

ESTADÍSTICA BÁSICA



Blaise Pascal



Pierre de Fermat



Thomas Bayes



Pierre S. Laplace



**Andrei N.
Kolmogorov**

- [Sucesos y Probabilidades](#)
 - [El espacio de los sucesos.](#)
 - [Azar, suceso aleatorio y probabilidad.](#)
- [Variables aleatorias](#)
 - [Variables aleatorias discretas](#)
 - [Variables aleatorias continuas](#)
 - [Distribución conjunta de dos variables](#)
- [Valor esperado de una variable](#)
- [Momentos de una variable](#)
 - [Momentos respecto del origen](#)
 - [Media aritmética de la variable](#)
 - [Momentos respecto a la media](#)
 - [Varianza](#)
 - [Desviación típica](#)
 - [Coeficiente de variación](#)
 - [Covarianza](#)
 - [Coeficiente de correlación](#)
 - [Propiedades de la varianza](#)
 - [Asimetría](#)
 - [Curtosis](#)

Sucesos y Probabilidades

El espacio de los sucesos.

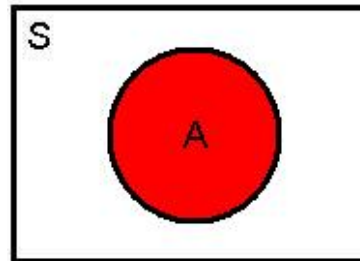
Un **experimento**, en estadística, es cualquier proceso que proporciona datos, numéricos o no numéricos.

Un conjunto cuyos elementos representan todos los posibles resultados de un experimento se llama **espacio muestral** y se representa como S . El espacio muestral de un experimento siempre existe y no es necesariamente único pues, dependiendo de nuestra valoración de los resultados, podemos construir diferentes espacios muestrales.

Los elementos del espacio muestral se llaman puntos muestrales y son los distintos resultados del experimento.

Si consideramos el conjunto de las partes de $(P(S))$ sus elementos son los sucesos. Un **suceso**, por tanto, es un subconjunto del espacio muestral.

Si $A \in P(S) \Rightarrow A$ es un suceso



Existen dos tipos de sucesos:

- Sucesos simples, que son aquellos que comprenden un sólo punto muestral.
- Sucesos compuestos, que son los que engloban más de un punto del espacio muestral. Todo suceso compuesto se puede considerar como unión de puntos del espacio muestral o unión de sucesos simples.

Azar, suceso aleatorio y probabilidad.

El **azar**, en el lenguaje normal, se considera como la característica de un suceso imprevisible.

En estadística esta definición se modifica añadiendo una propiedad adicional: El azar es la característica de un experimento que produce resultados diversos, impredecibles en cada situación concreta, pero cuyas frecuencias, a la larga, tienden a estabilizarse hacia un valor "límite" en el infinito.

Como consecuencia, se definen los **sucesos aleatorios** como los resultados de un experimento cuya variación (la de los resultados) es debida al azar.

La **probabilidad** de un suceso sólo se define para el caso de sucesos aleatorios.

Hay varias formas de definir la probabilidad.

En primer lugar podemos considerar la definición intuitiva que nos dice que la probabilidad de un suceso es la posibilidad de que éste ocurra. Esta primera definición no parece de gran utilidad por ser difícilmente cuantificable.

También podemos considerar la definición clásica de probabilidad. En esta definición se empieza por considerar todos los resultados posibles de un experimento; después se contabilizan los resultados favorables a nuestro suceso, es decir, todos aquellos en que el experimento resulta en el suceso considerado; por último, suponiendo que existe simetría recíproca de todos los resultados, es decir, que todos los resultados posibles son igualmente posibles, se define la probabilidad como el número de casos favorables dividido por el número de casos posibles.

Esta segunda definición presenta el inconveniente de que no siempre es posible saber cuantos son los resultados posibles de un experimento y no siempre todos los resultados posibles son igualmente probables.

Por tanto, consideraremos la probabilidad definida de otra forma. Supongamos que realizamos muchas veces un experimento y vamos anotando el valor de la frecuencia relativa que, como sabemos, tiende a estabilizarse. Suponiendo que pudiéramos realizar el experimento infinitas veces, el valor de estabilización de las frecuencias en el infinito sería la probabilidad de los sucesos. Es decir, la probabilidad es el valor de la frecuencia relativa en el infinito. Es importante señalar, que este valor de estabilización no es un límite en el sentido matemático de la expresión pues, por ser un suceso aleatorio, nadie puede garantizar una ecuación matemática para el valor de la frecuencia relativa.

Todo el cálculo de probabilidades y, con él, toda la estadística se basan en tres propiedades que se asignan a las probabilidades, que se llaman axiomas de Kolmogorov

1. La probabilidad de un suceso es siempre mayor o igual que cero y menor o igual que uno

Si A es un suceso $0 \leq \text{Probabilidad de A} = P(A) \leq 1$

2. La probabilidad del espacio muestral es igual a uno:

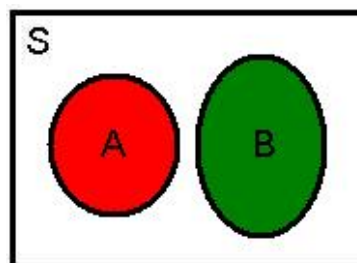
Si S es el espacio muestral $\text{Probabilidad de S} = P(S) = 1$

Es evidente, pues si realizamos un experimento siempre a de suceder alguna cosa. Esta propiedad se expresa como que la probabilidad de un suceso cierto es igual a uno. Si S tiene un único elemento ése es un suceso cierto. Como consecuencia, siguiendo el razonamiento anterior, la probabilidad de que no ocurra nada, lo cual es imposible, o en notación de conjuntos la probabilidad del conjunto vacío (F) es cero. $P(F) = 0$

Se llama suceso imposible a aquel cuya probabilidad vale cero.

3. Si A y B son sucesos mutuamente excluyentes, es decir, nunca ocurren simultáneamente ($A \cap B = F$) la probabilidad de su unión, es decir, de que ocurra uno u otro es la suma de sus probabilidades.

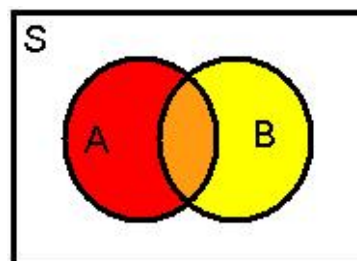
$$P(A \dot{\cup} B) = P(A) + P(B)$$



Otras propiedades de las probabilidades.

