

## Capítulo 4

# Momentos de una Distribución de Probabilidad

### 4.1. Esperanza de una variable aleatoria

**Definición 4.1** Sea  $X$  una v.a., entonces la **esperanza** de  $X$  se define como:

- $E(X) := \sum_x xf(x)$  si  $X$  tiene distribución discreta.
- $E(X) := \int_{-\infty}^{\infty} xf(x) dx$  si  $X$  tiene distribución continua.
- $E(X) := \sum_{x \in D} xf(x) + \int_{x \in C} xf(x) dx$  si  $X$  tiene distribución mixta.

Hemos de observar que la esperanza  $E(X)$  no es necesariamente igual a uno de los valores posibles de  $X$ .

En todos los casos anteriores la expresión valuada debe ser un número finito. Así, por ejemplo, para la distribución discreta, la serie debe de ser *absolutamente convergente*, si se extiende la suma a un número infinito de valores. Para la distribución continua, la integral debe ser, también, absolutamente convergente. Además siempre que  $X$ , v.a. continua, sea acotada su valor esperado ha de existir. Lo anterior se puede expresar del siguiente modo, las v.a. tienen su valores esperados, como ya los hemos descrito, si se cumple:

- $\sum_x |x|f(x) < \infty$  para  $X$  discreta.

- $\int_{-\infty}^{\infty} |x|f(x) \, dx < \infty$  para  $X$  continua.

A la esperanza de una v.a. también se le conoce con los nombres: *valor esperado*, *media* de  $X$ . Los términos son indistintamente utilizados.

## 4.2. Interpretación de la esperanza

A la media de una v.a. se le considera como el centro de gravedad de la distribución. Para entender este concepto obsérvese la figura (4.1), supongase que el eje de las  $x$ 's es una barra sin peso, y que las barras verticales representan las probabilidades en cada uno de los valores; estas barras tienen pesos de  $f(x_i)$  en cada punto  $x_i$  para todos los  $i$ . Entonces, el punto  $E(X)$  es el punto donde podemos apoyar la barra, de tal modo que ésta permanezca en equilibrio.

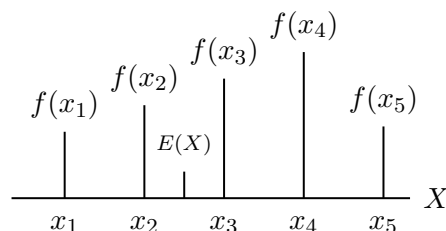


Figura 4.1: La media de una distribución discreta

Para una distribución continua, figura (4.2), la interpretación es similar, salvo que la masa de probabilidad varía continuamente.

Vemos de la gráfica que el valor medio de una distribución continua puede verse afectado por un cambio pequeño a la masa de probabilidad a un muy valor alejado de  $x$  con respecto de donde se concentra la mayor parte de la masa.

Para distribuciones, discretas y continuas, simétricas entorno a un punto  $x_0$ , pareciera que este punto puede ser la media; sin embargo, siempre hay que asegurarse de que así es pues, por ejemplo, la distribución Cauchy es simétrica y carece de media.

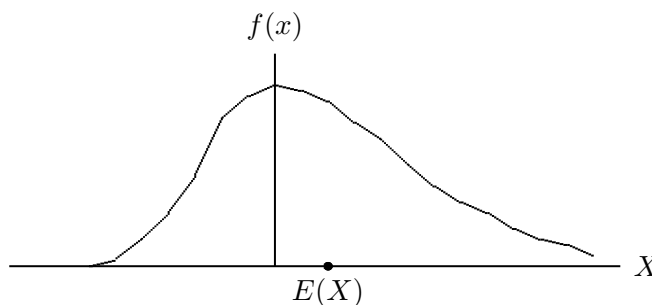


Figura 4.2: La media de una distribución continua

### 4.3. Esperanza de una función

Supongamos que  $X$  es una v.a. continua y que tiene f.d.p. igual a  $f_X$ , entonces el valor esperado de  $Y = r(X)$ , según la definición, es:

$$E[r(X)] = E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy.$$

Sin embargo, este valor puede ser calculado del siguiente modo:

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

Además, para existir, debe de cumplir con la condición de convergencia absoluta:

$$\int_{-\infty}^{\infty} |r(x)| f(x) dx < \infty.$$

En el caso de distribuciones discretas el proceso es el mismo.

Para funciones de varias variables es similar.

Supongamos que  $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$  es la f.d.p. conjunta de  $(X_1, \dots, X_n)$ , entonces:

$$\begin{aligned} E[r(X_1, \dots, X_n)] &= E(Y) \\ &= \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= \int \cdots \int_{\mathbb{R}^n} r(x_1, \dots, x_n) f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n. \end{aligned}$$

Y de este modo no evitamos el determinar la distribución de  $Y = r(X_1, \dots, X_n)$ .

