MA212 Probability and Statistics

Chapter Seven: Point Estimation and Testing of Hypotheses

Wenwu GONG & Qianqian WANG
Department of Statistics and Data Science
Email: 12031299@mail.sustech.edu.cn
12032005@mail.sustech.edu.cn

December 13, 2021

Point Estimation

Assume a population is $X \sim F(x;\theta_1,\theta_2,\dots,\theta_m)$, the function F is unknown. $\theta_1,\theta_2,\dots,\theta_m$ are **unknown parameters** in the population. $X_1,X_2,\dots,X_n$ are the samples from population X.

- **Parameter Space （参数空间）**: the value range of θ is parameter space. Noted as Θ.

  Parametric inference: The function of F is known. To infer the unknown parameters θ.

- **The point estimation of θ**: Build **a statistic $\hat{\theta}(X_1,X_2,\dots,X_n)$.**

Using the observation value $\hat{\theta}(x_1,x_2,\dots,x_n)$ of the statistic as the estimated value of unknown parameter θ.

$\hat{\theta}(X_1,X_2,\dots,X_n)$ is the estimator（估计量） of θ

$\hat{\theta}(x_1,x_2,\dots,x_n)$ is the estimated value（估计值） of θ

Duality

- **Methods:**
  - Moment Estimation（矩估计法）
  - Maximum Likelihood Estimation （最大似然估计法）
  - Least square estimation （最小二乘估计法）
  - Others…… (Iteration Algorithms)

Assuming the population is $X \sim F(x;\theta_1,\theta_2,\ldots,\theta_m)$. $\theta_1,\theta_2,\ldots,\theta_m$ are unknown parameters. $X_1,X_2,\ldots,X_n$ are the samples from the population X. If the following population Moment does exist

$$\alpha_k \triangleq E(X^k) \qquad (k=1,\ 2,\cdots,m)$$

Based on Khinchine law of large numbers

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} E(X^k) = \alpha_k \quad (n\to\infty, k=1,2,\cdots,m)$$

The Moment Estimation of μ *and* σ^2 are

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

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

- Assuming $X_1, X_2, \ldots, X_n$ are the samples of population $X \sim f(x;\theta)$, let

$$L(\theta) = L(\theta\,;X_1, X_2, \cdots, X_n) = \prod_{i=1}^{n} f(X_i;\theta)$$

Then we define $L(\theta)$ as Likelihood Function（似然函数）

- If the statistic $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ exists, then

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta\,; X_1, X_2, \cdots, X_n)$$

Thus $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ is the Maximum Likelihood Estimation, or MLE

- **Moment estimation method:**

  (1) Find the population Moment

  (2) Sample moment replace population moment

  (3) Find Moment Estimation

- **Maximum likelihood estimation method:**

  (1) Find Likelihood Function

  (2) Make the Likelihood Equation(s)

  (3) Find the maximum of the Likelihood Function or the *Ln* Likelihood Function by solving the Likelihood Equation(s).

Unbiased estimator

Assuming $X_1, X_2, \ldots, X_n$ are the samples of population X, $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ are the Point Estimation of unknown parameter θ. If the mathematic Expectation（数学期望） of estimator $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ exists. And $\forall \theta \in \Theta$ has

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

Thus $\hat{\theta}$ is the Unbiased Estimation（无偏估计）of θ

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

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

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

Define $b_n(\hat{\theta}) = E_\theta(\hat{\theta}) - \theta$ as the Bias（偏差）of estimator $\hat{\theta}$

- If $b_n(\hat{\theta}) = 0$, then $\hat{\theta}$ is the Unbiased Estimation of θ（无偏估计）
- If $b_n(\hat{\theta}) \neq 0$, then $\hat{\theta}$ is the Biased Estimation of θ（有偏估计）
- If $\lim_{n\to\infty} b_n(\hat{\theta}) = 0$, then $\hat{\theta}$ is the Asymptotic Unbiased Estimation of θ（渐进无偏估计）

Efficiency

Assuming $X_1, X_2, \dots, X_n$ are the samples of population $X \sim F(x, \theta); \theta \in \Theta$.

