Clase 10: Varianza y covarianza

Matías Carrasco

16 de agosto de 2019

Resumen Haciendo una analogía con la mecánica veremos cómo el momento de inercia sirve para medir la dispersión de una variable alrededor de su valor esperado. Esto nos llevará a los conceptos de varianza y covarianza.

Índice

La varianza 1
Propiedades de la varianza 4
Covarianza y correlación 8
Propiedades de la covarianza 11
Interpretación de la covarianza 12

La varianza

Nuestro objetivo ahora es definir una medida de la dispersión de una variable aleatoria. Para esto usaremos una analogía con la mecánica.

Imaginemos una distribución de bloques cualquiera, en la cual hemos ubicado el centro de masa, y supongamos que deseamos girar el tablón respecto del eje vertical que pasa por el centro masa. ¿Cuándo es más difícil girarlo?

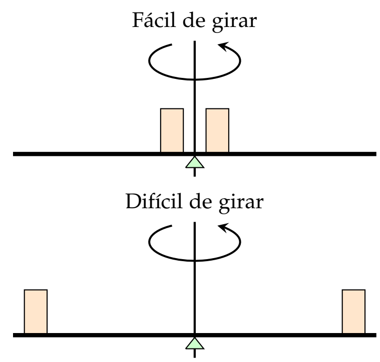

Figura 1: Cuanto más dispersión, más difícil es girar en torno al eje que pasa por el centro de masa del conjunto de bloques.

Intuitivamente es claro que si los bloques están muy concentrados sobre su centro de masa, entonces girarlos resulta sencillo, y ésto se hace cada vez más difícil a medida que dispersamos los bloques, ver la Figura 1. Podemos medir entonces la dispersión de un conjunto de bloques alrededor de su centro de masa por la dificultad de girarlos.

En física existe una cantidad que mide justamente la dificultad que tiene un objeto a ser girado en torno a un eje. Se llama *momento de inercia*, y es como la masa para movimientos rotatorios. Sin entrar en detalles sobre la definición, recordar que el momento de una partícula de masa m y velocidad v es $p = mv$. El momento angular es la fórmula análoga para movimientos alrededor de un eje: si la partícula gira entorno a un eje a distancia r, entonces

$$L = p \cdot r = (mv) \cdot r = (mr^2) \cdot \omega = I\omega,$$

en donde ω es la velocidad angular e I es el momento de inercia.

Si hay muchas partículas de diferentes masas, el momento de inercia es la suma

$$I = \sum_i m_i r_i^2.$$

de los momentos de inercia de cada partícula.

Si se trata de una distribución continua de masa, entonces la definición del momento de inercia es

$$I = \int r^2 dm.$$

Es decir, simplemente cambiamos la suma por una integral.

Apliquemos esta fórmula a nuestros bloques en el tablón. Supongamos que X es una variable discreta, con recorrido $\{x_1, x_2, \ldots\}$ y función de probabilidad puntual $p_i = p_X(x_i)$. Disponemos como antes, para cada $i \geq 1$ un bloque de peso p_i en la posición x_i.

El centro de masa está en la posición $E(X)$, por lo que la distancia de cada bloque al centro de masa es

$$r_i = x_i - E(X).$$

Como la masa es $m_i = p_i$, el momento de inercia de X es

$$I = \sum_{i \geq 1} (x_i - E(X))^2 p_i.$$

Si se trata de una variable continua, en lugar de bloques tenemos una densidad de probabilidad. En este caso la masa que se encuentra en la posición x del tablón viene representada por $dm(x) = p(x)dx$ en donde $p(x)$ es la densidad de X. Por lo tanto, el momento de inercia se escribe

$$I = \int_{-\infty}^{\infty} (x - E(X))^2 p(x) dx.$$

En probabilidad I se llama la *varianza* de X y se escribe $Var(X)$.

Definición de varianza

- Sea X una variable discreta con función de probabilidad puntual $p(x)$. La varianza de X es por definición

$$Var(X) = \sum_{x \in R_X} (x - E(X))^2 p(x).$$

- Sea X una variable continua con densidad $p(x)$. La varianza de X es por definición

$$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 p(x) dx.$$

Esta cantidad mide la dispersión de X entorno a su valor esperado.

Más allá del significado físico de la varianza, es importante recordar que la varianza mide cuán dispersos están los valores de X respecto de su valor esperado. Dicho de forma sencilla, la varianza mide el "ancho" de la gráfica de la función de probabilidad puntual.

