Lecture 11: Correlation and independence

Definition 4.5.1.

The covariance of random variables X and Y is defined as

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

provided that the expectation exists.

Definition 4.5.2.

The correlation (coefficient) of random variables X and Y is defined as

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

- By Cauchy-Schwartz's inequality, $[\mathrm{Cov}(X,Y)]^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y)$ and, hence, $|\rho_{X,Y}| \leq 1$.
- If large values of X tend to be observed with large (or small) values of Y and small values of X with small (or large) values of Y, then $\mathrm{Cov}(X,Y) > 0$ (or < 0).

- If $\text{Cov}(X, Y) = 0$, then we say that X and Y are uncorrelated.
- The correlation is a standardized value of the covariance.

Theorem 4.5.6.

If X and Y are random variables and a and b are constants, then

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

Theorem 4.5.6 with $a = b = 1$ implies that, if X and Y are positively correlated, then the variation in $X + Y$ is greater than the sum of the variations in X and Y; but if they are negatively correlated, then the variation in $X + Y$ is less than the sum of the variations.

This result is useful in statistical applications.

Multivariate expectation

The expectation of a random vector $X = (X_1, ..., X_n)$ is defined as $E(X) = (E(X_1), ..., E(X_n))$, provided that $E(X_i)$ exists for any i.

When M is a matrix whose (i, j)th element if a random variable X_{ij}, $E(M)$ is defined as the matrix whose (i, j)th element is $E(X_{ij})$, provided that $E(X_{ij})$ exists for any (i, j).

Variance-covariance matrix

The concept of mean and variance can be extended to random vectors: for an n-dimensional random vector $X = (X_1, ..., X_n)$, its mean is $E(X)$ and its variance-covariance matrix is

$$\mathrm{Var}(X) = E\{[X - E(X)][X - E(X)]'\} = E(XX') - E(X)E(X')$$

which is an $n \times n$ symmetric matrix whose ith diagonal element is the variance $\mathrm{Var}(X_i)$ and (i,j)th off-diagonal element is the covariance $\mathrm{Cov}(X_i, X_j)$.

$\mathrm{Var}(X)$ is nonnegative definite.

If the rank of $\mathrm{Var}(X)$ is $r < n$, then, with probability equal to 1, X is in a subspace of $\mathscr{R}^n$ with dimension r.

If A is a constant $m \times n$ matrix, then

$$E(AX) = AE(X) \qquad \text{and} \qquad \mathrm{Var}(AX) = A\mathrm{Var}(X)A'$$

Example 4.5.4.

The joint pdf of (X, Y) is

$$f(x,y) = \begin{cases} 1 & 0 < x < 1,\ x < y < x+1 \\ 0 & \text{otherwise} \end{cases}$$

The marginal of X is *uniform*$(0,1)$, since for $0 < x < 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_x^{x+1} dy = x + 1 - x = 1$$

For $0 < y < 1$,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^y dx = y - 0 = y$$

and for $1 \le y < 2$,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_{y-1}^1 dx = 1 - (y-1) = 2 - y$$

i.e.,

$$f_Y(y) = \begin{cases} 2 - y & 1 \le y < 2 \\ y & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Thus, $E(X) = 1/2$ and $\mathrm{Var}(X) = 1/12$, and

$$E(Y) = \int_0^1 y^2 dy + \int_1^2 y(2-y)dy = \frac{1}{3} + 4 - 1 - \frac{1}{3}(8-1) = 1$$

$$\mathrm{Var}(Y) = E(Y^2) - 1 = \int_0^1 y^3 dy + \int_1^2 y^2(2-y)dy - 1 = \frac{1}{4} + \frac{14}{3} - \frac{15}{4} - 1 = \frac{1}{6}$$

Also,

$$E(XY) = \int_0^1 \int_x^{x+1} xydydx = \int_0^1 \left(x^2 + \frac{x}{2}\right) dx = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$$

Hence,

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{7}{12} - \frac{1}{2} = \frac{1}{12}$$

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{1/12}{\sqrt{1/12 \times 1/6}} \frac{1}{\sqrt{2}}$$

Theorem 4.5.7.

For random variables X and Y, $|\rho_{X,Y}| = 1$ iff $P(Y = aX + b) = 1$ for constants a and b, where $a > 0$ if $\rho_{X,Y} = 1$ and $a < 0$ if $\rho_{X,Y} = -1$.

- The proof of this theorem is actually discussed when we study Cauchy-Schwartz's inequality (when the equality holds).
- If there is a line, $y = ax + b$ with $a \neq 0$, such that the values of the 2-dimensional random vector (X,Y) have a high probability of being near this line, then the correlation between X and Y will be near 1 or -1.

