Multivariate Statistical Analysis

Lecture 05

Fudan University

luoluo@fudan.edu.cn

Outline

Outline

Characteristic Function

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

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

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

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

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

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

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

Characteristic Function

Theorem

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

$$f(\mathbf{x}) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(-\mathrm{i}\,\mathbf{t}^\top \mathbf{x})\, \phi(\mathbf{t})\, \mathrm{d}t_1 \ldots \mathrm{d}t_p.$$

- 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*

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

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 $\phi_0(\mathbf{t}) = \exp\left(-\mathbf{t}^\top \mathbf{t}/2\right)$.
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

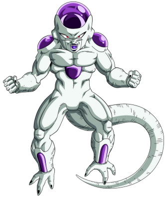
The characteristic function of $\mathbf{x}$ *distributed according to* $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *is*

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

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

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

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

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

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

Outline

1. Characteristic Function

2. Maximum Likelihood Estimation

The Maximum Likelihood Estimators

Theorem

If $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ constitute a sample from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $N > p$, the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are

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

respectively.

The Maximum Likelihood Estimators

The likelihood function is

$$L = \frac{1}{(2\pi)^{\frac{pN}{2}} \left(\det(\mathbf{\Sigma})\right)^{\frac{N}{2}}} \exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(\mathbf{x}_\alpha - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{x}_\alpha - \boldsymbol{\mu})\right].$$

The vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ are fixed at the sample values and L is a function of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$.

The logarithm of the likelihood function is

$$\ln L = -\frac{pN}{2}\ln 2\pi - \frac{N}{2}\ln\left(\det(\mathbf{\Sigma})\right) - \frac{1}{2}\sum_{\alpha=1}^{N}(\mathbf{x}_\alpha - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{x}_\alpha - \boldsymbol{\mu}).$$

The Maximum Likelihood Estimators

There are some results for estimating the covariance.

Theorem

The function $h : \mathbb{S}_{++}^p \to \mathbb{R}$ *such that*

$$h(\mathbf{X}) = -\log\det(\mathbf{X})$$

is convex, where $\mathbb{S}_{++}^p = \{\mathbf{X} \in \mathbb{R}^{p \times p} : \mathbf{X} \succ \mathbf{0}\}$.

Theorem

If $\mathbf{D} \in \mathbb{R}^{p \times p}$ *is positive definite, the maximum of*

$$f(\mathbf{G}) = -N \ln\det(\mathbf{G}) - \mathrm{tr}(\mathbf{G}^{-1}\mathbf{D})$$

with respect to positive definite matrices $\mathbf{G}$ *exists, occurs at* $\mathbf{G} = \frac{1}{N}\mathbf{D}$.