Multivariate Statistical Analysis

Lecture 14

Fudan University

luoluo@fudan.edu.cn

Outline

Outline

Likelihood Ratio Criterion and T^2-Statistic

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ constitute a sample from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $N > p$.

We shall derive T^2-Statistic

$$T^2 = N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$$

from likelihood ratio criterion

$$\lambda = \frac{\max_{\boldsymbol{\Sigma} \in \mathbb{S}_p^{++}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} \in \mathbb{S}_p^{++}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}.$$

We have

$$\lambda^{\frac{2}{N}} = \frac{1}{1 + T^2/(N-1)},$$

where

$$T^2 = N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0), \qquad \bar{\mathbf{x}} = \frac{1}{N}\sum_{\alpha=1}^{N} \mathbf{x}_\alpha$$

and

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

Likelihood Ratio Criterion and T^2-Statistic

The condition $\lambda^{2/N} > c$ for some $c \in (0, 1)$ is equivalent to

$$T^2 < \frac{(N-1)(1-c)}{c}.$$

Outline

Multivariate Analysis of Variance

We consider testing the equality of means with common covariance.

Let $\mathbf{x}_\alpha^{(g)}$ be an observation from the g-th population $\mathcal{N}_p(\boldsymbol{\mu}^{(g)}, \boldsymbol{\Sigma})$ for $\alpha = 1, \dots, N_g$ and $g = 1, \dots, q$. We wish to test the hypothesis

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_g.$$

Multivariate Analysis of Variance

The likelihood function is

$$
\begin{aligned}
&L(\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(g)}, \boldsymbol{\Sigma}) \\
=& \prod_{g=1}^{q} \frac{1}{(2\pi)^{\frac{pN_g}{2}} (\det(\boldsymbol{\Sigma}))^{\frac{N_g}{2}}} \exp\left(-\frac{1}{2} \sum_{\alpha=1}^{N_g} \left(\mathbf{x}_\alpha^{(g)} - \boldsymbol{\mu}^{(g)}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\mathbf{x}_\alpha^{(g)} - \boldsymbol{\mu}^{(g)}\right) \right).
\end{aligned}
$$

1. We let $\boldsymbol{\theta} = \{\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(g)}, \boldsymbol{\Sigma}\}$ be the parameters.
2. The set Ω is the space in which $\boldsymbol{\Sigma}$ is positive definite and each $\boldsymbol{\mu}^{(g)}$ is any p-dimensional vector.
3. The set ω is the space in which $\boldsymbol{\mu}^{(1)} = \cdots = \boldsymbol{\mu}^{(g)}$ (p-dimensional vectors) and $\boldsymbol{\Sigma}$ is positive definite matrix.

Multivariate Analysis of Variance

The likelihood ratio criterion is

$$\lambda = \frac{\sup_{\boldsymbol{\theta} \in \omega} L(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Omega} L(\boldsymbol{\theta})} = \frac{(\det(\hat{\boldsymbol{\Sigma}}_\Omega))^{\frac{N}{2}}}{(\det(\hat{\boldsymbol{\Sigma}}_\omega))^{\frac{N}{2}}},$$

where

$$\hat{\boldsymbol{\Sigma}}_\Omega = \frac{1}{N} \sum_{g=1}^{q} \sum_{\alpha=1}^{N_g} \left(\mathbf{x}_\alpha^{(g)} - \bar{\mathbf{x}}^{(g)}\right)\left(\mathbf{x}_\alpha^{(g)} - \bar{\mathbf{x}}^{(g)}\right)^\top$$

and

$$\hat{\boldsymbol{\Sigma}}_\omega = \frac{1}{N} \sum_{g=1}^{q} \sum_{\alpha=1}^{N_g} \left(\mathbf{x}_\alpha^{(g)} - \bar{\mathbf{x}}\right)\left(\mathbf{x}_\alpha^{(g)} - \bar{\mathbf{x}}\right)^\top.$$

Multivariate Analysis of Variance

We can write

$$N\hat{\mathbf{\Sigma}}_\omega = \mathbf{A} + \mathbf{B},$$

where

$$\mathbf{A} = N\hat{\mathbf{\Sigma}}_\Omega = \sum_{g=1}^{q}\sum_{\alpha=1}^{N_g}\left(\mathbf{x}_\alpha^{(g)} - \bar{\mathbf{x}}^{(g)}\right)\left(\mathbf{x}_\alpha^{(g)} - \bar{\mathbf{x}}^{(g)}\right)^\top \sim \mathcal{W}_p(\mathbf{\Sigma}, N-q)$$

and

$$\mathbf{B} = \sum_{g=1}^{q} N_g(\bar{\mathbf{x}}^{(g)} - \bar{\mathbf{x}})(\bar{\mathbf{x}}^{(g)} - \bar{\mathbf{x}})^\top \sim \mathcal{W}_p(\mathbf{\Sigma}, q-1)$$

are independent.