Muchas veces la varianza de una variable X se denota por σ^2 o σ_X^2. La raíz cuadrada de la varianza $\sigma = \sqrt{Var\,(X)}$ se llama *desvío estándar* de X. El desvío σ tiene las mismas unidades que X, mientras que la varianza tiene las unidades del cuadrado de X. Por ejemplo, si X se mide en metros, entonces σ^2 tiene unidades de metros cuadrados. Como σ y X tienen las mismas unidades, se suele usar el desvío como medida de dispersión.

Ejemplo 1 Para cada una de las variables X, Y, Z y W cuyas f.p.p. se muestran abajo, calcular la varianza.

1.

valor x	1	2	3	4	5
f.p.p. $p(x)$	1/5	1/5	1/5	1/5	1/5

2.

valor y	1	2	3	4	5
f.p.p. $p(y)$	1/10	2/10	4/10	2/10	1/10

3.

valor z	1	2	3	4	5
f.p.p. $p(z)$	1/2	0	0	0	1/2

4.

valor w	1	2	3	4	5
f.p.p. $p(w)$	0	0	1	0	0

Cada una de las variables tiene el mismo valor esperado, igual a 3, pero la probabilidad está distribuida de modo diferente. En los gráficos de la Figura 2, hemos ordenado las f.p.p. de mayor a menor varianza: Z, X, Y, W.

Ahora vamos a verificar nuestra intuición visual calculando la varianza de cada una de las variables. Lo haremos usando tablas.

1.

valor x	1	2	3	4	5
f.p.p. $p(x)$	1/5	1/5	1/5	1/5	1/5
$(X-3)^2$	4	1	0	1	4

$Var(X) = \frac{4}{5} + \frac{1}{5} + \frac{0}{5} + \frac{1}{5} + \frac{4}{5} = 2$

2.

valor y	1	2	3	4	5
f.p.p. $p(y)$	1/10	2/10	4/10	2/10	1/10
$(Y-3)^2$	4	1	0	1	4

$Var(Y) = \frac{4}{10} + \frac{2}{10} + \frac{0}{10} + \frac{2}{10} + \frac{4}{10} = 1.2$

3.

valor z	1	2	3	4	5
f.p.p. $p(z)$	1/2	0	0	0	1/2
$(Z-3)^2$	4	1	0	1	4

$Var(Z) = \frac{20}{10} + \frac{20}{10} = 4$

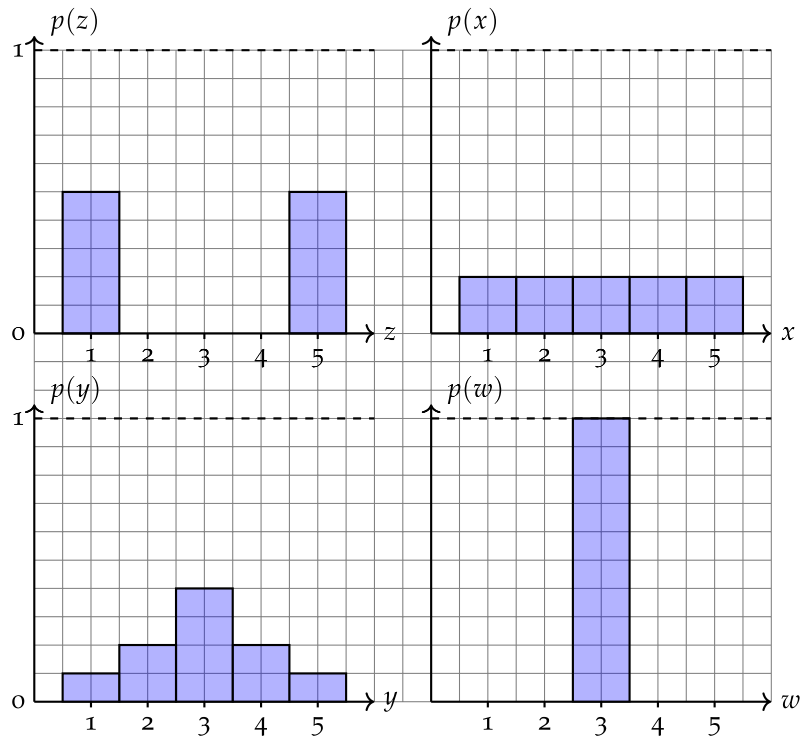

Figura 2: Varias distribuciones con el mismo valor esperado pero con diferentes varianzas.

