

第5-6章

多元正态分布的参数估计和置信域

李高荣

北京师范大学统计学院

E-mail: ligaorong@bnu.edu.cn



1 多元正态分布的参数估计

- 多元正态分布样本统计量
- 极大似然估计

2 多元正态分布的参数估计的性质

3 相关系数的极大似然估计

4 样本偏相关系数

5 总体均值向量的置信域估计

- 单个多元正态总体
- 两个多元正态总体



- 扫描二维码获取在线课件和相关教学电子资源
- 请遵守电子资源使用协议

多元正态分布样本统计量

- 设 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ 为 p 元正态总体 $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的样本容量为 n 的简单随机样本矩阵, 其中 $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\Sigma} > 0$, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ 和 $n > p$ 。

样本均值向量

设 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})' (i = 1, \dots, n)$ 为一组随机样本, 则称

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \frac{1}{n} \mathbf{X}' \mathbf{1}_n = (\bar{x}_1, \dots, \bar{x}_p)'$$

为 **样本均值向量**, 其中 $\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik}, k = 1, \dots, p$ 。

样本离差阵

设 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ ($i = 1, \dots, n$) 为一组随机样本, 则称

$$\begin{aligned}\mathbf{V} &= \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' = \mathbf{X}'\mathbf{X} - n\bar{\mathbf{x}}\bar{\mathbf{x}}' \\ &= \mathbf{X}'\left[\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right]\mathbf{X}\end{aligned}$$

为样本离差阵, 其中 \mathbf{X} 为 $n \times p$ 的样本矩阵。

$$\begin{aligned}\sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})' &= \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})' \\ &= \mathbf{V} + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'\end{aligned}$$

样本协方差阵

设 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ ($i = 1, \dots, n$) 为一组随机样本, 则称

$$\mathbf{S} = \frac{1}{n-1} \mathbf{V} = (s_{ij})_{p \times p}$$

为**样本协方差阵**。其中 $s_{kk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2$ ($k = 1, \dots, p$) 称为第 k 个样本变量 \mathbf{x}_k 的**样本方差**; $\sqrt{s_{kk}}$ 称为变量 \mathbf{x}_k 的**样本标准差**。

作用:

- 估计多元正态总体分布的协方差阵
- 对有关总体分布均值向量和协方差阵的假设进行检验

样本相关阵

设 $\mathbf{S} = (s_{ij})_{p \times p}$ 为样本协方差阵，则称 $\tilde{\mathbf{R}} = (r_{ij})_{p \times p}$ 为样本相关阵，其中

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}, \quad i, j = 1, \dots, p.$$

若记 $\tilde{\mathbf{D}}^{1/2} = \text{diag}(\sqrt{s_{11}}, \dots, \sqrt{s_{pp}})$ 为样本标准差对角阵，则

$$\tilde{\mathbf{R}} = \tilde{\mathbf{D}}^{-1/2} \mathbf{S} \tilde{\mathbf{D}}^{-1/2}, \quad \mathbf{S} = \tilde{\mathbf{D}}^{1/2} \tilde{\mathbf{R}} \tilde{\mathbf{D}}^{1/2}.$$

例1: 设从某小学某班随机抽取10个同学了解该班学生身高和体重的情况, 每个同学的身高为 X_1 (单位: 厘米), 体重为 X_2 (单位: 公斤), 具体数值如下:

$$\mathbf{X} = \begin{pmatrix} 142.0 & 32.7 \\ 143.0 & 33.5 \\ 143.0 & 33.0 \\ 145.0 & 34.2 \\ 148.0 & 37.4 \\ 149.0 & 43.1 \\ 138.0 & 31.5 \\ 144.0 & 43.2 \\ 150.0 & 37.5 \\ 153.0 & 51.3 \end{pmatrix}.$$

多元正态分布样本统计量

```
X1 = c(142.0, 143.0, 143.0, 145.0, 148.0,  
       149.0, 138.0, 144.0, 150.0, 153.0)  
X2 = c(32.7, 33.5, 33.0, 34.2, 37.4,  
       43.1, 31.5, 43.2, 37.5, 51.3)  
X = cbind(X1, X2)  
> apply(X, 2, mean)  
      X1      X2  
145.50  37.74  
> cov(X)  
      X1      X2  
X1 19.83333 22.11111  
X2 22.11111 39.98933  
> cor(X)  
      X1      X2  
X1 1.0000000 0.7851283  
X2 0.7851283 1.0000000
```

重要定理：定理5.1.1

定理5.1.1

设 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ ($i = 1, \dots, n$) 为来自 p 元正态总体 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的一组随机样本， $\bar{\mathbf{x}}$ 为样本均值向量， \mathbf{V} 为样本离差阵，则

- ① $\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$;
- ② $\mathbf{V} \sim W_p(n-1, \boldsymbol{\Sigma})$ ，其中 $n > p$ ，并且 $W_p(n-1, \boldsymbol{\Sigma})$ 是自由度是 $n-1$ 的Wishart分布；
- ③ $\bar{\mathbf{x}}$ 与 \mathbf{V} 相互独立。
- ④ $\Pr(\mathbf{V} > 0) = 1$ 的充要条件是 $n > p$ 。

定理5.1.1的证明

证明： 设 Γ 是第1个行向量为 $(1/\sqrt{n}, \dots, 1/\sqrt{n})$ 的 n 阶正交矩阵，具有如下形式：

$$\Gamma = \begin{pmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \gamma_{21} & \cdots & \gamma_{2n} \\ \vdots & & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nn} \end{pmatrix} = (\gamma_{ij})_{n \times n}.$$

令

$$\mathbf{Y} = \begin{pmatrix} y'_1 \\ \vdots \\ y'_n \end{pmatrix} = \Gamma \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} = \Gamma \mathbf{X},$$

定理5.1.1的证明

证明(续): 则有

$$\mathbf{y}_i = \sum_{k=1}^n \gamma_{ik} \mathbf{x}_k, \quad (i = 1, \cdots, n)$$

为 p 维随机向量。

因为 \mathbf{y}_i 是 p 维正态随机向量 $\mathbf{x}_1, \cdots, \mathbf{x}_n$ 的线性组合, 则 \mathbf{y}_i 也是 p 维正态随机向量。由 Γ 的定义, 则有

$$E(\mathbf{y}_i) = \sum_{k=1}^n \gamma_{ik} E(\mathbf{x}_k) = \begin{cases} \sqrt{n}\boldsymbol{\mu}, & \text{当 } i = 1 \text{ 时,} \\ \mathbf{0}, & \text{当 } i \neq 1 \text{ 时;} \end{cases}$$

$$\text{Cov}(\mathbf{y}_i, \mathbf{y}_j) = \sum_{k=1}^n \gamma_{ik} \gamma_{jk} \boldsymbol{\Sigma} = \begin{cases} \boldsymbol{\Sigma}, & \text{当 } i = j \text{ 时,} \\ \mathbf{0}, & \text{当 } i \neq j \text{ 时.} \end{cases}$$

定理5.1.1的证明

(1) 因为 $\mathbf{y}_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i = \sqrt{n} \bar{\mathbf{x}} \sim N_p(\sqrt{n} \boldsymbol{\mu}, \boldsymbol{\Sigma})$ 。故有：

$$\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n).$$

(2) 由于 $\mathbf{y}_1 = \sqrt{n} \bar{\mathbf{x}}$ ，则

$$\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' + n \bar{\mathbf{x}} \bar{\mathbf{x}}' = \mathbf{V} + \mathbf{y}_1 \mathbf{y}_1'.$$

