第3章
多元正态分布

李高荣

北京师范大学统计学院

E-mail: *ligaorong@bnu.edu.cn*

1. 随机向量
   - 随机向量及其分布表示
   - 随机向量的数字特征
   - 条件分布
   - 变量变换
2. 多元正态分布
   - 多元正态分布的定义及性质
   - 条件分布和独立性
3. 偏相关系数
4. 拉直运算与Kronecker积
5. 矩阵多元正态分布
6. 多元正态分布样本统计量
7. 特征函数*

- 扫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

在多元统计分析中，多元正态分布占有相当重要的地位，主要原因：

- 许多实际问题涉及的随机向量服从正态分布或近似服从正态分布；
- 当样本量很大时，许多统计量的极限分布往往和正态分布有关；
- 对多元正态分布，理论与实践都比较成熟，已有一整套行之有效的统计推断方法。

定义3.1：随机向量

将p个随机变量$X_1, X_2, \cdots, X_p$的整体称为p维随机向量，记为$\boldsymbol{X} = (X_1, X_2, \cdots, X_p)'$。

定义3.2：联合分布函数

设$\boldsymbol{X} = (X_1, X_2, \cdots, X_p)'$为$p$维随机向量，对任意的$p$维实数向量$\boldsymbol{x} = (x_1, x_2, \cdots, x_p)' \in \mathbb{R}^p$，称$p$元函数

$$
\begin{aligned}
F(\boldsymbol{x}) &= F(x_1, x_2, \cdots, x_p) \\
&= \Pr\{X_1 \leq x_1, X_2 \leq x_2, \cdots, X_p \leq x_p\}
\end{aligned}
$$

为$\boldsymbol{X}$的联合分布函数，并记成$\boldsymbol{X} \sim F(\boldsymbol{x})$，其中$\mathbb{R}^p$表示$p$维欧氏空间。

- 多维随机向量的统计特性可以通过联合分布函数来完整刻画和描述。

定义3.3：联合概率分布函数

设$\boldsymbol{X} = (X_1, X_2, \cdots, X_p)'$为$p$维随机向量，若存在有限个或可列个$p$维向量$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots$，记$\Pr(\boldsymbol{X} = \boldsymbol{x}_k) = p_k \ (k = 1, 2, \cdots)$且满足$p_1 + p_2 + \cdots = 1$，则称$\boldsymbol{X}$为离散型随机向量，称$\Pr(\boldsymbol{X} = \boldsymbol{x}_k) = p_k (k = 1, 2, \cdots)$为离散型随机向量$\boldsymbol{X}$的联合概率分布函数。

当随机向量$\boldsymbol{X}$中每一个分量都是连续型随机变量时，我们把$\boldsymbol{X}$称为连续型随机向量，它可以通过联合概率密度函数来描述。

定义3.4：联合概率密度函数

设$\boldsymbol{X} = (X_1, \cdots, X_p)'$为$p$维随机向量，若存在非负函数$f(x_1, \cdots, x_p)$，使得随机向量$\boldsymbol{X}$的联合分布函数对一切$(x_1, \cdots, x_p)' \in \mathbb{R}^p$均可表示为

$$F(x_1, \cdots, x_p) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f(u_1, \cdots, u_p) \mathrm{d}u_1 \cdots \mathrm{d}u_p,$$

则称$\boldsymbol{X}$为连续型随机向量，称$f(x_1, \cdots, x_p)$为连续型随机向量$\boldsymbol{X}$的联合概率密度函数，或称为多元密度函数或密度函数。

多元密度函数$f(x_1,x_2,\cdots,x_p)$具有以下四条性质：

1. $f(x_1,x_2,\cdots,x_p)\geq 0$，对一切实数$x_1,x_2,\cdots,x_p$；
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,x_2,\cdots,x_p)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_p=1$；
3. 若$f(x_1,x_2,\cdots,x_p)$在$(x_1,x_2,\cdots,x_p)$处连续，则有

$$\frac{\partial^p F(x_1,x_2,\cdots,x_p)}{\partial x_1\partial x_2\cdots\partial x_p}=f(x_1,x_2,\cdots,x_p);$$

4. 设D是$\mathbb{R}^p$空间的任意一个空间，则随机向量$\boldsymbol{X}$落在D内的概率为

$$\Pr(\boldsymbol{X}\in D)=\int\cdots\int_D f(x_1,x_2,\cdots,x_p)\mathrm{d}x_1\cdots\mathrm{d}x_p.$$

- $\boldsymbol{X} = (X_1, \cdots, X_p)'$联合分布函数为$F(x_1, \cdots, x_p)$，其任一个分量$\boldsymbol{X}^{(m)} = (X_1, \cdots, X_m)'$也是一个随机向量(其中$1 \le m < p$)，也有自己的联合分布函数，记为$F_{\boldsymbol{X}^{(m)}}(x_1, \cdots, x_m)$，称为$\boldsymbol{X}^{(m)}$的边缘分布函数，记为

$$\begin{aligned} &\Pr\{X_1 \le x_1, \cdots, X_m \le x_m\} \\ =&\Pr\{X_1 \le x_1, \cdots, X_m \le x_m, X_{m+1} \le \infty, \cdots, X_p \le \infty\} \\ =&F(x_1, \cdots, x_m, \infty, \cdots, \infty). \end{aligned}$$

- 由$\boldsymbol{X}$的联合概率密度函数$f(x_1, \cdots, x_p)$，得到$\boldsymbol{X}^{(m)}$的边缘概率密度函数为

$$f_{\boldsymbol{X}^{(m)}}(x_1, \cdots, x_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_p) \mathrm{d}x_{m+1} \cdots \mathrm{d}x_p.$$

定义3.5：相互独立

设$\boldsymbol{X}$和$\boldsymbol{Y}$是两个随机向量，对一切的$\boldsymbol{x}$和$\boldsymbol{y}$，若

$$\Pr(\boldsymbol{X} \le \boldsymbol{x}, \boldsymbol{Y} \le \boldsymbol{y}) = \Pr(\boldsymbol{X} \le \boldsymbol{x}) \Pr(\boldsymbol{Y} \le \boldsymbol{y})$$

成立，则称$\boldsymbol{X}$和$\boldsymbol{Y}$相互独立。若$F(\boldsymbol{x}, \boldsymbol{y})$为$(\boldsymbol{X}, \boldsymbol{Y})$的联合分布函数，$F_{\boldsymbol{X}}(\boldsymbol{x})$和$F_{\boldsymbol{Y}}(\boldsymbol{y})$分别是$\boldsymbol{X}$和$\boldsymbol{Y}$的边缘分布函数，则$\boldsymbol{X}$和$\boldsymbol{Y}$相互独立当且仅当

$$F(\boldsymbol{x}, \boldsymbol{y}) = F_{\boldsymbol{X}}(\boldsymbol{x}) F_{\boldsymbol{Y}}(\boldsymbol{y}).$$

若$f(\boldsymbol{x}, \boldsymbol{y})$为$(\boldsymbol{X}, \boldsymbol{Y})$的联合概率密度函数，$f_{\boldsymbol{X}}(\boldsymbol{x})$和$f_{\boldsymbol{Y}}(\boldsymbol{y})$分别是$\boldsymbol{X}$和$\boldsymbol{Y}$的边缘概率密度函数，则$\boldsymbol{X}$和$\boldsymbol{Y}$相互独立当且仅当

$$f(\boldsymbol{x}, \boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{Y}}(\boldsymbol{y}).$$

思考：如何定义$X_1, \cdots, X_p$的相互独立性？

均值向量

设$\boldsymbol{X}=(X_1,\cdots,X_p)'$是$p$维随机向量，若$E(X_i)=\mu_i(i=1,\cdots,p)$存在，定义随机向量$\boldsymbol{X}$的均值为：

$$E(\boldsymbol{X})=\begin{bmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{bmatrix}=\begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}=\boldsymbol{\mu},$$

其中$\boldsymbol{\mu}=(\mu_1,\cdots,\mu_p)'$是一个$p$维向量，称为均值向量。

定义3.6：协方差矩阵

设$\boldsymbol{X} = (X_1, \cdots, X_p)'$是$p$维的随机向量，若$X_i$和$X_j$的协方差$\mathrm{Cov}(X_i, X_j)$存在$(i, j = 1, \cdots, p)$，则称

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{X}) &= E[(\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{X} - E(\boldsymbol{X}))'] \\ &= \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \cdots & \mathrm{Cov}(X_1, X_p) \\ \mathrm{Cov}(X_2, X_1) & \cdots & \mathrm{Cov}(X_2, X_p) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, X_1) & \cdots & \mathrm{Cov}(X_p, X_p) \end{pmatrix}\end{aligned}$$

为随机向量$\boldsymbol{X}$的方差矩阵或协方差矩阵，记为$\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X}) = (\sigma_{ij})_{p \times p}$，其中$\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$。

定义3.6(续): 协方差矩阵

设$\boldsymbol{X}=(X_1,\cdots,X_p)'$和$\boldsymbol{Y}=(Y_1,\cdots,Y_q)'$分别是$p$维和$q$维的两个随机向量，若$X_i$和$Y_j$的协方差$\mathrm{Cov}(X_i,Y_j)$存在$(i=1,\cdots,p,j=1,\cdots,q)$，则称

$$
\begin{aligned}
\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y}) &= E[(\boldsymbol{X}-E(\boldsymbol{X}))(\boldsymbol{Y}-E(\boldsymbol{Y}))'] \\
&= \begin{pmatrix} \mathrm{Cov}(X_1,Y_1) & \cdots & \mathrm{Cov}(X_1,Y_q) \\ \mathrm{Cov}(X_2,Y_1) & \cdots & \mathrm{Cov}(X_2,Y_q) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p,Y_1) & \cdots & \mathrm{Cov}(X_p,Y_q) \end{pmatrix}
\end{aligned}
$$

为随机向量$\boldsymbol{X}$和$\boldsymbol{Y}$的**协方差矩阵**。

协方差矩阵

- 当$\boldsymbol{X} = \boldsymbol{Y}$和$p = q$时，$\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{Cov}(\boldsymbol{X})$
- 若$\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathbf{0}_{p\times q}$ (其中$\mathbf{0}_{p\times q}$表示$p \times q$的零矩阵)，则称随机向量$\boldsymbol{X}$和$\boldsymbol{Y}$不相关
- 当$\boldsymbol{X}$和$\boldsymbol{Y}$独立时，则$\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathbf{0}_{p\times q}$，反之则不成立

