http://staff.ustc.edu.cn/~ynyang/vector

第十讲 多元线性模型

2024.4.8

多元线性回归模型

多元线性回归模型将多元响应变量与自变量以线性形式联系起来，
其求解方式与一元(一个响应)线性回归模型相同，也是最小二乘法。
MANOVA（包括Hotelling T2检验)是多元线性模型的特殊情形。

例3. 同一对象测量有3个相关的响应(response) $\mathbf{y}=\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}$，比如血压、脉搏、心率，假设这些响应与自变量$x$（比如年龄）满足一元线性模型：

$$y_k=a_k+b_kx+\varepsilon_k,\ E\varepsilon_k=0, var(\varepsilon_k)=\sigma_k^2,\quad k=1,2,3$$

3个模型合在一起即是多元线性回归模型(总体模型)：

$\varepsilon_k\sim(0,\sigma_k^2)$

$$\mathbf{y}=\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}=\begin{pmatrix}a_1&b_1\\a_2&b_2\\a_3&b_3\end{pmatrix}\begin{pmatrix}1\\x\end{pmatrix}+\begin{pmatrix}\varepsilon_1\\\varepsilon_2\\\varepsilon_3\end{pmatrix}\triangleq B^\top\mathbf{x}+\boldsymbol{\varepsilon}, B=\begin{pmatrix}a_1&a_2&a_3\\b_1&b_2&b_3\end{pmatrix}$$

其中$\boldsymbol{\varepsilon}\sim(\mathbf{0},\Sigma)$，误差方差$\Sigma$是$3\times3$未知参数矩阵(未必对角）,回归系数$B$是$2\times3$未知参数矩阵。

现假设有n个研究对象，样本 $(\mathbf{y}_i, x_i)$，$i = 1, \dots, n$, 满足前述模型：

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix} = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{pmatrix} \begin{pmatrix} 1 \\ x_i \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{pmatrix} \triangleq B^\top \mathbf{x}_i + \boldsymbol{\varepsilon}_i,$$

按行排列样本：

$$Y = \begin{pmatrix} \mathbf{y}_1^\top \\ \vdots \\ \mathbf{y}_n^\top \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} & y_{13} \\ \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & y_{n3} \end{pmatrix}, \quad X = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, E = \begin{pmatrix} \boldsymbol{\varepsilon}_1^\top \\ \vdots \\ \boldsymbol{\varepsilon}_n^\top \end{pmatrix}$$

则$Y_{n\times 3} = \begin{pmatrix} \mathbf{y}_1^\top \\ \vdots \\ \mathbf{y}_n^\top \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_1^\top \end{pmatrix} B + E = X_{n\times 2} B_{2\times 3} + E_{n\times 3}$

一般地，若每个个体有q个响应，p个自变量，

总体模型： $\mathbf{y}_{q\times 1} = B^\top \mathbf{x}_{p\times 1} + \boldsymbol{\varepsilon}_{q\times 1}$，$\boldsymbol{\varepsilon} \sim (\mathbf{0}, \Sigma)$ 或$\boldsymbol{\varepsilon} \sim N_q(\mathbf{0}, \Sigma)$

样本模型： $Y_{n\times q} = X_{n\times p} B_{p\times q} + E_{n\times q}$

多元线性回归模型(总体版本)

假设响应为$q\times 1$向量$\mathbf{y}$，自变量为$p\times 1$向量$\mathbf{x}$（第一分量为1），（总体版本的）多元线性回归模型假设：

$$\mathbf{y}_{q\times 1} = B^{\top}\mathbf{x}_{p\times 1} + \boldsymbol{\varepsilon}_{q\times 1},\ \boldsymbol{\varepsilon}\sim(\mathbf{0},\Sigma)\ \text{与}\mathbf{x}\text{独立} \qquad (*)$$

其中$\boldsymbol{\varepsilon}$是误差向量，$E(\boldsymbol{\varepsilon})=0, \mathrm{var}(\boldsymbol{\varepsilon})=\Sigma$，$B$是$p\times q$回归系数矩阵。$B^{\top}$的第$k$行，$B$的第$k$列是第$k$个响应关于自变量$\mathbf{x}$的所有回归系数。