因为 $\mathbf{Y}'\mathbf{Y} = (\mathbf{X}'\boldsymbol{\Gamma}')(\boldsymbol{\Gamma}\mathbf{X}) = \mathbf{X}'\mathbf{X}$ ，则

$$\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' = \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i'.$$

定理5.1.1的证明

因此, 有:

$$\begin{aligned}\mathbf{V} &= \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' - \mathbf{y}_1 \mathbf{y}_1' = \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i' - \mathbf{y}_1 \mathbf{y}_1' \\ &= \sum_{i=2}^n \mathbf{y}_i \mathbf{y}_i' \sim W_p(n-1, \Sigma).\end{aligned}$$

- (3) 因为 $\bar{\mathbf{x}} = \mathbf{y}_1 / \sqrt{n}$ 是 \mathbf{y}_1 的函数, 而 $\mathbf{V} = \sum_{i=2}^n \mathbf{y}_i \mathbf{y}_i'$ 是 $\mathbf{y}_2, \dots, \mathbf{y}_n$ 的函数, 且 \mathbf{y}_1 与 $\mathbf{y}_2, \dots, \mathbf{y}_n$ 相互独立, 故 $\bar{\mathbf{x}}$ 与 \mathbf{V} 相互独立。
- (4) 令 $\mathbf{Y}_* = (\mathbf{y}_2, \dots, \mathbf{y}_n)'$ 。由 $\mathbf{V} \stackrel{d}{=} \mathbf{Y}_*' \mathbf{Y}_*$, 知: $\text{rank}(\mathbf{V}) = \text{rank}(\mathbf{Y}_*' \mathbf{Y}_*) = \text{rank}(\mathbf{Y}_*)$ 。而 \mathbf{Y}_* 为 $(n-1) \times p$ 随机阵, 各行向量独立同分布。故, $\Pr(\mathbf{V} > 0) = \Pr(\text{rank}(\mathbf{V}) = p) = \Pr(\text{rank}(\mathbf{Y}_*) = p) = 1 \Leftrightarrow n-1 \geq p$, 即 $n > p$ 。

极大似然估计(MLE)

- 设 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 是来自多元正态总体 $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的随机样本, 其中 $n > p, \boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} > 0$
- 样本 $\mathbf{x}_i (i = 1, \dots, n)$ 的似然函数记为

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^n \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right] \\ &= \frac{1}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right] \\ &= \frac{1}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2}} \exp \left[-\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \left\{ \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})' \right\} \right) \right] \\ &= \frac{1}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2}} \exp \left[-\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \{ \mathbf{V} + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})' \} \right) \right]. \end{aligned}$$

- 首先给定 $\Sigma > 0$ 时, 求 μ 的极大似然估计, 即求对数似然函数 $\ln L(\mu, \Sigma)$ 的极大值点
- 给定 $\Sigma > 0$, 关于 μ 的对数似然函数为:

$$\begin{aligned}
 \ln L(\mu, \Sigma) &= -\frac{np}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{tr} \left(\Sigma^{-1} \{ \mathbf{V} + n(\bar{\mathbf{x}} - \mu)(\bar{\mathbf{x}} - \mu)' \} \right) \\
 &= -\frac{np}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{V}) - \frac{n}{2} [(\bar{\mathbf{x}} - \mu)' \Sigma^{-1} (\bar{\mathbf{x}} - \mu)] \\
 &\leq -\frac{np}{2} \ln 2\pi - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{V}).
 \end{aligned}$$

- 不等式中等号成立当且仅当 $\mu = \bar{\mathbf{x}}$ 。因此, 总体均值向量 μ 的MLE为 $\hat{\mu} = \bar{\mathbf{x}}$ 。

- 由定理5.1.1的结论(1): $\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$, 可知 $\bar{\mathbf{x}}$ 是 $\boldsymbol{\mu}$ 的无偏估计。
- 将 $\boldsymbol{\mu}$ 用它的MLE替换, 得到 $\boldsymbol{\Sigma}$ 的似然函数为:

$$L(\bar{\mathbf{x}}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2}} \exp \left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{V}) \right].$$

- 利用矩阵的性质: 存在可逆对称阵 \mathbf{C} , 使得 $\mathbf{B} = \mathbf{C}\mathbf{C}$ 。记 $\tilde{\boldsymbol{\Sigma}} = \mathbf{C}^{-1}\boldsymbol{\Sigma}\mathbf{C}^{-1}$, 则 $|\boldsymbol{\Sigma}| = |\mathbf{B}||\tilde{\boldsymbol{\Sigma}}|$ 。
- 令 $\boldsymbol{\Sigma}^{-1/2}\mathbf{V}\boldsymbol{\Sigma}^{-1/2} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'$, 其中 \mathbf{U} 是正交矩阵, $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$ 是对角矩阵。

- 利用 $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$, 可知:

$$|\mathbf{V}|^{n/2} = |\boldsymbol{\Sigma}|^{n/2} \prod_{k=1}^p \lambda_k^{n/2},$$

$$\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{V}) = \text{tr}(\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Sigma}^{-1/2} \mathbf{V}) = \sum_{k=1}^p \lambda_k.$$

- 则似然函数可简化为:

$$L(\bar{\mathbf{x}}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2} |\mathbf{V}|^{n/2}} \prod_{k=1}^p \left[\lambda_k^{n/2} \exp \left\{ -\frac{\lambda_k}{2} \right\} \right].$$

- 由于 $f(x) = x^{n/2} \exp\{-x/2\}$ 在 $x = n$ 处取最大值, 可知上式在 $\lambda_1 = \dots = \lambda_p = n$ 时取最大值。
- 因此, Σ 的极大似然估计 $\hat{\Sigma}$ 满足条件:

$$\hat{\Sigma}^{-1/2} \mathbf{V} \hat{\Sigma}^{-1/2} = n \mathbf{I}_p.$$

- 则, Σ 的MLE为: $\hat{\Sigma} = \mathbf{V}/n$ 。
- 可见, 似然函数的最大值为:

$$L(\bar{\mathbf{x}}, \hat{\Sigma}) = \frac{1}{(2\pi)^{np/2}} e^{-np/2} \frac{1}{|\hat{\Sigma}|^{n/2}}.$$

判断估计量好坏的准则

在统计学中，评价参数的点估计通常有一些准则：

- ① 无偏性
- ② 充分性
- ③ 相合性
- ④ 完备性
- ⑤ 有效性
- ⑥ minimax性

无偏性

令 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 是来自总体 $X \sim f(\mathbf{x}, \boldsymbol{\theta})$ 的一组独立随机样本，其中 $\boldsymbol{\theta}$ 是未知的参数向量。假设 $\hat{\boldsymbol{\theta}}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ 是 $\boldsymbol{\theta}$ 的一个估计，并且它是一个统计量。对于不同的样本 $\mathbf{x}_1, \dots, \mathbf{x}_n$ ，估计 $\hat{\boldsymbol{\theta}}$ 取不同的值，如果 $\hat{\boldsymbol{\theta}}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ 的均值等于未知参数向量 $\boldsymbol{\theta}$ ，即

$$E[\hat{\boldsymbol{\theta}}(\mathbf{x}_1, \dots, \mathbf{x}_n)] = \boldsymbol{\theta}, \quad \text{对一切可能的 } \boldsymbol{\theta} \text{ 成立,}$$

