

Multivariate Statistical Analysis

Lecture 06

Fudan University

luoluo@fudan.edu.cn

Outline

- 1 Maximum Likelihood Estimation
- 2 Distribution Theory

Outline

1 Maximum Likelihood Estimation

2 Distribution Theory

The Maximum Likelihood Estimators

If $\mathbf{x}_1, \dots, \mathbf{x}_N$ constitutes a sample from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $N > p$ and define

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}},$$

then what is the maximum likelihood estimator of ρ_{ij} ?

We can replace σ_{ii} , σ_{jj} and σ_{ij} with

$$\begin{cases} \hat{\sigma}_{ii} = \frac{1}{N} \sum_{\alpha=1}^N (x_{i\alpha} - \bar{x}_i)^2, \\ \hat{\sigma}_{ij} = \frac{1}{N} \sum_{\alpha=1}^N (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j), \\ \hat{\sigma}_{jj} = \frac{1}{N} \sum_{\alpha=1}^N (x_{j\alpha} - \bar{x}_j)^2, \end{cases}$$

leading to

$$\hat{\rho}_{ij} = \frac{\sum_{\alpha=1}^N (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)}{\sqrt{\sum_{\alpha=1}^N (x_{i\alpha} - \bar{x}_i)^2} \sqrt{\sum_{\alpha=1}^N (x_{j\alpha} - \bar{x}_j)^2}}.$$

The Maximum Likelihood Estimators

Theorem

On the basis of a given sample, if

$$\hat{\theta}_1, \dots, \hat{\theta}_m$$

are maximum likelihood estimators of the parameters

$$\theta_1, \dots, \theta_m$$

of a distribution, then

$$\phi_1(\hat{\theta}_1, \dots, \hat{\theta}_m), \dots, \phi_m(\hat{\theta}_1, \dots, \hat{\theta}_m)$$

are maximum likelihood estimator of

$$\phi_1(\theta_1, \dots, \theta_m), \dots, \phi_m(\theta_1, \dots, \theta_m)$$

if the transformation from $\theta_1, \dots, \theta_m$ to ϕ_1, \dots, ϕ_m is one-to-one.

The Maximum Likelihood Estimators

If $\phi : \mathcal{S} \rightarrow \mathcal{S}^*$ is not one-to-one, we let

$$\phi^{-1}(\theta^*) = \{\theta : \theta^* = \phi(\theta)\}.$$

and define (the induced likelihood function)

$$g(\theta^*) = \sup\{f(\theta) : \theta^* = \phi(\theta)\}.$$

If $\theta = \hat{\theta}$ maximize $f(\theta)$, then $\theta^* = \phi(\hat{\theta})$ also maximize $g(\theta^*)$.

The Maximum Likelihood Estimators

The maximum likelihood estimator of ρ_{ij} is indeed

$$\hat{\rho}_{ij} = \frac{\sum_{\alpha=1}^N (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)}{\sqrt{\sum_{\alpha=1}^N (x_{i\alpha} - \bar{x}_i)^2} \sqrt{\sum_{\alpha=1}^N (x_{j\alpha} - \bar{x}_j)^2}}.$$

Outline

1 Maximum Likelihood Estimation

2 Distribution Theory

Theorem

Let $\mathbf{x}_1, \dots, \mathbf{x}_N$ be independent, each distributed according to $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then the mean of the sample

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \frac{1}{N} \sum_{\alpha=1}^N \mathbf{x}_{\alpha}$$

is distributed according to $\mathcal{N}(\boldsymbol{\mu}, \frac{1}{N} \boldsymbol{\Sigma})$ and independent of

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \sum_{\alpha=1}^N (\mathbf{x}_{\alpha} - \bar{\mathbf{x}})(\mathbf{x}_{\alpha} - \bar{\mathbf{x}})^{\top}.$$

Additionally, we have

$$N\hat{\boldsymbol{\Sigma}} = \sum_{\alpha=1}^{N-1} \mathbf{z}_{\alpha} \mathbf{z}_{\alpha}^{\top},$$

where $\mathbf{z}_{\alpha} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ for $\alpha = 1, \dots, N-1$, and $\mathbf{z}_1, \dots, \mathbf{z}_{N-1}$ are independent.

Lemma

Suppose $\mathbf{x}_1, \dots, \mathbf{x}_N$ are independent, where $\mathbf{x}_\alpha \sim \mathcal{N}_p(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma})$. Let $\mathbf{C} \in \mathbb{R}^{N \times N}$ be an orthogonal matrix, then

$$\mathbf{y}_\alpha = \sum_{\beta=1}^N c_{\alpha\beta} \mathbf{x}_\beta \sim \mathcal{N}_p(\boldsymbol{\nu}_\alpha, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\nu} = \sum_{\beta=1}^N c_{\alpha\beta} \boldsymbol{\mu}_\beta$ for $\alpha = 1, \dots, N$ and $\mathbf{y}_1, \dots, \mathbf{y}_N$ are independent.

Lemma

$$\text{If } \mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \dots & c_{pp} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1^\top \\ \mathbf{c}_2^\top \\ \vdots \\ \mathbf{c}_p^\top \end{bmatrix} \in \mathbb{R}^{p \times p} \text{ is orthogonal,}$$

then $\sum_{\alpha=1}^N \mathbf{x}_\alpha \mathbf{x}_\alpha^\top = \sum_{\beta=1}^N \mathbf{y}_\beta \mathbf{y}_\beta^\top$ where $\mathbf{y}_\alpha = \sum_{\beta=1}^N c_{\alpha\beta} \mathbf{x}_\beta$ for $\alpha = 1, \dots, N$.

Theorem

If $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ constitute a sample from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $N > p$, the estimator

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \sum_{\alpha=1}^N (\mathbf{x}_{\alpha} - \bar{\mathbf{x}})(\mathbf{x}_{\alpha} - \bar{\mathbf{x}})^{\top}$$

is positive definite with probability 1.

- 1 The matrix $\hat{\boldsymbol{\Sigma}}$ be must singular if $N \leq p$.
- 2 The proof indicates $\mathbf{U}^{\top} \mathbf{U}$ is non-singular with probability 1 for $\mathbf{U} \in \mathbb{R}^{d \times k}$ with $k \leq d$ and $u_{ij} \stackrel{i.i.d}{\sim} \mathcal{N}(0, 1)$.