0] = \frac{1024}{3125} + \frac{192}{625} = \frac{1984}{3125} \approx 0{,}63488$$

Ejercicio 7.6.7. Sea (X, Y) un vector aleatorio con distribución normal con parámetros $\mu_1 = 5$, $\mu_2 = 8$, $\sigma_1^2 = 16$, $\sigma_2^2 = 9$, $\rho = 0{,}6$. Calcular

$$P[5 < Y < 11 \mid X = 2].$$

Calculemos en primer lugar la distribución condicional de Y dado $X = 2$. Sabemos que:

$$Y \mid X = 2 \sim \mathcal{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(2-\mu_1), \sigma_2^2(1-\rho^2)\right) = \mathcal{N}\left(6{,}65, 5{,}76\right).$$

Por tanto, y notando por Z a la tipificada de $Y \mid X = 2$, tenemos que:

$$\begin{aligned}P[5 < Y < 11 \mid X = 2] &= P[-0{,}6875 < Z < 1{,}8125] = P[Z < 1{,}8125] - P[Z < -0{,}6875] = \\ &= P[Z < 1{,}8125] - 1 + P[Z > -0{,}6875] = \\ &= P[Z < 1{,}8125] - 1 + P[Z < 0{,}6875] \approx 0{,}96485 - 1 + 0{,}75490 = 0{,}71975.\end{aligned}$$

Ejercicio 7.6.8. Calcular la distribución de probabilidad de la suma $X + Y$ para $(X, Y) \sim \mathcal{N}_2((\mu_1, \mu_2), \Sigma)$, siendo

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

Sea la matriz A dada por:

$$A = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Sea $Z = X + Y = (X, Y)A$. Entonces, usando la normalidad para las combinaciones lineales de normales, tenemos:

$$X + Y \sim \mathcal{N}(\mu A, A^t\Sigma A)$$

Tenemos que:

$$\begin{aligned}\mu A &= \mu_1 + \mu_2 \\ A^t\Sigma A &= \begin{pmatrix} \sigma_1^2 + \sigma_1\sigma_2\rho & \sigma_2^2 + \sigma_1\sigma_2\rho \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\rho\end{aligned}$$

Por tanto, $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\rho)$.

Ejercicio 7.6.9. Si (X, Y) es un vector aleatorio con distribución normal, encontrar una condición necesaria y suficiente para que $X+Y$ y $X-Y$ sean independientes.

Tenemos que $(X, Y) \sim \mathcal{N}_2(\mu, \Sigma)$ para cierto vector de esperanzas μ y matriz de covarianzas Σ. Sea la siguiente matriz:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

De esta forma, $Z = (X, Y)A = (X+Y, X-Y)$, por lo que usando la normalidad para las combinaciones lineales de normales, tenemos que:

$$Z \sim \mathcal{N}_2\left(\mu A, A^t\Sigma A\right).$$

Calculamos la matriz de covarianzas de Z:

$$\begin{aligned}
A^t\Sigma A &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \\
&= \begin{pmatrix} \sigma_1^2 + \sigma_1\sigma_2\rho & \sigma_1\sigma_2\rho + \sigma_2^2 \\ \sigma_1^2 - \sigma_1\sigma_2\rho & \sigma_1\sigma_2\rho - \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \\
&= \begin{pmatrix} \sigma_1^2 + 2\sigma_1\sigma_2\rho + \sigma_2^2 & \sigma_1^2 - \sigma_2^2 \\ \sigma_1^2 - \sigma_2^2 & \sigma_1^2 - 2\sigma_1\sigma_2\rho + \sigma_2^2 \end{pmatrix}
\end{aligned}$$

Por tanto, tenemos que:

$$\mathrm{Cov}[X+Y, X-Y] = \sigma_1^2 - \sigma_2^2.$$

Al ser $Z = (X+Y, X-Y)$ una normal bidimensional, por la caracterización de la independencia de las componentes de normales bidimensionales, tenemos que:

$$X+Y \text{ y } X-Y \text{ son independientes} \Leftrightarrow \mathrm{Cov}[X+Y, X-Y] = 0 \Leftrightarrow \sigma_1^2 = \sigma_2^2.$$

Ejercicio 7.6.10. Para cada una de las siguientes densidades normales bidimensionales, hallar $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ y ρ:

1. $f_{(X,Y)}(x,y) = \dfrac{1}{2\pi} \exp\left(-\dfrac{1}{2}((x-1)^2 + (y-2)^2)\right)$

   Buscamos expresar la función de densidad de la forma:

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{(x-\mu_1, y-\mu_2)\Sigma^{-1}(x-\mu_1, y-\mu_2)^t}{2}\right)$$

   Tenemos que:

$$\begin{aligned}
f_{(X,Y)}(x,y) &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}((x-1)^2 + (y-2)^2)\right) \\
&= \frac{1}{2\pi} \exp\left(-\frac{1}{2} \begin{pmatrix} x-1 & y-2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x-1 \\ y-2 \end{pmatrix}\right)
\end{aligned}$$

   Por tanto, tomando $\Sigma = (Id_2)^{-1} = Id_2$, tenemos que $|\Sigma| = 1$ y cuadra. Por tanto,

$$(X,Y) \sim \mathcal{N}_2((1,2), Id_2)$$

   Por tanto:

$$\mu_1 = 1, \quad \mu_2 = 2, \quad \sigma_1^2 = 1, \quad \sigma_2^2 = 1, \quad \rho = 0.$$

2. $f_{(X,Y)}(x,y) = \dfrac{1}{2\pi} \exp\left(-\dfrac{1}{2}(x^2 + y^2 + 4x - 6y + 13)\right)$

   Tenemos que:

$$\begin{aligned}
f_{(X,Y)}(x,y) &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2 + 4x - 6y + 13)\right) \\
&= \frac{1}{2\pi} \exp\left(-\frac{1}{2}((x+2)^2 + (y-3)^2)\right) \\
&= \frac{1}{2\pi} \exp\left(-\frac{1}{2} \begin{pmatrix} x+2 & y-3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x+2 \\ y-3 \end{pmatrix}\right)
\end{aligned}$$

Por tanto, tomando $\Sigma = (Id_2)^{-1} = Id_2$, tenemos que $|\Sigma| = 1$ y cuadra. Por tanto,

$$(X,Y) \sim \mathcal{N}_2((-2,3), Id_2)$$

Por tanto:

$$\mu_1 = -2, \quad \mu_2 = 3, \quad \sigma_1^2 = 1, \quad \sigma_2^2 = 1, \quad \rho = 0.$$

3. $f_{(X,Y)}(x,y) = \dfrac{1}{2{,}4\pi} \exp\left(-\dfrac{1}{0{,}72}\left(\dfrac{x^2}{4} - 0{,}8xy + y^2\right)\right)$

Este caso es más complejo, porque $\Sigma \neq Id_2$. Buscamos ahora expresar la función de densidad de la forma:

$$f_X(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right)$$

Identificando términos, en primer lugar tenemos que:

$$\begin{cases} 2{,}4\pi = 2\pi\sigma_1\sigma_2\sqrt{1-\rho^2} \\ 0{,}72 = 2(1-\rho^2) \end{cases}$$

Por tanto, $1 - \rho^2 = 0{,}36$, de lo que obtenemos que:

$$\sqrt{1-\rho^2} = 0{,}6 \Longrightarrow \sigma_1\sigma_2 = \frac{2{,}4\pi}{2\pi \cdot 0{,}6} = 2,$$
$$1-\rho^2 = 0{,}36 \Longrightarrow \rho^2 = 0{,}64 \Longrightarrow |\rho| = 0{,}8$$

El signo de ρ no podemos determinarlo (ya que tampoco se puede determinar el signo de σ_1 y σ_2).