上述模型满足Gauss-Markov假设：

- ❑ 回归函数线性：$E(\mathbf{y}|\mathbf{x})=B^{\top}\mathbf{x}$
- ❑ 方差齐性： $\mathrm{var}(\mathbf{y}|\mathbf{x})=\Sigma$，与$\mathbf{x}$无关
- ❑ 外生性：$\boldsymbol{\varepsilon}=\mathbf{y}-B^{\top}\mathbf{x} \perp\!\!\!\perp \mathbf{x}$

注：

- 若$q=1$, (*)是一元线性回归模型。
- 若$p=2, \mathbf{x}=\begin{pmatrix}1\\ x\end{pmatrix}, x=0$或1，则模型(*)是两正态问题。

参数含义：假设$\mathbf{x}, \mathbf{y}$均值为$\mathbf{0}$ (中心化), 模型不含截距,则

$$B = \Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}, \quad \Sigma = \Sigma_{\mathbf{yy}} - \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}} = \Sigma_{\mathbf{yy\bullet x}}$$

（1）由独立性假设$\boldsymbol{\varepsilon} = \mathbf{y} - B^\top\mathbf{x} \perp\!\!\!\perp \mathbf{x}$

$$\Rightarrow 0 = \mathrm{cov}\,(\mathbf{y} - B^\top\mathbf{x}, \mathbf{x}) = \Sigma_{\mathbf{yx}} - B^\top\Sigma_{\mathbf{xx}}$$

$$\Rightarrow B = \Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}$$

（2）模型（*）两边求方差：$\Sigma_{\mathbf{yy}} = B^\top\Sigma_{\mathbf{xx}}B + \Sigma$

$$\Rightarrow \Sigma = \Sigma_{\mathbf{yy}} - B^\top\Sigma_{\mathbf{xx}}B = \Sigma_{\mathbf{yy}} - \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}} = \Sigma_{\mathbf{yy\bullet x}}$$

多元线性回归模型(样本版本)

数据$\mathbf{y}_i, \mathbf{x}_i, i=1,\dots,n$满足多元线性回归模型：

$$\mathbf{y}_i = B^\top \mathbf{x}_i + \boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_i, i=1,\dots,n\ iid\sim(\mathbf{0},\Sigma) \qquad (**)$$

按行排列数据

$$Y_{n\times q} = \begin{pmatrix} \mathbf{y}_1^\top \\ \vdots \\ \mathbf{y}_n^\top \end{pmatrix}, X_{n\times p} = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix}, E = \begin{pmatrix} \boldsymbol{\varepsilon}_1^\top \\ \vdots \\ \boldsymbol{\varepsilon}_n^\top \end{pmatrix}$$

则我们有多元线性回归模型

$$Y_{n\times q} = X_{n\times p} B_{p\times q} + E_{n\times q}, E\text{的各行iid}\sim(\mathbf{0},\Sigma), \Sigma = (\sigma_{ij})$$

记Y, B, E的第k列为$\mathbf{y}_{(k)}$，$\boldsymbol{\beta}_{(k)}$，$\boldsymbol{\varepsilon}_{(k)}$，即

$$Y = (\mathbf{y}_{(1)},\dots,\mathbf{y}_{(q)}),\ B = (\boldsymbol{\beta}_{(1)},\dots,\boldsymbol{\beta}_{(q)}), E = (\boldsymbol{\varepsilon}_{(1)},\dots,\boldsymbol{\varepsilon}_{(q)}),$$

第k个响应的一元模型：

$$\mathbf{y}_{(k)} = X\boldsymbol{\beta}_{(k)} + \boldsymbol{\varepsilon}_{(k)},\ \boldsymbol{\varepsilon}_{(k)}\sim N_n(\mathbf{0},\sigma_{kk}I_n),\ k=1,\dots,q$$

各个一元模型之间是相关的：$cov(\boldsymbol{\varepsilon}_{(k)}, \boldsymbol{\varepsilon}_{(j)}) = \sigma_{kj}I_n$。通常可以把多元模型拆分成$q$个一元模型（参见定理1注1）。

最小二乘法

Frobenius模

对任何矩阵$A, B \in R^{n\times q}$, 定义内积

$$\langle A, B\rangle = tr(A^\top B) = \sum_{i,j} a_{ij} b_{ij}$$

矩阵A的Frobenius/欧氏模

$$\|A\| = \sqrt{\langle A, A\rangle} = \sqrt{tr(A^\top A)} = \sqrt{\sum_{i,j} a_{ij}^2}$$