4.

valor w	1	2	3	4	5
f.p.p. $p(w)$	0	0	1	0	0
$(W-3)^2$	4	1	0	1	4

$Var(W) = 0$. Observar que W no varía, es constante igual a 3, por lo que su varianza es cero.

■

Propiedades de la varianza

Usando la fórmula del valor esperado de una función de una variable aleatoria podemos escribir de forma más compacta la definición de varianza. De hecho, notar que si tomamos la función $g(x) = (x - E(X))^2$, entonces

$$E(g(X)) = \sum_{x \in R_X} (x - E(X))^2 p(x) = Var(X) \qquad \text{(discreto)}$$

$$E(g(X)) = \int_{-\infty}^{\infty} (x - E(X))^2 p(x) dx = Var(X) \qquad \text{(continuo)}$$

Es decir, $Var(X) = E\left((X - E(X))^2\right)$.

Por ejemplo, si X es una variable *centrada*, lo cual quiere decir que $E(X) = 0$, entonces $Var(X) = E\left(X^2\right)$.

Veamos entonces, algunas propiedades de la varianza que nos permitan simplificar su cálculo.

La varianza no cambia si sumamos una constante

Sean X una variable aleatoria y c una constante cualquiera. Entonces $Var\left(X+c\right) = Var\left(X\right)$.

Demostración. La prueba es muy simple: notar que $E\left(X+c\right) = E\left(X\right) + c$ por la linealidad de la esperanza. Entonces

$$\begin{aligned} Var\left(X+c\right) &= E\left(\left(X+c-E\left(X+c\right)\right)^2\right) = E\left(\left(X+c-E\left(X\right)-c\right)^2\right) \\ &= E\left(\left(X-E\left(X\right)\right)^2\right) = Var\left(X\right). \end{aligned}$$

□

La varianza es cuadrática

Sean X una variable aleatoria y c una constante cualquiera. Entonces $Var\left(cX\right) = c^2 Var\left(X\right)$.

Demostración. La prueba también es muy fácil. Notar primero que $E\left(cX\right) = cE\left(X\right)$, de donde

$$\begin{aligned} Var\left(cX\right) &= E\left(\left(cX-E\left(cX\right)\right)^2\right) = E\left(\left(cX-cE\left(X\right)\right)^2\right) \\ &= E\left(c^2\left(X-E\left(X\right)\right)^2\right) = c^2 E\left(\left(X-E\left(X\right)\right)^2\right) = c^2 Var\left(X\right). \end{aligned}$$

□

Una fórmula útil para la varianza

La varianza de una variable X se puede calcular mediante la siguiente igualdad $Var\left(X\right) = E\left(X^2\right) - E\left(X\right)^2$.

Demostración. Para probarla, basta desarrollar el cuadrado

$$\left(X-E\left(X\right)\right)^2 = X^2 + E\left(X\right)^2 - 2E\left(X\right)X,$$

de donde al tomar esperanza

$$Var\left(X\right) = E\left(X^2\right) + E\left(X\right)^2 - 2E\left(X\right)^2 = E\left(X^2\right) - E\left(X\right)^2,$$

que es lo que queríamos probar. □

La varianza de la suma de independientes

Si X e Y son independientes: $Var(X+Y) = Var(X) + Var(Y)$.

Demostración. Para probar esta propiedad, recordar primero que si X e Y son independientes, entonces $E(XY) = E(X)E(Y)$. Entonces, usando que $E(X+Y) = E(X) + E(Y)$ vale siempre, tenemos

$$\begin{aligned} Var(X+Y) &= E\left((X+Y-E(X+Y))^2\right) \\ &= E\left((X-E(X)+Y-E(Y))^2\right). \end{aligned}$$

Desarrollando el cuadrado, tenemos que el lado derecho de la última ecuación es igual a

$$E\left((X-E(X))^2\right) + E\left((Y-E(Y))^2\right) + 2E\left((X-E(X))(Y-E(Y))\right).$$

