Appendix.03 Probability Theory

Definition.A.3.1 Probability

1. Subjective(Empirical) Definition :

> Our measure of how much we believe something to be true.

2. Relative Frequency Definition :

> Repeating the physical process an extreamly large number of times(trials) and then to look at the fraction of times that the outcome of interest occurs.

3. Mathematical(Classical) Definition :

> The ratio of a number of interested outcomes to the of number of all possible outcomes, presuming they all equally likely.

Axiom.A.3.1 Nonnegativity

$$P(A) \geq 0 \quad \forall \ \text{ event } \ A$$

Axiom.A.3.2 Additivity

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) \quad if \quad A_i \cap A_j \neq \varnothing, \ \forall \ i \neq j$$

Axiom.A.3.3 Normalization

$$P(\Omega) = 1$$

Definition.A.3.2 Conditional Probability of A given B

$$\Rightarrow P(A \cap B) = P(A|B)P(B)$$

Thorem.A.3.1 Total Probability Theorem

$$\text{then} \quad P(B) = P(A_1 \cap B) + \cdots + P(A_n \cap B) = P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n) = P(B) = \sum_i P(B|A_i)P(A_i)$$

Proof)
Trivial ■

Theorem.A.3.2 Bayes Rule

$$\text{then} \quad P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)} \quad for\ P(B)>0$$

Proof)

$$\begin{aligned} P(A_i \cap B) &= P(A_i|B)P(B) & \cdots \quad (1) \\ &= P(B|A_i)P(A_i) & \cdots \quad (2) \end{aligned}$$

$$\begin{aligned} P(A_i|B) &= \tfrac{P(B|A_i)P(A_i)}{P(B)} & (\because\ by\ (1),\ (2)\ ) \\ &= \tfrac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)} & (\because Theorem.A.3.1) \quad ■ \end{aligned}$$

We can express the posterior probability in terms of the prior probability and likelihood.

Definition.A.3.3 Independence

A and B are independence, if and only if,

$$P(A \cap B) = P(A)P(B)$$

$$\begin{aligned} \Rightarrow \quad & P(A|B) = P(A) \quad \text{for } P(B)>0 \\ & P(B|A) = P(B) \quad \text{for } P(A)>0 \end{aligned}$$

Theorem.A.3.3 Independence and disjoint

$$\text{Independent} \quad \not\rightleftarrows \quad \text{Disjoint}$$

Proof.
$i)$ $A,\ A^c$: disjoint, but dependent.
$ii)$

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) \quad (\because\ P(A \cap B) = 0) \end{aligned} \quad ■$$

Definition.A.3.4 Discrete random variables

If the random variable $X \in S$

$$s.t. \quad S = \{x_1,\ x_2,\ \cdots\} \quad S \text{ is a countable set,}$$

then X is called discrete random variable.

Definition.A.3.5 Probability mass functions(PMFs)

In a discrete random variable X,

$$P_X(x) = P(\{X = x\})$$

is called probability mass function of random variable X.
It has

$$\textstyle\sum_x P_X(x) = 1, \qquad P(X \in S) = \sum_{x \in S} P_X(x) = 1$$

Definition.A.3.6 Expectation

Suppose X is a discrete random variable.
Expectation of X is

$$\mathbb{E}[X] = \textstyle\sum_x x p_X(x).$$

Theorem.A.3.4 Expected value rule

$$\mathbb{E}[g(X)] = \textstyle\sum_x g(x) p_X(x)$$

Proof.
Let $Y = g(X)$

$$\begin{aligned} \mathbb{E}[g(X)] &= E[Y] \\ &= \textstyle\sum_y y P_Y(y) \\ &= \textstyle\sum_y \sum_{\{x\ :\ g(x)\ =\ y\}} y p_X(x) \\ &= \textstyle\sum_y \sum_{\{x\ :\ g(x)\ =\ y\}} g(x) p_X(x) \\ &= \textstyle\sum_x g(x) p_X(x) \qquad \blacksquare \end{aligned}$$

Definition.A.3.7 Variance, standard deviation and nth moment

Suppose X is a discrete random variable.
Variance of X is

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

Standard deviation of X is

$$\sigma_X = \sqrt{Var(X)}.$$

nth moment of X is

$$\mathbb{E}[X^n] = \sum_x x^n p_X(x).$$

Theorem.A.3.5 Variance rule

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

Proof.
Trivial.

Theorem.A.3.6 Properties of mean and variance

Let $Y = aX + b$

$$\begin{aligned}\mathbb{E}[Y] &= \sum_x (ax+b)p_X(x) \\ &= a\sum_x x p_X(x) + b\sum_x p_X(x) \\ &= a\mathbb{E}[X] + b\end{aligned}$$

$$\begin{aligned}Var[Y] &= \sum_x (ax + b - \mathbb{E}[aX+b])^2 p_X(x) \\ &= \sum_x (ax + b - a\mathbb{E}[X] - b)^2 p_X(x) \\ &= a^2\sum_x (x - \mathbb{E}[X])^2 p_X(x) \\ &= a^2 Var(X) \qquad \text{which is called shift-invariance.} \qquad \blacksquare\end{aligned}$$

Definition.A.3.8 Joint PMFs of multiple random variables

Suppose X and Y are discrete random variables.

$$P_{X,\ Y}(x,\ y) = P(X = x, Y = y)$$

Theorem.A.3.7 Marginal PMF of multiple random variables

Suppose X and Y are discrete random variables.

$$p_X(x) = \sum_y p_{X,\ Y}(x,\ y), \quad p_Y(y) = \sum_x p_{X,\ Y}(x,\ y)$$

Proof.

$$p_X(x) = P(\{X = x \cap Y = y_1\} \cup \{X = x \cap Y = y_2\} \cup \cdots \{X = x \cap Y = y_k\})$$

$$p_Y(y) = P(\{X = x_1 \cap Y = y\} \cup \{X = x_2 \cap Y = y\} \cup \cdots \{X = x_j \cap Y = y\})$$

■

Definition.A.3.9 Conditional PMF given an event and random variable

With given event A,

$$p_{X|A}(x) = P(X = x|A) = \frac{P(\{X = x\} \cap A)}{P(A)}.$$

With given a random variable Y,

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x,\ Y = y)}{P(Y = y)} = \frac{p_{X,\ Y}(x,\ y)}{P_Y(y)}$$

- Normalization :

$$\sum_x p_{X|Y}(x|y) = \frac{1}{P_Y(y)} \sum_x P_{X,\ Y}(x,\ y) = \frac{P_Y(y)}{P_Y(y)} = 1$$

- Joint PMF :

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

$$p_{X,\ Y,\ Z}(x,\ y,\ z) = p_X(x) p_{Y|X}(y|x) p_{Z|X,\ Y}(z|x,\ y)$$

- Marginal PMF

$$p_x(x) = \sum_y p_{X|Y}(x|y) p_Y(y)$$

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