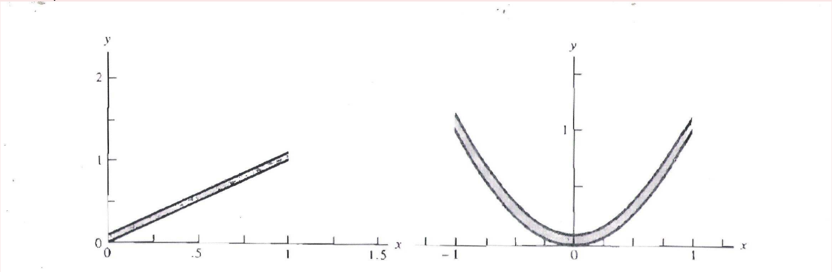
On the other hand, X and Y may be highly related but have no linear relationship, and the correlation could be nearly 0.

Example 4.5.8, 4.5.9.

Consider (X, Y) having pdf

$$f(x,y) = \begin{cases} 10 & 0 < x < 1,\ x < y < x + 0.1 \\ 0 & \text{otherwise} \end{cases}$$

This is the same as the pdf in Example 4.5.4 except that $x < y < x + 0.1$ (instead of $x < y < x + 1$) in the region $\{f(x,y) > 0\}$. The same calculation as in Example 4.5.4 shows that $\rho_{X,Y} = \sqrt{100/101}$, which is almost 1 (the first figure below).



Consider (X, Y) having a different pdf (the 2nd figure)

$$f(x,y) = \begin{cases} 5 & -1 < x < 1,\ x^2 < y < x^2 + 0.1 \\ 0 & \text{otherwise} \end{cases}$$

By symmetry, $E(X) = 0$ and $E(XY) = 0$; hence, $\text{Cov}(X, Y) = \rho_{X,Y} = 0$.
There is actually a strong relationship between X and Y.
But this relationship is not linear.
The correlation does not measure any nonlinear relationship.

Whether random variables are related or not related at all is described by the concept of independence.

First definition of independence of random variables/vectors

Let $X_1, ..., X_k$ be random variables, $F(x_1, ..., x_k)$ be their joint cdf., and $F_{X_i}(x_i)$ be the marginal cdf of X_i, $i = 1, ..., k$.
$X_1, ..., X_k$ are (statistically) independent iff

$$F(x_1, ..., x_k) = F_{X_1}(x_1) \cdots F_{X_k}(x_k) \qquad (x_1, ..., x_k) \in \mathscr{R}^k$$

We define the independence of random vectors $X_1, ..., X_k$ (they may have different dimensions) in the same way, except that F is the joint cdf of $(X_1, ..., X_k)$ and F_{X_i} is the (joint) cdf of X_i, $i = 1, ..., k$.

Clearly, this definition is equivalent to the following definition.

Second definition of independence of random variables/vectors

Random vectors $X_1,...,X_k$ are independent iff for any permutation $i_1,...,i_k$ of $1,...,k$, and any $r = 1,...,k-1$, the conditional distribution of $X_{i_1},...,X_{i_r}$ given $X_{i_{r+1}},...,X_{i_k}$ is the same as the distribution of $X_{i_1},...,X_{i_r}$.

- The joint cdf is determined by the n marginal cdf's if $X_1,...,X_n$ are independent; otherwise, the joint cdf depends on marginal cdf's and conditional distributions.
- If $X_1,...,X_k$ are independent, then $X_{i_1},...,X_{i_r}$ are independent for any subest $\{i_1,...,i_r\} \subset \{1,...,k\}$.
- When $k = 2$ and both X_1 and X_2 are discrete random variables, they are independent iff all $p_{ij} = p_{i\cdot}p_{\cdot j}$ in the probability table.
- If X_i has pmf (or pdf) f_{X_i}, $i = 1,...,k$, then $X_1,...,X_k$ are independent iff the joint pmf (or pdf) satisfies

$$f(x_1,...,x_k) = f_{X_1}(x_1)\cdots f_{X_k}(x_k) \qquad (x_1,...,x_k) \in \mathscr{R}^k$$

  In fact, sometimes the following lemma can be applied.

Lemma 4.2.7.

Let X and Y be random variables having joint pmf or pdf $f(x,y)$.
X and Y are independent iff there exist functions $g(x)$ and $h(y)$ with

$$f(x,y) = g(x)h(y) \qquad x \in \mathscr{R},\ y \in \mathscr{R}$$

A similar result holds for any fixed number of random variables/vectors.

In the definition of independence of random vectors, the components of each random vector may be dependent or independent.
For example, if X, Y, and Z are 3 random variables, we may have that (X,Y) and Z are independent but X and Y are not independent.

