

Multivariate Statistical Analysis

Lecture 02

Fudan University

luoluo@fudan.edu.cn

- 1 Random Vectors and Matrices
- 2 Random Samples
- 3 Generalized Variance

Outline

1 Random Vectors and Matrices

2 Random Samples

3 Generalized Variance

Random Vectors and Matrices

- ① A random matrix (vector) is a matrix (vector) whose elements are random variables.
- ② The expected value of a random matrix (or vector) is the matrix (vector) consisting of the expected values of each of its elements.
- ③ Let \mathbf{X} be an $m \times n$ random matrix, then its expected value, denoted by $\mathbb{E}[\mathbf{X}]$, is the $m \times n$ matrix of numbers (if they exist)

$$\mathbb{E}[\mathbf{X}] = \begin{bmatrix} \mathbb{E}[x_{11}] & \mathbb{E}[x_{12}] & \dots & \mathbb{E}[x_{1n}] \\ \mathbb{E}[x_{21}] & \mathbb{E}[x_{22}] & \dots & \mathbb{E}[x_{2n}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[x_{m1}] & \mathbb{E}[x_{m2}] & \dots & \mathbb{E}[x_{mn}] \end{bmatrix}.$$

Expectation of Random Matrices

Let \mathbf{X} and \mathbf{Y} be random matrices of the same dimension, and let \mathbf{A} and \mathbf{B} be conformable matrices of constants. Then we have

$$\mathbb{E}[\mathbf{X} + \mathbf{Y}] = \mathbb{E}[\mathbf{X}] + \mathbb{E}[\mathbf{Y}]$$

and

$$\mathbb{E}[\mathbf{AXB}] = \mathbf{A}\mathbb{E}[\mathbf{X}]\mathbf{B}.$$

Random Vector and Covariance Matrix

For random vector $\mathbf{x} = [x_1, \dots, x_p]^\top$, we denote $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}]$.

The expected value of the random matrix $(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top$ is

$$\text{Cov}[\mathbf{x}] = \mathbb{E} \left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top \right],$$

the covariance or covariance matrix of \mathbf{x} .

- 1 The i -th diagonal element of this matrix, $\mathbb{E}[(x_i - \mu_i)^2]$, is the variance of x_i .
- 2 The i, j -th off-diagonal element ($i \neq j$), $\mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)]$ is the covariance of x_i and x_j .
- 3 We have $\text{Cov}[\mathbf{x}] = \mathbb{E}[\mathbf{x}\mathbf{x}^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top$.

Theorem

Let $\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{f}$, where

- ① \mathbf{D} is an $n \times p$ constant matrix,
- ② \mathbf{x} is a p -dimensional random vector,
- ③ and \mathbf{f} is a n -dimensional constant vector.

Then we have

$$\mathbb{E}[\mathbf{y}] = \mathbf{D}\mathbb{E}[\mathbf{x}] + \mathbf{f} \quad \text{and} \quad \text{Cov}[\mathbf{y}] = \text{Cov}[\mathbf{D}\mathbf{x}] = \mathbf{D}\text{Cov}[\mathbf{x}]\mathbf{D}^T.$$

Example

Let $\mathbf{x} = [x_1, x_2]^\top$ be a random vector with

$$\mathbb{E}[\mathbf{x}] = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \text{and} \quad \text{Cov}[\mathbf{x}] = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}.$$

Let $\mathbf{z} = [z_1, z_2]^\top$ such that $z_1 = x_1 - x_2$ and $z_2 = x_1 + x_2$.

- 1 Find the $\mathbb{E}[\mathbf{z}]$ and $\text{Cov}[\mathbf{z}]$.
- 2 Find the condition that leads to z_1 and z_2 be uncorrelated.

For random vector $\mathbf{x} = [x_1, \dots, x_p]^\top$, we write its covariance as

$$\text{Cov}[\mathbf{x}] = \mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_{pp} \end{bmatrix}.$$

The correlation coefficient ρ_{ij} is defined as

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

which measures linear association between x_i and x_j .