Wilks' Lambda distribution

For two independent random matrices $\mathbf{A} \sim \mathcal{W}_p(\mathbf{\Sigma}, n)$ and $\mathbf{B} \sim \mathcal{W}_p(\mathbf{\Sigma}, m)$ with $n \geq p$, the ratio

$$\frac{\det(\mathbf{A})}{\det(\mathbf{A} + \mathbf{B})}$$

has Wilks' Lambda distribution with degrees of freedom n and m, which is typically written as

$$\frac{\det(\mathbf{A})}{\det(\mathbf{A} + \mathbf{B})} \sim \Lambda_{p,n,m}.$$

Wilks' Lambda distribution

Theorem

Let $\mathbf{A} \sim \mathcal{W}_p(\mathbf{\Sigma}, n)$ *and* $\mathbf{B} \sim \mathcal{W}_p(\mathbf{\Sigma}, m)$ *be two independent Wishart distributed variables, then we can write*

$$\frac{\det(\mathbf{A})}{\det(\mathbf{A}+\mathbf{B})} = \prod_{i=1}^{p} u_i \sim \Lambda_{p,n,m},$$

where $u_1, \dots, u_p$ *are independent distributed as*

$$u_i \sim \mathrm{Beta}\left(\frac{n+1-i}{2}, \frac{m}{2}\right).$$

Let $\mathbf{A} \sim \mathcal{W}_p(\mathbf{\Sigma}, n)$ and partition $\mathbf{A}$ and $\mathbf{\Sigma}$ into q and $p-q$ rows and columns as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \qquad \text{and} \qquad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix},$$

then we have

(a) $\mathbf{A}_{11} \sim \mathcal{W}_q(\mathbf{\Sigma}_{11}, n)$ and $\mathbf{A}_{22} \sim \mathcal{W}_{p-q}(\mathbf{\Sigma}_{22}, n)$;

(b) if $q = 1$, then

$$\mathbf{a}_{21} \mid \mathbf{A}_{22} \sim \mathcal{N}_{p-q}(\mathbf{A}_{22}\mathbf{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}, \sigma_{11.2}^2\mathbf{A}_{22})$$

where $\sigma_{11.2}^2 = \sigma_{11} - \boldsymbol{\sigma}_{21}^\top\mathbf{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$;

(c) if $n > p - q$, then

$$\mathbf{A}_{11.2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \sim \mathcal{W}_q(\mathbf{\Sigma}_{11.2}, n-p+q)$$

is independent on $\mathbf{A}_{22}$ and $\mathbf{A}_{12}$, where $\mathbf{\Sigma}_{11.2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$.

Quiz

Let $\mathbf{A} \sim \mathcal{W}_p(\mathbf{\Sigma}, n)$, we can follow above theorem to show

$$\det(\mathbf{A}) = \det(\mathbf{\Sigma}) \prod_{i=1}^{p} v_i$$

with some independent random variables $v_1, \dots, v_p$?

Outline

Multivariate Linear Regression

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

We suppose

$$\mathbf{y}_i = \mathbf{B}^\top \mathbf{x}_i + \boldsymbol{\epsilon}_i \quad \text{with} \quad \mathbf{B} \in \mathbb{R}^{p\times q} \qquad \text{and} \qquad \boldsymbol{\epsilon}_i \overset{\text{i.i.d}}{\sim} \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma})$$

for $i = 1, \dots, N$, $\boldsymbol{\Sigma} \succ 0$ and $N > p$.

We regard $\mathbf{B} \in \mathbb{R}^{p\times q}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{q\times q}$ as parameters, then

$$\boldsymbol{\epsilon}_i = \mathbf{y}_i - \mathbf{B}^\top \mathbf{x}_i \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}).$$

Multivariate Linear Regression

We denote

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_N^\top \end{bmatrix} \in \mathbb{R}^{N\times p}, \quad \mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^\top \\ \vdots \\ \mathbf{y}_N^\top \end{bmatrix} \in \mathbb{R}^{N\times q} \quad \text{and} \quad \mathbf{E} = \begin{bmatrix} \boldsymbol{\epsilon}_1^\top \\ \vdots \\ \boldsymbol{\epsilon}_N^\top \end{bmatrix} \in \mathbb{R}^{N\times q},$$

and suppose $\mathbf{X}$ is full rank.

