3 Mathematical Expectation

3.1 Univariate Random Variable

Definition of Mathematical Expectation (数学的期待値):

Let $g(X)$ be a function of random variable X. The mathematical expectation of $g(X)$, denoted by $\mathrm{E}(g(X))$, is defined as follows:

$$\mathrm{E}(g(X)) = \begin{cases} \displaystyle\sum_i g(x_i)p_i = \sum_i g(x_i)f(x_i), & \text{(Discrete RV)}, \\ \displaystyle\int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{d}x, & \text{(Continuous RV)}. \end{cases}$$

The following three functional forms of $g(X)$ are important.

1. $\boldsymbol{g(X) = X}$.

The expectation of X, $\mathrm{E}(X)$, is known as **mean (平均)** of random variable X.

$$\mathrm{E}(X) = \begin{cases} \displaystyle\sum_i x_i f(x_i), & \text{(Discrete RV)}, \\ \displaystyle\int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x, & \text{(Continuous RV)}, \end{cases}$$
$$= \mu, \quad (\text{or } \mu_x).$$

When a distribution of X is symmetric, mean indicates the center of the distribution.

2. $\boldsymbol{g(X) = (X - \mu)^2}$.

The expectation of $(X - \mu)^2$ is known as **variance (分散)** of random variable X, which is denoted by $\mathrm{V}(X)$.

$$\begin{aligned} \mathrm{V}(X) &= \mathrm{E}((X-\mu)^2) \\ &= \begin{cases} \displaystyle\sum_i (x_i - \mu)^2 f(x_i), & \text{(Discrete RV)}, \\ \displaystyle\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,\mathrm{d}x, & \text{(Continuous RV)}, \end{cases} \\ &= \sigma^2, \quad (\text{or } \sigma_x^2). \end{aligned}$$

If X is broadly distributed, $\sigma^2 = \mathrm{V}(X)$ becomes large. Conversely, if the distribution is concentrated on the center, σ^2 becomes small. Note that $\sigma = \sqrt{V(X)}$ is called the **standard deviation (標準偏差)**.

3. $\boldsymbol{g(X) = e^{\theta X}}$.

The expectation of $e^{\theta X}$ is called the **moment-generating function (積率母関数)**, which is denoted by $\phi(\theta)$.

$$\begin{aligned} \phi(\theta) &= \mathrm{E}(e^{\theta X}) \\ &= \begin{cases} \displaystyle\sum_i e^{\theta x_i} f(x_i), & \text{(Discrete RV)}, \\ \displaystyle\int_{-\infty}^{\infty} e^{\theta x} f(x)\,\mathrm{d}x, & \text{(Continuous RV)}. \end{cases} \end{aligned}$$

Some Formulas of Mean and Variance:

1. **Theorem:** $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$, where a and b are constant.

   Proof:

   When X is a discrete random variable,

$$\begin{aligned}\mathrm{E}(aX + b) &= \sum_i (ax_i + b) f(x_i) \\ &= a \sum_i x_i f(x_i) + b \sum_i f(x_i) \\ &= a\mathrm{E}(X) + b.\end{aligned}$$

Note that we have $\sum_i x_i f(x_i) = \mathrm{E}(X)$ from the definition of mean and $\sum_i f(x_i) = 1$ because $f(x_i)$ is a probability function.

If X is a continuous random variable,

$$\begin{aligned}\mathrm{E}(aX + b) &= \int_{-\infty}^{\infty} (ax + b) f(x)\,\mathrm{d}x \\ &= a \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x + b \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \\ &= a\mathrm{E}(X) + b.\end{aligned}$$

Similarly, note that we have $\int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \mathrm{E}(X)$ from

the definition of mean and $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$ because $f(x)$ is a probability density function.

2. **Theorem:** $\mathrm{V}(X) = \mathrm{E}(X^2) - \mu^2$, where $\mu = \mathrm{E}(X)$.

   Proof:

   $\mathrm{V}(X)$ is rewritten as follows:

$$\begin{aligned}\mathrm{V}(X) &= \mathrm{E}((X-\mu)^2) = \mathrm{E}(X^2 - 2\mu X - \mu^2) \\ &= \mathrm{E}(X^2) - 2\mu\mathrm{E}(X) + \mu^2 = \mathrm{E}(X^2) - \mu^2.\end{aligned}$$

   The first equality is due to the definition of variance.

3. **Theorem:** $\mathrm{V}(aX + b) = a^2\mathrm{V}(X)$, where a and b are constant.

