多元正态分布

张伟平

zwp@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~zwp/

论坛 http://fisher.stat.ustc.edu.cn

简介

1.1 多元随机变量

设 $X_1, X_2, \ldots, X_p$ 为 p 个随机变量, 它们组成的向量 $X = (X_1, X_2, \ldots, X_p)'$ 称为随机向量.

- **联合分布函数**(jcdf)

$$F(x_1, \ldots, x_p) = P(X_1 \le x_1, \ldots, X_p \le x_p)$$

  或者记 $\mathbf{x} = (x_1, \ldots, x_p)'$, 此时

$$F(\mathbf{x}) = P(X \le \mathbf{x})$$

- **联合概率密度函数**(jpdf) 如果存在非负函数 $f(x_1, \ldots, x_p)$, 使得对任意 $x_1, \ldots, x_p$ 有

$$F(x_1, \ldots, x_p) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f(x_1, \ldots, x_p) dx_1 \cdots dx_p$$

  则 $f(x_1, \ldots, x_p)$ 称为 X 的联合概率密度函数.

- X 的 q 个 $(q<p)$ 分量 $X^{(1)}=(X_1,\ldots,X_q)'$ 的分布称为**边际分布**:

$$P(X^{(1)}\le \mathbf{u})=P(X_1\le u_1,\ldots,X_q\le u_q)=F(u_1,\ldots,u_q,+\infty,\ldots,+\infty)$$

边际概率密度函数

$$g(\mathbf{u})=\int_{R^{p-q}} f(\mathbf{u},\mathbf{y})d\mathbf{y}$$

- 若 $X=(X^{(1)'},X^{(2)'})'$ 有概率密度函数 $f(\mathbf{x}_1,\mathbf{x}_2)$, $X^{(2)}$ 有密度函数 $g(\mathbf{u})$, 则 $X^{(1)}$ 在给定 $X^{(2)}=\mathbf{x}_2$ 条件下的**条件密度**为

$$f(\mathbf{x}_1|\mathbf{x}_2)=\frac{f(\mathbf{x}_1,\mathbf{x}_2)}{g(\mathbf{x}_2)}$$

- $X_1,X_2,\ldots,X_p$**相互独立**当且仅当 (F_i 为 X_i 的分布函数)

$$F(x_1,\ldots,x_p)=\prod_{i=1}^{p}F_i(x_i),\forall(x_1,\ldots,x_p)'\in R^p$$

- (**特征函数**), 设 p 元随机向量 $X \sim F(\mathbf{x})$, 则其特征函数定义为

$$f(\mathbf{t}) = Ee^{i\mathbf{t}'X}, \mathbf{t} \in R^p, i = \sqrt{-1}$$

矩:

- 期望 $EX = (EX_1, \ldots, EX_p)'$
- 协方差 $cov(X) = \Big((E(X_i - EX_i)(X_j - EX_j))_{ij}\Big)$
- 若 $X_{p\times 1}, Y_{q\times 1}$ 为随机向量, 则它们的协方差 $cov(X, Y) = \Big((E(X_i - EX_i)(Y_j - EY_j))_{ij}\Big)$
- $Etr(AXB) = tr(A(EX)B)$, $cov(AX) = Acov(X)A'$
- $E(X'AX) = \mu' A\mu + tr(A\Sigma)$, 其中 $\Sigma = cov(X)$
- $cov(AX, BY) = Acov(X, Y)B'$

1.2 多元正态分布密度及其性质

多元正态分布重要性:

- 许多多元统计技术基于多元正态的假设
- 正态分布数学上易于处理, 可以得到许多 “漂亮” 的结果
- 在实际某些问题里总体分布是正态分布的
- 即便总体分布不是正态分布, 许多统计量的分布渐近为正态分布

i.i.d 正态随机变量

设随机变量 $Y_1,\ldots,Y_p\ i.i.d \sim N(0,1)$, 则随机向量 $Y=(Y_1,\ldots,Y_p)'$ 的密度为

$$f(y_1,\ldots,y_p) = \prod_{i=1}^{p}\frac{1}{\sqrt{2\pi}}exp\{-\frac{1}{2}y_i^2\} = (2\pi)^{-p/2}exp\{-\frac{1}{2}\mathbf{y}'\mathbf{y}\}$$

