Example 1.13a (Efficient Estimator of μ): Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

Then, we show that $\overline{X}$ is an efficient estimator of μ.

$V(\overline{X})$ is given by $\dfrac{\sigma^2}{n}$, which does not depend on the distribution of X_i, $i = 1, 2, \cdots, n$. .... (A)

Because X_i is normally distributed with mean μ and variance σ^2, the density function of X_i is given by:

$$f(x;\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big).$$

The Cramer-Rao inequality is represented as:

$$V(\overline{X}) \ge \frac{1}{n\mathrm{E}\Big(\Big(\dfrac{\partial \log f(X;\mu)}{\partial \mu}\Big)^2\Big)},$$

where the logarithm of $f(X;\mu)$ is written as:

$$\log f(X;\mu) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(X-\mu)^2.$$

The partial derivative of $f(X;\mu)$ with respect to μ is:

$$\frac{\partial \log f(X;\mu)}{\partial \mu} = \frac{1}{\sigma^2}(X-\mu).$$

The Cramer-Rao inequality in this case is written as:

$$V(\overline{X}) \ge \frac{1}{n\mathrm{E}\Big(\Big(\dfrac{1}{\sigma^2}(X-\mu)\Big)^2\Big)} = \frac{1}{n\dfrac{1}{\sigma^4}\mathrm{E}((X-\mu)^2)} = \frac{\sigma^2}{n}. \quad \text{.................. (B)}$$

From (A) and (B), variance of $\overline{X}$ is equal to the lower bound of Cramer-Rao inequality, i.e., $V(\overline{X}) = \dfrac{\sigma^2}{n}$, which implies that the equality included in the Cramer-Rao inequality holds.

Therefore, we can conclude that the sample mean $\overline{X}$ is an efficient estimator of μ.

Example 1.13b (Efficient Estimator of σ^2): Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

Is S^2 is an efficient estimator of σ^2?

$\mathrm{E}(S^2) = \sigma^2$ .... Unbiased estimator

Under normality assumption, $V(S^2)$ is given by $\dfrac{2\sigma^4}{n-1}$, because $V(U) = 2(n-1)$ from $U = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

.... (A)

Because X_i is normally distributed with mean μ and variance σ^2, the density function of X_i is given by:

$$f(x;\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big).$$

The Cramer-Rao inequality is represented as:

$$V(S^2) \ge \frac{1}{n\mathrm{E}\Big(\Big(\dfrac{\partial \log f(X;\sigma^2)}{\partial \sigma^2}\Big)^2\Big)} = \frac{1}{-n\mathrm{E}\Big(\dfrac{\partial^2 \log f(X;\sigma^2)}{\partial (\sigma^2)^2}\Big)},$$

where the logarithm of $f(X;\sigma^2)$ is written as:

$$\log f(X;\sigma^2) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(X-\mu)^2.$$

The partial derivative of $f(X;\sigma^2)$ with respect to σ^2 is:

$$\frac{\partial \log f(X;\sigma^2)}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(X-\mu)^2.$$

The 2nd partial derivative of $f(X;\sigma^2)$ with respect to σ^2 is:

$$\frac{\partial^2 \log f(X;\sigma^2)}{\partial (\sigma^2)^2} = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(X-\mu)^2.$$

The Cramer-Rao inequality in this case is written as:

$$\mathrm{V}(S^2) \geq \frac{1}{-n\mathrm{E}\left(\frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(X-\mu)^2.\right)} = \frac{2\sigma^4}{n}. \;\ldots\ldots.(\mathrm{B})$$

From (A) and (B), variance of S^2 is not equal to the lower bound of Cramer-Rao inequality, i.e., $\mathrm{V}(S^2) = \dfrac{2\sigma^4}{n-1} > \dfrac{2\sigma^4}{n}$.

Therefore, we can conclude that the sample unbiased variance S^2 is not an efficient estimator of σ^2.

Example 1.14: Minimum Variance Linear Unbiased Estimator (最小分散線形不偏推定量): Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed with mean μ and variance σ^2 (note that the normality assumption is excluded from Example 1.13).

