First HW is posted on Tritoned, due next Thursday

X, Y discrete

$p_{X|Y}(\cdot \mid y)$ conditional p.m.f. of X

given $Y = y$,

$$p_{X|Y}(x_i \mid y) = \frac{p_{X,Y}(x_i, y)}{p_Y(y)}$$

$$= \mathbb{P}(X = x_i \mid Y = y)$$

provided $\mathbb{P}(Y = y) > 0$.

Example: a dice game

$1, 2, \cdots, 6$ sum $2, 3, 4, \cdots, 12$

$1, 2, \cdots, 6$

In a game, two dice are rolled and the sum of the two is observed.

$4 = 1+3 = 2+2 = 3+1$ $\frac{3}{6\times 6}$ $7 = 1+6 = 2+5 = 3+4$ $\frac{6}{36}$

If the sum is 4, the player loses. If the sum is 7 the player wins. Otherwise the game continues: the dice are rolled repeatedly until either a sum of 4 or 7 is observed.

Let X be the Bernoulli random variable that X=1 if in the end the player wins and X=0 if the player loses.

$X=1$ if 7 is observed in the end

$X=0$ if 4 is " "

Let N be the number of rolls until the end of the game.

$N =$ # rolls until we see 7 or 4.

What is the conditional distribution of X given N=k?

$N \sim \text{Geometric}(p)$, $p = P(\text{Sum}=4) + P(\text{sum}=7)$

$= \frac{3}{36} + \frac{1}{6} = \frac{3+6}{36} = \frac{1}{4}$.

$$p_{X|N}(\cdot \mid k)$$

$$p_{X|N}(0 \mid k) = \frac{\mathbb{P}(X=0, N=k)}{\mathbb{P}(N=k)}$$

$$= \frac{\mathbb{P}(\text{the first } k-1 \text{ rolls} \notin \{4,7\} \text{ and } k\text{-th roll is } 4)}{\mathbb{P}(\text{the first } k-1 \text{ rolls} \notin \{4,7\} \text{ and } k\text{-th roll} \in \{4,7\})}$$

$$= \frac{(1-p)^{k-1} \cdot \left(\frac{3}{36}\right)}{(1-p)^{k-1} \cdot \left(\frac{3}{36} + \frac{6}{36}\right)} = \frac{1}{3}$$

$$p_{X|N}(1 \mid k) = \frac{2}{3}$$ by similar calculation.

Recall X, Y discrete R.V.s,

X and Y are independent iff

$$P_{X,Y}(x, y) = P_X(x) P_Y(y). \quad \text{for any } x, y$$

Write it into conditioning form

if $P_Y(y) > 0$,

$$\boxed{P_{X|Y}(x|y) = P_X(x)}$$

That is X and Y are independent if and only if

for any y $P(Y=y) > 0$, we have

$$P_{X|Y}(\cdot | y) = P_X.$$

In the example above,

X and N are independent

Conditional expectation

$X : \Omega \to \{x_1, x_2, \ldots\}$ $\quad P_X \quad$ p.m.f.

$$\mathbb{E}(X) = \sum_i x_i \mathbb{P}(X = x_i) = \sum_i x_i P_X(x_i).$$

- X, Y discrete random variables.

- The conditional expectation of X given $\{Y = y\}$ is defined as

$$\sum_i x_i P_{X|Y}(x_i | y)$$

$$\mathbb{E}(X \,|\, Y = y) = \Sigma_{x_i} x_i \mathbb{P}_{X|Y}(x_i \,|\, y)$$

provided that $\underline{\mathbb{P}(Y = y)} > 0.$

Condition on $Y = y$, we have conditional p.m.f.

$$P_{X|Y}(\cdot \,|\, y)$$

Recall: $\mathbb{E}[g(X)] = \sum_i g(x_i) P_X(x_i)$.

- In the same way as for expectation E(X), we can compose the random variable X with a function $g : \mathbb{R} \to \mathbb{R}$, and the conditional expectation of g(X) given $Y = y$ is

$$\mathbb{E}(g(X) \,|\, Y = y) = \Sigma_{x_i} g(x_i) \mathbb{P}_{X|Y}(x_i \,|\, y) .$$

take $P_{X|Y}(\cdot \,|\, y)$

then we can compute $\mathbb{E}[g(X) \,|\, Y = y]$

$$= \sum_i g(x_i) P_{X|Y}(x_i \,|\, y).$$

Example

Let N be a random variable taking value in {0,1,2,…}.

Let $\xi_1, \xi_2, \ldots$ be a sequence of i.i.d. random variables with mean m, independent of N.

Let $\boxed{X = \xi_1 + \ldots + \xi_N}$. (a random sum)

The conditional expectation of X given N=n is

$$\mathbb{E}[X \mid N=n] = n \cdot \mathbb{E}[\xi_1] = n \cdot m.$$

The conditional distribution of X given $N=n$ is the same as $\xi_1 + \xi_2 + \cdots + \xi_n$.

