INDEPENDENCE, COVARIANCE AND CORRELATION

Independence:
Intuitive idea of "Y is independent of X": The distribution of Y doesn't depend on the value of X.

In terms of the conditional pdf's:
"f(y|x) doesn't depend on x."

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

If Y is independent of X, then:

1. \( \mu_Y = E(Y|X = x) \) 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. (Still assuming Y is independent of X) Let \( h(y) \) be the common pdf of the conditional distributions Y|X. Then for every x,

\[
h(y) = f(y|x) = \frac{f(x,y)}{f_X(x)},
\]

where \( f(x,y) \) is the joint pdf of X and Y.

Therefore

\[
f(x,y) = h(y) f_X(x)
\]

\[
f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx = \int_{-\infty}^{\infty} h(y) f_X(x) dx = h(y) \int_{-\infty}^{\infty} f_X(x) dx = h(y) f_Y(y)
\]

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 (still assuming $Y$ is independent of $X$) 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: \textit{If $Y$ is independent of $X$, then the joint distribution of $X$ and $Y$ is the product of the marginal distributions of $X$ and $Y$.}

\textit{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 $X$.

\textit{Observe:} 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."

\textbf{Summary:} The following conditions are all equivalent:

1. $X$ and $Y$ are independent.

2. $f_{X,Y}(x,y) = f_Y(y)f_X(x)$

3. The conditional distribution of $Y|X = x$ is independent of $x$.

4. The conditional distribution of $X|Y = y$ is independent of $y$.

5. $f(y|x) = f_Y(y)$ for all $y$.

6. $f(x|y) = f_X(x)$ for all $x$.

\textit{Additional property of independent random variables:} If $X$ and $Y$ are independent, then $E(XY) = E(X)E(Y)$. \textit{(Proof might be homework.)}

\textbf{Covariance:} For random variables $X$ and $Y$,

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

\textbf{Comments:}
• Cov (capital C) \(\leftrightarrow\) population
cov (or Cov-hat) \(\leftrightarrow\) sample.
Cov is a \textit{parameter} \(\leftrightarrow\) population
cov is a \textit{statistic} \(\leftrightarrow\) calculated from the sample.

1. \(\text{Cov}(X, X) =\)
2. \(\text{Cov}(Y, X) =\)
3. \(\text{Cov}(X, Y) = \mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]) =\)

• Compare and contrast with 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 \([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\) 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 \(Y\).

In words …

• Units of \(\text{Cov}(X,Y)\)?

• Compare with the alternate formula for \(\text{Var}(X)\).

4. Consequence: If \(X\) and \(Y\) are independent, then:

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

\textit{Note}: Converse false. (Future homework.)

5. \(\text{Cov}(cX, Y) =\)
Bounds on Covariance

\[ \sigma_X = \text{population standard deviation} \sqrt{\text{Var}(X)} \text{ of } X. \]

(Do not confuse with sample standard deviation \( s \) or s.d. or \( \hat{\sigma} \))

\( \sigma_Y \) defined similarly.

Consider the new random variable \( \frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} \).

Since Variance is always \( \geq 0 \),

\[ (*) \quad 0 \leq \text{Var} \left( \frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} \right) \]

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

\[ = \frac{1}{\sigma_X^2} \text{Var}(X) + \frac{1}{\sigma_Y^2} \text{Var}(Y) + \frac{2}{\sigma_X \sigma_Y} \text{Cov}(X,Y) \]

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

Therefore:
Equality in (**) $\iff$ equality in (*) -- i.e.,

$$\text{Var}\left(\frac{X}{\sigma_x} + \frac{Y}{\sigma_y}\right) = 0 \quad \iff \quad \frac{X}{\sigma_x} + \frac{Y}{\sigma_y} \text{ is constant}$$

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

$$Y = -\frac{\sigma_y}{\sigma_x} X + \sigma_y c$$

This says: The pairs $(X,Y)$ lie on a line with negative slope (namely, $-\sigma_y/\sigma_x$)

(Converse is also true -- details left to the student.)
Note: the line has slope \(-\frac{\sigma_y}{\sigma_x}\) and y-intercept \(\frac{\sigma_y}{\sigma_x} \mu_x + \mu_y\).

\[
\text{Cov}(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y} = +1 \text{ exactly when the pairs (X,Y) lie on a line with positive slope.}
\]

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

\[
\rho_{xy} = \frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y}.
\]

**Note:**
- \(\rho\) for short.
- \(\rho\) is a parameter (refers to the population).
- Do not confuse with the sample correlation coefficient (usually called \(r\): a statistic (calculated from the sample).

**Properties of \(\rho\):**
- Negative \(\rho\) indicates a tendency for the variables X and Y to co-vary negatively.
- Positive \(\rho\) indicates a tendency for the variables X and Y to co-vary positively.
- \(-1 \leq \rho \leq 1\)
- \(\rho = -1 \iff\) all pairs (X,Y) lie on a straight line with negative slope.
- \(\rho = 1 \iff\) all pairs (X,Y) lie on a straight line with positive slope.
- Units of \(\rho\) ?
- \(\rho\) is the Covariance of the standardized random variables \(\frac{X-\mu_x}{\sigma_x}\) and \(\frac{Y-\mu_y}{\sigma_y}\). (Details left to student.)
- \(\rho = 0 \iff\) Cov(X,Y) = 0.
Uncorrelated variables:

X and Y are uncorrelated means $\rho_{X,Y} = 0$ (or equivalently, $\text{Cov}(X,Y) = 0$).

Examples:

1. X and Y are independent $\Rightarrow$ uncorrelated. (Why?)

2. X uniform on the interval [-1, 1].
   - Y = $X^2$.
   - X and Y are uncorrelated (details homework)
   - X and Y not independent. (E(Y|X) not constant)

$\rho$ is a measure of the degree of a “overall” nonconstant linear relationship between X and Y.

Example 2 shows: Two variables can have a strong nonlinear relationship and still be uncorrelated.

Sample variance, covariance, and correlation

$(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n)$ sample of data from the joint distribution of X and Y

Sample covariance:

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

Sample correlation coefficient

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

- Estimates of the corresponding population parameters.
- Analogous properties.