   Proof:

   From the definition of the mathematical expectation, $\mathrm{V}(aX + b)$ is represented as:

$$\begin{aligned}\mathrm{V}(aX + b) &= \mathrm{E}\Big(((aX+b) - \mathrm{E}(aX+b))^2\Big) \\ &= \mathrm{E}\Big((aX - a\mu)^2\Big) = \mathrm{E}(a^2(X-\mu)^2) \\ &= a^2\mathrm{E}((X-\mu)^2) = a^2\mathrm{V}(X)\end{aligned}$$

   The first and the fifth equalities are from the definition of variance. We use $\mathrm{E}(aX + b) = a\mu + b$ in the second equality.

4. **Theorem:** The random variable X is assumed to be distributed with mean $\mathrm{E}(X) = \mu$ and variance $\mathrm{V}(X) = \sigma^2$. Define $Z = \dfrac{X-\mu}{\sigma}$. Then, we have $\mathrm{E}(Z) = 0$ and $\mathrm{V}(Z) = 1$.

   Proof:

   $\mathrm{E}(X)$ and $\mathrm{V}(X)$ are obtained as:

$$\mathrm{E}(Z) = \mathrm{E}\Big(\frac{X-\mu}{\sigma}\Big) = \frac{\mathrm{E}(X)-\mu}{\sigma} = 0,$$

$$\mathrm{V}(Z) = \mathrm{V}\Big(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\Big) = \frac{1}{\sigma^2}\mathrm{V}(X) = 1.$$

   The transformation from X to Z is known as normalization (正規化) or standardization (標準化).

Example 1.5: In Example 1.2 of flipping a coin three times (Section 1.1), we see in Section 2.1 that the probability function is written as the following binomial distribution:

$$P(X = x) = f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x},$$

$$\text{for } x = 0, 1, 2, \cdots, n,$$

where $n = 3$ and $p = 1/2$.
When X has the binomial distribution above, we obtain $\mathrm{E}(X)$, $\mathrm{V}(X)$ and $\phi(\theta)$ as follows.

First, $\mu = \mathrm{E}(X)$ is computed as:

$$\begin{aligned}\mu = \mathrm{E}(X) &= \sum_{x=0}^{n} x f(x) = \sum_{x=1}^{n} x f(x) = \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \\ &= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} \\ &= np \sum_{x'=0}^{n'} \frac{n'!}{x'!(n'-x')!} p^{x'} (1-p)^{n'-x'} = np,\end{aligned}$$

where $n' = n - 1$ and $x' = x - 1$ are set.

Second, in order to obtain $\sigma^2 = \mathrm{V}(X)$, we rewrite $\mathrm{V}(X)$ as:

$$\sigma^2 = \mathrm{V}(X) = \mathrm{E}(X^2) - \mu^2 = \mathrm{E}(X(X-1)) + \mu - \mu^2.$$

$\mathrm{E}(X(X-1))$ is given by:

$$\begin{aligned}\mathrm{E}(X(X-1)) &= \sum_{x=0}^{n} x(x-1) f(x) = \sum_{x=2}^{n} x(x-1) f(x) \\ &= \sum_{x=2}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} p^x (1-p)^{n-x}\end{aligned}$$

$$\begin{aligned}&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} (1-p)^{n-x} \\ &= n(n-1)p^2 \sum_{x'=0}^{n'} \frac{n'!}{x'!(n'-x')!} p^{x'} (1-p)^{n'-x'} \\ &= n(n-1)p^2,\end{aligned}$$

where $n' = n - 2$ and $x' = x - 2$ are re-defined.

Therefore, $\sigma^2 = \mathrm{V}(X)$ is obtained as:

$$\begin{aligned}\sigma^2 = \mathrm{V}(X) &= \mathrm{E}(X(X-1)) + \mu - \mu^2 \\ &= n(n-1)p^2 + np - n^2p^2 = -np^2 + np = np(1-p).\end{aligned}$$

Finally, the moment-generating function $\phi(\theta)$ is represented as:

$$\begin{aligned}\phi(\theta) = \mathrm{E}(e^{\theta X}) &= \sum_{x=0}^{n} e^{\theta x} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-p} \\ &= \sum_{x=0}^{n} \frac{n!}{x!(n-x)!} (pe^{\theta})^x (1-p)^{n-p} = (pe^{\theta} + 1 - p)^n.\end{aligned}$$

In the last equality, we utilize the following formula:

$$(a+b)^n = \sum_{x=0}^{n} \frac{n!}{x!(n-x)!} a^x b^{n-x},$$

which is called the **binomial theorem**.

