

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

| | |
|--|-----------|
| 0. Introducción | 5 |
| 1. Distribuciones de Probabilidad Continua | 7 |
| 1.1. Distribución Uniforme Continua | 7 |
| 1.2. Distribución Normal | 10 |
| 1.2.1. Aproximaciones | 15 |
| 1.3. Distribución Exponencial | 16 |
| 1.3.1. Relación con la Distribución Poisson | 19 |
| 1.4. Distribución de Erlang | 20 |
| 1.5. Distribución Gamma | 21 |
| 1.5.1. Función Gamma | 21 |
| 1.5.2. Distribución Gamma | 22 |
| 1.6. Distribución Beta | 24 |
| 1.6.1. Función Beta | 25 |
| 1.6.2. Distribución Beta | 25 |
| 2. Vectores Aleatorios | 29 |
| 2.1. Clasificación de vectores aleatorios | 32 |
| 2.1.1. Vectores aleatorios discretos | 33 |
| 2.1.2. Vectores aleatorios continuos | 35 |
| 2.2. Distribuciones marginales | 36 |
| 2.3. Distribuciones condicionadas | 38 |
| 2.4. Cambio de Variable | 39 |
| 2.4.1. Discreto a Discreto | 40 |
| 2.4.2. Continuo a Discreto | 40 |
| 2.4.3. Continuo a Continuo | 41 |
| 2.4.4. Distribución del Máximo y del Mínimo | 41 |
| 2.5. Esperanza | 43 |
| 2.6. Momentos | 43 |
| 2.6.1. Momentos No Centrados | 43 |
| 2.6.2. Momentos Centrados | 44 |
| 2.7. Función Generatriz de Momentos | 45 |
| 3. Independencia de Vectores Aleatorios | 47 |
| 3.1. Caracterizaciones de independencia para variables discretas | 47 |
| 3.2. Caracterizaciones de independencia para variables continuas | 49 |
| 3.3. Caracterización mediante conjuntos de Borel | 50 |
| 3.4. Propiedades de la independencia | 50 |

| | | |
|-----------|--|-----------|
| 3.4.1. | Teorema de la multiplicación de las esperanzas | 51 |
| 3.5. | Distribuciones Reproductivas | 52 |
| 3.6. | Independencia para familias de variables aleatorias | 56 |
| 3.7. | Independencia para vectores aleatorios | 57 |
| 4. | Esperanza Condicionada | 59 |
| 4.1. | Esperanza condicionada | 59 |
| 4.2. | Momentos condicionados | 65 |
| 4.2.1. | Momentos condicionados no centrados | 65 |
| 4.2.2. | Momentos condicionados centrados | 65 |
| 4.3. | Varianza condicionada | 66 |
| 4.4. | Regresión Mínimo Cuadrática | 68 |
| 4.4.1. | Búsqueda de la función de regresión óptima, φ_{opt} | 68 |
| 4.5. | Razones de Correlación | 72 |
| 5. | Relaciones de problemas | 77 |
| 5.0. | Introducción | 77 |

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 diriga 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 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:

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

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:

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

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 $t = \mu + u$:

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

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 $n+1$:

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

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' .

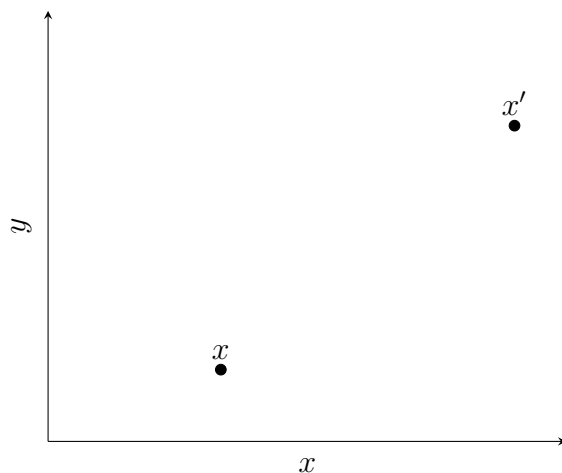


Figura 2.1: Relación de orden en \mathbb{R}^2 , donde $x \leq x'$.

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:

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

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:

$$\begin{aligned} \int_{\mathbb{R}^{n-1}} \frac{f_X(x_1, \dots, x_{i-1}, x_i^*, x_{i+1}, \dots, x_n)}{f_{X_i}(x_i^*)} dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n &= \frac{\int_{\mathbb{R}^{n-1}} f_X(x_1, \dots, x_{i-1}, x_i^*, x_{i+1}, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n}{f_{X_i}(x_i^*)} \stackrel{(*)}{=} \\ &\stackrel{(*)}{=} \frac{f_{X_i}(x_i^*)}{f_{X_i}(x_i^*)} = 1 \end{aligned}$$

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 $Y = (Y_1, \dots, Y_m)$ que depende de X mediante una transformación.