## 4.4. Propiedades de los valores esperados

Algunos resultados a lo que se pueden llegar de las definiciones, ya dadas, son:

**Teorema 4.1** Si  $Y = aX + b$ , donde  $a, b \in \mathbb{R}$ , entonces

$$E(Y) = aE(X) + b.$$

**Demostración.** Supóngase, que  $X$  tiene una distribución continua cuya f.d.p. es  $f_X$ . Entonces,

$$\begin{aligned} E(Y) &= E(aX + b) = \int_{-\infty}^{\infty} (ax + b)f_X(x) dx \\ &= a \int_{-\infty}^{\infty} xf_X(x) dx + b \int_{-\infty}^{\infty} f_X(x) dx \\ &= aE(X) + b \end{aligned}$$

La demostración cuando  $X$  es discreta es similar.  $\triangleleft$

Por el teorema anterior, resulta que para  $c \in \mathbb{R}$ ,  $E(c) = c$ .

**Teorema 4.2** Si existe una  $a \in \mathbb{R}$  tal que  $\Pr(X \geq a) = 1$ , entonces  $E(X) \geq a$ . Si existe una  $b \in \mathbb{R}$  tal que  $\Pr(X \leq b) = 1$ , entonces  $E(X) \leq b$ .

**Demostración.** Sea  $X$  con f.d.p. igual a  $f_X$  y supóngase en primer lugar que  $\Pr(X \geq a) = 1$ . Entonces,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x) dx = \int_a^{\infty} xf(x) dx \\ &\geq \int_a^{\infty} af(x) dx = a \Pr(X \geq a) = a \end{aligned}$$

La demostración de la otra parte del teorema y la demostración para una distribución discreta o un tipo de distribución más general son análogas.  $\triangleleft$

Por el teorema anterior resulta que si  $\Pr(a \leq X \leq b) = 1$ , entonces  $a \leq E(X) \leq b$ . Además se puede demostrar que si  $\Pr(X \geq a) = 1$  y si  $E(X) = a$ , entonces se debe verificar que  $\Pr(X > a) = 0$  y  $\Pr(X = a) = 1$ .

**Teorema 4.3** Si  $X_1, \dots, X_n$ , son  $n$  v.a. tal que  $E(X_i)$ , para  $i = \overline{1, n}$ , existe, entonces

$$E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n).$$

**Demostración.** Supóngase que  $n = 2$ , y que  $X_1$  y  $X_2$  tienen una distribución conjunta continua cuya f.d.p. es  $f_{X_1, X_2}$ . Entonces,

$$\begin{aligned} E(X_1 + X_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 + x_2) f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \\ &= E(X_1) + E(X_2). \end{aligned}$$

La demostración para una distribución discreta o una distribución conjunta más general es análoga. Y para cualquier  $n \in \mathbb{N}$  se deduce por inducción.  $\triangleleft$

El teorema anterior es verdadero independientemente de si las v.a. son o no independientes.

De los tres resultados anteriores resulta que para  $a_1, \dots, a_n, b \in \mathbb{R}$ ,

$$E(a_1 X_1 + \dots + a_n X_n + b) = a_1 E(X_1) + \dots + a_n E(X_n) + b.$$

**Teorema 4.4** Si  $X_1, \dots, X_n$  son variables aleatorias independientes cuyas esperanzas  $E(X_i)$  ( $i = \overline{1, n}$ ) existen, entonces

$$E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i).$$

**Demostración.** Supóngase que  $X_1, \dots, X_n$  tienen una distribución conjunta continua cuya f.d.p. conjunta es  $f_{X_1, \dots, X_n}$ . Además, se define  $f_i$  como la f.d.p. marginal de  $X_i$  ( $i = \overline{1, n}$ ). Entonces, puesto que las variables  $X_1, \dots, X_n$  son independientes resulta que para todo punto  $(x_1, \dots, x_n) \in \mathbb{R}^n$ ,

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i).$$

Por tanto,

$$\begin{aligned} E\left(\prod_{i=1}^n X_i\right) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\prod_{i=1}^n x_i\right) f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left[\prod_{i=1}^n x_i f_i(x_i)\right] dx_1 \dots dx_n \\ &= \prod_{i=1}^n \int_{-\infty}^{\infty} x_i f_i(x_i) dx_i = \prod_{i=1}^n E(X_i). \end{aligned}$$

La demostración para una distribución discreta o para un tipo de distribución más general es análoga.  $\triangleleft$

En resumen, los dos resultados anteriores dicen:

- La esperanza de la suma de v.a. es igual a la suma de sus esperanzas individuales, independientemente de si las variables son o no independientes.
- La esperanza del producto de v.a. no siempre es igual al producto de sus esperanzas individuales, a menos que las variables sean independientes.

Además tenemos los siguientes resultados:

- $E[r_1(X_1) \pm r_2(X_2)] = E[r_1(X_1)] \pm E[r_2(X_2)].$
- Si  $X_1$  y  $X_2$  son v.a. independientes estadísticamente

$$E[r_1(X_1)r_2(X_2)] = E[r_1(X_1)] \cdot E[r_2(X_2)].$$

- $E[a + br(X)] = a + bE[r(X)].$

### 4.5. Varianza

**Definición 4.2** Supóngase que  $X$  es una v.a. con media  $\mu = E(X)$ . La **varianza** de  $X$ , que se denotará por  $\text{Var}(X)$ , se define como sigue:

$$\text{Var}(X) := E[(X - \mu)^2]. \quad (4.1)$$

Dado que  $(X - \mu)^2 \geq 0 \implies \text{Var}(X) \geq 0$ . De nuevo es requisito que el valor de la ecuación (4.1) sea finito para existir.