Example

Suppose that (X,Y,Z) has pdf

$$f(x,y,z) = \begin{cases} e^{-y}e^{-z} & 0 < x < y, z > 0 \\ 0 & \text{otherwise} \end{cases}$$

It can be shown that $f(x,y,z) = g(x,y)h(z)$ with

$$g(x,y) = \begin{cases} e^{-y} & 0 < x < y \\ 0 & \text{otherwise} \end{cases} \qquad h(z) = \begin{cases} e^{-z} & z > 0 \\ 0 & z \le 0 \end{cases}$$

Hence, Z and $U = (X, Y)$ are independent.
In Example 4.2.4, we showed that $Y|X = x \sim exponential(x, 1)$,
The marginal pdf of Y is i.e., $gamma(2, 1)$, because

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,z)dxdz = \int_0^y e^{-y}dx = ye^{-y} \qquad y > 0$$

By the 2nd definition of independence, X and Y are not independent.
In fact, when $Y|X = x$ has a pdf varying with x, we have already known that X and Y are not independent.
Consequently, X, Y, and Z are not independent.

Note that “X, Y, and Z are independent" may be different from “(X, Y) and Z are independent".
Hence, when random vectors are involved, it is important to specify which variables/vectors are independent.

Third definition of independence of random variables/vectors

Random vectors $X_1, ..., X_k$ are independent iff events $\{X_1 \in B_1\}$, ...,$\{X_k \in B_k\}$ are independent for any Borel sets $B_1, ..., B_k$, where the dimension of B_i is the same as the dimension of X_i.

It can be shown that this definition is equivalent to the first definition. In fact, it is easy to show this definition implies the first definition, by considering $B_i = (-\infty, x_i]$.
A rigorous proof is out of the scope of this course.

Theorem 4.6.12.

If $X_1, ..., X_k$ are independent random vectors and $g_1, ..., g_k$ are functions, then
(i) $g_1(X_1), ..., g_k(X_k)$ are independent, and
(ii) $E[g_1(X_1)\cdots g_k(X_k)] = E[g_1(X_1)]\cdots E[g_k(X_k)]$.

The following is an important corollary.

Theorem 4.5.5.

If $X_1, ..., X_k$ are independent random variables, then X_i and X_j are uncorrelated for every pair (i,j).

The converse is not true, i.e., there are uncorrelated X and Y but they are not independent.
An example is X and Y in Example 4.5.9 with quadratic relationship.

Example

Let (X, Y) be a random vector on $\mathscr{R}^2$ with pdf

$$f(x,y) = \begin{cases} \pi^{-1} & x^2+y^2 \le 1 \\ 0 & x^2+y^2 > 1. \end{cases}$$

Let $D = \{(x,y) : x^2+y^2 \le 1\}$.
By changing variables, we obtain that

$$\begin{aligned} E(XY) &= \pi^{-1} \int_D dxdy \\ &= \pi^{-1} \left(\int_{D,xy>0} dxdy + \int_{D,xy<0} dxdy \right) \\ &= \pi^{-1} \left(\int_{D,xy>0} dxdy - \int_{D,xy>0} dxdy \right) \\ &= 0 \end{aligned}$$

Similar, $EX = EY = 0$ and, hence, $\mathrm{Cov}(X,Y) = 0$.

How do we show X and Y are not independent?

It is unnecessary to derive the marginal distributions for showing that X and Y are not independent.

By the 3rd definition of independence, we just need to find two Borel sets A and B such that

$$P(X \in A, Y \in B) \neq P(X \in A)P(Y \in B)$$

A direct calculation shows that

$$P(0 < X < 1/\sqrt{2}, 0 < Y < 1/\sqrt{2}) = \frac{1}{2\pi}$$

and

$$P(0 < X < 1/\sqrt{2}) = P(0 < Y < 1/\sqrt{2}) = \frac{1}{4} + \frac{1}{2\pi}.$$

Thus, X and Y are not independent because

$$\frac{1}{2\pi} \neq \left(\frac{1}{4} + \frac{1}{2\pi}\right)^2.$$

- This example indicates that if the distribution of (X, Y) is symmetric in some way, then $\text{Cov}(X, Y) = 0$ and X and Y don't have any linear relationship.
- Uncorrelated X and Y means that there is no **linear** relationship between X and Y, but X and Y may have a strong nonlinear relationship.