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

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

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. Independencia de Vectores Aleatorios

Estudiemos ahora la independencia de vectores aleatorios, que puede recordarnos a independencia de sucesos en estadística descriptiva.

Definición 3.1 (Independencia de vectores aleatorios). Sean X_1, \dots, X_n variables aleatorias definidas sobre el mismo espacio de probabilidad, con funciones de distribución F_{X_1}, \dots, F_{X_n} y función de distribución conjunta F_X . Se dice que dichas variables son mutuamente independientes (o, simplemente, independientes) si:

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

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, \dots, X_n variables aleatorias discretas definidas sobre el mismo espacio de probabilidad. Entonces, X_1, \dots, X_n son independientes si y solo si:

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

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

| $X_1 \backslash X_2$ | -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, \dots, X_n variables aleatorias discretas definidas sobre el mismo espacio de probabilidad. Entonces, X_1, \dots, X_n son independientes si y solo si:

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

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

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:

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

Comprobemos que X_1 y X_2 son independientes.

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

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

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:

$$h_1 : \mathbb{R} \longrightarrow \mathbb{R}$$

$$x_1 \longmapsto \begin{cases} \frac{1}{2^{x_1+1}} & \text{si } x_1 \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$$

$$h_2 : \mathbb{R} \longrightarrow \mathbb{R}$$

$$x_2 \longmapsto \begin{cases} 1 & \text{si } x_2 \in \{0, 1\} \\ 0 & \text{en otro caso} \end{cases}$$

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:*

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

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:*

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

siendo $h_i : \mathbb{R} \rightarrow \mathbb{R}$ 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:*

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

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:

$$P[X \leq x] = \begin{cases} 0 & \text{si } x < c \\ 1 & \text{si } x \geq c \end{cases} \quad P[X \leq x, Y \leq y] = \begin{cases} 0 & \text{si } x < c \\ P[Y \leq y] & \text{si } x \geq c \end{cases}$$

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

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

Proposición 3.8. *Sean X_1, \dots, 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, \dots, X_n variables aleatorias sobre el mismo espacio de probabilidad, y consideramos funciones medibles $g_i : \mathbb{R} \rightarrow \mathbb{R}$ para $i = 1, \dots, n$. Entonces:*

$$X_1, \dots, X_n \text{ son independientes} \implies g_1(X_1), \dots, g_n(X_n) \text{ son independientes}$$

Proposición 3.10 (Caracterización por funciones generatrices de momentos). *Sean X_1, \dots, 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 $i = 1, \dots, n$ en un entorno de 0. Entonces, X_1, \dots, X_n son independientes si y solo si:*

$$\begin{cases} \exists M_X(t_1, \dots, t_n) & \forall (t_1, \dots, t_n) \in I_1 \times \dots \times I_n \\ M_X(t_1, \dots, t_n) = M_{X_1}(t_1) \cdot \dots \cdot M_{X_n}(t_n) & \forall t_i \in I_i \end{cases}$$

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, \dots, X_n variables aleatorias sobre el mismo espacio de probabilidad, donde suponemos que $\exists E[X_i]$ para todo $i = 1, \dots, n$. Entonces:*

$$X_1, \dots, X_n \text{ son independientes} \implies \begin{cases} \exists E[X_1 \cdot \dots \cdot X_n] \\ \wedge \\ E[X_1 \cdot \dots \cdot X_n] = E[X_1] \cdot \dots \cdot E[X_n] \end{cases}$$

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

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

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:*

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

Demostración. Tenemos que:

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

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

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

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

Demostración.

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

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

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

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

Demostración.

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

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

3.5. Distribuciones Reproductivas

Definición 3.2 (Reproductividad de Variables Aleatorias). Una familia X_1, X_2, \dots, X_n de variables aleatorias independientes con el mismo tipo de distribución se dice que es reproductiva si la variable aleatoria $X = \sum_{i=1}^n X_i$ sigue una distribución del mismo tipo.

Observación. Notemos que, para ver que una familia de variables aleatorias es reproductiva, usaremos el Corolario 3.11.4 para calcular la función generatriz de momentos de la variable aleatoria $X = \sum_{i=1}^n X_i$.

