多元统计分析第三次作业

学习交流，无限进步

2024 年 9 月 16 日

Exercise 1

6. 假设随机向量 $(X,Y)'$ 服从二维正态分布，即

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right).$$

记 $(x_1,y_1)',\ldots,(x_m,y_m)'$ 为来自该二维正态分布的容量为 m 的一组独立同分布随机样本。此外，额外观测了来自总体 $X \sim N(\mu_1,\sigma_1^2)$ 容量为 $n-m$ 的一组独立同分布随机样本 $x_{m+1},\ldots,x_n$，其中 $n>m$。试给出所有未知参数的极大似然估计。

证明. 似然函数：

$$\begin{aligned}
L(\mu,\Sigma) =& \prod_{i=1}^{m} \frac{1}{(2\pi)^{1/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(x_i-\mu_1,y_i-\mu_2)'\Sigma^{-1}(x_i-\mu_1,y_i-\mu_2)\} \\
&\times \prod_{j=1}^{(} n-m) \frac{1}{(2\pi)^{1/2}\sigma_1} \exp\{-\frac{1}{2\sigma_1^2}(x_{m+j}-\mu_1)^2\} \\
=& \prod_{i=1}^{m} \frac{1}{(2\pi)^{1/2}\sqrt{1-\rho^2}\sigma_1\sigma_2} \exp\{-\frac{1}{2(1-\rho^2)}(\frac{(x_i-\mu_1)^2}{\sigma_1^2}-2\rho(x_i-u_1)(y_i-\mu_2)+\frac{(y_i-\mu_2)^2}{\sigma_2^2})\} \\
&\times \prod_{j=1}^{(} n-m) \frac{1}{(2\pi)^{1/2}\sigma_1} \exp\{-\frac{1}{2\sigma_1^2}(x_{m+j}-\mu_1)^2\} \\
=& \frac{1}{(2\pi)^{(n)/2}(1-\rho^2)^{m/2}\sigma_1^n\sigma_2^m} \exp\{-\frac{1}{2}\times \\
&(\Sigma_{i=1}^{m}\frac{(x_i-\mu_1)^2}{(1-\rho^2)\sigma_1^2}+\Sigma_{i=1}^{(} n-m)\frac{(x_{m+i}-\mu_1)^2}{\sigma_1^2}+\Sigma_{i=1}^{m} 2\frac{\rho}{1-\rho^2}(x_i-u_1)(y_i-\mu_2)+\Sigma_{i=1}^{m}\frac{(y_i-\mu_2)^2}{\sigma_2^2}
\end{aligned}$$

对数似然：

$$\ln(L(\mu,\Sigma)) = -\ln\left((2\pi)^{(n)/2}(1-\rho^2)^{m/2}\right) - (m+n)\ln(\sigma_1) - m\ln(\sigma_2) - \frac{1}{2}(\Sigma_{i=1}^m \frac{(x_i-\mu_1)^2}{(1-\rho^2)\sigma_1^2} + \Sigma_{i=1}^(n-m)\frac{(x_{m+i}-\mu_1)^2}{\sigma_1^2} + \Sigma_{i=1}^m 2\frac{\rho}{1-\rho^2}(x_i-u_1)(y_i-\mu_2) + \Sigma_{i=1}^m \frac{(y_i-\mu_2)^2}{(1-\rho^2)\sigma_2^2}))$$

于是：

$$\frac{\partial\ln(L(\mu,\Sigma))}{\partial\mu_1} = \frac{\Sigma_{i=1}^m x_i - m\mu_1}{(1-\rho^2)\sigma_1^2} + \frac{\Sigma_{i=1}^n x_{m+i} - (n-m)\mu_1}{\sigma_1^2} + 2\frac{\rho}{1-\rho^2}(\Sigma_{i=1}^m y_i - m\mu_2) = 0$$

$$\frac{\partial\ln(L(\mu,\Sigma))}{\partial\mu_2} = \frac{\Sigma_{i=1}^m y_i - m\mu_2}{(1-\rho^2)\sigma_2^2} + 2\frac{\rho}{1-\rho^2}(\Sigma_{i=1}^m x_i - m\mu_1) = 0$$