Por tanto, considerando los siguientes valores, tenemos que cuadraría:

$$\mu_1 = \mu_2 = 0, \quad \sigma_1^2 = 4, \quad \sigma_2^2 = 1$$

Ejercicio 7.6.11. Sea (X,Y) un vector aleatorio con función de densidad en $\mathbb{R}^2$ dada por:

$$f_{(X,Y)}(x,y) = f_1(x,y) + f_2(x,y),$$
$$f_1(x,y) = \frac{1}{4\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right), \quad |\rho| < 1,$$
$$f_2(x,y) = \frac{1}{4\pi\sqrt{1-\tau^2}} \exp\left(-\frac{1}{2(1-\tau^2)}(x^2+y^2-2\tau xy)\right), \quad |\tau| < 1.$$

1. Demostrar que las distribuciones marginales son normales de media cero y varianza uno.

Tenemos que:

$$f_1(x,y) = \frac{1}{2 \cdot} g_1(x,y), \qquad f_2(x,y) = \frac{1}{2} \cdot g_2(x,y)$$

donde:

- $g_1(x,y)$ es la función de densidad de $(X_1,Y_1) \sim \mathcal{N}_2(0,0,1,1,\rho)$.
- $g_2(x,y)$ es la función de densidad de $(X_2,Y_2) \sim \mathcal{N}_2(0,0,1,1,\tau)$.

Calculemos la función de densidad marginal de X:

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\,dy = \int_{-\infty}^{\infty} f_1(x,y)+f_2(x,y)\,dy = \\
&= \int_{-\infty}^{\infty} f_1(x,y)\,dy + \int_{-\infty}^{\infty} f_2(x,y)\,dy = \\
&= \frac{1}{2}\left(\int_{-\infty}^{\infty} g_1(x,y)\,dy + \int_{-\infty}^{\infty} g_2(x,y)\,dy\right) = \\
&= \frac{1}{2}\left(f_{X_1}(x)+f_{X_2}(x)\right)
\end{aligned}$$

donde, en la útlima igualdad, hemos usado que esas integrales son la función de densidad de las marginales de X_1 y X_2 respectivamente. Como sabemos que $X_1, X_2 \sim \mathcal{N}(0,1)$, tenemos que $f_{X_1} = f_{X_2}$, por lo que:

$$f_X(x) = \frac{1}{2}\left(2f_{X_1}(x)\right) = f_{X_1}(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$$

Por tanto, tenemos que:

$$X \sim \mathcal{N}(0,1)$$

Análogamnete, demostramos que $Y \sim \mathcal{N}(0,1)$.

2. Obtener la covarianza de X e Y.

Sabemos que $E[X] = E[Y] = 0$. Calculemos $E[XY]$:

$$\begin{aligned}
E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{(X,Y)}(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy(f_1(x,y)+f_2(x,y))\,dx\,dy = \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_1(x,y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_2(x,y)\,dx\,dy = \\
&= \frac{1}{2}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyg_1(x,y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyg_2(x,y)\,dx\,dy\right) = \\
&= \frac{1}{2}\left(E[X_1Y_1]+E[X_2Y_2]\right)
\end{aligned}$$

Como $E[X_1] = E[Y_1] = E[X_2] = E[Y_2] = 0$, tenemos que:

$$\begin{aligned}
E[XY] &= \frac{1}{2}\left(E[X_1Y_1] - E[X_1]E[Y_1] + E[X_2Y_2] - E[X_2]E[Y_2]\right) = \\
&= \frac{1}{2}\left(\mathrm{Cov}[X_1,Y_1] + \mathrm{Cov}[X_2,Y_2]\right)
\end{aligned}$$

Usando que $(X_1,Y_1) \sim \mathcal{N}_2(0,0,1,1,\rho)$ y $(X_2,Y_2) \sim \mathcal{N}_2(0,0,1,1,\tau)$, tenemos que:

$$\begin{aligned}
\mathrm{Cov}[X_1,Y_1] &= \rho, \\
\mathrm{Cov}[X_2,Y_2] &= \tau
\end{aligned}$$

Por tanto, tenemos que:

$$\mathrm{Cov}[X,Y] = \frac{1}{2}(\rho+\tau)$$

3. ¿Son X e Y independientes? ¿Son X e Y incorreladas?

   Veamos en primer lugar si son independientes, ya que en el caso de que lo sean sabemos que serán incorreladas. Tenemos que $X, Y \sim \mathcal{N}(0,1)$, por lo que:

$$\begin{aligned}
f_X(x)f_Y(y) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\cdot\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right) = \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right)\\
f_{(X,Y)}(x,y) &= \frac{1}{2}\left(g_1(x,y)+g_2(x,y)\right) =\\
&= \frac{1}{2}\left(\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right)+\right.\\
&\qquad \left.+\frac{1}{2\pi\sqrt{1-\tau^2}}\exp\left(-\frac{1}{2(1-\tau^2)}(x^2+y^2-2\tau xy)\right)\right)
\end{aligned}$$

   Evaluando por ejemplo en el $(0,0)$, tenemos que:

$$\begin{aligned}
f_X(0)f_Y(0) &= \frac{1}{2\pi}\\
f_{(X,Y)}(0,0) &= \frac{1}{2}\left(\frac{1}{2\pi\sqrt{1-\rho^2}}+\frac{1}{2\pi\sqrt{1-\tau^2}}\right) = \frac{1}{2\pi}\left(\frac{1}{2\sqrt{1-\rho^2}}+\frac{1}{2\sqrt{1-\tau^2}}\right)
\end{aligned}$$

   Por tanto, tenemos que:

$$\begin{aligned}
f_X(0)f_Y(0) = f_{(X,Y)}(0,0) &\Longleftrightarrow 1 = \frac{1}{2\sqrt{1-\rho^2}}+\frac{1}{2\sqrt{1-\tau^2}}\\
&\Longleftrightarrow 2 = \frac{1}{\sqrt{1-\rho^2}}+\frac{1}{\sqrt{1-\tau^2}}
\end{aligned}$$

   En este caso, como $\rho^2, \tau^2 > 0$, tenemos que ambos radicandos son menores o iguales que uno, por lo que ambos denomininadores son menores o iguales que uno. Por tanto, ambos sumandos son mayores o iguales que uno. Para que su suma sea 2, es necesario que ambos sumandos sean iguales a uno, y esto solo se tiene si $\rho = \tau = 0$. Por tanto, tenemos que:

$$f_X(0)f_Y(0) = f_{(X,Y)}(0,0) \Longleftrightarrow \rho = \tau = 0$$

   Por tanto, en caso contrario, tenemos que X e Y no son independientes. En el caso de que $\rho = \tau = 0$, tenemos que:

$$\begin{aligned}
f_{(X,Y)}(x,y) &= \frac{1}{2}\left(\frac{1}{2\pi}\exp\left(-\frac{1}{2}(x^2+y^2)\right)+\frac{1}{2\pi}\exp\left(-\frac{1}{2}(x^2+y^2)\right)\right) =\\
&= \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right) = f_X(x)f_Y(y) \qquad \forall (x,y)\in\mathbb{R}^2
\end{aligned}$$

Por tanto, concluimos que:

$$X \text{ e } Y \text{ son independientes} \Longleftrightarrow \rho = \tau = 0$$

Respecto a su correlación, como $\text{Cov}[X, Y] = \frac{1}{2}(\rho + \tau)$, tenemos que:

$$X \text{ e } Y \text{ son incorreladas} \Longleftrightarrow \text{Cov}[X, Y] = 0 \Longleftrightarrow \rho = -\tau$$

De nuevo, notemos que si son independientes, entonces son incorreladas, pero el recíproco no es cierto.

