

Probabilidad

FACULTAD
DE
CIENCIAS
UNIVERSIDAD DE GRANADA



Los Del DGIIM, losdeldgiim.github.io

Doble Grado en Ingeniería Informática y Matemáticas
Universidad de Granada



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional (CC BY-NC-ND 4.0).

Eres libre de compartir y redistribuir el contenido de esta obra en cualquier medio o formato, siempre y cuando des el crédito adecuado a los autores originales y no persigas fines comerciales.

Probabilidad

Los Del DGIIM, losdeldgiim.github.io

Arturo Olivares Martos

Granada, 2024-2025

Índice general

1. Distribuciones de Probabilidad Continua	5
1.1. Distribución Uniforme Continua	5
1.2. Distribución Normal	8
1.2.1. Aproximaciones	13
1.3. Distribución Exponencial	14
1.3.1. Relación con la Distribución Poisson	17
1.4. Distribución de Erlang	18
1.5. Distribución Gamma	19
1.5.1. Función Gamma	19
1.5.2. Distribución Gamma	20
1.6. Distribución Beta	22
1.6.1. Función Beta	23
1.6.2. Distribución Beta	23
2. Vectores Aleatorios	27
2.1. Clasificación de vectores aleatorios	30
2.1.1. Vectores aleatorios discretos	31
2.1.2. Vectores aleatorios continuos	33
2.2. Distribuciones marginales	34
2.3. Distribuciones condicionadas	36
2.4. Cambio de Variable	37
2.4.1. Discreto a Discreto	38
2.4.2. Continuo a Discreto	38
2.4.3. Continuo a Continuo	39
2.4.4. Distribución del Máximo y del Mínimo	39
2.5. Esperanza	41
2.6. Momentos	41
2.6.1. Momentos No Centrados	41
2.6.2. Momentos Centrados	42
2.7. Función Generatriz de Momentos	43
3. Relaciones de problemas	45

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 para variables aleatorias discretas, en este tema se presentarán las 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:

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

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:

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

Por tanto, la función de densidad de X es:

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

□

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

$$F_X: \mathbb{R} \longrightarrow [0, 1]$$

$$x \longmapsto \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b \end{cases}$$

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 $a \leq c \leq d \leq b$, 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:

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

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:

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

Para $t \neq 0$, tenemos que:

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

□

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:

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

□

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:

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

Demostración. Tenemos que:

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

Distinguimos ahora en función de la paridad de k :

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

□

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:

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

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

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

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x) dx &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \stackrel{(*)}{=} \\ &\stackrel{(*)}{=} \frac{1}{\sqrt{2\pi}\sigma} \cdot \sqrt{\frac{\pi}{\frac{1}{2\sigma^2}}} = 1 \end{aligned}$$

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

□

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

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

$$2. P[X \leq x] = P\left[Z \leq \frac{x - \mu}{\sigma}\right] \text{ para todo } x \in \mathbb{R}.$$

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. Tenemos que $Re_X = \mathbb{R}$, y definimos por comodidad la siguiente función:

$$\begin{aligned} g: \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{x - \mu}{\sigma} \end{aligned}$$

Tenemos por tanto que $Z = g(X)$, y como g es biyectiva tenemos que $Re_Z = Re_X = \mathbb{R}$. La inversa de g es:

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

La función de densidad de Z es, por tanto:

$$\begin{aligned} f_Z(z) &= f_X(g^{-1}(z)) |(g^{-1})(z)| = f_X(\sigma z + \mu) \cdot \sigma = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}} \cdot \sigma = \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \end{aligned}$$

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

2. Para demostrar esto, tenemos que:

$$\begin{aligned} P[X \leq x] &= \int_{-\infty}^x f_X(t) dt = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t - \mu)^2}{2\sigma^2}} dt \\ P\left[Z \leq \frac{x - \mu}{\sigma}\right] &= \int_{-\infty}^{\frac{x - \mu}{\sigma}} f_Z(t) dt = \int_{-\infty}^{\frac{x - \mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \end{aligned}$$

