3. **Theorem:** Let $\phi(\theta_1, \theta_2)$ be the moment-generating function of (X, Y).

The moment-generating function of X is given by $\phi_1(\theta_1)$ and that of Y is $\phi_2(\theta_2)$.

Then, we have the following facts:

$$\phi_1(\theta_1) = \phi(\theta_1, 0), \quad \phi_2(\theta_2) = \phi(0, \theta_2).$$

Proof:

Again, the definition of the moment-generating function of X and Y is represented as:

$$\phi(\theta_1, \theta_2) = \mathrm{E}(e^{\theta_1 X + \theta_2 Y}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\theta_1 x + \theta_2 y} f_{xy}(x, y)\, \mathrm{d}x\, \mathrm{d}y.$$

When $\phi(\theta_1, \theta_2)$ is evaluated at $\theta_2 = 0$, $\phi(\theta_1, 0)$ is rewritten as follows:

$$\begin{aligned}\phi(\theta_1, 0) = \mathrm{E}(e^{\theta_1 X}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\theta_1 x} f_{xy}(x, y)\, \mathrm{d}x\, \mathrm{d}y \\ &= \int_{-\infty}^{\infty} e^{\theta_1 x}\Big(\int_{-\infty}^{\infty} f_{xy}(x, y)\, \mathrm{d}y\Big)\, \mathrm{d}x\end{aligned}$$

$$= \int_{-\infty}^{\infty} e^{\theta_1 x} f_x(x)\, \mathrm{d}x = \mathrm{E}(e^{\theta_1 X}) = \phi_1(\theta_1).$$

Thus, we obtain the result: $\phi(\theta_1, 0) = \phi_1(\theta_1)$.

Similarly, $\phi(0, \theta_2) = \phi_2(\theta_2)$ can be derived.

4. **Theorem:** The moment-generating function of (X, Y) is given by $\phi(\theta_1, \theta_2)$.

Let $\phi_1(\theta_1)$ and $\phi_2(\theta_2)$ be the moment-generating functions of X and Y, respectively.

If X is independent of Y, we have:

$$\phi(\theta_1, \theta_2) = \phi_1(\theta_1)\phi_2(\theta_2).$$

Proof:

From the definition of $\phi(\theta_1, \theta_2)$, the moment-generating function of X and Y is rewritten as follows:

$$\phi(\theta_1, \theta_2) = \mathrm{E}(e^{\theta_1 X + \theta_2 Y}) = \mathrm{E}(e^{\theta_1 X})\mathrm{E}(e^{\theta_2 Y}) = \phi_1(\theta_1)\phi_2(\theta_2).$$

The second equality holds because X is independent of Y.

Multivariate Case: For multivariate random variables X_1, $X_2, \cdots, X_n$, the moment-generating function is defined as:

$$\phi(\theta_1, \theta_2, \cdots, \theta_n) = \mathrm{E}(e^{\theta_1 X_1 + \theta_2 X_2 + \cdots + \theta_n X_n}).$$

1. **Theorem:** If the multivariate random variables X_1, X_2, $\cdots$, X_n are mutually independent,

the moment-generating function of X_1, X_2, $\cdots$, X_n, denoted by $\phi(\theta_1, \theta_2, \cdots, \theta_n)$, is given by:

$$\phi(\theta_1, \theta_2, \cdots, \theta_n) = \phi_1(\theta_1)\phi_2(\theta_2)\cdots\phi_n(\theta_n),$$

where $\phi_i(\theta) = \mathrm{E}(e^{\theta X_i})$.

Proof:

From the definition of the moment-generating function in the multivariate cases, we obtain the following:

$$\begin{aligned}\phi(\theta_1, \theta_2, \cdots, \theta_n) &= \mathrm{E}(e^{\theta_1 X_1 + \theta_2 X_2 + \cdots + \theta_n X_n}) \\ &= \mathrm{E}(e^{\theta_1 X_1})\mathrm{E}(e^{\theta_2 X_2})\cdots\mathrm{E}(e^{\theta_n X_n}) \\ &= \phi_1(\theta_1)\phi_2(\theta_2)\cdots\phi_n(\theta_n).\end{aligned}$$

2. **Theorem:** Suppose that the multivariate random variables X_1, X_2, $\cdots$, X_n are mutually independently and identically distributed.

Suppose that $X_i \sim N(\mu, \sigma^2)$.