定义3.7：相关阵或相关系数矩阵

若X_i和X_j的协方差$\mathrm{Cov}(X_i, X_j)$存在$(i,j = 1,\cdots,p)$，称$\mathbf{R} = (\rho_{ij})_{p\times p}$为$X$的**相关阵**，其中

$$\rho_{ij} = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{Var}(X_i)}\sqrt{\mathrm{Var}(X_j)}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}, \quad i,j = 1,\cdots,p,$$

其中$\sigma_{ii} = \mathrm{Var}(X_i)$为随机变量$X_i$的**方差**，而$\sqrt{\sigma_{ii}}$为随机变量$X_i$的**标准差**$(i = 1,\cdots,p)$。

若记$\mathbf{D}^{1/2} = \mathrm{diag}(\sqrt{\sigma_{11}},\cdots,\sqrt{\sigma_{pp}})$为标准差对角阵，则

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

- 均值向量和协方差矩阵具有如下的性质：

性质3.1.1

设X和Y是两个随机向量，假定$\mathbf{A}$和$\mathbf{B}$为大小适合运算的两个常数矩阵，则

$$
\begin{aligned}
E(\mathbf{A}X) &= \mathbf{A}E(X),\\
E(\mathbf{A}X\mathbf{B}) &= \mathbf{A}E(X)\mathbf{B},\\
E(\mathbf{A}X+\mathbf{B}Y) &= \mathbf{A}E(X)+\mathbf{B}E(Y),\\
\mathrm{Cov}(\mathbf{A}X) &= \mathbf{A}\mathrm{Cov}(X)\mathbf{A}',\\
\mathrm{Cov}(\mathbf{A}X,\mathbf{B}Y) &= \mathbf{A}\mathrm{Cov}(X,Y)\mathbf{B}'.
\end{aligned}
$$

证明： 这里只给出最后一个公式的证明。

$$
\begin{aligned}
\mathrm{Cov}(\mathbf{A}\boldsymbol{X}, \mathbf{B}\boldsymbol{Y}) &= E[(\mathbf{A}\boldsymbol{X} - E(\mathbf{A}\boldsymbol{X}))(\mathbf{B}\boldsymbol{Y} - E(\mathbf{B}\boldsymbol{Y}))'] \\
&= E[\mathbf{A}(\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{Y} - E(\boldsymbol{Y}))'\mathbf{B}'] \\
&= \mathbf{A}E[(\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{Y} - E(\boldsymbol{Y}))']\mathbf{B}' \\
&= \mathbf{A}\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y})\mathbf{B}'.
\end{aligned}
$$

性质3.1.2

若$\boldsymbol{X}$和$\boldsymbol{Y}$是两个相互独立的随机向量，则$\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y})=\mathbf{0}_{p\times q}$，其中$p$是$\boldsymbol{X}$的维数并且$q$是$\boldsymbol{Y}$的维数。

性质3.1.3

随机向量$\boldsymbol{X}=(X_1,\cdots,X_p)'$的协方差矩阵$\boldsymbol{\Sigma}=\mathrm{Cov}(\boldsymbol{X})$是对称非负定矩阵。

证明： 因为$\mathrm{Cov}(X_i,X_j)=\mathrm{Cov}(X_j,X_i)$，所以$\boldsymbol{\Sigma}=\boldsymbol{\Sigma}'$。对任给$\boldsymbol{a}=(a_1,\cdots,a_p)'$，有

$$
\begin{aligned}
\boldsymbol{a}'\Sigma\boldsymbol{a} &= E[\boldsymbol{a}'(\boldsymbol{X}-E(\boldsymbol{X}))(\boldsymbol{X}-E(\boldsymbol{X}))'\boldsymbol{a}] \\
&= E[(\boldsymbol{a}'(\boldsymbol{X}-E(\boldsymbol{X})))^2] \geq 0,
\end{aligned}
$$

所以$\Sigma \geq 0$，即证明了Σ为非负定矩阵。

性质3.1.4

$\Sigma = \mathbf{L}^2$，其中$\mathbf{L}$为非负定矩阵。

证明：由性质3.1.3可知，$\mathbf{\Sigma}$是一个对称非负定矩阵，利用实对称矩阵的对角化定理，存在正交矩阵$\mathbf{\Gamma}$，使得

$$\mathbf{\Sigma} = \mathbf{\Gamma} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} \mathbf{\Gamma}'$$

$$= \mathbf{\Gamma} \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_p} \end{pmatrix} \mathbf{\Gamma}' \cdot \mathbf{\Gamma} \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_p} \end{pmatrix} \mathbf{\Gamma}' = \mathbf{L}^2,$$

其中$\mathbf{L} = \mathbf{\Gamma}\mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_p})\mathbf{\Gamma}'$，$\lambda_i \geq 0 (i = 1, \cdots, p)$，且$\mathbf{L} = \mathbf{L}'$，所以$\mathbf{L} \geq 0$。

性质3.1.5

设$\boldsymbol{X}$为p维随机向量，期望和协方差矩阵存在，记$\boldsymbol{\mu}=E(\boldsymbol{X})$，$\boldsymbol{\Sigma}=\mathrm{Cov}(\boldsymbol{X})$，$\mathbf{A}$为$p\times p$常数矩阵，则

$$E(\boldsymbol{X}'\mathbf{A}\boldsymbol{X})=\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})+\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}.$$

证明作为练习。

性质3.1.5

设$\boldsymbol{X}$为p维随机向量，期望和协方差矩阵存在，记$\boldsymbol{\mu} = E(\boldsymbol{X})$，$\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X})$，$\mathbf{A}$为$p \times p$常数矩阵，则

$$E(\boldsymbol{X}'\mathbf{A}\boldsymbol{X}) = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}.$$

证明作为练习。

证明：一些简单计算有：

$$\begin{aligned} \boldsymbol{X}'\mathbf{A}\boldsymbol{X} &= (\boldsymbol{X} - E\boldsymbol{X})'\mathbf{A}(\boldsymbol{X} - E\boldsymbol{X}) + \boldsymbol{X}'\mathbf{A}E\boldsymbol{X} \\ &\quad + (E\boldsymbol{X})'\mathbf{A}\boldsymbol{X} - (E\boldsymbol{X})'\mathbf{A}E\boldsymbol{X}. \end{aligned}$$

两边取数学期望，并简单化简，即完成性质3.1.5的证明。

矩阵性质：

令$\mathbf{A}$是一个$p \times p$对称矩阵，$\boldsymbol{x}$是一个$p \times 1$的向量，则

1. $\boldsymbol{x}'\mathbf{A}\boldsymbol{x} = \mathrm{tr}(\boldsymbol{x}'\mathbf{A}\boldsymbol{x}) = \mathrm{tr}(\mathbf{A}\boldsymbol{x}\boldsymbol{x}')$
2. $\mathrm{tr}(\mathbf{A}) = \sum_{k=1}^{p} \lambda_k$，其中$\lambda_k$是矩阵$\mathbf{A}$的特征值。

矩阵性质：

令$\mathbf{A}$是一个$p \times p$对称矩阵，$\boldsymbol{x}$是一个$p \times 1$的向量，则

1. $\boldsymbol{x}'\mathbf{A}\boldsymbol{x} = \mathrm{tr}(\boldsymbol{x}'\mathbf{A}\boldsymbol{x}) = \mathrm{tr}(\mathbf{A}\boldsymbol{x}\boldsymbol{x}')$
2. $\mathrm{tr}(\mathbf{A}) = \sum_{k=1}^{p} \lambda_k$，其中$\lambda_k$是矩阵$\mathbf{A}$的特征值。

证明：对(1)，注意到：$\boldsymbol{x}'\mathbf{A}\boldsymbol{x}$是常数，则$\boldsymbol{x}'\mathbf{A}\boldsymbol{x} = \mathrm{tr}(\boldsymbol{x}'\mathbf{A}\boldsymbol{x})$。利用结论：对$m \times p$矩阵$\mathbf{B}$和$p \times m$矩阵$\mathbf{C}$，有$\mathrm{tr}(\mathbf{BC}) = \mathrm{tr}(\mathbf{CB})$。令$\mathbf{B} = \boldsymbol{x}'$且$m = 1$，$\mathbf{C} = \mathbf{A}\boldsymbol{x}$，则完成了(1)的证明。

对(2)，利用谱分解，即$\mathbf{A} = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma}$，其中$\mathbf{\Gamma}\mathbf{\Gamma}' = \mathbf{I}$和$\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \cdots, \lambda_p\}$，则

$$\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma}) = \mathrm{tr}(\mathbf{\Lambda}\mathbf{\Gamma}\mathbf{\Gamma}') = \mathrm{tr}(\mathbf{\Lambda}) = \sum_{k=1}^{p} \lambda_k.$$

- 假设$f(x_1,\cdots,x_p)$是随机向量$\boldsymbol{X}$的联合概率密度函数，则联合矩为

$$E(X_1^{h_1}\cdots X_p^{h_p})=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_1^{h_1}\cdots x_p^{h_p}f(x_1,\cdots,x_p)\mathrm{d}x_1\cdots\mathrm{d}x_p.$$

- 对于$m<p$，变量子集$X_1,\cdots,X_m$的联合矩为

$$E(X_1^{h_1}\cdots X_m^{h_m})=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_1^{h_1}\cdots x_m^{h_m}f_{\boldsymbol{X}^{(m)}}(x_1,\cdots,x_m)\mathrm{d}x_1\cdots\mathrm{d}x_m.$$

- 如果$X_1,\cdots,X_p$相互独立，则

$$E(X_1^{h_1}\cdots X_p^{h_p})=\prod_{i=1}^{p}E(X_i^{h_i}).$$

- 二维随机向量(X,Y)，X与Y的相互关系除了独立性以外，还有相依关系，即随机变量的取值往往彼此是有影响的，这种关系用条件分布能更好地表达出来。
- 设给定二维随机向量(X,Y)，假设事件$A=\{\omega: x_1\le X\le x_2\}, B=\{\omega: y_1\le Y\le y_2\}$, 则给定事件$B$的条件下，事件$A$发生的条件概率为：

$$\begin{aligned}\Pr(A|B) &= \Pr(x_1\le X\le x_2|y_1\le Y\le y_2)\\ &= \frac{\Pr(x_1\le X\le x_2, y_1\le Y\le y_2)}{\Pr(y_1\le Y\le y_2)}\end{aligned}$$

离散型条件概率

如果二维离散随机向量(X,Y)的联合分布列为: $p_{ij}=\Pr(X=x_i,Y=y_j), i=1,2,\cdots;j=1,2,\cdots$。对一切使得$\Pr(Y=y_j)=\sum_{i=1}^{\infty}p_{ij}=p_{\cdot j}>0$的$y_j$，则称

$$p_{i|j}=\Pr(X=x_i|Y=y_j)=\frac{\Pr(X=x_i,Y=y_j)}{\Pr(Y=y_j)}=\frac{p_{ij}}{p_{\cdot j}}$$

