Chapter 7: Parameter Estimation (参数估计)

§1 Point estimation (点估计)

§2 The selection criteria of estimators

(估计量的评价标准)

§3 Interval estimation (区间估计)

Chapter 7: Parameter Estimation (参数估计)

§1 Point estimation (点估计)

§2 The selection criteria of estimators

(估计量的评价标准)

§3 Interval estimation (区间估计)

Recall some of the examples in the last lecture:

Example: Assume that $X_1, X_2, \dots, X_n$ are samples from the population $X \sim U(a,b)$. Based on the Method of Moment, the Point Estimation of a and b are

$$\hat{a} = \bar{X} - \sqrt{3}\tilde{S}, \hat{b} = \bar{X} + \sqrt{3}\tilde{S},$$

Based on the MLE, the Point Estimation of a and b are

$$\hat{a} = X_{(1)} = \min_{1 \le i \le n}\{X_i\}, \qquad \hat{b} = \max_{1 \le i \le n}\{X_i\}.$$

Example: Assume that $X_1, X_2, \dots, X_n$ are samples from the population $X \sim P(\lambda)$, since

$$E(X) = D(X) = \lambda$$

Therefore, $\hat{\lambda}_1 = \bar{X}$ and $\hat{\lambda}_2 = \tilde{S}^2$ can both be the Moment Estimation of the unknown parameter λ.

How to evaluate and select the Point Estimations of a parameter?

How to evaluate and select the Point Estimations of a parameter?

- Unbiasedness (无偏性)
- Efficiency (有效性)
- Consistency (一致性/相合性)

(1) Unbiasedness（无偏性）

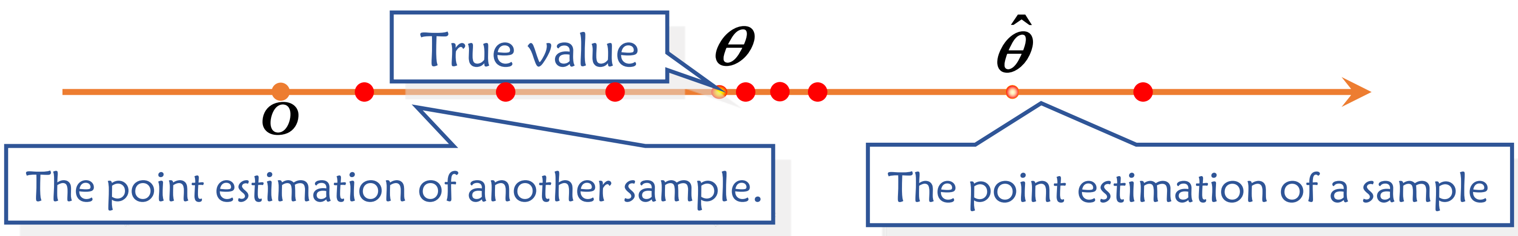
Assume that the population is $X \sim F(x;\theta)(\theta \in \Theta)$, Θ is the Parameter Space.

Let $X_1, X_2, \ldots, X_n$ be samples from the population X, $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ is a Point estimation of the unknown parameter θ.

Intuitively, what properties should a "great" estimation have?

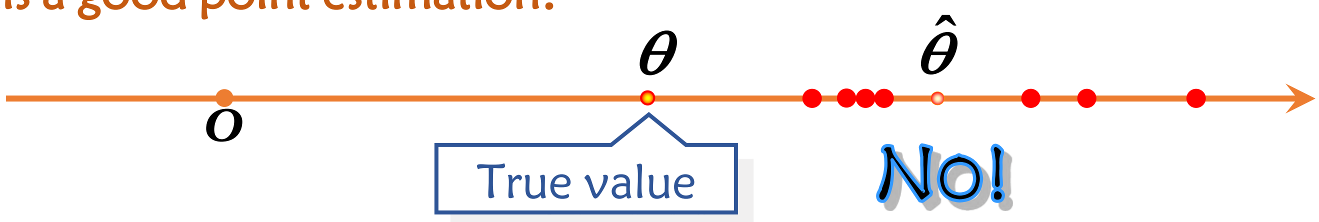
Analysis: $\because$ estimator (估计量) $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ is a random variable

$\therefore$ estimation (估计值) $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$ has volatility (波动性)



Note: For a good estimation, most of the estimations should fluctuate around the true value of the parameter being estimated.

Is this a good point estimation?



Definition: If the mathematical expectation of the estimator $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ exists, and for $\forall\theta \in \Theta$:

$$E_\theta(\hat{\theta}) = \theta$$

$E_\theta(\hat{\theta})$ is calculatd based on the population distribution $F(x;\theta)$. Its **mathematical expectation** is related to parameter θ.