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:

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

Usando el Corolario 3.11.4, tenemos que:

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

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

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

\square

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:

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

Usando el Corolario 3.11.4, tenemos que:

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

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

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

Usando el Corolario 3.11.4, tenemos que:

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

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

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

Usando el Corolario 3.11.4, tenemos que:

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

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:

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

Usando el Corolario 3.11.4, tenemos que:

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

Esta es la función generatriz de momentos de una variable aleatoria que sigue una distribución normal con parámetros $\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$. 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, \dots, X_n variables aleatorias independientes tal que $X_i \sim \exp(\lambda)$ para $i = 1, \dots, n$. Entonces:

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

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

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

Usando el Corolario 3.11.4, tenemos que:

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

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:

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

□

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

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

Usando el Corolario 3.11.4, tenemos que:

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

Esta es la función generatriz de momentos de una variable aleatoria que sigue una distribución de Erlang con parámetros $\left(\sum_{i=1}^n k_i, \lambda\right)$. 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, \dots, X_n variables aleatorias independientes tal que $X_i \sim \Gamma(u_i, \lambda)$ para $i = 1, \dots, n$. 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:

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

Usando el Corolario 3.11.4, tenemos que:

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

Esta es la función generatriz de momentos de una variable aleatoria que sigue una distribución gamma con parámetros $\left(\sum_{i=1}^n u_i, \lambda\right)$. 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 $E[X | Y]$, como la variable aleatoria siguiente (en el caso de que exista):

$$E[X | Y](y) := E[X | Y = y]$$

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:

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

- Si X e Y son continuas, entonces:

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

Notemos que la definición de $E[X | Y = y]$ 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 $E[X | Y]$ 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:*

$$\exists E[X] \implies \exists E[X | Y]$$

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

- Si X e Y son discretas, entonces:

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

donde en la desigualdad hemos hecho uso de que

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

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

- Si X e Y son continuas, entonces:

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

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

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

Por tanto, $E[X | Y]$ existe. □

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 $E[X | Y]$ y $E[Y | X]$.

Tenemos que:

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

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

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:

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

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

$$= \frac{1}{1-x} \left(\frac{1}{2} - \frac{x^2}{2} \right) = \frac{1}{2} \cdot \frac{1-x^2}{1-x} = \frac{1+x}{2}$$

Al igual que considerábamos $E[g(X)]$, consideramos ahora la esperanza condicionada de una función de X a Y .

Definición 4.2 (Esperanza condicionada de una función). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, y una función $g : \mathbb{R} \rightarrow \mathbb{R}$ medible, se define la esperanza condicionada de $g(X)$ a Y , notada por $E[g(X) | Y]$, como la variable aleatoria siguiente (en el caso de que exista):

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

Veamos en función de si es discreta o continua:

- Si X e Y son discretas, entonces:

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

- Si X e Y son continuas, entonces:

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

Al igual que ocurría con la esperanza condicionada de X a Y , para que exista basta con imponer que $E[g(X)] < \infty$. Es decir, la demostración de la siguiente proposición es análoga a la del proposición anterior.

Proposición 4.2. Sean dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, y una función $g : \mathbb{R} \rightarrow \mathbb{R}$ medible. Entonces:

$$\exists E[g(X)] \implies \exists E[g(X) | Y]$$

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 $E[X^2 | Y]$.

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 | $1/2^3$ | $1/2^3$ |
| $P[Y = y_j]$ | | | $3/2^2 \quad 1/2^2 \quad 1$ |

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

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

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

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

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

$$E[c | Y] = c$$

Demostración. Como Y es discreta, entonces:

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

□

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

$$E[(aX + b) | Y] = aE[X | Y] + b \quad \forall a, b \in \mathbb{R}$$

Proposición 4.5. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Entonces:*

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

Proposición 4.6. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Si $X \geq 0$, entonces:*

$$E[X | Y] \geq 0 \wedge E[X | Y] = 0 \iff P[X = 0] = 1$$

Proposición 4.7 (Conservación del Orden). *Sean X_1, X_2 e Y tres variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Si $X_1 \leq X_2$, entonces:*

$$E[X_1 | Y] \leq E[X_2 | Y]$$

Proposición 4.8. *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) , y consideramos una función $g : \mathbb{R} \rightarrow \mathbb{R}$ medible. Entonces:*

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

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

$$\text{Si } X, Y \text{ independientes} \implies E[X | Y] = E[X]$$

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:

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

□

Proposición 4.9. Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) , y consideramos una función $g : \mathbb{R} \rightarrow \mathbb{R}$ 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:

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

□

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 4.3 (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):

$$E[X^k | Y]$$

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

- Si X e Y son discretas, entonces:

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

- Si X e Y son continuas, entonces:

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

4.2.2. Momentos condicionados centrados

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

Definición 4.4 (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):

$$E[(X - E[X | Y])^k | Y]$$

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 $E[X | Y]$ toma el valor $E[X | Y = y]$ y, por tanto:

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

Entonces, ya que dado $Y = y$, $E[X | Y = y]$ es una constante, la variable $(X - E[X | Y = y])^k$ 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:

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

- Si X e Y son continuas, entonces:

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

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 4.5. 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 :

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

Veamos ahora algunas de sus propiedades.