为在给定$Y=y_j$条件下X的条件分布列，其中$i=1,2\cdots$。

条件分布函数与条件密度函数

设(X,Y)为连续型随机向量，联合密度函数为$f(x,y)$，边际分布函数分别为$f_X(x)$和$f_Y(y)$。对一切$f_Y(y)>0$的y，在给定$Y=y$条件下，X的条件分布函数和条件密度函数分别为:

$$F(x|y)=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}du,\quad f(x|y)=\frac{f(x,y)}{f_Y(y)}.$$

对一切$f_X(x)>0$的x，在给定$X=x$条件下，Y的条件分布函数和条件密度函数分别为:

$$F(y|x)=\int_{-\infty}^{y}\frac{f(x,v)}{f_X(x)}dv,\quad f(y|x)=\frac{f(x,y)}{f_X(x)}.$$

一般而言，随机向量$\boldsymbol{X}=(X_1,\cdots,X_p)'$的联合分布函数为$F(x_1,\cdots,x_p)$, 联合密度函数为$f(x_1,\cdots,x_p)$，则

- 给定$X_{m+1}=x_{m+1},\cdots,X_p=x_p$条件下，$(X_1,\cdots,X_m)'$ 的条件分布函数为：

$$\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_m}\frac{f(u_1,\cdots,u_m,x_{m+1},\cdots,x_p)}{f_{\boldsymbol{X}^{(-m)}}(x_{m+1},\cdots,x_p)}\mathrm{d}u_1\cdots\mathrm{d}u_m.$$

- 给定$X_{m+1}=x_{m+1},\cdots,X_p=x_p$条件下，$(X_1,\cdots,X_m)'$ 的条件概率密度函数为：

$$\frac{f(x_1,\cdots,x_p)}{f_{\boldsymbol{X}^{(-m)}}(x_{m+1},\cdots,x_p)},$$

其中$f_{\boldsymbol{X}^{(-m)}}(x_{m+1},\cdots,x_p)$为$\boldsymbol{X}^{-m}=(X_{m+1},\cdots,X_p)'$的边缘密度函数。

令随机向量$(X_1,\cdots,X_p)'$的联合密度函数为$f(x_1,\cdots,x_p)$，定义下面的p实值函数：

$$y_i = y_i(x_1,\cdots,x_p), \qquad i=1,\cdots,p.$$

满足：

- 假设从x空间到y空间的变换是一一对应的；
- 存在逆变换：$x_i = x_i(y_1,\cdots,y_p), i=1,\cdots,p$;
- 定义Jacobian(雅可比)矩阵$J(y_1,\cdots,y_p)$如下：

$$J(y_1,\cdots,y_p) = \mathrm{mod}\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_p}{\partial y_1} & \frac{\partial x_p}{\partial y_2} & \cdots & \frac{\partial x_p}{\partial y_p} \end{vmatrix},$$

其中mod是指模数(modulus)或绝对值。

多元函数的密度变换公式：

令随机变量$Y_1,\cdots,Y_p$定义为：

$$Y_i=y_i(X_1,\cdots,X_p),\qquad i=1,\cdots,p.$$

则$Y_1,\cdots,Y_p$的概率密度函数为：

$$\begin{aligned} g(y_1,\cdots,y_p) &= f\Big[x_1(y_1,\cdots,y_p),\cdots,x_p(y_1,\cdots,y_p)\Big] \\ &\quad \times J(y_1,\cdots,y_p). \end{aligned}$$

应用：如果给定了$X_1,\cdots,X_p$的联合密度函数，以及Y和X的函数关系，则可用上式计算$Y_1,\cdots,Y_p$落在一个给定区域S中的概率。

- 设$Y_1, \cdots, Y_p$为独立同分布(i.i.d.)的随机变量序列
- $Y_i \sim N(0,1)$，$i = 1, \cdots, p$，则随机向量$\boldsymbol{Y} = (Y_1, \cdots, Y_p)'$的概率密度函数为：

$$
\begin{aligned}
f(y_1, \cdots, y_p) &= \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} y_i^2 \right\} \\
&= (2\pi)^{-p/2} \exp\left\{ -\frac{1}{2} \boldsymbol{y}'\boldsymbol{y} \right\},
\end{aligned}
$$

其中$\boldsymbol{y} = (y_1, \cdots, y_p)'$。

设$\boldsymbol{Y} \sim N_p(\mathbf{0}, \mathbf{I}_p)$，则有

- $E(\boldsymbol{Y}) = \mathbf{0}$，$\mathrm{Cov}(\boldsymbol{Y}) = \mathbf{I}_p$;
- $E(\boldsymbol{Y}'\mathbf{A}\boldsymbol{Y}) = \mathrm{tr}(\mathbf{A})$;
- 设$\boldsymbol{a}$为p元向量，$\mathbf{A}$为对称矩阵，则

$$\mathrm{Cov}(\boldsymbol{a}'\boldsymbol{Y}, \boldsymbol{Y}'\mathbf{A}\boldsymbol{Y}) = 0;$$

- 设$\mathbf{A}, \mathbf{B}$为对称矩阵，则

$$\mathrm{Cov}(\boldsymbol{Y}'\mathbf{A}\boldsymbol{Y}, \boldsymbol{Y}'\mathbf{B}\boldsymbol{Y}) = 2\mathrm{tr}(\mathbf{AB}).$$

证明： 由于$\mathbf{Y}'\mathbf{A}\mathbf{Y} = \sum_{i,j} a_{ij} Y_i Y_j$, $\mathbf{Y}'\mathbf{B}\mathbf{Y} = \sum_{k,l} b_{kl} Y_k Y_l$，以及

$$E(Y_iY_jY_kY_l) = \begin{cases} 3, & i=j=k=l; \\ 1, & i=j\neq k=l \\ & i=k\neq j=l \\ & i=l\neq k=j; \\ 0, & \text{其他}. \end{cases}$$

则

$$\begin{aligned} E(\mathbf{Y}'\mathbf{A}\mathbf{Y}\mathbf{Y}'\mathbf{B}\mathbf{Y}) &= 3\sum_{i=1}^{p} a_{ii}b_{ii} + \sum_{1\le i\neq k\le p} a_{ii}b_{kk} + \sum_{1\le i\neq j\le p} a_{ij}b_{ij} + \sum_{1\le i\neq k\le p} a_{ik}b_{ki} \\ &= 3\sum_{i=1}^{p} a_{ii}b_{ii} + \sum_{1\le i\neq k\le p} a_{ii}b_{kk} + 2\sum_{1\le i\neq j\le p} a_{ij}b_{ij}. \end{aligned}$$

注意到

$$\sum_{1\le i\ne k\le p} a_{ii}b_{kk}=\sum_{k=1}^{p}\sum_{i=1}^{p}a_{ii}b_{kk}-\sum_{i=1}^{p}a_{ii}b_{ii}=\mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})-\sum_{i=1}^{p}a_{ii}b_{ii},$$

$$\sum_{1\le i\ne j\le p} a_{ij}b_{ij}=\sum_{j=1}^{p}\sum_{i=1}^{p}a_{ij}b_{ij}-\sum_{i=1}^{p}a_{ii}b_{ii}=\mathrm{tr}(\mathbf{AB})-\sum_{i=1}^{p}a_{ii}b_{ii}.$$

则

$$E(\boldsymbol{Y}'\mathbf{A}\boldsymbol{Y}\boldsymbol{Y}'\mathbf{B}\boldsymbol{Y})=2\mathrm{tr}(\mathbf{AB})+\mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B}).$$

最后可得到：

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{Y}'\mathbf{A}\boldsymbol{Y},\boldsymbol{Y}'\mathbf{B}\boldsymbol{Y})&=E(\boldsymbol{Y}'\mathbf{A}\boldsymbol{Y}\boldsymbol{Y}'\mathbf{B}\boldsymbol{Y})-E(\boldsymbol{Y}'\mathbf{A}\boldsymbol{Y})E(\boldsymbol{Y}'\mathbf{B}\boldsymbol{Y})\\&=2\mathrm{tr}(\mathbf{AB}).\end{aligned}$$

多元正态分布的定义

- 多元统计分析的主要理论都是建立在多元正态分布总体的基础上，因此多元正态分布是多元统计分析的基础。
- 思考：如何给出多元正态分布的定义？

- 多元统计分析的主要理论都是建立在多元正态分布总体的基础上，因此多元正态分布是多元统计分析的基础。
- 思考：如何给出多元正态分布的定义？

定义3.8: 多元正态分布

设$Y_1, \cdots, Y_q$为i.i.d.的随机变量序列，$Y_i \sim N(0,1), i = 1, \cdots, q$，$\mathbf{A}$是$p \times q$常数矩阵，$\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)'$为$p \times 1$常数向量。记$\boldsymbol{Y} = (Y_1, \cdots, Y_q)'$，则称随机向量$\boldsymbol{Y}$的线性组合$\boldsymbol{X} = \mathbf{A}\boldsymbol{Y} + \boldsymbol{\mu}$为**$p$元(维)正态分布**，记成$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，其中$\Sigma = \mathbf{A}\mathbf{A}'$。

可以从联合密度函数定义多元正态分布，定义如下：

定义3.9: 基于联合密度函数的多元正态分布定义

设$\boldsymbol{X}$是一个p维随机向量，具有联合概率密度函数

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \right\},$$

其中$\boldsymbol{x} = (x_1, \cdots, x_p)'$，$\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)'$是$p$维实向量，$\boldsymbol{\Sigma}$是$p$阶正定矩阵，$|\boldsymbol{\Sigma}|$为$\boldsymbol{\Sigma}$的行列式值，$\boldsymbol{\Sigma}^{-1}$为$\boldsymbol{\Sigma}$的逆矩阵，则称$\boldsymbol{X}$服从均值向量为$\boldsymbol{\mu}$，协方差阵为$\boldsymbol{\Sigma}$的$p$**元正态分布**，记成$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$。

作业: 如何推导得到上面的多元正态概率密度函数？

可以从联合密度函数定义多元正态分布，定义如下：

定义3.9: 基于联合密度函数的多元正态分布定义

设$\boldsymbol{X}$是一个p维随机向量，具有联合概率密度函数

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \right\},$$

其中$\boldsymbol{x} = (x_1, \cdots, x_p)'$，$\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)'$是$p$维实向量，$\Sigma$是$p$阶正定矩阵，$|\Sigma|$为$\Sigma$的行列式值，$\Sigma^{-1}$为$\Sigma$的逆矩阵，则称$\boldsymbol{X}$服从均值向量为$\boldsymbol{\mu}$，协方差阵为$\Sigma$的$p$元正态分布，记成$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$。