- Si A y B son dos sucesos cualesquiera:

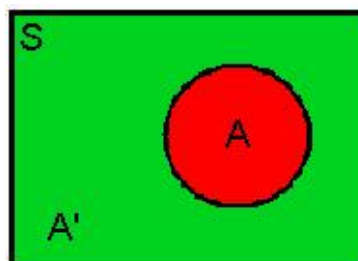
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



- Se llama suceso contrario del suceso A al suceso A' que se define como

$A' = S - A$. La probabilidad del suceso contrario es:

$$P(A') = 1 - P(A)$$



- Se llama probabilidad condicional del suceso B respecto del suceso A a la probabilidad de que, dado que el resultado de un experimento haya sido A sea, simultáneamente, B. Este valor se representa como $P(B|A)$.

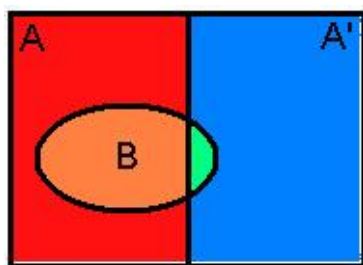
$$P(B|A) = \frac{\text{casos favorables a } A \cap B}{\text{casos favorables a } A} = \frac{P(A \cap B)}{P(A)} \quad \text{si } P(A) > 0$$

Por transposición de términos en la ecuación anterior y en la correspondiente a la probabilidad condicional de A respecto de B llegamos a:

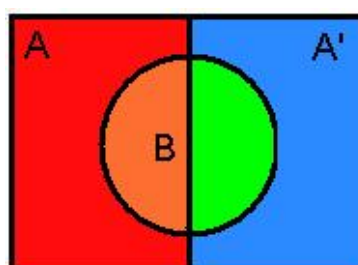
$$P(A \cap B) = P(A) \cdot P(B|A) = P(A|B) \cdot P(B)$$

- Se dice que dos sucesos A y B son independientes si y sólo si la probabilidad de su intersección es igual al producto de sus probabilidades

$$A \text{ y } B \text{ independientes} \Leftrightarrow P(A \cap B) = P(A) \cdot P(B) \Rightarrow \begin{cases} P(A|B) = P(A) \\ P(B|A) = P(B) \end{cases}$$



Sucesos dependientes



Sucesos independientes

Variables aleatorias

Como dijimos, un experimento estadístico es cualquier proceso que proporciona datos. Para su utilización en estadística, estos datos tienen que despojarse de detalles accesorios para convertirse en descripciones numéricas del resultado; la utilización de clasificaciones cualitativas, restringe a la mera descripción las posibilidades de manejo estadístico.

Estas descripciones numéricas son observaciones aleatorias. A las observaciones aleatorias se les considera como la expresión en cada caso concreto de una **variable aleatoria** que toma valores en los resultados del experimento.

Así pues, una variable aleatoria es una función cuyos valores son números reales determinados por los elementos del espacio muestral, es decir, una variable aleatoria es una variable matemática cuyos valores posibles son las descripciones numéricas de todos los resultados posibles de un experimento estadístico.

A los valores posibles de la variable aleatoria se les asigna una probabilidad que es la frecuencia del resultado al que corresponden.

Se pueden distinguir distintos tipos de variables aleatorias según dos criterios de clasificación:

1. **Variables cuantitativas** que son las que resultan de experimentos cuyos resultados son directamente numéricos.
2. **Variables cualitativas** que son las que proceden de experimentos cuyos resultados expresan una cualidad no numérica que necesita ser cuantificada.

Otra clasificación más operativa de las variables aleatorias sería:

- A. **Variable discreta:** Aquella que se define sobre un espacio muestral numerable, finito o infinito. Espacio numerable es aquel cuyos elementos se pueden ordenar, asignándoles a cada uno un número de la serie de los números naturales (del 1 al n ó del 1 al l). Todas las variables con un número finito de valores y todas las que tomen valores en números enteros o racionales (fraccionarios), son variables discretas.
- B. **Variable continua:** Es aquella que se define sobre un espacio asimilable al conjunto de los números reales, es decir, un espacio no numerable (o un espacio infinito de tipo C o infinito dos)

En general, la regla de oro es que todas las variables que proceden de experimentos en los que se cuenta son discretas y todas las variables que proceden de experimentos en los que se mide son continuas.

Variables aleatorias discretas

Función de probabilidad

Una variable aleatoria discreta toma cada uno de sus valores con una determinada probabilidad.

La relación entre valores y probabilidades en una variable X se puede expresar de forma tabular de la siguiente manera:

Valores de X	x_1	x_2	...	x_i
$P(X = x)$	$P(x_1)$	$P(x_2)$		$P(x_i)$

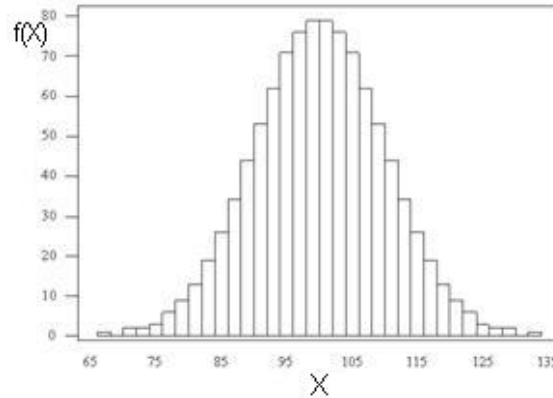
Este método puede ser complicado, e incluso imposible, si los valores de la variable son muchos o infinitos.

En algunos casos, existe una forma sistemática de aplicación de los valores de la probabilidad a los valores de la variable, de modo tal que se puede establecer una ecuación que ligue ambos. A esta ecuación se le llama función de probabilidad. Por tanto, la función de probabilidad de una variable aleatoria discreta X es una función tal que, al sustituir x por un valor de la variable, el valor que toma la función es la probabilidad de que la variable X asuma el valor x . Habitualmente, la función de probabilidad se representa como $f(x)$.

$$f(x) = P(X = x)$$

Las funciones de probabilidad sólo se definen para los valores de la variable aleatoria y deben cumplir tres propiedades:

1. $\forall x f(x) \geq 0$ Como consecuencia del primer axioma.
2. $\sum_x f(x) = 1$ Como consecuencia del segundo axioma.
3. $P(X = x) = f(x)$ Por definición.



