Multivariate Statistical Analysis

Lecture 11

Fudan University

luoluo@fudan.edu.cn

Outline

Outline

The Likelihood Ratio Criterion

The likelihood ratio criterion:

1. Let $L(\mathbf{x}, \boldsymbol{\theta})$ be the likelihood function of the observation $\mathbf{x}$ and the parameter vector $\boldsymbol{\theta} \in \Omega$.
2. Let a null hypothesis be defined by a proper subset ω of Ω. The likelihood ratio criterion is

$$\lambda(\mathbf{x}) = \frac{\sup_{\boldsymbol{\theta} \in \omega} L(\mathbf{x}, \boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Omega} L(\mathbf{x}, \boldsymbol{\theta})}.$$

3. The likelihood ratio test is the procedure of rejecting the null hypothesis when $\lambda(\mathbf{x})$ is less than a predetermined constant.

Test $\rho = \rho_0$ by the Likelihood Ratio Criterion

We consider the likelihood ratio test of the hypothesis that $\rho = \rho_0$ based on a sample $\mathbf{x}_1, \dots, \mathbf{x}_N$ from

$$\mathcal{N}_2\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2\end{bmatrix}\right).$$

Define the set

$$\Omega = \left\{(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) : \boldsymbol{\mu} \in \mathbb{R}^2, \sigma_1 > 0, \sigma_2 > 0, \boldsymbol{\Sigma} \succ \mathbf{0}\right\}$$

and its subset

$$\omega = \left\{(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) : \boldsymbol{\mu} \in \mathbb{R}^2, \sigma_1 > 0, \sigma_2 > 0, \boldsymbol{\Sigma} \succ \mathbf{0}, \rho = \rho_0\right\}.$$

We also follow the notation

$$r = \frac{a_{12}}{\sqrt{a_{11}}\sqrt{a_{22}}}, \quad \mathbf{A} = \sum_{\alpha=1}^{N}(\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})^\top \quad \text{and} \quad \bar{\mathbf{x}} = \frac{1}{N}\sum_{\alpha=1}^{N}\mathbf{x}_\alpha.$$

Test $\rho = \rho_0$ by the Likelihood Ratio Criterion

The likelihood ratio criterion is

$$\frac{\sup_{\omega} L(\mathbf{x}, \boldsymbol{\theta})}{\sup_{\Omega} L(\mathbf{x}, \boldsymbol{\theta})} = \left(\frac{(1-\rho_0^2)(1-r^2)}{(1-\rho_0 r)^2} \right)^{\frac{N}{2}}.$$

The likelihood ratio test is

$$\frac{(1-\rho_0^2)(1-r^2)}{(1-\rho_0 r)^2} \leq c$$

where c is chosen by the prescribed significance level.

The Maximum Likelihood Estimators

Let $\phi : \mathcal{S} \to \mathcal{S}^*$ (may be not one-to-one) and

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

and define (the induced likelihood function)

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

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

Test $\rho = \rho_0$ by the Likelihood Ratio Criterion

The critical region can be written equivalently as

$$(\rho_0^2 c - \rho_0^2 + 1)r^2 - 2\rho_0 cr + c - 1 + \rho_0^2 \geq 0,$$

that is,

$$r > \frac{\rho_0 c + (1-\rho_0^2)\sqrt{1-c}}{\rho_0^2 c - \rho_0^2 + 1} \quad \text{and} \quad r < \frac{\rho_0 c - (1-\rho_0^2)\sqrt{1-c}}{\rho_0^2 c - \rho_0^2 + 1}.$$

Thus the likelihood ratio test of $H : \rho = \rho_0$ against alternatives $\rho \neq \rho_0$ has a rejection region of the form $r > r_1$ and $r < r_2$.

