

Independence: For random variables X and Y , the intuitive idea behind " Y is independent of X " is that the distribution of Y shouldn't depend on what X is. This can be expressed in terms of the conditional pdf's to say " $f(y|x)$ doesn't depend on x ."

Caution: " Y is not independent of X " means simply that the *distribution* of X may vary as Y varies. It doesn't mean that X depends on Y .

If Y is independent of X , then:

1. $\mu_x = E(Y|X = x)$ does not depend on x .

(*Question:* Is the converse true? That is, if $E(Y|X = x)$ does not depend on x , can we conclude that Y is independent of X ?)

2. Let $h(y)$ be the common pdf of the conditional distributions $Y|X$. Then for every x ,

$$h(y) = f(y|x) = \frac{f(x,y)}{f_X(x)}, \text{ where } f(x,y) \text{ is the joint pdf of } X \text{ and } Y. \text{ Therefore}$$

$$f(x,y) = h(y) f_X(x)$$

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx \\ &= \int_{-\infty}^{\infty} h(y) f_X(x) dx \\ &= h(y) \int_{-\infty}^{\infty} f_X(x) dx = h(y) = f(y|x) \end{aligned}$$

In other words: *If Y is independent of X , then the conditional distributions of Y given X are the same as the marginal distribution of Y .*

3. Now we have

$$f_Y(y) = f(y|x) = \frac{f(x,y)}{f_X(x)},$$

so

$$f_Y(y)f_X(x) = f(x,y).$$

In other words: *If Y is independent of X , then the joint distribution of X and Y is the product of the marginal distributions of X and Y .*

Exercise: The converse of this last statement is true. That is: If the joint distribution of X and Y is the product of the marginal distributions of X and Y , then Y is independent of X .

Note that the condition $f_Y(y)f_X(x) = f(x,y)$ is symmetric in X and Y . Thus (3) and its converse imply that : Y is independent of X if and only if X is independent of Y . So it makes sense to say "X and Y are independent."

Putting this all together, have: The following conditions are all equivalent:

- i. X and Y are independent.
- ii. $F_{X,Y}(x,y) = f_Y(y)f_X(x)$
- iii. The conditional distribution of $Y|X = x$ is independent of x
- iv. The conditional distribution of $X|Y = y$ is independent of y .
- v. $f(y|x) = f_Y(y)$ for all y .
- vi. $f(x|y) = f_X(x)$ for all x .

Additional property of independent random variables: If X and Y are independent, then $E(XY) = E(X)E(Y)$. (The proof of this fact will be assigned as homework for October 8.)

Covariance: The *covariance* of two random variables X and Y is defined as

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

Comments:

- The capital C in Cov is consistent with the notation used in this class of capitalizing items that relate to the population, and using lower case (or a "hat") for items referring to a sample. There is a related notion of covariance for a sample, discussed briefly later. Consistent with general terminology, Cov is a *parameter* since it refers to the population, and the sample covariance (cov or Cov-hat) is a *statistic* since it is calculated from the sample.
- Compare and contrast with the definition of $\text{Var}(X)$.
- If X and Y both tend to be on the same side of their respective means (i.e., both greater than or both less than their means), then $[X - E(X)][Y - E(Y)]$ tends to be positive, so $\text{Cov}(X,Y)$ is positive. Similarly, if X and Y tend to be on opposite sides of their respective means, then $\text{Cov}(X,Y)$ is negative. If there is no trend of either sort, then $\text{Cov}(X,Y)$ should be zero. Thus covariance roughly measures the extent of a "positive" or "negative" trend in the joint distribution of X and Y .
- What are the units of $\text{Cov}(X,Y)$?

Properties:

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

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

- Why?
- In words ...
- Compare with the alternate formula for $\text{Var}(X)$.

3. Consequence: If X and Y are independent, then:

Note: The converse of this statement is false. This will be a problem on a future homework set.

4. $\text{Cov}(cX, Y) =$ and $\text{Cov}(X, cY) =$

5. $\text{Cov}(a + X, Y) =$ and $\text{Cov}(X, a + Y) =$

6. $\text{Cov}(X + Y, Z) =$ and $\text{Cov}(X, Y + Z) =$

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

- Why?
- Consequence: If X and Y are independent, then _____
- *Note:* The converse of this last statement is false.

Bounds on Covariance

Let σ_X denote the *population standard deviation* $\sqrt{\text{Var}(X)}$ of X . (Do not confuse with the sample standard deviation s !). Define the population standard deviation σ_Y of Y similarly.