Then,

$\hat{\theta}$ is called an Unbiased Estimator (无偏估计) of θ.

Otherwise, it is called a Biased Estimator (有偏估计) of θ .

Call $b_n(\hat{\theta}) = E_\theta(\hat{\theta}) - \theta$ **the Bias（偏差）** of estimator $\hat{\theta}$.

- If $b_n(\hat{\theta}) = 0$, then $\hat{\theta}$ is an Unbiased Estimator of θ (无偏估计)
- If $b_n(\hat{\theta}) \neq 0$, then $\hat{\theta}$ is a Biased Estimator of θ (有偏估计)
- If $\lim\limits_{n\to\infty} b_n(\hat{\theta}) = 0$, then $\hat{\theta}$ is an Asymptotic Unbiased Estimation of θ (渐近无偏估计)

Example: No matter what distribution a population X follows, if both

$$\mu \triangleq E(X), \sigma^2 \triangleq D(X)$$

exist, then $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = S^2$ are Unbiased estimators of μ and σ^2.

Proof: Based on the results from Chapter 6:

$$E(\hat{\mu}) = E(\bar{X}) = \mu$$

$$E(\hat{\sigma}^2) = E(S^2) = \sigma^2$$

Thus, $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = S^2$ are Unbiased estimators of μ and σ^2.

This explains why the definition of sample variance is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Note: The Adjusted Sample Variance is

$$\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}S^2$$

$$\therefore E(\tilde{S}^2) = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\sigma^2$$

Therefore $\tilde{S}^2$ is an Asymptotic unbiased estimator of σ^2.

Example: No matter what distribution a population X follows, if both

$$\mu \triangleq E(X), \sigma^2 \triangleq D(X)$$

exist, then $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = S^2$ are Unbiased estimators of μ and σ^2.

Example: Assume that $X_1, X_2, \dots, X_n$ are samples from the population $X \sim N(\mu, \sigma^2)$, are the **Moment estimation** or the **MLE** of μ and σ^2 unbiased?

Answer: The Moment estimation and MLE of μ are both

$$\hat{\mu} = \bar{X},$$

which is an **unbiased estimator** of μ.

The Moment estimation and MLE of σ^2 are both

$$\hat{\sigma}^2 = \tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

which is a **Biased estimator** of σ^2 or an **Asymptotic unbiased estimator.**

An Unbiased estimation of σ^2 is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Example: Assume that $X_1, X_2, \dots, X_n$ are samples from the population $X \sim U(0,\theta)(\theta > 0)$. Discuss the Moment estimation $\hat{\theta}_M$ and MLE $\hat{\theta}_L$ of θ in terms of unbiasedness.

Answer: $\because E(X) = \dfrac{\theta}{2}, \qquad \therefore \bar{X} \approx \dfrac{\theta}{2}$

Therefore, the Moment estimation of θ is $\hat{\theta}_M = 2\bar{X}$

$$\because E(\hat{\theta}_M) = 2E(\bar{X}) = 2E(X) = 2 \cdot \frac{\theta}{2} = \theta$$

$\therefore$ the Moment estimation $\hat{\theta}_M = 2\bar{X}$ is an **Unbiased estimator** of θ.

The **Likelihood** function is

$L(\theta) = \dfrac{1}{\theta^n} \quad (0 < X_1, \cdots, X_n < \theta)$. $L(\theta)$ is monotone decreasing on θ

Therefore, the MLE of θ is $\hat{\theta}_L = X_{(n)} = \max\limits_{1 \le i \le n} X_i$

- How to calculate $E(\hat{\theta}_L)$?

Answer: From Chapter 3 Extrema and Order Statistics, we have:

$$\because F_{\max}(z) = P\{X_{(n)} \le z\} = [F(z)]^n \Rightarrow f_{\max}(z) = nf(z)[F(z)]^{n-1}$$

$$\therefore E(\hat{\theta}_L) = \int_{-\infty}^{\infty} z \cdot f_{\max}(z) dz$$

Example: Assume that $X_1, X_2, \ldots, X_n$ are samples from the population $X \sim U(0,\theta)(\theta > 0)$. Discuss the Moment estimation $\hat{\theta}_M$ and MLE $\hat{\theta}_L$ of θ in terms of unbiasedness.