作业: 如何推导得到上面的多元正态概率密度函数？
(参考Anderson (2003)的2.3节。)

由定义3.9，多元正态分布密度函数在

$$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=c^2 \quad (c\text{为常数})$$

上是常数。因此，在p维欧氏空间上，所有满足上式的$\boldsymbol{x}$称为常数密度函数轮廓线，它是以$\boldsymbol{\mu}$为中心，$\pm c\sqrt{\lambda_i}\boldsymbol{\phi}_i\ (i=1,\cdots,p)$为轴的椭圆面。

- $\lambda_i(i=1,\cdots,p)$为$\boldsymbol{\Sigma}$的特征值
- $\boldsymbol{\phi}_i$为λ_i对应的单位特征向量，$\boldsymbol{\Sigma}\boldsymbol{\phi}_i=\lambda_i\boldsymbol{\phi}_i$
- 椭球的形状和方向由$\boldsymbol{\Sigma}$决定
- 椭球的大小由常数c^2决定
- 如果$\boldsymbol{\Sigma}=\sigma^2\mathbf{I}_p$，则定义3.9给出的密度函数为球密度函数

多元正态分布的性质：

性质3.2.1: 均值和协方差阵

设$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，则$E(\boldsymbol{X}) = \boldsymbol{\mu}$，$\mathrm{Cov}(\boldsymbol{X}) = \Sigma$。

性质3.2.2: 特征函数

设$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，则

$$\varphi_X(\boldsymbol{t}) = \exp\left\{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right\},$$

其中$\varphi_X(\boldsymbol{t}) = \varphi_X(t_1, \cdots, t_p)$为多元正态随机向量$\boldsymbol{X}$的**特征函数**，$\boldsymbol{t} = (t_1, \cdots, t_p)'$，$\mathrm{i} = \sqrt{-1}$。

证明： 令$\boldsymbol{z} = (z_1, \cdots, z_q)'$，标准正态随机向量$\boldsymbol{Y}$的特征函数为$\varphi_{\boldsymbol{Y}}(\boldsymbol{z}) = \varphi_{\boldsymbol{Y}}(z_1, \cdots, z_q)$，其中$\varphi_j(z_j)$为$Y_j$的特征函数，$j = 1, \cdots, q$，则

$$\varphi_{\boldsymbol{Y}}(\boldsymbol{z}) = \prod_{j=1}^{q} \varphi_j(z_j) = \prod_{j=1}^{q} \exp\left\{ -\frac{1}{2} z_j^2 \right\} = \exp\left\{ -\frac{1}{2} \boldsymbol{z}'\boldsymbol{z} \right\}.$$

由随机向量特征函数性质，得$\boldsymbol{X} = \mathbf{A}\boldsymbol{Y} + \boldsymbol{\mu}$的特征函数为

$$\begin{aligned}
\varphi_{\boldsymbol{X}}(\boldsymbol{t}) &= \exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}\}\varphi_{\boldsymbol{Y}}(\mathbf{A}'\boldsymbol{t}) \\
&= \exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}\} \exp\left\{ -\frac{1}{2} \boldsymbol{t}'\mathbf{A}\mathbf{A}'\boldsymbol{t} \right\} \\
&= \exp\left\{ \mathrm{i}\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2} \boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t} \right\}.
\end{aligned}$$

- 由性质3.2.2可知，特征函数$\varphi_X(\boldsymbol{t})$只与$\boldsymbol{\mu}$和$\Sigma$有关。
- 由随机向量与其特征函数一一对应关系可知，p元正态分布$N_p(\boldsymbol{\mu},\Sigma)$又可用特征函数定义如下。

定义3.10: 基于特征函数的多元正态分布定义

设$\boldsymbol{X}=(X_1,\cdots,X_p)'$为$p$维随机向量，若其特征函数为

$$\varphi_X(\boldsymbol{t})=\exp\left\{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right\},$$

其中$\boldsymbol{t}=(t_1,\cdots,t_p)'$，$\mathrm{i}=\sqrt{-1}$。则称$\boldsymbol{X}$服从$p$元正态分布，记成$\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\Sigma)$。

性质3.2.3: 线性变换

设$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，$\mathbf{B}$为$q \times p$的常数矩阵，$\boldsymbol{\theta}$为$q \times 1$常向量，令$\boldsymbol{Z} = \mathbf{B}\boldsymbol{X} + \boldsymbol{\theta}$，则$\boldsymbol{Z} \sim N_q(\mathbf{B}\boldsymbol{\mu} + \boldsymbol{\theta}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}')$。

性质3.2.3: 线性变换

设$X \sim N_p(\mu, \Sigma)$，$\mathbf{B}$为$q \times p$的常数矩阵，θ为$q \times 1$常向量，令$Z = \mathbf{B}X + \theta$，则$Z \sim N_q(\mathbf{B}\mu + \theta, \mathbf{B}\Sigma\mathbf{B}')$。

证明： 令$t = (t_1, \cdots, t_p)'$，$s = (s_1, \cdots, s_q)'$。由定义3.10 知：随机向量X的特征函数为

$$\varphi_X(t) = \exp\left\{ it'\mu - \frac{1}{2}t'\Sigma t \right\}.$$

由特征函数的性质，得$\mathbf{Z} = \mathbf{B}\mathbf{X} + \boldsymbol{\theta}$的特征函数为

$$\begin{aligned}
\varphi_Z(\boldsymbol{s}) &= \exp\{\mathrm{i}\boldsymbol{s}'\boldsymbol{\theta}\}\,\varphi_X(\mathbf{B}'\boldsymbol{s}) \\
&= \exp\{\mathrm{i}\boldsymbol{s}'\boldsymbol{\theta}\}\ \exp\left\{\mathrm{i}\boldsymbol{s}'\mathbf{B}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{s}'\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}'\boldsymbol{s}\right\} \\
&= \exp\left\{\mathrm{i}\boldsymbol{s}'(\mathbf{B}\boldsymbol{\mu} + \boldsymbol{\theta}) - \frac{1}{2}\boldsymbol{s}'\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}'\boldsymbol{s}\right\}.
\end{aligned}$$

由定义3.10可知，$\mathbf{Z} \sim N_q(\mathbf{B}\boldsymbol{\mu} + \boldsymbol{\theta}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}')$。

性质3.2.4:

设$\boldsymbol{X} = (X_1, \cdots, X_p)'$为$p$维随机向量，则$\boldsymbol{X}$服从$p$元正态分布$N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$的充分必要条件是它的任意一个线性组合$\boldsymbol{a}'\boldsymbol{X}$都服从一元正态分布，其中$\boldsymbol{a} \neq \boldsymbol{0}$。

性质3.2.4:

设$\boldsymbol{X}=(X_1,\cdots,X_p)'$为$p$维随机向量，则$\boldsymbol{X}$服从$p$元正态分布$N_p(\boldsymbol{\mu},\Sigma)$的充分必要条件是它的任意一个线性组合$\boldsymbol{a}'\boldsymbol{X}$都服从一元正态分布，其中$\boldsymbol{a}\neq\boldsymbol{0}$。

证明：“$\Rightarrow$"显然。现在证明“$\Leftarrow$"。对任意非零向量$\boldsymbol{a}\in\mathbb{R}^p$，由于$\boldsymbol{a}'\boldsymbol{X}$服从一元正态分布，故$\boldsymbol{X}$的二阶矩存在，记$E(\boldsymbol{X})=\boldsymbol{\mu}$，$\mathrm{Cov}(\boldsymbol{X})=\Sigma$，则$E(\boldsymbol{a}'\boldsymbol{X})=\boldsymbol{a}'\boldsymbol{\mu}$，$\mathrm{Var}(\boldsymbol{a}'\boldsymbol{X})=\boldsymbol{a}'\Sigma\boldsymbol{a}$。注意到$\boldsymbol{a}'\boldsymbol{X}$的特征函数

$$\varphi_{\boldsymbol{a}'\boldsymbol{X}}(t)=\exp\left\{\mathrm{i}t\boldsymbol{a}'\boldsymbol{\mu}-\frac{1}{2}t^2\boldsymbol{a}'\Sigma\boldsymbol{a}\right\},\quad t\in\mathbb{R}.$$

令$t=1$, 有$\varphi_{\boldsymbol{a}'\boldsymbol{X}}(1)=\exp\left\{\mathrm{i}\boldsymbol{a}'\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{a}'\Sigma\boldsymbol{a}\right\}=\varphi_{\boldsymbol{X}}(\boldsymbol{a})$。视其为$\boldsymbol{a}$的函数，则由定义3.10，可知$\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\Sigma)$。

性质3.2.5：可加性

设$\boldsymbol{X}_1, \cdots, \boldsymbol{X}_k$是相互独立的$k$组$p$维随机向量，$\boldsymbol{X}_i \sim N_p(\boldsymbol{\mu}_i, \Sigma_i), i = 1, \cdots, k$，则

$$\sum_{i=1}^{k} \boldsymbol{X}_i \sim N_p\left(\sum_{i=1}^{k} \boldsymbol{\mu}_i, \sum_{i=1}^{k} \Sigma_i\right).$$

性质3.2.6：二次型

设$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，$\boldsymbol{\Sigma} > 0$，对于$p$阶矩阵$\mathbf{A} \geq 0$，则二次型$(\boldsymbol{X}-\boldsymbol{\mu})'\mathbf{A}(\boldsymbol{X}-\boldsymbol{\mu})$服从中心$\chi^2_m$分布的充要条件为$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma} = \mathbf{A}\boldsymbol{\Sigma}$，且其自由度$m = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})$。特别地，在$\mathbf{A} = \boldsymbol{\Sigma}^{-1}$时，$(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$服从自由度为$p$的中心$\chi^2$分布。

证明留给大家课后作业。

mvtnorm程序包

- install.packages("mvtnorm")　　# 安装程序包
- library(mvtnorm)　　# 加载程序包
- dmvnorm()　　计算多元正态密度函数
- pmvnorm()　　计算分布函数
- qmvnorm()　　计算分位数
- rmvnorm()　　产生多元正态分布的随机数

二元正态分布

设二维随机向量$\boldsymbol{Z} = (X,Y)' \sim N_2(\boldsymbol{\mu},\boldsymbol{\Sigma})$，或记为$(X,Y)' \sim N_2(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，其中$\sigma_1>0,\sigma_2>0,|\rho|<1$，并且

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},\quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} =: \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} > 0.$$

二维正态随机向量$(X,Y)'$的联合密度函数为

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})\right] \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right\}, \end{aligned}$$