La varianza de una distribución es una medida de la variación o dispersión de una distribución entorno a su media  $\mu$ . Dependiendo del valor que tome la varianza tenemos:

- $\text{Var}(X)$  pequeña, indica que la distribución de probabilidad está muy concentrada entorno a  $\mu$ .
- $\text{Var}(X)$  grande, indica que la distribución de probabilidad tiene una dispersión amplia entorno a  $\mu$ .

**Definición 4.3** La **desviación típica** de una v.a. o de una distribución, denotada  $\sigma$ , se define como:

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

A la desviación típica también se le conoce como *desviación estándar*. De esta notación se sigue que la varianza sea denotada como  $\sigma^2$ .

Ahora presentamos algunos resultados relacionados con la varianza.

**Teorema 4.5**  $\text{Var}(X) = 0 \iff \exists c \in \mathbb{R} \text{ tal que } \Pr(X = c) = 1.$

**Demostración.** Supongamos que  $\exists c \in \mathbb{R}$  tal que  $\Pr(X = c) = 1$ . Así,  $E(X) = c$  y  $\Pr[(X - c)^2 = 0] = 1$ .  $\therefore$

$$\text{Var}(X) = E[(X - c)^2] = 0.$$

Inversamente, supóngase que  $\text{Var}(X) = 0$ . Entonces,  $\Pr[(X - \mu)^2 \geq 0] = 1$  pero  $E[(X - \mu)^2] = 0$ .  $\therefore$  observamos que

$$\Pr[(X - \mu)^2 = 0] = 1.$$

$\therefore \Pr(X = \mu) = 1$ .  $\triangleleft$

**Teorema 4.6**  $\forall a, b \in \mathbb{R},$

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$



**Demostración.** Si  $E(X) = \mu$ , entonces  $E(aX + b) = a\mu + b$ .  $\therefore$

$$\begin{aligned}\text{Var}(aX + b) &= E[(aX + b - a\mu - b)^2] = E[(aX - a\mu)^2] \\ &= a^2 E[(X - \mu)^2] = a^2 \text{Var}(X).\end{aligned}$$

$\triangleleft$

De este último resultado tenemos que  $\text{Var}(X + b) = \text{Var}(X) \forall b \in \mathbb{R}$ . Esto indica que el desplazamiento de una distribución en  $b$  unidades no afecta su dispersión entorno a  $\mu$ .

También, de este último teorema, resulta  $\text{Var}(-X) = \text{Var}(X)$ .

**Teorema 4.7** Para cualquier variable aleatoria  $X$ ,  $\text{Var}(X) = E(X^2) - [E(X)]^2$ .

**Demostración.** Sea  $E(X) = \mu$ . Entonces,

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

$\triangleleft$

**Teorema 4.8** Si  $X_1, \dots, X_n$  son v.a. independientes, entonces

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

**Demostración.** Supóngase que  $n = 2$ . Si  $E(X_1) = \mu_1$  y  $E(X_2) = \mu_2$ , así

$$E(X_1 + X_2) = \mu_1 + \mu_2.$$

Por tanto,

$$\begin{aligned}\text{Var}(X_1 + X_2) &= E[(X_1 + X_2 - \mu_1 - \mu_2)^2] \\ &= E[(X_1 - \mu_1)^2 + (X_2 - \mu_2)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= \text{Var}(X_1) + \text{Var}(X_2) + 2E[(X_1 - \mu_1)(X_2 - \mu_2)].\end{aligned}$$

Pero  $X_1$  y  $X_2$  son independientes,

$$\begin{aligned}E[(X_1 - \mu_1)(X_2 - \mu_2)] &= E(X_1 - \mu_1)E(X_2 - \mu_2) \\ &= (\mu_1 - \mu_1)(\mu_2 - \mu_2) \\ &= 0.\end{aligned}$$

Resulta que

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2).$$

Obtenemos el resultado para  $n \in \mathbb{N}$  por inducción.  $\triangleleft$

Recuérdese que el anterior teorema es únicamente para v.a. independientes.

**Corolario 4.1** *Si  $X_1, \dots, X_n$  son variables aleatorias independientes y si  $a_1, \dots, a_n$  y  $b$  son constantes arbitrarias, entonces*

$$\text{Var} \left( \sum_{i=1}^n a_i X_i + b \right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i).$$

## 4.6. Momentos

**Definición 4.4** Sea  $X$  cualquier v.a. y  $n \in \mathbb{N}$ , la esperanza

$$E(X^k)$$

se denomina **momento** de orden  $k$  de  $X$ .

Es requisito que  $E(|X|^k) < \infty$  para que el momento exista. Si la v.a.  $X$  está acotada, entonces deben existir necesariamente todos los momentos de  $X$ . Es posible, también, que existan todos los momentos de  $X$  aunque  $X$  no esté acotada.