记A的各行、列$A = (\mathbf{a}_1, \dots, \mathbf{a}_n)^\top = (\mathbf{a}_{(1)}, \dots, \mathbf{a}_{(q)})$,

$$\|A\|^2 = \|\mathbf{a}_1\|^2 + \cdots + \|\mathbf{a}_n\|^2 = \|\mathbf{a}_{(1)}\|^2 + \cdots + \|\mathbf{a}_{(q)}\|^2$$

最小二乘法

模型$Y = XB + E$ 的误差平方和

$$\|E\|^2 = \|Y - XB\|^2 = tr\big((Y - XB)^\top (Y - XB)\big)$$

最小二乘法(LS)极小化$\|E\|^2$。

按列排列

$$Y=\left(\mathbf{y}_{(1)},\ldots,\mathbf{y}_{(q)}\right), B=\left(\boldsymbol{\beta}_{(1)},\ldots,\boldsymbol{\beta}_{(q)}\right), E=(\boldsymbol{\varepsilon}_{(1)},\ldots,\boldsymbol{\varepsilon}_{(q)}),$$

则 $\|E\|^2=\|Y-XB\|^2=\sum_{k=1}^{q}\left\|\boldsymbol{\varepsilon}_{(k)}\right\|^2=\sum_{k=1}^{q}\left\|\mathbf{y}_{(k)}-X\boldsymbol{\beta}_{(k)}\right\|^2$，
为了极小化$\|E\|^2$，只需极小化加和中的每一项。

对于第k个响应，有一元线性模型：

$$\mathbf{y}_{(k)}=X\boldsymbol{\beta}_{(k)}+\boldsymbol{\varepsilon}_{(k)},\quad \boldsymbol{\varepsilon}_{(k)}\sim N_n(\mathbf{0},\sigma_{kk}I_n),\quad k=1,\ldots,q$$

一元回归的最小二乘法(LS)：

$$\min_{\boldsymbol{\beta}_{(k)}\in R^p}\left\|\boldsymbol{\varepsilon}_{(k)}\right\|^2=\min_{\boldsymbol{\beta}_{(k)}\in R^p}\left\|\mathbf{y}_{(k)}-X\boldsymbol{\beta}_{(k)}\right\|^2=\min_{\mathbf{u}_{(k)}\in C(X)}\left\|\mathbf{y}_{(k)}-\mathbf{u}_{(k)}\right\|^2$$

我们已知：$\mathbf{u}_{(k)}=\hat{\mathbf{y}}_{(k)}$ ($\mathbf{y}_{(k)}$在X的列张成的空间$C(X)$上的投影) 时，误差平方和达到最小, 其中

$$\hat{\mathbf{y}}_{(k)}=P_X\mathbf{y}_{(k)}=X(X^\top X)^{-1}X^\top\mathbf{y}_{(k)}=X\widehat{\boldsymbol{\beta}}_{(k)},$$

投影满足正则条件

$$X^\top\left(\mathbf{y}_{(k)}-\hat{\mathbf{y}}_{(k)}\right)=X^\top\left(\mathbf{y}_{(k)}-X\widehat{\boldsymbol{\beta}}_{(k)}\right)=0$$

此时$\boldsymbol{\beta}_{(k)}$的最优估计，即LS估计$\widehat{\boldsymbol{\beta}}_{(k)}=(X^\top X)^{-1}X^\top\mathbf{y}_{(k)}$,

残差 $\mathbf{e}_{(k)}=\mathbf{y}_{(k)}-\hat{\mathbf{y}}_{(k)}=\mathbf{y}_{(k)}-X\widehat{\boldsymbol{\beta}}_{(k)}$.

因此B的LS估计

$$\hat{B} = (\hat{\boldsymbol{\beta}}_{(1)}, \dots, \hat{\boldsymbol{\beta}}_{(q)}) = (X^\top X)^{-1} X^\top (\mathbf{y}_{(1)}, \dots, \mathbf{y}_{(q)}) = (X^\top X)^{-1} X^\top Y$$

最优逼近(拟合、投影): $\hat{Y} = (\hat{\mathbf{y}}_{(1)}, \dots, \hat{\mathbf{y}}_{(q)}) = X\hat{B} = X(X^\top X)^{-1} X^\top Y$

满足正则方程: $X^\top (Y - X\hat{B}) = 0$

残差$\hat{E} = Y - X\hat{B} = (\boldsymbol{e}_{(1)}, \dots, \boldsymbol{e}_{(q)})$

下述定理说明LS估计使得$\min_B \operatorname{tr}(E^\top E)$、$\min_B \operatorname{tr}(\Sigma^{-1} E^\top E)$都达到最小。

