

Multivariate Statistical Analysis

Lecture 05

Fudan University

luoluo@fudan.edu.cn

- 1 Singular Normal Distributions
- 2 Conditional Distribution
- 3 Characteristic Function

1 Singular Normal Distributions

2 Conditional Distribution

3 Characteristic Function

Singular Normal Distributions

In previous section, we focus on non-singular normal normally distributed variate $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} \succ \mathbf{0}$ whose density function is

$$n(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right).$$

What about the case of singular $\boldsymbol{\Sigma}$?

General Linear Transformation

- ① Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} \succ \mathbf{0}$. Then

$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

is distributed according to $\mathcal{N}_p(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\top)$ for non-singular $\mathbf{C} \in \mathbb{R}^{p \times p}$.

- ② Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} \succ \mathbf{0}$. Then

$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

is distributed according to $\mathcal{N}_q(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\top)$ for $\mathbf{C} \in \mathbb{R}^{q \times p}$ of rank $q \leq p$.

- ③ Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

is distributed according to $\mathcal{N}_q(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\top)$ for any $\mathbf{C} \in \mathbb{R}^{q \times p}$.

Transformation



5.3×10^5

$$c \neq 0$$

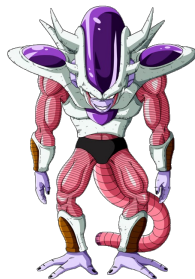
$$\sigma^2 > 0$$



$> 1.0 \times 10^6$

$$\mathbf{C} \in \mathbb{R}^{p \times p} \text{ is non-singular}$$

$$\Sigma \succ 0$$



$2.0 \times 10^6 \sim 3.0 \times 10^6$

$$\mathbf{C} \in \mathbb{R}^{q \times p} \text{ of rank } q \leq p$$

$$\Sigma \succ 0$$



$> 3.0 \times 10^7$

$$\mathbf{C} \in \mathbb{R}^{q \times p}$$

$$\Sigma \succeq 0$$

General Linear Transformation

Theorem

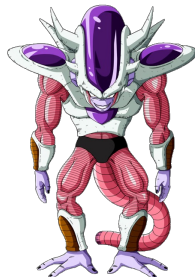
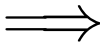
Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} \succ \mathbf{0}$. Then

$$\mathbf{z} = \mathbf{D}\mathbf{x}$$

is distributed according to $\mathcal{N}_q(\mathbf{D}\boldsymbol{\mu}, \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^\top)$ for $\mathbf{D} \in \mathbb{R}^{q \times p}$ of rank $q \leq p$.



non-singular



full-rank

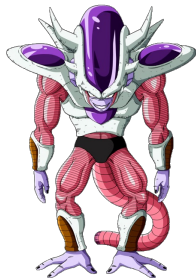
General Linear Transformation

Theorem

Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

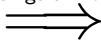
$$\mathbf{z} = \mathbf{D}\mathbf{x}$$

is distributed according to $\mathcal{N}_q(\mathbf{D}\boldsymbol{\mu}, \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^\top)$ for any $\mathbf{D} \in \mathbb{R}^{q \times p}$.



full-rank

understand the singular normal distribution



no limitation

Singular Normal Distribution

Singular normal distribution:

- 1 The mass is concentrated on a given lower dimensional set.
- 2 The probability associated with any set that does not intersecting the given low-dimensional set is 0.

For example, consider that

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right).$$

- 1 Probability of any set that does not intersecting the x_2 -axis is 0.
- 2 The measure of x_2 -axis in the space of \mathbb{R}^2 is zero.
- 3 The random vector \mathbf{x} has no density, but its distribution exists.

Singular Normal Distributions

Suppose that $\mathbf{y} \sim \mathcal{N}_q(\boldsymbol{\nu}, \mathbf{T})$, $\mathbf{A} \in \mathbb{R}^{p \times q}$ with $p > q$ and $\boldsymbol{\lambda} \in \mathbb{R}^p$; then we say that

$$\mathbf{x} = \mathbf{A}\mathbf{y} + \boldsymbol{\lambda}$$

has a singular (degenerate) normal distribution in p -space.