联立解得：

$$\hat{\mu_1} = \frac{m(1-4\sigma_1^2\sigma_2^2\rho^2)\bar{x_m} + (n-m)(1-\rho^2)\bar{x_n}}{m(1-4\sigma_1^2\sigma_2^2\rho^2) + (n-m)(1-\rho^2)} \triangleq \frac{\alpha\bar{x_m} + \beta\bar{x_n}}{\alpha+\beta} \quad (1)$$

$$\hat{\mu_2} = \frac{2(n-m)\sigma_2^2(1-\rho^2)\rho(\bar{x_m} - \bar{x_n})}{m(1-4\sigma_1^2\sigma_2^2\rho^2) + (n-m)(1-\rho^2)} + \bar{y} \quad (2)$$

又：

$$\frac{\partial\ln(L(\mu,\Sigma))}{\partial\rho} = \frac{m\rho(1-\rho^2)\sigma_1^2 - \rho\sigma_2^2\Sigma_{i=1}^m(x_i-\mu_1)^2 - (1+\rho^2)\sigma_1^2\sigma_2^2\Sigma_{i=1}^m(x_i-\mu_1)(y_i-\mu_2) - \rho\sigma_1^2\Sigma_{i=1}^m(y_i-\mu_2)^2}{(1-\rho^2)^2\sigma_1^2\sigma_2^2}$$

令对数似然函数对 σ_1, σ_2 的偏导为 0 可以求出所有参数的极大似然估计 □

Exercise 2

7. 假设从该二维正态总体 $N_2(\mu,\Sigma)$ 中随机产生 n 个模拟样本，其中

$$\mu=\begin{pmatrix}1\\2\end{pmatrix},\quad \Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix},$$

且 $\sigma_1=1,\ \sigma_2=2,\ \rho=0.6$。针对不同的样本量 $n=50,100,200$，重复模拟 1000 次。
(1) 试计算参数 $\mu_1,\mu_2,\sigma_1,\sigma_2$ 和 ρ 估计的平均值、偏差和标准差，并通过 QQ 图和直方图展示估计的好坏。进一步，随着样本量的变化，说明结果有什么变化；
(2) 基于式 (5.41) 和式 (5.42) 编写程序，分别计算相关系数 ρ 的 95% 的置信区间和区间长度，并进行比较哪个置信区间最优。进一步，随着样本量的变化，平均区间长度有什么变化。

证明. 采用极大似然估计

$$\hat{\mu_1}=\Sigma_{i=1}^n x_{i1} \tag{3}$$

$$\hat{\mu_2}=\Sigma_{i=1}^n x_{i2} \tag{4}$$

$$\hat{\sigma_1}=\sqrt{\frac{\Sigma_{i=1}^n(x_{i1}-\hat{\mu_1})^2}{n}} \tag{5}$$

$$\hat{\sigma_2}=\sqrt{\frac{\Sigma_{i=1}^n(x_{i2}-\hat{\mu_2})^2}{n}} \tag{6}$$

$$\hat{\rho}=\frac{\Sigma_{i=1}^n(x_{i1}-\mu_1)(x_{i2}-\mu_2)}{\sqrt{\Sigma_{i=1}^n(x_{i1}-\hat{\mu_1})^2\Sigma_{i=1}^n(x_{i2}-\hat{\mu_2})^2}} \tag{7}$$

$$\left[r(n)-\frac{1-r^2(n)}{\sqrt{n}}z_{1-\alpha/2},\ r(n)+\frac{1-r^2(n)}{\sqrt{n}}z_{1-\alpha/2}\right] \tag{8}$$