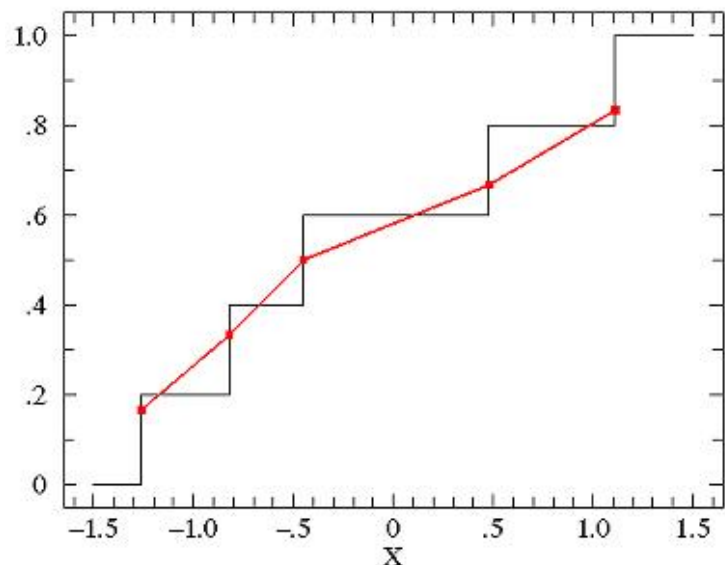
Función de distribución

La función de distribución $F(x)$ de una variable aleatoria discreta X , con función de probabilidad $f(x)$, es una función de la variable en la que al sustituir x por un valor, el valor de la función es la probabilidad de que la variable tome valores menores o iguales que dicho valor x .

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

La función de distribución se define para todos los números reales, no sólo para los valores de la variable. Su máximo es siempre 1 pues cuando el valor que se sustituye es mayor o igual que el valor máximo de la variable, la probabilidad de que ésta tome valores menores o iguales que el sustituido es la probabilidad del espacio muestral. Normalmente, sus valores se dan de forma tabular. Supongamos, por ejemplo que los valores de la variable X sean $x_1, x_2, x_3, \dots, x_n$

$$F(x) = \begin{cases} x < x_1 & 0 \\ x_1 \leq x < x_2 & f(x_1) \\ x_2 \leq x < x_3 & f(x_1) + f(x_2) \\ \dots & \dots \\ x_n \leq x & f(x_1) + f(x_2) + \dots + f(x_n) = 1 \end{cases}$$



VARIABLES ALEATORIAS CONTINUAS

Función de densidad

Una variable aleatoria continua tiene la característica de tomar cada uno de sus valores con probabilidad infinitesimal, a efectos prácticos, 0. Por tanto, no se pueden expresar en forma tabular. Sin embargo, aunque no se pueden considerar probabilidades de valores concretos, puede calcularse la probabilidad de que la variable tome valores en determinados intervalos (los intervalos en cuestión pueden ser abiertos o cerrados, sin que se modifique la probabilidad total).

$$P(a \leq X \leq b) = P(X = a) + P(a < X < b) + P(X = b) = P(a < X < b)$$

Tal como ocurría en el caso de las variables discretas, cuando existe una asignación regular de probabilidad se puede definir una función que nos permita calcular probabilidades para cualquier intervalo de valores, a esta función se le llama función de densidad, $f(x)$

La función de densidad de una variable aleatoria continua X es una función continua tal que su integral entre los extremos de un intervalo nos da el valor de la probabilidad de que X tome valores en ese intervalo.

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

La representación gráfica de la función de densidad en un sistema de ejes cartesianos es la de una curva continua, construida de forma tal que la altura de la curva, sobre el eje de las X , en cada punto es el cociente entre el diferencial de la probabilidad en dicho punto y el diferencial de x . Esta construcción es una extensión por diferenciación del concepto de histograma.

Como consecuencia, la integral de $f(x)$ sobre todo el campo de variación de X es igual a 1.

Es evidente que $f(x)$ es siempre positiva pues si no lo fuera cabría la posibilidad de encontrar intervalos para los cuales la integral sería negativa y eso significaría probabilidad negativa, en abierta contradicción con la definición de probabilidad.

La función de densidad siempre se define para todos los valores en el intervalo

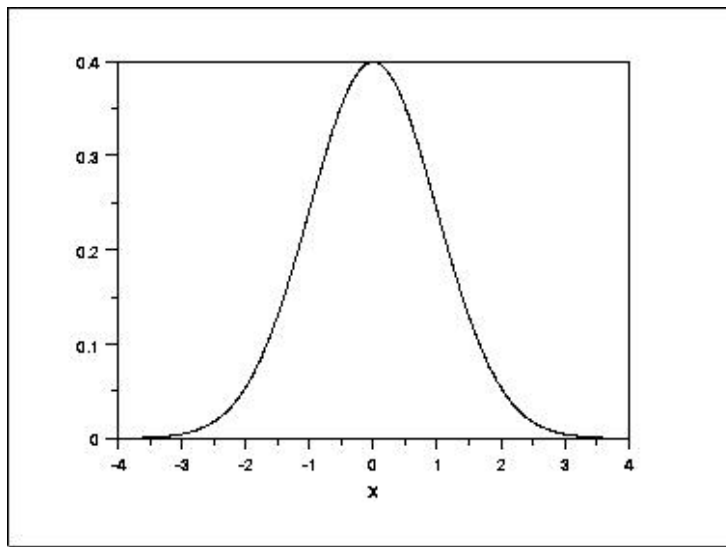
$(-\infty, \infty)$ Esto no ofrece problemas si el campo de variación de X se extiende por todo el intervalo; si no fuera así, la función se define como igual a cero para todos los valores no incluidos en el campo de variación de X .

