Multivariate Statistical Analysis

Lecture 07

Fudan University

luoluo@fudan.edu.cn

Outline

Outline

Unbiasedness

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

Sufficiency

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

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

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

Theorem

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

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

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

Outline

Completeness

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

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

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

Completeness

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
$$\mathbf{B} = \begin{bmatrix} \times & \dots & \times \\ \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{N}} & \dots & \frac{1}{\sqrt{N}} \end{bmatrix}$$
2. The condition $\mathbb{E}[g(\bar{\mathbf{x}}, n\mathbf{S})] \equiv 0$ implies the Laplace transform of
$$g\left(\bar{\mathbf{x}}, \mathbf{B} - N\bar{\mathbf{x}}\bar{\mathbf{x}}^\top\right) h(\bar{\mathbf{x}}, \mathbf{B})$$
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 $\mathbf{B}$.

Outline

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

Then 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 $\mathbf{t}$, where f is the density of the distribution with respect to the components of $\boldsymbol{\theta}$.

Efficiency

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 **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

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

Efficiency

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 **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 the unbiased estimator **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 the unbiased estimator **t**.

Multivariate Cramér-Rao Inequality

Theorem

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

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

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 $\mathbf{s} = \dfrac{\partial \ln g(\mathbf{X},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, where g is the joint density on N samples.
2. For unbiased estimator $\mathbf{t}$ of $\boldsymbol{\theta}$, we have $\mathrm{Cov}[\mathbf{t},\mathbf{s}] = \mathbf{I}$.

Efficiency

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

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

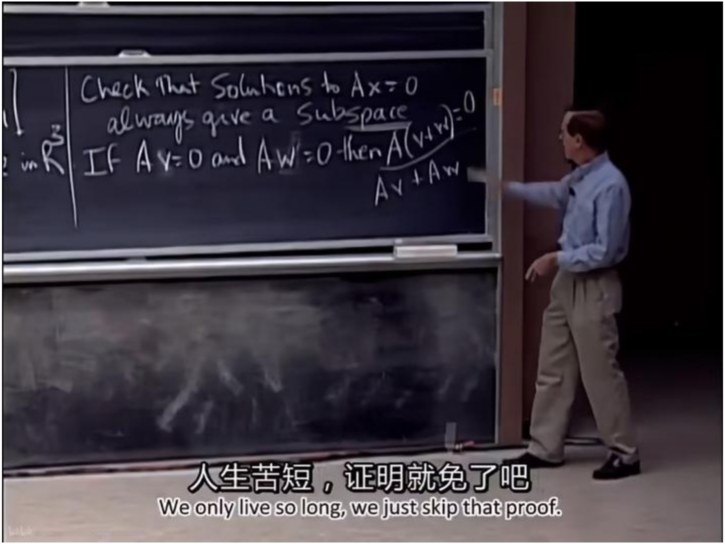
Efficiency

Consider the case of the multivariate normal distribution.

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

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

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


Check That Solutions to Ax=0
always give a Subspace
If Av=0 and Aw=0 then A(v+w)=0
Av+Aw
人生苦短，证明就免了吧
We only live so long, we just skip that proof.

Outline

Consistency

A sequence of random vectors $\mathbf{t}_n = [t_{1n}, \ldots, t_{pn}]^\top$ for $n = 1, 2, \ldots,$ is a consistent estimator of $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_p]^\top$ if

$$\operatorname*{plim}_{n \to +\infty} t_{in} = \theta_i$$

for $i = 1, \ldots, p$.

The definition of convergence in probability says

$$\lim_{n \to +\infty} \Pr\left(|t_{in} - \theta_i| < \epsilon\right) = 1$$

holds for any $\epsilon > 0$.

Consistency

The weak law of large numbers states that the sample means converges in probability towards the expected value.

For sample $\mathbf{x}_1, \mathbf{x}_2 \dots$ from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the estimators

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

are consistent estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively.

Outline

Let $x_1, \dots, x_n$ be independent and identically distributed random variables with the same arbitrary distribution, mean μ, and variance σ^2.

Let $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, then the random variable

$$z = \lim_{n \to \infty} \sqrt{n}\left(\frac{\bar{x}_n - \mu}{\sigma}\right)$$

is a standard normal distribution.

What about multivariate case?

Multivariate central limit theorem.

Theorem

Let p-component vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots$ be i.i.d with means $\mathbb{E}[\mathbf{y}_\alpha] = \boldsymbol{\nu}$ and covariance matrices $\mathbb{E}[(\mathbf{y}_\alpha - \boldsymbol{\nu})(\mathbf{y}_\alpha - \boldsymbol{\nu})^\top] = \mathbf{T}$. Then the limiting distribution of

$$\frac{1}{\sqrt{n}}\sum_{\alpha=1}^{n}(\mathbf{y}_\alpha - \boldsymbol{\nu})$$

as $n \to +\infty$ is $\mathcal{N}(\mathbf{0}, \mathbf{T})$.

Characteristic Function and Probability

Theorem

Let $\{F_j(\mathbf{x})\}$ be a sequence of cdfs, and let $\{\phi_j(\mathbf{t})\}$ be the sequence of corresponding characteristic functions. A necessary and sufficient condition for $F_j(\mathbf{x})$ to converge to a cdf $F(\mathbf{x})$ is that, for every $\mathbf{t}$, $\phi_j(\mathbf{t})$ converges to a limit $\phi(\mathbf{t})$ that is continuous at $\mathbf{t} = \mathbf{0}$. When this condition is satisfied, the limit $\phi(\mathbf{t})$ is identical with the characteristic function of the limiting distribution $F(\mathbf{x})$.