其中$\boldsymbol{z}=(x,y)'$。

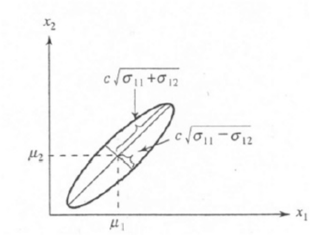
二元正态分布

例1： 对二元正态$N_2\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}\right)$，求其常数密度轮廓线椭圆。

解： 由$|\lambda \mathbf{I}_2 - \boldsymbol{\Sigma}| = 0$和$\boldsymbol{\Sigma}\phi = \lambda\phi$，则易得其特征值和特征向量为

$$\lambda_1 = \sigma_{11} + \sigma_{12}, \phi_1 = (1,1)'/\sqrt{2}$$
$$\lambda_2 = \sigma_{11} - \sigma_{12}, \phi_2 = (1,-1)'/\sqrt{2}$$

Bivariate Normal Contours $(\sigma_{11} = \sigma_{22})$



课下习题：

- 假设$p=2, \boldsymbol{X}=(X_1,X_2)' \sim N_2(\boldsymbol{\mu},\boldsymbol{\Sigma})$
- 其中$\boldsymbol{\mu}=(1,2)'$， $\boldsymbol{\Sigma}=\begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}$
- 问题： 用R语言写程序，根据上面二元正态分布产生100个随机数，画出密度轮廓线椭圆。

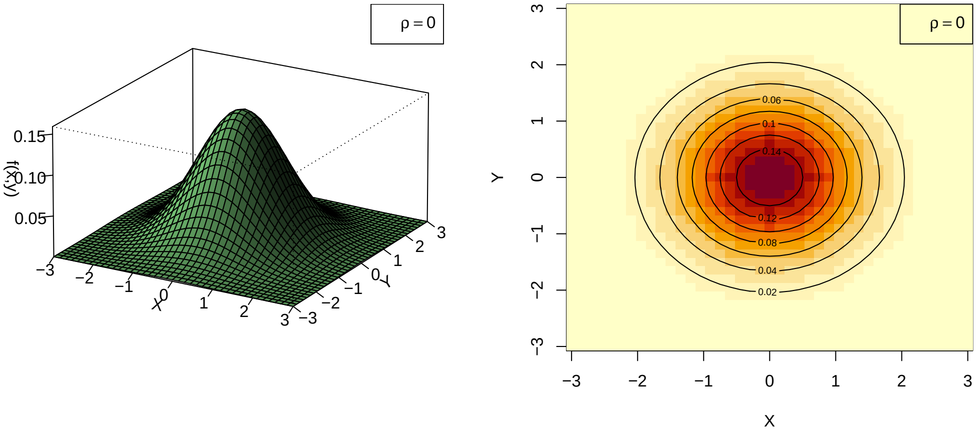


Figure 1: $\sigma_1^2 = 1, \sigma_2^2 = 1, \rho = 0$时的二元正态密度函数及其等高线图。

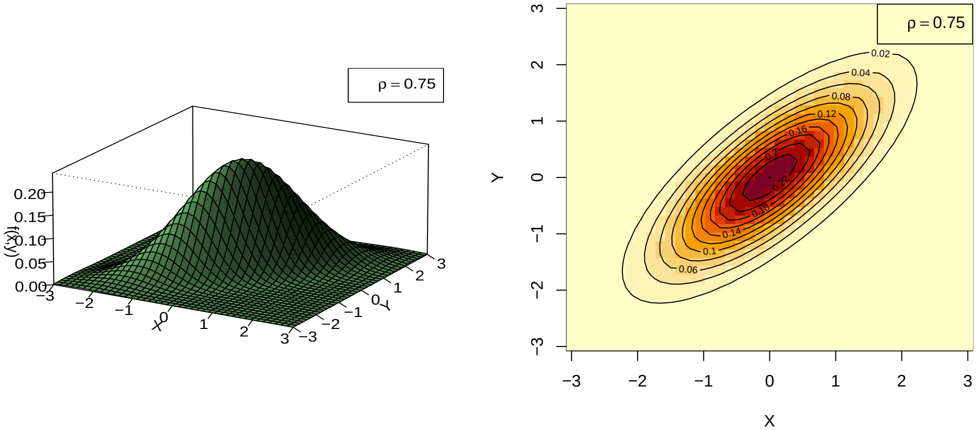
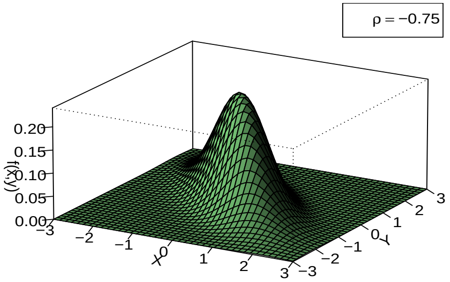


Figure 2: $\sigma_1^2 = 1, \sigma_2^2 = 1, \rho = 0.75$时的二元正态密度函数及其等高线图。



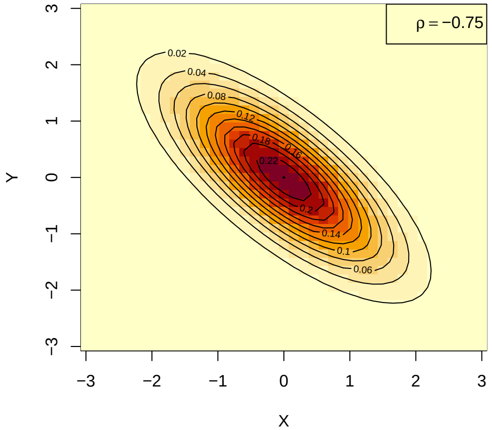


Figure 3: $\sigma_1^2 = 1, \sigma_2^2 = 1, \rho = -0.75$时的二元正态密度函数及其等高线图。

设$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $p \geq 2$，将$\boldsymbol{X}, \boldsymbol{\mu}$和$\boldsymbol{\Sigma}$剖分为

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

其中

- $\boldsymbol{X}^{(1)}$和$\boldsymbol{\mu}^{(1)}$为$q \times 1$的向量
- $\boldsymbol{\Sigma}_{11}$为$q \times q$矩阵
- $\boldsymbol{X}^{(2)}$和$\boldsymbol{\mu}^{(2)}$ 为$(p-q) \times 1$的向量
- $\boldsymbol{\Sigma}_{22}$为$(p-q) \times (p-q)$矩阵
- $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}'$为$q \times (p-q)$矩阵

性质3.2.7: 边际分布

设$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma), p \geq 2$，则

$$\boldsymbol{X}^{(1)} \sim N_q(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11}), \quad \boldsymbol{X}^{(2)} \sim N_{p-q}(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}).$$

性质3.2.7: 边际分布

设$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma), p \geq 2$，则

$$\boldsymbol{X}^{(1)} \sim N_q(\boldsymbol{\mu}^{(1)}, \Sigma_{11}), \quad \boldsymbol{X}^{(2)} \sim N_{p-q}(\boldsymbol{\mu}^{(2)}, \Sigma_{22}).$$

证明1： 令$\boldsymbol{X}^{(1)} = \mathbf{B}^{(1)}\boldsymbol{X} + \boldsymbol{\theta}^{(1)}$和$\boldsymbol{X}^{(2)} = \mathbf{B}^{(2)}\boldsymbol{X} + \boldsymbol{\theta}^{(2)}$，其中取

$$\mathbf{B}^{(1)}_{q\times p} = (\mathbf{I}_q, \mathbf{0}_{q\times(p-q)}), \quad \boldsymbol{\theta}^{(1)} = \mathbf{0},$$

$$\mathbf{B}^{(2)}_{(p-q)\times p} = (\mathbf{0}_{(p-q)\times q}, \mathbf{I}_{p-q}), \quad \boldsymbol{\theta}^{(2)} = \mathbf{0}.$$

由性质3.2.3，即可证明性质3.2.7。

证明2：　由分块矩阵的行列式和逆公式，有

$$|\mathbf{\Sigma}| = |\mathbf{\Sigma}_{11}||\mathbf{\Sigma}_{22.1}|,$$

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} \mathbf{I}_q & -\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} \\ \mathbf{0} & \mathbf{I}_{p-q} \end{pmatrix} \begin{pmatrix} \mathbf{\Sigma}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{22.1}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_q & \mathbf{0} \\ -\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1} & \mathbf{I}_{p-q} \end{pmatrix},$$

其中$\mathbf{\Sigma}_{22.1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$。则$\boldsymbol{X}$的密度函数$f(\boldsymbol{x}) = f(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)})$为：

$$f(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = f_1(\boldsymbol{x}^{(1)}) f(\boldsymbol{x}^{(2)}|\boldsymbol{x}^{(1)}),$$

其中

$$f_1(\boldsymbol{x}^{(1)}) = \frac{1}{(2\pi)^{q/2}|\mathbf{\Sigma}_{11}|^{1/2}} \exp\left[-\frac{1}{2}\left(\boldsymbol{x}^{(1)} - \boldsymbol{\mu}^{(1)}\right)' \mathbf{\Sigma}_{11}^{-1} \left(\boldsymbol{x}^{(1)} - \boldsymbol{\mu}^{(1)}\right)\right],$$

$$f(\boldsymbol{x}^{(2)}|\boldsymbol{x}^{(1)}) = \frac{1}{(2\pi)^{(p-q)/2}|\Sigma_{22.1}|^{1/2}} \exp\left[-\frac{1}{2}\left(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}_{2.1}\right)' \Sigma_{22.1}^{-1}\left(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}_{2.1}\right)\right],$$

其中$\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}^{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}^{(1)} - \boldsymbol{\mu}^{(1)})$。可知$f_1(\boldsymbol{x}^{(1)})$是$q$元正态分布$N_q(\boldsymbol{\mu}^{(1)}, \Sigma_{11})$的概率密度函数；此外，如果把$\boldsymbol{x}^{(1)}$当作一个常数，则$f(\boldsymbol{x}^{(2)}|\boldsymbol{x}^{(1)})$是$p-q$元正态分布$N_{p-q}(\boldsymbol{\mu}_{2.1}, \Sigma_{22.1})$的概率密度函数。由边缘概率密度函数的定义，$\boldsymbol{X}^{(1)}$的边缘密度函数为：

$$\int f(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)})\mathrm{d}\boldsymbol{x}^{(2)} = f_1(\boldsymbol{x}^{(1)}) \int f(\boldsymbol{x}^{(2)}|\boldsymbol{x}^{(1)})\mathrm{d}\boldsymbol{x}^{(2)} = f_1(\boldsymbol{x}^{(1)}).$$

由此可证明: $\boldsymbol{X}^{(1)} \sim N_q(\boldsymbol{\mu}^{(1)}, \Sigma_{11})$。

思考： 考虑用特征函数证明性质3.2.7?

