CSE 107 2-24-26

- Mid 2 : Th 2/26

(4.2) .... continue ....

<u>Defn</u>

The <u>correlation coefficient</u> of r.v.s X, Y is

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

Theorem

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

$\uparrow$ Perfect Neg. $\qquad\qquad$ $\uparrow$ Perfect Pos.

(SEE handout Proof.)

Ex.

Toss a coin n times, $P(\text{head}) = p$. Let X = heads, Y = tails. Find $\rho(X,Y)$.

solution:

observe $X + Y = n$, so

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

and

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

Also

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

so

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

Thus

$$\begin{aligned} Cov(X,Y) &= E\left[(X-E[X])(Y-E[Y])\right] \\ &= E\left[-(X-E[X])^2\right] \\ &= -E\left[(X-E[X])^2\right] \\ &= -Var(X) \end{aligned}$$

and

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

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

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

Variance of a Sum

Theorem

Let X, Y be r.v.s (not necc. indep.).

Then

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

.

More generally

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

Proof in case $n=2$

Let

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

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

Then

$$
\begin{aligned}
Var(X+Y) &= Var(U+V) \\
&= E\left[\left((U+V) - \cancelto{0}{E[U+V]}\right)^2\right] \\
&= E\left[(U+V)^2\right] \\
&= E\left[U^2 + V^2 + 2UV\right] \\
&= E[U^2] + E[V^2] + 2E[UV] \\
&= Var(X) + Var(Y) + 2\,Cov(X,Y)
\end{aligned}
$$

Exercise

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

4.3 Conditional Mean & Variance

Recall

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

If we replace y by Y, we get

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

What is the expected value of this r.v. ?

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

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

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

$$= E[X]$$

law of iterated expectations

Thus

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

Can do something similar with Variance.

Define a function

$$h(y) = Var(X \mid Y = y)$$

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

Replace y by Y to get

$$h(Y) = Var(X \mid Y) \leftarrow \underline{\text{conditional}}\ \underline{\text{Variance}}$$

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

Theorem (law of total Variance)

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

Proof.

by Var( ) in terms of moments

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

Thus

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

Then

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

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

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

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

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

$\blacksquare$

<u>skip to 4.5</u>

<u>4.4 is on transforms</u>

4.5 Sum of a random number of random variables

Let $X_1, X_2, X_3, \ldots$ be an (infinite) sequence of independent, identically distributed (i.i.d) r.v.s. i.e. all have same PDF, or PMF:

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

Let $E[X]$, $Var(X)$ be their common mean and variance

Let N be a r.v. with range$(N) \subseteq \{0,1,2,\dots\}$.

Assume $\{N, X_1, X_2, \dots\}$ is indep. Now define

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

<u>Goal</u>: find $E[Y]$, $Var(Y)$

let $n \geq 0$ be fixed. Then

$$
\begin{aligned}
&E[Y \mid N=n] \\
&= E[X_1 + \cdots + X_N \mid N=n] \\
&= E[X_1 + \cdots + X_n \mid N=n] \\
&= E[X_1 + \cdots + X_n] \quad \left\{ \begin{array}{l} \text{by} \\ \text{indep.} \end{array} \right. \\
&= E[X_1] + \cdots + E[X_n] \\
&= n\, E[X]
\end{aligned}
$$

$$\therefore \quad E[X \mid N] = N \cdot E[X]$$

By law of iterated expectation

$$E[Y] = E[E[Y|N]]$$

$$= E[N \cdot E[X]]$$

$$= E[N] \cdot E[X]$$

Also

$$Var(Y|N=n)$$

$$= Var(X_1 + \cdots + X_N | N=n)$$

$$= Var(X_1 + \cdots + X_n | N=n)$$

$$= Var(X_1 + \cdots + X_n) \quad \{ \text{by indep.}$$

$$= Var(X_1) + \cdots + Var(X_n) \quad \left\{ \begin{array}{l} \text{by} \\ \text{indep.} \end{array} \right.$$

$$= n\,Var(X)$$

so

$$Var(Y|N) = N \cdot Var(X)$$

by total Variance

$$Var(Y) = E[Var(Y|N)] + Var(E[Y|N])$$

$$= E[N \cdot Var(X)] + Var(N \cdot E[X])$$

$$= E[N] \cdot Var(X) + E[X]^2 \cdot Var(N)$$