Answer: The density function of $X_{(n)} = \max\limits_{1\le i\le n} X_i$ is

$$f_{\max}(z) = nf(z)[F(z)]^{n-1} = \begin{cases} \dfrac{n}{\theta}\left(\displaystyle\int_0^z \frac{1}{\theta}dx\right)^{n-1} & ,0 < z < \theta \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \dfrac{n}{\theta}\left(\dfrac{z}{\theta}\right)^{n-1} & ,0 < z < \theta \\ 0, & \text{otherwise} \end{cases}$$

$$\therefore E(\hat{\theta}_L) = \int_{-\infty}^{\infty} z \cdot f_{\max}(z)dz$$

$$= \int_0^{\theta} z \cdot \frac{n}{\theta}\left(\frac{z}{\theta}\right)^{n-1} dz = \frac{n}{n+1}\theta$$

$\therefore$ $\hat{\theta}_L = X_{(n)}$ is a Biased estimator (Asymptotic unbiased estimator) of θ.

If it is adjusted as: $\hat{\theta}_L = \dfrac{n+1}{n}X_{(n)}$,

then it is an **Unbiased estimator** of θ.

Practical meaning of Unbiasedness

Engineering	$E(\hat{\theta})-\theta$ is called the system error （系统误差）
Economy Activity	Unbiasedness represents the fairness of commercial activities（无偏性反映了商业行为的公平性）
Competition Grading	Unbiasedness represents the justice in competition grading（无偏性反映了评分的公正性）

Under what circumstances the unbiasedness is meaningful?

Unbiasedness is meaningful only when the number of trials is large.

(2) Efficiency (有效性)

Example: Assume that $X_1, X_2, \dots, X_n$ are samples from the population $X \sim P(\lambda)$.

$$\because E(X) = D(X) = \lambda$$

$$\therefore E(\bar{X}) = E(X) = \lambda, \qquad E(S^2) = D(X) = \lambda$$

Thus $\hat{\lambda}_1 = \bar{X}, \hat{\lambda}_2 = S^2$ are both **unbiased estimators** of λ.

For any constant c, the statistic

$$\hat{\lambda} = c\hat{\lambda}_1 + (1-c)\hat{\lambda}_2 = c\bar{X} + (1-c)S^2,$$

is also an **unbiased estimator** of λ.

How to compare two unbiased estimators?

The one with a smaller variance is better.

Definition: Assume that $X_1, X_2, \ldots, X_n$ are samples from the population $X \sim F(x, \theta); \theta \in \Theta$.

$$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n), \qquad \hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$$

are two unbiased estimators of θ, i.e., $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta \ (\forall \theta \in \Theta)$.

If for $\forall \theta \in \Theta$, we have:

$$D(\hat{\theta}_1) \leq D(\hat{\theta}_2),$$

then $\hat{\theta}_1$ is said to be **more efficient** than $\hat{\theta}_2$.

Example: Assume that X_1, X_2, X_3, X_4 are samples from the population $X \sim EXP(1/\theta)$ (θ is the unknown parameter). The following are estimators:

$$T_1 = \frac{1}{6}(X_1 + X_2) + \frac{1}{3}(X_3 + X_4)$$

$$T_2 = \frac{1}{5}(X_1 + 2X_2 + 3X_3 + 4X_4)$$

$$T_3 = \frac{1}{4}(X_1 + X_2 + X_3 + X_4)$$

a) Which one of the T_1 , T_2 , T_3 is the unbiased estimation of θ ?
b) Which one of the estimations T_1 , T_2 , T_3 has more efficient?

(3) Consistency (一致性/相合性)

Assume that $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ is a Point estimation of an unknown parameter θ.

When n increases, how to evaluate whether $\hat{\theta}$ is a better estimation?

Idea: When the sample size n increases, the information about the unknown parameter θ in the sample $X_1, X_2, \dots, X_n$ also increases. Therefore, the estimation should be more accurate.

Since $\hat{\theta}$ is r.v., how to describe the accuracy of an estimation?

Definition: Assume that $\hat{\theta}_n = \hat{\theta}(X_1, X_2, \dots, X_n)$ is a Point estimation of the unknown parameter θ. If for $\forall \theta \in \Theta$ and $\forall \varepsilon > 0$ we have:

$$\lim_{n\to\infty} P\{|\hat{\theta}_n - \theta| \geq \varepsilon\} = 0$$

then $\hat{\theta}_n$ is called a **Consistent Estimator (相合估计)** of θ.

Note: $\hat{\theta}_n$ is a consistent estimator of θ $\Leftrightarrow$ $\hat{\theta}_n \xrightarrow{P} \theta \ (n \to \infty)$

Example: No matter what distribution a population X follows, if both

$$\mu \triangleq E(X), \sigma^2 \triangleq D(X)$$

exist, then $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = S^2$ are Consistent estimators（相合估计）of μ and σ^2.

Proof: Based on the Khinchine's law of large numbers

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu \ (n \to \infty)$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n(\bar{X})^2\right)$$

$$= \frac{n}{n-1} \cdot \frac{1}{n}\sum_{i}^{n} X_i^2 - \frac{n}{n-1}(\bar{X})^2 \xrightarrow{P} E(X^2) - \mu^2 = \sigma^2$$

Thus $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = S^2$ are Consistent Estimators of μ and σ^2.

