多元统计分析 (2)

钟瑜　222018314210044

2020 年 10 月 30 日

2-6　设 $X \sim N_3(\mu, \Sigma)$，其中

$$X = (X_1, X_2, X_3)', \quad \mu = (2, -3, 1)',$$

$$\Sigma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}.$$

(1) 试求 $3X_1 - 2X_2 + X_3$ 的分布；

(2) 求二维向量 $a = (a_1, a_2)'$，使 X_3 与 $X_3 - a'\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ 相互独立.

解. *(1)* 由于 $3X_1 - 2X_2 + X_3 = (3, -2, 1)X = a'X$，故 $3X_1 - 2X_2 + X_3$ 服从 $N(a'\mu, a'\Sigma a) = N(25, 9)$，其中

$$a'\mu = (3, -2, 1)\mu = 13 \tag{1}$$

$$a'\Sigma a = (3, -2, 1)\begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{pmatrix}(3, -2, 1)' = 9$$

(2) 由于 $X_3 \sim N(1, 2)$，而

$$X_3 - a'(X_1, X_2)' = X_3 - a_1X_1 - a_2X_2 = (-a_1, -a_2, 1)X = b'X \sim N(b'\mu, b'\Sigma b)$$

要使得 X_3 与 $b'X$ 独立，则 $cov(X_3, b'X) = 0$

$$\begin{aligned} cov(X_3, b'X) &= \mathbb{E}[(X_3 - \mu_3)(b'X - b'\mu)] \\ &= \mathbb{E}[X_3 b'X] - 2b'\mu \\ &= \mathbb{E}[X_3(-a_1, -a_2, 1)X] - 2b'\mu \\ &= 0 \end{aligned} \tag{2}$$

即

$$\mathbb{E}[-a_1X_3X_1 - a_2X_2X_3 + X_3X_3] = -6a_1 + 9a_2 + 3$$

而

$$\mathbb{E}[X_3X_1] = cov(X_3, X_1) + \mathbb{E}X_1\mathbb{E}X_3 = 3$$

$$\mathbb{E}[X_3X_2] = cov(X_3, X_2) + \mathbb{E}X_2\mathbb{E}X_3 = -1$$

$$\mathbb{E}[X_3X_3] = var(X_3) + \mathbb{E}X_3^2 = 1$$

代入得 $a_1 + a_2 - 2 = 0$.

2-11 已知 $X=(X_1,X_2)'$ 的密度函数为

$$f(x_1,x_2)=\frac{1}{2\pi}\exp\left\{-\frac{1}{2}(2x_1^2+x_2^2+2x_1x_2-22x_1-14x_2+65)\right\},$$

试求 X 的均值向量和协方差阵.

解. 首先求出 X_1 和 X_2 的边缘分布:

$$\begin{aligned}
f_{X_1}(x)&=\int_{-\infty}^{\infty}f(x_1,x_2)dx_2\\
&=\frac{1}{2\pi}\int_{-\infty}^{\infty}exp\left\{-\frac{1}{2}(2x_1^2+x_2^2+2x_1x_2-22x_1-14x_2+65)\right\}dx_2\\
&=\frac{1}{2\pi}exp\left\{-\frac{1}{2}(2x_1^2-22x_1+65)\right\}\int_{-\infty}^{\infty}exp\left\{-\frac{1}{2}(x_2^2+2x_1x_2-14x_2)\right\}dx_2\\
&=\frac{1}{2\pi}exp\left\{-\frac{1}{2}(2x_1^2-22x_1+65)\right\}\int_{-\infty}^{\infty}exp\left\{-\frac{1}{2}(x_2^2+2x_2(x_1-7)+(x_1-7)^2)\right\}dexp\left\{-\frac{1}{2}(x_1-7)^2\right\}x_2\\
&=\frac{1}{2\pi}exp\left\{-\frac{1}{2}(2x_1^2-22x_1+65)\right\}exp\left\{-\frac{1}{2}(x_1-7)^2\right\}\int_{-\infty}^{\infty}exp\left\{-\frac{1}{2}(x_2^2+2x_2(x_1-7)+(x_1-7)^2)\right\}dx_2\\
&=\frac{1}{\sqrt{2\pi}}exp\left\{-\frac{1}{2}(x_1^2-8x_1+16)\right\}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}exp\left\{-\frac{1}{2}(x_2-x_1+7)^2\right\}dx_2\\
&=\frac{1}{\sqrt{2\pi}}exp\left\{-\frac{1}{2}(x_1-4)^2\right\}
\end{aligned}\tag{3}$$

因此 $X_1\sim N(4,1)$ 同理可得 $X_2\sim N(3,2)$ 故均值向量 $\mu=(4,3)'$.