Resolvamos la primera integral mediante el Cambio de Variable siguiente:

$$\left[\begin{array}{l} t = \sigma u + \mu \\ \frac{dt}{du} = \sigma \end{array} \right] \quad \left\{ \begin{array}{l} \text{Cuando } t = -\infty, u = -\infty \\ \text{Cuando } t = x, u = \frac{x - \mu}{\sigma} \end{array} \right.$$

Por tanto, tenemos que:

$$\begin{aligned} P[X \leq x] &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t - \mu)^2}{2\sigma^2}} dt \stackrel{(*)}{=} \int_{-\infty}^{\frac{x - \mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \cdot \sigma du = \\ &= \int_{-\infty}^{\frac{x - \mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = P\left[Z \leq \frac{x - \mu}{\sigma}\right] \end{aligned}$$

donde en $(*)$ hemos empleado el cambio de variable. \square

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

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

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 $X \sim \mathcal{N}(\mu, \sigma^2)$. Tenemos:

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

que

Por tanto, tenemos que:

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

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

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

□

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 \bar{X} o μ):

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

$$E[X] = \mu$$

Demostración. Tenemos que:

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

□

De igual forma, podemos ver que σ^2 es la varianza de la variable aleatoria:

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

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

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

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

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

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

□

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:

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

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

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

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

$$\begin{aligned} P[X \geq \mu + x] &= 1 - P[X \leq \mu + x] = 1 - \int_{-\infty}^{\mu+x} f_X(t) dt = \\ &= \int_{-\infty}^{+\infty} f_X(t) dt - \int_{-\infty}^{\mu+x} f_X(t) dt = \int_{\mu+x}^{+\infty} f_X(t) dt \end{aligned}$$

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

$$\left[\begin{array}{l} t = \mu + u \\ \frac{dt}{du} = 1 \end{array} \right] \quad \left\{ \begin{array}{l} \lim_{u \rightarrow x} t = \mu + x \\ \lim_{u \rightarrow \infty} t = +\infty \end{array} \right.$$

Por tanto, tenemos que:

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

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

$$\left[\begin{array}{l} u = -v + \mu \\ \frac{du}{dv} = -1 \end{array} \right] \quad \left\{ \begin{array}{l} \lim_{v \rightarrow \mu-x} t = x \\ \lim_{v \rightarrow -\infty} t = +\infty \end{array} \right.$$

Por tanto, tenemos que:

$$\begin{aligned} P[X \geq \mu + x] &= \int_x^{\infty} f_X(\mu - u) du = - \int_{\mu-x}^{-\infty} f_X(v) dv = \int_{-\infty}^{\mu-x} f_X(v) dv = \\ &= P[X \leq \mu - x] \end{aligned}$$

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. \square

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 $P[X \leq \mu] = P[X \geq \mu] = 1/2$.

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

$$P[X \geq \mu] = 1 - P[X \leq \mu] = 1 - P[X \geq \mu] \implies P[X \geq \mu] = \frac{1}{2}$$

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

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

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

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, $\mu = Mo[X]$. \square

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. $P[\mu - \sigma \leq X \leq \mu + \sigma] \approx 0,6826$
2. $P[\mu - 2\sigma \leq X \leq \mu + 2\sigma] \approx 0,9544$
3. $P[\mu - 3\sigma \leq X \leq \mu + 3\sigma] \approx 0,9974$

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

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

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

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

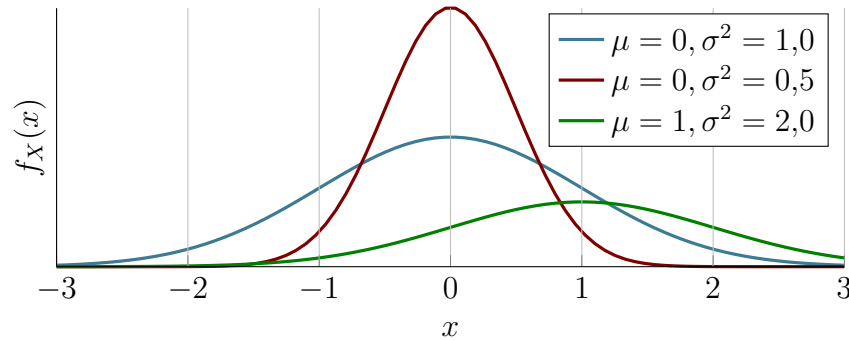


Figura 1.1: Función de densidad de una v. a. con distribución normal.