设 $Y\sim N_p(0,I)$, 则有

- $EY=0, cov(Y)=I_p$.
- $E(Y'AY)=tr(A)$.
- 设 $\mathbf{a}$ 为 p 元向量, A 为对称矩阵, 则 $cov(\mathbf{a}'Y, Y'AY)=0$.
- 设 A,B 为对称矩阵, 则 $cov(Y'AY, Y'BY)=2tr(AB)$.

证明. 由于 $Y'AY = \sum_{i,j} a_{ij} Y_i Y_j, Y'BY = \sum_{k,l} b_{kl} Y_k Y_l$, 以及

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

于是

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

注意到

$$\sum\sum_{1\leq i\neq k\leq p} a_{ii}b_{kk} = \sum_{k=1}^{p}\sum_{i=1}^{p} a_{ii}b_{kk} - \sum_{i=1}^{p} a_{ii}b_{ii} = tr(A)tr(B) - \sum_{i=1}^{p} a_{ii}b_{ii}$$

$$\sum\sum_{1\leq i\neq j\leq p} a_{ij}b_{ij} = \sum_{j=1}^{p}\sum_{i=1}^{p} a_{ij}b_{ij} - \sum_{i=1}^{p} a_{ii}b_{ii} = tr(AB) - \sum_{i=1}^{p} a_{ii}b_{ii}$$

从而

$$E(Y'AYY'BY) = 2tr(AB) + tr(A)tr(B).$$

最后

$$cov(Y'AY, Y'BY) = E(Y'AYY'BY) - E(Y'AY)E(Y'BY) = 2tr(AB).$$

$\square$

定理 1. 设 p 元随机向量 $X=\mu+A'Y$, 其中 $\mu\in R^p$, A 为 $p\times p$ 实满秩矩阵, 随机向量 $Y=(Y_1,\ldots,Y_p)'$ 的各分量 $Y_1,\ldots,Y_p$ $i.i.d\sim N(0,1)$, 则 X 有概率密度

$$g(\mathbf{x})=(2\pi)^{-p/2}|\Sigma|^{-1/2}exp\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\}$$

其中 $\Sigma=A'A>0$.

证明. 由于 A 为满秩方阵, 因此变换 $\mathbf{y}\to\mathbf{x}=\mu+A'\mathbf{y}$ 为一一变换, 因此 $\mathbf{y}(\mathbf{x})=(A')^{-1}(\mathbf{x}-\mu)$. 变换的 Jacobian 行列式为

$$J(\mathbf{y}\to\mathbf{x})=|A^{-1}|_+=|A|_+^{-1}=|A'A|^{-1/2}=|\Sigma|^{-1/2}$$

从而由多元函数的密度变换公式有 ξ 的密度函数为

$$\begin{aligned} g(\mathbf{x}) &= f(\mathbf{y}(\mathbf{x}))J(\mathbf{y}\to\mathbf{x}) \\ &= (2\pi)^{-p/2}|\Sigma|^{-1/2}exp\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\} \end{aligned}$$

□

此定理表明多元正态分布密度函数由参数 μ 和 Σ 确定, 因此定义一般的多元正态分布为

称 p 元随机变量 X 服从参数为 μ 和 Σ 的多元正态分布，如果其有概率密度函数

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

Definition

其中 $\mu \in R^p$，Σ 为 p 阶正定矩阵. 此时，记 $X \sim N_p(\mu, \Sigma)$.
常称 $N_p(0, I)$ 为 p 元标准正态分布.