则称 $\hat{\boldsymbol{\theta}}$ 为 $\boldsymbol{\theta}$ 的 **无偏估计**。

- 根据定理5.1.1的结论(2): $\mathbf{V} \sim W_p(n-1, \Sigma)$, 则 $E(\mathbf{V}) = (n-1)\Sigma$, 所以极大似然估计 $\hat{\Sigma} = \mathbf{V}/n$ 并不是 Σ 的无偏估计。
- 记样本协方差阵 $\mathbf{S} = \mathbf{V}/(n-1)$, 则样本协方差阵 \mathbf{S} 才是 Σ 的无偏估计。

定理5.2.1

设 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ ($i = 1, \dots, n$) 为来自 p 元正态总体 $N_p(\mu, \Sigma)$ 的一组随机样本, $n > p$, $\bar{\mathbf{x}}$ 为样本均值向量, \mathbf{V} 为样本离差阵, \mathbf{S} 为样本协方差阵, 则 μ 和 Σ 的MLE分别为 $\hat{\mu} = \bar{\mathbf{x}}$ 和 $\hat{\Sigma} = \mathbf{V}/n$ 。进一步, μ 和 Σ 的无偏估计分别为 $\bar{\mathbf{x}}$ 和 \mathbf{S} 。

- **充分统计量**：就是包含了样本中对兴趣待估参数向量全部信息的统计量。
- 如何判断一个统计量是充分统计量？

Neyman-Fisher因子判别法则

令 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 是来自总体 $X \sim f(\mathbf{x}, \boldsymbol{\theta})$ 的一组独立随机样本，其中 $\boldsymbol{\theta}$ 是未知的兴趣待估参数向量。设 $t \equiv t(\mathbf{x}_1, \dots, \mathbf{x}_n)$ 为一统计量，若样本的分布密度函数可以分解为：

$$\prod_{i=1}^n f(\mathbf{x}_i, \boldsymbol{\theta}) = g(t, \boldsymbol{\theta}) h(\mathbf{x}_1, \dots, \mathbf{x}_n),$$

其中 $h(\mathbf{x}_1, \dots, \mathbf{x}_n)$ 与 $\boldsymbol{\theta}$ 无关； $g(t, \boldsymbol{\theta})$ 可能与参数向量 $\boldsymbol{\theta}$ 有关，但与样本有关是通过统计量 t 发生关系，则称 t 是 $\boldsymbol{\theta}$ 的充分统计量。

定理5.2.2

设 $X \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 并且 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'(i = 1, \dots, n)$ 为来自 p 元正态总体 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的一组随机样本, $n > p$, $\bar{\mathbf{x}}$ 为样本均值向量, \mathbf{S} 为样本协方差阵, 则

- ① 当 $\boldsymbol{\Sigma}$ 已知时, $\bar{\mathbf{x}}$ 是 $\boldsymbol{\mu}$ 的充分统计量;
- ② 当 $\boldsymbol{\mu}$ 已知时, $\sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})'$ 是 $\boldsymbol{\Sigma}$ 的充分统计量;
- ③ 当 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 未知时, $\bar{\mathbf{x}}$ 和 \mathbf{S} 分别为 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的充分统计量。

几点说明:

- 根据样本 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 的联合密度函数和Neyman-Fisher 因子判别法则, 很容易完成定理5.2.2的证明;
- 定理5.2.2说明, $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 一切“好的”估计都是 $(\boldsymbol{\mu}, \mathbf{S})$ 的函数;
- 对多元正态总体而言, 充分统计量的重要性体现在关于 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的所有信息包含在 $\bar{\mathbf{x}}$ 和 \mathbf{S} 中;
- 对于非多元正态总体的情形, 一般是不正确的, 除了 $\bar{\mathbf{x}}$ 和 \mathbf{S} 的信息外, 还有其他有用的样本信息。

相合性

令 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 是来自总体 $X \sim f(\mathbf{x}, \boldsymbol{\theta})$ 的一组独立随机样本，其中 $\boldsymbol{\theta}$ 是未知的兴趣待估参数向量， $\boldsymbol{\theta}$ 的变化范围为参数空间 Θ ， $g(\boldsymbol{\theta})$ 是 $\boldsymbol{\theta}$ 的函数。设 $T_n \equiv T_n(\mathbf{x}_1, \dots, \mathbf{x}_n)$ 为样本 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 的统计量。如果对任给的 $\varepsilon > 0$ ，有

$$\lim_{n \rightarrow \infty} \Pr\{|T_n - g(\boldsymbol{\theta})| > \varepsilon\} = 0,$$

则称 T_n 是 $g(\boldsymbol{\theta})$ 的弱相合估计，记为 $T_n \xrightarrow{P} g(\boldsymbol{\theta})$ 。如果

$$\Pr\left\{\lim_{n \rightarrow \infty} T_n = g(\boldsymbol{\theta})\right\} = 1,$$

则称 T_n 是 $g(\boldsymbol{\theta})$ 的强相合估计，记为 $T_n \xrightarrow{a.s.} g(\boldsymbol{\theta})$ 。

- 注: $T_n \xrightarrow{a.s.} g(\boldsymbol{\theta}) \Rightarrow T_n \xrightarrow{P} g(\boldsymbol{\theta})$

定理5.2.3

在定理5.2.1的假设及记号下, $\bar{\mathbf{x}}$ 和 \mathbf{V}/n 分别为 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的强(弱)相合估计。

证明： 由于 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})' (i = 1, \dots, n)$ 为来自 p 元正态总体 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的一组随机样本， $n > p$, $\bar{\mathbf{x}}$ 为样本均值向量。从而由 **Kolmogorov 强大数律** (若 $\{X_i\}$ 为相互独立同分布的随机变量, 则

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{a.s.} c \iff E|X_1| < \infty, \quad c = EX_1).$$

显然 $\bar{\mathbf{x}}$ 为 $\boldsymbol{\mu}$ 的强相合估计。

由于 $\boldsymbol{\Sigma}$ 的极大似然估计为 $\frac{1}{n}\mathbf{V}$, 记

$$\mathbf{V}/n = (v_{ij}/n)_{1 \leq i, j \leq p}$$

且

$$\begin{aligned}\frac{1}{n}v_{ij} &= \frac{1}{n} \sum_{k=1}^n (x_{ki} - \bar{x}_i) (x_{kj} - \bar{x}_j) \\ &= \frac{1}{n} \sum_{k=1}^n x_{ki}x_{kj} - \left(\frac{1}{n} \sum_{k=1}^n x_{ki} \right) \left(\frac{1}{n} \sum_{k=1}^n x_{kj} \right).\end{aligned}$$

显然 $\frac{1}{n} \sum_{k=1}^n x_{ki} \xrightarrow{a.s.} \mu_i$, $\frac{1}{n} \sum_{k=1}^n x_{kj} \xrightarrow{a.s.} \mu_j$, 而

$$E(x_{ki}x_{kj}) = E(x_{ki} - \mu_i)(x_{kj} - \mu_j) + \mu_i\mu_j = \sigma_{ij} + \mu_i\mu_j.$$

由Cauchy-Schwarz不等式, 则有

$$E |x_{ki}x_{kj}| \leq [Ex_{ki}^2 Ex_{kj}^2]^{1/2} = [(\sigma_{ii} + \mu_i^2)(\sigma_{jj} + \mu_j^2)]^{1/2} < \infty.$$

