cse 107 2-24-25

- mid 2: Fri 2/28
- labs soon

4.2 Covariance and Correlation

Defn

The covariance of r.v.s X, Y is

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

<u>note</u>: $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.

we say X, Y are <u>uncorrelated</u> iff $\mathrm{Cov}(X,Y) = 0$.

if $\mathrm{Cov}(X,Y) > 0$ we say X, Y are <u>positively correlated</u>.

if $\mathrm{Cov}(X,Y) < 0$ we say X, Y are <u>negatively correlated</u>.

Observe:

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

$$= E\left[XY - E[X]Y - E[Y]X + E[X]E[Y]\right]$$

$$= E[XY] - E[X]E[Y] \cancel{- E[X]E[Y]} + \cancel{E[X]E[Y]}$$

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

Note:

X, Y independent

$$\Rightarrow E[XY] = E[X] \cdot E[Y]$$

$$\Rightarrow \text{Cov}(X, Y) = 0$$

$\Rightarrow$ X, Y uncorrelated.

The converse is false! See Ex. 4.13 p. 218-219 in text.

Recall logical contrapositive:

$$p \Rightarrow q \equiv \neg q \Rightarrow \neg p$$

so

Positive (or negative) correlation

$\Rightarrow X, Y$ not independent.

<u>Defn</u>

The <u>correlation coefficient</u> is

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X) \cdot \operatorname{Var}(Y)}}$$

It can be shown

$$-1 \le \rho(X,Y) \le 1$$

$$\underset{\text{high neg. correlation}}{\uparrow} \qquad\qquad \underset{\text{high Pos. correlation}}{\uparrow}$$

high neg. correlation

high Pos. correlation.

Ex. We toss a coin n times with $P(\text{head}) = p$. Let

$X = \#\text{heads}$

$Y = \#\text{tails}$

find $\rho(X,Y)$.

Soln

observe $X+Y=n$, so

$$E[X]+E[Y]=E[X+Y]=E[n]=n$$

and

$$\begin{aligned}Var(Y)&=Var(-X+n)\\&=(-1)^2Var(X)=Var(X)\end{aligned}$$

also

$$(X-E[X])+(Y-E[Y])=n-n=0$$

so

$$Y-E[Y]=-(X-E[X])$$

Thus

$$\mathrm{Cov}(X,Y) = E\left[-(X-E[X])^2\right]$$

$$= -\mathrm{Var}(X)$$

and

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}}$$

$$= \frac{-\mathrm{Var}(X)}{\sqrt{\mathrm{Var}(X)^2}} = \frac{-\mathrm{Var}(x)}{\mathrm{Var}(X)}$$

$$= \boxed{-1}$$

Perfect negative correlation.

Variance of a Sum

Let X, Y be r.v.'s (not necessarily independent.) Then

Theorem

$$\boxed{\begin{aligned} &\mathrm{Var}(X+Y) = \\ &\qquad \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y) \end{aligned}}$$

more generally

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$$

Proof of case $n=2$:

Let

$$U = X - E[X] \quad \therefore E[U] = 0$$

$$V = Y - E[Y] \quad \therefore E[V] = 0.$$

Then

$$Var(X+Y) = Var(U+V)$$

$$= E[((U+V) - 0)^2]$$

$$= E[(U+V)^2]$$

$$= E[U^2 + V^2 + 2UV]$$

$$= E[U^2] + E[V^2] + 2E[UV]$$

$$= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$$

$\blacksquare$

<u>Exercise</u>

Prove general formula.

See Ex. 4.15 on P. 221.

4.3 Conditional expectation and variance

Recall

$$g(y) = E[X \mid Y = y]$$

is a function of y. If we substitute Y for y we get a new r.v.

$$g(Y) = E[X \mid Y]$$

called: conditional expectation

The expected value of this r.v. is

$$E[E[X|Y]] = E[g(Y)]$$

$$= \int_{-\infty}^{\infty} g(y) \cdot f_Y(y)\, dy \quad \left\{ \begin{array}{l} \text{expected} \\ \text{value} \\ \text{rule} \end{array} \right.$$

$$= \int_{-\infty}^{\infty} E[X|Y=y] \cdot f_Y(y)\, dy$$

$$= E[X] \quad \left\{ \begin{array}{l} \text{total expectation} \\ \text{theorem} \end{array} \right.$$

so we get the law of iterated expectations.

$$\boxed{E[E[X \mid Y]] = E[X]}$$