Proposición 4.10. Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Si $\exists E[X^2]$, entonces:

1. $\exists \text{Var}[X | Y]$ y $\text{Var}[X | Y] \geq 0$.
2. $\text{Var}[X | Y] = E[X^2 | Y] - (E[X | Y])^2$.

Demostración. Demostramos cada uno de los apartados:

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

Por otro lado, como $E[(X - E[X | Y])^2 | Y]$, es la esperanza condicionada de una variable aleatoria no negativa, entonces $\text{Var}[X | Y] \geq 0$.

2. Partiendo de la definición de varianza condicionada, fijado $Y = y$, tenemos:

$$\begin{aligned} \text{Var}[X | Y = y] &= E[(X - E[X | Y = y])^2 | Y = y] = \\ &= E[X^2 - 2XE[X | Y = y] + (E[X | Y = y])^2 | Y = y] \stackrel{(*)}{=} \\ &\stackrel{(*)}{=} E[X^2 | Y = y] - 2E[E[X | Y = y] \cdot X | Y = y] + E[(E[X | Y = y])^2 | Y = y] \stackrel{(**)}{=} \\ &\stackrel{(**)}{=} E[X^2 | Y = y] - 2E[X | Y = y] \cdot E[X | Y = y] + (E[X | Y = y])^2 = \\ &= E[X^2 | Y = y] - 2(E[X | Y = y])^2 + (E[X | Y = y])^2 = \\ &= E[X^2 | Y = y] - (E[X | Y = y])^2 \end{aligned}$$

donde en $(*)$ hemos usado la linealidad de la esperanza condicionada, y en $(**)$ hemos usado que, como ya hemos condicionado a $Y = y$, $E[X | Y = y]$ es una constante.

□

Introducimos además esta última proposición, que cobrará gran importancia en la siguiente sección.

Proposición 4.11 (Descomposición de la varianza). *Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Si $\exists E[X^2]$, entonces se tiene que $\exists \text{Var}[E[X | Y]]$ y $\exists E[\text{Var}[X | Y]]$, y se cumple que:*

$$\text{Var}[X] = \text{Var}[E[X | Y]] + E[\text{Var}[X | Y]]$$

Demostración. En primer lugar, teneos que $\exists E[X^2]$ implica $\exists E[E[X^2 | Y]]$. Por tanto, del segundo apartado de la proposición anterior, tenemos que:

$$0 \leq \text{Var}[X | Y] = E[X^2 | Y] - (E[X | Y])^2 \implies 0 \leq (E[X | Y])^2 \leq E[X^2 | Y]$$

Por tanto, como $(E[X | Y])^2$ es una variable aleatoria acotada por variables aleatorias con esperanza, entonces $\exists E[(E[X | Y])^2]$. Tenemos por tanto, que:

- $\exists E[(E[X | Y])^2] \implies \exists \text{Var}[E[X | Y]] = E[(E[X | Y])^2] - (E[E[X | Y]])^2$.
- $\exists E[(E[X | Y])^2]$ y $\exists E[E[X^2 | Y]]$, lo que implica que:

$$\exists E[E[X^2 | Y]] - E[(E[X | Y])^2] = E[E[X^2 | Y] - (E[X | Y])^2] = E[\text{Var}[X | Y]]$$

Por tanto, hemos demostrado las dos existencias. Para demostrar la igualdad, tenemos que:

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

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 4.6 (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 4.12. Sea Y una variable aleatoria sobre el espacio de probabilidad (Ω, \mathcal{A}, P) , y X una variable aleatoria sobre el mismo espacio de probabilidad. Entonces, $E[Y | X]$ minimiza el E.C.M. Es decir:

$$\varphi_{\text{opt}}(X) = E[Y | X]$$

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:

$$\begin{aligned} \text{E.C.M.}(\varphi_{\text{opt}}) &= \text{E.C.M.}(E[Y | X]) = E[(Y - E[Y | X])^2] = \\ &= E[E[(Y - E[Y | X])^2 | X]] = E[\text{Var}[Y | X]] \end{aligned}$$

Considerando la descomposición de la varianza, tenemos:

$$\text{Var}[Y] = \text{Var}[E[Y | X]] + E[\text{Var}[Y | X]] = \text{Var}[\varphi_{\text{opt}}(X)] + \text{E.C.M.}(\varphi_{\text{opt}})$$

Definición 4.7 (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 :

$$\hat{Y}(x) = E[Y | X = x] \quad \forall x \in E_X$$

- Curva de Regresión Mínimo Cuadrática de X sobre Y :

$$\hat{X}(y) = E[X | Y = y] \quad \forall y \in E_Y$$

Veamos algunos casos particulares, para los que antes debemos introducir las siguientes definiciones.

Definición 4.8 (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 $f : E_X \rightarrow E_Y$ tal que:

$$Y = f(X) \quad \forall x \in E_X$$

Dicha función f se denomina curva de dependencia.