因此由Kolmogorov 强大数律有

$$\frac{1}{n}v_{ij} \xrightarrow{a.s.} \sigma_{ij}.$$

综合上面的结果, 可证得 \mathbf{V}/n 为 Σ 的强相合估计。

定理5.2.4

假设 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 是来自 p 元正态分布 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的 i.i.d. 随机样本, 记 $\mathbf{V} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$, 则

$$\mathbf{B}_n = \frac{1}{\sqrt{n}}(\mathbf{V} - n\boldsymbol{\Sigma})$$

收敛于一个正态随机矩阵 $\mathbf{B} = (b_{ij})$, 其元素均值为 0, 协方差为

$$E(b_{ij}b_{kl}) = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}.$$

- 完备统计量的理解：
- 不管参数 θ 怎么变化，统计量 T 的任何一种构造(比如函数 $g(T)$)，都不可能为0的无偏估计，即无法得到

$$E_{\theta}(g(T)) = 0.$$

除非 $g(T)$ 本身就是0。

- **完备统计量**和**充分统计量**一样，也是为寻求参数估计的优良统计量起重要的作用。
- 考虑参数统计模型 $(\mathcal{X}, \mathcal{B}, \mathcal{F})$ ，其中
 - ① \mathcal{X} 为样本空间，为样本 X 的一切可能取值；
 - ② \mathcal{B} 是 \mathcal{X} 的某些子集构成的 σ 域，并称 $(\mathcal{X}, \mathcal{B})$ 为**可测空间**；
 - ③ $\mathcal{F} = \{F_{\theta} : \theta \in \Theta\}$ 为关于参数向量 θ 的分布族。
- 设 $\phi(\mathbf{x})$ 是定义在 \mathcal{X} 上的 **\mathcal{B} 可测函数**
- 则 $\phi(X)$ 的期望为：

$$E_{\theta}[\phi(X)] = \int_{\mathcal{X}} \phi(\mathbf{x}) dF_{\theta}(\mathbf{x}), \quad \theta \in \Theta,$$

其中 E_{θ} 是强调期望是在参数为 θ 的分布下进行的。

- 希望 $E_{\theta}[\phi(X)]$ 对每一个固定的 θ 都有唯一确定的值。
- 这时考虑下面的命题成立：

$$\phi_1(\mathbf{x}) = \phi_2(\mathbf{x}), \quad a.s. F_{\theta} \Leftrightarrow E_{\theta}[\phi_1(X)] = E_{\theta}[\phi_2(X)].$$

- 必要性“ \Rightarrow ”是显然的；
- 充分性“ \Leftarrow ”：如果命题

$$E_{\theta}[\phi(X)] = 0 \Rightarrow \phi(\mathbf{x}) = 0, \quad a.s. F_{\theta}$$

成立，就可证明上面的充分性。

- 因为 $E_{\theta}[\phi_1(X)] = E_{\theta}[\phi_2(X)]$ ，则 $E_{\theta}[\phi_1(X) - \phi_2(X)] = 0$
- 故可利用上面的命题，可证得：

$$\phi_1(\mathbf{x}) = \phi_2(\mathbf{x}), \quad a.s. F_{\theta}.$$

完备性

设 T 是一个连续的随机变量，其密度函数为 $f_T(t, \theta)$, $\theta \in \Theta$, Θ 为参数 θ 变化的范围。分布密度族 $\{f_T(t, \theta) : \theta \in \Theta\}$ 称为**完备的**，如果对任意的实数函数 $g(T)$ ，当

$$E_{\theta}[g(T)] = \int g(t)f_T(t, \theta)dt = 0$$

对每个 $\theta \in \Theta$ 成立时，能推出 $g(T)$ 几乎处处为0，或 $\Pr_{\theta}(g(T) = 0) = 1$ 对任何 $\theta \in \Theta$ 。如果一个充分统计量的分布密度族是完备的，则称它是**完备充分统计量**。

注：定理5.2.2证明了 (\bar{x}, S) 是 (μ, Σ) 的充分统计量，同样也可以证明它们是完备的，详细的证明可见Anderson (2003)。因此称 **(\bar{x}, S) 是 (μ, Σ) 的完备充分统计量**。

有效性

令 X_1, \dots, X_n 是来自总体 $X \sim f(x, \boldsymbol{\theta})$ 的一组独立随机样本，其中 $\boldsymbol{\theta}$ 是未知的兴趣待估参数向量， $\boldsymbol{\theta}$ 的变化范围为参数空间 Θ ， $g(\boldsymbol{\theta})$ 是 $\boldsymbol{\theta}$ 的函数。设 T 为样本 X_1, \dots, X_n 的统计量，且是 $g(\boldsymbol{\theta})$ 的无偏估计。如果对 $g(\boldsymbol{\theta})$ 的任一无偏估计 \tilde{T} 都有

$$E[T - g(\boldsymbol{\theta})]^2 \leq E[\tilde{T} - g(\boldsymbol{\theta})]^2, \quad \text{对任意 } \boldsymbol{\theta} \in \Theta,$$

则称 T 为 $g(\boldsymbol{\theta})$ 的**有效估计**。

- **推广：**当待估函数 $\mathbf{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), \dots, g_r(\boldsymbol{\theta}))'$ 是一个向量时，设它的某个无偏估计 $\mathbf{T} = (T_1, \dots, T_r)'$ 有协方差矩阵。如果对于任意一个具有协方差矩阵的无偏估计 $\tilde{\mathbf{T}}$ ，有

$$\text{Cov}(\mathbf{T}) \leq \text{Cov}(\tilde{\mathbf{T}}), \quad \text{对任意 } \boldsymbol{\theta} \in \Theta,$$

即 $\text{Cov}(\tilde{\mathbf{T}}) - \text{Cov}(\mathbf{T})$ 是非负定矩阵，则称 \mathbf{T} 为待估函数 $\mathbf{g}(\boldsymbol{\theta})$ 的有效估计。

定理5.2.5

在定理1的假设及记号下， $(\bar{\mathbf{x}}, \mathbf{S})$ 为 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的一致最小方差无偏估计(UMVUE)。

相关系数的极大似然估计

- 在有了 μ 和 Σ 的极大似然估计 $\hat{\mu} = \bar{x}$ 和 $\hat{\Sigma} = V/n$ 后, 我们是否可以通过使用 $\hat{\mu}$ 和 $\hat{\Sigma}$ 替换前面定义过的回归系数, 相关系数, 条件协方差阵和偏相关系数等中的 μ 和 Σ 来得到相应的MLE? 由下面引理知道这是可以的。

引理5.4.1

设 θ 的极大似然估计为 $\hat{\theta}$, 若 $\theta \rightarrow \phi(\theta)$ 为一一变换, 则 $\phi(\theta)$ 的极大似然估计为 $\phi(\hat{\theta})$ 。

问题

求相关系数的极大似然估计?

