Lecture 12

Covariance, Correlation and Conditional Expectation

Def. (Covariance)

Let X be a r.v. and denote $E(X) = \mu_X$.
Let Y be a r.v. with $E(Y) = \mu_Y$.

The covariance of X and Y is

(1) $\operatorname{cov}(X, Y) = E[(X-\mu_X)(Y-\mu_Y)]$.

Result (How to simplify the calculation in (1)).

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

- The covariance of two variables is a measure of their dependency.
- If $\operatorname{cov}(X,Y)$ is high, we can be confident that X and Y are dependent. Special care is needed to interpret a low covariance value.

How to calculate cov(X, Y)?

Example 4.6.2

$$(X, Y) \sim f(x, y) = \begin{cases} 2xy + 0.5; & \text{if } 0 \le x, y \le 1 \\ 0 & ; \text{ otherwise.} \end{cases}$$

Find cov(X, Y).

Solution We use the formula $E(XY) - E(X) \cdot E(Y) = \text{cov}(X, Y)$.

- $f_X(x) = \int_0^1 f(x,y)\,dy$. Therefore, for $x \in (0, 1)$;

$$E(X) = \mu_X = \int_0^1 x \cdot f_X(x)\,dx$$

$$= \int_0^1 x \left[\int_0^1 (2xy + 0.5)\,dy \right] dx.$$

$$= \int_0^1 x \left[2x \cdot \frac{y^2}{2} \Big|_{y=0}^{y=1} + 0.5 \right] dx = \int_0^1 (x^2 + 0.5)\,dx.$$

$$= \frac{7}{12}.$$

- Similarly, $E(Y) = \frac{7}{12}$.

- $$E(XY) = \int_0^1\int_0^1 xy\,[2xy + 0.5]\,dx\,dy$$
$$= \frac{50}{144}$$

Thus $\quad cov(XY) = \frac{50}{144} - \frac{7}{12}\cdot\frac{7}{12} = \frac{1}{144}$ //.

The correlation of X and Y

To measure the ~~s~~association between X and Y in a way that is not influenced by the scales of X and/or Y, we introduce:

Definition (Correlation between X and Y).

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y}, \quad \text{where}$$

$\sigma_X^2 = var(X)$ and $\sigma_Y^2 = var(Y)$.

- We can check that:

$$\rho(aX, bY) = \rho(X, Y).$$

$$\rho(aX, bY) = \frac{cov(aX, bY)}{\sqrt{var(aX)} \cdot \sqrt{var(bY)}}. \quad (3)$$

$$cov(aX, bY) = E(abXY) - [a E(X)] \cdot [b \cdot E(Y)]$$
$$= ab[E(XY) - E(X) \cdot E(Y)]$$
$$= a \cdot b \cdot cov(X, Y) \quad (1)$$

$$var(aX) = a^2 var(X); \quad var(bY) = b^2 \cdot var(Y). \quad (2)$$

By (1), (2) and (3):

$$\rho(aX, bY) = \frac{\cancel{a \cdot b} \cdot cov(X, Y)}{\cancel{a \cdot b} \cdot \sqrt{var(X)} \cdot \sqrt{var(Y)}} = \rho(X, Y).$$

Very important result

(Display (4.6.10) page 251).

$$\boxed{-1 \leq \rho(X,Y) \leq 1, \text{ for } \underline{\text{any}} \; X \text{ and } Y \text{ with finite variance.}}$$

Proof See book. It uses a very important inequality, the Cauchy - Schwarz inequality
(Not required).

$$[E(UV)]^2 \leq E(U^2) \cdot E(V^2).$$

Jargon :

- $\rho(X,Y) > 0$: X and Y are positively correlated
- $\rho(X,Y) < 0$: Negatively correlated.
- $\rho(X,Y) = 0$: Uncorrelated.

How to use/interpret the concept of correlation?

Theorem   If X and Y are independent, then X and Y are uncorrelated:

$$\mathrm{cov}(X,Y) = \rho(X,Y) = 0.$$

NOTE.   The other direction is NOT true: We can have $\rho(X,Y)=0$ for dependent variables.

Moral   Lack of correlation between random variables is $\boxed{\text{NOT EQUIVALENT}}$ with lack of independence.

Example. Let $X \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$ and

let $Y = X^2 \begin{pmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$.