Ejercicio 7.6.12. Sea (X, Y) un vector aleatorio con distribución normal con parámetros $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ y ρ. Demostrar que X e Y son independientes $\Leftrightarrow \rho = 0$.

Demostrado en la Proposición 5.16.

Ejercicio 7.6.13. Sea (X, Y) un vector aleatorio con distribución normal bidimensional con las siguientes características:

- La mediana de X es 1.
- $\text{Var}(X) = \text{Var}(Y)$.
- $\rho_{X,Y} = 0{,}5$.
- $\text{Cov}(X, Y) = 2/3$.
- $P[X \leqslant -1 \mid Y = 1] = 0{,}06681$.

Determinar el vector de medias y la matriz de covarianzas de (X, Y).

Tenemos que:

$$0{,}5 = \rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}} \overset{(*)}{=} \frac{2/3}{\text{Var}(X)} \Longrightarrow \text{Var}(X) = \text{Var}(Y) = \frac{2}{3} \cdot 2 = \frac{4}{3}$$

donde en $(*)$ he empleado que $\text{Var}(X) = \text{Var}(Y)$. Por tanto, la matriz de covarianzas es:

$$\Sigma = \begin{pmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & \text{Var}(Y) \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}$$

Como (X, Y) es una normal bidimensional, X es una normal, luego su mediana es igual a su media. Por tanto, $E[X] = 1$. Para calcular $E[Y]$, usamos el único resultado que no hemos empleado, la distribución $X \mid Y = 1$. Como tenemos que $(X, Y) \sim \mathcal{N}_2(1, E[Y], 4/3, 4/3, 1/2)$, entonces:

$$\begin{aligned}(X \mid Y = 1) &\sim \mathcal{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(1 - \mu_2), \sigma_1^2(1 - \rho^2)\right) = \mathcal{N}\left(1 + \frac{1}{2} \cdot 1(1 - E[Y]), \frac{4}{3} \cdot \frac{3}{4}\right) = \\ &= \mathcal{N}\left(1{,}5 - \frac{E[Y]}{2}, 1\right)\end{aligned}$$

Como la normal es simétrica respecto a su media, tenemos que:

$$0{,}06681 = P[X \leqslant -1 \mid Y = 1] = 1 - P[X > -1 \mid Y = 1] = 1 - P[X < 1 \mid Y = 1] \Longrightarrow$$
$$\Longrightarrow P[X < 1 \mid Y = 1] = 1 - 0{,}06681 = 0{,}93319$$

Sea Z la tipificada de $X \mid Y = 1$. Entonces, tenemos que:

$$0{,}93319 = P[X < 1 \mid Y = 1] = P\left[Z < \frac{1 - 1{,}5 + {}^{E[Y]}\!/_2}{1}\right] = P\left[Z < \frac{-0{,}5 + {}^{E[Y]}\!/_2}{1}\right]$$

Consultando la tabla de la normal estándar, como $P[Z \leqslant 1{,}5] = 0{,}93319$, tenemos que:

$$-0{,}5 + \frac{E[Y]}{2} = 1{,}5 \Longrightarrow E[Y] = 4$$

Por tanto, el vector de medias es:

$$\mu = \begin{pmatrix} 1 & 4 \end{pmatrix}$$

7.7. Teorema Central del Límite

Ejercicio 7.7.1. Se realizan 2500 lanzamientos independientes de una moneda correcta. Calcular la probabilidad aproximada de obtener $1/2$ como frecuencia relativa de cara con un error máximo de 0,02.

De aquí en adelante, fijemos $n = 2500$, aunque lo haremos dependiente de n para una mayor generalidad. Para cada $i = 1, \dots, n$, sea X_i una variable aleatoria que toma el valor 1 si en el i-ésimo lanzamiento se obtiene cara, y 0 en caso contrario, de forma que $X_i \sim B(1, 1/2)$. Consideramos la variable aleatoria:

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

De esta forma, por la Reproductividad de la Binomial, $S_n \sim B(n, 1/2)$. Tiene sentido, puesto que S_n es el número de caras obtenidas en n lanzamientos de una moneda correcta. Nos piden:

$$\begin{aligned} P\left[\left|\frac{S_n}{n} - \frac{1}{2}\right| \leqslant 0{,}02\right] &= P\left[-0{,}02 \leqslant \frac{S_n}{n} - \frac{1}{2} \leqslant 0{,}02\right] = \\ &= P\left[0{,}48 \leqslant \frac{S_n}{n} \leqslant 0{,}52\right] \end{aligned}$$

Para calcular esta probabilidad, hay diversas opciones:

Usar la distribución de S_n Dado que $S_n \sim B(n, 1/2)$, podemos calcular la probabilidad directamente. Por el Teorema de Cambio de Variable, tenemos que:

$$\begin{aligned} P\left[\left|\frac{S_n}{n} - \frac{1}{2}\right| \leqslant 0{,}02\right] &= P\left[0{,}48 \leqslant \frac{S_n}{n} \leqslant 0{,}52\right] = \\ &= P\left[n \cdot 0{,}48 \leqslant S_n \leqslant n \cdot 0{,}52\right] \end{aligned}$$

Esta probabilidad teóricamente sabemos calcularla de forma exacta, aunque en la práctica es complicado para valores grandes de n, ya que obtenemos factoriales de números grandes. Probemos para $n = 2500$:

$$\begin{aligned} P\left[\left|\frac{S_{2500}}{2500} - \frac{1}{2}\right| \leqslant 0{,}02\right] &= P\left[0{,}48 \leqslant \frac{S_{2500}}{2500} \leqslant 0{,}52\right] = \\ &= P\left[1200 \leqslant S_{2500} \leqslant 1300\right] = \\ &= \sum_{k=1200}^{1300} P[S_{2500} = k] = \sum_{k=1200}^{1300} \binom{2500}{k} \left(\frac{1}{2}\right)^{2500} \end{aligned}$$

Mediante cálculo numérico con herramientas disponibles en internet, tenemos que:

$$P\left[\left|\frac{S_{2500}}{2500} - \frac{1}{2}\right| \leqslant 0{,}02\right] \approx 0{,}95664$$

No obstante, usaremos lo desarrollado en teoría para obtener otras aproximaciones.