相关系数的极大似然估计

解: 由相关系数矩阵 $\mathbf{R} = \mathbf{D}^{-1/2} \boldsymbol{\Sigma} \mathbf{D}^{-1/2}$, 其中

$$\mathbf{D}^{1/2} = \text{diag}(\sqrt{\sigma_{11}}, \dots, \sqrt{\sigma_{pp}}),$$

$\boldsymbol{\Sigma} \rightarrow (\mathbf{D}, \mathbf{R})$ 一一对应, 因此由 $\boldsymbol{\Sigma}$ 的极大似然估计 $\hat{\boldsymbol{\Sigma}}$, 可知 $\hat{\mathbf{D}} = \sqrt{\text{diag}(\hat{\boldsymbol{\Sigma}})}$, 则 \mathbf{R} 的MLE为:

$$\hat{\mathbf{R}} = \hat{\mathbf{D}}^{-1/2} \hat{\boldsymbol{\Sigma}} \hat{\mathbf{D}}^{-1/2},$$

其 (i, j) 元素为, $i, j = 1, \dots, p$

$$\hat{\rho}_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i) (x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} = \frac{v_{ij}}{\sqrt{v_{ii} v_{jj}}} =: r_{ij}.$$

■ 假设随机向量 $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $p > 2$, 将 \mathbf{X} , $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 剖分为

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

- ▷ $\mathbf{X}^{(1)}$ 和 $\boldsymbol{\mu}^{(1)}$ 为 $q \times 1$ 的向量
- ▷ $\boldsymbol{\Sigma}_{11}$ 为 $q \times q$ 矩阵
- ▷ $\mathbf{X}^{(2)}$ 和 $\boldsymbol{\mu}^{(2)}$ 为 $(p - q) \times 1$ 的向量
- ▷ $\boldsymbol{\Sigma}_{22}$ 为 $(p - q) \times (p - q)$ 矩阵
- ▷ $\boldsymbol{\Sigma}_{12}$ 为 $q \times (p - q)$ 矩阵

样本偏相关系数

- 给定 $\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$ 时 $\mathbf{X}^{(1)}$ 的条件分布服从 q 元正态分布, 即 $(\mathbf{X}^{(1)} | \mathbf{X}^{(2)} = \mathbf{x}^{(2)}) \sim N_q(\boldsymbol{\mu}^{(1)} + \boldsymbol{\beta}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}), \boldsymbol{\Sigma}_{11.2})$.
- $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ 为 $\mathbf{X}^{(1)}$ 对 $\mathbf{X}^{(2)}$ 的 **回归系数**。
- $\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ 为 **条件协方差矩阵**, 它的元素用 $\sigma_{ij \cdot q+1, \dots, p}$ 表示, $i, j = 1, \dots, q$ 。
- 给定 $\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$ 条件下, X_i 与 X_j 的 **偏相关系数** $(\text{corr}(\mathbf{X}^{(1)} | \mathbf{X}^{(2)}))$ 定义为:

$$\rho_{ij \cdot q+1, \dots, p} = \frac{\sigma_{ij \cdot q+1, \dots, p}}{(\sigma_{ii \cdot q+1, \dots, p} \sigma_{jj \cdot q+1, \dots, p})^{1/2}}, \quad i, j = 1, \dots, q.$$

- 类似地，对样本均值 $\bar{\mathbf{x}}$ ，离差矩阵 \mathbf{V} 和样本协方差矩阵 \mathbf{S} 进行如下剖分：

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{\mathbf{x}}^{(1)} \\ \bar{\mathbf{x}}^{(2)} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix}, \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix},$$
$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix}.$$

- 由定理5.1.2，在给定 $\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$ 时，则条件期望 $\mu_{1.2}$ 的MLE为：

$$\begin{aligned} \hat{\mu}_{1.2} &= \bar{\mathbf{x}}^{(1)} + (\mathbf{V}_{12}/n)(\mathbf{V}_{22}/n)^{-1}(\mathbf{x}^{(2)} - \bar{\mathbf{x}}^{(2)}) \\ &= \bar{\mathbf{x}}^{(1)} + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{x}^{(2)} - \bar{\mathbf{x}}^{(2)}). \end{aligned}$$

■ 由 $\mathbf{V}_{12}\mathbf{V}_{22}^{-1} = \mathbf{S}_{12}\mathbf{S}_{22}^{-1}$, 所以条件期望 $\mu_{1.2}$ 的MLE也可表示为:

$$\hat{\mu}_{1.2} = \bar{\mathbf{x}}^{(1)} + \mathbf{S}_{12}\mathbf{S}_{22}^{-1}(\mathbf{x}^{(2)} - \bar{\mathbf{x}}^{(2)}).$$

■ 条件协方差阵 $\Sigma_{11.2}$ 的MLE为:

$$\hat{\Sigma}_{11.2} = \frac{\mathbf{V}_{11}}{n} - \frac{\mathbf{V}_{12}}{n} \left(\frac{\mathbf{V}_{22}}{n} \right)^{-1} \frac{\mathbf{V}_{21}}{n} = \frac{\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}}{n} = \frac{\mathbf{V}_{11.2}}{n}.$$

■ 可得偏相关系数的极大似然估计为:

$$\hat{\rho}_{ij \cdot q+1, \dots, p} = \frac{\hat{\sigma}_{ij \cdot q+1, \dots, p}}{(\hat{\sigma}_{ii \cdot q+1, \dots, p} \hat{\sigma}_{jj \cdot q+1, \dots, p})^{1/2}}, \quad i, j = 1, \dots, q,$$

其中 $\hat{\sigma}_{ij \cdot q+1, \dots, p}$ 是 $\hat{\Sigma}_{11.2}$ 的第 (i, j) 元素。

定理5.4.4

设 $X \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 并且 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ ($i = 1, \dots, n$) 为来自 p 元正态总体 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的一组随机样本, $n > p$, $\bar{\mathbf{x}}$ 为样本均值向量, \mathbf{V} 为样本离差阵, \mathbf{S} 为样本协方差阵, 在给定 $\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$ 时, 则

- ① 条件期望 $\boldsymbol{\mu}_{1.2}$ 和条件协方差阵 $\boldsymbol{\Sigma}_{11.2}$ 的MLE分别为 $\hat{\boldsymbol{\mu}}_{1.2} = \bar{\mathbf{x}}^{(1)} + \mathbf{S}_{12}\mathbf{S}_{22}^{-1}(\mathbf{x}^{(2)} - \bar{\mathbf{x}}^{(2)})$ 和 $\hat{\boldsymbol{\Sigma}}_{11.2} = \mathbf{V}_{11.2}/n$, 同时 $\boldsymbol{\beta}$ 的MLE为 $\hat{\boldsymbol{\beta}} = \mathbf{V}_{12}\mathbf{V}_{22}^{-1}$;
- ② 进一步, 条件期望 $\boldsymbol{\mu}_{1.2}$ 和条件协方差阵 $\boldsymbol{\Sigma}_{11.2}$ 的无偏估计分别为 $\bar{\mathbf{x}}^{(1)} + \mathbf{S}_{12}\mathbf{S}_{22}^{-1}(\mathbf{x}^{(2)} - \bar{\mathbf{x}}^{(2)})$ 和 $\mathbf{S}_{11.2}$, 其中 $\mathbf{S}_{11.2} = \mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21} = \mathbf{V}_{11.2}/(n-1)$ 。

定理5.4.5

设 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 为来自 p 元正态总体 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的一组随机样本, $n > p$, 则给定后 $p - q$ 个分量条件下, 前 q 个分量的偏相关系数 $\rho_{ij \cdot q+1, \dots, p}$ 的MLE为

$$\hat{\rho}_{ij \cdot q+1, \dots, p} = \frac{v_{ij \cdot q+1, \dots, p}}{(v_{ii \cdot q+1, \dots, p} v_{jj \cdot q+1, \dots, p})^{1/2}}, \quad i, j = 1, \dots, q,$$