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

are the unbiased estimation of θ, which is $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta (\forall \theta \in \Theta)$.

If $\forall \theta \in \Theta$

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

then $\hat{\theta}_1$ is **more efficient** than $\hat{\theta}_2$

Consistency

Assuming $\hat{\theta}_n = \hat{\theta}(X_1, X_2, \dots, X_n)$ is the Point estimation of unknown parameter θ.
If $\forall \theta \in \Theta$ has: $\forall \varepsilon > 0$

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

Then $\hat{\theta}_n$ is the Mutual Estimation（相合估计） of θ, consistency.

Example: No matter what distribution the population X follows, if

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

exist, then $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = S^2$ are the Mutual estimation of μ *and* σ^2.

1.6 Summary

- Based on Khinchine Law of large numbers, the Moment estimation $\hat{\theta}$ of θ is Mutual estimation.
- The MLE $\hat{\theta}$ of θ is always Mutual estimation.
- The Mutual estimation of θ may not be Unbiased estimation
- If $\hat{\theta}$ is the unbiased estimation of θ, based on Chebyshev Inequality （切比雪夫不等式）, it has

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

Therefore, when $\lim_{n\to\infty} D(\hat{\theta}) = 0$, $\hat{\theta}$ is the **Mutual Estimation** of θ.

Definition of interval estimation

Assuming a population $X \sim F(x;\theta)(\theta \in \Theta)$, $\forall 0 < \alpha < 1$, if there are two statistic

$$\underline{\theta} = \underline{\theta}(X_1, X_2, \cdots, X_n),\ \overline{\theta} = \overline{\theta}(X_1, X_2, \cdots, X_n) \quad (\underline{\theta} < \overline{\theta})$$

Then $\forall \theta \in \Theta$

$$P\{\underline{\theta} \le \theta \le \overline{\theta}\} \ge 1 - \alpha$$

Further, a random interval $(\underline{\theta}, \overline{\theta})$ (θ) with Confidence Level （置信水平） $1-\alpha$ is the Confidence Interval（置信区间）. $\underline{\theta}$ *and* $\overline{\theta}$ are the confidence lower limit（置信下限） and confidence upper limit（置信上限）.

- Confidence Level is also called Degree of confidence（置信度）, when α is small, then $1-\alpha$ is large.
- Interpretation (α=0.05): if sampling randomly 100 times, the number of values which is between the interval $(\underline{\theta}, \overline{\theta})$ of θ is about 95. The probability that interval covers true parameter θ.
- Estimation accuracy of confidence interval: CI is not unique, the shorter, the higher the accuracy.
- The general method to find interval estimation: using the **Pivot** Variable method（枢轴变量法） based on Wave theory（波动理论）.

Pivot Variable Method

Assuming θ is the unknown parameter to be estimated, φ is another unknown parameter

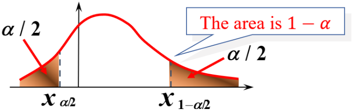
- Find the point estimations $\hat{\theta}$ *and* $\hat{\varphi}$ for θ *and* φ.

Pivot which distribution is irrelevant to unknown parameters

- Create sampling function

$$T = T(\theta, \hat{\theta}, \hat{\varphi}) \sim f(x)$$

- For the given confidence level $1-\alpha$, using $f(x)$ to find the two quantiles $x_{1-\alpha/2}$ and $x_{\alpha/2}$



$$P\{x_{\alpha/2} < T(\theta, \hat{\theta}, \hat{\varphi}) < x_{1-\alpha/2}\} = 1-\alpha$$

Equivalently

$$P\{\ \underline{\theta} < \theta < \overline{\theta}\ \} = 1-\alpha$$

- The Confidence interval of θ is $(\underline{\theta}, \overline{\theta})$.