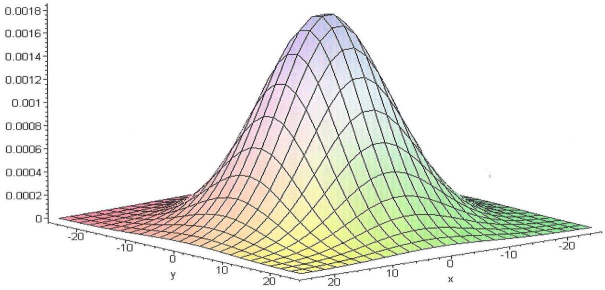
Example 4.5.10 (bivariate normal distribution).

Consider $(X, Y) \sim$ the bivariate normal distribution pdf on $\mathscr{R}^2$:

$$f(x,y) = \frac{\exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2(1-\rho^2)} + \frac{\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)} - \frac{(y-\mu_2)^2}{2\sigma_2^2(1-\rho^2)}\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \qquad (x,y) \in \mathscr{R}^2$$

where $\mu_1 \in \mathscr{R}$, $\mu_2 \in \mathscr{R}$, $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 < \rho < 1$ are constants.

This 2-dimensional pdf has a very nice shape, as the following figure shows.



The marginal pdf of X is

$$f_X(x) = \frac{\exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2(1-\rho^2)}\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(\frac{\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)} - \frac{(y-\mu_2)^2}{2\sigma_2^2(1-\rho^2)}\right) dy$$

$$= \frac{\exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{[(y-\mu_2) - \frac{\rho\sigma_2}{\sigma_1}(x-\mu_1)]^2}{2\sigma_2^2(1-\rho^2)}\right) dy$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right)$$

which is the pdf of $N(\mu_1, \sigma_1^2)$ (as expected).

Similarly, the marginal pdf of Y is that of $N(\mu_2, \sigma_2^2)$.

We now calculate the correlation coefficient

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = E\left(\frac{X-\mu_1}{\sigma_1}\right)\left(\frac{Y-\mu_2}{\sigma_2}\right)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) f(x,y)\,dxdy$$

Letting $s = \left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)$ and $t = \frac{x-\mu_1}{\sigma_1}$, we obtain

$$\begin{aligned}
\rho_{X,Y} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} sf\left(\sigma_1 t + \mu_1, \sigma_2 \frac{s}{t} + \mu_2\right)\frac{\sigma_1\sigma_2}{|t|}dsdt \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{s\exp\left(-\frac{t^2-2\rho s+s^2/t^2}{2(1-\rho^2)}\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\frac{\sigma_1\sigma_2}{|t|}dsdt \\
&= \int_{-\infty}^{\infty} \frac{\exp(-t^2/2)}{2\pi\sqrt{(1-\rho^2)t^2}}\left[\int_{-\infty}^{\infty} s\exp\left(-\frac{(s-\rho t^2)^2}{2(1-\rho^2)t^2}\right)ds\right]dt \\
&= \int_{-\infty}^{\infty} \frac{\exp(-t^2/2)}{\sqrt{2\pi}}\rho t^2 dt = \frac{\rho}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t^2\exp(-t^2/2)dt \\
&= \rho
\end{aligned}$$

Thus, all 5 parameters in $f(x,y)$ have meanings:
μ_1 and μ_2 are two marginal means.
σ_1^2 and σ_2^2 are two marginal variances.
ρ is the correlation coefficient for the two marginal variables.

The conditional pdf of $Y|X$ can be derived as follows.

$$\frac{f(x,y)}{f_X(x)} = \frac{\exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2(1-\rho^2)} + \frac{\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)} - \frac{(y-\mu_2)^2}{2\sigma_2^2(1-\rho^2)}\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \Bigg/ \frac{\exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right)}{\sqrt{2\pi}\sigma_1}$$

which is equal to

$$\frac{\exp\left(-\frac{\rho^2(x-\mu_1)^2}{2\sigma_1^2(1-\rho^2)} + \frac{\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)} - \frac{(y-\mu_2)^2}{2\sigma_2^2(1-\rho^2)}\right)}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} = \frac{\exp\left(-\frac{[y-\mu_2-\frac{\rho\sigma_2}{\sigma_1}(x-\mu_1)]^2}{2\sigma_2^2(1-\rho^2)}\right)}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}$$

Hence,

$$Y|X \sim N\left(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(X-\mu_1), \sigma_2^2(1-\rho^2)\right)$$

From the second definition of independence, X and Y are independent if they are uncorrelated ($\rho = 0$).

On the other hand, if $\rho \neq 0$, then X and Y are correlated and they cannot be independent (Theorem 4.5.5).

This means that for bivariate normal (X, Y), independence of X and Y is equivalent to $\rho = 0$: either X and Y have a linear relationship or they don't have any relationship at all (they cannot have a nonlinear relationship).