Pero este es a su vez igual a

$$Var(X) + Var(Y) + 2E\left((X-E(X))(Y-E(Y))\right).$$

Luego, basta probar que

$$E\left((X-E(X))(Y-E(Y))\right) = 0.$$

Esto se deduce de

$$E\left((X-E(X))(Y-E(Y))\right) = E(XY) - E(X)E(Y) = 0,$$

pues X e Y son independientes. □

Ejemplo 2 — Varianza de una Bernoulli. Sea X una variable con distribución Bernoulli de parámetro p. El valor esperado es $E(X) = p$. Entonces, de la definición tenemos

$$\begin{aligned} Var(X) &= (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p \\ &= (1-p)\left[p^2 + (1-p)p\right] = p(1-p). \end{aligned}$$

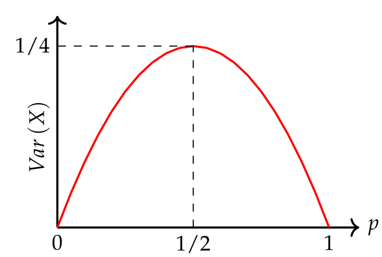

Figura 3: Varianza de una variable Bernoulli en función del parámetro p.

En la Figura 3 se muestra la varianza de X en función de p. Notar que el máximo se da cuando $p = 1/2$ y vale $Var(X) = 1/4$. ■

Ejemplo 3 Una moneda tiene probabilidad p de salir cara. Se la lanza hasta obtener cara, y sea X la cantidad de lanzamientos necesarios. Calculemos la varianza de X.

Recordar que la esperanza de X es igual a $1/p$. Entonces, de la definición tenemos

$$Var(X) = E\left(X^2\right) - E(X)^2 = E\left(X^2\right) - \frac{1}{p^2}.$$

Falta evaluar el primer término. Notar que éste se puede descomponer en

$$E\left(X^2\right) = E\left(X(X-1)\right) + E\left(X\right) = E\left(X(X-1)\right) + \frac{1}{p}.$$

La esperanza de $X(X-1)$ es por definición igual a

$$\begin{aligned}\sum_{k=1}^{\infty} k(k-1)p(1-p)^{k-1} &= p(1-p)\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} \\ &= p(1-p)\frac{d^2}{dp^2}\left[\frac{1}{p}\right].\end{aligned}$$

Es decir, tenemos que

$$E\left(X^2\right) = \frac{2(1-p)}{p^2} + \frac{1}{p}.$$

Juntando todo, nos queda

$$Var\left(X\right) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}.$$

Hemos probado entonces que la varianza es $(1-p)/p^2$. ■

Ejemplo 4 — Varianza de una uniforme. Sea X con distribución uniforme en (a,b). Calcular $E\left(X\right)$ y $Var\left(X\right)$.

Un simple cambio de escala transforma el intervalo (a,b) en $(0,1)$. La distribución uniforme en (a,b) se transforma en la distribución uniforme en $(0,1)$, cuya densidad es simplemente 1 en $(0,1)$ y 0 en cualquier otro lugar.

En términos de variables aleatorias, cualquier problema que involucre una variable aleatoria X, uniforme en (a,b), se reduce fácilmente a uno que involucre una variable aleatoria U uniforme en $(0,1)$, definida por

$$U = \frac{X-a}{b-a}, \quad \text{o} \quad X = a + (b-a)U.$$

Este tipo de cambio de escala, o cambio lineal de variable, es una técnica básica para reducir los problemas al caso más simple y evitar cálculos innecesarios.

Para ilustrarlo, el valor esperado de X es

$$\begin{aligned}E\left(X\right) &= E\left(a+(b-a)U\right) = a + (b-a)E\left(U\right) \\ &= a + (b-a)\int_0^1 u du = a + (b-a)\frac{1}{2} = \frac{a+b}{2}\end{aligned}$$

Esto es obvio de todas maneras debido a la simetría de la distribución respecto al punto medio del intervalo (a,b). La varianza de X es

$$\begin{aligned}Var\left(X\right) &= Var\left(a+(b-a)U\right) = (b-a)^2 Var\left(U\right) \\ &= (b-a)^2\left[E\left(U^2\right) - E\left(U\right)^2\right] \\ &= (b-a)^2\left[E\left(U^2\right) - \frac{1}{4}\right]\end{aligned}$$

La esperanza de U^2 es la integral

$$\int_0^2 u^2 du = \left[\frac{u^3}{3}\right]_0^1 = \frac{1}{3},$$

de donde

$$Var\,(X) = (b-a)^2\left[\frac{1}{3} - \frac{1}{4}\right] = \frac{(b-a)^2}{12}.$$

■

Covarianza y correlación

¿Cómo hacemos para medir la dependencia entre dos variables? En esta sección veremos una primera aproximación a este problema.

