

# Multivariate Statistical Analysis

## Lecture 07

Fudan University

luoluo@fudan.edu.cn

# Outline

- 1 Unbiasedness
- 2 Sufficiency
- 3 Completeness
- 4 Efficiency

# Outline

1 Unbiasedness

2 Sufficiency

3 Completeness

4 Efficiency

An estimator  $\mathbf{t}$  of a parameter vector  $\boldsymbol{\theta}$  is unbiased if and only if

$$\mathbb{E}[\mathbf{t}] = \boldsymbol{\theta}.$$

For the estimators obtain from MLE for normal distribution,

- ① the vector  $\hat{\boldsymbol{\mu}}$  is an unbiased estimator of  $\boldsymbol{\mu}$ ;
- ② the matrix  $\hat{\boldsymbol{\Sigma}}$  is a biased estimator of  $\boldsymbol{\Sigma}$ .

# Outline

- 1 Unbiasedness
- 2 Sufficiency
- 3 Completeness
- 4 Efficiency

# Sufficiency

A statistic  $\mathbf{t}(\mathbf{y})$  is sufficient for a family of distributions of random variable  $\mathbf{y}$  with parameter  $\theta$ , if the conditional distribution of  $\mathbf{y}$  given  $\mathbf{t}(\mathbf{y}) = \mathbf{t}_0$  does not depend on  $\theta$ .

- 1 The statistic  $\mathbf{t}$  gives as much information about  $\theta$  as the entire sample  $\mathbf{y}$ .
- 2 For the MLE of normal distribution, we check the sufficiency by taking

$$\theta = \{\mu, \Sigma\}, \quad \mathbf{y} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \quad \text{and} \quad \mathbf{t}(\mathbf{y}) = \{\bar{\mathbf{x}}, \mathbf{S}\}.$$

## Theorem

*A statistic  $\mathbf{t}(\mathbf{y})$  is sufficient for  $\theta$  if and only if the density  $f(\mathbf{y}; \theta)$  can be factored as*

$$f(\mathbf{y}; \theta) = g(\mathbf{t}(\mathbf{y}); \theta)h(\mathbf{y})$$

*where  $g(\mathbf{t}(\mathbf{y}); \theta)$  and  $h(\mathbf{y})$  are nonnegative and  $h(\mathbf{y})$  does not depend on  $\theta$ .*

# Outline

- 1 Unbiasedness
- 2 Sufficiency
- 3 Completeness**
- 4 Efficiency

A family of distributions of statistics  $\mathbf{t}$  indexed by  $\theta$  is complete if for every real-valued function  $g(\mathbf{t})$ , we have

$$\mathbb{E}[g(\mathbf{t})] \equiv 0$$

identically in  $\theta$  implies  $g(\mathbf{t}) = 0$  except for a set of  $\mathbf{t}$  of probability 0 for every  $\theta$ .



## Theorem

*The sufficient set of statistics  $\bar{\mathbf{x}}, \mathbf{S}$  is complete for  $\boldsymbol{\mu}, \boldsymbol{\Sigma}$  when the sample is drawn from  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .*

Sketch of the proof:

- ① We have  $N\hat{\boldsymbol{\Sigma}} = \sum_{\alpha=1}^{N-1} \mathbf{z}_{\alpha} \mathbf{z}_{\alpha}^{\top}$ , where  $\mathbf{z}_{\alpha} = \sum_{\beta=1}^N b_{\alpha\beta} \mathbf{x}_{\beta}$  and

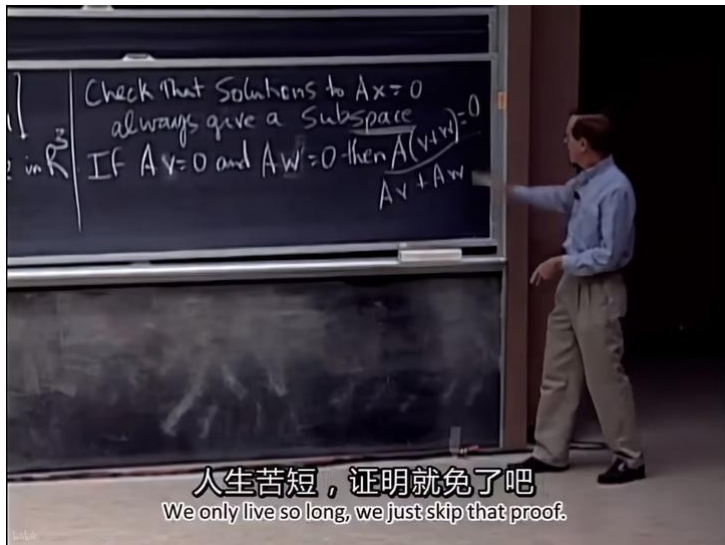
$$\mathbf{B} = \begin{bmatrix} \times & \cdots & \times \\ \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{N}} & \cdots & \frac{1}{\sqrt{N}} \end{bmatrix}$$

- ② The condition  $\mathbb{E}[g(\bar{\mathbf{x}}, n\mathbf{S})] \equiv 0$  implies the Laplace transform of

$$g(\bar{\mathbf{x}}, \mathbf{B} - N\bar{\mathbf{x}}\bar{\mathbf{x}}^{\top}) h(\bar{\mathbf{x}}, \mathbf{B})$$

is zero, where  $\mathbf{B} = \sum_{\alpha=1}^{N-1} \mathbf{z}_{\alpha} \mathbf{z}_{\alpha}^{\top} + N\bar{\mathbf{x}}\bar{\mathbf{x}}^{\top}$  and  $h(\bar{\mathbf{x}}, \mathbf{B})$  is the joint density of  $\bar{\mathbf{x}}$  and  $\mathbf{B}$ .

