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. μ_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. (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

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

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

$$= \int_{-\infty}^{\infty} h(y)\, f_X(x)\,dx$$

$$= h(y)\int_{-\infty}^{\infty} f_X(x)\,dx = h(y) = f(y|x)$$

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: *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.

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."

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.

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

Covariance: For random variables X and Y,

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

Comments:

- Cov (capital C) $\longleftrightarrow$ population
  cov (or Cov-hat) $\longleftrightarrow$ sample.
  Cov is a *parameter* $\longleftrightarrow$ population
  cov is a *statistic* $\longleftrightarrow$ calculated from the sample.

- Compare and contrast with definition of 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 Cov(X,Y) is positive. Similarly, if X and Y tend to be on opposite sides of their respective means, then Cov(X,Y) is negative. If there is no trend of either sort, then 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.

- Units of Cov(X,Y)?

***Properties*:**

1. Cov(X, X) =

2. Cov (Y, X) =

3. Cov (X, Y) = E([X - E(X)][Y - E(Y)]) =

   In words …

   (Compare with the alternate formula for Var(X).)

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

   Cov(X,Y) =

   Note: Converse false. (Future homework.)

5. Cov(cX, Y) =

Cov(X, cY) =

6. Cov(a + X,Y) =

Cov(X, a +Y) =

7. Cov(X + Y, Z)) =

Cov(X, Y + Z) =

8. Var(X + Y) =

Consequence: If X and Y are independent, then

Var(X + Y) =

(converse false!)

When else might this be true?

Bounds on Covariance

σ_X = *population standard deviation* $\sqrt{Var(X)}$ of X.

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

σ_Y defined similarly.

Consider the new random variable $\dfrac{X}{\sigma_X} + \dfrac{Y}{\sigma_Y}$.

Since Variance is always ≥ 0,

$$(*) \quad 0 \le \mathrm{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right)$$

$$= \mathrm{Var}\left(\frac{X}{\sigma_X}\right) + \mathrm{Var}\left(\frac{Y}{\sigma_Y}\right) + 2\mathrm{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right)$$

$$= \frac{1}{\sigma_x^2}\mathrm{Var}(X) + \frac{1}{\sigma_Y^2}\mathrm{Var}(Y) + \frac{2}{\sigma_X\sigma_Y}\mathrm{Cov}(X,Y)$$

=

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

Therefore:

(**) $$\frac{Cov(X,Y)}{\sigma_X\sigma_Y} \geq -1$$

Equivalently: Cov(X, Y) ≥ - $\sigma_X\sigma_Y$.

Looking at $\mathrm{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right)$ similarly shows:

(***) $$\frac{Cov(X,Y)}{\sigma_X\sigma_Y} \leq 1$$

Equivalently: Cov(X, Y) ≤ $\sigma_X\sigma_Y$.

(details left to the student)

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

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

Equivalently: |Cov(X, Y)| ≤ $\sigma_X\sigma_Y$

Equality in (**) ⇔ equality in (*) -- i.e,

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

$\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$ 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

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

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

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

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

Note:
- ρ for short.
- ρ 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 ρ:

- Negative ρ indicates a tendency for the variables X and Y to co-vary negatively.
- Positive ρ indicates a tendency for the variables X and Y to co-vary positively.
- $-1 \leq \rho \leq 1$
- $\rho = -1 \Leftrightarrow$ all pairs (X,Y) lie on a straight line with negative slope.
- $\rho = 1 \Leftrightarrow$ all pairs (X,Y) lie on a straight line with positive slope.
- Units of ρ ?
- 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 student.)
- $\rho = 0 \Leftrightarrow Cov(X,Y) = 0$.

Uncorrelated variables:

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

Examples:

1. X and Y are independent ⇒ 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)

ρ 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:

cov(x,y) (or)

$$= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Sample correlation coefficient

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

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