Consider the following linear estimator: $\hat{\mu} = \sum_{i=1}^{n} a_i X_i$.

Then, we want to show $\hat{\mu}$ (i.e., $\overline{X}$) is a **minimum variance linear unbiased estimator** if $a_i = \dfrac{1}{n}$ for all i, i.e., if $\hat{\mu} = \overline{X}$.

Utilizing Theorem on p.134, when $\mathrm{E}(X_i) = \mu$ and $\mathrm{V}(X_i) = \sigma^2$ for all i, we have: $\mathrm{E}(\hat{\mu}) = \mu \sum_{i=1}^{n} a_i$ and $\mathrm{V}(\hat{\mu}) = \sigma^2 \sum_{i=1}^{n} a_i^2$.

Since $\hat{\mu}$ is linear in X_i, $\hat{\mu}$ is called a **linear estimator (線形推定量)** of μ.

In order for $\hat{\mu}$ to be unbiased, we need to have the condition: $\mathrm{E}(\hat{\mu}) = \mu \sum_{i=1}^{n} a_i = \mu$.

That is, if $\sum_{i=1}^{n} a_i = 1$ is satisfied, $\hat{\mu}$ gives us a **linear unbiased estimator (線形不偏推定量)**.

Thus, as mentioned in Example 1.12 of Section 7.2, there are numerous unbiased estimators.

The variance of $\hat{\mu}$ is given by $\sigma^2 \sum_{i=1}^{n} a_i^2$.

We obtain the value of a_i which minimizes $\sum_{i=1}^{n} a_i^2$ with the constraint $\sum_{i=1}^{n} a_i = 1$.

Construct the Lagrange function as follows:

$$L = \frac{1}{2}\sum_{i=1}^{n} a_i^2 + \lambda(1 - \sum_{i=1}^{n} a_i),$$

where λ denotes the Lagrange multiplier.

The $\frac{1}{2}$ in the first term makes computation easier.

For minimization, the partial derivatives of L with respect to a_i and λ are equal to zero, i.e.,

$$\frac{\partial L}{\partial a_i} = a_i - \lambda = 0, \qquad i = 1, 2, \cdots, n,$$
$$\frac{\partial L}{\partial \lambda} = 1 - \sum_{i=1}^{n} a_i = 0.$$

Solving the above equations, $a_i = \lambda = \frac{1}{n}$ is obtained.

When $a_i = \frac{1}{n}$ for all i, $\hat{\mu}$ has minimum variance in a class of linear unbiased estimators.

$\overline{X}$ is a **minimum variance linear unbiased estimator**.

The **minimum variance linear unbiased estimator** is different from the **efficient estimator**.

The former does not requires the normality assumption.

The latter gives us the unbiased estimator which variance is equal to the Cramer-Rao lower bound, which is not restricted to a class of the linear unbiased estimators.

Under normality assumption, he linear unbiased minimum variance estimator leads to the efficient estimator.

Note that the efficient estimator does not necessarily exist.

Consistency (一致性): Let $\hat{\theta}_n$ be an estimator of θ.

Suppose that for any $\epsilon > 0$ we have the following:

$$P(|\hat{\theta}_n - \theta| \geq \epsilon) \longrightarrow 0, \qquad \text{as} \quad n \longrightarrow \infty,$$

which implies that $\hat{\theta} \longrightarrow \theta$ as $n \longrightarrow \infty$.

We say that $\hat{\theta}_n$ is a **consistent estimator (一致推定量)** of θ.

Example 1.15: Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed with mean μ and variance σ^2.

Assume that σ^2 is known.

Then, it is shown that $\overline{X}$ is a consistent estimator of μ.

For RV X, Chebyshev's inequality is given by:

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

Here, replacing X by $\overline{X}$, we obtain $\mathrm{E}(\overline{X})$ and $\mathrm{V}(\overline{X})$ as follows:

$$\mathrm{E}(\overline{X}) = \mu, \quad \mathrm{V}(\overline{X}) = \frac{\sigma^2}{n},$$

because $\mathrm{E}(X_i) = \mu$ and $\mathrm{V}(X_i) = \sigma^2 < \infty$ for all i.

