Applied Multivariate Statistical Analysis: A Note

2017 年 12 月 28 日

1 Linear Algebra

Theorem 给定正定矩阵 $B_{p\times p}$ 以及标量 $b>0$，我们有

$$\frac{1}{|\Sigma|^b}e^{-\mathrm{tr}(\Sigma^{-1}B)/2}\leq\frac{1}{|B|^b}(2b)^{pb}e^{-pb}$$

等号成立当且仅当 $\Sigma=\frac{1}{2b}B$.

Theorem 设 A,B 均为 J 阶方阵，记其特征值从大到小排列为 $\{\lambda_j(A)\}_{j=1}^J,\{\lambda_j(B)\}_{j=1}^J$，则有

$$\sum_{j=1}^J[\lambda_j(A)-\lambda_j(B)]^2\leq\mathrm{tr}[(A-B)(A-B)^T]$$

Theorem 设 A 为 J 阶方阵，记其特征值从大到小排列为 $\{\lambda_j(A)\}_{j=1}^J$，U 为 $J\times K$ 阶列正交矩阵，则有

$$\lambda_j(U^TAU)\leq\lambda_j(A)$$

等号成立当且仅当 U 的列向量是 A 的前 K 个特征值.

2 Multivariate Normal Distribution

Theorem 设 $X=\begin{pmatrix}X_1\\X_2\end{pmatrix}$ 服从多元正态分布 $N_P(\mu,\Sigma)$，其中 $\mu=\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix},\Sigma=\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&\Sigma_{22}\end{pmatrix}$. 则在给定 $X_2=x_2$ 的条件下，X_1 服从均值为 $\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)$，协方差为 $\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ 的多元正态分布.

Theorem 设 $X\sim N_{n\times p}(M,I_n\otimes\Sigma),A\in\mathbb{R}^{m\times n},B\in\mathbb{R}^{p\times q}$，则 Y 是正态矩阵当且仅当下列条件同时成立:

(i) 存在常数 α 使得 $A1_n=\alpha1_n$ 或者 $B^T\mu=0$；

(ii) 存在常数 β 使得 $AA^T=\beta I_n$ 或者 $B^T\Sigma B=0$.

如果上述条件同时成立，则 $Y\sim N_{m\times q}(\alpha B^T\mu,\beta B^T\Sigma B)$.

Theorem 设 $X \sim N_{n\times p}(M, I_n \otimes \Sigma)$，$C_1, \cdots, C_k$ 是对称矩阵. 如果 $C_r C_s = 0$ 对于一切 $r \neq s$ 均成立，则 $X^T C_1 X, \cdots, X^T C_k X$ 相互独立.

Theorem 设 $X \sim N_{n\times p}(M, I_n \otimes \Sigma)$，$A, B$ 均为 n 阶对称幂等矩阵，则 $X^T AX$ 与 $X^T BX$ 相互独立 $\Leftrightarrow AB = O$.

3 Cochran's Theorem

Theorem 设 x 是一个 n 维向量,且存在一列半正定矩阵 $A_j, j = 1, 2, \cdots, k$ 使得 $x^T x = \sum_{j=1}^k x^T A_j x$. 记 $Q_j = x^T A_j x, r_j = \text{rank}(A_j), S_j = \sum_{i=1}^j r_i, j = 1, 2, \cdots, k$，如果 $\sum_{j=1}^k r_j = n$，则存在正交矩阵 C 和向量 y 使得 $x = Cy$，且有 $Q_j = \sum_{i=S_{j-1}+1}^{S_j} y_i^2$.

Proof 以 $n = 2$ 为例，设 $x^T x = x^T A_1 x + x^T A_2 x$，并设 $A_1 = C\Lambda_1 C^T$，其中 C 是正交矩阵，Λ_1 是对角矩阵. 不妨设 Λ_1 的前 r_1 个对角元不为 0，记 $x = Cy$，则有

$$y^T C^T A_1 C y + y^T C^T A_2 C y = y^T \Lambda_1 y + y^T C^T A_2 C y = \sum_{j=1}^{r_1} \lambda_j y_j^2 + y^T C^T A_2 C y$$

从而

$$y^T C^T A_2 C y = x^T x - \sum_{j=1}^{r_1} \lambda_j y_j^2 = y^T y - \sum_{j=1}^{r_1} \lambda_j y_j^2 = \sum_{j=1}^{r_1} (1 - \lambda_j) y_j^2 + \sum_{j=r_1+1}^{n} y_j^2$$

等号左边是秩为 r_2 的二次型，为了使等号两边的秩相等，必须有 $1 - \lambda_j = 0, j = 1, 2, \cdots, r_1$，从而进一步得到 $C^T A_2 C = I$.