La función de densidad debe cumplir tres condiciones análogas a las de la función de probabilidad:

1) $\forall x \in \mathbb{R} \quad f(x) \geq 0$ como consecuencia del primer axioma

2) $\int_{-\infty}^{\infty} f(x) dx = 1$ como consecuencia del segundo axioma

3) $P(a \leq X \leq b) = \int_a^b f(x) dx$ por definición



Función de distribución

Para variables continuas también se define la función de distribución, de la siguiente manera:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x) dx$$

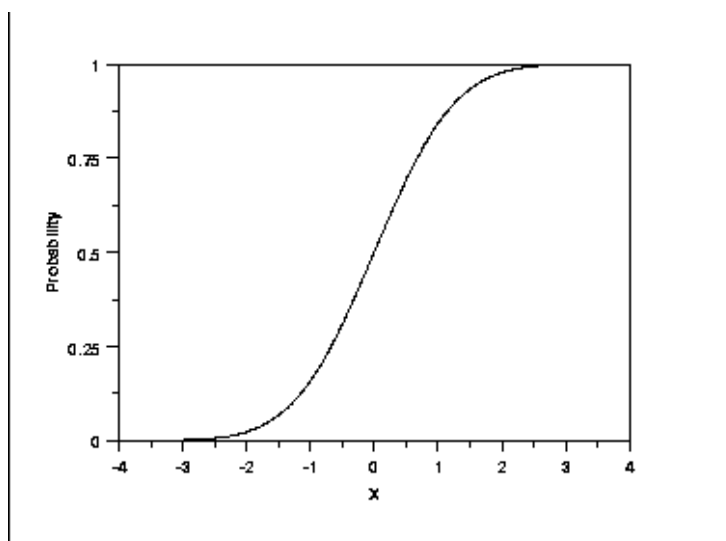
Las características de $F(x)$ son iguales a las expuestas para el caso de las variables discretas, salvo que, obviamente, nunca se expresan en forma tabular.

En general, cualquiera que sea el tipo de variable, las funciones de distribución nos pueden servir para calcular probabilidades. Por ejemplo, en el caso de las variables continuas:

$$P(a \leq X \leq b) = \int_a^b f(x) dx = \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx = F(b) - F(a)$$

Dada su definición, resulta que, para variables continuas, la función de densidad es la derivada respecto a X de la función de distribución.

Las funciones de distribución de las variables continuas más interesantes están tabuladas.



Distribución conjunta de dos variables

Cuando tenemos dos variables aleatorias X e Y , si queremos estudiarlas conjuntamente debemos establecer una relación que ligue los valores de una con los de la otra. Esta relación podrá ser lógica o no, útil o no, en cualquier caso, dadas dos variables cualesquiera y una relación que las ligue se puede pensar en realizar un estudio estadístico conjunto, es decir, aun cuando en la práctica sólo se utilicen variables unidas por nexos lógicos, desde un punto de vista puramente teórico, toda relación imaginable puede ser estudiada.

Así pues, en una situación como esta, para variables discretas, se puede establecer una función de probabilidad para las posibles parejas de valores de ambas variables; a esta función se le llama función de probabilidad conjunta, $f(x,y)$.

Una función de probabilidad conjunta de las variables X e Y es una función de las dos variables tal que, al sustituir la x por un valor de la variable X y la y por un valor de la variable Y , el valor de la función nos da la probabilidad de que X e Y tomen simultáneamente esa pareja de valores anteriormente citados.

$$P[(X = x) \cap (Y = y)] = f(x, y)$$

Las propiedades que debe cumplir la función de probabilidad conjunta son:

1. $\forall x, y \quad f(x, y) \geq 0$ Como consecuencia del primer axioma.
2. $\sum_x \sum_y f(x, y) = 1$ Como consecuencia del segundo axioma.
3. $P[(x, y) \in A, A \subset X \times Y] = \sum_A \sum f(x, y)$ Por definición.

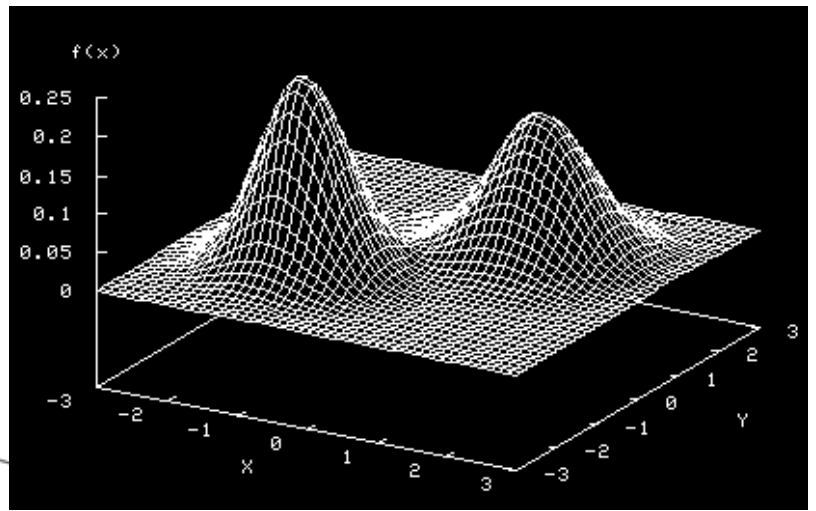
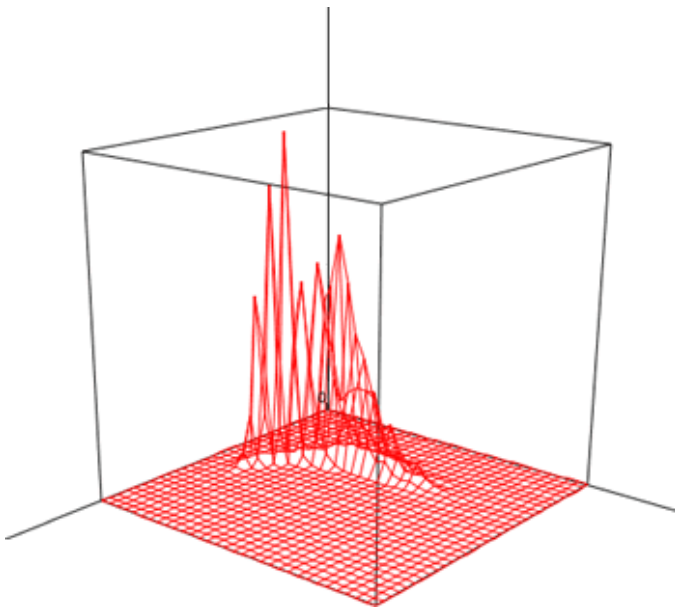
Donde $X \times Y$ es el producto cartesiano de X por Y , o sea, el conjunto de todos las parejas de valores x, y .