**Teorema 4.9** Si  $E(|X|^k) < \infty$  para  $k \in \mathbb{N}$ , entonces  $E(|X|^j) < \infty$  para cualquier  $j \in \mathbb{N}$  tal que  $j < k$ .

**Demostración.** Supongamos que  $X$  es v.a. continua y que la f.d.p. es  $f_X$ . Así

$$\begin{aligned} E(|X|^j) &= \int_{-\infty}^{\infty} |x|^j f(x) dx \\ &= \int_{|x| \leq 1} |x|^j f(x) dx + \int_{|x| > 1} |x|^j f(x) dx \leq \\ &= \int_{|x| \leq 1} 1 \cdot f(x) dx + \int_{|x| > 1} |x|^k f(x) dx \leq \\ &= \Pr(|X| \leq 1) + E(|X|^k). \end{aligned}$$

Por hipótesis,  $E(|X|^k) < \infty$ . Por tanto,  $E(|X|^j) < \infty$ . Para una distribución discreta, se aplica una demostración análoga.  $\triangleleft$

El anterior teorema nos dice que si  $E(X^2) < \infty$ , entonces existen la media y la varianza de  $X$ .

**Definición 4.5** Sea  $X$  cualquier v.a. con  $E(X) = \mu$  y  $n \in \mathbb{N}$ , la esperanza

$$E[(X - \mu)^k]$$

se llama **momento central** de orden  $k$  de  $X$

De acuerdo con lo anterior la varianza es el segundo momento central.

Y el momento central de orden  $k = 1$  es cero, pues:

$$E(X - \mu) = \mu - \mu = 0.$$

También, si la distribución de  $X$  es simétrica respecto a  $E(X) = \mu$  y si existe el momento central de orden  $k$ ,  $E[(X - \mu)^k]$  para un entero impar  $k$ , entonces  $E[(X - \mu)^k] = 0$ , porque los términos positivos y negativos de esta expresión se cancelan unos con otros.

### 4.7. Función generatriz de momentos

**Definición 4.6** Dada la v.a.  $X$  con distribución de probabilidad  $f_X(x)$ , se llama **función generatriz de momentos** de  $X$ , abreviada f.g.m., a la función real de variable real  $\psi_X$  definida por:

$$\psi_X(t) := E(e^{tX}). \quad (4.2)$$

Recuérdese que hemos dicho que si la distribución es acotada entonces existe la esperanza, en este caso, la de  $e^{tX}$ .

Podemos ver que la f.g.m.  $\psi_X(t)$  debe existir en el punto  $t = 0$  y ahí su valor es:

$$\psi_X(0) = E(1) = 1 \quad .$$

En general, siempre que exista la f.g.m.  $\psi_X(t)$  de  $X$  para todos los valores de  $t$  en un intervalo entorno del punto  $t = 0$ , entonces existen todos los momentos  $E(X^k)$  de  $X$  ( $k \in \mathbb{N}$ ). Además, es posible derivar  $\psi_X(t)$  un número arbitrario de veces en el punto  $t = 0$ . Para  $k \in \mathbb{N}$ , la  $k$ -ésima derivada  $\psi_X^{(k)}(0)$  en  $t = 0$  satisface:

$$\begin{aligned} \psi_X^{(k)}(0) &= \left[ \frac{d^k}{dt^k} E(e^{tX}) \right]_{t=0} \\ &= E \left[ \left( \frac{d^k}{dt^k} e^{tX} \right)_{t=0} \right] \\ &= E[(X^k e^{tX})_{t=0}] \\ &= E(X^k) \\ &= \alpha_k. \end{aligned}$$

Cuando  $k = 1$  tenemos:

$$\begin{aligned} \psi_X'(0) &= \left[ \frac{d}{dt} E(e^{tX}) \right]_{t=0} \\ &= E \left[ \left( \frac{d}{dt} e^{tX} \right)_{t=0} \right] \\ &= E[(X e^{tX})_{t=0}] \\ &= E(X) \quad . \end{aligned}$$

Es decir, la primera derivada de la f.g.m.  $\psi_X(t)$  valuada en  $t = 0$  es la media de  $X$ .

Y para los demás  $k$ , tenemos:

$$\begin{aligned}\psi'_X(0) &= E(X) \\ \psi''_X(0) &= E(X^2) \\ \psi'''_X(0) &= E(X^3)\end{aligned}$$

**Teorema 4.10** Sea  $X$  una v.a. con f.g.m.  $\psi_X$ ; sea  $Y = aX + b$ , donde  $a, b \in \mathbb{R}$ ; y sea  $\psi_Y$  la f.g.m. de  $Y$ . Entonces, para cualquier valor de  $t$  tal que existe  $\psi_X(at)$ ,

$$\psi_Y(t) = e^{bt}\psi_X(at).$$

**Demostración.** De la definición de una f.g.m.

$$\psi_Y(t) = E(e^{tY}) = E[e^{t(aX+b)}] = e^{bt}E(e^{atX}) = e^{bt}\psi_X(at).$$