$\mathbb{E}[X \mid N=n] = \mathbb{E}[\xi_1 + \cdots + \xi_n] \overset{\text{linearity}}{=} \mathbb{E}[\xi_1] + \cdots + \mathbb{E}[\xi_n] \overset{\text{i.i.d.}}{=} n \cdot m.$

Properties of conditional expectation

- (similar to the law of total probability, a special case of the so called "tower property" of conditional expectation)

$$\mathbb{E}(g(X)) = \Sigma_{y_i} \mathbb{E}(g(X) \mid Y = y_i) \mathbb{P}(Y = y_i) \qquad ①$$

This identity is usually written in a more compact form as

$$\boxed{\mathbb{E}(g(X)) = \mathbb{E}(\mathbb{E}(g(X) \mid Y))}.$$

Why ① is true:

$$\mathbb{E}[g(X)] = \sum_{x_i} g(x_i) \underline{\mathbb{P}(X = x_i)}$$

$$= \sum_{x_i} g(x_i) \left(\sum_{y_j} \mathbb{P}(X = x_i \mid Y = y_j)\, \mathbb{P}(Y = y_j) \right)$$

$$= \sum_{y_j} \left(\sum_{x_i} g(x_i)\, P(X = x_i \mid Y = y_j) \right) P(Y = y_j)$$

$$= \sum_{y_j} \mathbb{E}[g(X) \mid Y = y_j]\, P(Y = y_j).$$

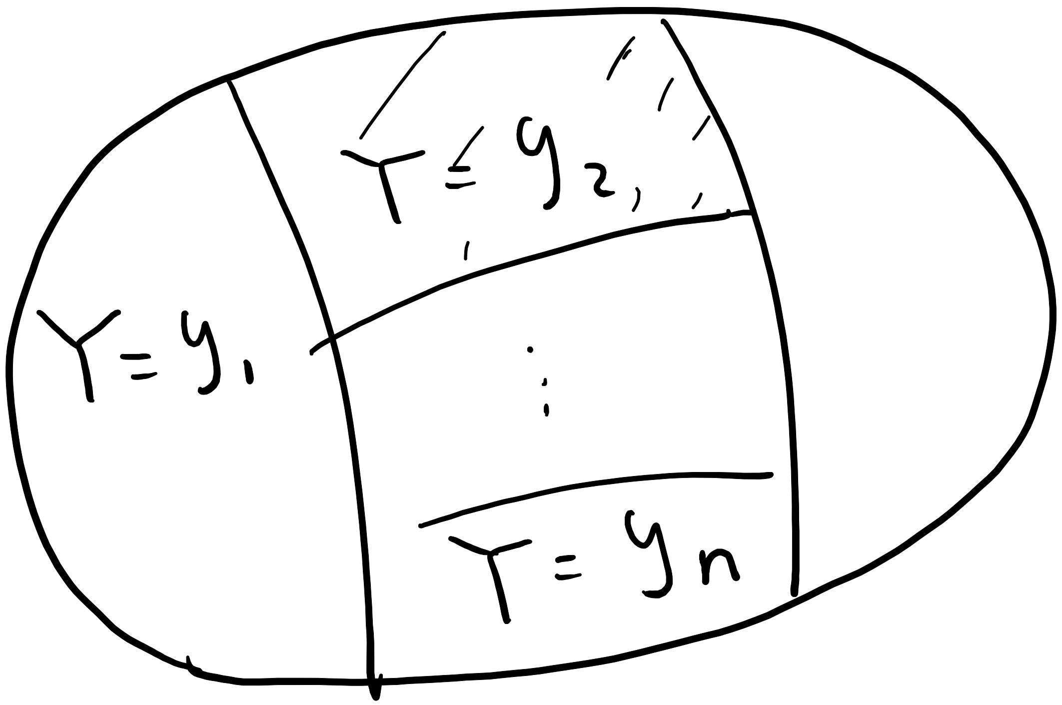
We can understand

$\mathbb{E}[g(X) \mid Y]$ as a Random Variable :

Ω sample space $\quad X : \Omega \to \mathbb{R}$

$Y : \Omega \to \mathbb{R}$

Y takes value in $\{y_1, y_2, \dots\}$



the events $\{Y = y_i\}$ form a partition of Ω

and we know what $\mathbb{E}[g(X) \mid Y = y_i]$ means.

The R.V. $\mathbb{E}[g(X) \mid Y] : \Omega \to \mathbb{R}$ is

defined as $\big(\mathbb{E}[g(X) \mid Y]\big)(\omega) = \mathbb{E}[g(X) \mid Y = Y(\omega)]$

That is, the R.V. $\mathbb{E}[g(X)|Y]$ respects the partition $\Omega = \bigcup_j \{Y = y_j\}$,

and on the piece $\{Y = y_j\}$, it takes value $\mathbb{E}[g(X)|Y = y_j]$.

Exercise: check that for the R.V. $\mathbb{E}[g(X)|Y]$ defined above,

$$\mathbb{E}\big[\mathbb{E}[g(X)|Y]\big] = \mathbb{E}[g(X)].$$

Solution to Exercise above (think about it yourself first, it's just about seeing definitions right)

Formally, let's define a function

$f : \{y_1, y_2, \dots\} \to \mathbb{R}$ by setting

$$f(y_i) = \mathbb{E}[g(X) \mid Y = y_i] \quad \text{if } \mathbb{P}(Y = y_i) > 0.$$

Then $\mathbb{E}[g(X) \mid Y] = f(Y)$ by definition.

Now
$$\mathbb{E}[f(Y)] = \sum_j f(y_j)\, \mathbb{P}(Y = y_j)$$

$$= \sum_j \mathbb{E}[g(X) \mid Y = y_j]\, \mathbb{P}(Y = y_j)$$

$$= \mathbb{E}[g(X)] \quad \text{by formula (i).}$$