↑Example

二元正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}+\frac{(y-\mu_2)^2}{\sigma_2^2} -2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2}\right]\right\}$$

↓Example

可以看出 $\mu=(\mu_1,\mu_2)'$ 以及

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

多元正态分布的性质: 设随机向量 X 服从多元正态分布 $N(\mu, \Sigma)$, 则

- $EX = \mu$, $cov(X) = \Sigma$
- $(X - \mu)'\Sigma^{-1}(X - \mu) \sim \chi_p^2$
- 协方差矩阵的零元素表明相应的 X 分量相互独立
- X 的任意子集服从 (多元) 正态分布
- 分量的条件分布为 (多元) 正态分布

常数密度轮廓线

从 p 元正态密度函数可以看出, 多元密度函数在

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

上是常数. 因此所有满足上式的 $\mathbf{x}$ 称为常数密度轮廓线, 它是以 μ 为中心, $\pm c\sqrt{\lambda_j}$ $(j = 1, \ldots, p)$ 为轴的椭圆面.

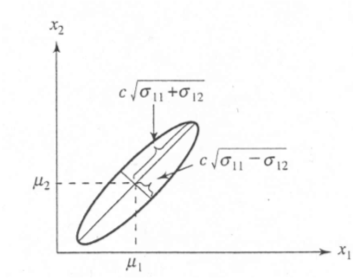
对二元正态 $N(\mu, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{11} \end{pmatrix})$, 求其常数密度轮廓线椭圆.

↑Example
↓Example

解: 由 $|\lambda I - \Sigma| = 0$ 和 $\Sigma\phi = \lambda\phi$, 则易得其特征根和特征向量为

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

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



定理 2. 设 $X \sim N_p(\mu, \Sigma)$, $Z = B'X + \mathbf{d}$, 其中 $B: p \times s$ 列满秩, $\mathbf{d} \in R^s$, 则

$$Z \sim N(B'\mu + \mathbf{d}, B'\Sigma B)$$

证明. 由于 $X \stackrel{d}{=} \mu + \Sigma^{1/2}Y$, 其中 $Y \sim N_p(0, I_p)$, 因此 $Z = \mathbf{d} + B'(\mu + \Sigma^{1/2}Y) = (B'\mu + \mathbf{d}) + (\Sigma^{1/2}B)'Y \sim N(B'\mu + \mathbf{d}, B'\Sigma B)$. □

由此定理可得

定理 3. 设 $X \sim N_p(\mu, \Sigma)$, 则

$$P((X-\mu)'\Sigma^{-1}(X-\mu) \le \chi_p^2(\alpha)) = 1 - \alpha.$$

证明. 由于 $Y = \Sigma^{-1/2}(X-\mu) \sim N_p(0, I)$, 因此

$$(X-\mu)'\Sigma^{-1}(X-\mu) = \sum_{i=1}^{p} Y_i^2 \sim \chi_p^2$$

从而得证. □

定理 4. 多元正态分布 $X \sim N_p(\mu, \Sigma)$ 的特征函数为

$$g(\mathbf{t}) = exp\{i\mathbf{t}'\mu - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\}$$

证明. 由于对标准多元正态分布 $Y \sim N_p(0, I_p)$ 的特征函数为

$$\begin{aligned} g_Y(\mathbf{t}) &= Ee^{i\mathbf{t}'Y} = Eexp\{i\sum_{k=1}^{p} t_k Y_k\} = \prod_{k=1}^{p} Ee^{it_kY_k} \\ &= \prod_{k=1}^{p} e^{-\frac{1}{2}t_k^2} = e^{-\frac{1}{2}\mathbf{t}'\mathbf{t}}. \end{aligned}$$

而 $X \stackrel{d}{=} \mu + \Sigma^{1/2}Y$, 所以

$$Ee^{i\mathbf{t}'X} = Ee^{i\mathbf{t}'(\mu+\Sigma^{1/2}Y)} = e^{i\mathbf{t}'\mu}Ee^{i\mathbf{t}'\Sigma^{1/2}Y} = exp\{i\mathbf{t}'\mu - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\}.$$

□

定理 5. 设 X 为 p 维随机变量, 则 X 服从 p 元正态分布的充分必要条件是它的任意线性组合 $\mathbf{a}'X$(其中 $\mathbf{a}\neq 0$) 都服从一元正态分布.