Outline

The Asymptotic Distribution of Sample Correlation

For a sample $\mathbf{x}_1, \dots, \mathbf{x}_N$ from a normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we are interested in the asymptotic behavior of sample correlation coefficient

$$r(n) = \frac{a_{ij}(n)}{\sqrt{a_{ii}(n)}\sqrt{a_{jj}(n)}}$$

where $n = N - 1$,

$$a_{ij}(n) = \sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) = \sum_{\alpha=1}^{n}\begin{bmatrix} z_{i\alpha} \\ z_{j\alpha} \end{bmatrix}\begin{bmatrix} z_{i\alpha} & z_{j\alpha} \end{bmatrix}$$

with

$$\begin{bmatrix} z_{i\alpha} \\ z_{j\alpha} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ji} & \sigma_{jj} \end{bmatrix}\right) \quad \text{and} \quad \bar{x}_i = \frac{1}{N}\sum_{\alpha=1}^{N} x_{i\alpha}.$$

The Asymptotic Distribution of Sample Correlation

Theorem

Let

$$\mathbf{A}(n) = \sum_{\alpha=1}^{N} \left(\mathbf{x}_\alpha - \bar{\mathbf{x}}_N\right)\left(\mathbf{x}_\alpha - \bar{\mathbf{x}}_N\right)^\top,$$

where $\mathbf{x}_1, \dots, \mathbf{x}_N$ *are independently distributed according to* $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and* $n = N - 1$*. Then the limiting distribution of*

$$\mathbf{B}(n) = \frac{1}{\sqrt{n}}\left(\mathbf{A}(n) - n\boldsymbol{\Sigma}\right)$$

is normal with mean $\mathbf{0}$ *and covariance of the entries of* $\mathbf{B}(n)$ *is*

$$\mathbb{E}\left[b_{ij}(n) b_{kl}(n)\right] = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}.$$

The Asymptotic Distribution of Sample Correlation

We achieve

$$\lim_{n\to\infty} \frac{\sqrt{n}(r(n)-\rho)}{1-\rho^2} \sim \mathcal{N}(0,1).$$

by applying the following theorem.

Theorem (Serfling (1980), Section 3.3)

Let $\{\mathbf{u}(n)\}$ be a sequence of m-component random vectors and $\mathbf{b}$ a fixed vector such that

$$\lim_{n\to\infty} \sqrt{n}(\mathbf{u}(n)-\mathbf{b}) \sim \mathcal{N}(\mathbf{0},\mathbf{T}).$$

Let $\mathbf{f}(\mathbf{u})$ be a vector-valued function of $\mathbf{u}$ such that each component $f_j(\mathbf{u})$ has a nonzero differential at $\mathbf{u}=\mathbf{b}$, and define $\mathbf{\Phi_b}$ with its (i,j)-th component being

$$\frac{\partial f_j(\mathbf{u})}{\partial u_i}\bigg|_{\mathbf{u}=\mathbf{b}}.$$

Then $\sqrt{n}(\mathbf{f}(\mathbf{u}(n))-f(\mathbf{b}))$ has the limiting distribution $\mathcal{N}(\mathbf{0},\mathbf{\Phi_b}^\top\mathbf{T}\mathbf{\Phi_b})$.

Outline

John Wishart

John Wishart (November 28th, 1898 – July 14th, 1956)

The Wishart Distribution

We consider the distribution of

$$\mathbf{A} = \sum_{\alpha=1}^{N} (\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})^\top,$$

where $\mathbf{x}_1, \dots, \mathbf{x}_N$ are independent, each with the distribution $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $N > p$.

The distribution of $\mathbf{A}$ is often called Wishart distribution with n degrees of freedom and scale parameter $\boldsymbol{\Sigma}$, written as

$$\mathbf{A} \sim \mathcal{W}_p(\boldsymbol{\Sigma}, n) \qquad \text{where} \quad n = N - 1 \geq p.$$

The Wishart Distribution

We can write

$$\mathbf{A} = \sum_{\alpha=1}^{n} \mathbf{z}_\alpha^\top \mathbf{z}_\alpha,$$

where $\mathbf{z}_1, \dots, \mathbf{z}_n$ are independent, each with the distribution $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ and $N = n - 1$.