定理3.2.1

设$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，$\Sigma > 0$，对$\boldsymbol{X}$和$\Sigma$作剖分，则

1. 给定$\boldsymbol{X}^{(2)} = \boldsymbol{x}^{(2)}$时，$\boldsymbol{X}^{(1)}$的条件分布服从$q$元正态分布，即

$$(\boldsymbol{X}^{(1)} | \boldsymbol{X}^{(2)} = \boldsymbol{x}^{(2)}) \sim N_q(\boldsymbol{\mu}_{1.2}, \Sigma_{11.2}),$$

   其中$\boldsymbol{\mu}_{1.2} = \boldsymbol{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)}), \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$。

2. 给定$\boldsymbol{X}^{(1)} = \boldsymbol{x}^{(1)}$时，$\boldsymbol{X}^{(2)}$的条件分布服从$p-q$元正态分布，即

$$(\boldsymbol{X}^{(2)} | \boldsymbol{X}^{(1)} = \boldsymbol{x}^{(1)}) \sim N_{p-q}(\boldsymbol{\mu}_{2.1}, \Sigma_{22.1}),$$

   其中$\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}^{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}^{(1)} - \boldsymbol{\mu}^{(1)}), \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$。

证明： 由性质3.2.7的证明，以及条件密度定义及分块矩阵的逆易证。

由性质3.2.7和定理3.2.1可知:

- $\boldsymbol{X}^{(1)}$和$(\boldsymbol{X}^{(1)}|\boldsymbol{X}^{(2)})$的分布都服从正态分布，它们的协方差阵分别为$\boldsymbol{\Sigma}_{11}$和$\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$;
- 由于$\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \geq 0$，故$\boldsymbol{\Sigma}_{11} \geq \boldsymbol{\Sigma}_{11.2}$，等号成立当且仅当$\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' \equiv \boldsymbol{0}$;
- 协方差阵描述变量之间关系及散布程度的，$\boldsymbol{\Sigma}_{11} \geq \boldsymbol{\Sigma}_{11.2}$表明：在已知$\boldsymbol{X}^{(2)}$的条件下，减少了$\boldsymbol{X}^{(1)}$的散布程度，而且只有当$\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$时，两者相等;
- $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$等价于$\boldsymbol{X}^{(1)}$与$\boldsymbol{X}^{(2)}$相互独立。独立性表明，即使给定$\boldsymbol{X}^{(2)}$的条件下，对$\boldsymbol{X}^{(1)}$的分布没有影响。

在二元正态场合$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right)$,

则$X_1|X_2 = x_2 \sim N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right)$。

注：$\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} = \sigma_{11}(1-\rho^2)$. $|\rho|$越大，方差越小。

例2

假设大学生男生的体重(lbs) 和身高(inches，英寸) 服从多元正态分布，均值$\boldsymbol{\mu} = (175, 71)'$和协方差矩阵$\Sigma = \begin{pmatrix} 550 & 40 \\ 40 & 8 \end{pmatrix}$，试求给定身高为$X_2 = x_2$条件下体重$X_1$ 的分布。

解：由定理3.2.1，$X_1|X_2 = x_2 \sim N(-180 + 5x_2, 350)$，例如当身高为$x_2 = 70$ 时，体重则服从$N(170, 350)$。

定理3.2.2：

设$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma} > 0$，对$\boldsymbol{X}$和$\boldsymbol{\Sigma}$作(1)的剖分，则$\boldsymbol{X}^{(1)}$与$\boldsymbol{X}^{(2)}$相互独立当且仅当$\boldsymbol{\Sigma}_{12} = \mathbf{0}$。

定理3.2.2:

设$X \sim N_p(\mu, \Sigma), \Sigma > 0$，对$X$和$\Sigma$作(1)的剖分，则$X^{(1)}$与$X^{(2)}$相互独立当且仅当$\Sigma_{12} = \mathbf{0}$。

证明："⇒" 当$X^{(1)}$与$X^{(2)}$相互独立时，显然

$$\mathrm{Cov}(X^{(1)}, X^{(2)}) = \Sigma_{12} = \mathbf{0}.$$

"⇐"当$\Sigma_{12} = \mathbf{0}$时，由性质3.2.7的证明可知，$\Sigma_{22.1} = \Sigma_{22}$，则

$$f(x^{(1)}, x^{(2)}) = f_{X^{(1)}}(x^{(1)}) f_{X^{(2)}}(x^{(2)}).$$

可知，$X^{(1)}$与$X^{(2)}$相互独立。

思考：对任何分布，定理3.2.2成立吗？

定理3.2.3:

设$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，$\Sigma > 0$，对$\boldsymbol{X}$和$\Sigma$作类似(1)的剖分：

$$\boldsymbol{X} = \begin{array}{c} \begin{pmatrix} \boldsymbol{X}^{(1)} \\ \vdots \\ \boldsymbol{X}^{(m)} \end{pmatrix} \end{array} \begin{array}{c} q_1 \\ \vdots \\ q_m \end{array}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1m} \\ \vdots & & \vdots \\ \Sigma_{m1} & \cdots & \Sigma_{mm} \end{pmatrix} \begin{array}{c} q_1 \\ \vdots \\ q_m \end{array},$$

其中$q_1 + \cdots + q_m = p$，则$\mathrm{Cov}(\boldsymbol{X}^{(i)}, \boldsymbol{X}^{(j)}) = \Sigma_{ij} = \mathbf{0} (1 \leq i < j \leq m)$都成立，是$\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(m)}$相互独立的充要条件。

推论3.2.1：

在定理3.2.1的条件下，若$\boldsymbol{X}^{(1)}$与$\boldsymbol{X}^{(2)}$相互独立，即$\Sigma_{12} = \Sigma_{21}' = \mathbf{0}$，则存在$q$维实数向量$\boldsymbol{a} = (a_1, \cdots, a_q)'$和$p-q$维的实数向量$\boldsymbol{b} = (b_1, \cdots, b_{p-q})'$，使得变量$\xi = \boldsymbol{a}'\boldsymbol{X}^{(1)}$和$\eta = \boldsymbol{b}'\boldsymbol{X}^{(2)}$相互独立。

推论3.2.2：

设$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，$\Sigma > 0$，若$\Sigma$为对角矩阵，则$X_1, \cdots, X_p$相互独立。

定理3.2.4：

在定理3.2.1的条件下，则

(1) $\boldsymbol{X}^{(1)}$与$\boldsymbol{X}^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{X}^{(1)}$相互独立，并且

$$\begin{aligned}
&\boldsymbol{X}^{(1)} \sim N_q(\boldsymbol{\mu}^{(1)}, \Sigma_{11}),\\
&\boldsymbol{X}^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{X}^{(1)} \sim N_{p-q}(\boldsymbol{\mu}^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}^{(1)}, \Sigma_{22.1});
\end{aligned}$$

(2) $\boldsymbol{X}^{(2)}$与$\boldsymbol{X}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{X}^{(2)}$相互独立，并且

$$\begin{aligned}
&\boldsymbol{X}^{(2)} \sim N_{p-q}(\boldsymbol{\mu}^{(2)}, \Sigma_{22}),\\
&\boldsymbol{X}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{X}^{(2)} \sim N_q(\boldsymbol{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}^{(2)}, \Sigma_{11.2}).
\end{aligned}$$

证明： 令

$$\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{Y}^{(1)}\\ \boldsymbol{Y}^{(2)}\end{pmatrix}=\begin{pmatrix}\mathbf{I}_q & \mathbf{0}\\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I}_{p-q}\end{pmatrix}\begin{pmatrix}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\end{pmatrix},$$

则$\boldsymbol{Y}^{(1)}=\boldsymbol{X}^{(1)},\boldsymbol{Y}^{(2)}=\boldsymbol{X}^{(2)}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}^{(1)}$。由于$\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$，利用性质3.2.3，通过简单计算有

$$\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{Y}^{(1)}\\ \boldsymbol{Y}^{(2)}\end{pmatrix}\sim N_p\left(\begin{pmatrix}\boldsymbol{\mu}^{(1)}\\ \boldsymbol{\mu}^{(2)}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}^{(1)}\end{pmatrix},\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_{22.1}\end{pmatrix}\right).$$

由定理3.2.2知，$\boldsymbol{Y}^{(1)}=\boldsymbol{X}^{(1)}$和$\boldsymbol{Y}^{(2)}=\boldsymbol{X}^{(2)}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}^{(1)}$相互独立，且

$$\boldsymbol{X}^{(1)}\sim N_q(\boldsymbol{\mu}^{(1)},\boldsymbol{\Sigma}_{11}),$$
$$\boldsymbol{X}^{(2)}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}^{(1)}\sim N_{p-q}(\boldsymbol{\mu}^{(2)}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}^{(1)},\boldsymbol{\Sigma}_{22.1}).$$

由定理3.2.1，可知

$$E(\boldsymbol{X}^{(1)}|\boldsymbol{X}^{(2)}) = \boldsymbol{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{X}^{(2)} - \boldsymbol{\mu}^{(2)})$$

- 这表明$\boldsymbol{X}^{(1)}$和$\boldsymbol{X}^{(2)}$之间有回归关系；
- 于是称矩阵$\beta = \Sigma_{12}\Sigma_{22}^{-1}$为$\boldsymbol{X}^{(1)}$对$\boldsymbol{X}^{(2)}$的回归系数;
- $\Sigma_{11.2}$为条件协方差矩阵，它的元素用$\sigma_{ij\cdot q+1,\cdots,p}$表示，$i,j = 1,\cdots,q$。

- 偏相关系数 (partial correlation of coefficient): 由于$(\boldsymbol{X}^{(1)}|\boldsymbol{X}^{(2)} = \boldsymbol{x}^{(2)}) \sim N_q(\boldsymbol{\mu}_{1.2}, \boldsymbol{\Sigma}_{11.2})$, 故可定义条件相关系数

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

为在给定$\boldsymbol{X}^{(2)} = \boldsymbol{x}^{(2)}$条件下, X_i 与X_j 的偏相关系数 $(\mathrm{corr}(\boldsymbol{X}^{(1)}|\boldsymbol{X}^{(2)}))$。

- 显然

$$\mathbf{R}_{11.2} = \left[\mathrm{diag}\left(\boldsymbol{\Sigma}_{11.2}\right)\right]^{-1/2} \boldsymbol{\Sigma}_{11.2} \left[\mathrm{diag}\left(\boldsymbol{\Sigma}_{11.2}\right)\right]^{-1/2}.$$