其中 $(v_{ij \cdot q+1, \dots, p}) = \mathbf{V}_{11} - \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21} = \mathbf{V}_{11.2}$ 。

- 为了简单, 把极大似然估计 $\hat{\rho}_{ij \cdot q+1, \dots, p}$ 记为 $r_{ij \cdot q+1, \dots, p}$, 称为**样本偏相关系数**。
- 由估计的回归系数 $\hat{\beta}$, 矩阵 $\mathbf{V}_{11.2}$ 也可以表示为:

$$\begin{aligned}\mathbf{V}_{11.2} &= \sum_{k=1}^n \left[\mathbf{x}_k^{(1)} - \bar{\mathbf{x}}^{(1)} - \hat{\beta}(\mathbf{x}_k^{(2)} - \bar{\mathbf{x}}^{(2)}) \right] \left[\mathbf{x}_k^{(1)} - \bar{\mathbf{x}}^{(1)} - \hat{\beta}(\mathbf{x}_k^{(2)} - \bar{\mathbf{x}}^{(2)}) \right]' \\ &= \mathbf{V}_{11} - \hat{\beta} \mathbf{V}_{22} \hat{\beta}'\end{aligned}$$

- 其中向量 $\hat{\boldsymbol{\varepsilon}}_k^{(1.2)} = \mathbf{x}_k^{(1)} - \bar{\mathbf{x}}^{(1)} - \hat{\beta}(\mathbf{x}_k^{(2)} - \bar{\mathbf{x}}^{(2)})$ 称为 $\mathbf{x}_k^{(1)}$ 对 $\mathbf{x}_k^{(2)}$ 的回归残差, 样本偏相关系数就是这些回归残差之间简单的相关系数。

Σ 已知时, μ 的置信域

- 设 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 是来自 p 元正态总体 $N_p(\mu, \Sigma)$ 的随机样本, 其中 $\mu \in \mathbb{R}^p$, $\Sigma > 0$, $n > p$.
- 样本均值 $\bar{\mathbf{x}}$ 和协方差阵 \mathbf{S} 分别是 μ 和 Σ 的无偏估计.
- 当 Σ 已知时, 由于 $\bar{\mathbf{x}} \sim N_p(\mu, \Sigma/n)$, 根据多元正态分布的性质, 有

$$n(\bar{\mathbf{x}} - \mu)' \Sigma^{-1} (\bar{\mathbf{x}} - \mu) \sim \chi_p^2.$$

- 可得 μ 的置信水平为 $1 - \alpha$ 的置信域为超椭圆:

$$D = \{\mu^* \in \mathbb{R}^p : n(\bar{\mathbf{x}} - \mu^*)' \Sigma^{-1} (\bar{\mathbf{x}} - \mu^*) \leq c_\alpha\},$$

这里 c_α 表示自由度为 p 的 χ^2 分布的上侧 α 分位点, 即 $\Pr(\chi_p^2 > c_\alpha) = \alpha$.

Σ 已知时, μ 的置信域

R语言中有两个函数可以构造:

- 程序包mixtools中的函数ellipse()
- 程序包car中的函数dataEllipse()

用两种数据产生方法从二元正态总体 $N_2(\mu, \Sigma)$ 中生成1000个模拟数据, 其中

$$\mu = (0, 2)', \quad \Sigma = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}.$$

- 第一种: 直接从函数mvrnorm()生成数据
- 第二种: 基于多元正态分布的定义生成数据, 即 $\mathbf{X} = \mu + \mathbf{A}\mathbf{Y} \sim N_p(\mu, \mathbf{A}\mathbf{A}')$, 其中

$$\mathbf{Y} = (Y_1, \dots, Y_p)' \sim N_p(\mathbf{0}, \mathbf{I}_p),$$

$$\mathbf{A} = \Gamma \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p}) \Gamma'.$$

```
library(MASS); library(car)
set.seed(1); n = 1000
mu = c(0, 2); Sig = matrix(c(1, 0.7, 0.7, 1), 2, 2)
## 第一种数据产生方法:
biv1=mvrnorm(n, mu, Sig);colnames(biv1)=c("X","Y")
## 第二种由定义3.8产生方法:
sig.eigen = eigen(Sig)
Sigma=sig.eigen$vec%*%diag(sqrt(sig.eigen$val))%*%t(sig.eigen$vec)
biv2 = mu+t(matrix(rnorm(2*n),ncol=2)%*%Sigma)
biv2 = t(biv2); colnames(biv2) = c("X","Y")
par(mfrow = c(1, 2))
dataEllipse(biv1[, 1], biv1[, 2], levels = 0.95)
points(t(mu), col = 'red', pch = 8)
dataEllipse(biv2[, 1], biv2[, 2], levels = 0.95)
points(t(mu), col = 'red', pch = 8)
```

Σ 已知时, μ 的置信域

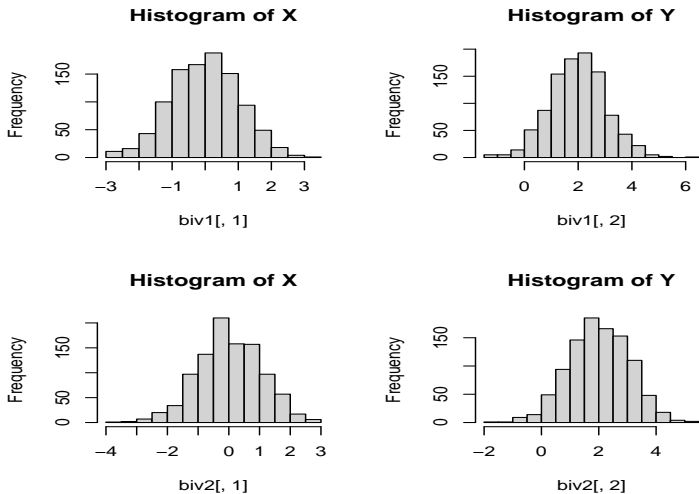


Figure: 两种数据的直方图。

Σ 已知时, μ 的置信域

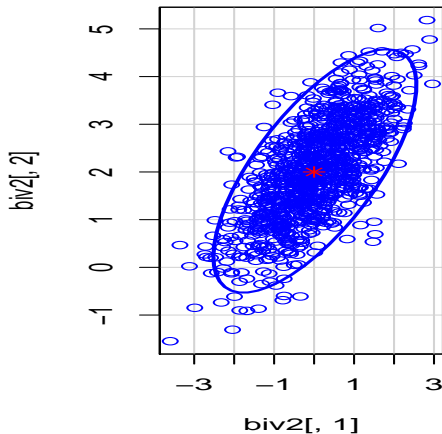
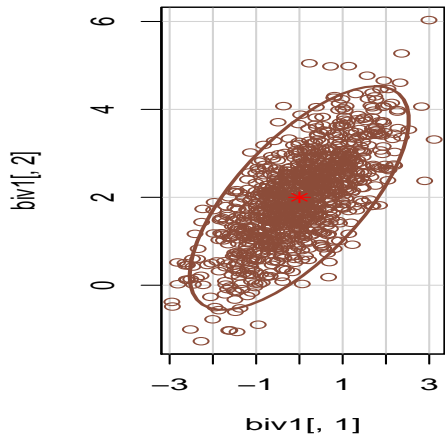