$$\left[\frac{\frac{1+r(n)}{1-r(n)}\exp\left(-\frac{2}{\sqrt{n}}z_{1-\alpha/2}\right)-1}{\frac{1+r(n)}{1-r(n)}\exp\left(-\frac{2}{\sqrt{n}}z_{1-\alpha/2}\right)+1},\ \frac{\frac{1+r(n)}{1-r(n)}\exp\left(\frac{2}{\sqrt{n}}z_{1-\alpha/2}\right)-1}{\frac{1+r(n)}{1-r(n)}\exp\left(\frac{2}{\sqrt{n}}z_{1-\alpha/2}\right)+1}\right]. \tag{9}$$

其中 $z_{1-\alpha/2}$ 为正态分布的上 $\frac{\alpha}{2}$ 分位数 □

Exercise 3

9. 假设 $x_1,\ldots,x_n$ 为来自 0-1 分布的独立同分布的简单随机样本，其分布律为 $\Pr(x_1=1)=p$, $\Pr(x_1=0)=1-p$，其中 $0<p<1$。根据中心极限定理，有

$$\sqrt{n}(\bar{x}-p)\xrightarrow{d}N(0,p(1-p)),$$

其中 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$。

(1) 试用 Fisher Z 变换方法构造 p 的置信水平为 $1-\alpha$ 的置信区间；

(2) 取 $p=0.6$，从 0-1 分布中随机产生样本量 $n=50,100,200$ 的随机样本，重复 1000 次试验，编写程序，计算 p 的 95% 的平均置信区间和区间长度，并观察随着样本量的变化，平均区间长度有什么变化。

证明. 函数 f 满足 $\sqrt{n}(f(\bar{x})-f(p))\xrightarrow{d}N(0,(f'(p))^2p(1-p))$ ，其中 $(f'(p))^2p(1-p)=1$ 于是：$(f'(p))^2=\frac{1}{p(1-p)},p\in(0,1)$ ，令 $p=sin^2\theta,\theta\in(0,\frac{\pi}{2})$

$$\int\frac{1}{\sqrt{(p(1-p)}}\,dp=\int\frac{2sin\theta\cos\theta}{sin\theta cos\theta}\,d\theta=2\theta=2\arcsin\left(\sqrt{p}\right)$$

于是可取 $f(x)=2\arcsin\left(\sqrt{x}\right)$ ，于是 $\sqrt{n}(f(\bar{x})-f(p))\in[-z_{1-\alpha/2},z_{1-\alpha/2}]$ 。由此可推知置信区间为：

$$[sin^2(\frac{1}{2}(2\arcsin\left(\sqrt{\bar{x}}\right)-\frac{z_{1-\alpha/2}}{\sqrt{n}}),sin^2(\frac{1}{2}(2\arcsin\left(\sqrt{\bar{x}}\right)+\frac{z_{1-\alpha/2}}{\sqrt{n}})]$$

□

Exercise 4

11. 设 X 和 Y 是相互独立的随机向量，且 $X\sim N_p(\mu_1,\Sigma)$, $Y\sim N_p(\mu_2,\Sigma)$，其中 $\Sigma>0$。进一步，假设 $x_1,\ldots,x_n$ 为来自总体 X 的独立同分布的随机样本，$y_1,\ldots,y_m$ 为来自总体 Y 的独立同分布的随机样本，$n,m>p$。

(1) 试证明参数 (μ_1,μ_2,Σ) 的充分完备统计量为 $(\bar{x},\bar{y},V_1+V_2)$，其中

$$V_1=\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})',$$

$$V_2=\sum_{i=1}^{m}(y_i-\bar{y})(y_i-\bar{y})';$$

(2) 试求参数 (μ_1,μ_2,Σ) 的极大似然估计，它们是无偏估计吗?

(3) 试求参数 (μ_1,μ_2,Σ) 的已知最小协方差矩阵无偏估计，它们是不是唯一存在的?