The population correlation matrix of \mathbf{x} is defined as

$$\begin{aligned}\boldsymbol{\rho} &= \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}\sigma_{11}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sigma_{pp}}} \\ \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sqrt{\sigma_{pp}\sigma_{11}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}\sigma_{pp}}} \end{bmatrix} \\ &= \begin{bmatrix} 1 & \cdots & \rho_{1p} \\ \vdots & \ddots & \vdots \\ \rho_{p1} & \cdots & 1 \end{bmatrix}.\end{aligned}$$

Transformation of Variables

Let the density of p -dimensional random vector $\mathbf{x} = [x_1, \dots, x_p]^\top$ be $f(\mathbf{x})$.

Consider the p -dimensional random vector $\mathbf{y} = [y_1, \dots, y_p]^\top$ such that $y_i = u_i(\mathbf{x})$ for $i = 1, \dots, p$. Let the density function of \mathbf{y} be $g(\mathbf{y})$.

Assume the transformation $\mathbf{u}(\mathbf{x}) = [u_1(\mathbf{x}), \dots, u_p(\mathbf{x})]^\top : \mathbb{R}^p \rightarrow \mathbb{R}^p$ from the space of \mathbf{x} to the space of \mathbf{y} is smooth and one-to-one.

Then we have $f(\mathbf{x}) = g(\mathbf{u}(\mathbf{x})) |\det(\mathbf{J}(\mathbf{x}))|$ where

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \frac{\partial u_1(\mathbf{x})}{\partial x_1} & \frac{\partial u_1(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial u_1(\mathbf{x})}{\partial x_p} \\ \frac{\partial u_2(\mathbf{x})}{\partial x_1} & \frac{\partial u_2(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial u_2(\mathbf{x})}{\partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_p(\mathbf{x})}{\partial x_1} & \frac{\partial u_p(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial u_p(\mathbf{x})}{\partial x_p} \end{bmatrix}.$$

Transformation of Variables

Similarly, we also have $g(\mathbf{y}) = f(\mathbf{u}^{-1}(\mathbf{y}))|\det(\mathbf{J}^{-1}(\mathbf{y}))|$ where

$$\mathbf{J}^{-1}(\mathbf{y}) = \begin{bmatrix} \frac{\partial u_1^{-1}(\mathbf{y})}{\partial y_1} & \frac{\partial u_1^{-1}(\mathbf{y})}{\partial y_2} & \cdots & \frac{\partial u_1^{-1}(\mathbf{y})}{\partial y_p} \\ \frac{\partial u_2^{-1}(\mathbf{y})}{\partial y_1} & \frac{\partial u_2^{-1}(\mathbf{y})}{\partial y_2} & \cdots & \frac{\partial u_2^{-1}(\mathbf{y})}{\partial y_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_p^{-1}(\mathbf{y})}{\partial y_1} & \frac{\partial u_p^{-1}(\mathbf{y})}{\partial y_2} & \cdots & \frac{\partial u_p^{-1}(\mathbf{y})}{\partial y_p} \end{bmatrix}.$$

Outline

1 Random Vectors and Matrices

2 Random Samples

3 Generalized Variance

Random Samples

We use the notation $x_{\alpha j}$ to indicate the value of the α -th variable that is observed on the j -th item, or trial.

We display the N measurements on p variables as the $N \times p$ matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{\alpha 1} & x_{\alpha 2} & \dots & x_{\alpha j} & \dots & x_{\alpha p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nj} & \dots & x_{Np} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_i^\top \\ \vdots \\ \mathbf{x}_N^\top \end{bmatrix}.$$

We mainly focus on the following case.

- 1 The random p variables in a single trial, such as $\mathbf{x}_i = [x_{i1}, \dots, x_{ip}]^\top$ will usually be correlated, and the ones from different trials be independent. (It may not hold when the variables are likely to drift over time.)

Sample Mean and Covariance

Let $\mathbf{x}_1, \dots, \mathbf{x}_N$ be a random sample from a joint distribution that has mean vector $\boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Sigma}$. Then the sample means

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

is an unbiased estimator of $\boldsymbol{\mu}$, and its covariance matrix is

$$\text{Cov}[\bar{\mathbf{x}}] = \frac{1}{N} \boldsymbol{\Sigma}.$$

However, the matrix

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

is a biased estimator of $\boldsymbol{\Sigma}$.