然后计算 X_1 和 X_2 的协方差.

$$\begin{aligned}
cov(X_1,X_2)&=\mathbb{E}[(X_1-4)(X_2-3)]\\
&=\iint_{-\infty}^{\infty}(x_1-4)(x_2-3)f(x_1,x_2)dx_1dx_2\\
&=\iint_{-\infty}^{\infty}a_1a_2\frac{1}{2\pi}exp\left\{-\frac{1}{2}(2a_1^2+a_2^2+2a_1a_2)\right\}da_1da_2(\text{其中}a_1=x_1-4,a_2=x_2-3)\\
&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}a_1exp\left\{-\frac{a_1^2}{2}\right\}\left[\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}a_2exp\left\{-\frac{1}{2}(a_1+a_2)^2\right\}da_2\right]da_1\\
&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}a_1exp\left\{-\frac{a_1^2}{2}\right\}\frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{\infty}(a_1+a_2-a_1)exp\left\{-\frac{1}{2}(a_1+a_2)^2\right\}da_2\right]da_1\\
&=-1
\end{aligned}\tag{4}$$

因此协方差矩阵为:

$$\Sigma=\begin{pmatrix}1&-1\\-1&2\end{pmatrix}$$

2-18 设 $X_{(1)},\cdots,X_{(n)}$是来自 $N_p(\mu,\Sigma)$的随机样本，$c_i\geqslant 0$ $(i=$

1,⋯,n)，$\sum_{i=1}^{n} c_i = 1$，令 $Z = \sum_{i=1}^{n} c_i X_{(i)}$. 试证明：

(1) Z 是 μ 的无偏估计量；

(2) $Z \sim N_p(\mu, c'c\Sigma)$，其中 $c = (c_1, \cdots, c_n)'$；

(3) 当 $c = \frac{1}{n}\mathbf{1}_n$ 时，Z 的协方差阵在非负定的意义下达到极小.

解. *(1)* $\mathbb{E}Z = \mathbb{E}(\sum_{i=1}^{n} c_i X_{(i)}) = \sum_{i=1}^{n} c_i \mathbb{E}X_{(i)} = \sum_{i=1}^{n} c_i \mu = \mu$

(2) 因为 $X_{(i)} \sim N_p(\mu, \Sigma)$, 故对每个 $i, X_{(i)}$ 的特征函数为:

$$\phi(t) = \mathbb{E}(e^{itX_{(i)}}) = exp\left[it'\mu - \frac{1}{2}t'\Sigma t\right] \tag{5}$$

由于 $Z{=}\sum_{i=1}^{n} c_i X(i)$, 故 Z 的特征函数为:

$$\begin{aligned}
\phi_Z(t) &= \mathbb{E}(e^{itZ}) \\
&= \mathbb{E}(e^{it\sum_{i=1}^{n} c_i X_{(i)}}) \\
&= \prod_{i=1}^{n} \mathbb{E}(e^{ic_i t X_{(i)}}) \\
&= \prod_{i=1}^{n} exp\left[ic_i t'\mu - \frac{1}{2}(c_i t)'\Sigma(c_i t)\right] \\
&= \prod_{i=1}^{n} exp\left[ic_i t'\mu - \frac{1}{2}c_i^2 t'\Sigma t\right] \\
&= exp\left[i\sum_{i=1}^{n} c_i t'\mu - \frac{1}{2}\sum_{i=1}^{n} c_i^2 t'\Sigma t\right] \\
&= exp\left[it'\mu - \frac{1}{2}t'(c_i' c_i \Sigma)t\right]
\end{aligned} \tag{6}$$

因此 $Z \sim N_p(\mu, c_i' c_i \Sigma)$.

(3) 由均值不等式,

$$c'c = \sum_{i=1}^{n} c_i^2 \quad \geq (\sum_{i=1}^{n} c_i)^2/n \tag{7}$$

当且仅当 $c_1 = ... = c_n$ 时, 即 $c = \frac{1}{n}\mathbf{1}_n$,$Z$ 的协方差阵在非负定意义下达到极小, 为 $c_i' c_i \Sigma = \frac{1}{n}\Sigma$

2-19　为了了解某种橡胶的性能，今抽取 10 个样品，每个测量三项指标：硬度、变形和弹性，其数据如下表：

序号	硬度(X_1)	变形(X_2)	弹性(X_3)
1	65	45	27.6
2	70	45	30.7
3	70	48	31.8
4	69	46	32.6
5	66	50	31.0
6	67	46	31.3
7	68	47	37.0
8	72	43	33.6
9	66	47	33.1
10	68	48	34.2

试计算样本均值、样本离差阵、样本协方差阵和样本相关阵.

解. R 运算结果如下:

```
> rubber<-read.csv(”E:/4.多元统计分析/zuoye/2/1.csv”)
> rubber
   num X1 X2   X3
1    1 65 45 27.6
2    2 70 45 30.7
3    3 70 48 31.8
4    4 69 46 32.6
5    5 66 50 31.0
6    6 67 46 31.3
7    7 68 47 37.0
8    8 72 43 33.6
9    9 66 47 33.1
10  10 68 48 34.2

> mean(rubber$X1)    #样本均值
[1] 68.1
> mean(rubber$X2)
[1] 46.5
> mean(rubber$X3)
[1] 32.29

> cov(rubber[-1])    #样本协方差阵
          X1         X2        X3
X1  4.766667 -1.9444444 1.9344444
X2 -1.944444  3.8333333 0.6166667
X3  1.934444  0.6166667 6.1898889

```

```
> 9*cov(rubber[-1])     #样本离差阵
        X1     X2     X3
X1   42.90 -17.50 17.410
X2  -17.50  34.50  5.550
X3   17.41   5.55 55.709

> cor(rubber[-1])    #样本相关阵
            X1         X2        X3
X1   1.0000000 -0.4548832 0.3561291
X2  -0.4548832  1.0000000 0.1265962
X3   0.3561291  0.1265962 1.0000000
```