Figure: 均值向量 $(\mu_1, \mu_2)'$ 的95%置信域, 其中左图是第一种数据产生方法的95% 置信域, 右图是第二种数据产生方法的95%置信域。

- 当 Σ 未知时, 用无偏估计 \mathbf{S} 替换 Σ , 则有

$$\begin{aligned}T^2 &= n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\&= n(n-1)(\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{V}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\&\sim T^2(p, n-1).\end{aligned}$$

- 或者由Hotelling T^2 分布的性质4.3.4可知

$$\begin{aligned}\frac{n-p}{(n-1)p} T^2(p, n-1) &= \frac{n(n-p)}{p} (\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{V}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\&\sim F_{p, n-p}.\end{aligned}$$

- 此结果可得 μ 的置信水平为 $1 - \alpha$ 的置信域为超椭圆:

$$\mathcal{D} = \left\{ \mu^* \in \mathbb{R}^p : \frac{n(n-p)}{p} (\bar{\mathbf{x}} - \mu^*)' \mathbf{V}^{-1} (\bar{\mathbf{x}} - \mu^*) \leq c_{\alpha}^* \right\}$$

或

$$\mathcal{D} = \left\{ \mu^* \in \mathbb{R}^p : \frac{n(n-p)}{p(n-1)} (\bar{\mathbf{x}} - \mu^*)' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \mu^*) \leq c_{\alpha}^* \right\}, \quad (1)$$

这里 c_{α}^* 表示自由度为 p 和 $n-p$ 的 F 分布的上侧 α 分位点, 即 $\Pr(F_{p,n-p} > c_{\alpha}^*) = \alpha$ 。

Σ 未知时, μ 的置信域

- 显然, 置信域是以 $\bar{\mathbf{x}}$ 为中心, 椭球的轴为:

$$\pm \sqrt{\lambda_i} \sqrt{\frac{p}{n(n-p)} F_{p,n-p}(\alpha)} \mathbf{e}_i,$$

其中 $\mathbf{V}\mathbf{e}_i = \lambda_i \mathbf{e}_i, \quad i = 1, \dots, p$ 。

- 等价的有:

$$\pm \sqrt{\lambda_i} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)} \mathbf{e}_i,$$

其中 $\mathbf{S}\mathbf{e}_i = \lambda_i \mathbf{e}_i, \quad i = 1, \dots, p$ 。

Σ 未知时, μ 的置信域

例子

考虑45个雄性钩嘴鸢的尾巴长度(X , 单位: 毫米)和翅膀长度(Y , 单位: 毫米)的数据, 该数据来源于Johnson 和Wichern (2008), 并见下表。假设该数据来自一个二元正态分布 $(X, Y)' \sim N_2(\mu, \Sigma)$ 。试编程绘制尾巴长度(X) 和翅膀长度(Y)的均值向量 $\mu = (\mu_1, \mu_2)'$ 置信水平为95%的置信域。

Σ 未知时, μ 的置信域

序号	X	Y	序号	X	Y	序号	X	Y
1	180	278	16	185	282	31	284	277
2	186	277	17	195	285	32	176	281
3	206	308	18	183	276	33	185	287
4	184	290	19	202	308	34	191	295
5	177	273	20	177	254	35	177	267
6	177	284	21	177	268	36	197	310
7	176	267	22	170	260	37	199	299
8	200	281	23	186	274	38	190	273
9	191	287	24	177	272	39	180	278
10	193	271	25	178	266	40	189	280
11	212	302	26	192	281	41	194	290
12	181	254	27	204	276	42	186	287
13	195	297	28	191	290	43	191	286
14	187	281	29	178	265	44	187	288
15	190	284	30	177	275	45	186	275

```
conf.ellipse = function(xdata, alpha){  
  if(ncol(xdata)!=2) stop("Only for bivariate normal")  
  n = nrow(xdata); xbar = colMeans(xdata)  
  S = cov(xdata); es = eigen(S)  
  e1 = es$vec %*% diag(sqrt(es$val))  
  r1 = sqrt(qf(alpha,2,n-2))*sqrt(2*(n-1)/(n*(n-2)))  
  theta = seq(0,2*pi, len=250)  
  v1 = cbind(r1*cos(theta), r1*sin(theta))  
  pts = t(xbar-(e1%*%t(v1)))  
  plot(pts,type="l",col='red',lwd=3,main="Confidence Region for  
    Bivariate Normal", xlab="X", ylab="Y", asp=1)  
  grid(lty = 1, equilogs = FALSE)
```

```
segments(-0.2,xbar[2],xbar[1],xbar[2],lty=2,lwd=2)
segments(xbar[1],0,xbar[1],xbar[2],lty=2,lwd=2)
th2 = c(0,pi/2,pi,3*pi/2,2*pi)
v2  = cbind(r1*cos(th2), r1*sin(th2))
pts2 = t(xbar-(e1%*%t(v2)))
segments(pts2[3,1],pts2[3,2],pts2[1,1],pts2[1,2],lty=4,lwd=2)
segments(pts2[2,1],pts2[2,2],pts2[4,1],pts2[4,2],lty=4,lwd=2)
}

source("conf.ellipse.R")
XY.data = read.table("Hook.DAT")
colnames(XY.data) = c("Tail length", "Wing length")
conf.ellipse(XY.data,alpha=0.95)
```

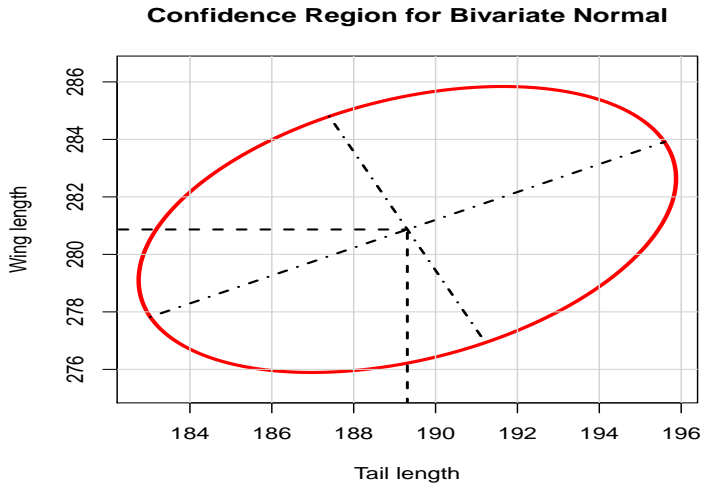


Figure: 95%的置信域。

大样本置信区间

当样本量 n 很大时，对均值向量的推断可以不假设多元正态成立。由大样本理论，知

$$\Pr[n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \leq \chi_p^2(\alpha)] \approx 1 - \alpha.$$

定理6.1.1

设 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 为来自均值 $\boldsymbol{\mu}$ ，协方差矩阵 $\boldsymbol{\Sigma}$ 的总体的简单随机样本，其中 $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\Sigma} > 0$ 均未知，令 $\bar{\mathbf{x}}, \mathbf{S}$ 分别为样本均值和样本协方差矩阵，则当 $n - p \rightarrow \infty$ 时， $\boldsymbol{\mu}$ 的一个渐近 $1 - \alpha$ 的置信域为：

$$\left\{ \boldsymbol{\mu}^* \in \mathbb{R}^p \mid n(\bar{\mathbf{x}} - \boldsymbol{\mu}^*)' \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}^*) \leq \chi_p^2(\alpha) \right\}.$$