最小二乘估计

定理1. 线性模型$Y = XB + E, E \sim (0, \Sigma)$,不假设正态误差，最小二乘法

$$\min_B \operatorname{tr}(E^\top E) = \min_B \operatorname{tr}((Y - XB)^\top (Y - XB))$$

$$\Leftrightarrow \min_B \operatorname{tr}(\Sigma^{-1} E^\top E) = \min_B \operatorname{tr}(\Sigma^{-1} (Y - XB)^\top (Y - XB))$$

得到LS估计： $\hat{B} = (X^\top X)^{-1} X^\top Y,$

Σ的LS估计通常定义为 $\hat{\Sigma} = \dfrac{1}{n-p}(Y - X\hat{B})^\top (Y - X\hat{B}) = \dfrac{1}{n-p} Y^\top (\mathrm{I}_n - P_X) Y$。

定理1的证明：

记$\hat{Y} = P_X Y = X(X^\top X)^{-1} X^\top Y = X\hat{B}$为$Y$在$X$列张成的空间$C(X)$中的投影，则$Y - X\hat{B} = (I_n - P_X)Y, (I_n - P_X)X = 0,$所以

$$\begin{aligned}
\mathrm{tr}(Y - XB)\Sigma^{-1}(Y - XB)^\top &= \mathrm{tr}(Y - X\hat{B} + X\hat{B} - XB)\Sigma^{-1}(Y - X\hat{B} + X\hat{B} - XB)^\top \\
&= \mathrm{tr}(Y - X\hat{B})\Sigma^{-1}(Y - X\hat{B})^\top + \mathrm{tr}(X\hat{B} - XB)\Sigma^{-1}(X\hat{B} - XB)^\top \\
&\geq \mathrm{tr}(Y - X\hat{B})\Sigma^{-1}(Y - X\hat{B})^\top,
\end{aligned}$$

其中交叉项$\mathrm{tr}(Y - X\hat{B})\Sigma^{-1}(X\hat{B} - XB)^\top = \mathrm{tr}(I_n - P_X)Y\Sigma^{-1}(X\hat{B} - XB)^\top$

$= trX^\top(I_n - P_X)Y\Sigma^{-1}(\hat{B} - B)^\top = 0.$

当上面没有Σ^{-1}时也成立，即$\mathrm{tr}(Y - XB)(Y - XB)^\top \geq \mathrm{tr}(Y - X\hat{B})(Y - X\hat{B})^\top$，

所以$B = \hat{B}$时，$\mathrm{tr}(Y - XB)\Sigma^{-1}(Y - XB)^\top$或$\mathrm{tr}(Y - XB)(Y - XB)^\top$最小。

定理2. 多元线性(回归)模型

$$Y_{n\times q}=X_{n\times p}B_{p\times q}+E_{n\times q}，\ E\text{的各行iid}\sim N_q(\mathbf{0},\Sigma),\text{与}X\text{独立，}$$

B和Σ的极大似然估计为$\hat{B}=(X^{\top}X)^{-1}X^{\top}Y,\quad \tilde{\Sigma}=\frac{1}{n}(Y-X\hat{B})^{\top}(Y-X\hat{B})$.

注1：回归系数的LS估计与正态下的MLE相同，但方差估计略有差异。

<u>定理2的证明：</u>$\mathbf{x}_i$给定时，$\mathbf{y}_i\sim N_q(B^{\top}\mathbf{x}_i,\Sigma)$，所以似然函数为

$$L(B,\Sigma)=\prod_{i=1}^{n}\frac{1}{(2\pi)^{q/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}\left(\mathbf{y}_i-B^{\top}\mathbf{x}_i\right)^{\top}\Sigma^{-1}\left(\mathbf{y}_i-B^{\top}\mathbf{x}_i\right)\right)$$

$$=\frac{1}{(2\pi)^{nq/2}|\Sigma|^{n/2}}\exp\left(-\frac{1}{2}\mathrm{tr}(Y-XB)\Sigma^{-1}(Y-XB)^{\top}\right)$$

$$\log L(B,\Sigma)=-n/2\log|\Sigma|-\frac{1}{2}\mathrm{tr}(Y-XB)\Sigma^{-1}(Y-XB)^{\top}\Rightarrow B\text{的MLE与LS估计相同。}$$