Si X e Y son variables continuas, la función que se define es una función de densidad conjunta y es una función que al integrarla respecto de x e y sobre unos intervalos nos da la probabilidad de que la variable tome valores en esos intervalos.

$$\int_a^b \int_c^d f(x, y) dx dy = P[(a < X < b) \cap (c < Y < d)]$$

Que debe de cumplir unas condiciones similares a las anteriores:

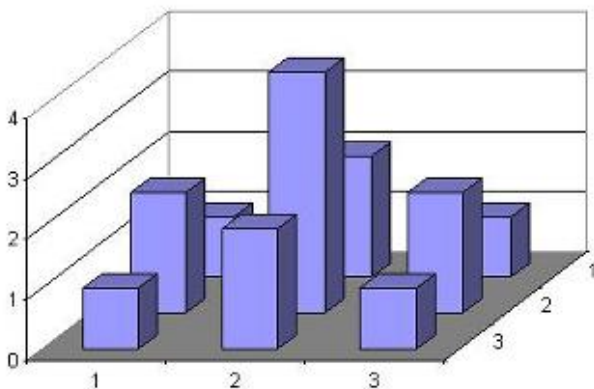
1. $\forall x, y \quad f(x, y) \geq 0$ Como consecuencia del primer axioma.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$ Como consecuencia del segundo axioma.
3. $P[(x, y) \in A, A \subset X \times Y] = \iint_A f(x, y) dx dy$ Por definición.



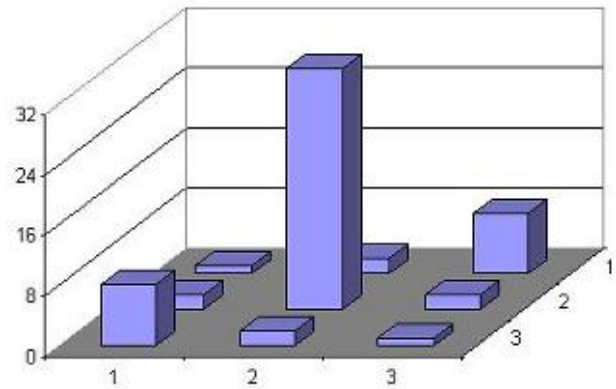
VARIABLES ALEATORIAS INDEPENDIENTES

Dos variables aleatorias X e Y , discretas o continuas cuyas funciones de probabilidad o densidad son $g(x)$ y $h(y)$, respectivamente, con función de probabilidad o densidad conjunta $f(x, y)$, son estadísticamente independientes si y sólo si

$$f(x, y) = g(x) \cdot h(y)$$



VARIABLES INDEPENDIENTES



VARIABLES DEPENDIENTES

VALOR ESPERADO DE UNA VARIABLE

Supongamos que hemos realizado n veces un experimento aleatorio que genera una variable X . El valor medio del experimento en estas n repeticiones es la suma de los productos de los valores de la variable por su frecuencia relativa. Cuando n sea igual a infinito, el valor medio del experimento se llama valor esperado o esperanza matemática, $E[X]$.

Si X es una variable discreta con función de probabilidad $f(x)$, el valor esperado de X se calcula según decíamos anteriormente sumando los productos de los valores de la variable por sus respectivas probabilidades.

$$E[X] = \sum_x x f(x)$$

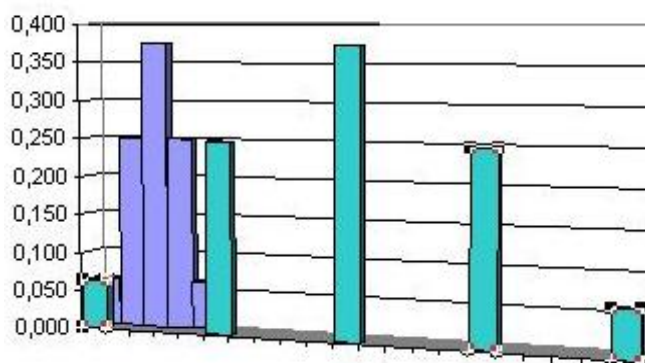
En el caso de una variable continua

$$\mu_3 = E[(X - \mu)^3]$$

Propiedades del valor esperado

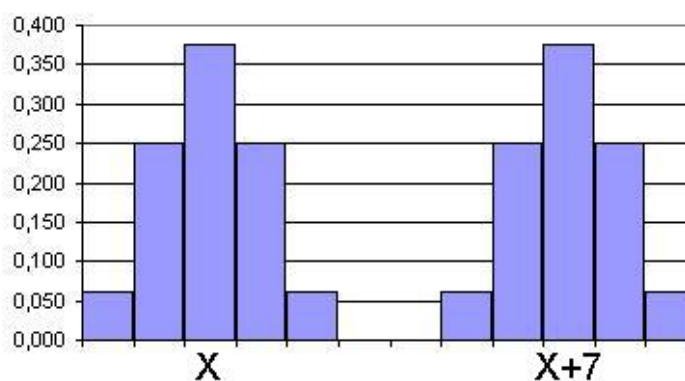
- Al multiplicar todos los valores de una variable por una misma constante, el valor esperado de ésta queda multiplicado por el valor de la constante.

$$\forall a \in \mathbb{R} \quad E[aX] = aE[X]$$



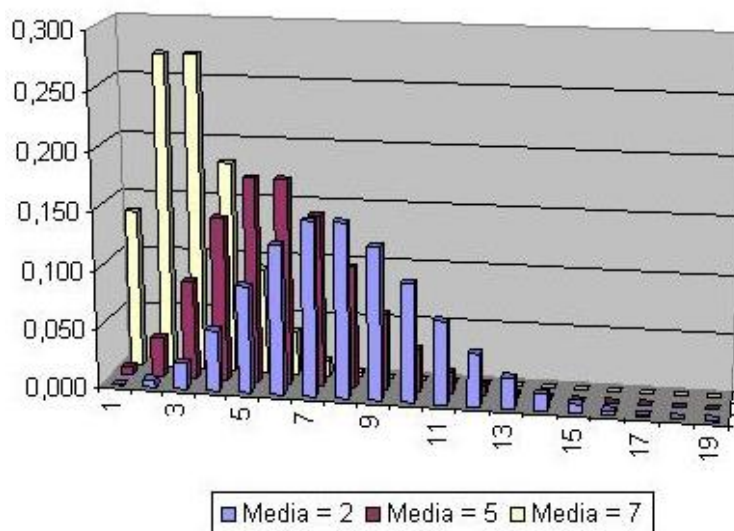
- Al sumar a todos los valores de una variable una misma constante, el valor esperado de ésta queda incrementado por el valor de la constante.