◁

**Teorema 4.11** Supóngase que  $X_1, \dots, X_n$  son v.a. independientes; y para  $i = \overline{1, n}$ , sea  $\psi_i$  la f.g.m. de  $X_i$ . Sea  $Y = X_1 + \dots + X_n$  y denótese por  $\psi_Y$  la f.g.m. de  $Y$ . Entonces, para cualquier valor de  $t$  tal que existe  $\psi_i(t)$  para  $i = \overline{1, n}$ ,

$$\psi_Y(t) = \prod_{i=1}^n \psi_i(t).$$

**Demostración.** Por definición,

$$\psi_Y(t) = E(e^{tY}) = E[e^{t(X_1 + \dots + X_n)}] = E\left(\prod_{i=1}^n e^{tX_i}\right) = \prod_{i=1}^n E(e^{tX_i}).$$

Recuérdese que la esperanza de un producto de v.a. independientes es el producto de las esperanzas individuales. ◁

**Teorema 4.12** Si las f.g.m. de dos v.a.  $X_1$  y  $X_2$  son idénticas para todos los valores de  $t$  en un intervalo alrededor del punto  $t = 0$ , entonces las distribuciones de probabilidad de  $X_1$  y  $X_2$  deben ser idénticas.

**Definición 4.7** Dada la v.a.  $X$  con distribución de probabilidad  $f_X(x)$ , se llama **función generatriz de momentos respecto de la media** a la función real de variable real  $\bar{\psi}_X$  definida por:

$$\bar{\psi}_X(t) := E[e^{t(X-\mu)}]$$

en donde  $\mu$  es la media de  $X$  y se supone que existe.

Obviamente:

$$\bar{\psi}_X(t) = \begin{cases} \sum e^{t(x-\mu)} f_X(x) & \text{para variables discretas.} \\ \int_{-\infty}^{+\infty} e^{t(x-\mu)} f_X(x) dx & \text{para variables continuas.} \end{cases}$$

Recuérdese el requisito de convergencia de la serie o de la integral, para la existencia de  $\bar{\psi}_X$ .

**Teorema 4.13** *Sea  $X$  una v.a. con distribución de probabilidad  $f_X(x)$ , media  $\mu$  y f.g.m. respecto de la media  $\bar{\psi}_X$ . Si existe el momento centrado de orden  $k \in \mathbb{N}$ ,  $\mu_k$ , entonces:*

$$\left. \frac{d^k \bar{\psi}_X(t)}{dt^k} \right|_{t=0} = \mu_k.$$

## 4.8. La media y la mediana

### 4.8.1. La mediana

Hemos interpretado la media de una distribución como el centro de gravedad de la distribución; pues bien, la mediana se puede interpretar como el punto que divide la distribución en dos puntos. Hemos de mencionar, que para no todas las distribuciones discretas existirá un punto en que la distribución de probabilidad se divida en dos partes iguales.

**Definición 4.8** *Para cualquier v.a.  $X$ , la **mediana** de la distribución de  $X$  se define como un punto  $m$  tal que*

$$\Pr(X \leq m) \leq 1/2 \quad y \quad \Pr(X \geq m) \geq 1/2 .$$

Habrán casos en los que al no existir un punto que satisfaga exactamente la definición se puedan considerar dos medianas.

### 4.9. Asimetría y curtosis

Los valores de asimetría y curtosis, generalmente, nos permitirán ver el perfil o forma que tiene la distribución. Éstas son llamadas *medidas de forma*.

**Definición 4.9** Una v.a.  $X$  presenta una **distribución simétrica** respecto a un eje perpendicular al de abscisas en el punto  $s$  cuando

$$\Pr(X \geq s + x) = \Pr(X \leq s - x) \quad \forall x.$$

En caso contrario es una distribución no simétrica o **asimétrica**.

Cuando la distribución de probabilidad es continua, y su f.d.p. es  $f_X(x)$ , dicha distribución será simétrica si

$$f_X(s + x) = f_X(s - x), \quad \forall x.$$

Generalmente se mide la simetría o asimetría de una distribución de probabilidad respecto al eje que pase por  $\mu$ , es decir, tomando  $s = \mu$ .

Ahora deduciremos un coeficiente de asimetría. Si la distribución es simétrica en  $\mu$  las desviaciones a la derecha (i.e. las positivas) tendrán el mismo peso que las desviaciones negativas (i.e. las de la izquierda), por lo que todos los momentos de orden impar respecto a  $\mu$  serán nulos. Es decir, si una distribución es simétrica en  $\mu$  se cumplirá:

$$\mu_{2t+1} = E[(X - \mu)^{2t+1}] = 0 \quad \text{para } t = 0, 1, 2, \dots$$

como  $\mu_1 = 0$  siempre, sea la distribución simétrica o no, utilizaremos el siguiente momento impar,  $\mu_3$ , para medir la asimetría.

Así se presentarán los siguientes casos:

- $\mu_3 = 0$  la distribución es simétrica.
- $\mu_3 > 0$  tienen mayor peso las desviaciones positivas sobre las negativas. En este caso la distribución es asimétrica positiva o a la derecha.
- $\mu_3 < 0$  tienen mayor peso las desviaciones negativas sobre las positivas. En este caso la distribución es asimétrica negativa o a la izquierda.