$$\mathrm{logL}(\widehat{\boldsymbol{\beta}},\Sigma)=c+\frac{n}{2}\log|\Sigma|^{-1}-\frac{1}{2}\mathrm{tr}(Y-X\hat{B})^{\top}(Y-X\hat{B})\Sigma^{-1}$$

$$=c+\frac{n}{2}\log|\Omega|-\frac{1}{2}\mathrm{tr}A\Omega,\quad（\text{令}\Omega=\Sigma^{-1},\ A=(Y-X\hat{B})^{\top}(Y-X\hat{B})）$$

$$\Rightarrow\tilde{\Sigma}=\tilde{\Omega}^{-1}=\frac{1}{n}A.$$

综上，为了求解模型$Y = XB + E$ 中的回归系数矩阵B, 我们对每个响应，即$Y = \left(\mathbf{y}_{(1)}, \dots, \mathbf{y}_{(q)}\right)$， $B = \left(\boldsymbol{\beta}_{(1)}, \dots, \boldsymbol{\beta}_{(q)}\right)$, $E = \left(\boldsymbol{\varepsilon}_{(1)}, \dots, \boldsymbol{\varepsilon}_{(q)}\right)$的每一列，拟合一元线性模型：

$$\mathbf{y}_{(k)} = X\boldsymbol{\beta}_{(k)} + \boldsymbol{\varepsilon}_{(k)}，\ \boldsymbol{\varepsilon}_{(k)} \sim N_n(\mathbf{0}, \sigma_{kk} I_n)，\ k = 1, \dots, q$$

得到LS估计和残差

$$\hat{\boldsymbol{\beta}}_{(k)} = (X^\top X)^{-1} X^\top \mathbf{y}_{(k)}，\ \mathbf{e}_{(k)} = \mathbf{y}_{(k)} - X\hat{\boldsymbol{\beta}}_{(k)}$$

$\Rightarrow$

- 所有LS估计、所有残差合在一起即得到B的LS估计、残差:

$$\hat{B} = \left(\hat{\boldsymbol{\beta}}_{(1)}, \dots, \hat{\boldsymbol{\beta}}_{(q)}\right) = (X^\top X)^{-1} X^\top Y,$$

$$\hat{E} = \left(\mathbf{e}_{(1)}, \dots, \mathbf{e}_{(q)}\right) = Y - X\hat{B}$$

- $\mathbf{e}_{(i)}, \mathbf{e}_{(j)}$的样本协方差估计$\sigma_{ij}$:

$$\hat{\sigma}_{ij} = \mathbf{e}_{(i)}^\top \mathbf{e}_{(j)} / (n - p)$$

Σ的LS估计

$$\hat{\Sigma} = \hat{E}^\top \hat{E} / (n - p) = \left(\mathbf{e}_{(i)}^\top \mathbf{e}_{(j)}\right) / (n - p)$$

定理3(课本Johnson & Wichern, Result 7.10). 假设正态线性模型，

$$Y_{n\times q} = X_{n\times p}B_{p\times q} + E_{n\times q}，E\text{的各行iid} \sim N_q(\mathbf{0},\Sigma),\text{与}X\text{独立},$$

$\hat{B} = (X^\top X)^{-1}X^\top Y$为LS估计，其第$k$列为$\hat{\boldsymbol{\beta}}_{(k)}$，则在给定$X$的条件下

(1)矩阵$\hat{B}$的所有元素联合服从正态分布

$$E(\hat{B}) = B, \text{var}(\hat{\boldsymbol{\beta}}_{(k)}) = \sigma_{kk}(X^\top X)^{-1}, \text{cov}(\hat{\boldsymbol{\beta}}_{(k)}, \hat{\boldsymbol{\beta}}_{(j)}) = \sigma_{kj}(X^\top X)^{-1}$$

(2)$(n-p)\hat{\Sigma} \sim W_q(n-p,\Sigma)$，与$\hat{B}$独立。

$vec(\hat{\mathrm{B}}) \sim N_{mp}(vec(B), \Sigma \otimes (X^\top X)^{-1})$

证：（1）$\hat{B} = (X^\top X)^{-1}X^\top Y$，$\hat{\boldsymbol{\beta}}_{(k)} = (X^\top X)^{-1}X^\top \mathbf{y}_{(k)}$，

$$\Rightarrow E(\hat{B} \mid X) = (X^\top X)^{-1}X^\top E(Y \mid X) = (X^\top X)^{-1}X^\top XB = B,$$

$$\text{cov}(\hat{\boldsymbol{\beta}}_{(k)}, \hat{\boldsymbol{\beta}}_{(j)} \mid X) = (X^\top X)^{-1}X^\top \text{cov}(\mathbf{y}_{(k)}, \mathbf{y}_{(j)})X(X^\top X)^{-1} = \sigma_{kj}(X^\top X)^{-1}.$$