For $p = 1$, we have

$$a \sim \mathcal{W}_1(\sigma^2, n) \qquad \text{and} \qquad \frac{a}{\sigma^2} \sim \chi^2(n).$$

Properties of Wishart Distribution

Theorem

Let $\mathbf{A} \sim \mathcal{W}_p(\mathbf{\Sigma}, n)$ *and* $\mathbf{C} \in \mathbb{R}^{q \times p}$*, then*

$$\mathbf{C}\mathbf{A}\mathbf{C}^\top \sim \mathcal{W}_p(\mathbf{C}\mathbf{\Sigma}\mathbf{C}^\top, n)$$

For any $\mathbf{t} \in \mathbb{R}^p$, we have

$$\mathbf{t}^\top\mathbf{A}\mathbf{t} \sim \mathcal{W}_1(\mathbf{t}^\top\mathbf{\Sigma}\mathbf{t}, n) \qquad \text{and} \qquad \frac{\mathbf{t}^\top\mathbf{A}\mathbf{t}}{\mathbf{t}^\top\mathbf{\Sigma}\mathbf{t}} \sim \chi^2(n).$$

Properties of Wishart Distribution

Theorem

If $\mathbf{A}_1, \ldots, \mathbf{A}_k$ *are independently distributed with* $\mathbf{A}_i \sim \mathcal{W}(\mathbf{\Sigma}, n_i)$ *for* $i = 1, \ldots, k$, *then*

$$\mathbf{A} = \sum_{i=1}^{k} \mathbf{A}_i \sim \mathcal{W}\left(\mathbf{\Sigma}, \sum_{i=1}^{k} n_i\right).$$

Properties of Wishart Distribution

Theorem

Let

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1^\top \\ \vdots \\ \mathbf{z}_n^\top \end{bmatrix} \in \mathbb{R}^{n \times p}$$

where $\mathbf{z}_1, \dots, \mathbf{z}_n$ *are independent, each with the distribution*

$$\mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma}).$$

For projection matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ *with rank-*r*, we have*

$$\mathbf{Z}^\top \mathbf{Q} \mathbf{Z} \sim \mathcal{W}_p(\mathbf{\Sigma}, r).$$

The density of $\mathbf{A} \sim \mathcal{W}_p(\mathbf{\Sigma}, n)$ is

$$\frac{(\det(\mathbf{A}))^{\frac{n-p-1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{A}\right)\right)}{2^{\frac{np}{2}}\pi^{\frac{p(p-1)}{4}}(\det(\mathbf{\Sigma}))^{\frac{n}{2}}\prod_{i=1}^{p}\Gamma\left(\frac{1}{2}(n+1-i)\right)}.$$

for positive definite $\mathbf{A}$ and 0 elsewhere.

Quiz 1

Let $\mathbf{A} \in \mathbb{R}^{p\times p}$ and define $\mathbf{F} : \mathbb{R}^{p\times q} \to \mathbb{R}^{p\times q}$ as

$$\mathbf{F}(\mathbf{X}) = \mathbf{A}\mathbf{X}.$$

What is the determinant of Jacobian of $\mathbf{F}(\mathbf{X})$?

Quiz 1.5

Let $\mathbf{A} \in \mathbb{R}^{p\times p}, \mathbf{B} \in \mathbb{R}^{q\times q}$ and define $\mathbf{F} : \mathbb{R}^{p\times q} \to \mathbb{R}^{p\times q}$ as

$$\mathbf{F}(\mathbf{X}) = \mathbf{A}\mathbf{X}\mathbf{B}.$$

What is the determinant of Jacobian of $\mathbf{F}(\mathbf{X})$?

Quiz 2

Let $\mathbf{A} \in \mathbb{R}^{p \times p}$ be non-singular and define $\mathbf{F} : \mathbb{S}^p \to \mathbb{R}^{p \times p}$ as

$$\mathbf{F}(\mathbf{X}) = \mathbf{A}\mathbf{X}\mathbf{A}^\top,$$

where $\mathbb{S}^p = \{\mathbf{X} \in \mathbb{R}^{p \times p} : \mathbf{X} = \mathbf{X}^\top\}$.

What is the determinant of Jacobian of $\mathbf{F}(\mathbf{X})$?