Examples

Interval estimation of the unknown parameter in Single normal population
（单正态总体的未知参数的区间估计）

$$\left(\bar{X}-\frac{u_{1-\alpha/2}}{\sqrt{n}},\ \bar{X}+\frac{u_{1-\alpha/2}}{\sqrt{n}}\right)$$

$$\left(\bar{X}-\frac{S}{\sqrt{n}}t_{1-\alpha/2}(n-1),\ \bar{X}+\frac{S}{\sqrt{n}}t_{1-\alpha/2}(n-1)\right)$$

$$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)},\ \frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right)$$

Interval estimation of the unknown parameters in Double normal populations
（双正态总体的未知参数的区间估计）

$$\left((\bar{X}-\bar{Y})-u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ (\bar{X}-\bar{Y})+u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

$$\left((\bar{X}-\bar{Y})\pm t_{1-\frac{\alpha}{2}}(n_1+n_2-2)\,S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$$

1.10 Testing of Hypotheses

Basic concept

Two types of error
- Type I: H_0 is true, but being rejected.
- Type II: H_0 is not true, but being accepted

P - value

Let $W(\mathbf{X})$ be a test statistic such that large values of W give evidence that H_1 is true. For each sample point $\mathbf{x}$, define

$$p(\mathbf{x}) = \sup_{\theta \in \Theta_0} P_\theta \left(W(\mathbf{X}) \geq W(\mathbf{x}) \right).$$

Then, $p(\mathbf{X})$ is a valid p-value.

Framework

- Make hypotheses $\mathbf{H_0}: \boldsymbol{\mu} = \boldsymbol{\mu_0}$, $\mathbf{H_1}: \boldsymbol{\mu} \neq \boldsymbol{\mu_0}$

Rejection region.

Find the critical value C $\left\{ \text{When } |\overline{X} - \mu_0| \geq C, \text{ reject } H_0 \right\}$

- Make decision guided by type error I or p-value Reject H_0 or Accept H_0

How to testify the hypotheses?

Principles

Test principle 1: Protect H_0

The status of H_0 and H_1 is unequal.

Reasons

- The content of H_0 is important which related to the benefit of the tester.
- H_0 is based on something.
- The tester prefers H_0 is true or wrong.
- The content of H_0 is true. But if the judgement is wrong, the consequence is fateful.

Test principle 2: Control type I error (ignore type II error)

Rejection region.

Given a small number $\alpha (0 < \alpha < 1)$, let $P\{reject\ H_0 | H_0\ is\ true\} \leq \alpha$

- α is significant level, the value of α usually is 0.01, 0.05, 0.10.
- Using the test principle above with the significant level α to make decision to H_0. This process is doing test of significance to H_0.(Fisher test)

Test principle 3: Probability contradiction

Point out null hypothesis.

Single population $N(\mu, \sigma^2)$

Assuming $X_1, X_2, \dots, X_n$ are the samples of population $X \sim N(\mu, \sigma^2)$. μ *and* σ^2 are unknown. When the significant level is α, test hypothesis

Two-sided test

$$H_0: \mu = \mu_0, \quad H_1: \mu \neq \mu_0 \qquad (\mu_0 \text{ is known})$$

The rejection region of H_0 is

$$|\bar{X} - \mu_0| > \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1) \ .$$

One-sided test

$$H_0: \mu \leq \mu_0, \quad H_1: \mu > \mu_0 \qquad (\mu_0 \text{ is known})$$

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > t_{1-\alpha}(n-1)$$

Single population $N(\mu, \sigma^2)$

Assuming $X_1, X_2, \ldots, X_n$ are the samples of population $X \sim N(\mu, \sigma^2)$. μ *and* σ^2 are unknown. When the significant level is α, test hypothesis