Proposición 4.13. Sean X, Y dos variables aleatorias sobre el mismo espacio de probabilidad.

$$\text{Var}[Y | X] = 0 \iff Y \text{ depende funcionalmente de } X$$

Demostración. Demostramos mediante la doble implicación:

\implies) Partimos de:

$$\text{Var}[Y | X = x] = 0 \quad \forall x \in E_X$$

Por tanto, fijado $x \in E_X$, $\exists!$ $c_x \in E_Y$ tal que:

$$[Y | X = x] = c_x$$

Por tanto, definimos la función $f : E_X \rightarrow E_Y$ como:

$$f(x) = c_x \quad \forall x \in E_X$$

Por tanto, Y depende funcionalmente de X .

\Leftarrow) Partimos de que Y depende funcionalmente de X , es decir, $\exists f : E_X \rightarrow E_Y$ tal que:

$$Y = f(X)$$

Entonces, tenemos que:

$$\begin{aligned} \text{Var}[Y \mid X = x] &= \text{Var}[f(X) \mid X = x] = \\ &= \text{Var}[f(x) \mid X = x] = 0 \quad \forall x \in E_X \end{aligned}$$

donde hemos usado que la varianza de una constante es 0.

□

Definición 4.9 (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 $f : E_X \rightarrow E_Y$ tal que:

$$\begin{aligned} Y &= f(X) & \forall x \in E_X \\ X &= f^{-1}(Y) & \forall y \in E_Y \end{aligned}$$

Dicha función f se denomina curva de dependencia.

Corolario 4.13.1. Sean X, Y dos variables aleatorias sobre el mismo espacio de probabilidad.

$$\text{Var}[Y \mid X] = \text{Var}[X \mid Y] = 0 \iff Y \text{ y } X \text{ dependen recíprocamente}$$

Demostración. Demostramos mediante doble implicación:

\Rightarrow) Como corolario de la proposición anterior, sabemos que $\exists f : E_X \rightarrow E_Y$, $g : E_Y \rightarrow E_X$ tales que:

$$\begin{aligned} Y &= f(X) & \forall x \in E_X \\ X &= g(Y) & \forall y \in E_Y \end{aligned}$$

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.

\Leftarrow) Se tiene de forma directa como corolario de la proposición anterior.

□

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:

$$y = f(x) = E[Y | X = x] \quad \forall x \in E_X$$

Demostración. Dado que $Y = f(X)$, entonces:

$$E[Y | X = x] = E[f(X) | X = x] = E[f(x) | X = x] = f(x) \quad \forall x \in E_X$$

□

- 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:

$$\begin{aligned} y = f(x) &= E[Y | X = x] & \forall x \in E_X \\ x = f^{-1}(y) &= E[X | Y = y] & \forall y \in E_Y \end{aligned}$$

- 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 :

$$\begin{aligned} \hat{Y}(x) &= E[Y] & \forall x \in E_X \\ \hat{X}(y) &= E[X] & \forall y \in E_Y \end{aligned}$$

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:

$$\begin{aligned} E[Y | X = x] &= E[Y] & \forall x \in E_X \\ E[X | Y = y] &= E[X] & \forall y \in E_Y \end{aligned}$$

□

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 $X = x$ |
|-----------|------------------|---------------------|-------------------------|
| \hat{Y} | $E[Y]$ | $E[Y X]$ | $E[Y X = x]$ |
| E.C.M. | $\text{Var}[Y]$ | $\text{Var}[Y X]$ | $\text{Var}[Y X = x]$ |

Tabla 4.1: Resumen de las predicciones de Y sobre en función de lo que se observe.

4.5. 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:

$$Y = E[Y | X] + (Y - E[Y | X])$$

Sabemos que $\hat{Y} = E[Y | X]$ es la mejor aproximación que hemos obtenido. Vemos por tanto que el error es la variable aleatoria $(Y - E[Y | X])$ unidades, que denominaremos *residuo*.

Definición 4.10 (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:

$$R = Y - E[Y | X]$$

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:

$$E[R] = E[Y - E[Y | X]] = E[Y] - E[E[Y | X]] = E[Y] - E[Y] = 0$$

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 4.11 (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:

$$\text{Var}[R] = \text{Var}[Y - E[Y | X]]$$

Calculemos el valor de la varianza residual.

$$\begin{aligned} \text{Var}[R] &= \text{Var}[Y - E[Y | X]] = E[(Y - E[Y | X])^2] - (E[Y - E[Y | X]])^2 = \\ &= E[(Y - E[Y | X])^2] - \cancel{E[R]^2} = E[(Y - E[Y | X])^2] = \\ &= E[E[(Y - E[Y | X])^2 | X]] = E[\text{Var}[Y | X]] = \text{E.C.M.}(\varphi_{\text{opt}}) \end{aligned}$$