MLE for Multivariate Linear Regression

We construct the likelihood function for $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_N$ as follows

$$
\begin{aligned}
& L(\mathbf{B}, \boldsymbol{\Sigma}) \\
=& \prod_{\alpha=1}^{N} \frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{B}^\top \mathbf{x}_\alpha - \mathbf{y}_\alpha)^\top \boldsymbol{\Sigma}^{-1}(\mathbf{B}^\top \mathbf{x}_\alpha - \mathbf{y}_\alpha)\right) \\
=& \frac{1}{(2\pi)^{Np/2}(\det(\boldsymbol{\Sigma}))^{N/2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left((\mathbf{XB} - \mathbf{Y})\boldsymbol{\Sigma}^{-1}(\mathbf{XB} - \mathbf{Y})^\top\right)\right).
\end{aligned}
$$

The maximum likelihood estimators are

$$
\hat{\mathbf{B}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y} \qquad \text{and} \qquad \hat{\boldsymbol{\Sigma}} = \frac{1}{N}\mathbf{Y}^\top(\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top)\mathbf{Y}.
$$

MLE for Multivariate Linear Regression

We write

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_q \end{bmatrix} \in \mathbb{R}^{q \times p} \quad \text{and} \quad \hat{\mathbf{B}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 & \cdots & \hat{\boldsymbol{\beta}}_q \end{bmatrix} \in \mathbb{R}^{q \times p}.$$

Then the joint distribution of $\hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_N$ is normal and we have

1. $\mathbb{E}[\hat{\boldsymbol{\beta}}_i] = \boldsymbol{\beta}_i$;
2. $\mathrm{Cov}[\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\beta}}_j] = \sigma_{ij}(\mathbf{X}^\top \mathbf{X})^{-1}$;
3. $\hat{\boldsymbol{\Sigma}} \sim \mathcal{W}_q\left(\frac{1}{N}\boldsymbol{\Sigma}, N - p\right)$.

Bayesian Multivariate Linear Regression

We can additionally suppose each b_{ij} independently follows

$$b_{ij} \sim \mathcal{N}(0, \tau^2),$$

then the posterior likelihood function is

$$\begin{aligned}
& L(\mathbf{B}, \mathbf{\Sigma}) \\
=& \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^p \det(\mathbf{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{B}^\top \mathbf{x}_i - \mathbf{y}_i)^\top \mathbf{\Sigma}^{-1} (\mathbf{B}^\top \mathbf{x}_i - \mathbf{y}_i)\right) \\
& \cdot \prod_{i=1}^{p} \prod_{j=1}^{q} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{b_{ij}^2}{2\tau^2}\right) \\
\propto & \frac{1}{(\det(\mathbf{\Sigma}))^{N/2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left((\mathbf{XB} - \mathbf{Y})\mathbf{\Sigma}^{-1}(\mathbf{XB} - \mathbf{Y})^\top\right) - \frac{1}{2\tau^2}\left\|\mathbf{B}\right\|_F^2\right),
\end{aligned}$$

which leads to solving Sylvester equation.

We typically suppose

$$\boldsymbol{\beta}_{(i)} \overset{\text{i.i.d}}{\sim} \mathcal{N}_q(\mathbf{0}, \tau^2\boldsymbol{\Sigma}), \qquad \text{where} \qquad \mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_{(1)}^\top \\ \vdots \\ \boldsymbol{\beta}_{(p)}^\top \end{bmatrix} \in \mathbb{R}^{p\times q},$$

then the posterior likelihood function is

$$\begin{aligned} & L(\mathbf{B}, \boldsymbol{\Sigma}) \\ = & \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{B}^\top \mathbf{x}_i - \mathbf{y}_i)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{B}^\top \mathbf{x}_i - \mathbf{y}_i)\right) \\ & \cdot \prod_{j=1}^{p} \frac{1}{\sqrt{(2\pi)^q \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2\tau^2} \boldsymbol{\beta}_{(j)}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}_{(j)}\right) \\ \propto & \frac{1}{(\det(\boldsymbol{\Sigma}))^{N/2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left((\mathbf{XB} - \mathbf{Y})\boldsymbol{\Sigma}^{-1}(\mathbf{XB} - \mathbf{Y})^\top\right) - \frac{1}{2\tau^2}\mathbf{B}\boldsymbol{\Sigma}^{-1}\mathbf{B}^\top\right). \end{aligned}$$

Bayesian Multivariate Linear Regression

We have

$$\hat{\mathbf{B}} = (\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{Y},$$

and

$$\hat{\mathbf{\Sigma}} = \frac{1}{N}\left((\mathbf{XB} - \mathbf{Y})^\top(\mathbf{XB} - \mathbf{Y}) + \lambda\mathbf{B}^\top\mathbf{B}\right),$$

where $\lambda = 1/\tau^2$.