注：由置信区间和假设检验的等价性，也可用于 $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0 \leftrightarrow H_1: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ 的检验问题。

两个多元正态总体的置信域

- 设 X 和 Y 相互独立, 且 $X \sim N_p(\mu_1, \Sigma)$ 和 $Y \sim N_p(\mu_2, \Sigma)$, $\mu_1, \mu_2 \in \mathbb{R}^p$, $\Sigma > 0$ 。
- $\mathbf{x}_1, \dots, \mathbf{x}_n$ 和 $\mathbf{y}_1, \dots, \mathbf{y}_m$ 是分别来自总体 X 和 Y 的随机样本, $n, m > p$ 。
- 记 $\bar{\mathbf{x}}$ 和 $\bar{\mathbf{y}}$ 分别为它们的样本均值向量。
- 样本 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 的离差矩阵和样本协方差阵为

$$\mathbf{V}_1 = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})', \quad \mathbf{S}_1 = \frac{1}{n-1} \mathbf{V}_1.$$

- 样本 $\mathbf{y}_1, \dots, \mathbf{y}_m$ 的离差矩阵和样本协方差阵分别为

$$\mathbf{V}_2 = \sum_{i=1}^m (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})', \quad \mathbf{S}_2 = \frac{1}{m-1} \mathbf{V}_2.$$

Σ 已知时, δ 的置信域

问题: 讨论总体均值差 $\delta = \mu_1 - \mu_2$ 的置信域估计问题。

- 当 Σ 已知时, 由于 $\bar{\mathbf{x}} \sim N_p(\mu_1, \Sigma/n)$, $\bar{\mathbf{y}} \sim N_p(\mu_2, \Sigma/m)$, 并且 $\bar{\mathbf{x}}$ 和 $\bar{\mathbf{y}}$ 相互独立, 则有

$$\bar{\mathbf{x}} - \bar{\mathbf{y}} \sim N_p\left(\delta, \left(\frac{1}{n} + \frac{1}{m}\right)\Sigma\right).$$

- 根据多元正态分布的性质, 有

$$\frac{nm}{n+m} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \delta]' \Sigma^{-1} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \delta] \sim \chi^2(p).$$

- 由此结果可得 δ 的置信水平为 $1 - \alpha$ 的置信域为超椭圆:

$$\mathcal{D} = \left\{ \delta^* \in \mathbb{R}^p : \frac{nm}{n+m} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \delta^*]' \Sigma^{-1} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \delta^*] \leq c_\alpha \right\},$$

这里, c_α 表示自由度为 p 的 χ^2 分布的上侧 α 分位点, 即 $\Pr(\chi_p^2 > c_\alpha) = \alpha$ 。

Σ 未知时, δ 的置信域

- 当 Σ 未知时, 为了构造 δ 的置信域, 首先需要找到 Σ 的最优无偏估计量。 Σ 唯一的最小协方差阵无偏估计为

$$\hat{\Sigma} = \frac{\mathbf{V}_1 + \mathbf{V}_2}{n + m - 2} = \frac{n - 1}{n + m - 2} \mathbf{S}_1 + \frac{m - 1}{n + m - 2} \mathbf{S}_2.$$

- 考虑下面的枢轴统计量

$$\frac{(n + m - 2)nm}{n + m} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \delta]' (\mathbf{V}_1 + \mathbf{V}_2)^{-1} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \delta].$$

- 由 $\bar{\mathbf{x}} - \bar{\mathbf{y}}$ 与 $\mathbf{V}_1 + \mathbf{V}_2$ 相互独立, 以及

$$\sqrt{\frac{nm}{n + m}} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \delta] \sim N_p(\mathbf{0}, \Sigma),$$
$$\mathbf{V}_1 + \mathbf{V}_2 \sim W_p(n + m - 2, \Sigma).$$

- 由上式和Hotelling T^2 分布的定义4.4, 有

$$\begin{aligned} T^2 &= \frac{(n+m-2)nm}{n+m} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \boldsymbol{\delta}]' (\mathbf{V}_1 + \mathbf{V}_2)^{-1} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \boldsymbol{\delta}] \\ &= (n+m-2) \left(\sqrt{\frac{nm}{n+m}} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \boldsymbol{\delta}] \right)' (\mathbf{V}_1 + \mathbf{V}_2)^{-1} \\ &\quad \cdot \left(\sqrt{\frac{nm}{n+m}} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \boldsymbol{\delta}] \right) \sim T^2(p, n+m-2). \end{aligned}$$

- 由Hotelling T^2 分布的性质4.3.3和4.3.4, 有

$$\begin{aligned}\frac{1}{n+m-2}T^2 &= \frac{nm}{n+m}[(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \boldsymbol{\delta}]'(\mathbf{V}_1 + \mathbf{V}_2)^{-1}[(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \boldsymbol{\delta}] \\ &\stackrel{d}{=} \frac{\chi_p^2}{\chi_{n+m-p-1}^2},\end{aligned}$$

其中分子和分母的 χ^2 分布相互独立。

- 进一步有

$$\begin{aligned}& \frac{n+m-p-1}{(n+m-2)p} T^2 \\&= \frac{(n+m-p-1)nm}{(n+m)p} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \boldsymbol{\delta}]' (\mathbf{V}_1 + \mathbf{V}_2)^{-1} [(\bar{\mathbf{x}} - \bar{\mathbf{y}}) - \boldsymbol{\delta}] \\&\sim F_{p, n+m-p-1}.\end{aligned}$$

- 此结果可得 δ 的置信水平为 $1 - \alpha$ 的置信域为超椭圆:

$$\mathcal{D} = \left\{ \delta^* \in \mathbb{R}^p : \frac{n + m - p - 1}{(n + m - 2)p} T^2 \leq c_{\alpha}^* \right\},$$

这里, c_{α}^* 表示自由度为 p 和 $n + m - p - 1$ 的 F 分布的上侧 α 分位点,

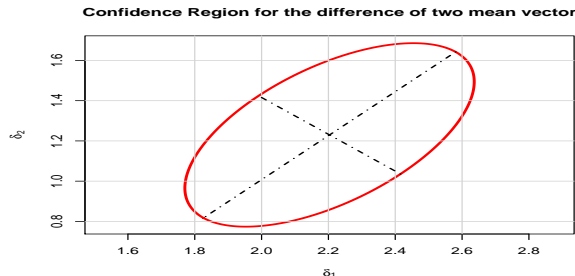
即 $\Pr(F_{p, n+m-p-1} > c_{\alpha}^*) = \alpha$ 。

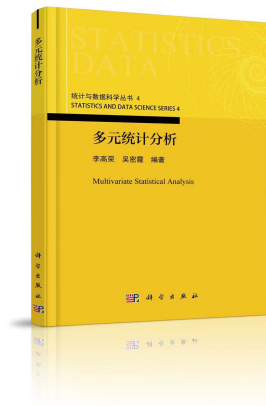
Σ 未知时, δ 的置信域

分别从二元正态总体 $X \sim N_2(\mu_1, \Sigma)$ 和 $Y \sim N_2(\mu_2, \Sigma)$ 中生成50个和45个模拟数据, 其中

$$\mu_1 = (10, 4)', \quad \mu_2 = (8, 3)', \quad \Sigma = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}.$$

编程序并构造均值差 $\mu_1 - \mu_2$ 的95%置信域。





谢谢，请多提宝贵意见！