Then, when $n \longrightarrow \infty$, we obtain the following result:

$$P(|\overline{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0,$$

which implies that $\overline{X} \longrightarrow \mu$ as $n \longrightarrow \infty$.

Therefore, $\overline{X}$ is a consistent estimator of μ.

Summary:

When the distribution of X_i is **not** assumed for all i, $\overline{X}$ is an **minimum variance linear unbiased and consistent estimator** of μ.

When the distribution of X_i is assumed to be **normal** for all i, $\overline{X}$ leads to an **efficient and consistent estimator** of μ.

Example 1.16a: Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

Consider $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, which is an unbiased estimator of σ^2.

We obtain the following Chebyshev's inequality:

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

We compute $\mathrm{E}((S^2 - \sigma^2)^2) \equiv \mathrm{V}(S^2)$.

$$U = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

$\mathrm{E}(U) = n-1$ and $\mathrm{V}(U) = 2(n-1)$.

$$\mathrm{V}(U) = \mathrm{V}(\frac{(n-1)S^2}{\sigma^2}) = 2(n-1)$$

$$\frac{(n-1)^2}{\sigma^4}\mathrm{V}(S^2) = 2(n-1)$$

$$\mathrm{V}(S^2) = \frac{2\sigma^2}{n-1}$$

$$P(|S^2 - \sigma^2| \geq \epsilon) \leq \frac{\mathrm{E}((S^2 - \sigma^2)^2)}{\epsilon^2} = \frac{2\sigma^2}{(n-1)\epsilon^2} \longrightarrow 0,$$

which implies that $S^2 \longrightarrow \sigma^2$ as $n \longrightarrow \infty$.

Threfore, S^2 is a consistent estimator of σ^2.

Example 1.16b: Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

Consider $S^{**2} = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$, which is an estimate of σ^2.

We obtain the following Chebyshev's inequality:

$$P(|S^{**2} - \sigma^2| \geq \epsilon) \leq \frac{\mathrm{E}((S^{**2} - \sigma^2)^2)}{\epsilon^2}.$$

We compute $\mathrm{E}((S^{**2} - \sigma^2)^2)$.

Define $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ as an estimator σ^2.

From $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, we obtain $\mathrm{E}\Big(\dfrac{(n-1)S^2}{\sigma^2}\Big) = n-1$ and $\mathrm{V}\Big(\dfrac{(n-1)S^2}{\sigma^2}\Big) = 2(n-1)$.

Therefore, $\mathrm{E}(S^2) = \sigma^2$ and $\mathrm{V}(S^2) = \dfrac{2\sigma^4}{n-1}$ can be derived.

Using $S^{**2} = \dfrac{n-1}{n}S^2$, we have the following:

$$\begin{aligned}\mathrm{E}((S^{**2} - \sigma^2)^2) &= \mathrm{E}\Big(\Big(\frac{n-1}{n}S^2 - \sigma^2\Big)^2\Big) \\ &= \mathrm{E}\Big(\Big(\frac{n-1}{n}(S^2 - \sigma^2) - \frac{\sigma^2}{n}\Big)^2\Big) \\ &= \frac{(n-1)^2}{n^2}\mathrm{E}((S^2 - \sigma^2)^2) + \frac{\sigma^4}{n^2} \\ &= \frac{(n-1)^2}{n^2}\mathrm{V}(S^2) + \frac{\sigma^4}{n^2} = \frac{(2n-1)}{n^2}\sigma^4.\end{aligned}$$

Therefore, as $n \longrightarrow \infty$, we obtain:

$$P(|S^{**2} - \sigma^2| \geq \epsilon) \leq \frac{1}{\epsilon^2}\frac{(2n-1)}{n^2}\sigma^4 \longrightarrow 0.$$

Because $S^{**2} \longrightarrow \sigma^2$, S^{**2} is a consistent estimator of σ^2.

S^{**2} is biased (see Section 7.3, p.273), but is is consistent.