Sean X e Y dos variables aleatorias con varianza finita. Comencemos por notar lo siguiente: si la varianza $Var\,(X-Y)$ es pequeña, entonces intuitivamente Y es muy parecida a X. En general, Y puede depender fuertemente de X sin ser igual a ella.

Por ejemplo, la forma más fuerte de dependencia es cuando Y es una función de X, es decir, cuando existe una función $g : \mathbb{R} \to \mathbb{R}$ tal que $Y = g(X)$.

Una forma de medir cuán cerca está Y de ser una función de X es entonces minimizar la varianza $Var\,(Y - g(X))$ entre todas las funciones posibles g.

Naturalmente, si $Y = g(X)$ para alguna g, entonces el mínimo es cero. Por otro lado, si Y es independiente de X, entonces

$$Var\,(Y - g(X)) = Var\,(Y) + Var\,(g(X)) \geq Var\,(Y)\,,$$

por lo que el mínimo es la $Var\,(Y)$.

Consideremos entonces el problema de encontrar

$$v = \frac{\min_g \{Var\,(Y - g(X))\}}{Var\,(Y)}.$$

Hemos dividido entre la varianza de Y para que $v \in [0,1]$.[1]

El problema de calcular v lo veremos más adelante en toda su generalidad. Por ahora, empecemos por calcular el mínimo entre aquellas funciones que son *lineales*.

Nos restringiremos entonces a funciones $g(x) = ax$, y buscamos[2]

$$v_L = \frac{\min_{a\in\mathbb{R}} \{Var\,(Y - aX)\}}{Var\,(Y)}.$$

De la definición de varianza, tenemos que

$$\begin{aligned} Var\,(Y - aX) &= E\left((Y - aX - E\,(Y - aX))^2\right) \\ &= E\left((Y - E\,(Y) - a(X - E\,(X)))^2\right). \end{aligned}$$

[1] Esto lo probaremos en breve. Notar que cuanto más cerca esté v de cero, más cerca está Y de ser una función de X; y recíprocamente, cuanto más cerca esté v de 1, más lejos está Y de ser una función de X.

[2] La L indica que nos restringimos al caso lineal.

Desarrollando el cuadrado obtenemos que la varianza de $Y - aX$ es igual a

$$a^2 Var\,(X) - 2aE\,((Y - E\,(Y))(X - E\,(X))) + Var\,(Y)\,.$$

Este es un polinomio cuadrático en a, por lo que derivando e igualando a cero, vemos que el mínimo se da en

$$a_0 = \frac{E\,((Y - E\,(Y))(X - E\,(X)))}{Var\,(X)}.$$

El numerador de esta última expresión es tan importante que tiene nombre propio, se llama la *covarianza* de X e Y y se escribe $Cov\,(X,Y)$. Un cálculo sencillo, y similar a los que ya hemos hecho, muestra que podemos escribir la covarianza como

$$Cov\,(X,Y) = E\,(XY) - E\,(X)\,E\,(Y)\,.$$

Para aliviar un poco la notación, denotemos por

$$\sigma_X^2 = Var\,(X)\,,\ \sigma_Y^2 = Var\,(Y)\,,\ \sigma_{XY} = Cov\,(X,Y)\,.$$

Reemplazando el valor de a_0 en la ecuación para v_L vemos entonces que

$$v_L = 1 - \left(\frac{\sigma_{XY}}{\sigma_X \sigma_Y}\right)^2$$

El número que aparece elevado al cuadrado se conoce con el nombre de *coeficiente de correlación lineal* entre X e Y, y se denota por ρ_{XY}, o simplemente ρ si está claro con qué variables se está trabajando. Así que

$$v_L = 1 - \rho^2 \text{ con } \rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Resumimos en el siguiente cuadro las ecuaciones más importantes hasta ahora.

Covarianza y correlación

Sean X e Y dos variables aleatorias con varianza finita. La covarianza entre X e Y es por definición

$$\sigma_{XY} = Cov\,(X,Y) = E\,(XY) - E\,(X)\,E\,(Y)\,.$$

El coeficiente v_L está dado por $v_L = 1 - \rho^2$, en donde

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

es el coeficiente de correlación lineal.