(2) 由$Y - X\hat{B} = (I_n - X(X^\top X)^{-1}X^\top)Y = (I_n - X(X^\top X)^{-1}X^\top)(XB + E)$

$= (I_n - P_{\mathrm{X}})E$，$\Rightarrow (n-p)\hat{\Sigma} = (Y - X\hat{B})^\top (Y - X\hat{B}) = E^\top (I_n - P_{\mathrm{X}})E$。

因为$I_n - P_{\mathrm{X}}$对称幂等，$tr(I_n - P_{\mathrm{X}}) = n - p$，由Cochran定理，

$$(n-p)\hat{\Sigma} \sim W_q(n-p,\Sigma)，\text{且与}\hat{B}\text{独立。}$$

假设检验

定理4(课本Johnson & Wichern, Result7.11)假设正态线性模型

$$Y_{n\times q} = X_{n\times p}B_{p\times q} + E_{n\times q},\quad E\text{的各行iid} \sim N_q(\mathbf{0},\Sigma),$$

划分$X=(X_1,X_2)$, X_1为$n\times(k+1)$, X_2为$n\times(p-k)$, 划分$B=\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}\begin{matrix}(k+1)\times q \\ (p-k)\times q\end{matrix}$,

模型为$Y = X_1B_1 + X_2B_2 + E$，原假设$H_0: B_2=\mathbf{0}$的似然比检验为

$$-2\log(\Lambda) = -n\log\left(|\hat{\Sigma}|/|\hat{\Sigma}_0|\right) \xrightarrow[\mathrm{H}_0]{d} \chi^2_{q(p-k)},\quad n\to\infty.$$

其中$\hat{\Sigma} = \frac{1}{n}Y^\top(I_n - P_X)Y$，$\hat{\Sigma}_0 = \frac{1}{n}Y^\top(I_n - P_{X_1})Y \ge \hat{\Sigma}$。

注:$|\hat{\Sigma}|/|\hat{\Sigma}_0| = |\hat{\Sigma}|/|\hat{\Sigma}+(\hat{\Sigma}_0-\hat{\Sigma})| = |W|/|W+B|$

证：全模型下似然函数

$$L(B,\Sigma) = \frac{1}{(2\pi)^{nq/2}|\Sigma|^{n/2}}\exp\left(-\frac{1}{2}\mathrm{tr}(Y-XB)\Sigma^{-1}(Y-XB)^\top\right)$$

$$\hat{B} = (X^\top X)^{-1}X^\top Y,\ \hat{\Sigma} = \frac{1}{n}(Y-X\hat{B})^\top(Y-X\hat{B}) = \frac{1}{n}Y^\top(I_n-P_X)Y$$

最大似然$L(\hat{B},\hat{\Sigma}) = \dfrac{1}{(2\pi)^{nq/2}|\hat{\Sigma}|^{n/2}}\exp\left(-\dfrac{nq}{2}\right).$

原假设下$Y = X_1 B_1 + E$，E各行$iid \sim N_q(\mathbf{0}, \Sigma)$,

似然函数$L(B_1, \Sigma) = \dfrac{1}{(2\pi)^{nq/2} |\Sigma|^{n/2}} \exp\left(-\dfrac{1}{2} \mathrm{tr}(Y - X_1 B_1)\Sigma^{-1}(Y - X_1 B_1)^{\top} \right)$

$\widetilde{B}_1 = (X_1{}^{\top} X_1)^{-1} X_1{}^{\top} Y, \hat{\Sigma}_0 = \dfrac{1}{n}(Y - X_1 \widetilde{B}_1)^{\top}(Y - X_1 \widetilde{B}_1) = \dfrac{1}{n} Y^{\top}(I_n - P_{X_1})Y$

最大似然$L(\widetilde{B}_1, \hat{\Sigma}_0) = \dfrac{1}{(2\pi)^{nq/2} |\hat{\Sigma}_0|^{n/2}} \exp\left(-\dfrac{nq}{2} \right)$。