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.

Proposición 1.13 (Aproximación de la Binomial por la Normal). *Sea X una variable aleatoria tal que $X \sim B(n, p)$. Entonces, si n es suficientemente grande y p*

no está muy cerca de 0 o 1, se tiene que X se puede aproximar por una distribución normal con parámetros $\mu = np$ y $\sigma^2 = np(1 - p)$. Es decir:

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

Empíricamente se ha comprobado que esta aproximación es buena si $n \geq 30$ y $0,1 \leq p \leq 0,9$.

Proposición 1.14 (Aproximación de la Poisson por la Normal). Sea X una variable aleatoria tal que $X \sim P(\lambda)$. Entonces, si λ es suficientemente grande, se tiene que X se puede aproximar por una distribución normal con parámetros $\mu = \lambda$ y $\sigma^2 = \lambda$. Es decir:

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

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

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 aproximarla 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 $P[x_i - 0,5 \leq X \leq x_i + 0,5]$ en la normal.
- Para aproximar $P[X \leq x_i]$ en la binomial, se calculará $P[X \leq x_i + 0,5]$ en la normal.
- Para aproximar $P[X \geq x_i]$ en la binomial, se calculará $P[X \geq x_i - 0,5]$ 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:

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

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:

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

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

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

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

□

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:

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

Demostración. Distinguimos dos casos:

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

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

□

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

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

Demostración. Tenemos que:

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

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 tal 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:

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

■ Caso base: $k = 1$.

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

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

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

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:

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

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:*

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

Demostración. Tenemos que:

$$\begin{aligned} P(X \geq t + s \mid X \geq s) &= \frac{P(X \geq t + s, X \geq s)}{P(X \geq s)} = \frac{P(X \geq t + s)}{P(X \geq s)} \stackrel{(*)}{=} \\ &\stackrel{(*)}{=} \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} \stackrel{(*)}{=} P(X \geq t) \end{aligned}$$

donde en $(*)$ hemos usado que:

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

□

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.

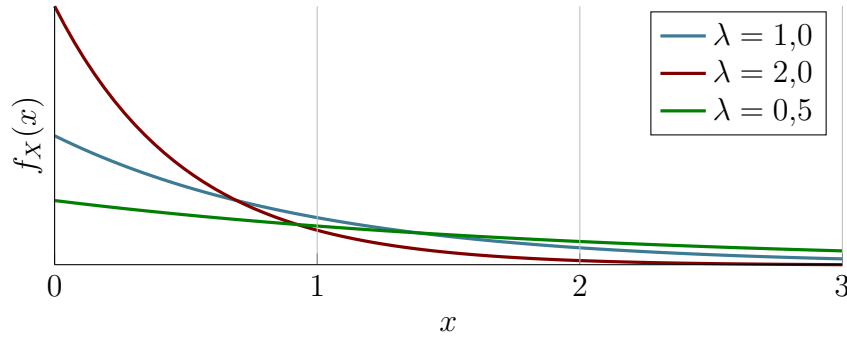


Figura 1.2: Función de densidad de una v. a. con distribución exponencial.

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:

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

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 ($P[X \geq t]$).

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:

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

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

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

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

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

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

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

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

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

- 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 $n+1$:

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

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

Por tanto, tenemos que:

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

□

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

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

Demostración. Tenemos que:

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

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

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

Por tanto, tenemos que:

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

□

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:

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

Algunas propiedades son:

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

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

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

Mediante integración por partes, tenemos que:

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

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

Se deduce directamente de las dos propiedades anteriores.

