

Multivariate Statistical Analysis

Lecture 08

Fudan University

luoluo@fudan.edu.cn

Outline

- 1 Consistency
- 2 Asymptotic Normality
- 3 Bayesian Estimation

Outline

- 1 Consistency
- 2 Asymptotic Normality
- 3 Bayesian Estimation

Consistency

A sequence of random vectors $\mathbf{t}_n = [t_{1n}, \dots, t_{pn}]^\top$ for $n = 1, 2, \dots$, is a consistent estimator of $\boldsymbol{\theta} = [\theta_1, \dots, \theta_p]^\top$ if

$$\text{plim}_{n \rightarrow +\infty} t_{in} = \theta_i$$

for $i = 1, \dots, p$.

The definition of convergence in probability says

$$\lim_{n \rightarrow +\infty} \Pr(|t_{in} - \theta_i| < \epsilon) = 1$$

holds for any $\epsilon > 0$.

The weak law of large numbers states that the sample means converges in probability towards the expected value.

For sample $\mathbf{x}_1, \mathbf{x}_2 \dots$ are independently and identically distributed with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, the estimators

$$\bar{\mathbf{x}}_N = \frac{1}{N} \sum_{\alpha=1}^N \mathbf{x}_{\alpha} \quad \text{and} \quad \mathbf{S}_N = \frac{1}{N-1} \sum_{\alpha=1}^N (\mathbf{x}_{\alpha} - \bar{\mathbf{x}}_N)(\mathbf{x}_{\alpha} - \bar{\mathbf{x}}_N)^{\top}$$

are consistent estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively.

Outline

- 1 Consistency
- 2 Asymptotic Normality
- 3 Bayesian Estimation

Asymptotic Normality

Let x_1, \dots, x_n be independent and identically distributed random variables with the same arbitrary distribution, mean μ , and variance σ^2 .

Let $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$, then the random variable

$$z = \lim_{n \rightarrow \infty} \sqrt{n} \left(\frac{\bar{x}_n - \mu}{\sigma} \right)$$

is a standard normal distribution.

What about multivariate case?

Asymptotic Normality

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n$$



Multivariate Central Limit Theorem

Theorem

Let p -component vectors $\mathbf{y}_1, \mathbf{y}_2, \dots$ be i.i.d with means $\mathbb{E}[\mathbf{y}_\alpha] = \boldsymbol{\nu}$ and covariance matrices $\mathbb{E}[(\mathbf{y}_\alpha - \boldsymbol{\nu})(\mathbf{y}_\alpha - \boldsymbol{\nu})^\top] = \mathbf{T}$. Then the limiting distribution of

$$\frac{1}{\sqrt{n}} \sum_{\alpha=1}^n (\mathbf{y}_\alpha - \boldsymbol{\nu})$$

as $n \rightarrow +\infty$ is $\mathcal{N}(\mathbf{0}, \mathbf{T})$.

Theorem

Let $\{F_j(\mathbf{x})\}$ be a sequence of cdfs, and let $\{\phi_j(\mathbf{t})\}$ be the sequence of corresponding characteristic functions. A necessary and sufficient condition for $F_j(\mathbf{x})$ to converge to a cdf $F(\mathbf{x})$ is that, for every \mathbf{t} , $\phi_j(\mathbf{t})$ converges to a limit $\phi(\mathbf{t})$ that is continuous at $\mathbf{t} = \mathbf{0}$. When this condition is satisfied, the limit $\phi(\mathbf{t})$ is identical with the characteristic function of the limiting distribution $F(\mathbf{x})$.

Outline

- 1 Consistency
- 2 Asymptotic Normality
- 3 Bayesian Estimation**

Revisiting Linear Regression

Given dataset $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, where $\mathbf{x}_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$ are the feature and the corresponding label of the i -th data.

We suppose

$$y_i = \boldsymbol{\beta}^\top \mathbf{x}_i + \epsilon_i$$

with

$$\boldsymbol{\beta} \in \mathbb{R}^p \quad \text{and} \quad \epsilon_i \stackrel{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$$

for $i = 1, \dots, N$, where $\sigma > 0$.

Revisiting Linear Regression

Maximizing the likelihood function leads to optimization problem

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{X}\beta - \mathbf{y}\|_2^2.$$

Suppose $\mathbf{X}^\top \mathbf{X}$ is non-singular, then

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y},$$

which has distribution

$$\hat{\beta} \sim \mathcal{N}_p(\beta, \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}).$$

Revisiting Linear Regression

We define the sample error as

$$\hat{\epsilon} = \mathbf{y} - \mathbf{X}\hat{\beta},$$

which is uncorrelated to $\hat{\beta}$.

Ridge Regression

In Bayesian statistics, we regard the parameters as a random variable with prior distribution.

For linear regression, we additionally suppose the parameter has a prior distribution

$$\boldsymbol{\beta} \sim \mathcal{N}_p(\mathbf{0}, \tau^2 \mathbf{I}),$$

which leads to optimization problem

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2^2 + \frac{\sigma^2}{2\tau^2} \|\boldsymbol{\beta}\|_2^2.$$

Theorem

If $\mathbf{x}_1, \dots, \mathbf{x}_N$ are independently distributed and each \mathbf{x}_α has distribution $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and if $\boldsymbol{\mu}$ has an a prior distribution $\mathcal{N}(\boldsymbol{\nu}, \boldsymbol{\Phi})$, then the a posterior distribution of $\boldsymbol{\mu}$ given $\mathbf{x}_1, \dots, \mathbf{x}_N$ is normal with mean

$$\boldsymbol{\Phi} \left(\boldsymbol{\Phi} + \frac{1}{N} \boldsymbol{\Sigma} \right)^{-1} \bar{\mathbf{x}} + \frac{1}{N} \boldsymbol{\Sigma} \left(\boldsymbol{\Phi} + \frac{1}{N} \boldsymbol{\Sigma} \right)^{-1} \boldsymbol{\nu}$$

and covariance matrix

$$\boldsymbol{\Phi} - \boldsymbol{\Phi} \left(\boldsymbol{\Phi} + \frac{1}{N} \boldsymbol{\Sigma} \right)^{-1} \boldsymbol{\Phi}.$$