Sample Covariance

We define the sample (variance-covariance) covariance matrix as

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

which is an unbiased estimator of $\mathbf{\Sigma}$.

Let $\mathbf{1}_N = [1, \dots, 1]^{\top} \in \mathbb{R}^N$, then we have

$$\mathbf{S} = \frac{1}{N-1} \mathbf{X}^{\top} \left(\mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^{\top} \right) \mathbf{X} \quad (2)$$

$$= \frac{1}{N-1} \left(\mathbf{X}^{\top} \mathbf{X} - \frac{1}{N} \mathbf{X}^{\top} \mathbf{1}_N \mathbf{1}_N^{\top} \mathbf{X} \right). \quad (3)$$

It provides a more efficient implementation.

Sample Correlation

Given sample covariance matrix

$$\mathbf{S} = \begin{bmatrix} s_{11} & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_{pp} \end{bmatrix} \in \mathbb{R}^{p \times p},$$

we define the sample correlation matrix as

$$\mathbf{R} = \begin{bmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{p1} & \cdots & r_{pp} \end{bmatrix} \in \mathbb{R}^{p \times p},$$

$$\text{where } r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}.$$

Geometrical Interpretation

We display p -dimensional random vectors $\mathbf{x}_1, \dots, \mathbf{x}_N$ as follows

$$\mathbf{X} = \begin{bmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{N1} & \dots & x_{Np} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_N^\top \end{bmatrix} = [\mathbf{y}_1 \quad \dots \quad \mathbf{y}_p] \in \mathbb{R}^{N \times p}.$$

We denote $\bar{\mathbf{x}} = [\bar{x}_1 \quad \dots \quad \bar{x}_p]^\top$ and $\mathbf{d}_i = \mathbf{y}_i - \bar{x}_i \mathbf{1}_N$.

- 1 The projection of \mathbf{y}_i onto the equal angular vector $\mathbf{1}_N$ is the vector $\bar{x}_i \mathbf{1}_N$.
- 2 The information comprising \mathbf{S} is obtained from the deviation vectors $\{\mathbf{d}_i\}$.
- 3 The sample correlation r_{ij} is the cosine of the angle between \mathbf{d}_i and \mathbf{d}_j .

Outline

1 Random Vectors and Matrices

2 Random Samples

3 Generalized Variance

Sample Covariance

When all variables are observed, the variation is described by the sample covariance matrix

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \in \mathbb{R}^{p \times p},$$

where $s_{ij} = \frac{1}{N-1} \sum_{\alpha=1}^N (x_{\alpha i} - \bar{x}_i)(x_{\alpha j} - \bar{x}_j).$

The sample covariance matrix contains p variances and $p(p-1)/2$ potentially different covariances.

Generalized Sample Variance

The value of $\det(\mathbf{S})$ reduces to usual sample variance when $p = 1$.

This determinant is called the generalized sample variance:

$$\text{generalized sample variance} = \det(\mathbf{S}).$$

Geometrical Interpretation: Parallelotope

Theorem

Define $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_p] \in \mathbb{R}^{N \times p}$ and let

$$\text{Vol}(\mathbf{v}_1, \dots, \mathbf{v}_p)$$

be the p -dimensional volume of the parallelotope with $\mathbf{v}_1, \dots, \mathbf{v}_p \in \mathbb{R}^N$ as principal edges ($N \geq p$), then

$$(\text{Vol}(\mathbf{v}_1, \dots, \mathbf{v}_p))^2 = \det(\mathbf{V}^\top \mathbf{V}).$$

For $\mathbf{d}_i = \mathbf{y}_i - \bar{x}_i \mathbf{1}_N$, we have

$$\det(\mathbf{S}) = (N - 1)^{-p} (\text{Vol}(\mathbf{d}_1, \dots, \mathbf{d}_p))^2.$$