Teorema de Lévy. Distribución Degenerada. Dado que n es grande, podemos aplicar el Teorema de Lévy. Tenemos que:

$$\begin{aligned}
P\left[\left|\frac{S_n}{n}-\frac{1}{2}\right| \leqslant 0{,}02\right] &= P\left[0{,}48 \leqslant \frac{S_n}{n} \leqslant 0{,}52\right] = \\
&= P\left[0{,}48-\frac{E[S_n]}{n} \leqslant \frac{S_n-E[S_n]}{n} \leqslant 0{,}52-\frac{E[S_n]}{n}\right] = \\
&= P\left[\frac{S_n-E[S_n]}{n} \leqslant 0{,}52-\frac{E[S_n]}{n}\right] - P\left[\frac{S_n-E[S_n]}{n} < 0{,}48-\frac{E[S_n]}{n}\right]
\end{aligned}$$

Por la Convergencia en Ley a la Distribución Degenerada, tenemos que:

$$\left\{\frac{S_n-E[S_n]}{n}\right\} \xrightarrow{L} 0$$

Por tanto, tenemos que:

$$\begin{aligned}
P\left[\left|\frac{S_n}{n}-\frac{1}{2}\right| \leqslant 0{,}02\right] &= P\left[\frac{S_n-E[S_n]}{n} \leqslant 0{,}52-\frac{E[S_n]}{n}\right] - P\left[\frac{S_n-E[S_n]}{n} < 0{,}48-\frac{E[S_n]}{n}\right] \approx \\
&\approx P\left[0 \leqslant 0{,}52-\frac{E[S_n]}{n}\right] - P\left[0 < 0{,}48-\frac{E[S_n]}{n}\right] = \\
&= P\left[0 \leqslant 0{,}52-p\right] - P\left[0 < 0{,}48-p\right] = 1-0 = 1
\end{aligned}$$

Teorema de Lévy. Distribución Normal. Dado que n es grande, podemos aplicar el Teorema de Lévy. Tenemos que:

$$\begin{aligned}
P\left[\left|\frac{S_n}{n}-\frac{1}{2}\right| \leqslant 0{,}02\right] &= P\left[n\cdot 0{,}48 \leqslant S_n \leqslant n\cdot 0{,}52\right] = \\
&= P\left[\frac{n\cdot 0{,}48-E[S_n]}{\sqrt{\operatorname{Var}[S_n]}} \leqslant \frac{S_n-E[S_n]}{\sqrt{\operatorname{Var}[S_n]}} \leqslant \frac{n\cdot 0{,}52-E[S_n]}{\sqrt{\operatorname{Var}[S_n]}}\right]
\end{aligned}$$

Por la Convergencia en Ley a la Normal, tenemos que:

$$\left\{\frac{S_n-E[S_n]}{\sqrt{\operatorname{Var}[S_n]}}\right\} \xrightarrow{L} \mathcal{N}(0,1)$$

Por tanto, notando por Z a la tipificada de la normal, tenemos que:

$$\begin{aligned}
P\left[\left|\frac{S_n}{n}-\frac{1}{2}\right| \leqslant 0{,}02\right] &= P\left[\frac{n\cdot 0{,}48-E[S_n]}{\sqrt{\operatorname{Var}[S_n]}} \leqslant \frac{S_n-E[S_n]}{\sqrt{\operatorname{Var}[S_n]}} \leqslant \frac{n\cdot 0{,}52-E[S_n]}{\sqrt{\operatorname{Var}[S_n]}}\right] \approx \\
&\approx P\left[\frac{n\cdot 0{,}48-E[S_n]}{\sqrt{\operatorname{Var}[S_n]}} \leqslant Z \leqslant \frac{n\cdot 0{,}52-E[S_n]}{\sqrt{\operatorname{Var}[S_n]}}\right]
\end{aligned}$$

Usando ahora que $S_n \sim B(n, 1/2)$ y que $n = 2500$, tenemos que:

$$\begin{aligned}
P\left[\left|\frac{S_{2500}}{2500}-\frac{1}{2}\right| \leqslant 0{,}02\right] &\approx P\left[\frac{2500\cdot 0{,}48-2500\cdot 0{,}5}{\sqrt{2500\cdot 0{,}5\cdot 0{,}5}} \leqslant Z \leqslant \frac{2500\cdot 0{,}52-2500\cdot 0{,}5}{\sqrt{2500\cdot 0{,}5\cdot 0{,}5}}\right] = \\
&= P\left[-2 \leqslant Z \leqslant 2\right] = \\
&= 2\cdot P\left[Z \leqslant 2\right]-1 \approx 2\cdot 0{,}97725-1 = 0{,}9545
\end{aligned}$$

Por tanto, hemos obtenido tres aproximaciones distintas. Notemos que la primera y la tercera coinciden a excepción de la corrección por continuidad, mientras que la segunda es ligeramente distinta.

Ejercicio 7.7.2. Calcular, aproximadamente, la probabilidad de que al lanzar 100 veces un dado, la media de los puntos obtenidos sea mayor que 3,7.

De aquí en adelante, fijemos $n = 100$, aunque lo haremos dependiente de n para una mayor generalidad.

Para cada $i = 1, \dots, n$, sea X_i una variable aleatoria que toma el valor $1, 2, 3, 4, 5$ o 6 según el resultado del i-ésimo lanzamiento de un dado, de forma que se tiene $X_i \sim \mathcal{U}(\{1, 2, 3, 4, 5, 6\})$. Consideramos la variable aleatoria:

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

Por tanto, nos piden:

$$P\left[\frac{S_n}{n} > 3{,}7\right] = 1 - P\left[\frac{S_n}{n} \leqslant 3{,}7\right]$$

Calculemos dicha probabilidad. Tenemos que, para $n = 100$:

$$P\left[\frac{S_n}{n} \leqslant 3{,}7\right] = P\left[S_n \leqslant 3{,}7 \cdot n\right] = P\left[S_{100} \leqslant 370\right] = \sum_{k=100}^{370} P[S_{100} = k]$$

Calcular de forma directa dicha probabilidad no es sencillo, por lo que haremos aproximaciones. Tenemos que:

$$\begin{aligned}
E[X_1] &= \frac{1+2+3+4+5+6}{6} = \frac{7}{2} = 3{,}5 \\
E[X_1^2] &= \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} = \frac{91}{6} \\
\mathrm{Var}[X_1] &= E[X_1^2] - E[X_1]^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12} \\
E[S_n] &= E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = n \cdot E[X_1] = 3{,}5 \cdot n \\
\mathrm{Var}[S_n] &= \mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = n \cdot \mathrm{Var}[X_1] = \frac{35}{12} \cdot n
\end{aligned}$$

Por el Teorema de Lévy, tenemos que:

$$\left\{\frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right\} \xrightarrow{L} \mathcal{N}(0, 1)$$

Por tanto, sea Z la tipificada de la normal, tenemos que:

$$\begin{aligned}
P\left[\frac{S_n}{n} > 3{,}7\right] &= 1 - P\left[\frac{S_n}{n} \leqslant 3{,}7\right] = 1 - P\left[\frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \leqslant \frac{3{,}7 \cdot n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right] \approx \\
&\approx 1 - P\left[Z \leqslant \frac{3{,}7 \cdot n - 3{,}5 \cdot n}{\sqrt{\frac{35}{12} \cdot n}}\right]
\end{aligned}$$

En concreto, para $n = 100$, tenemos que:

$$P\left[\frac{S_{100}}{100} > 3{,}7\right] \approx 1 - P\left[Z \leqslant 1{,}17\right] \approx 1 - 0{,}87900 = 0{,}121$$

Ejercicio 7.7.3. El número de piezas correctas elaboradas en una fábrica cuadruplica el de piezas defectuosas, y la probabilidad de producir una pieza defectuosa se mantiene constante durante todo el proceso de fabricación. Se eligen al azar 200 piezas, calcular aproximadamente la probabilidad de que el número de defectuosas oscile entre 40 y 50.