$$\forall b \in \mathbb{R} \quad E[X+b] = E[X] + b$$



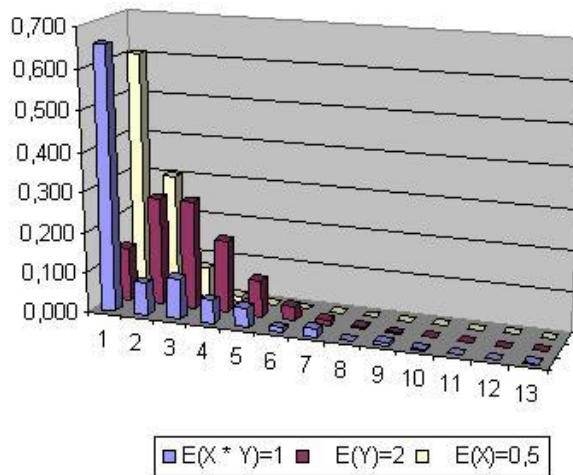
- Si tenemos dos variables X e Y, discretas o continuas, el valor esperado de su suma o diferencia es la suma o diferencia de sus valores esperados

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



- Si las variables anteriores, X e Y son variables aleatorias independientes ocurre que el valor esperado de su producto es igual al producto de sus valores esperados.

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



Es importante indicar que la independencia de las variables es condición suficiente pero no necesaria para que el valor esperado del producto de dos variables sea igual al producto de sus valores esperados, es decir, ésta es una propiedad de las variables independientes pero se cumple en variables que no son independientes.

Momentos de una variable

Momentos respecto del origen

Dada una variable aleatoria X con función de probabilidad o densidad $f(x)$ podemos definir una función de X que sea igual a la variable elevada a un exponente entero no negativo.

$$z(x) = x^k \text{ siendo } k \in \mathbb{Z} \quad k \geq 0$$

El valor esperado de $z(x)$ es el k-ésimo momento de la variable X respecto a su origen y se llama

$$\mu'_k = E[x^k] = \begin{cases} \sum_x x^k f(x) & \text{Si X es discreta} \\ \int_{-\infty}^{\infty} x^k f(x) dx & \text{Si X es continua} \end{cases}$$

- $k = 0 \quad \mu'_0 = E[x^0] = 1$

- $k = 1 \quad \mu'_1 = E[x^1] = E[x] = \mu_x = \mu$

a este primer momento respecto al origen que es igual al valor esperado se le llama también **media aritmética de la variable** y se le denomina μ_x , simplemente μ .

En la mayoría de los casos, la media μ expresa la tendencia central de la variable o el orden de magnitud de sus valores.

El resto de los momentos respecto al origen tienen escaso interés en la mayoría de los casos.

Momentos respecto a la media

Dada una variable aleatoria X con función de probabilidad o densidad $f(x)$ podemos definir una función de X que sea igual a la diferencia entre la variable y su media aritmética elevada a un exponente entero no negativo.

$$z(x) = (x - \mu)^k \text{ siendo } k \in \mathbb{Z} \quad k \geq 0$$

El valor esperado de $z(x)$ es el k-ésimo momento de la variable X respecto a la media y se llama μ_k .

$$\mu_k = E[(x - \mu)^k] = \begin{cases} \sum_x (x - \mu)^k f(x) & \text{Si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx & \text{Si } X \text{ es continua} \end{cases}$$

- $k = 0$ $\mu_0 = E[(x - \mu)^0] = 1$
- $k = 1$ $\mu_1 = E[(x - \mu)^1] = E[(x - \mu)] = E[X] - \mu = 0$

es decir, en cualquier variable aleatoria su primer momento respecto de la media es igual a 0. Esta propiedad se utiliza reiteradamente en las demostraciones estadísticas.

- $k = 2$ $\mu_2 = E[(x - \mu)^2] = \sigma_x^2 = \sigma^2$

este segundo momento respecto de la media se le llama también **varianza**.

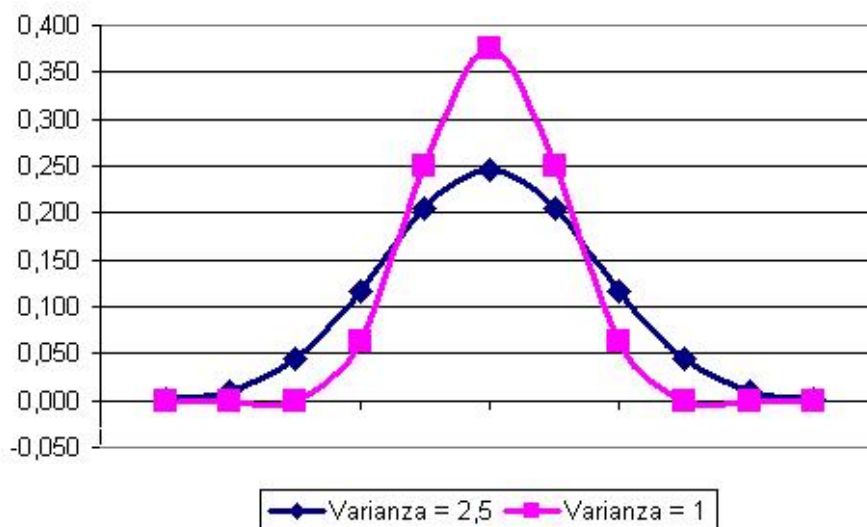
$$\sigma^2 = E[(x - \mu)^2] = \begin{cases} \sum_x (x - \mu)^2 f(x) & \text{Si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx & \text{Si } X \text{ es continua} \end{cases}$$

La varianza de una variable mide la dispersión de sus valores respecto al valor central μ .

Para calcular la varianza por un método más sencillo se utiliza la expresión:

$$\sigma^2 = E[(x - \mu)^2] = E[X^2] - \mu^2 = E[X^2] - E[X]^2$$

Es decir, la varianza de una variable es igual a la media de los cuadrados menos el cuadrado de la media.