De lo anterior Fisher definió:

**Definición 4.10** El **coeficiente de asimetría** se define como:

$$\gamma_1 := \frac{\mu_3}{\sigma^3}.$$

Y según su valor:



- Si  $\gamma_1 = 0$  la distribución es **simétrica**.
- Si  $\gamma_1 > 0$  la distribución es **asimétrica positiva**.
- Si  $\gamma_1 < 0$  la distribución es **asimétrica negativa**.

El anterior coeficiente tiene la propiedad de ser adimensional e invariante a cambios de origen y de escala.

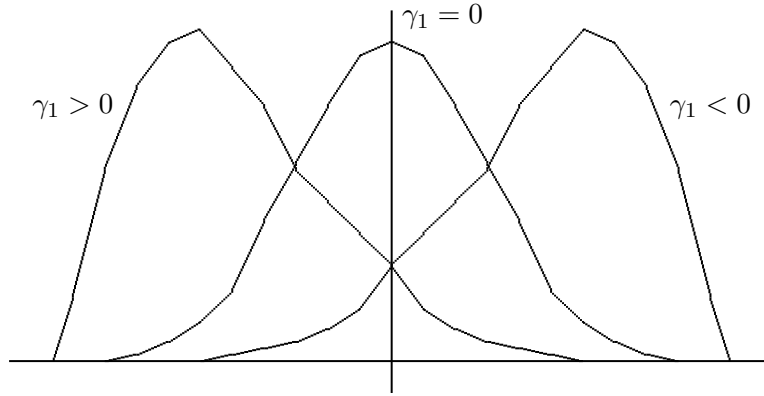


Figura 4.3: Coeficiente de asimetría  $\gamma_1$ .

Ahora estudiemos el concepto de curtosis. Fisher definió el siguiente coeficiente:

**Definición 4.11** *El coeficiente de curtosis se define como:*

$$\gamma_2 := \frac{\mu_4}{\sigma^4} - 3 .$$

*Y según su valor:*

- Si  $\gamma_2 = 0$ , se dice que es **mesocúrtica**.
- Si  $\gamma_2 > 0$ , se dice que es **leptocúrtica**.
- Si  $\gamma_2 < 0$ , se dice que es **platicúrtica**.

El anterior coeficiente es invariante a cambios de origen y de escala. La distribución normal tiene f.d.p. igual a:

$$f_X(x) = \frac{1}{(2\pi)^{1/2}\sigma} \exp \left[ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right] \mathbf{I}_{\mathbb{R}}(x)$$

y a su gráfica se le conoce *como campana de Gauss*. Ésta distribución está relacionada con este coeficiente pues el coeficiente de curtosis nos permite ver que tan apuntada es una distribución comparada con la normal. Para la normal:

$$\mu_4 = 3\sigma^4 .$$

Así tenemos la siguiente interpretación:

- Si  $\gamma_2 = 0$ , la distribución tendrá la forma de la normal.
- Si  $\gamma_2 > 0$ , la distribución es más apuntada que la normal.
- Si  $\gamma_2 < 0$ , la distribución es más achatada que la normal.

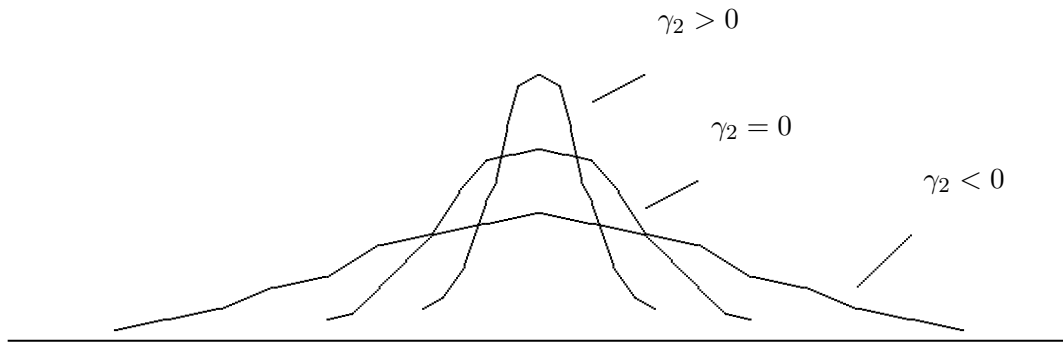


Figura 4.4: Coeficiente de curtosis  $\gamma_2$ .

### 4.10. Covarianza y correlación

Una cantidad que mide la relación que hay entre dos v.a. que tienen distribución conjunta es la covarianza. Ahora la definimos.

**Definición 4.12** Sean  $X$  e  $Y$  v.a. que tienen una distribución conjunta cuyos primeros momentos y varianzas son  $E(X) = \mu_X, E(Y) = \mu_Y, \text{Var}(X) = \sigma_X^2$  y  $\text{Var}(Y) = \sigma_Y^2$ . La **covarianza de  $X$  e  $Y$** , que se denota por  $\text{Cov}(X, Y)$ , se define por:

$$\text{Cov}(X, Y) := E[(X - \mu_X)(Y - \mu_Y)]. \quad (4.3)$$

También se utiliza  $\sigma_{X,Y} = \text{Cov}(X, Y)$ .