似然比:

$$\Lambda = \frac{\max L(B_1, \Sigma)}{\max L(B, \Sigma)} = \frac{L(\widetilde{B}_1, \widehat{\Sigma}_0)}{L(\widetilde{B}_1, \widehat{\Sigma})} = \left(\frac{|\widehat{\Sigma}|}{|\widehat{\Sigma}_0|} \right)^{n/2}, \text{Wilks' lambda } \Lambda^* = \Lambda^{2/n} = \frac{|\widehat{\Sigma}|}{|\widehat{\Sigma}_0|},$$

由上一讲Wilks定理，H_0成立时，$-n \log\left(\dfrac{|\widehat{\Sigma}|}{|\widehat{\Sigma}_0|} \right) \xrightarrow{d} \chi^2_{q(p-k)}$，$n \to \infty$.

至此，我们介绍了古典多元统计分析的基本内容，其中较少涉及的Jacobian计算是多元分析和多变量微积分计算中的重要工具，下面简单介绍外积微分形式及其在Jacobian计算中的应用。

附录：外积微分形式与Jacobian

我们的目标是从$\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ iid $\sim N_m(0, \Sigma) \Leftrightarrow Z \sim N_{nm}(0, I_n \otimes \Sigma)$ 的概率密度，应用变量变换求$W = Z^\top Z$ 的概率密度。

为了求变换的Jacobian，一个比求变换的导数更简单的方法是利用外积微分形式(exterior product, exterior differential forms)

Exterior products and exterior differential forms were given a systematic treatment by Cartan (1922) in his theory of integral invariants. Since then they have found wide use in differential geometry and mathematical physics; see, for example, Sternberg (1964), Cartan (1967), and Flanders (1963).

Definition 2.1.2 can be extended to define exterior differential forms on differentiable and analytic manifolds and, under certain conditions, these in turn can be used to construct invariant measures on such manifolds. Details of this construction can be found in James (1954) for manifolds of particular interest in multivariate analysis. We will not go further into the formal

本附录材料来源：

Muirhead (1982) Aspects of multivariate statistical theory. Wiley. Chapter 2

外积微分

考虑多重积分

$$I = \int_{A\subset R^m} f(x_1, \dots, x_m) dx_1 \dots dx_m \qquad (*)$$

考虑变量1:1变换$\mathbf{x} \to \mathbf{y}$:

$$x_1 = x_1(y_1, .., y_m)$$
$$\vdots$$
$$x_m = x_m(y_1, .., y_m)$$

变换的Jacobian

$$J(\mathbf{x} \to \mathbf{y}) = det\begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_m} \\ \vdots & & \vdots \\ \frac{\partial x_m}{\partial y_1} & \cdots & \frac{\partial x_m}{\partial y_m} \end{pmatrix}$$

则(*)变为

$$I = \int_{A'} f(\mathbf{x}(\mathbf{y}))|J(\mathbf{x} \to \mathbf{y})| dy_1 \dots dy_n$$

例如m=2时，

$$I_2 = \int_{A'} f(\mathbf{x}(\mathbf{y})) \left|\frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1}\right| dy_1 dy_2.$$

另外一方面，我们可将如下线性微分形式代入(*),

$$dx_i = \frac{\partial x_i}{\partial y_1} dy_1 + \cdots + \frac{\partial x_i}{\partial y_n} dy_n$$

仅考虑$m=2$情形：

$$
\begin{aligned}
I_2 &= \int_{A'} f(\mathbf{x}(\mathbf{y}))\left(\frac{\partial x_1}{\partial y_1}dy_1+\frac{\partial x_1}{\partial y_2}dy_2\right)\left(\frac{\partial x_2}{\partial y_1}dy_1+\frac{\partial x_2}{\partial y_2}dy_2\right)\\
&= \int_{A'} f(\mathbf{x}(\mathbf{y}))\left(\frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_1}dy_1dy_1+\frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2}dy_1dy_2+\frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1}dy_2dy_1+\frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_2}dy_2dy_2\right)\\
&\stackrel{?}{=} \int_{A'} f(\mathbf{x}(\mathbf{y}))\left(\frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2}-\frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1}\right)dy_1dy_2
\end{aligned}
$$

为使最后一式成立，只需定义微分的乘积（外积）满足下述条件

$$dy_1dy_2=-dy_2dy_1\,,dy_1dy_1=dy_2dy_2=0\,，$$

为了与一般乘积区分，我们用符号$\wedge$ (wedge)表示外积。

$$I_2=\int_{A'} f(\mathbf{x}(\mathbf{y}))\left(\frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2}-\frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1}\right)dy_1\wedge dy_2$$