El principal problema de la varianza es que se expresa en unidades cuadráticas que no siempre tienen una interpretación clara. Para obviar este problema se define otra medida de la dispersión que es la **desviación típica**, σ_x , o simplemente σ , que se calcula como la raíz cuadrada positiva de la varianza; evidentemente, la desviación típica se mide en las mismas unidades que la variable

$$\sigma_x = +\sqrt{\sigma_x^2}$$

No obstante, la desviación típica no resuelve todos los problemas que se pueden plantear, como por ejemplo la comparación de situaciones en las que la unidad de medida o el orden de magnitud de esta sea diferente. Para resolver esta cuestión se define una medida adimensional de la variabilidad que es el **coeficiente de variación**, C V, que se calcula como el cociente entre la desviación típica y la media (a veces este cociente se expresa en tanto por ciento multiplicándolo por 100).

$$C.V. = \frac{\sigma}{\mu} \quad C.V. = 100 \cdot \frac{\sigma}{\mu}$$

En este contexto de la medida de la variación se plantea el problema de medir la variación conjunta de variables de variables asociadas.

Supongamos que tenemos dos variables aleatorias X e Y, discretas o continuas, con función de probabilidad o densidad conjunta $f(x,y)$ y definimos una función $z(x,y)$ igual al producto de las desviaciones de cada valor a su media respectiva (es decir, $z(x,y)$ tiene la misma estructura que $(X - \mu)^2 = (X - \mu)(X - \mu)$ si sustituimos una vez a X por Y).

$$z(x,y) = (x - \mu_x)(y - \mu_y)$$

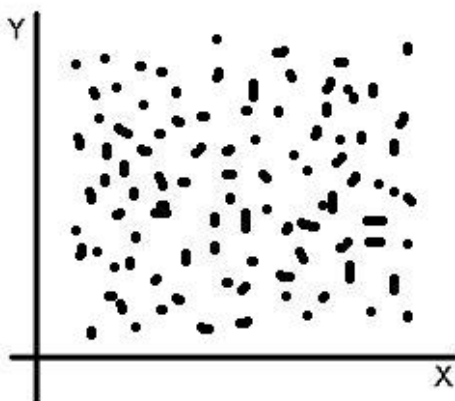
Al valor esperado de $z(x,y)$ se le llama **covarianza** de las variables X e Y y se representa como σ_{xy} o $cov(x,y)$.

$$X \text{ e } Y \text{ discretas} \Rightarrow \sigma_{xy} = E[(x - \mu_x)(y - \mu_y)] = \sum_x \sum_y (x - \mu_x)(y - \mu_y) f(x,y)$$

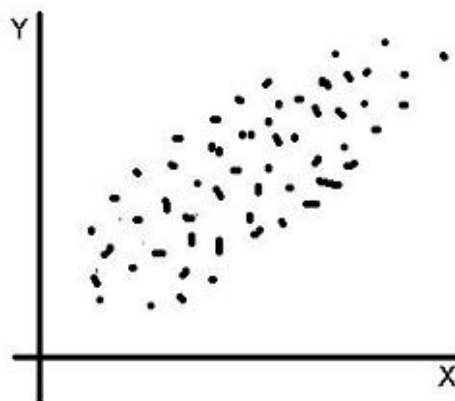
$$X \text{ e } Y \text{ continuas} \Rightarrow \sigma_{xy} = E[(x - \mu_x)(y - \mu_y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x,y) dx dy$$

La covarianza es una medida de la variación común a dos variables y, por tanto, una medida del grado y tipo de su relación.

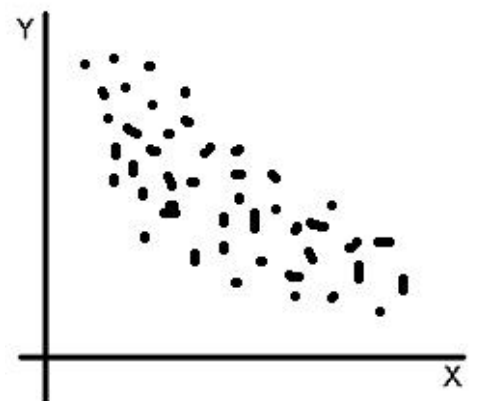
- σ_{xy} es positiva si los valores altos de X están asociados a los valores altos de Y y viceversa.
- σ_{xy} es negativa si los valores altos de X están asociados a los valores bajos de Y y viceversa.
- Si X e Y son variables aleatorias independientes $cov(x,y) = 0$.
- La independencia es condición suficiente pero no necesaria para que la $cov(x,y)$ sea nula.



$cov(x,y) = 0$



$cov(x,y) > 0$



$cov(x,y) < 0$

Se puede deducir, algebraicamente, un medio más sencillo para calcular la covarianza de dos variables.

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E[X \cdot Y] - \mu_X \mu_Y = E[X \cdot Y] - E[X] \cdot E[Y]$$

En el caso de la covarianza tenemos el mismo problema que se nos presentó con la varianza, es decir, la covarianza se expresa en términos del producto de las unidades de medida de ambas variables, lo cual no siempre es fácilmente interpretable. Por otra parte también es difícil comparar situaciones diferentes entre sí. En este caso, ambos problemas se solucionan de una vez mediante la definición del **coeficiente de correlación**, ρ , que se define como el cociente entre la covarianza y el producto de las desviaciones típicas de las dos variables.

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

La correlación toma valores entre -1 y 1, siendo su signo igual al de la covarianza. Correlaciones con valor absoluto 1 implican que existe una asociación matemática lineal perfecta, positiva o negativa, entre las dos variables y correlaciones iguales a 0 implican ausencia de asociación. Obviamente, las variables independientes tienen correlación 0, pero nuevamente, la independencia es condición suficiente pero no necesaria.

Correlaciones con valores absolutos intermedios indican cierto grado de asociación entre los valores de las variables.