De aquí en adelante, fijemos $n = 200$, aunque lo haremos dependiente de n para una mayor generalidad.

Obtengamos en primer lugar la probabilidad de que una pieza sea defectuosa. Sea k el número de piezas defectuosas, de forma que el número de piezas correctas es $4k$. Por tanto, tenemos que:

$$P[\text{defectuosa}] = \frac{k}{4k+k} = \frac{k}{5k} = \frac{1}{5}$$

Para cada $i = 1, \dots, n$, sea X_i una variable aleatoria que toma el valor 1 si la i-ésima pieza es defectuosa, y 0 en caso contrario, de forma que se tiene $X_i \sim B(1, 1/5)$. Consideramos la variable aleatoria:

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

Notemos que $S_n \sim B(n, 1/5)$ por la propiedad de reproductividad de la binomial. Además, S_n es el número de piezas defectuosas en n piezas elegidas al azar. Por tanto, nos piden:

$$P[40 \leqslant S_n \leqslant 50]$$

Calcular de forma directa dicha probabilidad no es sencillo, por lo que hemos de aproximarlo, y lo haremos por la Convergencia en Ley a la Normal demostrada en el Teorema de Lévy. Tenemos que:

$$\left\{\frac{S_n - E[S_n]}{\sqrt{\operatorname{Var}[S_n]}}\right\} \xrightarrow{L} \mathcal{N}(0,1)$$

Por tanto, sea Z la tipificada de la normal, tenemos que:

$$\begin{aligned}
P[40 \leqslant S_n \leqslant 50] &= P\left[\frac{40 - E[S_n]}{\sqrt{\operatorname{Var}[S_n]}} \leqslant \frac{S_n - E[S_n]}{\sqrt{\operatorname{Var}[S_n]}} \leqslant \frac{50 - E[S_n]}{\sqrt{\operatorname{Var}[S_n]}}\right] \approx \\
&\approx P\left[\frac{40 - E[S_n]}{\sqrt{\operatorname{Var}[S_n]}} \leqslant Z \leqslant \frac{50 - E[S_n]}{\sqrt{\operatorname{Var}[S_n]}}\right]
\end{aligned}$$

Usando que $S_n \sim B(n, 1/5)$, tenemos que:

$$E[S_n] = n \cdot \frac{1}{5} = \frac{n}{5}$$
$$\operatorname{Var}[S_n] = n \cdot \frac{1}{5} \cdot \frac{4}{5} = \frac{4n}{25}$$

Por tanto, para $n = 200$, tenemos que:

$$\begin{aligned}
P[40 \leqslant S_{200} \leqslant 50] &\approx P\left[\frac{40-\frac{200}{5}}{\sqrt{\frac{4\cdot 200}{25}}} \leqslant Z \leqslant \frac{50-\frac{200}{5}}{\sqrt{\frac{4\cdot 200}{25}}}\right] = \\
&= P\left[0 \leqslant Z \leqslant 1{,}7678\right] = \\
&= P\left[Z \leqslant 1{,}7678\right] - P\left[Z < 0\right] = \\
&= 0{,}96164 - 0{,}5 = 0{,}46164
\end{aligned}$$

Ejercicio 7.7.4. Se elige un punto aleatorio $(X_1, \dots, X_{100})$ en el espacio $\mathbb{R}^{100}$. Suponiendo que las variables aleatorias son independientes e idénticamente distribuidas según una $\mathcal{U}([-1,1])$, calcular, aproximadamente, la probabilidad de que el cuadrado de la distancia del punto al origen sea menor que 40.

De aquí en adelante, fijemos $n = 100$, aunque lo haremos dependiente de n para una mayor generalidad.

Para cada $i = 1, \dots, n$, sea X_i una variable aleatoria que toma valores en $[-1,1]$, de forma que se tiene $X_i \sim \mathcal{U}([-1,1])$. Además, para cada $i = 1, \dots, n$, sea $Y_i = X_i^2$. Consideramos la variable aleatoria que mide el cuadrado de la distancia al origen:

$$S_n = \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i$$

Por tanto, nos piden:

$$P[S_n < 40]$$

Calcular de forma directa dicha probabilidad no es sencillo, por lo que hemos de aproximarlo, y lo haremos por la Convergencia en Ley a la Normal demostrada en el Teorema de Lévy. Comprobemos las hipótesis. Como $X_1, \dots, X_{100}$ son independientes e idénticamente distribuidas, tenemos que $Y_1, \dots, Y_{100}$ también lo son. La función de densidad de X_1 es:

$$f_{X_1}(x) = \begin{cases} \frac{1}{1-(-1)} = 1/2 & x \in [-1,1] \\ 0 & \text{en otro caso} \end{cases}$$

Por tanto, tenemos que:

$$\begin{aligned}
E[Y_1] &= E[X_1^2] = \int_{-1}^{1} x^2 \cdot \frac{1}{2}\, dx = \frac{1}{2}\int_{-1}^{1} x^2\, dx = \frac{1}{2}\left[\frac{x^3}{3}\right]_{-1}^{1} = \frac{1}{2}\left(\frac{1}{3} - \frac{-1}{3}\right) = \frac{1}{3} \\
E[Y_1^2] &= E[X_1^4] = \int_{-1}^{1} x^4 \cdot \frac{1}{2}\, dx = \frac{1}{2}\int_{-1}^{1} x^4\, dx = \frac{1}{2}\left[\frac{x^5}{5}\right]_{-1}^{1} = \frac{1}{2}\left(\frac{1}{5} - \frac{-1}{5}\right) = \frac{1}{5} \\
\mathrm{Var}[Y_1] &= E[Y_1^2] - E[Y_1]^2 = \frac{1}{5} - \frac{1}{9} = \frac{4}{45} \\
E[S_n] &= E\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} E[Y_i] = n \cdot E[Y_1] = \frac{n}{3} \\
\mathrm{Var}[S_n] &= \mathrm{Var}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \mathrm{Var}[Y_i] = n \cdot \mathrm{Var}[Y_1] = \frac{4n}{45}
\end{aligned}$$

Por tanto, por el Teorema de Lévy, tenemos que:

$$\left\{\frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right\} \xrightarrow{L} \mathcal{N}(0,1)$$

Por tanto, sea Z la tipificada de la normal, tenemos que:

$$P[S_n < 40] = P\left[\frac{S_n}{n} < \frac{40}{n}\right] = P\left[\frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}} < \frac{40 - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right] \approx P\left[Z < \frac{40 - \frac{n}{3}}{\sqrt{\frac{4n}{45}}}\right]$$

En concreto, para $n = 100$, tenemos que:

$$P[S_{100} < 40] \approx P\left[Z < \frac{20 \cdot 3}{3 \cdot 4\sqrt{5}}\right] = P\left[Z < \sqrt{5}\right] \approx P\left[Z < 2{,}2361\right] \approx 0{,}98745$$

Ejercicio 7.7.5. Cierta enfermedad afecta al 0,5 % de una población. Existe una prueba para la detección de la enfermedad, que da positiva en los individuos enfermos con probabilidad 0,99 y da negativa en los individuos sanos con probabilidad 0,99.