证明. 充分性显然. 下证必要性. 对任意非零 $\mathbf{a}\in R^p$, 由于 $\mathbf{a}'X$ 服从一元正态分布, 故 X 的二阶矩存在. 记 $EX=\mu, cov(X)=\Sigma$, 则 $E\mathbf{a}'X=\mathbf{a}'\mu, cov(\mathbf{a}'X)=\mathbf{a}'\Sigma\mathbf{a}$. 注意到 $\mathbf{a}'X$ 的特征函数

$$g_{a'X}(\theta)=exp\{i\theta\mathbf{a}'\mu-\frac{1}{2}\theta^2\mathbf{a}'\Sigma\mathbf{a}\},\theta\in R$$

令 $\theta=1$, 有

$$g_{a'X}(1)=exp\{i\mathbf{a}'\mu-\frac{1}{2}\mathbf{a}'\Sigma\mathbf{a}\}$$

视其为 $\mathbf{a}$ 的函数, 恰好是参数为 μ,Σ 的 p 元正态分布的特征函数, 因此得证. □

注 因此有些时候也用此定理作为多元正态分布的定义.

定理 6. 设 $X \sim N_p(\mu, \Sigma)$, $\mathbf{a}$ 为 p 维向量, A 和 B 为两个 p 阶对称矩阵, 则有

(1) $EX'AX = \mu' A\mu + tr(A\Sigma)$.

(2) $cov(\mathbf{a}'X, X'AX) = 2\mathbf{a}'\Sigma A\mu$.

(3) $cov(X'AX, X'BX) = 2tr(A\Sigma B\Sigma) + 4\mu' A\Sigma B\mu$. 特别, $Var(X'AX) = 2tr(A\Sigma A\Sigma) + 4\mu' A\Sigma A\mu$.

证明. (1) 记 $X = \mu + CY$, 其中 $Y \sim N_p(0, I_p)$, $\Sigma = CC'$, 则有

$$\begin{aligned} E(X'AX) &= \mu' A\mu + 2E\mu' ACY + EY'C'ACY \\ &= \mu' A\mu + tr(C'AC) = \mu' A\mu + tr(A\Sigma). \end{aligned}$$

(2) 易知

$$cov(\mathbf{a}'X, X'AX) = cov(\mathbf{a}'\mu + \mathbf{a}'CY, \mu' A\mu + 2\mu' ACY + Y'C'ACY)$$
$$= cov(\mathbf{a}'CY, 2\mu' ACY) + cov(\mathbf{a}'CY, Y'C'ACY)'$$
$$= 2\mathbf{a}'CC'A\mu = 2\mathbf{a}'\Sigma A\mu.$$

(3) 由 (2) 的结论有

$$\begin{aligned}
&cov(X'AX, X'BX) \\
&= E(4\mu' ACYY'C'B\mu) + cov(Y'C'ACY, Y'C'BCY) \\
&= 4\mu' A\Sigma B\mu + 2tr(A\Sigma B\Sigma).
\end{aligned}$$

□

1.3 条件分布

设 $X \sim N_p(\mu, \Sigma), p \geq 2$, 将 X, μ, Σ 分块为

$$X = \begin{pmatrix} X^{(1)}_{q\times 1} \\ X^{(2)}_{(p-q)\times 1} \end{pmatrix}, \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}$$

其中 $\mu^{(1)}$ 为 $q \times 1$ 维, Σ_{11} 为 $q \times q$ 维.