Propiedades de la varianza

Si X es una variable aleatoria con función de probabilidad o densidad $f(x)$, la varianza de una función de la variable X , $m(x)$, se calcula según la expresión:

$$\sigma_{m(X)}^2 = E\left[\left(m(X) - \mu_{m(X)}\right)^2\right] = E\left[\left(m(X) - E[m(X)]\right)^2\right]$$

Casos concretos:

1. Cuando a todos los valores de una variable se les suma una constante, la varianza de la variable conserva el mismo valor (ver imagen en las propiedades de la media)

$$\forall b \in \mathbb{R} \quad m(X) = X + b \Rightarrow \sigma^2(X + b) = \sigma_X^2$$

2. Cuando a todos los valores de una variable se les multiplica por una constante, la varianza de la variable queda multiplicada por el valor de la constante elevado al cuadrado (ver imagen en las propiedades de la media)

$$\forall a \in \mathbb{R} \quad m(X) = a \cdot X \Rightarrow \sigma^2(a \cdot X) = a^2 \cdot \sigma_X^2$$

3. Si X e Y son dos variables aleatorias con función de densidad o probabilidad conjunta $f(x,y)$, la varianza de la función $m(x,y) = aX \pm bY$, donde a y b son constantes reales se calcula como:

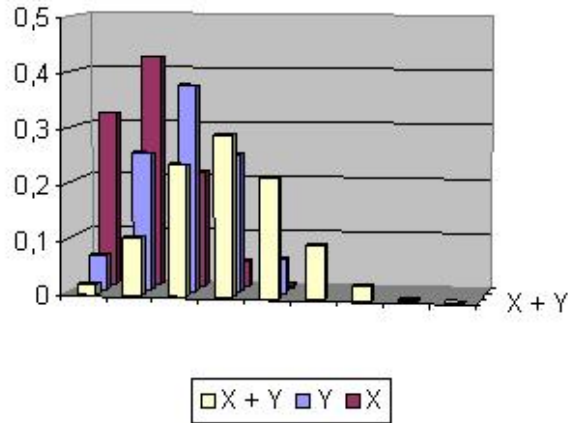
$$\sigma^2(aX \pm bY) = a^2 \cdot \sigma_X^2 + b^2 \cdot \sigma_Y^2 \pm 2 \cdot a \cdot b \cdot \sigma_{XY}$$

En el caso de que $a = b = 1$ $\sigma^2(X \pm Y) = \sigma_X^2 + \sigma_Y^2 \pm 2 \cdot \sigma_{XY}$

Si además ocurre que X e Y sean independientes $\sigma_{xy} = 0$, luego

$$X \text{ Y independientes } \Rightarrow \sigma^2(X \pm Y) = \sigma^2(X) + \sigma^2(Y)$$

Varianza de la suma



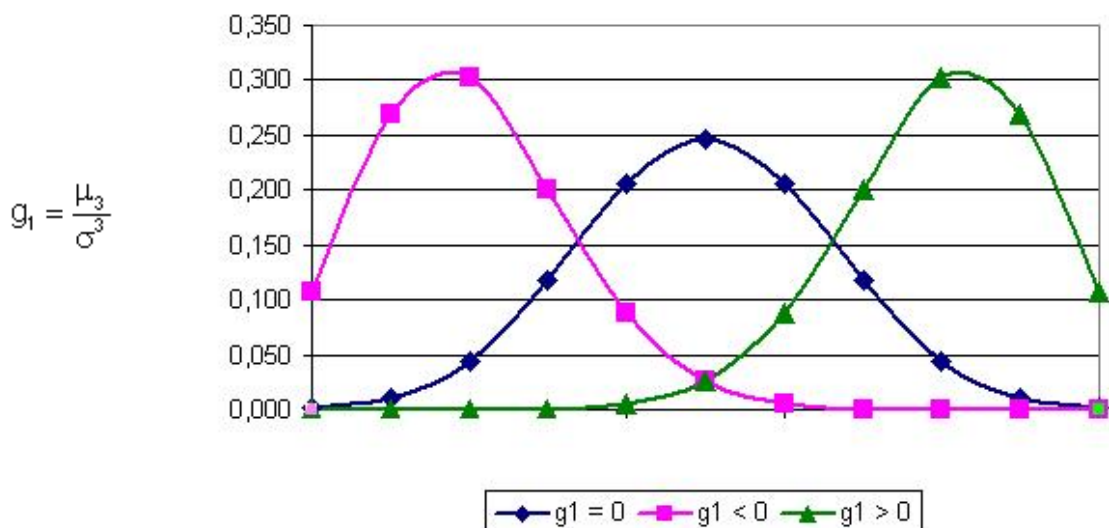
Volviendo al tema de los momentos respecto al origen, veamos los dos siguientes que también son interesantes,

➤ $k = 3$

$$\mu_3 = E[(x - \mu)^3] = \text{asimetría}$$

El tercer momento respecto de la media mide la asimetría de la distribución, es decir, si existen o no observaciones muy extremas en algún sentido con frecuencias razonablemente altas. Si la asimetría es negativa, la variable toma valores muy bajos con mayor frecuencia que valores muy altos y se dice que tiene una cola izquierda pesada o que es asimétrica hacia la izquierda. Si la asimetría es positiva, la variable toma valores muy altos con mayor frecuencia que valores muy bajos y se dice que tiene una cola derecha pesada o que es asimétrica hacia la derecha. Si la asimetría es cero, los valores bajos y altos de la variable tienen probabilidades iguales (el ejemplo más típico de variable simétrica es la variable normal)

La asimetría tiene el mismo problema que la varianza y la covarianza en cuanto a sus unidades de medida y, por ello, normalmente se utiliza una medida adimensional de la asimetría que es el **coeficiente de asimetría**, g_1 , que se calcula como el cociente entre el tercer momento y el cubo de la desviación típica.



➤ $k = 4$ $\mu_4 = E[(X - \mu)^4] = \text{curtosis}$

El cuarto momento respecto de la media mide la curtosis de la distribución, es decir, la forma de la distribución de probabilidad. Al representar gráficamente variables con curtosis pequeña, platicúrticas, se observan curvas o histogramas con colas cortas y aspecto aplanado o en meseta; si la variable tiene curtosis grande, es decir, si es leptocúrtica, su gráfica ser alta y estilizada, con colas largas y pesadas.

La curtosis de una variable siempre es positiva y se mide en la unidades de la variable elevadas a potencia 4. Por tanto, nuevamente se nos plantean los problemas relacionados con las unidades de medida y las escalas y necesitamos una medida adimensional de la curtosis. Esta medida adimensional de la curtosis es el **coeficiente de curtosis**, g_2 , que se calcula como el cociente entre el cuarto momento y el cuadrado de la varianza, al que se le resta 3 unidades. Esta corrección se debe a que, sin ella, las variables normales tendrían coeficiente de curtosis igual a 3; al restar 3 conseguimos que el coeficiente de curtosis de la variable normal sea 0 y que las variables platicúrticas tengan coeficiente de curtosis negativo y la leptocúrticas positivo, lo cual es más mnemotécnico que la distinción entre curtosis pequeña y grande.

$$g_2 = \frac{\mu_4}{\sigma^4} - 3$$