7.5 Estimation Methods

- **Maximum Likelihood Estimation Method (最尤推定法)**
- **Least Squares Estimation Method (最小二乗法)**
- **Method of Moment (積率法)**

7.5.1 Maximum Likelihood Estimator (最尤推定量)

In Section 7.4, the properties of the estimators $\overline{X}$ and S^2 are discussed.

It is shown that $\overline{X}$ is an unbiased, efficient and consistent estimator of μ under normality assumption and that S^2 is an unbiased and consistent estimator of σ^2.

The parameter θ is included in the underlying distribution $f(x;\theta)$.

$\theta = (\mu, \sigma^2)$ in the case of the normal distribution.

Now, in more general cases, we want to consider how to estimate θ.

The **maximum likelihood estimator (最尤推定量)** gives us one of the solutions.

Let $X_1, X_2, \cdots, X_n$ be mutually independently and identically distributed random samples.

X_i has the probability density function $f(x;\theta)$.

The joint density function of $X_1, X_2, \cdots, X_n$ is given by:

$$f(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta),$$

where θ denotes the unknown parameter.
Given the actually observed data $x_1, x_2, \cdots, x_n$, the joint density $f(x_1, x_2, \cdots, x_n; \theta)$ is regarded as a function of θ, i.e.,

$$l(\theta) = l(\theta; x) = l(\theta; x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i; \theta).$$

$l(\theta)$ is called the **likelihood function (尤度関数)**.

Let $\hat{\theta}_n$ be the θ which maximizes the likelihood function.
Given data $x_1, x_2, \cdots, x_n$, $\hat{\theta}_n(x_1, x_2, \cdots, x_n)$ is called the **maximum likelihood estimate (MLE, 最尤推定値)**.
Replacing $x_1, x_2, \cdots, x_n$ by $X_1, X_2, \cdots, X_n$, $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$ is called the **maximum likelihood estimator (MLE, 最尤推定量)**.
That is, solving the following equation:

$$\frac{\partial l(\theta)}{\partial \theta} = 0,$$

MLE $\hat{\theta}_n \equiv \hat{\theta}_n(X_1, X_2, \cdots, X_n)$ is obtained.

Example 1.17a: Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

We derive the maximum likelihood estimators of μ and σ^2.

The joint density (or the likelihood function) of $X_1, X_2, \cdots, X_n$ is:

$$f(x_1, x_2, \cdots, x_n; \mu, \sigma^2) = \prod_{i=1}^{n} f(x_i; \mu, \sigma^2)$$

$$= \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\Big(-\frac{1}{2\sigma^2}(x_i - \mu)^2\Big)$$

$$= (2\pi\sigma^2)^{-n/2} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\Big)$$

$$= l(\mu, \sigma^2).$$

The logarithm of the likelihood function is given by:

$$\log l(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2,$$

which is called the **log-likelihood function (対数尤度関数)**.

For maximization of the likelihood function, differentiating the log-likelihood function $\log l(\mu, \sigma^2)$ with respect to μ and σ^2, the first derivatives should be equal to zero, i.e.,

$$\frac{\partial \log l(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0,$$

$$\frac{\partial \log l(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0.$$

Let $\hat{\mu}$ and $\hat{\sigma}^2$ be the solution which satisfies the above two equations.

Solving the two equations, we obtain the maximum likelihood estimates as follows:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x},$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = s^{**2}.$$

Replacing x_i by X_i for $i = 1, 2, \cdots, n$, the maximum likelihood estimators of μ and σ^2 are given by $\overline{X}$ and S^{**2}.

Since $\mathrm{E}(\overline{X}) = \mu$, the maximum likelihood estimator of μ, $\overline{X}$, is an unbiased estimator.

We have checked that $\overline{X}$ is efficient and consistent.

However, because of $\mathrm{E}(S^{**2}) = \dfrac{n-1}{n}\sigma^2 \neq \sigma^2$ as shown in Section 7.3, the maximum likelihood estimator of σ^2, S^{**2}, is not an unbiased estimator.

We have checked that S^{**2} is inefficient but consistent.