Por tanto, usando la descomposición de la varianza, tenemos que:

$$\text{Var}[Y] = \text{Var}[E[Y | X]] + E[\text{Var}[Y | X]] = \text{Var}[\hat{Y}] + \text{Var}[R] = \text{Var}[\hat{Y}] + \text{E.C.M.}(\varphi_{\text{opt}}) \quad (4.1)$$

Vemos que el $\text{E.C.M.}(\varphi_{\text{opt}})$ será menor conforme mayor sea $\text{Var}[\hat{Y}]$, por lo que la aproximación será mejor conforme mayor sea $\text{Var}[\hat{Y}] = \text{Var}[E[Y | X]]$. No obstante, usar este valor para comparar la bondad de la aproximación introduce los siguientes problemas:

- $\text{Var}[E[Y | X]]$ 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 $Y' = 100Y$. No obstante, tenemos que:

$$\text{Var}[E[Y' | X]] = \text{Var}[E[100Y | X]] = \text{Var}[100E[Y | X]] = 100^2 \text{Var}[E[Y | X]]$$

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 4.12 (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 :

$$\eta_{Y/X}^2 = \frac{\text{Var}[E[Y | X]]}{\text{Var}[Y]}$$

De forma análoga se define la razón de correlación de X sobre Y :

$$\eta_{X/Y}^2 = \frac{\text{Var}[E[X | Y]]}{\text{Var}[X]}$$

Usando la Ecuación 4.1, tenemos:

$$\eta_{Y/X}^2 = \frac{\text{Var}[E[Y | X]]}{\text{Var}[Y]} = \frac{\text{Var}[Y] - E[\text{Var}[Y | X]]}{\text{Var}[Y]} = 1 - \frac{E[\text{Var}[Y | X]]}{\text{Var}[Y]}$$

Veamos que $\eta_{Y/X}^2$ solventa los dos problemas que presentábamos al usar $\text{Var}[E[Y | X]]$ para medir la bondad de la aproximación.

- $\eta_{Y/X}^2$ es adimensional, ya que las unidades de medida de $\text{Var}[E[Y | X]]$ y $\text{Var}[Y]$ 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:

$$\eta_{Y'/X}^2 = \eta_{aY/X}^2 = \frac{\text{Var}[E[aY | X]]}{\text{Var}[aY]} = \frac{\text{Var}[aE[Y | X]]}{a^2 \text{Var}[Y]} = \frac{a^2 \text{Var}[E[Y | X]]}{a^2 \text{Var}[Y]} = \eta_{Y/X}^2$$

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, $\eta_{Y/X}^2$ mide la proporción de la varianza de Y que queda explicada por la función de regresión $\hat{Y} = E[Y | X]$. 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 4.14. Sean X e Y dos variables aleatorias sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Entonces:

1. $0 \leq \eta_{Y/X}^2, \eta_{X/Y}^2 \leq 1$.
2. $\eta_{Y/X}^2 = 0 \iff$ La curva de regresión de Y sobre X es $Y = E[Y]$.
 $\eta_{X/Y}^2 = 0 \iff$ La curva de regresión de X sobre Y es $X = E[X]$.
3. $\eta_{Y/X}^2 = 1 \iff Y$ depende funcionalmente de X .
 $\eta_{X/Y}^2 = 1 \iff X$ depende funcionalmente de Y .
4. $\eta_{Y/X}^2 = \eta_{X/Y}^2 = 1 \iff X$ e Y tienen dependencia recíproca.

Demostración. Demostraremos tan solo los resultados para $\eta_{Y/X}^2$, ya que los resultados para $\eta_{X/Y}^2$ son análogos.

- Tenemos en primer lugar $\text{Var}[E[Y | X]] \geq 0$ y $\text{Var}[Y] \geq 0$ por ser varianzas. Por tanto lado, usamos que:

$$\eta_{Y/X}^2 = 1 - \frac{E[\text{Var}[Y | X]]}{\text{Var}[Y]}$$

Tenemos que $\text{Var}[Y | X], \text{Var}[Y] \geq 0$ por ser varianzas. Además, como la esperanza de una variable aleatoria no negativa es no negativa, tenemos que $E[\text{Var}[Y | X]] \geq 0$. Por tanto, $\eta_{Y/X}^2 \leq 1$ y tenemos lo buscado.

- Tenemos que:

$$\eta_{Y/X}^2 = 0 \iff \text{Var}[E[Y | X]] = 0 \iff \exists c \in \mathbb{R} \text{ tal que } E[Y | X] = c \quad \forall x \in E_X$$

donde la última implicación se debe a las propiedades de la varianza. Por tanto, tenemos que:

$$c = E[c] = E[E[Y | X]] = E[Y]$$

Por tanto, tenemos que:

$$\eta_{Y/X}^2 = 0 \iff E[Y | X] = E[Y]$$

- Tenemos que:

$$\begin{aligned} \eta_{Y/X}^2 = 1 &\iff E[\text{Var}[Y | X]] = 0 \iff P[\text{Var}[Y | X] = 0] = 1 \iff \\ &\iff \text{Var}[Y | X] = 0 \end{aligned}$$

Por tanto, tenemos que Y depende funcionalmente de X .