Example 1.6: As an example of continuous random variables, in Section 2.1 the uniform distribution is introduced, which is given by:

$$f(x) = \begin{cases} 1, & \text{for } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

When X has the uniform distribution above, $\mathrm{E}(X)$, $\mathrm{V}(X)$ and $\phi(\theta)$ are computed as follows:

$$\mu = \mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \int_0^1 x\,\mathrm{d}x = \Big[\frac{1}{2}x^2\Big]_0^1 = \frac{1}{2},$$

$$\sigma^2 = \mathrm{V}(X) = \mathrm{E}(X^2) - \mu^2 = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x - \mu^2$$
$$= \int_0^1 x^2\,\mathrm{d}x - \mu^2 = \Big[\frac{1}{3}x^3\Big]_0^1 - \Big(\frac{1}{2}\Big)^2 = \frac{1}{12},$$

$$\phi(\theta) = \mathrm{E}(e^{\theta X}) = \int_{-\infty}^{\infty} e^{\theta x} f(x)\,\mathrm{d}x = \int_0^1 e^{\theta x}\,\mathrm{d}x$$
$$= [\frac{1}{\theta}e^{\theta x}]_0^1 = \frac{1}{\theta}(e^{\theta} - 1).$$

Example 1.7: As another example of continuous random variables, we take the standard normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad \text{for } -\infty < x < \infty$$

When X has a standard normal distribution, i.e., when $X \sim N(0,1)$, $\mathrm{E}(X)$, $\mathrm{V}(X)$ and $\phi(\theta)$ are as follows.

$\mathrm{E}(X)$ is obtained as:

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}x^2}\,\mathrm{d}x$$
$$= \frac{1}{\sqrt{2\pi}} \Big[-e^{-\frac{1}{2}x^2}\Big]_{-\infty}^{\infty} = 0,$$

because $\lim_{x\to\pm\infty} -e^{-\frac{1}{2}x^2} = 0$.

$\mathrm{V}(X)$ is computed as follows:

$$\mathrm{V}(X) = \mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x$$
$$= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \frac{\mathrm{d}(-e^{-\frac{1}{2}x^2})}{\mathrm{d}x}\,\mathrm{d}x$$
$$= \frac{1}{\sqrt{2\pi}} \Big[x(-e^{-\frac{1}{2}x^2})\Big]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,\mathrm{d}x$$
$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x = 1.$$

The first equality holds because of $\mathrm{E}(X) = 0$.

In the fifth equality, use the following integration formula, called the **integration by parts**:

$$\int_a^b h(x)g'(x)\,\mathrm{d}x = \Big[h(x)g(x)\Big]_a^b - \int_a^b h'(x)g(x)\,\mathrm{d}x,$$

where we take $h(x) = x$ and $g(x) = -e^{-\frac{1}{2}x^2}$ in this case.

In the sixth equality, $\lim_{x\to\pm\infty} -xe^{-\frac{1}{2}x^2} = 0$ is utilized.

The last equality is because the integration of the standard normal probability density function is equal to one.

$\phi(\theta)$ is derived as follows:

$$\begin{aligned}\phi(\theta) &= \int_{-\infty}^{\infty} e^{\theta x} f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} e^{\theta x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2+\theta x}\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left((x-\theta)^2-\theta^2\right)}\,\mathrm{d}x \\ &= e^{\frac{1}{2}\theta^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}\,\mathrm{d}x = e^{\frac{1}{2}\theta^2}.\end{aligned}$$

The last equality holds because the integration indicates the normal density with mean θ and variance one.

Example 1.8: When the moment-generating function of X is given by $\phi_x(\theta) = e^{\frac{1}{2}\theta^2}$ (i.e., X has a standard normal distribution), we want to obtain the moment-generating function of $Y = \mu + \sigma X$.