# Completeness



# Outline

- 1 Unbiasedness
- 2 Sufficiency
- 3 Completeness
- 4 Efficiency

# Concentration Ellipsoid

If a  $p$ -dimensional random vector  $\mathbf{y}$  has mean vector

$$\boldsymbol{\nu} = \mathbb{E}[\mathbf{y}]$$

and covariance matrix

$$\boldsymbol{\Psi} = \mathbb{E} \left[ (\mathbf{y} - \boldsymbol{\nu})(\mathbf{y} - \boldsymbol{\nu})^\top \right] \succ \mathbf{0},$$

then

$$\left\{ \mathbf{z} : (\mathbf{z} - \boldsymbol{\nu})^\top \boldsymbol{\Psi}^{-1} (\mathbf{z} - \boldsymbol{\nu}) = p + 2 \right\}$$

is called the concentration ellipsoid of  $\mathbf{y}$ .

# Concentration Ellipsoid

Let  $\theta$  be a vector of  $p$  parameters in a distribution, and let  $\mathbf{t}$  be a vector of unbiased estimators (that is,  $\mathbb{E}[\mathbf{t}] = \theta$ ) based on  $N$  observations from that distribution with covariance matrix  $\Psi$ .

Then the ellipsoid

$$\left\{ \mathbf{z} : (\mathbf{z} - \theta)^\top \mathbb{E} \left[ N \cdot \frac{\partial \ln f(\mathbf{x}, \theta)}{\partial \theta} \left( \frac{\partial \ln f(\mathbf{x}, \theta)}{\partial \theta} \right)^\top \right] (\mathbf{z} - \theta) = p + 2 \right\}$$

lies entirely within the ellipsoid of concentration of  $\mathbf{t}$ , where  $f$  is the density of the distribution with respect to the components of  $\theta$ .

The ellipsoid

$$\left\{ \mathbf{z} : (\mathbf{z} - \boldsymbol{\theta})^\top \mathbb{E} \left[ N \cdot \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left( \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \right] (\mathbf{z} - \boldsymbol{\theta}) = p + 2 \right\}$$

lies entirely within the ellipsoid of concentration of  $\mathbf{t}$

$$\left\{ \mathbf{z} : (\mathbf{z} - \boldsymbol{\theta})^\top \left( \mathbb{E} \left[ (\mathbf{t} - \boldsymbol{\theta})(\mathbf{t} - \boldsymbol{\theta})^\top \right] \right)^{-1} (\mathbf{z} - \boldsymbol{\theta}) = p + 2 \right\},$$

that is

$$\left( N \mathbb{E} \left[ \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left( \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \right] \right)^{-1} \preceq \mathbb{E} \left[ (\mathbf{t} - \boldsymbol{\theta})(\mathbf{t} - \boldsymbol{\theta})^\top \right].$$

The ellipsoid

$$\left\{ \mathbf{z} : (\mathbf{z} - \boldsymbol{\theta})^\top \mathbb{E} \left[ N \cdot \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left( \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \right] (\mathbf{z} - \boldsymbol{\theta}) = p + 2 \right\} \quad (1)$$

lies entirely within the ellipsoid of concentration of  $\mathbf{t}$

$$\left\{ \mathbf{z} : (\mathbf{z} - \boldsymbol{\theta})^\top \left( \mathbb{E} \left[ (\mathbf{t} - \boldsymbol{\theta})(\mathbf{t} - \boldsymbol{\theta})^\top \right] \right)^{-1} (\mathbf{z} - \boldsymbol{\theta}) = p + 2 \right\}. \quad (2)$$

- 1 If the ellipsoid (1) and the ellipsoid (2) are identical, then  $\mathbf{t}$  is said to be efficient.
- 2 In general, the ratio of the volume of ellipsoid (1) to that of the ellipsoid (2) defines the efficiency of  $\mathbf{t}$ .

# Multivariate Cramér-Rao Inequality

## Theorem

*Under the regularity condition (everything is well-defined, integration and differentiation can be swapped), we have*

$$N\mathbb{E}[(\mathbf{t} - \boldsymbol{\theta})(\mathbf{t} - \boldsymbol{\theta})^\top] \succeq \left( \mathbb{E} \left[ \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left( \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \right] \right)^{-1},$$

*where  $\mathbb{E}[\mathbf{t}] = \boldsymbol{\theta}$  and  $f(\mathbf{x}, \boldsymbol{\theta})$  is the density of the distribution with respect to the components of  $\boldsymbol{\theta}$ .*

- 1 Let  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  and  $\mathbf{s} = \frac{\partial \ln g(\mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ , where  $g$  is the joint density on  $N$  samples.
- 2 For unbiased estimator  $\mathbf{t}$  of  $\boldsymbol{\theta}$ , we have  $\text{Cov}[\mathbf{t}, \mathbf{s}] = \mathbf{I}$ .



We define the Fisher information matrix as

$$\mathbb{E} \left[ \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left( \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \right].$$

Under the regularity condition, we have

$$\mathbb{E} \left[ \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left( \frac{\partial \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^\top \right] = -\mathbb{E} \left[ \frac{\partial^2 \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right].$$

Consider the case of the multivariate normal distribution.

- ① If  $\theta = \mu$ , then  $\bar{\mathbf{x}}$  is efficient.
- ② If  $\theta = \{\mu, \Sigma\}$ , then  $\{\bar{\mathbf{x}}, \mathbf{S}\}$  has efficiency

$$\left(\frac{N-1}{N}\right)^{p(p+1)/2},$$

which converges to 1 if  $N \rightarrow +\infty$ .