定理 7. 设 $X \sim N_p(\mu, \Sigma), \Sigma > 0$, 则

$$X^{(1)}|X^{(2)} = \mathbf{x}^{(2)} \sim N_q(\mu_{11\cdot 2}, \Sigma_{11\cdot 2})$$

其中 $\mu_{11\cdot 2} = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}^{(2)} - \mu^{(2)})$, $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

证明. 由条件密度定义及分块矩阵的逆易证. $\square$

易知 $\Sigma_{11} \geq \Sigma_{11\cdot 2}$, 即在已知 $X^{(2)}$ 条件下, $X^{(1)}$ 的散布比不知道 $X^{(2)}$ 的情况缩小了.

在二元正态场合 $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix})$, 则 $X_1|X_2 = x_2 \sim N(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}})$.

↑Example

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

↓Example

解: 易得 $X_1|X_2 = x_2 \sim N(175+\frac{40}{8}(x_2-71), 550-\frac{40^2}{8}) = N(-180+5x_2, 350)$. 例如当身高为 $x_2 = 70$ 时候, 体重则服从 $N(170, 350)$.

推论 1. 在上述定理的条件下, $X^{(2)}$ 与 $X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}X^{(2)}$ 相互独立, $X^{(1)}$ 与 $X^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}X^{(1)}$ 相互独立.

由定理7知

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

这表明 $X^{(1)}$ 和 $X^{(2)}$ 之间有回归关系。于是称 $\Sigma_{12}\Sigma_{22}^{-1}$ 为 $X^{(1)}$ 对 $X^{(2)}$ 的**回归系数**, 它的元素用 $\beta_{ij\cdot q+1,\cdots,p}$ 表示; $\Sigma_{11\cdot 2}$ 为**条件协方差**, 它的元素用 $\sigma_{ij\cdot q+1,\cdots,p}$ 表示.

偏相关系数(partial correlation of coefficient)

由于 $X^{(1)}|X^{(2)} = \mathbf{x}^{(2)} \sim N_q(\mu_{11\cdot 2}, \Sigma_{11\cdot 2})$, 故可定义条件相关系数

$$r_{ij\cdot q+1,\cdots,p} = \frac{\sigma_{ij\cdot q+1,\cdots,p}}{(\sigma_{ii\cdot q+1,\cdots,p}\sigma_{jj\cdot q+1,\cdots,p})^{1/2}}$$

为在给定 $X^{(2)} = \mathbf{x}^{(2)}$ 条件下, X_i 与 $X_j (i,j = 1,2,\ldots,q)$ 的**偏相关系数**$(corr(X^{(1)}|X^{(2)}))$. 显然

$$R_{11\cdot 2} = [diag(\Sigma_{11\cdot 2})]^{-1/2}\Sigma_{11\cdot 2}[diag(\Sigma_{11\cdot 2})]^{-1/2}$$

有时候我们感兴趣的是 X_i, X_j 在给定**所有其他**$X_k, k=1,\dots,p$, $k\neq i, k\neq j$ 的相关系数, 也称为**偏相关系数**. 此时, 可以证明在多元正态分布假设下

$$\rho_{ij\backslash\{i,j\}} = -\frac{\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}}$$

其中 σ^{ij} 为 Σ^{-1} 的 (i,j) 元.

事实上, 记 $X^{(1)} = (X_i, X_j)'$, $X^{(2)} = X_{\backslash\{i,j\}}$, $K = \Sigma^{-1} = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}$, 则利用分块矩阵的逆矩阵表达我们有

$$K_{11}^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11\cdot 2}$$

记 $K_{11} = \begin{pmatrix} \sigma^{ii} & \sigma^{ij} \\ \sigma^{ij} & \sigma^{jj} \end{pmatrix}$, 则可以得到

$$\Sigma_{11\cdot 2} = K_{11}^{-1} = \frac{1}{detK_{11}} = \begin{pmatrix} \sigma^{jj} & -\sigma^{ij} \\ -\sigma^{ij} & \sigma^{ii} \end{pmatrix}$$

