Multivariate Statistics

Lecture 05

Fudan University

Outline

Outline

The Maximum Likelihood Estimators

Theorem 1

If $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ constitute a sample from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $p < N$, 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.

Lemma 1

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 $\mathbf{G} = \frac{1}{N}\mathbf{D}$.

The Maximum Likelihood Estimators

Theorem 1

If $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ constitute a sample from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $p < N$, 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.

Can we guarantee $\hat{\boldsymbol{\Sigma}} = \frac{1}{N}\sum_{\alpha=1}^{N} (\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})^\top$ is positive definite?

In the univariate case, the mean of a sample is distributed normally and independently of the sample variance.

In the multivariate case, the sample mean $\hat{\boldsymbol{\mu}}$ is also distributed normally and independently of $\hat{\boldsymbol{\Sigma}}$.

Lemma 1

Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_N$ are independent, where $\mathbf{x}_\alpha \sim \mathcal{N}_p(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma})$. Let $\mathbf{C} \in \mathbb{R}^{N\times N}$ be an orthogonal matrix, then

$$\mathbf{y}_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta}\mathbf{x}_\beta \sim \mathcal{N}_p(\boldsymbol{\nu}_\alpha, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\nu} = \sum_{\beta=1}^{N} c_{\alpha\beta}\boldsymbol{\mu}_\beta$ for $\alpha = 1, \ldots, N$ and $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are independent.

Lemma 2

If $\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \dots & c_{pp} \end{bmatrix} = \begin{bmatrix} c_1^\top \\ c_2^\top \\ \vdots \\ c_p^\top \end{bmatrix} \in \mathbb{R}^{p \times p}$ is orthogonal, then $\sum_{\alpha=1}^{N} \mathbf{x}_\alpha \mathbf{x}_\alpha^\top = \sum_{\beta=1}^{N} \mathbf{y}_\alpha \mathbf{y}_\alpha^\top$ where $\mathbf{y}_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta} \mathbf{x}_\beta$ for $\alpha = 1, \dots, N$.

Let $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_p^\top \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^\top \\ \mathbf{y}_2^\top \\ \vdots \\ \mathbf{y}_p^\top \end{bmatrix}$, then $\mathbf{y}_\alpha = \mathbf{X}^\top \mathbf{c}_\alpha$ and $\mathbf{Y} = \mathbf{CX}$.

Distribution Theory

Theorem 2

Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be independent, each distributed according to $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then the mean of the sample

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

is distributed according to $\mathcal{N}(\boldsymbol{\mu}, \frac{1}{N}\boldsymbol{\Sigma})$ and independent of

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

Additionally, we have $N\hat{\boldsymbol{\Sigma}} = \sum_{\alpha=1}^{N-1}\mathbf{z}_\alpha\mathbf{z}_\alpha^\top$, where $\mathbf{z}_\alpha \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ for $\alpha = 1, \ldots, N-1$, and $\mathbf{z}_1, \ldots, \mathbf{z}_{N-1}$ are independent.

Distribution Theory

Theorem 1

If $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ constitute a sample from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $p < N$, 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.

Theorem 3

Using the notation of Theorem 1, if $N > p$, the probability is 1 of drawing a sample so that

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

is positive definite.

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, the vector $\hat{\boldsymbol{\mu}}$ is an unbiased estimator of $\boldsymbol{\mu}$ and $\hat{\boldsymbol{\Sigma}}$ is a biased estimator of $\boldsymbol{\Sigma}$.

Consider the result of MLE for normal distribution:

1. We have

$$\mathbb{E}[\hat{\boldsymbol{\mu}}] = \mathbb{E}[\bar{\mathbf{x}}] = \mathbb{E}\left[\sum_{\alpha=1}^{N} \mathbf{x}_\alpha\right] = \boldsymbol{\mu}$$

   and (not limited to normal distribution)

$$\mathbb{E}\big[\hat{\boldsymbol{\Sigma}}\big] = \mathbb{E}\left[\frac{1}{N}\sum_{\alpha=1}^{N-1} \mathbf{z}_\alpha \mathbf{z}_\alpha^\top\right] = \frac{N-1}{N}\boldsymbol{\Sigma}.$$

2. The sample covariance

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

   is an unbiased estimator of $\boldsymbol{\Sigma}$.

Outline

Let

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

We shall show that $\bar{\mathbf{x}}$ and $\mathbf{S}$ are sufficient statistics and are complete.

Sufficiency

A statistic $\mathbf{t}$ is sufficient for a family of distributions of $\mathbf{y}$ or for a parameter $\boldsymbol{\theta}$ if the conditional distribution of $\mathbf{y}$ given $\mathbf{t}$ does not depend on $\boldsymbol{\theta}$.

The statistic $\mathbf{t}$ gives as much information about $\boldsymbol{\theta}$ as the entire sample $\mathbf{y}$.

Theorem 4

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

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

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

For the MLE of normal distribution, we apply this theorem with

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

Sufficiency

Theorem 5

If $\mathbf{x}_1, \ldots, \mathbf{x}_N$ are observations from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

1. $\bar{\mathbf{x}}$ and $\mathbf{S}$ are sufficient for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$;
2. if $\boldsymbol{\mu}$ is given, $\sum_{\alpha=1}^{N}(\mathbf{x}_\alpha - \boldsymbol{\mu})(\mathbf{x}_\alpha - \boldsymbol{\mu})^\top$ is sufficient for $\boldsymbol{\Sigma}$;
3. if $\boldsymbol{\Sigma}$ is given, $\bar{\mathbf{x}}$ is sufficient for $\boldsymbol{\mu}$.

Outline

Completeness

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

$$\mathbb{E}[g(\mathbf{y})] = 0$$

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

If the family of distributions of a sufficient set of statistics is complete, the set is called a complete sufficient set.

Completeness

Theorem 6

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 $\mathbf{B}$.