(4) $\Delta^2=(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)$ 通常用来表示两个正态分布 $N_p(\mu_1,\Sigma)$ 和 $N_p(\mu_2,\Sigma)$ 之间的距离。试求 Δ^2 的极大似然估计，并且是无偏估计吗? 若不是，请给出 Δ^2 的无偏估计。

证明. (1)

充分性:

$$\begin{aligned}
&f(x_1,\cdots,x_n,y_1,\cdots,y_m)\\
=&\prod_{i=1}^{n}\frac{exp\{-\frac{1}{2}(x_i-\mu_1)'\Sigma^{-1}(x_i-\mu_1)\}}{(2\pi)^{1/2}|\Sigma|^{1/2}}\prod_{i=1}^{m}\frac{exp\{-\frac{1}{2}(y_i-\mu_2)'\Sigma^{-1}(y_i-\mu_2)\}}{(2\pi)^{1/2}|\Sigma|^{1/2}}\\
=&\frac{exp\{-\frac{1}{2}(\Sigma_{i=1}^{n}(x_i-\mu_1)'\Sigma^{-1}(x_i-\mu_1)+\Sigma_{i=1}^{m}(y_i-\mu_2)'\Sigma^{-1}(y_i-\mu_2))\}}{(2\pi)^{(m+n)/2}|\Sigma|^{(m+n)/2}}\\
=&\frac{exp\{-\frac{1}{2}(tr(\Sigma_{i=1}^{n}(x_i-\mu_1)'\Sigma^{-1}(x_i-\mu_1))+tr(\Sigma_{i=1}^{m}(y_i-\mu_2)'\Sigma^{-1}(y_i-\mu_2)))\}}{(2\pi)^{(m+n)/2}|\Sigma|^{(m+n)/2}}\\
=&\frac{exp\{-\frac{1}{2}(tr(\Sigma^{-1}(V_1+V_2))+n(\bar{x}-\mu_1)'\Sigma^{-1}(\bar{x}-\mu_1)+m(\bar{y}-\mu_2)'\Sigma^{-1}(\bar{y}-\mu_2))\}}{(2\pi)^{(m+n)/2}|\Sigma|^{(m+n)/2}}
\end{aligned}$$

由因式分解定理知充分性成立，由指数分布族的性质知其完备性。

(2) 似然函数:

$$L=\frac{exp\{-\frac{1}{2}(tr(\Sigma^{-1}(V_1+V_2))+n(\bar{x}-\mu_1)'\Sigma^{-1}(\bar{x}-\mu_1)+m(\bar{y}-\mu_2)'\Sigma^{-1}(\bar{y}-\mu_2))\}}{(2\pi)^{(m+n)/2}|\Sigma|^{(m+n)/2}}$$

对数似然函数:

$$\begin{aligned}
\ln L=&-\ln((2\pi)^{(m+n)/2}|\Sigma|^{(m+n)/2})\\
&-\frac{1}{2}(tr(\Sigma^{-1}(V_1+V_2))+n(\bar{x}-\mu_1)'\Sigma^{-1}(\bar{x}-\mu_1)+m(\bar{y}-\mu_2)'\Sigma^{-1}(\bar{y}-\mu_2))
\end{aligned}$$

对于任意给定的 μ_2,Σ，由 Σ^{-1} 的正定性知，$\mu_1=\bar{x}$ 时 $n(\bar{x}-\mu_1)'\Sigma^{-1}(\bar{x}-\mu_1)$ 取得最小值，于是 $\hat{\mu_1}=\bar{x}$ 。同理 $\hat{\mu_2}=\bar{y}$ 。代入对数似然函数有:

$$\begin{aligned}\ln L =& -\ln((2\pi)^{(m+n)/2}|\Sigma|^{(m+n)/2}) - \frac{1}{2}(tr(\Sigma^{-1}(V_1+V_2)))\\ =& -\ln((2\pi)^{(m+n)/2}|\Sigma|^{(m+n)/2}) - \frac{1}{2}(tr(\Sigma^{-1/2}(V_1+V_2)\Sigma^{-1/2}))\end{aligned}$$

