CSE 107 5-20-25

- midterm 2: Th 5/22

4.2 Covariance & Correlation

Defn

The covariance of r.v.s X, Y

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

observe

$$\mathrm{Cov}(X,X) = \mathrm{Var}(X).$$

we say X, Y are

- <u>uncorrelated</u> iff $\mathrm{Cov}(X,Y)=0$
- <u>Positively correlated</u> iff $\mathrm{Cov}(X,Y)>0$
- <u>negatively correlated</u> iff $\mathrm{Cov}(X,Y)<0$.

<u>note</u>:

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

$$=E[XY-E[X]Y-XE[Y]+E[X]E[Y]]$$

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

$$=E[XY]-E[X]E[Y]$$

Recall:

X, Y indep. $\Rightarrow E[XY] = E[X] \cdot E[Y]$

$\Rightarrow Cov(X,Y) = 0$

$\Rightarrow X, Y$ uncorrelated.

Converse is false!!

(example 4.13 on p. 218–219.)

Defn

The correlation coefficient $\rho(X,Y)$ is

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

Thm

$$\underset{\substack{\text{Perfect}\\\text{negative}\\\text{correlation}}}{\underset{\uparrow}{-1}} \le \rho(X,Y) \le \underset{\substack{\text{Perfect}\\\text{positive}\\\text{correlation.}}}{\underset{\uparrow}{1}}$$

Ex.

We toss a coin n times, $P(\text{head}) = p$. let

$$X = \#\text{heads}$$

and

$$Y = \#\text{tails}$$

find $\rho(X, Y)$.

Soln

observe $X + Y = n$ so

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

also

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

also

$$\begin{aligned} &(X-E[X])+(Y-E[Y]) \\ &= (X+Y) - (E[X]+E[Y]) \\ &= n-n=0 \end{aligned}$$

$$\therefore \quad Y-E[Y] = -(X-E[X])$$

7

Thus

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

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

and

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

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

$$= \frac{-\mathrm{Var}(X)}{\mathrm{Var}(X)} = \boxed{-1}$$

Exercise

- Prove $\mathrm{Cov}(aX+b, Y) = a\,\mathrm{Cov}(X,Y)$
- Prove $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
- Prove if $a>0$, then

$$\rho(aX+b, Y) = \rho(X,Y)$$

- Prove $\mathrm{Cov}(aX+bY, Z)$

$$= a\,\mathrm{Cov}(X,Z) + b\,\mathrm{Cov}(Y,Z)$$

Variance of a Sum

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

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

more generally

$$\boxed{\begin{aligned} &Var\left(\sum_{i=1}^{n} X_i\right) \\ &= \sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j) \end{aligned}}$$

Proof of case $n=2$

let

$$U = X - E[X] \Rightarrow E[U] = 0$$

$$V = Y - E[Y] \Rightarrow E[V] = 0$$

$$\therefore U+V = (X+Y) - E[X+Y] \Rightarrow E[U+V] = 0$$

Then

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

$$= Var(U+V)$$

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

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

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

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

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

$\blacksquare$

<u>Exercise</u>

Prove case $n=3$.

<u>Example</u> 4.15 on p. 221.

4.3 Conditional Expection & Conditional Variance

Recall that

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

is a function. substitute Y in for y.

$$g(Y) = E[X \mid Y] \quad \left\{ \begin{array}{l} \text{Conditional} \\ \text{expectation} \end{array} \right.$$

To find $E[g(Y)]$, use expected value rule

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

$$= \int_{-\infty}^{\infty} g(y) \cdot f_Y(y)\, dy$$

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

$$= E[X] \quad \left\{ \text{by the total expectation theorem} \right.$$

we call this the <u>law of iterated expecations</u>

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

we do something similar for variances

$$h(y) = Var(X|Y=y)$$

$$= E[(X - E[X|Y=y])^2 | Y=y]$$

now subsititute Y for y

$$Var(X|Y) = h(Y)$$

$$= E\left[(X - E[X|Y])^2 \mid Y\right]$$

We call $Var(X|Y)$ the <u>conditional variance</u>.

<u>Law of total Variance</u>

$$\boxed{Var(X) = E[Var(X|Y)] + Var(E[X|Y])}$$

(See P. 225-227 for a proof.)

Proof

By Variance in terms of moments formula (conditional version)

$$Var(X|Y) = E[X^2|Y] - E[X|Y]^2$$

Thus

$$E[X^2|Y] = Var(X|Y) + E[X|Y]^2$$

by law of iterated expectation:

$$Var(X) = E[X^2] - E[X]^2$$

$$= E[E[X^2|Y]] - E[E[X|Y]]^2$$

$$= E\left[\mathrm{Var}(X|Y) + E[X|Y]^2\right]$$

$$- E\left[E[X|Y]\right]^2$$

$$= E[\mathrm{Var}(X|Y)]$$

$$E\left[E[X|Y]^2\right] - E\left[E[X|Y]\right]^2$$

$$= E[\mathrm{Var}(X|Y)] + \mathrm{Var}(E[X|Y])$$

$\blacksquare$

4.5 Sum of a random number of random variables

Let $X_1, X_2, \ldots$ be an (infinite) seq. of independent, identically distributed (iid) r.v.s. This means all have same PDF, denoted

$$f_X(x) \quad \text{or} \quad p_X(x).$$

Let N be a r.v. taking values in $\{0, 1, 2, 3, \ldots\}$.

Assume

$\{N, X_1, X_2, \ldots\}$

are independent. Define

$$Y = X_1 + X_2 + \cdots + X_N$$

denote by $E[X]$, $Var(X)$ the common mean and variance of the X_i $(i = 1, 2, \ldots)$.

Goal: find $E[Y]$, $Var(Y)$.