1. Calcular la probabilidad de que un individuo elegido al azar esté realmente enfermo si la prueba da resultado positivo.

   Notemos de la siguiente forma los siguientes sucesos:

   - E: El individuo está enfermo.
   - S: El individuo está sano.
   - $+$: La prueba da positiva.
   - $-$: La prueba da negativa.

   Sabemos que:

   $$P[E] = 0{,}005 \quad P[S] = 0{,}995 \quad P[+ \mid E] = 0{,}99 \quad P[- \mid S] = 0{,}99$$

   Nos piden calcular $P[E \mid +]$. Por la regla de Bayes, tenemos que:

   $$\begin{aligned} P[E \mid +] &= \frac{P[+ \mid E] \cdot P[E]}{P[+]} = \frac{P[+ \mid E] \cdot P[E]}{P[+ \mid E] \cdot P[E] + P[+ \mid S] \cdot P[S]} = \\ &= \frac{0{,}99 \cdot 0{,}005}{0{,}99 \cdot 0{,}005 + 0{,}01 \cdot 0{,}995} = \frac{99}{298} \approx 0{,}3322 \end{aligned}$$

2. Calcular, aproximadamente, el número mínimo de personas con resultado positivo en la prueba que deben ser elegidas, de forma aleatoria e independiente, para asegurar una proporción de personas realmente enfermas en la muestra inferior a un $1/2$, con probabilidad mayor o igual que 0,95.

   Sea n el número de personas elegidas. Para cada $i = 1, \ldots, n$, sea X_i una variable aleatoria que toma el valor 1 si el i-ésimo individuo está enfermo, y 0 en caso contrario, de forma que se tiene $X_i \sim B(1, P[E \mid +])$. Consideramos la variable aleatoria:

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

Por la reproducción de la binomial, tenemos que $S_n \sim B(n, P[E \mid +])$. Además, S_n es el número de personas realmente enfermas en n elegidas al azar con resultado positivo en la prueba. Nos piden:

$$P\left[\frac{S_n}{n} < \frac{1}{2}\right] \geqslant 0{,}95$$

Como se pide un resultado aproximado, usaremos el Teorema de Lévy. Tenemos que:

$$\begin{aligned} E[S_n] &= n \cdot P[E \mid +] \\ \operatorname{Var}[S_n] &= n \cdot P[E \mid +] \cdot P[S \mid +] \end{aligned}$$

Por tanto, por el Teorema de Lévy, tenemos que:

$$\left\{\frac{S_n - E[S_n]}{\sqrt{\operatorname{Var}[S_n]}}\right\} \xrightarrow{L} \mathcal{N}(0,1)$$

Por tanto, sea Z la tipificada de la normal, tenemos que:

$$\begin{aligned} P\left[\frac{S_n}{n} < \frac{1}{2}\right] \geqslant 0{,}95 \Longleftrightarrow P\left[\frac{S_n - E[S_n]}{\sqrt{\operatorname{Var}[S_n]}} < \frac{{}^{n}\!/\!{}_{2} - E[S_n]}{\sqrt{\operatorname{Var}[S_n]}}\right] \geqslant 0{,}95 \Longleftrightarrow \\ \Longleftrightarrow P\left[Z < \frac{{}^{n}\!/\!{}_{2} - n \cdot P[E \mid +]}{\sqrt{n \cdot P[E \mid +] \cdot P[S \mid +]}}\right] \geqslant 0{,}95 \end{aligned}$$

Por tanto, consultando en la tabla de la normal, tenemos que:

$$\begin{aligned} \frac{{}^{n}\!/\!{}_{2} - n \cdot P[E \mid +]}{\sqrt{n \cdot P[E \mid +] \cdot P[S \mid +]}} \geqslant 1{,}65 \Longleftrightarrow \frac{{}^{n}\!/\!{}_{2} - n \cdot {}^{99}\!/\!{}_{298}}{\sqrt{n \cdot {}^{99}\!/\!{}_{298} \cdot {}^{199}\!/\!{}_{298}}} \geqslant 1{,}65 \Longleftrightarrow \\ \Longleftrightarrow \sqrt{n} \cdot \frac{{}^{1}\!/\!{}_{2} - {}^{99}\!/\!{}_{298}}{\sqrt{{}^{99}\!/\!{}_{298} \cdot {}^{199}\!/\!{}_{298}}} \geqslant 1{,}65 \Longleftrightarrow \\ \Longleftrightarrow \sqrt{n} \geqslant 4{,}6319 \Longleftrightarrow n \geqslant 21{,}45 \end{aligned}$$

Por tanto, como n ha de ser un número entero, el número mínimo de personas con resultado positivo en la prueba que deben ser elegidas es 22.

Ejercicio 7.7.6. Una empresa necesita adquirir al menos 100 vehículos. Para ello realiza una prueba a una población de coches compuesta de dos tipos distintos (un 40 % de tipo A y un 60 % de tipo B) y un coche es adquirido si supera la prueba. Un coche de tipo A supera la prueba con probabilidad $1/3$, mientras que para uno de tipo B, dicha probabilidad es $2/3$.

1. Calcular la probabilidad de que un coche elegido al azar supere la prueba.

   Notemos de la siguiente forma los siguientes sucesos:

   - A: El coche es de tipo A.
   - B: El coche es de tipo B.
   - S: El coche supera la prueba.
   - F: El coche no supera la prueba.

   Sabemos que:

   $$P[A] = 0{,}4 \quad P[B] = 0{,}6 \quad P[S \mid A] = 1/3 \quad P[S \mid B] = 2/3$$

   Nos piden calcular $P[S]$. Por la regla de la probabilidad total, tenemos que:

   $$P[S] = P[S \mid A] \cdot P[A] + P[S \mid B] \cdot P[B] = \frac{1}{3} \cdot 0{,}4 + \frac{2}{3} \cdot 0{,}6 = \frac{8}{15} = 0{,}5333$$

2. Calcula el número de coches que deben examinarse para cubrir las necesidades de la empresa con probabilidad mayor o igual que 0,90147.