Two-sided test

$$H_0: \sigma^2 = \sigma_0^2, \quad H_1: \sigma^2 \neq \sigma_0^2$$

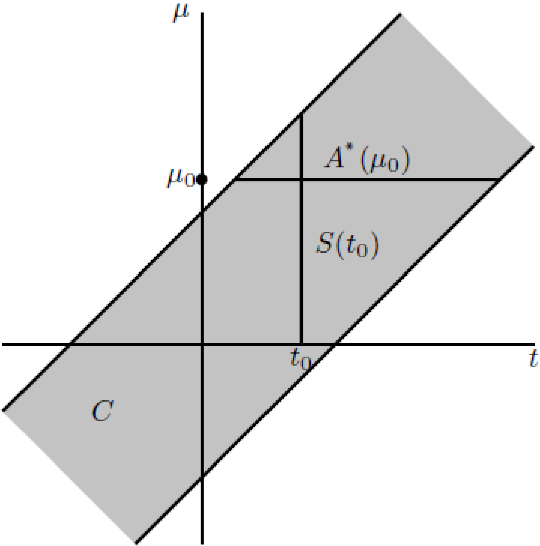
$$S^2 < \frac{\sigma_0^2}{n-1}\chi^2_{\alpha/2}(n-1) \quad .\text{or}. \quad S^2 > \frac{\sigma_0^2}{n-1}\chi^2_{1-\alpha/2}(n-1)$$

One-sided test

$$H_0: \sigma^2 \geq \sigma_0^2, H_1: \sigma^2 < \sigma_0^2$$

$$\frac{(n-1)S^2}{\sigma_0^2} < \chi^2_\alpha(n-1)$$

1.13 CI and Hypothesis



Point estimator

- Suppose $X_1, X_2, \cdots, X_n$ are sampled from $N(\mu, \sigma^2)$. find out the constant c such that $Y = c\sum_{i=1}^{n-1}(X_{i+1} - X_i)^2$ is a unbiased estimator of σ^2.
- Suppose that $x_1, x_2, \cdots, x_n$ are sampled from $f(x;\theta)$, find out the MLE of θ and $1/\theta$.
    - $f(x;\theta) = 1, \quad \theta - 1/2 < x < \theta + 1/2$
    - $f(x;\theta) = \theta x^{\theta-1}, \quad 0 < x < 1$

5. Suppose that X is a discrete random variable with $P(X=1)=\theta$ and $P(X=2) = 1-\theta$. Three independent observations of X are made: $x_1=1, x_2=2, x_3=2$.

a. Find the method of moments estimate of θ.

b. What is the likelihood function?

c. What is the maximum likelihood estimate of θ?

Supplementary Questions

1. Assume that $X_1, X_2, \dots, X_n$ is a sample from the population X with density function:

$$f(x;\theta)=\begin{cases}\frac{2}{\theta^2}(\theta-x), 0<x<\theta \\ 0, \qquad \text{otherwise}\end{cases}$$

Find the moment estimation of θ.

2. Assume that $X_1, X_2, \dots, X_n$ is a sample from the population X. Find the maximum likelihood estimate (MLE) of θ for the following density function of X:

(1) $f(x;\theta)=\begin{cases}\frac{\theta^x}{x!}e^{-\theta}, x=0,1,2,\cdots \\ 0, \qquad \text{otherwise}\end{cases}$ $(\theta>0)$

(2) $f(x;\theta)=\begin{cases}\theta\alpha x^{\alpha-1}e^{-\theta x^\alpha}, x>0 \\ 0, \qquad \text{otherwise}\end{cases}$ (α is known)

3. Suppose that the population X has density function:

$$f(x;\theta)=\begin{cases}\theta(1-x)^{\theta-1}, & 0\le x\le 1 \\ 0, & \text{otherwise}\end{cases}.$$

$X_1, X_2, \cdots, X_n$ is a sample from X, find the moment estimation and the maximum likelihood estimation of θ.

2.2 Homework1

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?