Geometrical Interpretation: Parallelotope

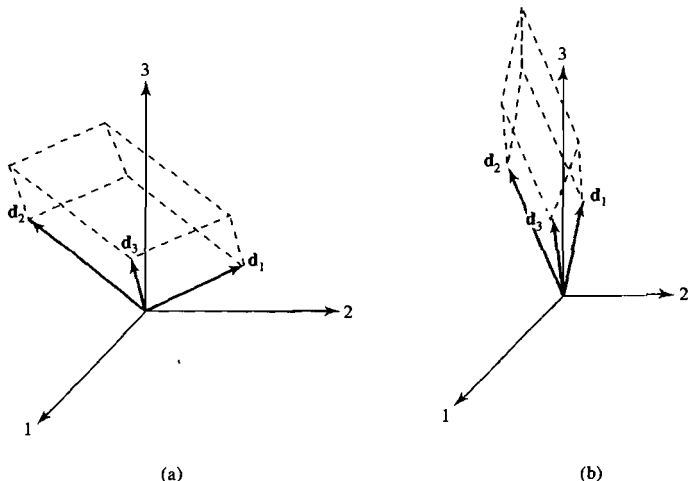


Figure 3.6 (a) "Large" generalized sample variance for $p = 3$.
(b) "Small" generalized sample variance for $p = 3$.

Geometrical Interpretation: Hyperellipsoid

The coordinates

$$\mathbf{x} = [x_1, x_2, \dots, x_p]^\top$$

of the points has a constant distance $c > 0$ from $\bar{\mathbf{x}}$ satisfy (suppose $\mathbf{S} \succ \mathbf{0}$)

$$(\mathbf{x} - \bar{\mathbf{x}})^\top \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) = c^2,$$

which defines hyperellipsoid centered at $\bar{\mathbf{x}}$.

The volume of this hyperellipsoid is

$$\frac{2\pi^{p/2}}{p\Gamma(p/2)} \cdot c^p (\det(\mathbf{S}))^{1/2},$$

where

$$\Gamma(p) = \int_0^\infty t^{p-1} \exp(-t) dt.$$

When the Generalized Sample Variance is Zero

The generalized variance is zero when, and only when, at least one of

$$\{\mathbf{d}_1, \dots, \mathbf{d}_p\}$$

lies in the hyperplane formed by all linear combinations of the others.

That is, the columns of the matrix of deviations

$$\begin{aligned}\mathbf{X} - \mathbf{1}_N \bar{\mathbf{x}}^\top &= \begin{bmatrix} (\mathbf{x}_1 - \bar{\mathbf{x}})^\top \\ \vdots \\ (\mathbf{x}_N - \bar{\mathbf{x}})^\top \end{bmatrix} = [\mathbf{y}_1 - \bar{x}_1 \mathbf{1}_N \quad \dots \quad \mathbf{y}_p - \bar{x}_p \mathbf{1}_N] \\ &= [\mathbf{d}_1 \quad \dots \quad \mathbf{d}_p] \in \mathbb{R}^{N \times p}\end{aligned}$$

are linearly dependent.

Generalized Sample Variance Determined by Correlation

We can also define generalized variance by

$$\det(\mathbf{R}),$$

where \mathbf{R} is the sample correlation matrix

$$\mathbf{R} = \begin{bmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{p1} & \cdots & r_{pp} \end{bmatrix} \in \mathbb{R}^{p \times p},$$

where $r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}.$

It holds that

$$\det(\mathbf{S}) = \det(\mathbf{R}) \prod_{i=1}^p s_{ii}.$$

Total Sample Variance

We define the total sample variance as the sum of the diagonal elements of the sample covariance matrix, that is

$$\text{total sample variance} = \sum_{i=1}^p s_{ii}.$$

- ① It is the sum of the squared lengths of the p deviation vectors

$$\mathbf{d}_1 = \mathbf{y}_1 - \bar{x}_1 \mathbf{1}_N, \mathbf{d}_2 = \mathbf{y}_2 - \bar{x}_2 \mathbf{1}_N, \dots, \mathbf{d}_p = \mathbf{y}_p - \bar{x}_p \mathbf{1}_N$$

divided by $N - 1$.

- ② It pays no attention to the orientation of \mathbf{d}_i .