Let us define $\hat{\mu} = \sum_{i=1}^n a_i X_i$, where a_i, $i = 1, 2, \cdots, n$, are assumed to be known.

Then, $\hat{\mu} \sim N(\mu \sum_{i=1}^n a_i, \sigma^2 \sum_{i=1}^n a_i^2)$.

Proof:

From Example 1.8 (p.111) and Example 1.9 (p.147), it is shown that the moment-generating function of X is given by: $\phi_x(\theta) = \exp(\mu\theta + \frac{1}{2}\sigma^2\theta^2)$, when X is normally distributed as $X \sim N(\mu, \sigma^2)$.

Let $\phi_{\hat{\mu}}$ be the moment-generating function of $\hat{\mu}$.

$$\begin{aligned}\phi_{\hat{\mu}}(\theta) &= \mathrm{E}(e^{\theta\hat{\mu}}) = \mathrm{E}(e^{\theta \sum_{i=1}^n a_i X_i}) = \prod_{i=1}^n \mathrm{E}(e^{\theta a_i X_i}) \\ &= \prod_{i=1}^n \phi_x(a_i\theta) = \prod_{i=1}^n \exp(\mu a_i \theta + \frac{1}{2}\sigma^2 a_i^2 \theta^2) \\ &= \exp(\mu \sum_{i=1}^n a_i \theta + \frac{1}{2}\sigma^2 \sum_{i=1}^n a_i^2 \theta^2)\end{aligned}$$

which is equivalent to the moment-generating function of the normal distribution with mean $\mu \sum_{i=1}^n a_i$ and variance $\sigma^2 \sum_{i=1}^n a_i^2$, where μ and σ^2 in $\phi_x(\theta)$ is simply

replaced by $\mu \sum_{i=1}^n a_i$ and $\sigma^2 \sum_{i=1}^n a_i^2$ in $\phi_{\hat{\mu}}(\theta)$, respectively.

Moreover, note as follows.

When $a_i = 1/n$ is taken for all $i = 1, 2, \cdots, n$, i.e., when $\hat{\mu} = \overline{X}$ is taken, $\hat{\mu} = \overline{X}$ is normally distributed as: $\overline{X} \sim N(\mu, \sigma^2/n)$.

6 Law of Large Numbers (対数の法則) and Central Limit Theorem (中心極限定理)

6.1 Chebyshev's Inequality (チェビシェフの不等式)

Theorem: Let $g(X)$ be a nonnegative function of the random variable X, i.e., $g(X) \geq 0$.

If $\mathrm{E}(g(X))$ exists, then we have:

$$P(g(X) \geq k) \leq \frac{\mathrm{E}(g(X))}{k}, \tag{6}$$

for a positive constant value k.

Proof:

We define the discrete random variable U as follows:

$$U = \begin{cases} 1, & \text{if } g(X) \geq k, \\ 0, & \text{if } g(X) < k. \end{cases}$$

Thus, the discrete random variable U takes 0 or 1.

Suppose that the probability function of U is given by:

$$f(u) = P(U = u),$$

where $P(U = u)$ is represented as:

$$P(U = 1) = P(g(X) \geq k),$$

$$P(U = 0) = P(g(X) < k).$$

Then, in spite of the value which U takes, the following equation always holds:

$$g(X) \geq kU,$$

which implies that we have $g(X) \geq k$ when $U = 1$ and $g(X) \geq 0$ when $U = 0$, where k is a positive constant value.

Therefore, taking the expectation on both sides, we obtain:

$$\mathrm{E}(g(X)) \geq k\mathrm{E}(U), \tag{7}$$

where $\mathrm{E}(U)$ is given by:

$$\begin{aligned} \mathrm{E}(U) &= \sum_{u=0}^{1} uP(U = u) = 1 \times P(U = 1) + 0 \times P(U = 0) \\ &= P(U = 1) = P(g(X) \geq k). \end{aligned} \tag{8}$$

Accordingly, substituting equation (8) into equation (7), we have the following inequality:

$$P(g(X) \geq k) \leq \frac{\mathrm{E}(g(X))}{k}.$$

Chebyshev's Inequality: Assume that $\mathrm{E}(X) = \mu$, $\mathrm{V}(X) = \sigma^2$, and λ is a positive constant value. Then, we have the following inequality:

$$P(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2},$$

or equivalently,

$$P(|X - \mu| < \lambda\sigma) \geq 1 - \frac{1}{\lambda^2},$$

which is called **Chebyshev's inequality**.