定义3.12: 残差向量和回归系数

向量$\varepsilon^{(1.2)} = X^{(1)} - \mu^{(1)} - \beta(X^{(2)} - \mu^{(2)})$称为$X^{(1)}$ 对$X^{(2)}$ 的残差向量，其中$\beta = \Sigma_{12}\Sigma_{22}^{-1}$ 为$X^{(1)}$ 对$X^{(2)}$ 的回归系数，它的第i行第$(k-q)$列元素用$\beta_{ik\cdot q+1,\cdots,p}$ 表示，$i = 1,\cdots,q, k = q+1,\cdots,p$。

定理3.3.1

残差向量$\varepsilon^{(1.2)}$与$X^{(2)}$是不相关的。

证明： 由定理3.2.4知，$\varepsilon^{(1.2)} = Y^{(1)} - E(Y^{(1)})$, 其中

$$Y^{(1)} = X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}X^{(2)}.$$

结果可由定理3.2.4得到。

定理3.3.2

对于任意一个向量$\boldsymbol{a}$，则有$\mathrm{Var}(\varepsilon_i^{(1.2)}) \leq \mathrm{Var}(X_i - \boldsymbol{a}'\boldsymbol{X}^{(2)})$，其中$\varepsilon_i^{(1.2)}$为$\boldsymbol{\varepsilon}^{(1.2)}$的第$i$个分量。

证明： 由定理3.3.1和简单计算，有

$$\begin{aligned}\mathrm{Var}(X_i - \boldsymbol{a}'\boldsymbol{X}^{(2)}) &= E[X_i - \mu_i - \boldsymbol{a}'(\boldsymbol{X}^{(2)} - \boldsymbol{\mu}^{(2)})]^2 \\ &= E[\varepsilon_i^{(1.2)} - E(\varepsilon_i^{(1.2)}) + (\boldsymbol{\beta}_{(i)} - \boldsymbol{a})'(\boldsymbol{X}^{(2)} - \boldsymbol{\mu}^{(2)})]^2 \\ &= \mathrm{Var}(\varepsilon_i^{(1.2)}) + (\boldsymbol{\beta}_{(i)} - \boldsymbol{a})'\Sigma_{22}(\boldsymbol{\beta}_{(i)} - \boldsymbol{a}),\end{aligned}$$

其中$\boldsymbol{\beta}_{(i)}'$是$\boldsymbol{\beta}$的第$i$行。因为$\Sigma_{22}$是正定矩阵，则上式第二项二次型$(\boldsymbol{\beta}_{(i)} - \boldsymbol{a})'\Sigma_{22}(\boldsymbol{\beta}_{(i)} - \boldsymbol{a}) \geq 0$。等号当且仅当$\boldsymbol{\beta}_{(i)} = \boldsymbol{a}$时成立。

注：定理3.3.2表明$\mu_i + \boldsymbol{\beta}_{(i)}'(\boldsymbol{X}^{(2)} - \boldsymbol{\mu}^{(2)})$是$X_i$的一个最佳线性预测。

定理3.3.3

设$\boldsymbol{X} = (X_1, \cdots, X_p)' \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，$\boldsymbol{\Sigma} > 0$. 令$\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} = (\omega_{ij})_{i,j=1,\cdots,p}$为精度矩阵。有下面两个结论：

(1) 精度矩阵$\boldsymbol{\Omega}$的对角线上的元素ω_{ii}是其余变量给定后X_i的条件方差的倒数，即

$$\omega_{ii} = [\mathrm{Var}(X_i | X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_p)]^{-1}, \quad i = 1, \cdots, p.$$

(2) 类似于相关系数矩阵的计算。令

$$\mathbf{C} = \begin{pmatrix} c_{11} & \cdots & c_{1p} \\ \vdots & \ddots & \vdots \\ c_{p1} & \cdots & c_{pp} \end{pmatrix}, \qquad \boldsymbol{\Lambda} = \begin{pmatrix} \frac{1}{\sqrt{\omega_{11}}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\omega_{pp}}} \end{pmatrix}.$$

定理3.3.3(续)

并且令

$$\mathbf{C} = \mathbf{\Lambda\Omega\Lambda}.$$

也就是有

$$c_{ij} = \frac{\omega_{ij}}{\sqrt{\omega_{ii}}\sqrt{\omega_{jj}}}, \quad i, j = 1, \cdots, p.$$

可见矩阵$\mathbf{C}$对角线元素为1，而非对角线元素为c_{ij}，是其余变量给定后，X_i和X_j的偏相关系数的相反数，即

$$\begin{aligned} c_{ij} &= -\rho_{X_i, X_j | X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_{j-1}, X_{j+1}, \cdots, X_p} \\ &= -\rho_{ij.1, \cdots, i-1, i+1, \cdots, j-1, j+1, \cdots, p}, \quad 1 \le i < j \le p. \end{aligned}$$

精度矩阵

证明： 对于(1)，仅证明$i=1$的情况，即

$$\omega_{11} = [\mathrm{Var}(X_1|X_2,\cdots,X_p)]^{-1},$$

其他情况可类似证明。将协方差矩阵$\mathbf{\Sigma}$和精度矩阵$\mathbf{\Omega}$进行剖分：

$$\mathbf{\Sigma} = \begin{array}{cc} \begin{pmatrix} \sigma_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix} & \begin{array}{c} 1 \\ p-1 \end{array} \end{array}, \quad \mathbf{\Omega} = \begin{array}{cc} \begin{pmatrix} \omega_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{pmatrix} & \begin{array}{c} 1 \\ p-1 \end{array} \end{array},$$

由定理3.2.1，有：$\mathrm{Var}(X_1|X_2,\cdots,X_p) = \sigma_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$。
由$\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$, 根据分块矩阵的逆矩阵公式，有

$$\omega_{11} = (\sigma_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})^{-1}.$$

由上面两个公式，则证明了$i=1$的情况。

证明(续)：对于(2)，仅证明$i=1,j=2$的情况，即给定$X_3,\cdots,X_p$后，X_1与X_2的偏相关系数为：

$$\rho_{X_1,X_2|X_3,\cdots,X_p}=\rho_{12.3,\cdots,p}=-c_{12},$$

其他情况可类似证明。将协方差矩阵$\boldsymbol{\Sigma}$和精度矩阵$\boldsymbol{\Omega}$进行剖分：

$$\boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}\begin{array}{c}2\\ p-2\end{array},\quad \boldsymbol{\Omega}=\begin{pmatrix}\boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12}\\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22}\end{pmatrix}\begin{array}{c}2\\ p-2\end{array},$$

根据分块矩阵的逆矩阵公式，有

$$\boldsymbol{\Omega}_{11}=(\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1}.$$

证明(续)： 记

$$\mathbf{\Omega}_{11} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}, \quad \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

则

$$\omega_{11} = \frac{a_{22}}{a_{11}a_{22} - a_{12}a_{21}}, \quad \omega_{22} = \frac{a_{11}}{a_{11}a_{22} - a_{12}a_{21}}, \quad \omega_{12} = -\frac{a_{12}}{a_{11}a_{22} - a_{12}a_{21}},$$

在给定$X_3, \cdots, X_p$后，X_1与X_2的偏相关系数为

$$\rho_{12.3,\cdots,p} = \frac{a_{12}}{\sqrt{a_{11}a_{22}}}.$$

由于

$$c_{12} = \frac{\omega_{12}}{\sqrt{\omega_{11}\omega_{22}}} = -\frac{a_{12}}{\sqrt{a_{11}a_{22}}} = -\rho_{12.3,\cdots,p}.$$

- 由定理3.3.3(2)可知：精度矩阵Ω的非对角线元素$\omega_{ij}=0$, $\omega_{ij}>0$ 和$\omega_{ij}<0$分别表示其余变量给定后X_i与X_j条件独立，条件负相关和条件正相关的充要条件。
- 假设$\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma}), \boldsymbol{\Sigma}>0$, 将$\boldsymbol{X},\boldsymbol{\Sigma},\boldsymbol{\Omega}$类似进行剖分：

$$\boldsymbol{X}=\begin{array}{cc}\begin{pmatrix}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\\ \boldsymbol{X}^{(3)}\end{pmatrix} & \begin{array}{c}q_1\\ q_2\\ q_3\end{array}\end{array},\quad \boldsymbol{\Sigma}=\begin{array}{cc}\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23}\\ \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33}\end{pmatrix} & \begin{array}{c}q_1\\ q_2\\ q_3\end{array}\end{array},$$

$$\boldsymbol{\Omega}=\begin{array}{cc}\begin{pmatrix}\boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} & \boldsymbol{\Omega}_{13}\\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} & \boldsymbol{\Omega}_{23}\\ \boldsymbol{\Omega}_{31} & \boldsymbol{\Omega}_{32} & \boldsymbol{\Omega}_{33}\end{pmatrix} & \begin{array}{c}q_1\\ q_2\\ q_3\end{array}\end{array}.$$

定理3.3.4

在$\boldsymbol{X}^{(3)}$给定后，$\boldsymbol{X}^{(1)}$与$\boldsymbol{X}^{(2)}$相互条件独立的充要条件为$\boldsymbol{\Omega}_{12}=\boldsymbol{0}$。

证明：在$\boldsymbol{X}^{(3)}$给定后，$\boldsymbol{X}^{(1)}$与$\boldsymbol{X}^{(2)}$的条件协方差矩阵为

$$\begin{aligned}\boldsymbol{\Sigma}_{12.3} &= \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{23} \end{pmatrix} \boldsymbol{\Sigma}_{33}^{-1}(\boldsymbol{\Sigma}_{31}, \boldsymbol{\Sigma}_{32}) \\ &= \begin{pmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32} \\ \boldsymbol{\Sigma}_{21} - \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32} \end{pmatrix}.\end{aligned}$$

所以在给定$\boldsymbol{X}^{(3)}$后，$\boldsymbol{X}^{(1)}$与$\boldsymbol{X}^{(2)}$相互独立的充要条件是

$$\boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32} = \boldsymbol{0}.$$

由于$\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$，根据分块矩阵逆矩阵的计算公式，有

$$\begin{aligned}\begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}^{-1} &= \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{23} \end{pmatrix} \boldsymbol{\Sigma}_{33}^{-1}(\boldsymbol{\Sigma}_{31}, \boldsymbol{\Sigma}_{32}) \\ &= \begin{pmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32} \\ \boldsymbol{\Sigma}_{21} - \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32} \end{pmatrix}.\end{aligned}$$

由此可见：$\boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32} = \boldsymbol{0}$的充要条件是$\boldsymbol{\Omega}_{12} = \boldsymbol{0}$。