We have $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}] = \mathbf{A}\boldsymbol{\nu} + \boldsymbol{\lambda}$ and

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}] = \mathbf{A}\boldsymbol{\nu} + \boldsymbol{\lambda} \quad \text{and} \quad \boldsymbol{\Sigma} = \text{Cov}(\mathbf{x}) = \mathbf{A}\mathbf{T}\mathbf{A}^\top.$$

The matrix $\boldsymbol{\Sigma}$ is singular and we cannot write density for \mathbf{x} .

Singular Normal Distributions

Now we give a formal definition of a normal distribution that includes the singular distribution.

Definition

A p -dimensional random vector \mathbf{x} with $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}$ and $\text{Cov}[\mathbf{x}] = \boldsymbol{\Sigma}$ is said to be normally distributed if there is a transformation

$$\mathbf{x} = \mathbf{A}\mathbf{y} + \boldsymbol{\lambda},$$

where $\mathbf{A} \in \mathbb{R}^{p \times r}$, $\boldsymbol{\lambda} \in \mathbb{R}^p$, r is the rank of $\boldsymbol{\Sigma}$ and \mathbf{y} has r -dimensional non-singular normal distribution, e.g., $\mathbf{y} \sim \mathcal{N}_r(\boldsymbol{\nu}, \mathbf{T})$ with $\mathbf{T} \succ \mathbf{0}$.

We also use the notation $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ even if $\boldsymbol{\Sigma}$ is singular.

If $\boldsymbol{\Sigma}$ has rank p , we can take $\mathbf{A} = \mathbf{I}$ and $\boldsymbol{\lambda} = \mathbf{0}$.

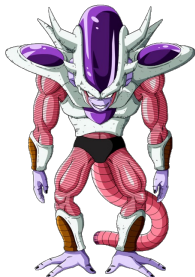
General Linear Transformation

Theorem

Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

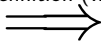
$$\mathbf{z} = \mathbf{D}\mathbf{x}$$

is distributed according to $\mathcal{N}_q(\mathbf{D}\boldsymbol{\mu}, \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^\top)$ for any $\mathbf{D} \in \mathbb{R}^{q \times p}$.



full-rank

only use the definition (without density)



no limitation

Outline

1 Singular Normal Distributions

2 Conditional Distribution

3 Characteristic Function

Conditional Distribution

Let \mathbf{x} be distributed according to $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} \succ \mathbf{0}$.

We partition

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{bmatrix} \quad \text{with } \mathbf{x}^{(1)} \in \mathbb{R}^q \text{ and } \mathbf{x}^{(2)} \in \mathbb{R}^{p-q},$$
$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix} \quad \text{with } \boldsymbol{\mu}^{(1)} \in \mathbb{R}^q \text{ and } \boldsymbol{\mu}^{(2)} \in \mathbb{R}^{p-q},$$

and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

with $\boldsymbol{\Sigma}_{11} \in \mathbb{R}^{q \times q}$, $\boldsymbol{\Sigma}_{12} \in \mathbb{R}^{q \times (p-q)}$, $\boldsymbol{\Sigma}_{21} \in \mathbb{R}^{(p-q) \times q}$ and $\boldsymbol{\Sigma}_{22} \in \mathbb{R}^{(p-q) \times (p-q)}$.

Conditional Distribution

The conditional density of $\mathbf{x}^{(1)}$ given that $\mathbf{x}^{(2)}$ is

$$\begin{aligned} f(\mathbf{x}^{(1)} | \mathbf{x}^{(2)}) &= \frac{f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})}{f(\mathbf{x}^{(2)})} \\ &= \frac{1}{\sqrt{(2\pi)^q \det(\boldsymbol{\Sigma}_{11.2})}} \exp \left(-\frac{1}{2} (\mathbf{x}_{11.2} - \boldsymbol{\mu}_{11.2})^\top \boldsymbol{\Sigma}_{11.2}^{-1} (\mathbf{x}_{11.2} - \boldsymbol{\mu}_{11.2}) \right), \end{aligned}$$

where

$$\begin{aligned} \mathbf{x}_{11.2} &= \mathbf{x}^{(1)} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \mathbf{x}^{(2)}, \\ \boldsymbol{\mu}_{11.2} &= \boldsymbol{\mu}^{(1)} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\mu}^{(2)}, \end{aligned}$$

and

$$\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}.$$

Examples

Theorem

Let \mathbf{U} be a $d \times k$ random matrix ($k \leq d$) and each of its entry is independent distributed according to $\mathcal{N}(0, 1)$, then it holds that

$$\mathbb{E} \left[\mathbf{U}(\mathbf{U}^\top \mathbf{U})^{-1} \mathbf{U}^\top \right] = \frac{k}{d} \mathbf{I}_d.$$