Confidence Interval

- Suppose 0.5, 1.25, 0.8, 2 are sampled from population X and $Y = \ln X \sim N(\mu, 1)$.
  - Find the confidence interval of population mean μ with the confidence level 0.95.
  - Find the confidence interval of E(X) with the confidence level 0.95.
- Suppose that $x_1, x_2, \cdots, x_n$ are sampled from $f(x;\theta) = \theta e^{-\theta x},\ x > 0$, find out the confidence interval of θ with the confidence level 1-α.

1. The drying time X (unit: h, represent hours) of a certain varnish follows a normal distribution $X \sim N(\mu, \sigma^2)$. Suppose a sample of size 9 of X is observed: 6.0, 5.7, 5.8, 6.5, 7.0, 6.3, 5.6, 6.1, 5.0. Find the confidence interval of μ with confidence level 0.95 under the following two conditions:

$$(1)\ \sigma = 0.6(h);\ (2)\ \sigma \text{ is unknown.}$$

2. Choose 16 bags of candies randomly and find the weight (unit: gram) of each as following:

506	508	499	503	504	510	497	512
514	505	493	496	506	502	509	496

Assume that the weight of a bag of candy follows a normal distribution.
(1) find the confidence interval of the population mean μ with confidence level 0.95.
(2) find the confidence interval of the population standard deviation σ with confidence level 0.95.

3. In order to compare the muzzle velocity of two types of rifle bullets I and II, 10 type I bullets are randomly selected, the average muzzle velocity is $\overline{x_1} = 500(m/s)$ and the variance is $s_1^2 = 1.10(m/s)^2$. 20 type II bullets are randomly selected, the average muzzle velocity is $\overline{x_2} = 496(m/s)$ and the variance is $s_2^2 = 1.20(m/s)^2$. Suppose that the two populations both approximately follow the normal distribution, and the production process can be considered as having the same variance. Find the confidence interval with confidence level 0.95 for the difference between the two population means $\mu_1 - \mu_2$.

1. The pulse (脉搏) of ordinary people follows a normal distribution with mean 62 times/minute. Assume that there are 10 patients, and their pulse measures are 54, 68, 65, 77, 70, 64, 69, 72, 62, 71 (times per minute). Are the pulse of the 10 patients significantly different from ordinary people at significant level $\alpha = 0.05$?

2. There are two methods A and B for studying the latent heat of ice. The following data are collected which measures the heat absorption per gram of ice when the temperature of ice increased from initial -0.72℃ to temperature 0℃:

   Method A : 79.98, 80.04, 80.02, 80.03, 80.03, 80.04, 80.04
   79.97, 80.05, 80.03, 80.02, 80.00, 80.02

   Method B: 80.02, 79.94, 79.97, 79.98, 79.97, 80.03, 79.95, 79.97

   Assume that the data from the two methods follow two normal distributions with equal variance. Test the hypothesis H_0: the mean of the two methods are the same. ($\alpha = 0.05$)

3. Assume the size of a type of device follows the normal distribution $N(\mu, \sigma^2)$. According to the requirements, the standard deviation shouldn't be more than 0.9. 19 devices were taken for assessment and found that the standard deviation is 1.2. Is the standard deviation acceptable at level $\alpha = 0.05$?

4. Assume that the IQ of students from each district of a city follows a normal distribution. Take a sample of 16 students from a district and the mean and standard deviation of their IQs are found to be 107 and 10. The sample mean and standard deviation of 16 students from another district are 112 and 8. Does the IQ of students from the two districts differ significantly at level $\alpha = 0.05$?

5. The following data provide the rates of three-word phrases from 8 articles written by Mark Twain and 10 articles by Snodgrass:

   Mark Twain
   0.225 0.262 0.217 0.240 0.230 0.229 0.235 0.217

   Snodgrass
   0.209 0.205 0.196 0.210 0.202 0.207 0.224 0.223 0.220 0.201

   Assume that the two samples were separately taken from two normal distributed populations which have the same variance and are independent to each other. Are the rates of three-word phrases significantly different between the two populations ($\alpha = 0.05$)?