- Por lo razonado anteriormente, tenemos que:

$$\eta_{Y/X}^2 = \eta_{X/Y}^2 = 1 \iff \text{Var}[Y | X] = 0 = \text{Var}[X | Y]$$

Por tanto, X e Y tienen dependencia recíproca.

□

En esta última proposición, vemos que nos falta estudiar el caso en el que $\eta_{Y/X}^2 = \eta_{X/Y}^2 = 0$. En este caso, tenemos que ambas curvas de regresión son las esperanzas, y ninguna de las variables explica a la otra. Introducimos por tanto el siguiente concepto:

Definición 4.13 (Incorrelación). Dadas dos variables aleatorias X e Y sobre el mismo espacio de probabilidad, se dice que X e Y son incorreladas si

$$\eta_{Y/X}^2 = \eta_{X/Y}^2 = 0.$$

Como consecuencia de lo anteriormente visto, tenemos que si X e Y son independientes, entonces son incorreladas. No obstante, el recíproco no es cierto, como muestra el Ejercicio ??.

5. Relaciones de problemas

5.0. 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:

$$P[2 \leq Y \leq 5] = F_Y(5) - F_Y(1) = P[Y \leq 5] - P[Y \leq 1] \stackrel{(*)}{=} 0,9464 - 0,1211 = 0,8253$$

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:

$$\begin{aligned} P[X = 10] &= P[X \leq 10] - P[X \leq 9] = (1 - (1 - 0,15)^{11}) - (1 - (1 - 0,15)^{10}) = \\ &= (1 - 0,15)^{10} - (1 - 0,15)^{11} = 0,85^{10} \cdot (1 - 0,85) \approx 0,029 \end{aligned}$$

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:

$$5 + E[X] = 5 + \frac{k(1-p)}{p} = 5 + \frac{35}{3} = 16.\bar{6}$$

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:

$$\begin{aligned} P[X = 10] &= \frac{(10 + 5 - 1)!}{10!(5 - 1)!} \cdot 0,3^5 \cdot 0,7^{10} = \frac{14!}{10!4!} \cdot 0,3^5 \cdot 0,7^{10} = \\ &= \frac{14 \cdot 13 \cdot 12 \cdot 11}{4 \cdot 3 \cdot 2} \cdot 0,3^5 \cdot 0,7^{10} \approx 0,0687 \end{aligned}$$

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:

$$P[X \geq 15] = 1 - P[X \leq 14] = 1 - 0,3^5 \cdot \sum_{i=0}^{14} \binom{i+4}{4} 0,7^i \approx 0,2822$$

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:

$$\begin{aligned} P[X \geq 1] &= 1 - P[X = 0] = 1 - \frac{\binom{100}{0} \cdot \binom{9900}{100}}{\binom{10000}{100}} = \\ &= 1 - \frac{1 \cdot \frac{9900!}{100!9800!}}{\frac{10000!}{100!9900!}} = 1 - \frac{9900! \cdot 100! \cdot 9900!}{10000! \cdot 100! \cdot 9800!} \end{aligned}$$

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 $p = \frac{N_1}{N} = \frac{100}{10000} = 0,01$. 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:

$$P[X \geq 1] = 1 - P[X = 0] = 1 - \binom{100}{0} \cdot 0,01^0 \cdot 0,99^{100} \approx 1 - 0,366 = 0,634$$

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 aproximación 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:

$$P[X \geq 5] = 1 - P[X \leq 4] \stackrel{(*)}{=} 1 - 0,998 = 0,002$$

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:

$$\text{Var}[Y + 4] \stackrel{(*)}{=} \text{Var}[Y] = \frac{k(1-p)}{p^2} = \frac{4 \cdot 0,95}{0,05^2} = 1520$$

donde en $(*)$ hemos usado que:

$$\text{Var}[aX + b] = a^2 \text{Var}[X] \quad \forall a, b \in \mathbb{R}$$

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:

$$P(A \cap B) = P(A) \cdot P(B | A)$$

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 $P(B | A)$, 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:

$$P(B | A) = \frac{3}{6} = 0,5$$

Por tanto, la probabilidad de que se necesiten cinco extracciones para obtener tres números pares es:

$$P(A \cap B) = P(A) \cdot P(B | A) = 0,476 \cdot 0,5 = 0,238$$

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:

$$P(C \cap D) = P(C) \cdot P(D | C)$$

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 $P(D | C)$, 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:

$$P(D | C) = \frac{1}{7}$$

Por tanto, la probabilidad de obtener el número 7 en la cuarta extracción es:

$$P(C \cap D) = P(C) \cdot P(D | C) = \frac{7}{10} \cdot \frac{1}{7} = \frac{1}{10} = 0,1$$