Consider the new random variable $\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$. Since Variance is always ≥ 0 ,

$$\begin{aligned}
 (*) \quad 0 &\leq \text{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) \\
 &= \text{Var}\left(\frac{X}{\sigma_X}\right) + \text{Var}\left(\frac{Y}{\sigma_Y}\right) + 2\text{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) \\
 &= \frac{1}{\sigma_X^2} \text{Var}(X) + \frac{1}{\sigma_Y^2} \text{Var}(Y) + \frac{2}{\sigma_X \sigma_Y} \text{Cov}(X, Y)
 \end{aligned}$$

$$= 2\left[1 + \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y}\right].$$

Therefore

$$(**) \quad \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} \geq -1 \quad (\text{or: } \text{Cov}(X, Y) \geq -\sigma_X\sigma_Y).$$

Looking at $\text{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right)$ similarly shows (details left to the student):

$$(***) \quad \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} \leq 1, \quad (\text{or: } \text{Cov}(X, Y) \leq \sigma_X\sigma_Y).$$

Combining (**) and (***) gives:

$$\left| \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} \right| \leq 1, \quad (\text{or: } |\text{Cov}(X, Y)| \leq \sigma_X\sigma_Y)$$

Moreover, the only way we can have equality in inequality (**) is to have equality in (*) -- i.e, when

$$\text{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = 0$$

This can happen if and only if the random variable $\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$ is constant -- say,

$$\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} = c.$$

(Note that c must be the mean of $\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$, which is $\frac{\mu_X}{\sigma_X} + \frac{\mu_Y}{\sigma_Y}$)

This in turn is equivalent to

$$Y = \sigma_Y \left(-\frac{X}{\sigma_X} + c \right)$$

or

$$Y = -\frac{\sigma_Y}{\sigma_X} X + \sigma_Y c,$$

which says: The pairs (X,Y) lie on a line with negative slope. (The converse is also true -- details left to the student. Also note that the slope of the line is $-\frac{\sigma_Y}{\sigma_X}$ and the y-intercept

is $\frac{\sigma_Y}{\sigma_X}\mu_X + \mu_Y$.)

Similarly, $\frac{Cov(X,Y)}{\sigma_X\sigma_Y} = +1$ exactly when the pairs (X,Y) lie on a line with *positive* slope.

Correlation: The *correlation coefficient* of the random variables X and Y is

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}.$$

Note:

- $\rho_{X,Y}$ is often called ρ for short.
- ρ is a parameter (since it refers to the population). There is also a sample correlation coefficient (usually called r) that is a statistic (calculated from the sample).

Stated in terms of ρ , the discussion above says:

- Negative ρ indicates a tendency for the variables X and Y to co-vary in a negative way.
- Positive ρ indicates a tendency for the variables X and Y to co-vary in a positive way.
- $-1 \leq \rho \leq 1$
- $\rho = -1$ if and only if all pairs (X,Y) lie on a straight line with negative slope.
- $\rho = 1$ if and only if all pairs (X,Y) lie on a straight line with positive slope.
- ρ is unitless.
- ρ is the Covariance of the standardized random variables $\frac{X - \mu_X}{\sigma_X}$ and $\frac{Y - \mu_Y}{\sigma_Y}$.

(Details left to the student.)

Also, from the definition, we see that $\rho = 0$ if and only if $Cov(X,Y) = 0$.

Uncorrelated variables: We say that two random variables are *uncorrelated* if $\rho_{X,Y} = 0$ (or equivalently, if $Cov(X,Y) = 0$).

Examples:

- If X and Y are independent, then they are uncorrelated. (Why?)
- Suppose that the random variable X is uniform on the interval $[-1, 1]$. Let $Y = X^2$. Then X and Y are uncorrelated, but not independent. (To see that X and Y are not independent, note that $E(Y|X)$ is not constant. Details of why X and Y are uncorrelated will be on the next homework assignment.)

In general, ρ is a measure of the degree of a *nonconstant linear* relationship between X and Y . Example 2 above shows that two variables can have a strong nonlinear relationship and still be uncorrelated.

Sample variance, covariance, and correlation

If we have a sample of data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ from the joint distribution of X and Y , we can define the statistics

$$\text{sample covariance } \text{cov}(x, y) \text{ (or } \widehat{\text{Cov}}(x, y)) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

and

$$\text{sample correlation coefficient } r \text{ (or } \hat{\rho}) = \frac{\text{cov}(x, y)}{sd(x)sd(y)}.$$

These are estimators of the corresponding population parameters. We can establish properties of the sample covariance and correlation coefficient analogous to those of the population covariance and correlation coefficient.