Si  $\sigma_X^2 < \infty$  y  $\sigma_Y^2 < \infty$ , entonces existe la esperanza de la ecuación (4.3) y  $\text{Cov}(X, Y) < \infty$ . Además, el valor de  $\text{Cov}(X, Y) \in \mathbb{R}$ .

**Definición 4.13** Si  $\sigma_X^2, \sigma_Y^2 \in \mathbb{R}^2$ , entonces la **correlación de  $X$  e  $Y$** , denotada por  $\text{Corr}(X, Y)$ , se define como :

$$\text{Corr}(X, Y) := \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}. \quad (4.4)$$

También se utiliza  $\rho_{X,Y} = \text{Corr}(X, Y)$ .

La determinación del rango de los valores posibles de la correlación  $\rho_{X,Y}$ , necesita de el siguiente resultado:

**Teorema 4.14 (Desigualdad de Schwarz)** Para cualesquiera v.a.  $U$  y  $V$ ,

$$[E(UV)]^2 \leq E(U^2)E(V^2).$$

Definiendo  $U = X - \mu_X$  y  $V = Y - \mu_Y$ , de la desigualdad de Schwarz resulta:

$$[\text{Cov}(X, Y)]^2 \leq \sigma_X^2 \sigma_Y^2.$$

Luego, de la ecuación (4.4) tenemos que  $[\text{Corr}(X, Y)]^2 \leq 1$  o, equivalente a:

$$-1 \leq \rho_{X,Y} \leq 1.$$

**Definición 4.14** Según el valor que tome el  $\text{Corr}(X, Y)$  tenemos:

- $\rho_{X,Y} > 0$ , entonces  $X$  e  $Y$  están **correlacionados positivamente**.
- $\rho_{X,Y} < 0$ , entonces  $X$  e  $Y$  están **correlacionados negativamente**.
- $\rho_{X,Y} = 0$ , entonces  $X$  e  $Y$  están **no correlacionados**.

De las definiciones para  $\text{Cov}(X, Y)$  y  $\text{Corr}(X, Y)$ , vemos que ambas tienen el mismo signo.

Ahora daremos algunos resultados para las definiciones anteriores.

**Teorema 4.15** *Para cualesquiera v.a.  $X$  e  $Y$  tales que  $\sigma_X^2, \sigma_Y^2 \in \mathbb{R}^+$ ,*

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y). \quad (4.5)$$

**Demostración.** De la definición tenemos:

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y. \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

◁

**Teorema 4.16** *Si  $X$  e  $Y$  son v.a. independientes con  $\sigma_X^2, \sigma_Y^2 \in \mathbb{R}^+$ , entonces*

$$\text{Cov}(X, Y) = 0.$$

**Demostración.** Por hipótesis  $X$  e  $Y$  son independientes, así  $E(XY) = E(X)E(Y)$ . Luego, de la ecuación (4.5),  $\text{Cov}(X, Y) = 0$ . Por tanto  $\text{Corr}(X, Y) = 0$ . ◁

El inverso del teorema anterior no es cierto en general. Dos variables aleatorias dependientes pueden ser no correlacionadas. De hecho, aún considerando que  $Y$  sea una función explícita de  $X$ , es posible que  $\text{Corr}(X, Y) = 0$ .

**Teorema 4.17** *Supóngase que  $X$  es una v.a. tal que  $\sigma_X^2 \in \mathbb{R}$  y que  $Y = aX + b$  para  $a, b \in \mathbb{R}$ , donde  $a \neq 0$ . Si  $a \in \mathbb{R}^+$ , entonces  $\text{Corr}(X, Y) = 1$ . Si  $a \in \mathbb{R}^-$ , entonces  $\text{Corr}(X, Y) = -1$ .*

**Demostración.** Si  $Y = aX + b$ , entonces  $\mu_Y = a\mu_X + b$  e  $Y - \mu_Y = a(X - \mu_X)$ . Por tanto, de la definición de Cov,

$$\text{Cov}(X, Y) = aE[(X - \mu_X)^2] = a\sigma_X^2.$$

Dado que  $\sigma_Y = |a|\sigma_X$ , el teorema se deduce de la definición de Corr. ◁

$\text{Corr}(X, Y)$  es una medida que nos dice que tan relacionadas están estas dos variables que tienen una distribución conjunta. Cuando este valor es positivo, significa que la distribución se encuentra concentrada entorno a una línea recta con pendiente positiva. Y cuando tiene valor negativo, la distribución se encuentra concentrada entorno a una recta con pendiente negativa.

Ahora presentamos otra serie de resultados.