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:

$$P[X \geq 12] = 1 - P[X \leq 11] \stackrel{(*)}{=} 1 - 0,6968 = 0,3032$$

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 $\hat{x} \in \mathbb{N}$ tal que:

$$P[X \leq \hat{x}] \geq 0,95$$

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 $P_X : \Omega \rightarrow [0, 1]$, 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 $F_X : \mathbb{R} \rightarrow [0, 1]$, es:

$$F_X(x) = P[X \leq x] = \begin{cases} 0, & x < 1 \\ \frac{\lfloor x \rfloor}{6}, & 1 \leq x < 7 \\ 1, & x \geq 7 \end{cases}$$

3. Función generatriz de momentos.

La función generatriz de momentos de X , notada por $M_X : \mathbb{R} \rightarrow \mathbb{R}$, 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 $t = 0$:

$$\begin{aligned} E[X] &= M'_X(t) \Big|_{t=0} = \frac{d}{dt} \left(\frac{1}{6} \sum_{x=1}^6 e^{tx} \right) \Big|_{t=0} \\ &= \frac{1}{6} \sum_{x=1}^6 \frac{d}{dt} (e^{tx}) \Big|_{t=0} = \frac{1}{6} \sum_{x=1}^6 x e^{tx} \Big|_{t=0} = \frac{1}{6} \sum_{x=1}^6 x = \frac{7}{2} = 3,5 \end{aligned}$$

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.

$$X \sim \mathcal{U}(1, \dots, 6)$$

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:

$$\begin{aligned} P[X = -3] &= P[X = 3] = \frac{1}{2^3} = \frac{1}{8} \\ P[X = -1] &= P[X = 1] = \frac{3}{2^3} = \frac{3}{8} \end{aligned}$$

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:

$$F_X(x) = P[X \leq x] = \begin{cases} 0, & x < -3 \\ 1/8, & -3 \leq x < -1 \\ 1/2, & -1 \leq x < 1 \\ 7/8, & 1 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

6. Probabilidad de que X sea positiva.

Tenemos que:

$$P[X > 0] = 1 - P[X \leq 0] = 1 - 1/2 = 1/2$$

Ejercicio 12. Dado $k \in \mathbb{R}$, sea X una variable aleatoria con función de densidad dada por:

$$f_X(x) = \begin{cases} k/x^2 & 1 \leq x \leq 8 \\ 0, & \text{en otro caso} \end{cases}$$

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 $f_X(x) \geq 0$ 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:

$$F_X(x) = P[X \leq x] = \begin{cases} 0, & x < 1 \\ \int_1^x \frac{8}{7x^2} dx = \left[-\frac{8}{7x} \right]_1^x = -\frac{8}{7x} + \frac{8}{7}, & 1 \leq x < 8 \\ 1, & x \geq 8 \end{cases}$$

3. Valor esperado.

El valor esperado de X es:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \int_1^8 x \cdot \frac{8}{7x^2} dx = \frac{8}{7} \int_1^8 \frac{1}{x} dx = \frac{8}{7} [\ln |x|]_1^8 = \frac{8}{7} \ln(8)$$

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:

$$\text{Var}[X] = E[X^2] - E[X]^2 = 8 - \frac{64}{49} \cdot \ln(8)^2 \approx 2,352$$

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:

$$f_X(x) = \begin{cases} 3/8 \cdot x^2 & 0 \leq x \leq 2 \\ 0, & \text{en otro caso} \end{cases}$$

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 cantidad. 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:

$$G(x) = \begin{cases} 100 \cdot x, & 0 \leq x \leq 1 \\ 140 \cdot x & 1 < x \leq 2 \end{cases}$$

Calculamos el valor esperado de la ganancia de la gasolinera en un día:

$$\begin{aligned} E[G(X)] &= \int_{-\infty}^{\infty} G(x) \cdot f_X(x) dx = \int_0^2 G(x) \cdot \frac{3}{8} x^2 dx = \\ &= \int_0^1 100x \cdot \frac{3}{8} x^2 dx + \int_1^2 140x \cdot \frac{3}{8} x^2 dx = \\ &= \frac{3}{8} \int_0^1 100x^3 dx + \frac{3}{8} \int_1^2 140x^3 dx = \frac{3}{8} [25x^4]_0^1 + \frac{3}{8} [35x^4]_1^2 = \\ &= \frac{3}{8} \cdot 25 + \frac{3}{8} \cdot 35 \cdot 15 = \frac{825}{4} = 206,25 \end{aligned}$$

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.