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

Hacemos el cambio de variable $t = u/\lambda$:

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

5. $\Gamma(1/2) = \sqrt{\pi}.$

$$\begin{aligned} \Gamma(1/2) &= \int_0^\infty t^{-1/2} e^{-t} dt = \left[\begin{array}{l} t = u^2/2 \\ \frac{dt}{du} = u \end{array} \right] = \int_0^\infty \sqrt{2} \cdot u^{-1} \cdot e^{-u^2/2} \cdot u du = \\ &= \sqrt{2} \int_0^\infty e^{-u^2/2} du \end{aligned}$$

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

$$\begin{aligned} \Gamma(1/2) &= \sqrt{2} \int_0^\infty e^{-u^2/2} du = \frac{\sqrt{2}}{2} \int_{-\infty}^\infty e^{-u^2/2} du = \frac{1}{\sqrt{2}} \int_{-\infty}^\infty e^{-u^2/2} du = \\ &= \frac{\sqrt{\pi}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-u^2/2} du = \sqrt{\pi} \cdot \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \stackrel{(*)}{=} \sqrt{\pi} \end{aligned}$$

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:

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

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:

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

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

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

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x) dx &= \int_0^{\infty} \frac{\lambda^u}{\Gamma(u)} x^{u-1} e^{-\lambda x} dx = \frac{\lambda^u}{\Gamma(u)} \int_0^{\infty} x^{u-1} e^{-\lambda x} dx \stackrel{(*)}{=} \\ &\stackrel{(*)}{=} \frac{\lambda^u}{\Gamma(u)} \cdot \frac{\Gamma(u)}{\lambda^u} = 1 \end{aligned}$$

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:

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

Demostración. Tenemos que:

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

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:

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

- 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 $k + 1$:

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

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

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

Opción 2 Usar la definición de los momentos no centrados.

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

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

□

Como consecuencia, tenemos que:

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

Tiene muchas aplicaciones en experimentos o fenómenos aleatorios que tienen asociadas v.a. 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:

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

Algunas propiedades son:

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

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

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

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:

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

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:

- $f_X(x) \geq 0$ 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.

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

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

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:*

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

Demostración. Tenemos que:

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

□

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 $q \geq 1$, es decreciente y cóncava, mientras que si $p \geq 1$ 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 σ -álgebra de Borel en \mathbb{R}^n .

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

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

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

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:

$$x \leq y \iff x_i \leq y_i, \quad \forall i = 1, \dots, n.$$

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

Gráficamente, en el plano tenemos que $x \leq x'$ si y solo si x está a la izquierda y por debajo de x' .



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

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 (Ω, \mathcal{A}, P) se define como una función medible:

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

tal que se cumple que:

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

Es decir:

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

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 $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$. 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:

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

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 $P[X \in B]$.

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. No negatividad: $P_X(B) = P(X^{-1}(B)) \geq 0$, $\forall B \in \mathcal{B}^n$ ya que P es una medida de probabilidad.
2. Suceso seguro: $P_X(\mathbb{R}^n) = P(X^{-1}(\mathbb{R}^n)) = P(\Omega) = 1$.
3. σ -aditividad: Sean $B_1, B_2, \dots \in \mathcal{B}^n$ disjuntos dos a dos. Entonces, como $X^{-1}(B_1), X^{-1}(B_2), \dots$ son disjuntos dos a dos, se tiene:

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

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

Así, tenemos que todo vector aleatorio X transforma el espacio de probabilidad (Ω, \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:

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

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

$$F_X(x) = P[X_1 \leq x_1, \dots, X_n \leq x_n].$$

Algunas de las propiedades que cumple esta son:

1. Es monótona no decreciente 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:

$$x_i \leq x'_i \implies F_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) \leq F_X(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n).$$

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:

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

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:

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

4. Tenemos que:

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

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

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

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:

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

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.

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

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:

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

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 (Ω, \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 $\exists A \subset \mathbb{R}^n$ 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.