Common conclusions on the Consistent Estimation

1. Based on the Khinchine's Law of large numbers, the Moment estimation $\hat{\theta}_M$ of θ is a consistent estimation.
2. The MLE $\hat{\theta}_L$ of θ is also generally a consistent estimation.
3. A consistent estimator of θ may not be an unbiased estimator.

For the population $X \sim U(0,\theta)(\theta > 0)$, $\hat{\theta}_L = X_{(n)}$ is the MLE of θ. Since for $\forall \varepsilon > 0$:

$$P\{|\hat{\theta}_L - \theta| \geq \varepsilon\} = P\{X_{(n)} \leq \theta - \varepsilon\} = [F(\theta - \varepsilon)]^n = \left(\frac{\theta - \varepsilon}{\theta}\right)^n \to 0\ (n \to \infty)$$

$\therefore$ $\hat{\theta}_L$ is a consistent estimator. However, it is a biased estimator of θ since:

$$E(\hat{\theta}_L) = \frac{n}{n+1}\theta.$$

4. If $\hat{\theta}$ is an unbiased estimator of θ, based on the Chebyshev's Inequality (切比雪夫不等式), it satisfies

$$P\{|\hat{\theta} - \theta| \geq \varepsilon\} \leq \frac{D(\hat{\theta})}{\varepsilon^2}$$

Therefore, when $\lim_{n\to\infty} D(\hat{\theta}) = 0$, $\hat{\theta}$ is a **consistent estimator** of θ (sufficient condition, not necessary condition).

Example: Assume that $X_1, X_2, \dots, X_n$ is a sample from the population $X \sim b(m, p)(0 < p < 1)$, find the MLE $\hat{p}$ of the unknown parameter p. Prove that $\hat{p}$ is an unbiased estimator and a consistent estimator of p.

Answer: The Likelihood function is

$$L(p) = \prod_{i=1}^{n} C_m^{X_i} p^{X_i}(1-p)^{m-X_i} = \left(\prod_{i=1}^{n} C_m^{X_i}\right) \cdot p^{\sum_{i=1}^{n} X_i} \cdot (1-p)^{\sum_{i=1}^{n}(m-X_i)}$$

$$= \left(\prod_{i=1}^{n} C_m^{X_i}\right) \cdot p^{n\bar{X}} \cdot (1-p)^{mn-n\bar{X}}$$

Let $\frac{\partial \ln L}{\partial p} = \frac{n\bar{X}}{p} - \frac{mn - n\bar{X}}{1-p} = 0$, the MLE of p is obtained as $\hat{p} = \frac{\bar{X}}{m}$

$\because E(\hat{p}) = \frac{E(\bar{X})}{m} = \frac{mp}{m} = p$, $\therefore$ $\hat{p}$ is an unbiased estimator of p.

$\because$ $\hat{p} = \frac{\bar{X}}{m} \xrightarrow{P} \frac{mp}{m} = p \ (n \to \infty)$, $\therefore$ $\hat{p} = \frac{\bar{X}}{m}$ is a consistent estimator of p.

Summary

There are three selection criteria:

- Unbiasedness (无偏性)

$$E_\theta(\hat{\theta}) = \theta$$

- Efficiency (有效性)

$$D_\theta(\hat{\theta}_1) \le D_\theta(\hat{\theta}_2)$$

- Consistency (一致性/相合性)

$$\hat{\theta}_n \xrightarrow{P} \theta \ (n \to \infty)$$

Homework

1. Assume that $X_1, X_2, \dots, X_n$ is a sample from $X \sim N(\mu, \sigma^2)$. Compute the constant k such that $\sigma^2 = \frac{1}{k}\sum_{i=1}^{n-1}(X_{i+1} - X_i)^2$ is an unbiased estimator of σ^2.

2. Pick two independent samples with sample size n_1, n_2 from a population with mean μ and variance $\sigma^2 > 0$. Let $\bar{X}_1$ and $\bar{X}_2$ be the means of these two samples, respectively. Prove that for any a and b $(a + b = 1)$, $Y = a\bar{X}_1 + b\bar{X}_2$ is an unbiased estimator of μ and then calculate the constants a and b that minimize $D(Y)$.

3. Suppose that the population $X \sim EXP(1/\theta)(\theta > 0)$ with density function

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, x > 0 \\ 0, \qquad x \le 0 \end{cases}.$$

(a) Prove: $\bar{X}$ and $n \cdot \min\{X_1, X_2, \cdots, X_n\}$ are both unbiased estimators of θ.
(b) Which of the two unbiased estimators is more efficient?