Proof:

Take $g(X) = (X-\mu)^2$ and $k = \lambda^2\sigma^2$. Then, we have:

$$P((X-\mu)^2 \geq \lambda^2\sigma^2) \leq \frac{\mathrm{E}(X-\mu)^2}{\lambda^2\sigma^2},$$

which implies $P(|X-\mu| \geq \lambda\sigma) \leq \dfrac{1}{\lambda^2}$.

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

Since we have $P(|X-\mu| \geq \lambda\sigma) + P(|X-\mu| < \lambda\sigma) = 1$, we can derive the following inequality:

$$P(|X-\mu| < \lambda\sigma) \geq 1 - \frac{1}{\lambda^2}. \qquad (9)$$

An Interpretation of Chebyshev's inequality: $1/\lambda^2$ is an upper bound for the probability $P(|X-\mu| \geq \lambda\sigma)$.

Equation (9) is rewritten as:

$$P(\mu - \lambda\sigma < X < \mu + \lambda\sigma) \geq 1 - \frac{1}{\lambda^2}.$$

That is, the probability that X falls within $\lambda\sigma$ units of μ is greater than or equal to $1 - 1/\lambda^2$.

Taking an example of $\lambda = 2$, the probability that X falls within two standard deviations of its mean is at least 0.75.

Furthermore, note as follows.

Taking $\epsilon = \lambda\sigma$, we obtain as follows:

$$P(|X-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2},$$

i.e.,

$$P(|X-\mathrm{E}(X)| \geq \epsilon) \leq \frac{\mathrm{V}(X)}{\epsilon^2}, \qquad (10)$$

which inequality is used in the next section.

Remark: Equation (10) can be derived when we take $g(X) = (X-\mu)^2$, $\mu = \mathrm{E}(X)$ and $k = \epsilon^2$ in equation (6).

Even when we have $\mu \neq \mathrm{E}(X)$, the following inequality still hold:

$$P(|X-\mu| \geq \epsilon) \leq \frac{\mathrm{E}((X-\mu)^2)}{\epsilon^2}.$$

Note that $\mathrm{E}((X-\mu)^2)$ represents the mean square error (MSE).

When $\mu = \mathrm{E}(X)$, the mean square error reduces to the variance.

6.2 Law of Large Numbers (対数の法則) and Convergence in Probability (確率収束)

Law of Large Numbers 1: Assume that $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed with mean $\mathrm{E}(X_i) = \mu$ for all i.

Supopose that the moment-generating function of X_i is finite.

Define $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

Then, $\overline{X}_n \longrightarrow \mu$ as $n \longrightarrow \infty$.

Proof: The moment-generating function is written as:

$$\begin{aligned}\phi(\theta) &= 1 + \mu_1'\theta + \frac{1}{2!}\mu_2'\theta^2 + \frac{1}{3!}\mu_3'\theta^3 + \cdots \\ &= 1 + \mu_1'\theta + O(\theta^2)\end{aligned}$$

where $\mu_k' = \mathrm{E}(X^k)$ for all k. That is, all the moments exist.

$$\begin{aligned}\phi_{\overline{x}}(\theta) &= \Big(\phi\big(\frac{\theta}{n}\big)\Big)^n = \Big(1 + \mu_1'\frac{\theta}{n} + O(\frac{\theta^2}{n^2})\Big)^n \\ &= \Big(1 + \mu_1'\frac{\theta}{n} + O(\frac{1}{n^2})\Big)^n = \Big((1+x)^{\frac{1}{x}}\Big)^{\mu\theta + O(n^{-1})} \\ &\longrightarrow \exp(\mu\theta) \quad \text{as } x \longrightarrow 0,\end{aligned}$$

which is the following probability function:

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

$$\phi(\theta) = \sum e^{\theta x} f(x) = e^{\theta\mu} f(\mu) = e^{\theta\mu}$$

Law of Large Numbers 2: Assume that $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed with mean $\mathrm{E}(X_i) = \mu$ and variance $\mathrm{V}(X_i) = \sigma^2 < \infty$ for all i.

Then, for any positive value ϵ, as $n \longrightarrow \infty$, we have the following result:

$$P(|\overline{X}_n - \mu| > \epsilon) \longrightarrow 0,$$

where $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

We say that $\overline{X}_n$ converges in probability to μ.