Covarianza e independencia

Si X e Y son independientes, entonces $Cov\,(X,Y) = 0$. Eso es debido a que en este caso $E\,(XY) = E\,(X)\,E\,(Y)$. El recíproco no es cierto en general, como lo muestra el siguiente ejemplo.

Sea X una variable que vale 1 y -1 con probabilidad $1/2$. Sea Y una variable que vale 0 si $X = -1$, y que vale 1 y -1 con probabilidad $1/2$ si $X = 1$.

Notar que la distribución conjunta de X e Y está dada por la tabla de contingencia que se muestra en la Tabla 1.

		X		
		-1	1	p_Y
	-1	0	$1/4$	$1/4$
Y	0	$1/2$	0	$1/2$
	1	0	$1/4$	$1/4$
	p_X	$1/2$	$1/2$	1

Tabla 1: Distribución conjunta de X e Y.

De aquí se ve fácilmente que $E\,(X) = E\,(Y) = 0$. Más aún

$$E\,(XY) = 0 \cdot 1/2 + 1 \cdot 1/4 - 1 \cdot 1/4 = 0.$$

Así que $Cov\,(X,Y) = 0$. Pero X e Y no son independientes: esto se ve facilmente de la tabla, ya que la conjunta no es el producto de las marginales.

Covarianza entre variables Bernoulli

Sean X e Y dos variables Bernoulli. Calculemos la covarianza entre ellas. Notar que el producto XY también es una variable Bernoulli, con probabilidad de éxito igual a $P\,(X = 1, Y = 1)$. Entonces

$$Cov\,(X,Y) = P\,(X = 1, Y = 1) - P\,(X = 1)\,P\,(Y = 1)\,.$$

Notar que en este caso sí es cierto que covarianza cero implica independencia.

Covarianza entre Bernoulli

Sean X e Y dos variables Bernoulli. Entonces

$$X \text{ e } Y \text{ son independientes} \Leftrightarrow Cov\,(X,Y) = 0.$$

Más aún, la covarianza es positiva si

$$P\,(Y = 1 | X = 1) > P\,(Y = 1)$$

y negativa si

$$P\,(Y = 1 | X = 1) < P\,(Y = 1)\,.$$

Esto es, si la covarianza es positiva, entonces $X = 1$ aumenta las chances de que $Y = 1$, y recíprocamente si es negativa.

Correlación 1 o -1

Mencionamos una importante propiedad del coeficiente de correlación lineal. Si el coeficiente de correlación es igual a 1, o igual a -1, entonces el coeficiente $v_L = 0$. Esto quiere decir que existe a que minimiza la varianza $Var\,(Y - aX)$, y que este mínimo es igual a cero.

Pero $Var\,(Y - aX) = 0$ implica que $Y - aX$ es constante. Es decir, existe una constante b tal que $Y = aX + b$. El signo de a depende del signo de ρ_{XY}.

Correlación ± 1

- Si $\rho_{XY} = 1$ entonces $Y = aX + b$ con $a > 0$.
- Si $\rho_{XY} = -1$ entonces $Y = aX + b$ con $a < 0$.

Propiedades de la covarianza

Observemos primero una desigualdad notable que se deduce de lo que hemos hecho hasta ahora. Claramente, el coeficiente v_L es mayor o igual a cero por definición. Esto implica que $|\rho_{XY}| \leq 1$, que en términos de covarianza, se puede reescribir como

$$|Cov\,(X, Y)| \leq \sqrt{Var\,(X)}\sqrt{Var\,(Y)} \leq 1.$$

Esta es la famosa *Desigualdad de Cauchy-Schwarz*.

Esto sugiere que la covarianza tiene las propiedades de un producto interno. Claramente la covarianza es simétrica

$$Cov\,(X, Y) = Cov\,(Y, X)\,.$$

Además, si $Cov\,(X, X) = 0$ entonces $Var\,(X) = 0$, y esto implica que X es una constante. Como las variables aleatorias que son constantes no son muy interesantes, la trataremos como variables triviales. Así

$$Cov\,(X, X) = 0 \Leftrightarrow X \text{ es trivial.}$$

Y por último, la covarianza es bi-lineal. Es decir, si X, Y y Z son tres variables discretas y a es una constante, entonces

$$Cov\,(aX + Y, Z) = a \cdot Cov\,(X, Z) + Cov\,(Y, Z)\,.$$

La prueba es directa: notar primero que

$$E\,(aX + Y) = aE\,(X) + E\,(Y)\,.$$

Luego, desarrollando el producto

$$\begin{aligned} &\big(aX + Y - (aE\,(X) + E\,(Y))\big)\big(Z - E\,(Z)\big) \\ &= a(X - E\,(X))(Z - E\,(Z)\big) + (Y - E\,(Y))(Z - E\,(Z)\big) \end{aligned}$$

Tomando esperanza se prueba la propiedad.

