Lecture 9

(4 September 2025)

Conditional Expectation

The conditional expectation of X given $Y=y$ is defined as

$$E[X|Y=y] = \sum_x x\, p_{X|Y}(x|y).$$

Similarly for an event A with $P(A)>0$,

$$E[X|A] = \sum_x x\, P_{X|A}(x).$$

Theorem. $E[g(X)|A] = \sum_x g(x)\, p_{X|A}(x).$

The proof of this theorem is exactly similar to the proof of LOTUS $E[g(X)] = \sum_x g(x) p_X(x)$.

Total Expectation Theorem

If the events $A_1, A_2, ---, A_n$ form a partition of the sample space, with $P(A_i)>0$ for all $i \in [1:n]$, then

$$E[X] = \sum_{i=1}^{n} P(A_i)\, E[X|A_i].$$

Proof. $\sum_{i=1}^{n} P(A_i) E[x|A_i]$

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

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

$= \sum_{x} x \sum_{i=1}^{n} P(X=x \cap A_i)$

$= \sum_{x} x P(X=x \cap (\bigcup_{i=1}^{n} A_i))$

$= \sum_{x} x \; P(X=x \cap \Omega)$

$= \sum_{x} x \; p_X(x).$

Taking $\{Y=y\}_{y \in Y}$ as the partition of Ω, the above total expectation theorem gives

$$E[x] = \sum_{y} p_Y(y) E[X|Y=y].$$

Conditional Expectation as a RV

Let $\phi(y) = E[X|Y=y]$

Y is a RV, so $\phi(Y)$ is also a RV.

We denote $\phi(Y) \equiv E[X|Y]$.

$E[X|Y]$ is a function of RV Y,

$E[X|Y](\omega) = \phi(Y(\omega))$, $\forall \omega \in \Omega$.

Law of Iterated Expectations

$$E[E[X|Y]] = E[X].$$

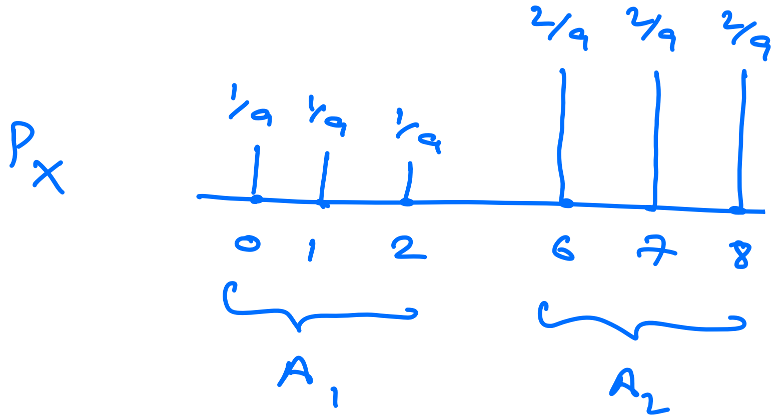
$\phi(y) = E[X|Y=y]$, $y \in \mathcal{Y}$.

$\phi(Y) = E[X|Y]$.

$$\begin{aligned} E[E[X|Y]] &= E[\phi(Y)] \\ &= \sum_{y} \phi(y) P_Y(y) \\ &= \sum_{y} E[X|Y=y] P_Y(y) \\ &= E[X]. \end{aligned}$$

(by total exp. theorem)

Example.



$$A_1 = \{0 \le x \le 2\}, \quad A_2 = \{6 \le x \le 8\}.$$

$$P(A_1) = \frac{1}{3}, \quad P(A_2) = \frac{2}{3}$$

$$p_{X|A_1}(x) = \begin{cases} 1/3, & \text{if } x = 0, 1, 2 \\ 0, & \text{o.w.} \end{cases}$$

$$p_{X|A_2}(x) = \begin{cases} 1/3, & \text{if } x = 6, 7, 8 \\ 0, & \text{o.w.} \end{cases}$$

$$E[X|A_1] = \sum_x x\, p_{X|A_1}(x) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3}$$

$$= 1.$$

$$E[X|A_2] = \sum_x x\, p_{X|A_2}(x) = 6 \cdot \frac{1}{3} + 7 \cdot \frac{1}{3} + 8 \cdot \frac{1}{3}$$

$$= \frac{21}{3} = 7.$$

$$E[X] = E[X|A_1]P(A_1) + E[X|A_2]P(A_2)$$

$$= 1\left(\frac{1}{3}\right) + 7\left(\frac{2}{3}\right) = 5.$$

Conditional Independence

Two random variables x and y are conditionally independent given an event A if
$(p(A)>0)$

$$p_{\underline{x}y|A}(\underline{x}\,y) = p_{x|A}(x)\; p_{y|A}(\underline{y}) \quad \forall\, \underline{x}\, y,$$

Here $p_{\underline{x}y|A}(\underline{x}\,y) = \dfrac{p(X=x \cap Y=y \cap A)}{p(A)}$,

Exercise

Are x and y independent?