Let $\phi_x(\theta)$ and $\phi_y(\theta)$ be the moment-generating functions of X and Y, respectively. Then, the moment-generating function of Y is obtained as follows:

$$\begin{aligned}\phi_y(\theta) &= \mathrm{E}(e^{\theta Y}) = \mathrm{E}(e^{\theta(\mu+\sigma X)}) = e^{\theta\mu}\mathrm{E}(e^{\theta\sigma X}) = e^{\theta\mu}\phi_x(\theta\sigma) \\ &= e^{\theta\mu} e^{\frac{1}{2}\sigma^2\theta^2} = \exp(\mu\theta + \frac{1}{2}\sigma^2\theta^2).\end{aligned}$$

Example 1.8(b): When $X \sim N(\mu, \sigma^2)$, what is the moment-generating function of X?

$$\begin{aligned}\phi(\theta) &= \int_{-\infty}^{\infty} e^{\theta x} f(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} e^{\theta x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(\theta x - \frac{1}{2\sigma^2}(x-\mu)^2\Big)\,\mathrm{d}x \\ &= \exp(\mu\theta + \frac{1}{2}\sigma^2\theta^2) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(\frac{1}{2\sigma^2}(x-\mu-\sigma^2\theta)^2\Big)\,\mathrm{d}x \\ &= \exp(\mu\theta + \frac{1}{2}\sigma^2\theta^2).\end{aligned}$$

3.2 Bivariate Random Variable

Definition: Let $g(X, Y)$ be a function of random variables X and Y. The mathematical expectation of $g(X, Y)$, denoted by $\mathrm{E}(g(X, Y))$, is defined as:

$$\mathrm{E}(g(X,Y)) = \begin{cases} \displaystyle\sum_i \sum_j g(x_i, y_j) f(x_i, y_j), & \text{(Discrete)}, \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\,\mathrm{d}x\,\mathrm{d}y, & \text{(Continuous)}. \end{cases}$$

The following four functional forms are important, i.e., mean, variance, covariance and the moment-generating function.

1. **$g(X, Y) = X$:**

   The expectation of random variable X, i.e., $\mathrm{E}(X)$, is given by:

$$\begin{aligned}\mathrm{E}(X) &= \begin{cases} \displaystyle\sum_i \sum_j x_i f(x_i, y_j), & \text{(Discrete)}, \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,\mathrm{d}x\,\mathrm{d}y, & \text{(Continuous)}, \end{cases} \\ &= \mu_x.\end{aligned}$$

   The case of $g(X, Y) = Y$ is exactly the same formulation as above, i.e., $\mathrm{E}(Y) = \mu_y$.

2. **$g(X, Y) = (X - \mu_x)^2$:**

   The expectation of $(X - \mu_x)^2$ is known as variance of X.

$$\begin{aligned}\mathrm{V}(X) &= \mathrm{E}((X-\mu_x)^2) \\ &= \begin{cases} \displaystyle\sum_i \sum_j (x_i-\mu_x)^2 f(x_i, y_j), & \text{(Discrete)} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)^2 f(x,y)\,\mathrm{d}x\,\mathrm{d}y, & \text{(Continuous)} \end{cases} \\ &= \sigma_x^2.\end{aligned}$$

   The variance of Y is also obtained in the same way, i.e., $\mathrm{V}(Y) = \sigma_y^2$.

3. $\boldsymbol{g(X,Y) = (X-\mu_x)(Y-\mu_y)}$:

The expectation of $(X-\mu_x)(Y-\mu_y)$ is known as **covariance** of X and Y, which is denoted by $\mathrm{Cov}(X,Y)$ and written as:

$$\mathrm{Cov}(X,Y) = \mathrm{E}((X-\mu_x)(Y-\mu_y))$$
$$= \begin{cases} \displaystyle\sum_i \sum_j (x_i-\mu_x)(y_j-\mu_y) f(x_i, y_j), & \text{(Discrete)}, \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y) f(x,y)\,\mathrm{d}x\,\mathrm{d}y, & \text{(Continuous)}. \end{cases}$$

Thus, covariance is defined in the case of bivariate random variables.

4. $\boldsymbol{g(X,Y) = e^{\theta_1 X + \theta_2 Y}}$:

The mathematical expectation of $e^{\theta_1 X + \theta_2 Y}$ is called the moment-generating function, which is denoted by:

$$\phi(\theta_1,\theta_2) = \mathrm{E}(e^{\theta_1 X + \theta_2 Y})$$
$$= \begin{cases} \displaystyle\sum_i \sum_j e^{\theta_1 x_i + \theta_2 y_j} f(x_i, y_j), & \text{(Discrete)} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\theta_1 x + \theta_2 y} f(x,y)\,\mathrm{d}x\,\mathrm{d}y, & \text{(Continuous)} \end{cases}$$

In Section 5, the moment-generating function in the multivariate cases is discussed in more detail.