Lemma

Assume $\mathbf{P} \in \mathbb{R}^{d \times k}$ is column orthonormal ($k \leq d$) and $\mathbf{v} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{P}\mathbf{P}^\top)$ is a d -dimensional multivariate normal distributed vector. Then we have

$$\mathbb{E} \left[\frac{\mathbf{v}\mathbf{v}^\top}{\mathbf{v}^\top \mathbf{v}} \right] = \frac{1}{k} \mathbf{P}\mathbf{P}^\top.$$

Outline

1 Singular Normal Distributions

2 Conditional Distribution

3 Characteristic Function

Characteristic Function

The characteristic function of a p -dimensional random vector \mathbf{x} is

$$\phi(\mathbf{t}) = \mathbb{E} \left[\exp(\mathbf{i} \mathbf{t}^\top \mathbf{x}) \right]$$

defined for every real vector $\mathbf{t} \in \mathbb{R}^p$.

For the complex-valued function $g(z)$ be written as

$$g(z) = g_1(z) + \mathbf{i} g_2(z),$$

where $g_1(z)$ and $g_2(z)$ are real-valued, the expected value of $g(z)$ is

$$\mathbb{E}[g(z)] = \mathbb{E}[g_1(z)] + \mathbf{i} \mathbb{E}[g_2(z)].$$

Theorem

If the p -dimensional random vector \mathbf{x} has the density $f(\mathbf{x})$ and the characteristic function $\phi(\mathbf{t})$, then

$$f(\mathbf{x}) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(-i \mathbf{t}^\top \mathbf{x}) \phi(\mathbf{t}) dt_1 \dots dt_p.$$

- If the random variable have a density, the characteristic function determines the density function uniquely.
- If the random variable does not have a density, the characteristic function uniquely defines the probability of any continuity interval.

Characteristic Function

Theorem

The characteristic function of \mathbf{x} distributed according to $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is

$$\phi(\mathbf{t}) = \exp \left(i \mathbf{t}^\top \boldsymbol{\mu} - \frac{1}{2} \mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t} \right).$$

for every $\mathbf{t} \in \mathbb{R}^p$.

Sketch of the proof

- 1 The characteristic function of $\mathbf{y} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$ is $\phi_0(\mathbf{t}) = \exp \left(-\frac{1}{2} \mathbf{t}^\top \mathbf{t} \right)$.
- 2 For $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we have $\mathbf{x} = \mathbf{A}\mathbf{y} + \boldsymbol{\mu}$ such that $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^\top$.
- 3 Using $\phi_0(\mathbf{t})$ to present the characteristic function of $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Characteristic Function

Theorem

The characteristic function of \mathbf{x} distributed according to $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is

$$\phi(\mathbf{t}) = \exp \left(i \mathbf{t}^\top \boldsymbol{\mu} - \frac{1}{2} \mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t} \right).$$

for every $\mathbf{t} \in \mathbb{R}^p$.

This theorem directly implies $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ leads to $\mathbf{C}\mathbf{x} \sim \mathcal{N}(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\top)$.



characteristic function



trick of matrix

Theorem

If every linear combination of the components of a random vector \mathbf{y} is normally distributed, then \mathbf{y} is normally distributed.

In other words, if the p -dimensional random vector \mathbf{y} leads to the univariate random variable

$$\mathbf{u}^\top \mathbf{y}$$

is normally distributed for any fixed $\mathbf{u} \in \mathbb{R}^p$, then \mathbf{y} is normally distributed.

This is another definition of multivariate normal distribution.

Example

Theorem

We let

$$\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \quad \mathbf{y} \sim \mathcal{N}_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) \quad \text{and} \quad \mathbf{z} = \mathbf{x} + \mathbf{y}.$$

Suppose that \mathbf{x} and \mathbf{y} are independent, then we have

$$\mathbf{z} \sim \mathcal{N}_p(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2).$$



characteristic function



this result