Are x and y conditionally x independent given

$A = \{X \geq \underline{3} \;\; Y \leq 2\}$?

$p_{\underline{x}y}$	1	2	y 3	4
1	0	1/20	0	0
2	0	1/20	3/20	1/20
3	2/20	4/20	1/20	2/20
4	1/20	2/20	2/20	0

- X and Y are said to be conditionally independent given Z if

$$p_{\underline{x}y|z}(\underline{x}\,y|z) = p_{x|z}(x|z)\, p_{y|z}(\underline{y}|z) \quad \forall\, z \text{ s.t. } p_z(z)>0,$$

Conditional Variance

Recall variance of X,

$$var(X) = E[(X-E[X])^2].$$

$$\sigma_X = \sqrt{var(X)}$$ (standard deviation)

Let $\psi(y) = E[(X-E[X|Y=y])^2 \mid Y=y]$.

$$= E[X^2 + E[X|Y=y]^2 - 2XE[X|Y=y] \mid Y=y]$$

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

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

Define $var(X|Y) = \psi(Y)$ is a RV, a function of Y.

Law of Total Variance

$$Var(X) = E[Var(X|Y)] + var(E[X|Y]).$$

Proof. $E[Var(X|Y)] = E[\psi(Y)]$

$$= \sum_y \psi(y) p_Y(y)$$

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

$$= E[X^2] - \sum_y E[X|Y=y]^2 P_Y(y).$$

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

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

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

$$= \sum_y (\phi(y) - E[X])^2 P_Y(y)$$

$$= \sum_y E[X|Y=y]^2 P_Y(y) + E[X]^2$$

$$- 2E[X] \sum_y \phi(y) P_Y(y)$$

$$= \sum_y E[X|Y=y]^2 P_Y(y) - E[X]^2.$$

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

$$= E[X^2] - E[X]^2 = \mathrm{Var}(X).$$

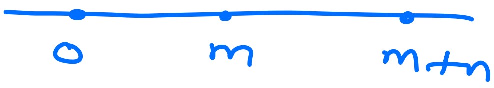
Memoryless Property of Geometric RV

Let X be a geometric RV with parameter p.

$$P(X>n) = \sum_{i=n+1}^{\infty} (1-p)^{i-1} p = \cancel{p} \cdot \frac{(1-p)^n}{\cancel{p}} = (1-p)^n .$$

For $m, n \in N$

$$P(X > m+n \mid X > m)$$



$$= \frac{P(X > m+n, X > m)}{P(X>m)}$$

$$= \frac{P(X>m+n)}{P(X>m)} = \frac{(1-p)^{m+n}}{(1-p)^m} = (1-p)^n$$

$$= P(X>n).$$

This is called memoryless property of the geometric random variable.

Exercise. If X is a positive integer valued RV satisfying $P(X>m+n \mid X>m) = P(X>n)$, then show that X is a geometric RV.

- We have seen that $F_X(x) = \sum_{i : x_i \le x} p_X(x_i)$ if $x \in \{x_1, x_2, \dots\}$. Then

$$p_X(x_i) = F_X(x_i) - F_X(x_{i-1}).$$

Q) $X = \max\{X_1, X_2, X_3\}$

where X_1, X_2 and X_3 are independend and identically distributed RV with common PMF $p_{X_i}(k) = \frac{1}{10}$. $\forall\, k \in [1:10]$, $\forall\, i \in [1:3]$.

Sol:- $F_X(k) = P(X \le k)$

$$= P(\max\{X_1, X_2, X_3\} \le k)$$

$$= P(X_1 \le k, X_2 \le k, X_3 \le k)$$

$$= P(X_1 \le k)\, P(X_2 \le k)\, P(X_3 \le k)$$

$$= F_{X_1}(k)\, F_{X_2}(k)\, F_{X_3}(k)$$

$$p_X(k) = F_X(k) - F_X(k-1)$$

$$= \left(\frac{k}{10}\right)^3 - \left(\frac{k-1}{10}\right)^3$$

$$= \frac{k^3 - (k-1)^3}{1000} = \frac{3k^2 - 3k + 1}{1000}.$$