由于 $\Sigma^{-1/2}(V_1+V_2)\Sigma^{-1/2}$ 的对称性，故存在正交矩阵 U ，对角矩阵 $\Lambda = diag(\lambda_1 \cdots \lambda_p)$ 使得 $\Sigma^{-1/2}(V_1+V_2)\Sigma^{-1/2} = U\Lambda U'$ ，代入对数似然函数有：

$$\begin{aligned}\ln L =& -\frac{m+n}{2}\ln(2\pi) - \frac{m+n}{2}\ln(\frac{|V_1+V_2|}{\prod\lambda}) - \frac{1}{2}(\Sigma_{i=1}^{p}\lambda_i)\\ =& -\frac{m+n}{2}\ln(2\pi) - \frac{m+n}{2}\ln(|V_1+V_2|) + \frac{m+n}{2}\Sigma_{i=1}^{p}\ln\lambda_i - \frac{1}{2}(\Sigma_{i=1}^{p}\lambda_i)\\ =& -\frac{m+n}{2}\ln(2\pi) - \frac{m+n}{2}\ln(|V_1+V_2|) + \frac{1}{2}\Sigma_{i=1}^{p}[(m+n)\ln\lambda_i - \lambda_i]\end{aligned}$$

由于 $(m+n)\ln\lambda - \lambda$ 在 $\lambda = m+n$ 时取到最大值，此时 $\Lambda = (m+n)I_p, \Sigma = \frac{V_1+V_2}{m+n}$ 。即 $\hat{\Sigma} = \frac{V_1+V_2}{m+n}$.

$E(\bar{x}) = \mu_1$ 于是 $\bar{x}$ 是无偏估计

$E(\bar{y}) = \mu_2$ 于是 $\bar{y}$ 是无偏估计

$E(\frac{V_1+V_2}{m+n}) = \frac{m+n-2}{m+n}\Sigma$ 于是 $\frac{V_1+V_2}{m+n}$ 不是无偏估计

(3) 由于 UMVUE 是充分完备统计量的函数，可得以下结论：

$\bar{x}, \bar{y}$ 是 μ_1, μ_2 的无偏估计，因此也是其 UMVUE

$E(\frac{V_1+V_2}{m+n}) = \frac{m+n-2}{m+n}\Sigma$ 于是 Σ 的 UMVUE 为 $\frac{V_1+V_2}{m+n-2}$

由 UMVUE 的定义知其存在即唯一。

(4) 由于函数变换保持极大似然估计，于是

$$\hat{\Delta^2} = (\bar{x}-\bar{y})'(\frac{V_1+V_2}{m+n})^{-1}(\bar{x}-\bar{y}) = (m+n)(\bar{x}-\bar{y})'(V_1+V_2)^{-1}(\bar{x}-\bar{y})$$

由于 $\bar{x}-\bar{y} \sim N_p(\mu_1-\mu_2, \frac{\Sigma}{n}+\frac{\Sigma}{m}), V_1+V_2 \sim W_p(m+n-2, \Sigma)$ ，且两者独立。于是：

$$\begin{aligned}E(\hat{\Delta^2}) =&(m+n)E(\bar{x}-\bar{y})'E((V_1+V_2)^{-1})E(\bar{x}-\bar{y})\\ =&(m+n)(\mu_1-\mu_2)'E((V_1+V_2)^{-1})(\mu_1-\mu_2)\\ =&(m+n)(\mu_1-\mu_2)'(\frac{\Sigma^{-1}}{m+n-p-3})(\mu_1-\mu_2)\\ =&\frac{m+n}{m+n-p-3}(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)\end{aligned}$$

因此不是无偏估计，其无偏估计为 $(m+n-p-3)(\bar{x}-\bar{y})'(V_1+V_2)^{-1}(\bar{x}-\bar{y})$ □