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

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

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

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

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

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

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

\Leftarrow) 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 $E_{X_1} \times \dots \times E_{X_n} \subset \mathbb{R}^n$ numerable, y tenemos que:

$$\begin{aligned} P[X \in E_{X_1} \times \dots \times E_{X_n}] &= P[X_1 \in E_{X_1}, \dots, X_n \in E_{X_n}] = P\left[\bigcap_{i=1}^n X_i \in E_{X_i}\right] \geq \\ &\geq 1 - \sum_{i=1}^n P[X_i \notin E_{X_i}] = 1. \end{aligned}$$

donde hemos usado la desigualdad de Boole. Por tanto, como las probabilidades son menores o iguales que 1, tenemos que $P[X \in E_{X_1} \times \dots \times E_{X_n}] = 1$, 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:

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

Esta satisface:

1. $p_X(x) \geq 0$.
2. $\sum_{x \in E_X} p_X(x) = 1$.

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

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

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

$$\begin{aligned} X_1 &= \begin{cases} -1 & \text{si sale impar} \\ 1 & \text{si sale par} \end{cases} \\ X_2 &= \begin{cases} -2 & \text{si sale 1, 2, 3} \\ 0 & \text{si sale 4} \\ 3 & \text{si sale 5, 6} \end{cases} \end{aligned}$$

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

1. Calcular la función masa de probabilidad de X .

Tenemos que:

$$\begin{aligned} E_{X_1} &= \{-1, 1\}, \\ E_{X_2} &= \{-2, 0, 3\}, \end{aligned}$$

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:

$$F_X(x_1, x_2) = \begin{cases} 0 & x_1 < -1 \text{ o } x_2 < -2 \\ 2/6 & x_1 \in [-1, 1[, x_2 \in [-2, 3[\\ 3/6 & x_1 \in [-1, 1], x_2 \geq 3 \\ 3/6 & x_1 \geq 1, x_2 \in [-2, 0[\\ 4/6 & x_1 \geq 1, x_2 \in [0, 3[\\ 1 & x_1 \geq 1, x_2 \geq 3 \end{cases}$$

3. Calcular $P[X_1 + X_2 \leq 1]$.

En este caso, los valores de $X_1 + X_2$ que cumplen que $X_1 + X_2 \leq 1$ son:

$$B = \{(-1, -2), (1, -2), (1, 0)\}.$$

Por tanto,

$$\begin{aligned} P[X_1 + X_2 \leq 1] &= P[X \in B] = P[X = (-1, -2)] + P[X = (1, -2)] + P[X = (1, 0)] = \\ &= 2/6 + 1/6 + 1/6 = 4/6. \end{aligned}$$

2.1.2. Vectores aleatorios continuos

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

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

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

Además, si f_X es continua en un punto $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, entonces 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. $f_X(x) \geq 0$.
2. Es integrable en \mathbb{R}^n .
3. $\int_{\mathbb{R}^n} f_X(x) dx = 1$.

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

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.

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

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

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

Definimos por tanto:

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

Por tanto, tenemos que:

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

Como f_i es integrable y no negativa, tenemos que es la función de densidad de probabilidad de X_i . Por tanto, X_i es una variable aleatoria continua. \square

2.2. Distribuciones marginales

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

Definición 2.9. Sea $X = (X_1, \dots, X_n)$ 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 $X = (X_1, \dots, X_n)$ un vector aleatorio con función de distribución F_X . Entonces, la función de distribución de X_i es:

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

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:

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

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

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

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

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

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

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

2.3. Distribuciones condicionadas

Dado un vector aleatorio $X = (X_1, \dots, X_n)$, 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 $X = (X_1, \dots, X_n)$ 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 $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ a $X_i = x_i^*$ es la distribución con función de masa de probabilidad:

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

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:

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

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:

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

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

Dado un vector aleatorio $X = (X_1, \dots, X_n)$, podemos estudiar la distribución de otro vector aleatorio $Y = (Y_1, \dots, Y_n)$ que depende de X mediante una transformación.