Pitágoras y la varianza

El hecho de que $Var\,(X) = Cov\,(X, X)$, y que la covarianza sea un producto interno, implican que la raíz de la varianza

$$\sigma_X = \sqrt{Var\,(X)}$$

es una *norma* que en cierto modo mide la longitud de X. Si pensamos a las variables como vectores en un espacio vectorial (de dimensión infinita), entonces su norma está dada por σ_X.

Notar que como toda norma, σ_X verifica la desigualdad triangular:

$$\sigma_{X+Y} \leq \sigma_X + \sigma_Y,$$

que se puede probar fácilmente a partir de la desigualdad de Cauchy-Schwarz.

El coeficiente de correlación ρ_{XY} se puede identificar entonces con el coseno del ángulo entre dos variables:

$$\rho_{XY} = \cos(\alpha)$$

en donde α es el ángulo entre X e Y.

De esta forma, la covarianza nula no es otra cosa que decir que las variables X e Y son *ortogonales*. Esta terminología es muy usada. En particular, si X e Y son independientes, entonces son ortogonales.

Podemos interpretar la varianza de la suma de variables ortogonales como un caso particular del Teorema de Pitágoras.

Teorema de Pitágoras

Sean X e Y dos variables ortogonales, esto es con $\sigma_{XY} = 0$. Entonces

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y.$$

Es decir, la hipotenusa al cuadrado es igual a la suma de los cuadrados de los catetos.

En general se tiene que

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\sigma_{XY}.$$

Recordando la analogía entre la raíz de la varianza y la longitud de vectores esta fórmula es muy fácil de retener en la memoria.

Interpretación de la covarianza

Veremos un caso muy sencillo en el cual podremos dar una interpretación visual de la covarianza entre dos variables.

Sean X e Y dos variables discretas con distribución uniforme en $\{x_1, \ldots, x_n\}$ e $\{y_1, \ldots, y_n\}$ respectivamente. Claramente esta información no es suficiente para determinar su covarianza, pues no sabemos cuál es la distribución conjunta. Veremos tres casos distintos, pero antes haremos unos cálculos generales.

Los valores posibles del par (X, Y) son obviamente los pares (x_i, y_j). Para cada par de pares (x_i, y_j) y (x_k, y_l) podemos construir un rectángulo que los tenga como vértices. Dos situaciones se pueden dar, como se muestra en la Figura 4.

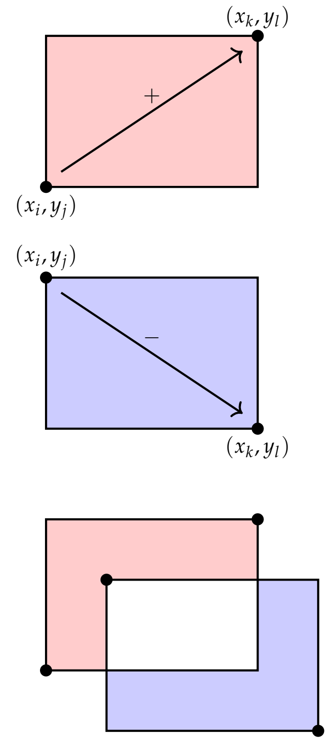

Figura 4: Las dos posibilidades para los rectángulos. En el primer caso lo pintamos de rojo para indicar la asociación positiva, y en el segundo de azul para indicar la asociación negativa. Cuando dos rectángulos opuestos y de la misma intensidad se solapan, la intersección se representa en blanco como color neutro.

Lo que haremos es construir un rectángulo para cada par de pares posible. ¿Qué representa cada rectángulo?