因此

$$\rho_{ij\backslash\{i,j\}} = \frac{\sigma_{ij\backslash\{i,j\}}}{\sqrt{\sigma_{ii\backslash\{i,j\}}\sigma_{jj\backslash\{i,j\}}}} = -\frac{\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}}$$

复相关系数(multiple correlation of coefficient)

用于度量一个随机变量 (比如 Y) 与一个随机向量之间的相关性 (比如 $X = (X_1, \ldots, X_p)'$), 其定义为

$$R = \left(\frac{\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}}{\sigma_{YY}}\right)^{1/2} = \max_{var(a'X)=1} corr(Y, \mathbf{a}'X)$$

其中 $\Sigma_{YX} = cov((Y, X')')$. R^2 称为总体判定系数.(population coefficient of determination)

由 $Y|X \sim N(EY + \beta'(X - EX), \sigma_{11\cdot 2})$, 其中 $\beta = \Sigma_{XX}^{-1}\Sigma_{XY}$.

易知对任意给定的非零常数向量 $\mathbf{a}$ 有

$$
\begin{aligned}
&Var(Y-\mathbf{a}'X) = Var(Y-\beta'X+(\beta-\mathbf{a})'X) \\
&= Var(Y-\beta'X) + 2cov(Y-\beta'X,(\beta-\mathbf{a})'X) + Var((\beta-\mathbf{a})'X) \\
&= Var(Y-\beta'X) + Var((\beta-\mathbf{a})'X) \geq Var(Y-\beta'X)
\end{aligned}
$$

因此

$$Var(Y)-2cov(Y,\mathbf{a}'X)+Var(\mathbf{a}'X) \geq Var(Y)-2cov(Y,\beta'X)+Var(\beta'X)$$

注意到 $corr(Y,\mathbf{a}'X)$ 具有刻度不变性 (变量乘任意一个正常数不变),
因此不妨设 $Var(\mathbf{a}'X)=Var(\beta'X)$, 利用上式有

$$corr(Y,\mathbf{a}'X) \leq corr(Y,\beta'X)$$

对任意给定的非零常向量 $\mathbf{a}$ 成立.

于是

$$\max_{var(a'X)=1} corr(Y, \mathbf{a}'X) = \max_{var(a'X)=var(\beta'X)} corr(Y, \mathbf{a}'X)$$
$$= corr(Y, \beta'X) = \left(\frac{\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}}{\sigma_{YY}} \right)^{1/2}$$

在有了 (Y, X) 的样本 $(y_i, \mathbf{x}_i)_{i=1}^n$ 后, 易得 R^2 的估计为

$$\hat{R}^2 = \frac{SS_{reg}}{SS_{total}}$$

1.4 二次型的独立性

形如 $X'AX$(A 为对称阵) 的量称为二次型。

定理 8. 设 $X \sim N_p(\mu, \Sigma)$, $\Sigma > 0$, A 为对称阵, C 为一 $p \times q$ 矩阵, 则 $C'X$ 和 $X'AX$ 相互独立的充要条件是 $C'\Sigma A = 0$.

证明. (充分性) 由于 A 为对称阵, 记其秩为 r, 以及 $B = \Sigma^{1/2} A \Sigma^{1/2}$ 和其所有非零特征根组成的对角阵为 Λ_r, 则存在正交矩阵 P, 使得

$$P'BP = \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix}$$

若 $C'\Sigma A = 0$, 则有 $C'\Sigma^{1/2} B = 0 = C'\Sigma^{1/2} PP'B$, 记 $D = C'\Sigma^{1/2} P$, 并将其分块, 则

$$\begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix} = 0 \Longrightarrow D_{11} = 0, D_{21} = 0$$

于是

$$D = \begin{pmatrix} 0 & D_{12} \\ 0 & D_{22} \end{pmatrix} := [0, D_2] \Longrightarrow C' = DP'\Sigma^{-1/2} = [0, D_2]P'\Sigma^{-1/2}$$

注意到 $Y = P'\Sigma^{-1/2}X = [Y_1', Y_2']'$, 其中 Y_1 为 r 维, 则 Y_1 与 Y_2 相互独立, 于是由

$$C'X = [0, D_2]P'\Sigma^{-1/2}X = D_2Y_2$$

$$X'AX = X'\Sigma^{-1/2}B\Sigma^{-1/2}X = Y_1'\Lambda_r Y_1$$

从而 $C'X$ 和 $X'AX$ 相互独立.