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

Demostración. Como la composición de funciones medibles es medible, tenemos que $Y = g(X)$ es medible por serlo X y g . Por tanto, Y es un vector aleatorio. □

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

Proposición 2.10. *Sea $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$ un vector aleatorio n -dimensional y $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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:

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

Veamos ahora algunos casos particulares de transformaciones de vectores aleatorios.

2.4.1. Discreto a Discreto

Sea $X = (X_1, \dots, X_n)$ un vector aleatorio discreto con valores en E_X , y sea $g : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^m, \mathcal{B}^m)$ 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:

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

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

Sea $X = (X_1, \dots, X_n)$ un vector aleatorio continuo con función de densidad f_X , y sea $g : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^m, \mathcal{B}^m)$ 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:

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

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:

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

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

$$\begin{aligned} P[Y = 1] &= P[X_1 > X_2] = \int_0^\infty \int_{x_2}^\infty \lambda \mu e^{-\lambda x_1 - \mu x_2} dx_1 dx_2 = \int_0^\infty \mu e^{-\mu x_2} \int_{x_2}^\infty \lambda e^{-\lambda x_1} dx_1 dx_2 = \\ &= \int_0^\infty \mu e^{-\mu x_2} [-e^{-\lambda x_1}]_{x_2}^\infty dx_2 = \int_0^\infty \mu e^{-\mu x_2} e^{-\lambda x_2} dx_2 = \int_0^\infty \mu e^{-(\mu+\lambda)x_2} dx_2 = \\ &= \left[-\frac{\mu}{\mu+\lambda} e^{-(\mu+\lambda)x_2} \right]_0^\infty = \frac{\mu}{\mu+\lambda}. \\ P[Y = 0] &= 1 - P[Y = 1] = 1 - \frac{\mu}{\mu+\lambda} = \frac{\lambda}{\mu+\lambda}. \end{aligned}$$

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

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

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

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

- 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:

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

2.4.4. Distribución del Máximo y del Mínimo

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

$$\min\{f_1, f_2, \dots, f_n\}, \max\{f_1, f_2, \dots, f_n\} : A \rightarrow B$$

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

$$\begin{aligned} \min\{f_1, f_2, \dots, f_n\}(x) &= \min\{f_1(x), f_2(x), \dots, f_n(x)\} \\ \max\{f_1, f_2, \dots, f_n\}(x) &= \max\{f_1(x), f_2(x), \dots, f_n(x)\} \end{aligned}$$

Dada una función $F : A^n \rightarrow B$ de n variables, podemos definir $\min F$ (análogamente $\max F$) como la función:

$$\min F = \min\{F_1, F_2, \dots, F_n\}$$

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

Proposición 2.11. *Sea $g : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^m, \mathcal{B}^m)$ una función medible, entonces $\min g$ y $\max g$ 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 $Z = \min X$ y $W = \max X$. Dado $x \in \mathbb{R}$, tenemos que:

- $Z = \min = \min X$: $Z = \min X = \min\{X_1, \dots, X_n\}$.

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

$$Z(\omega) > x \iff \min\{X_1(\omega), \dots, X_n(\omega)\} > x \iff X_1(\omega) > x, \dots, X_n(\omega) > x.$$

Por tanto, tenemos que:

$$P[\min \leq x] = 1 - P[\min > x] = 1 - P[X_1 > x, \dots, X_n > x] \quad \forall x \in \mathbb{R}.$$

- $W = \max = \max X$: $W = \max X = \max\{X_1, \dots, X_n\}$.

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

$$W(\omega) < x \iff \max\{X_1(\omega), \dots, X_n(\omega)\} < x \iff X_1(\omega) < x, \dots, X_n(\omega) < x.$$

Por tanto, tenemos que:

$$P[\max \leq x] = P[X_1 \leq x, \dots, X_n \leq x] = P[X \leq (x, \dots, x)] \quad \forall x \in \mathbb{R}.$$

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

$$(\max X, \min X)$$

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

$$P[(\max X, \min X) \leq (x, y)] = P[\max X \leq x, \min X \leq y]$$

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

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

- Si $x > y$, entonces:

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

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

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

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:

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

Usando el Teorema de Cambio de Variable en cada uno de los casos, deducimos de forma directa el siguiente teorema:

Teorema 2.12. Sea $X = (X_1, \dots, X_n)$ una variable aleatoria y $g : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 :

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

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.