- Clearly Y and X are dependent, as one is a function of the other.
- However: $E(XY) = E(X^3) = E(X) = 0$.

Thus $\operatorname{cov}(X,Y) = E(XY) - E(X) \cdot E(Y)$

$= 0 - 0 \cdot 0 = 0 \Rightarrow$

$\rho(X,Y) = 0$, therefore X and Y

are uncorrelated, but dependent.

Reading example Example 4.6.5. (Required).

Theorem

$|\rho(X,Y)| = 1$ <u>if and only if</u>

X and Y are linearly related.

<u>Remark</u>:
- If $Y = aX + b$ and $a > 0 \Rightarrow \rho(X,Y) = 1$
- If $Y = aX + b$ and $a < 0 \Rightarrow \rho(X,Y) = -1$

<u>Conclusion</u> <u>Correlation measures only linear relationships</u>.

- $\rho(X,Y)$ ~~large~~ suggests that X and Y are linearly related, therefore dependent, with a <u>linear</u> dependency structure.
- $|\rho(X,Y)| \neq 1$ (for instance, $=0$) only means that X and Y are <u>not linearly related</u>, but that does not mean that they cannot be related quadratically, for instance.

ρ –

Formula for variances of sums of random variables.

Result 1 $var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$, for any X and Y with finite variances.

Result 2 If X and Y are independent, then (1) $var(X+Y) = var(X) + var(Y)$.

(2) $var(X-Y) = var(X) + var(Y)$.

- For generalizations to more than 2 random variables, see Th. 4.6.7 and Corollary 4.6.2.

Result 2 (For 3 variables)

$var(X+Y+Z) = var(X) + var(Y) + 2cov(X,Y) + 2cov(X,Z) + 2cov(Y,Z)$.

Conditional expectation and variance

- $$E(Y \mid X = x) = \int_{-\infty}^{\infty} y \cdot g_2(y \mid x)\, dy,$$

for continuous distributions.

Here $g_2(y|x)$ is the conditional distribution of Y given $X = x$.

- Notice that $E(Y/X = x)$ changes with x and is therefore a function of x.

It is a random variables that is just a function of X and we denote it by $E(Y/X)$.

Result $E(E(Y|X)) = E(Y)$.

Definition (4.7.3)

- $var(Y \mid X = x) =: var(Y \mid x)$ A number

$$= E\{[Y - E(Y \mid x)]^2 \mid x\}.$$

- $var(Y \mid X) = E[[Y - E(Y \mid X)]^2 \mid X]$. A random variable

Prediction.

The function $d(X)$ that minimizes the mean squared error

$$E(Y - d(X))^2$$ is

$$d(X) = E(Y \mid X).$$

- **Result**. $E(var(Y \mid X)) = E[Y - E(Y \mid X)]^2$.

Therefore the prediction of X via $E(Y \mid X)$ minimizes $var(Y \mid X)$, averaged over all possible values of X.

- Recall that

$$E(Y) = E(E(Y|X)), \text{ for any}$$

r.v. X, Y for which this can be defined.

- Theorem

- $Var(Y) = E(\,var(Y|X)) + Var(E(Y|X))$, for any ~~values~~ r.v.'s X, Y for which the above quantities can be defined.

IMPORTANT RULES FOR PREDICTING Y via X.

(1) "Best" prediction of Y, $\boxed{\text{before}}$ X is observed, is the random variable

→ $E(Y|X)$. Its MSE is

$$E(Y - E(Y|X))^2 = E(\,var(\cancel{Y}|X)).$$

This MSE is called the overall MSE.

(2) Best prediction of Y $\boxed{\text{after } X = x}$ has been observed is the <u>number</u> (which depends on the value of x) $\underline{E(Y | X = x)}$.

The MSE for <u>this</u> prediction is $var(Y|x)$. A useful formula for it is:

$$\boxed{Var(Y|x) = E(Y^2|x) - [E(Y|x)]^2}$$

93 or more	A	90 - less than 93	A-
86 - less than 90	B+	83 - less than 86	B
80 -less than 83	B-	76 - less than 80	C+
73 - less than 76	C	70 -less than 73	C-
66 - less than 70	D+	63 - less than 66	D
60 - less than 63	D-	less than 60	F