   Sea n el número de coches examinados. Para cada $i = 1, \ldots, n$, sea X_i una variable aleatoria que toma el valor 1 si el i-ésimo coche supera la prueba, y 0 en caso contrario, de forma que se tiene $X_i \sim B(1, P[S])$. Consideramos la variable aleatoria:

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

   Por la reproducción de la binomial, tenemos que $S_n \sim B(n, P[S])$. Además, S_n es el número de coches que superan la prueba en n examinados. Nos piden:

   $$P[S_n \geqslant 100] \geqslant 0{,}90147$$

   Como se pide un resultado aproximado, usaremos el Teorema de Lévy. Tenemos que:

   $$E[S_n] = n \cdot P[S]$$
   $$\text{Var}[S_n] = n \cdot P[S] \cdot P[F]$$

   Por tanto, por el Teorema de Lévy, tenemos que:

   $$\left\{ \frac{S_n - E[S_n]}{\sqrt{\text{Var}[S_n]}} \right\} \xrightarrow{L} \mathcal{N}(0,1)$$

   Por tanto, sea Z la tipificada de la normal, tenemos que:

   $$\begin{aligned} P[S_n \geqslant 100] \geqslant 0{,}90147 &\iff P\left[\frac{S_n - E[S_n]}{\sqrt{\text{Var}[S_n]}} \geqslant \frac{100 - E[S_n]}{\sqrt{\text{Var}[S_n]}} \right] \geqslant 0{,}90147 \iff \\ &\iff P\left[Z \geqslant \frac{100 - n \cdot P[S]}{\sqrt{n \cdot P[S] \cdot P[F]}} \right] \geqslant 0{,}90147 \iff \\ &\iff P\left[Z < -\frac{100 - n \cdot P[S]}{\sqrt{n \cdot P[S] \cdot P[F]}} \right] \geqslant 0{,}90147 \end{aligned}$$

Por tanto, consultando en la tabla de la normal, tenemos que:

$$\begin{aligned}
-\frac{100 - n \cdot P[S]}{\sqrt{n \cdot P[S] \cdot P[F]}} \geqslant 1{,}29 \Longleftrightarrow \frac{100 - n \cdot P[S]}{\sqrt{n \cdot P[S] \cdot P[F]}} &\leqslant -1{,}29 \Longleftrightarrow \\
\Longleftrightarrow \frac{100 - n \cdot {}^8\!/_{15}}{\sqrt{n \cdot {}^8\!/_{15} \cdot {}^7\!/_{15}}} &\leqslant -1{,}29 \Longleftrightarrow \\
\Longleftrightarrow 100 - n \cdot {}^8\!/_{15} \leqslant -\sqrt{n} \cdot 0{,}6435 &\Longleftrightarrow \\
\Longleftrightarrow \frac{375}{2} - n + \sqrt{n} \cdot 1{,}2067 &\leqslant 0
\end{aligned}$$

Resolvemos para el caso de la igualdad. Tenemos que:

$$\sqrt{n} = \frac{-1{,}2067 \pm \sqrt{1{,}2067^2 + 4 \cdot {}^{375}\!/_2}}{-2} = \frac{-1{,}2067 \pm 27{,}4127}{-2} \Longrightarrow \sqrt{n} = 14{,}3097 \Longrightarrow n = 204{,}76$$

Por tanto, evaluando vemos que los valores que cumplen la desigualdad han de ser mayores o iguales que 204,76. Como n ha de ser un número entero, el número de coches que deben examinarse para cubrir las necesidades de la empresa con probabilidad mayor o igual que 0,90147 es 204.

Ejercicio 7.7.7. Para realizar una compra de un determinado material eléctrico del que se sabe que es defectuoso con probabilidad 0,25, se le somete a una determinada prueba que da los resultados A, B y C, con probabilidades 0,8, 0,15 y 0,05, si el material es válido, y probabilidades 0,2, 0,3 y 0,5, si el material es defectuoso. Si cada lote se somete a seis pruebas de forma independiente, y se acepta para su compra si no aparece nunca el resultado C ni más de dos veces el resultado B, calcular:

1. La probabilidad de que un lote elegido al azar sea aceptado para su compra.

   Notemos de la siguiente forma los siguientes sucesos:

   - V: El material es válido.
   - D: El material es defectuoso.
   - A: Aparece el resultado A.
   - B: Aparece el resultado B.
   - C: Aparece el resultado C.

   Sabemos que:

$$\begin{gathered}
P[V] = 0{,}75 \quad P[D] = 0{,}25 \\
P[A \mid V] = 0{,}8 \quad P[B \mid V] = 0{,}15 \quad P[C \mid V] = 0{,}05 \\
P[A \mid D] = 0{,}2 \quad P[B \mid D] = 0{,}3 \quad P[C \mid D] = 0{,}5
\end{gathered}$$

   Por la regla de la probabilidad total, tenemos que:

$$P[A] = P[A \mid V] \cdot P[V] + P[A \mid D] \cdot P[D] = 0{,}8 \cdot 0{,}75 + 0{,}2 \cdot 0{,}25 = \frac{13}{20} = 0{,}65$$

$$P[B] = P[B \mid V] \cdot P[V] + P[B \mid D] \cdot P[D] = 0{,}15 \cdot 0{,}75 + 0{,}3 \cdot 0{,}25 = \frac{3}{16} = 0{,}1875$$

$$P[C] = P[C \mid V] \cdot P[V] + P[C \mid D] \cdot P[D] = 0{,}05 \cdot 0{,}75 + 0{,}5 \cdot 0{,}25 = \frac{13}{80} = 0{,}1625$$

Dado $i \in \{A, B, C\}$, sea X_i la variable aleatoria que modela el número de veces que aparece el resultado i en las seis pruebas de un lote. Tenemos que:

$$(X_B, X_C) \sim \mathcal{M}_2(6, P[B], P[C])$$

De esta forma, la probabilidad de que un lote elegido al azar sea aceptado para su compra es:

$$\begin{aligned} P[\text{aceptado}] &= P[X_B \leqslant 2, X_C = 0] = \\ &= P[X_B = 2, X_C = 0] + P[X_B = 1, X_C = 0] + P[X_B = 0, X_C = 0] \end{aligned}$$

Calculamos cada una de las probabilidades:

$$\begin{aligned} P[X_B = 2, X_C = 0] &= \binom{6!}{2!4!} \cdot 0{,}1875^2 \cdot 0{,}65^4 \approx 0{,}0941 \\ P[X_B = 1, X_C = 0] &= \binom{6!}{1!5!} \cdot 0{,}1875 \cdot 0{,}65^5 \approx 0{,}1305 \\ P[X_B = 0, X_C = 0] &= \binom{6!}{0!6!} \cdot 0{,}65^6 \approx 0{,}07542 \end{aligned}$$

Por tanto, la probabilidad de que un lote elegido al azar sea aceptado para su compra es:

$$P[\text{aceptado}] \approx 0{,}0941 + 0{,}1305 + 0{,}07542 = 0{,}3$$

2. El número de lotes que hay que probar para comprar al menos 20 con probabilidad mayor o igual que 0,9452.

   Sea n el número de lotes probados. Para cada $i = 1, \dots, n$, sea X_i una variable aleatoria que toma el valor 1 si el i-ésimo lote es aceptado, y 0 en caso contrario, de forma que se tiene $X_i \sim B(1, P[\text{aceptado}])$. Consideramos la variable aleatoria:

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

   Por la reproducción de la binomial, tenemos que $S_n \sim B(n, P[\text{aceptado}])$. Además, S_n es el número de lotes aceptados en n probados. Nos piden:

$$P[S_n \geqslant 20] \geqslant 0{,}9452$$

   Como se pide un resultado aproximado, usaremos el Teorema de Lévy. Tenemos que:

$$\begin{aligned} E[S_n] &= n \cdot P[\text{aceptado}] \\ \operatorname{Var}[S_n] &= n \cdot P[\text{aceptado}] \cdot P[\text{no aceptado}] \end{aligned}$$

   Por tanto, por el Teorema de Lévy, tenemos que:

$$\left\{ \frac{S_n - E[S_n]}{\sqrt{\operatorname{Var}[S_n]}} \right\} \xrightarrow{L} \mathcal{N}(0, 1)$$

Por tanto, sea Z la tipificada de la normal, tenemos que:

$$
\begin{aligned}
P[S_n \geqslant 20] \geqslant 0{,}9452 &\iff P\left[\frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \geqslant \frac{20 - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right] \geqslant 0{,}9452 \iff \\
&\iff P\left[Z \geqslant \frac{20 - n \cdot P[\text{aceptado}]}{\sqrt{n \cdot P[\text{aceptado}] \cdot P[\text{no aceptado}]}}\right] \geqslant 0{,}9452 \iff \\
&\iff P\left[Z < -\frac{20 - n \cdot P[\text{aceptado}]}{\sqrt{n \cdot P[\text{aceptado}] \cdot P[\text{no aceptado}]}}\right] \geqslant 0{,}9452 \iff \\
&\iff P\left[Z < -\frac{20 - 0{,}3n}{\sqrt{0{,}21n}}\right] \geqslant 0{,}9452
\end{aligned}
$$

Por tanto, consultando en la tabla de la normal, tenemos que:

$$
\begin{aligned}
-\frac{20 - 0{,}3n}{\sqrt{0{,}21n}} \geqslant 1{,}6 &\iff \frac{0{,}3n - 20}{\sqrt{0{,}21n}} \geqslant 1{,}6 \iff \\
&\iff 0{,}3n - 20 \geqslant 0{,}733\sqrt{n} \iff \\
&\iff 0{,}3n - 0{,}733\sqrt{n} - 20 \geqslant 0
\end{aligned}
$$

Resolvemos para el caso de la igualdad. Tenemos que:

$$
\sqrt{n} = \frac{0{,}733 \pm \sqrt{0{,}733^2 + 4 \cdot 0{,}3 \cdot 20}}{0{,}6} = \frac{0{,}733 \pm 4{,}9535}{0{,}6} \Longrightarrow \sqrt{n} = 9{,}477 \Longrightarrow n = 89{,}823
$$

Por tanto, evaluando vemos que los valores que cumplen la desigualdad han de ser mayores o iguales que 89,823. Como n ha de ser un número entero, el número de lotes que hay que probar para comprar al menos 20 con probabilidad mayor o igual que 0,9452 es 90.

Ejercicio 7.7.8. La longitud de vida (en horas) de una determinada pieza de cierta máquina es una variable aleatoria que se distribuye de acuerdo a la función de densidad

$$
f(x) = \exp(1 - x), \quad x > 1.
$$

Cuando falla la pieza, se sustituye por otra de las mismas características, y se supone que las longitudes de vida de distintas piezas son independientes. Calcular aproximadamente el número de piezas de recambio necesario para asegurar el funcionamiento de la máquina al menos durante 1000 horas, con probabilidad mayor que 0,95.

Sea n el número de piezas de recambio necesarias. Para cada $i = 1, \ldots, n$, sea X_i una variable aleatoria que modela la longitud de vida de la i-ésima pieza de recambio. Consideramos la variable aleatoria:

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

Vemos que S_n es la longitud de vida total de las n piezas de recambio. Nos piden:

$$
P[S_n \geqslant 1000] \geqslant 0{,}95
$$

Como se pide un resultado aproximado, usaremos el Teorema de Lévy. Tenemos que:

$$\begin{aligned}
E[X_1] &= \int_1^\infty x \cdot \exp(1-x)\ dx = \int_1^\infty x \cdot \exp(1) \cdot \exp(-x)\ dx = e \cdot \int_1^\infty x \cdot e^{-x}\ dx = \\
&= e \cdot \left[-(x+1)e^{-x}\right]_1^\infty = e \cdot 2e^{-1} = 2 \\
E[X_1^2] &= \int_1^\infty x^2 \cdot \exp(1-x)\ dx = \int_1^\infty x^2 \cdot \exp(1) \cdot \exp(-x)\ dx = e \cdot \int_1^\infty x^2 \cdot e^{-x}\ dx = \\
&= e \cdot \left[-(x^2+2x+2)e^{-x}\right]_1^\infty = e \cdot 5e^{-1} = 5 \\
\mathrm{Var}[X_1] &= E[X_1^2] - E[X_1]^2 = 5 - 4 = 1 \\
E[S_n] &= n \cdot E[X_1] = 2n \\
\mathrm{Var}[S_n] &= n \cdot \mathrm{Var}[X_1] = n
\end{aligned}$$

Por tanto, por el Teorema de Lévy, tenemos que:

$$\left\{\frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right\} \xrightarrow{L} \mathcal{N}(0,1)$$

Por tanto, sea Z la tipificada de la normal, tenemos que:

$$\begin{aligned}
P[S_n \geqslant 1000] \geqslant 0{,}95 &\Longleftrightarrow P\left[\frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \geqslant \frac{1000 - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right] \geqslant 0{,}95 \Longleftrightarrow \\
&\Longleftrightarrow P\left[Z \geqslant \frac{1000-2n}{\sqrt{n}}\right] \geqslant 0{,}95 \Longleftrightarrow \\
&\Longleftrightarrow P\left[Z < -\frac{1000-2n}{\sqrt{n}}\right] \geqslant 0{,}95
\end{aligned}$$

Por tanto, consultando en la tabla de la normal, tenemos que:

$$\begin{aligned}
-\frac{1000-2n}{\sqrt{n}} \geqslant 1{,}645 \Longleftrightarrow \frac{2n-1000}{\sqrt{n}} &\geqslant 1{,}645 \Longleftrightarrow \\
&\Longleftrightarrow 2n - 1000 \geqslant 1{,}645\sqrt{n} \Longleftrightarrow \\
&\Longleftrightarrow 2n - 1{,}645\sqrt{n} - 1000 \geqslant 0
\end{aligned}$$

Resolvemos para el caso de la igualdad. Tenemos que:

$$\sqrt{n} = \frac{1{,}645 \pm \sqrt{1{,}645^2 + 4 \cdot 2 \cdot 1000}}{4} = \frac{1{,}645 \pm 89{,}4578}{4} \Longrightarrow \sqrt{n} = 22{,}7757 \Longrightarrow n = 518{,}73$$

Por tanto, evaluando vemos que los valores que cumplen la desigualdad han de ser mayores o iguales que 518,73. Como n ha de ser un número entero, el número de piezas de recambio necesario para asegurar el funcionamiento de la máquina al menos durante 1000 horas, con probabilidad mayor que 0,95 es 519.