La cantidad de color de un rectángulo es el área del mismo. De modo que $(x_i - x_k) \times (y_j - y_l)$ representa la cantidad de color, siendo rojo cuando es positivo y azul cuando es negativo. Sin embargo, debemos tener en cuenta la probabilidad con que (X, Y) es igual a cada uno de éstos. Así que multiplicamos

$$(x_i - x_k)(y_j - y_l)p_{ij}p_{kl}$$

en donde hemos denotado $p_{ij} = P\left(X = x_i, Y = y_j\right)$. Esto equivale a hacer más clarito o más fuerte el color del rectángulo según sea grande o chico el producto $p_{ij}p_{kl}$.

Cuando sumamos en todos los rectángulos, obtenemos la cantidad total de rojo menos la de azul, teniendo en cuenta que cuando un rectángulo rojo se solapa con uno azul, la intersección queda con el color correspondiente a la resta de las intensidades. Por ejemplo, si un rectángulo es azul y el otro rojo, ambos con la misma intensidad, la intersección se anula y queda en blanco.

Pero observar que

$$\begin{aligned}
&\sum_{\text{rectángulos}} (x_i - x_k)(y_j - y_l)p_{ij}p_{kl} \\
&= \frac{1}{2}\sum_{i,j}\sum_{k,l}(x_i - x_k)(y_j - y_l)p_{ij}p_{kl} \\
&= \frac{1}{2}\sum_{i,j}\sum_{k,l}(x_iy_j - x_iy_l - x_ky_j + x_ky_l)p_{ij}p_{kl}
\end{aligned}$$

El primer y último término suman lo mismo, y lo mismo vale para el segundo y el tercero. Entonces, la suma es igual a

$$\sum_{i,j}\sum_{k,l} x_iy_jp_{ij}p_{kl} - \sum_{i,j}\sum_{k,l} x_iy_lp_{ij}p_{kl}.$$

Notar que la primer suma es igual a

$$\sum_{i,j} x_iy_jp_{ij} = E\left(XY\right),$$

ya que $\sum_{k,l} p_{kl} = 1$. Del mismo modo se puede ver que la segunda suma es igual a $E(X)\,E(Y)$. Es decir, hemos probado que

$$Cov\,(X,Y) = \text{(cantidad de rojo)} - \text{(cantidad de azul)}.$$

Veamos unos ejemplos sencillos. En la Figura 5 se muestra la distribución conjunta de X e Y en tres casos distintos. En cada caso son $n = 10$ puntos en total, todos igualmente probables para que la elección de los colores sea simple. Es decir, todos los rectángulos tienen la misma intensidad de color. La única diferencia es la disposición de los puntos.

En la gráfica de arriba se muestra una distribución conjunta con covarianza positiva. Esto se ve fácilmente por la cantidad de rojo sobre el azul. En la del centro, se muestra una distribución con covarianza casi nula. Aquí también se puede apreciar casi la misma cantidad de rojo que de azul. Y por último, en la de abajo, se muestra una con covarianza negativa.

Es decir, el signo de la covarianza nos indica la tendencia de X a crecer o a decrecer cuando Y crece.

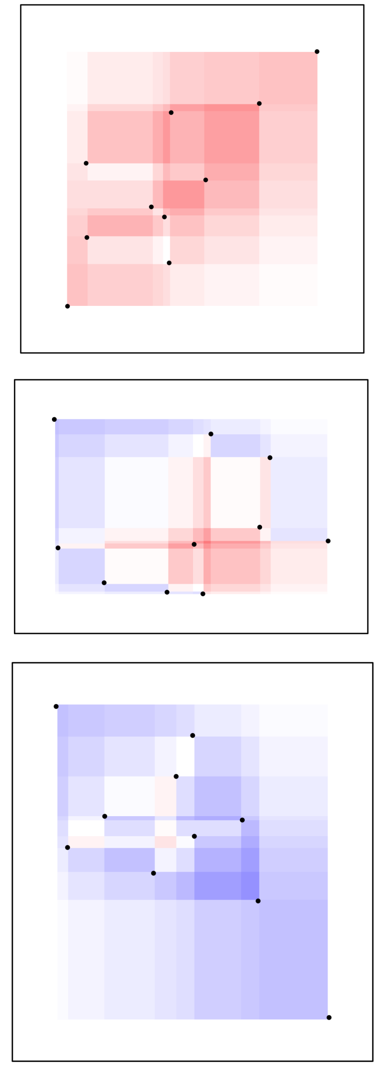

Figura 5: Arriba: covarianza positiva. Centro: covarianza nula. Abajo: covarianza negativa.