Theorem 设 $X \sim N_n(0, \sigma^2 I)$,且存在一列半正定矩阵 $A_j, j = 1, 2, \cdots, k$ 使得 $X^T X = \sum_{j=1}^k X^T A_j X$. 记 $Q_j = X^T A_j X, r_j = \text{rank}(A_j), S_j = \sum_{i=1}^j r_i, j = 1, 2, \cdots, k$，如果 $\sum_{j=1}^k r_j = n$，则

i. $Q_1, Q_2, \cdots, Q_k$ 相互独立,

ii. $Q_j \sim \sigma^2 \chi^2_{r_j}, j = 1, 2, \cdots, k$,

4 Distributions

4.1 Wishart Distribution

Definition 设 $X_{(i)} \sim N_p(0, \Sigma), i = 1, 2, \cdots, n$ 相互独立，则称随机矩阵

$$W = \sum_{i=1}^n X_{(i)} X_{(i)}^T$$

服从 Wishart 分布，记为 $W \sim W_p(n, \Sigma)$.

Theorem 设 $X_{(i)} \sim N_p(0, \Sigma), i = 1, 2, \cdots, n$ 相互独立，则

$$A = \sum_{i=1}^n (X_{(i)} - \bar{X})(X_{(i)} - \bar{X})^T \sim W_p(n-1, p)$$

Theorem 设 $W_i \sim W_p(n_i, \Sigma), i = 1, 2, \cdots, k$ 相互独立，则

$$\sum_{i=1}^{k} W_i \sim W_p(\sum_{i=1}^{k} n_i, \Sigma)$$

Theorem 设 $W \sim W_p(n, \Sigma)$，C 是 $m \times p$ 矩阵，则

$$CWC^T \sim W_m(n, C\Sigma C^T)$$

4.2 Hotelling's T^2 Distribution

Definition 设 $X \sim N_p(0, \Sigma), W \sim W_p(n, \Sigma)$，且 X 与 W 相互独立，则称统计量

$$T^2 = X^T (\frac{W}{n})^{-1} X$$

服从 n 个自由度的 T^2 分布，记为 $T^2 \sim T^2(p, n)$.

Theorem 设 $X_{(i)} \sim N_p(\mu, \Sigma), i = 1, 2, \cdots, n$ 相互独立,记 $\bar{X} = \frac{\sum_{i=1}^{n} X_{(i)}}{n}, S = \frac{\sum_{i=1}^{n} (X_{(i)} - \bar{X})(X_{(i)} - \bar{X})^T}{n-1}$，则

$$T^2 = n(\bar{X} - \mu) S^{-1} (\bar{X} - \mu)^T \sim T^2(p, n-1)$$

Theorem 设 $T^2 \sim T^2(p.n)$，则

$$\frac{n-p+1}{np} T^2 \sim F_{p, n-p+1}$$

5 Linear Regression

Theorem 设 $Y = Z\beta + \varepsilon$，其中 $\mathbb{E}\varepsilon = 0, \mathrm{Cov}(\varepsilon) = \sigma^2 I$，则对任意 c，估计值 $c^T \hat{\beta}$ 是 $c^T \beta$ 所有线性无偏估计中方差最小的.

Proof 设 $a^T Y$ 是 $c^T \beta$ 的无偏估计，则有 $\mathbb{E}(a^T Y) = a^T Z \beta = c^T \beta$，从而 $(a^T Z - c^T)\beta = 0$ 对一切 β 均成立，因此令 $\beta = Z^T a - c$，我们得到 $a^T Z = c^T$. 再注意到 $c^T \hat{\beta} = c^T (Z^T Z)^{-1} Z^T Y = (a^*)^T Y$，其中 $a^* = Z(Z^T Z)^{-1} c$，注意到 $(a - a^*)^T a^* = (a^T Z - (a^*)^T Z)(Z^T Z)^{-1} c = (c^T - c^T)(Z^T Z)^{-1} c = 0$，于是有

$$\begin{aligned}
\mathrm{Var}(a^T Y) = \mathrm{Var}(a^T \varepsilon) = \sigma^2 a^T a \\
&= \sigma (a - a^* + a^*)^T (a - a^* + a^*) \\
&= \sigma [(a - a^*)^T (a - a^*) + (a^*)^T a^*] \\
&\geq \sigma^2 (a^*)^T a^* = \mathrm{Var}((a^*)^T Y)
\end{aligned}$$