Haciendo uso de la esperanza de una función de un vector aleatorio, tenemos:

- Caso discreto:

$$m_{k_1, \dots, k_n} = E[X_1^{k_1} \cdots X_n^{k_n}] = \sum_{x \in E_X} x_1^{k_1} \cdots x_n^{k_n} P[X = x].$$

- Caso continuo:

$$m_{k_1, \dots, k_n} = E[X_1^{k_1} \cdots X_n^{k_n}] = \int_{\mathbb{R}^n} x_1^{k_1} \cdots x_n^{k_n} f_X(x) dx.$$

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.

Haciendo uso de la esperanza de una función de un vector aleatorio, tenemos:

■ Caso discreto:

$$\begin{aligned} \mu_{k_1, \dots, k_n} &= E[(X_1 - E[X_1])^{k_1} \cdots (X_n - E[X_n])^{k_n}] = \\ &= \sum_{x \in E_X} (x_1 - E[X_1])^{k_1} \cdots (x_n - E[X_n])^{k_n} P[X = x]. \end{aligned}$$

■ Caso continuo:

$$\begin{aligned} \mu_{k_1, \dots, k_n} &= E[(X_1 - E[X_1])^{k_1} \cdots (X_n - E[X_n])^{k_n}] = \\ &= \int_{\mathbb{R}^n} (x_1 - E[X_1])^{k_1} \cdots (x_n - E[X_n])^{k_n} f_X(x) dx. \end{aligned}$$

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:

$$\text{Cov}(X_i, X_j) = \mu_{\substack{i \\ 0 \dots 0 \ 1 \ 0 \dots 0 \ 0 \ 1 \ 0 \dots 0}}^j = E[(X_i - E[X_i])(X_j - E[X_j])].$$

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

Lema 2.13. Sean X, Y dos variables aleatorias. Entonces:

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

Demostración. Tenemos que:

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

□

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

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

Definición 2.17 (Matriz de Covarianzas). Sea $X = (X_1, \dots, X_n)$ 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:

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

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:

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

2.7. Función Generatriz de Momentos

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

$$M_X : \begin{array}{ccc} N & \longrightarrow & \mathbb{R} \\ (t_1, \dots, t_n) & \longmapsto & E[e^{t_1 X_1 + \dots + t_n X_n}] \end{array}$$

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

Teorema 2.14 (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 2.15. *Sea $X = (X_1, \dots, X_n)$ un vector aleatorio. Entonces, si existe la función generatriz de momentos de X :*

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

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

Demostración. Supuesta cierta la primera afirmación, tenemos que:

$$\begin{aligned}
 \left. \frac{\partial^{k_1+\dots+k_n} M_X(t_1, \dots, t_n)}{\partial t_1^{k_1} \dots \partial t_n^{k_n}} \right|_{t_1=\dots=t_n=0} &= \left. \frac{\partial^{k_1+\dots+k_n} E[e^{t_1 X_1 + \dots + t_n X_n}]}{\partial t_1^{k_1} \dots \partial t_n^{k_n}} \right|_{t_1=\dots=t_n=0} = \\
 &= E \left[\left. \frac{\partial^{k_1+\dots+k_n} e^{t_1 X_1 + \dots + t_n X_n}}{\partial t_1^{k_1} \dots \partial t_n^{k_n}} \right|_{t_1=\dots=t_n=0} \right] = \\
 &= E \left[X_1^{k_1} \dots X_n^{k_n} e^{t_1 X_1 + \dots + t_n X_n} \right]_{t_1=\dots=t_n=0} = \\
 &= E [X_1^{k_1} \dots X_n^{k_n} e^0] = \\
 &= E[X_1^{k_1} \dots X_n^{k_n}].
 \end{aligned}$$

□

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:

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

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. Relaciones de problemas