Lecture 8

(4 September 2025)

1) $Y = aX + b$

$$E[Y] = aE[X] + b \quad Var(Y) = a^2 Var(X).$$

2) Linearity of Expectation

$$E[X_1 + X_2] = E[X_1] + E[X_2].$$

$$E[X_1 + X_2] = \sum_{x_1, x_2} (x_1 + x_2)\, p_{X_1, X_2}(x_1, x_2)$$

(by LOTUS)

$$= \sum_{x_1} x_1 \left(\sum_{x_2} p_{X_1, X_2}(x_1, x_2) \right)$$

$$+$$

$$\sum_{x_2} x_2 \left(\sum_{x_1} p_{X_1, X_2}(x_1, x_2) \right)$$

$$= \sum_{x_1} x_1 p_{X_1}(x_1) + x_2 p_{X_2}(x_2)$$

$$= E[X_1] + E[X_2].$$

Similarly $E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i].$

Example: Consider a binomial random variable Y with parameters n and p.

$$Y = \sum_{i=1}^{n} X_i \quad \text{where} \quad X_i = \begin{cases} 1 & \text{if } i^{th} \text{ trial is heads} \\ 0 & \text{if } i^{th} \text{ trial is tails.} \end{cases}$$

$X_1, X_2, \ldots, X_n$ are independent.

$$E[Y] = E\left[\sum_{i=1}^{n} X_i\right]$$

$$= \sum_{i=1}^{n} E[X_i] = np, \quad \text{as } E[X_i] = p \ \forall i.$$

3) $Var(X+Y) = Var(X) + Var(Y) + Cov(X, Y)$.

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

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

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

$$+ 2E\left[(X-E[X])(Y-E[Y])\right]$$

$$= Var(X) + Var(Y) + Cov(X, Y).$$

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

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

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

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

The concept of covariance generalizes that of variance in that $\mathrm{cov}(X,X) = \mathrm{var}(X)$.

- X and Y are uncorrelated if $\mathrm{cov}(X,Y) = 0$.
- Independent random variables are always uncorrelated, although the converse is not true.

The correlation coefficient of X and Y is defined as

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \cdot \sigma_Y}.$$

<u>Theorem</u>. $|\rho(X,Y)| \leq 1$ with equality if and only if $Y = aX + b$ with probability 1, for some $a, b \in \mathbb{R}$.

The proof of this theorem is an application of Cauchy-Schwarz inequality.

Cauchy-Schwarz Inequality

For random variables $\underline{x}$ and $\underline{y}$

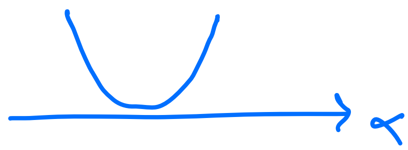
$$|E[xy]| \leq \sqrt{E[x^2]E[y^2]},$$

with equality if and only if $x = \alpha y$ with probability $\underline{1}$ for some $\alpha \in \mathbb{R}$.

Proof. $0 \leq E[(x-\alpha y)^2]$

$$= E[x^2] + \alpha^2 E[y^2] - 2\alpha E[xy]$$

Discriminant is non-positive



$$\Longrightarrow \cancel{4} E[xy]^2 \leq \cancel{4} E[x^2] E[y^2]$$

$$\Longrightarrow |E[xy]| \leq \sqrt{E[x^2]E[y^2]}.$$

Discriminant $= 0$ if and only if the quadratic has a real root if and only if $x = \alpha \underline{y}$ for $\alpha \in \mathbb{R}$

$\Longrightarrow$ equality holds if and only if $x = \alpha \underline{y}$ for $\alpha \in \mathbb{R}$.

Proof of $|\rho(\underline{x},\underline{y})| \leq 1$. Apply Cauchy-Schwarz to the random variables $x - E[x]$ and $y - E[y]$.

4) If X and Y are independent then

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y).$$

If $X_1, X_2, \ldots, X_n$ are independent then

$$\mathrm{var}(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{var}(X_i).$$

<u>Example</u>. $Y \sim \mathrm{Binomial}(n, p)$, $Y = \sum_{i=1}^{n} X_i$

$$\text{where } X_i = \begin{cases} 1 & \text{if } i^{th} \text{ trial is heads} \\ 0 & \text{if } i^{th} \text{ trial is tails,} \end{cases}$$

$X_1, X_2, \ldots, X_n$ are independent.

$$\mathrm{Var}(Y) = \sum_{i=1}^{n} \mathrm{var}(X_i) = np(1-p).$$

5) $Z = X + Y$

$$p_Z(z) = \sum_{(x,y):\, z = x+y} p_{XY}(x, y) = \sum_{x} p_{XY}(x, z-x).$$

If X and Y are independent

$$p_Z(z) = \sum_{x} p_X(x)\, p_Y(z-x) = \sum_{y} p_X(z-y)\, p_Y(y).$$

$$= p_X * p_Y \quad (\text{convolution}).$$

Exercise. If X_1 and X_2 are independent geometric random variables with common pmf

$$p_X(k) = (1-p)^{k-1}p \quad k = 1, 2, \ldots$$

$$p_Z(z) = (z-1)p^2(1-p)^{z-2} \quad z = 2, 3, \ldots$$

$$\sum_{k=1}^{z-1} (1-p)^{k-1}p \, (1-p)^{z-k-1}p = (z-1)p^2(1-p)^{z-2}.$$

Conditioning

Conditioning a RV on an event:

The conditional pmf of a RV X conditioned on a particular event A with $P(A)>0$ is defined as

$$p_{X|A}(x) = P(X = x \mid A)$$

$$= \frac{P(\{\omega : X = x\} \cap A)}{P(A)}.$$

$$\sum_x p_{X|A}(x) = 1.$$

Example. Let X = roll of a fair die, $A = \{2, 4, 6\}$.

$$p_{X|A}(k) = \begin{cases} 1/3 & \text{if } k = 2, 4, 6 \\ 0 & \text{o.w.} \end{cases}$$

Exercise. If $A_1, A_2, \dots, A_n$ form a partition of the sample space, with $P(A_i)>0 \; \forall i$, then

$$p_X(x) = \sum_{i=1}^{n} P(A_i)\, p_{X|A_i}(x).$$

Conditioning one RV on another:

Consider two jointly discrete RVs X & Y. If we know that the value of Y is some particular y with $p_Y(y)>0$, this provides partial knowledge about the value of X. This knowledge is captured by the conditional PMF $p_{X|Y}$ defined as

$$p_{X|Y}(x|y) = P(X=x \mid Y=y)$$

$$= \frac{p_{X,Y}(x,y)}{p_Y(y)}, \quad \text{if } p_Y(y)>0.$$

$$\sum_x p_{X|Y}(x|y) = 1.$$

$$p_{X,Y}(x,y) = p_{X|Y}(x|y)\, p_Y(y) = p_{Y|X}(y|x)\, p_X(x).$$