必要性由矩母函数可证, 此处略. $\square$

定理 9. 设 $X \sim N_p(\mu, \Sigma), \Sigma > 0$, A, B 为对称阵. 则 $X'AX$ 与 $X'BX$ 相互独立的充要条件为 $A\Sigma B = 0$.

证明. 记 $\Sigma^{1/2}A\Sigma^{1/2} = \tilde{A}$, $\Sigma^{1/2}B\Sigma^{1/2} = \tilde{B}$, $Y = \Sigma^{-1/2}X$, 则 $X'AX = Y'\tilde{A}Y$, $X'BX = Y'\tilde{B}Y$ 且 $Y \sim N_p(\Sigma^{-1/2}\mu, I_p)$. 当 $A\Sigma B = 0$ 时, 有 $\tilde{A}\tilde{B} = \tilde{B}\tilde{A} = 0$, 故可以对 $\tilde{A}$ 和 $\tilde{B}$ 同时进行对角化:

$$P'\tilde{A}P = \Lambda_1, P'\tilde{B}P = \Lambda_2$$

其中 P 为正交阵, Λ_i 为对角阵, $i = 1, 2$. 从而由 $\tilde{A}\tilde{B} = 0$ 知 $\Lambda_1\Lambda_2 = 0$, 因此

$$X'AX = (PY)'\Lambda_1(PY), X'BX = (PY)'\Lambda_2(PY)$$

注意到 $PY \sim N_p(P\Sigma^{-1/2}\mu, I_p)$, 故 PY 个分量相互独立, 而由 $\Lambda_1\Lambda_2 = 0$ 知 $X'AX$ 和 $X'BX$ 依赖 PY 不同的分量, 因此相互独立.

必要性证明略. □

1.5 矩阵正态分布 *

当一维样本 $X_1, \ldots, X_n, i.i.d \sim N(\mu, \sigma^2)$ 时, 样本均值 $\bar{X}$ 和样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 为 μ 和 σ 的估计, 相互独立, 且有分布 $\bar{X} \sim N(\mu, \sigma^2/n)$ 和 $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$.

对多元正态分布场合, $\mathbf{X}_1, \ldots, \mathbf{X}_n$ 为来自多元正态总体 $N_p(\mu, \Sigma)$ 的样本, 则

- μ 和 Σ 如何估计?
- 样本均值向量 $\bar{X}$ 和样本方差 S 又有何种性质?
- 样本均值向量 $\bar{X}$ 和样本方差 S 有何种分布?

记 $X_1,\ldots,X_n$ 为来自多元正态总体 $N_p(\mu,\Sigma)$ 的样本, 令

$$\mathbf{X}=\begin{pmatrix} X_1' \\ X_2' \\ \vdots \\ X_n' \end{pmatrix}=\begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{12} & X_{22} & \cdots & X_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}=[X_{(1)},X_{(2)},\ldots,X_{(p)}]$$

称 $\mathbf{X}$ 为资料矩阵或观测矩阵.

设 $\mathbf{Y}=(Y_{ij})$ 为 $m\times q$ 的随机矩阵，其元素相互独立同分布，均服从 $N(0,1)$ 分布. 矩阵 $M:n\times p$，$A:p\times q$ 和 $B:n\times m$ 为常数矩阵，令 $\mathbf{X}:=M+B\mathbf{Y}A'$，则称 $\mathbf{X}$ 服从矩阵正态分布，其分布为 $N_{n\times p}(M,(BB')\otimes(AA'))$.