**Teorema 4.18** Si  $X$  e  $Y$  son v.a. con varianza finita, entonces

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y). \quad (4.6)$$

**Demostración.** Puesto que  $E(X + Y) = \mu_X + \mu_Y$ , entonces

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

◁

Además,  $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$  para  $a, b \in \mathbb{R}$ . Así también tenemos:

$$\text{Var}(aX + bY + c) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y).$$

En particular,

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

**Teorema 4.19** Si  $X_1, \dots, X_n$  son v.a. tales que  $\text{Var}(X_i) < \infty$  para  $i = \overline{1, n}$ , entonces

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2\sum_{i < j} \text{Cov}(X_i, X_j).$$

**Demostración.** Para cualquier v.a.  $Y$ ,  $\text{Cov}(Y, Y) = \text{Var}(Y)$ . Por tanto, se puede obtener la siguiente relación:

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) = \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j).$$

Separaremos la última suma en dos sumas: (1) la suma de aquellos términos cuando  $i = j$  y (2) la suma de aquellos términos cuando  $i \neq j$ . Entonces, sabiendo que  $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$ , obtenemos:

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Var}(X_i) + 2\sum_{i < j} \text{Cov}(X_i, X_j) \end{aligned}$$

◁

Un caso particular de este teorema ya lo demostramos al decir que la varianza de la suma de v.a. independientes entre si es igual a la suma de sus varianzas individuales.

**Matriz de covarianzas**

Sea  $\mathbf{X}$  un vector aleatorio de  $n$ -dimensional, i.e. de orden  $n \times 1$ , y denotado por:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

y con vector de esperanzas matemáticas:

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \boldsymbol{\mu}$$

Suponiendo que existen, la *matriz de covarianzas* entre estas variables aleatorias se define como:

**Definición 4.15** Definimos la *matriz de covarianzas* como:

$$\begin{aligned} \text{Cov}(\mathbf{X}) &= E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t] \\ &= E \left[ \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_n - \mu_n \end{bmatrix} [X_1 - \mu_1, X_2 - \mu_2, \dots, X_n - \mu_n] \right] \\ &= \begin{bmatrix} E[(X_1 - \mu_1)^2] & E[(X_1 - \mu_1)(X_2 - \mu_2)] & \cdots & E[(X_1 - \mu_1)(X_n - \mu_n)] \\ E[(X_2 - \mu_2)(X_1 - \mu_1)] & E[(X_2 - \mu_2)^2] & \cdots & E[(X_2 - \mu_2)(X_n - \mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(X_n - \mu_n)(X_1 - \mu_1)] & E[(X_n - \mu_n)(X_2 - \mu_2)] & \cdots & E[(X_n - \mu_n)^2] \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix} \end{aligned}$$

Vemos de la definición que los elementos de la diagonal principal son las varianzas marginales de cada una de las variables aleatorias en el vector aleatorio. Pues:

$$\sigma_{ii} = \sigma_i^2 = \text{Var}(X_i)$$

y  $\sigma_{ij}$ , donde  $i \neq j$ , es la covarianza entre las dos variables aleatorias  $X_i$  y  $X_j$ .

La matriz de covarianzas tiene las siguientes propiedades:

1. La matriz de covarianzas es simétrica, ya que,  $\sigma_{ij} = \sigma_{ji}$  para todo  $i \neq j$ .
2. La matriz de covarianzas es siempre definida no negativa (semidefinida positiva).

En efecto, sea  $\mathbf{C} \neq \mathbf{0}$ , un vector de constantes de orden  $n \times 1$ . Entonces:

$$\begin{aligned} \mathbf{C}^t \text{Cov}(\mathbf{X}) \mathbf{C} &= \mathbf{C}^t E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t] \mathbf{C} \\ &= E[\mathbf{C}^t (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t \mathbf{C}] \\ &= E[\mathbf{C}^t (\mathbf{X} - \boldsymbol{\mu})]^2 \geq 0 \end{aligned}$$

puesto que  $\mathbf{C}^t (\mathbf{X} - \boldsymbol{\mu})$  y  $(\mathbf{X} - \boldsymbol{\mu})^t \mathbf{C}$  son escalares y la esperanza de su cuadrado será no negativa.

3. Si las variables  $X_1, X_2, \dots, X_n$  del vector aleatorio  $\mathbf{X}$  son linealmente independientes, la matriz de covarianzas  $\text{Cov}(\mathbf{X})$  es definida positiva.
4. Si todas las covarianzas  $\sigma_{ij} = 0$ , para todo  $i \neq j$ , la matriz de covarianzas  $\text{Cov}(\mathbf{X})$  es diagonal.

En este caso, la traza de la matriz será:

$$\text{tr}(\text{Cov}(\mathbf{X})) = \sum_i \sigma_i^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 + \dots + \sigma_{X_n}^2 = \sigma_{X_1 + X_2 + \dots + X_n}^2.$$

5. Si las variables  $X_1, X_2, \dots, X_n$  son estadísticamente independientes entre sí, o independientes dos a dos, también la matriz de covarianzas será diagonal, y su traza será la calculada en la propiedad anterior.
6. Si todas las variables tienen la misma varianza  $\sigma^2$ , y sus covarianzas son nulas, entonces:

$$\text{Cov} \mathbf{X} = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_{n \times n}$$

siendo  $\mathbf{I}_{n \times n}$  la matriz identidad de orden  $n \times n$ .