2.Probability Theory

- Conditional and Unconditional Probability

When we observe the probability that an event (A) occurs without taking into account whether it is necessarily preceded by any other specific events, it is known as an **unconditional probability**.

- Also known as a **marginal probability**.
- Notation: $\mathbb{P}(A)$

When we observe the probability of a given event after taking into account that another event has already occurred (A occurs given that B has occurred), it is known as a **conditional probability**.

- Notation: $\mathbb{P}(A|B)$ "probability of A given B"
- Example: Event A: *the stock return of the company ABC falls below* -10%
  Event B: *the market return falls below* -10%
  $A|B$: *the stock return of the company ABC falls below* -10% *given that the market return falls below* -10%

Marginal Probability - probability of any single event occurring unconditioned on any other events. Whenever someone asks you whether the weather is going to be rainy or sunny today, you are computing a marginal probability.

Joint Probability - probability of more than one event occurring simultaneously. If I ask you whether the weather is going to be rainy and the temperature is going to be above a certain number, you are computing a joint probability.

Conditional probability - probability of an event occurring given some events that you have already observed. When I ask you what's the probability that today is rainy or sunny given that I noticed the temperature is going to be above 80 degrees, you are computing a conditional probability.

Conditional probability:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \ \mathbb{P}(B) \neq 0$$

Multiplication rule: the joint probability of A and B is

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)}$$

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$$

- Independent Events

When events are **independent**, the occurrence of one does not affect the probability of the other. In other words, $\mathbb{P}(A|B) = \mathbb{P}(A)$ and $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Then **the multiplication rule**: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

More generally, if A_1, A_2,..., A_n are independent, then

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2)\ldots\mathbb{P}(A_n)$$

[E(A)*E(B) = E(A*B)]

- Bayes' Rule

Bayes' Rule is used to determine how a subjective belief should change given new evidence.

$$\mathbb{P}(Event|Information) = \left(\frac{\mathbb{P}(Information|Event)}{\mathbb{P}(Information)}\right)\mathbb{P}(Event)$$

- Updated probability given new information (posterior probability)
- Ratio of probability of the new information given the event and unconditional probability of the new information
- Prior probability of the event

Bayes' Rule:

$$\mathbb{P}(E|A) = \frac{\mathbb{P}(A|E)\mathbb{P}(E)}{\mathbb{P}(A|E)\mathbb{P}(E) + \mathbb{P}(A|E^c)\mathbb{P}(E^c)}$$

More generally,

$$\mathbb{P}(E_j|A) = \frac{\mathbb{P}(A|E_j)\mathbb{P}(E_j)}{\underbrace{\sum_{i=1}^{n} \mathbb{P}(A|E_i)\mathbb{P}(E_i)}_{}} \quad = P(A)$$

where $E_1, E_2, \ldots, E_n$ are mutually exclusive and exhaustive events.

- Covariance and Correlation

Covariance and correlation are measures of **linear dependence** between random variables.

Covariance of two random variables X and Y is

$$Cov(X,Y) = \sigma_{X,Y} = \mathbb{E}\left([X-\mathbb{E}(X)][Y-\mathbb{E}(Y)]\right)$$

Correlation is covariance scaled by standard deviations of random variables:

$$\rho(X,Y) = \rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{\sigma^2(X)\sigma^2(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

- $-1 \le \rho_{X,Y} \le 1$

- Properties of Expected Value and Variance

$\mathbb{E}(aX+bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$, where a and b are deterministic.

$\mathbb{E}(aX+c) = a\mathbb{E}(X) + c$, where a and c are deterministic.

$\sigma^2(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$

$\sigma^2(aX+bY) = a^2\sigma^2(X) + b^2\sigma^2(Y) + 2ab\, Cov(X,Y)$, where a and b are deterministic.

$\sigma^2(aX+c) = a^2\sigma^2(X)$, where a and c are deterministic.

$Cov(X,Y) = Cov(Y,X)$, $\rho(X,Y) = \rho(Y,X)$,
$Cov(X,X) = \sigma^2(X)$

$Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

If $X_1, X_2, \ldots, X_n$ are independent, then
$\sigma^2\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \sigma^2(X_i)$

- Conditional Expectations

Conditional distribution that $X = x_i$ given that $Y = y_i$ is described by a set of conditional probabilities:

$$\mathbb{P}(X = x_i | Y = y_i) = \frac{\mathbb{P}(X = x_i, Y = y_i)}{\mathbb{P}(Y = y_i)}$$

Notation: $p_{X|Y}(x|y)$

The **conditional expectation** of X given $Y = y$ is

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

The Total Probability Rule for Expected Value

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(X|E)\mathbb{P}(E) + \mathbb{E}(X|E^C)\mathbb{P}(E^C) \\ &= \mathbb{E}(X|E_1)\mathbb{P}(E_1) + \mathbb{E}(X|E_2)\mathbb{P}(E_2) + \cdots + \mathbb{E}(X|E_n)\mathbb{P}(E_n), \end{aligned}$$

where $E_1, E_2, \ldots, E_n$ are mutually exclusive and exhaustive events.

- Prediction

Predict Y by means of a constant value c

Minimize the **mean squared error** $MSE = \mathbb{E}\left((Y - c)^2\right)$

Optimal $c = \mathbb{E}(Y)$

Predict Y by some function of $h(X)$, that is minimize $MSE = \mathbb{E}\left((Y - h(X))^2\right)$

Optimal in the mean squared sense $h(X) = \mathbb{E}(Y|X)$

If $h(X)$ is a linear function, the conditional expectation represent the linear regression of Y on X.