Definition

可以看出

- 该定义等同于 $vec(\mathbf{X}') = vec(M') + (B \otimes A)vec(Y')$ 的分布
- $EX = M$
- $cov(vec(\mathbf{X}')) = BB' \otimes AA'$
- $cov(vec(\mathbf{X}) = AA' \otimes BB'$
- 其特征函数为

  $$g(T) = Ee^{i\sum_j \sum_k t_{jk}x_{jk}} = Eetr[iT'X] = etr(iT'M - \frac{1}{2}T'WTV)$$

  其中 $T : n \times p$, $W = BB'$, $V = AA'$, $etr(U) = exp(tr(U))$.
- 矩阵正态分布 $N_{n\times p}(M, W \otimes V), W > 0, V > 0$ 的密度函数为

  $$f(X) = (2\pi)^{-np/2}|W|^{-p/2}|V|^{-n/2}etr[-\frac{1}{2}W^{-1}(X - M)V^{-1}(X - M)']$$

边际分布

定理 10. 设 $\mathbf{X} = N_{n\times p}(M, W\otimes V)$, 记 $W = (\omega_{ij})$, $V = (v_{ij})$,

$$\mathbf{X} = \begin{pmatrix} X_1' \\ X_2' \\ \vdots \\ X_n' \end{pmatrix} = (X_{(1)}, \ldots, X_{(p)}), M = \begin{pmatrix} \mu_1' \\ \mu_2' \\ \vdots \\ \mu_n' \end{pmatrix} = (\mu_{(1)}, \ldots, \mu_{(p)})$$

则

- $X_i \sim N_p(\mu_i, \omega_{ii}V), i = 1, \ldots, n$
- $X_{(j)} \sim N_n(\mu_{(j)}, v_{jj}W), j = 1, \ldots, p$
- $cov(X_i, X_j) = \omega_{ij}V, i, j = 1, \ldots, n$
- $cov(X_{(i)}, X_{(j)}) = v_{ij}W, i, j = 1, \ldots, p.$

推论 2. 设 $\mathbf{X} \sim N_{n\times p}(M, W\otimes V), Z = CXD' + U$, 其中 $C: m\times n$, $D: q\times p$, $U: m\times q$, 则

$$Z \sim N_{m\times q}(CMD' + U, (CWC')\otimes(DVD'))$$

对资料矩阵 $\mathbf{X}$, 因为 $X_i \sim N_p(\mu, \Sigma)$, 故由正态分布定义

$$X_i \stackrel{d}{=} \mu + A'Y_i, \qquad A'A = \Sigma, \qquad Y_i \sim N_p(0, I_p), i = 1, \ldots, n$$

记 $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, 则 $\mathbf{Y} \sim N_{n\times p}(0, I_n \otimes I_p)$. 从而

$$\mathbf{X} \stackrel{d}{=} \mathbf{1}_n\mu' + \mathbf{Y}A \sim N_{n\times p}(\mathbf{1}_n\mu', I_n\otimes\Sigma)$$

从而显然有

- 上述矩阵正态表明资料矩阵的每一行是来自多元正态 $N_p(\mu, \Sigma)$ 的样本, 即 $X_i \sim N_p(\mu, \Sigma), i = 1, \ldots, n$.
- $X_{(j)} \sim N_n(\mu_j\mathbf{1}_n, \sigma_{jj}I_n), j = 1, \ldots, p$.

- $cov(X_i, X_j) = \delta_{ij}\Sigma, i, j = 1, \cdots, p$
- $cov(X_{(i)}, X_{(j)}) = \sigma_{ij} I_n$
- $vec(X') \sim N_{np}(\mathbf{1}_n \otimes \mu, I_n \otimes \Sigma)$

其中 $\mu = (\mu_1, \ldots, \mu_p)'$, $\Sigma = (\sigma_{ij})$, $\delta_{ij} = 1$ 若果 $i = j$; 否则为 0.