思考题：对于给定的相关系数矩阵$\mathbf{R} = (\rho_{ij})_{1\le i,j\le p}$，考虑下面两种结构：

(1) 当$i = j$时，$\rho_{ij} = 1$；当$i \neq j$时，$\rho_{ij} = \rho$；

(2) AR(1)自回归结构，即$\rho_{ij} = \rho^{|i-j|}$。

问题：观察当维数p变大时，精度矩阵$\mathbf{\Omega} = \mathbf{R}^{-1}$中元素的变化。

```
n = 100; p = 30; rho = 0.5
corrmat=diag(rep(1-rho,p))+matrix(rho,p,p)
solve(corrmat)
ar1mat=rho^outer(1:p,1:p, function(x,y) abs(x-y))
solve(ar1mat)
```

- 设总体$\boldsymbol{X}=(X_1,\cdots,X_p)'\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$
- 假设$\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$是来自总体$\boldsymbol{X}$的样本容量为$n$的随机样本，记为$n\times p$的样本矩阵$\mathbf{X}$，即

$$\mathbf{X}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)'=\begin{pmatrix}\boldsymbol{x}_1'\\ \vdots\\ \boldsymbol{x}_n'\end{pmatrix}=\begin{pmatrix}x_{11}&\cdots&x_{1p}\\ \vdots&&\vdots\\ x_{n1}&\cdots&x_{np}\end{pmatrix}.$$

- 讨论样本随机矩阵$\mathbf{X}$的分布时，通常把$\mathbf{X}$的行向量拉直构成一个长度为np的向量，然后讨论这个np维向量的分布。
- 为了讨论np维向量的分布，我们首先需要介绍矩阵的拉直运算和克罗内克(Kronecker)积运算。

- 矩阵拉直： 令$n\times p$阶矩阵$\mathbf{X}=(\mathbf{x}_1,\cdots,\mathbf{x}_p)$，其中$\mathbf{x}_i=(x_{1i},\cdots,x_{ni})'$，$i=1,\cdots,p$。所谓矩阵的拉直运算就是将矩阵按列拉直为向量，$\mathbf{X}$按列拉直后的向量记为$\mathrm{Vec}(\mathbf{X})$，它等于

$$\mathrm{Vec}(\mathbf{X})=\begin{pmatrix}\mathbf{x}_1\\ \vdots\\ \mathbf{x}_p\end{pmatrix}=(x_{11},x_{21},\cdots,x_{n1},\cdots,x_{1p},x_{2p},\cdots,x_{np})'.$$

由于$\mathbf{X}$是$n\times p$阶矩阵，所以$\mathrm{Vec}(\mathbf{X})$是$np\times 1$维向量。另外

$$\mathrm{Vec}(\mathbf{X}')=\begin{pmatrix}\boldsymbol{x}_1\\ \vdots\\ \boldsymbol{x}_n\end{pmatrix}=(x_{11},x_{12},\cdots,x_{1p},\cdots,x_{n1},x_{n2},\cdots,x_{np})'.$$

- 克罗内克(Kronecker)积： 令$\mathbf{A}$和$\mathbf{B}$分别是$n \times p$和$m \times q$阶矩阵，并记$\mathbf{A} = (a_{ij})$，其中a_{ij}为矩阵$\mathbf{A}$的元素，$i = 1, \cdots, n$, $j = 1, \cdots, p$。矩阵$\mathbf{A}$和$\mathbf{B}$的Kronecker乘积，记为$\mathbf{A} \otimes \mathbf{B}$，它等于

$$\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B}) = \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1p}\mathbf{B} \\ \vdots & & \vdots \\ a_{n1}\mathbf{B} & \cdots & a_{np}\mathbf{B} \end{pmatrix},$$

可知，$\mathbf{A} \otimes \mathbf{B}$是$nm \times pq$阶矩阵。

- 利用矩阵拉直运算和Kronecker积的定义和性质，可知

$$\mathrm{Vec}(\mathbf{X}') \sim N_{np}(\mathbf{1}_n \otimes \boldsymbol{\mu}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}),$$

其中$\mathbf{1}_n$表示元素均为1的n维向量，$\mathbf{I}_n$表示$n \times n$的单位阵。

- np维向量$\mathrm{Vec}(\mathbf{X}')$的联合密度函数可表示为

$$\begin{aligned} f(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) &= \prod_{i=1}^{n} \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right] \\ &= \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right] \\ &= \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left[-\frac{1}{2}\begin{pmatrix} \boldsymbol{x}_1 - \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{x}_n - \boldsymbol{\mu} \end{pmatrix}' \begin{pmatrix} \boldsymbol{\Sigma} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Sigma} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{x}_1 - \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{x}_n - \boldsymbol{\mu} \end{pmatrix}\right]. \end{aligned}$$

定义3.13：矩阵正态分布

设$\mathbf{X}$为$n \times p$随机矩阵，若$\mathbf{X}$按行拉直后服从下面的分布

$$\mathrm{Vec}(\mathbf{X}') \sim N_{np}(\mathbf{1}_n \otimes \boldsymbol{\mu}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}),$$

则称$\mathbf{X}$服从**矩阵正态分布**，记作

$$\mathbf{X} \sim N_{n\times p}(\mathbf{M}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}),$$

其中

$$\mathbf{M} = \begin{pmatrix} \mu_1 & \cdots & \mu_p \\ \vdots & \ddots & \vdots \\ \mu_1 & \cdots & \mu_p \end{pmatrix}.$$

性质3.4.1

设$\mathbf{X} \sim N_{n\times p}(\mathbf{M}, \mathbf{I}_n \otimes \boldsymbol{\Sigma})$，$\mathbf{A}$为$k\times n$常数矩阵，$\mathbf{B}$为$p\times q$常数矩阵，$\mathbf{D}$为$k\times q$常数矩阵，令$\mathbf{Z}=\mathbf{AXB}+\mathbf{D}$，则

$$\mathbf{Z} \sim N_{k\times q}\Big(\mathbf{AMB}+\mathbf{D}, (\mathbf{AA}') \otimes (\mathbf{B}'\boldsymbol{\Sigma}\mathbf{B})\Big).$$

证明留作作业，课下完成。

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

样本均值向量

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

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

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

样本离差阵

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

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

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

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

样本协方差阵

设$\boldsymbol{x}_i = (x_{i1}, \cdots, x_{ip})'(i = 1, \cdots, n)$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$s_{kk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ik} - \bar{x}_k)^2 (k = 1, \ldots, p)$称为第$k$个样本变量$\boldsymbol{x}_k$的样本方差；$\sqrt{s_{kk}}$称为变量$\boldsymbol{x}_k$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$\widetilde{\mathbf{R}}=(r_{ij})_{p\times p}$为样本相关阵，其中

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

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

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

定义：多元特征函数

随机向量$\boldsymbol{X}=(X_1,\cdots,X_p)'$的特征函数称为多元特征函数，它等于

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t})=E[\exp(\mathrm{i}\boldsymbol{t}'\boldsymbol{X})],$$

其中$\boldsymbol{t}=(t_1,\cdots,t_p)'$，$\mathrm{i}=\sqrt{-1}$。

下面介绍多元特征函数的一些性质：

性质1'：矩

若随机向量$X = (X_1, \cdots, X_p)'$的特征函数为$\varphi_X(t)$，并且$E(X_1^{k_1} \cdots X_p^{k_p})$ 存在，则

$$E(X_1^{k_1} \cdots X_p^{k_p}) = \mathrm{i}^{k_1+\cdots+k_p} \left[\frac{\partial^{k_1+\cdots+k_p} \varphi_X(t)}{\partial t_1^{k_1} \cdots \partial t_p^{k_p}} \right]_{t_1=\cdots=t_p=0}.$$

特别地，
(1) 若X_j的均值$E(X_j)$存在，则

$$E(X_j) = \mathrm{i} \left[\frac{\partial \varphi_X(t_1, \cdots, t_p)}{\partial t_j} \right]_{t_1=\cdots=t_p=0};$$

性质1'(续)：矩

(2) 若X_j的二阶矩$E(X_j^2)$存在，则

$$E(X_j^2) = -\left[\frac{\partial^2 \varphi_{\mathbf{X}}(t_1, \cdots, t_p)}{\partial t_j^2}\right]_{t_1=\cdots=t_p=0};$$

(3) 若X_j和$X_k(j \neq k)$的混合矩$E(X_jX_k)$存在，则

$$E(X_jX_k) = -\left[\frac{\partial^2 \varphi_{\mathbf{X}}(t_1, \cdots, t_p)}{\partial t_j \partial t_k}\right]_{t_1=\cdots=t_p=0}.$$

特征函数*

性质2’：边际分布

若随机向量$\boldsymbol{X}=(X_1,\cdots,X_p)'$的特征函数为$\varphi_{\boldsymbol{X}}(t_1,\cdots,t_p)$，则$\boldsymbol{X}^{(m)}=(X_1,\cdots,X_m)'$的特征函数为$\varphi_{\boldsymbol{X}^{(m)}}(t_1,\cdots,t_m,0,\cdots,0)$，其中$m<p$。

性质3’：唯一性

特征函数与分布函数相互唯一确定。

性质4’：独立性

如果$\boldsymbol{X}=(X_1,\cdots,X_p)'$的分量是相互独立的，则

$$\varphi_{\boldsymbol{X}}(t_1,\cdots,t_p)=\prod_{i=1}^{p}E(\exp(\mathrm{i}t_iX_i)).$$

性质5'：和

假设$\boldsymbol{X}_k \in \mathbb{R}^p$，$\boldsymbol{X}_k$的特征函数为$\varphi_{\boldsymbol{X}_k}(\boldsymbol{t}), k = 1, \cdots, m$。如果$\boldsymbol{X}_1, \cdots, \boldsymbol{X}_m$相互独立，则$\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_m$ 的特征函数为$\varphi(\boldsymbol{t})$ 等于$\boldsymbol{X}_1, \cdots, \boldsymbol{X}_m$的特征函数的乘积，即

$$\varphi(\boldsymbol{t}) = \varphi_{\boldsymbol{X}_1}(\boldsymbol{t}) \cdots \varphi_{\boldsymbol{X}_m}(\boldsymbol{t}).$$

性质6'：密度函数

如果$\boldsymbol{X} = (X_1, \cdots, X_p)'$有密度函数$f(\boldsymbol{x})$和特征函数$\varphi_{\boldsymbol{X}}(\boldsymbol{t})$，则

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(-\mathrm{i}\boldsymbol{t}'\boldsymbol{x}) \varphi_{\boldsymbol{X}}(\boldsymbol{t}) dt_1 \cdots dt_p,$$

其中$\boldsymbol{x} = (x_1, \cdots, x_p)'$和$\boldsymbol{t} = (t_1, \cdots, t_p)'$。

多元统计分析
STATISTICS
DATA
统计与数据科学丛书 4
STATISTICS AND DATA SCIENCE SERIES 4
多元统计分析
李高荣 吴密霞 编著
Multivariate Statistical Analysis
科学出版社

谢谢，请多提宝贵意见！