$d\mathbf{x}=(dx_1,\dots,dx_m)^\top$是$m\times 1$微分向量,微分外积(exterior product,wedge product)

$$(d\mathbf{x})=\wedge_{i=1}^m dx_i$$

满足$dx_i\wedge dx_j=-dx_j\wedge dx_i$

定理A1. 若$d\mathbf{y}=(dy_1,\dots,dy_m)^\top$是$m\times 1$微分向量， $B_{m\times m}$可逆， $d\mathbf{x}=Bd\mathbf{y}$, 则

$$\wedge_{i=1}^m dx_i=\det(B)\wedge_{i=1}^m dy_i$$

证明：$\wedge_{i=1}^m dx_i=p(B)\wedge_{i=1}^m dy_i, p(B)$ 是B的元素的多项式。

- $p(B)$关于B的每一行线性
- 改变dx_i，dx_j的次序⇔互换B的i，j行， $p(B)$改变符号
- $p(I_m)=1$

以上是行列式的定义， 故$p(B)=\det(B)$.

第一讲引理1：球坐标变换$\mathbf{x} \in R^n \to (r, \theta_1, \dots, \theta_{n-1})$，

$$x_1 = r\cos(\theta_1) \qquad r \ge 0, 0 \le \theta_1, \dots, \theta_{n-2} \le \pi, 0 \le \theta_{n-1} < 2\pi$$

$$\dots$$

$$x_{n-1} = r\sin(\theta_1)\cdots\sin(\theta_{n-2})\ \cos(\theta_{n-1})$$

$$x_n \quad = r\sin(\theta_1)\cdots\sin(\theta_{n-2})\ \sin(\theta_{n-1})$$

Jacobian：$J(\mathbf{x} \to (r, \boldsymbol{\theta})) = r^{n-1}\sin^{n-2}(\theta_1)\sin^{n-3}(\theta_2)\cdots\sin(\theta_{n-2})$.

证明：直接求导计算得到下三角矩阵（参见Bilodeau P32）.但我们采用外积方法：

$$x_n^2 = r^2\sin^2(\theta_1)\cdots\sin^2(\theta_{n-2})\ \sin^2(\theta_{n-1})$$

$$x_n^2 + x_{n-1}^2 = r^2\sin^2(\theta_1)\cdots\sin^2(\theta_{n-2})$$

$$\dots$$

$$x_n^2 + \cdots + x_1^2 = r^2$$

第一式求微分

$$2x_n dx_n = 2r^2\sin^2(\theta_1)\cdots\sin^2(\theta_{n-2})\ sin(\theta_{n-1})\, cos(\theta_{n-1})\mathrm{d}\theta_{n-1} + dr, d\theta_1, \dots, d\theta_{n-2}\, terms$$

第二式求微分

$$2x_n dx_n + 2x_{n-1}dx_{n-1} = 2r^2\sin^2(\theta_1)\cdots sin(\theta_{n-2})\, cos(\theta_{n-2})\,\mathrm{d}\theta_{n-2} + dr, d\theta_1, \dots, d\theta_{n-3}\, terms$$

最后一式求微分

$$2x_n dx_n + \cdots + 2x_1 dx_1 = 2rdr$$

两边同时求外积

$$2^n x_1\cdots x_n \wedge_{i=1}^{n} dx_i = 2^n r^{2n-1}\sin^{2n-3}(\theta_1)\sin^{2n-5}(\theta_2)\cdots\sin(\theta_{n-1})\cos(\theta_1)\cdots\cos(\theta_{n-1})\ dr \wedge_{i=1}^{n-1} d\theta_i$$

$$= 2^n x_1\cdots x_n r^{n-1}\sin^{n-2}(\theta_1)\sin^{n-3}(\theta_2)\cdots\sin(\theta_{n-2})\ dr \wedge_{i=1}^{n-1} d\theta_i$$

$$\Rightarrow \wedge_{i=1}^{n} dx_i = r^{n-1}\sin^{n-2}(\theta_1)\sin^{n-3}(\theta_2)\cdots\sin(\theta_{n-2})\ dr \wedge_{i=1}^{n-1} d\theta_i$$

矩阵的外积微分

对任何矩阵$X=(x_{ij})$，其微分$dX=(dx_{ij})$，dX的外积微分形式记作

$$(dX)=\wedge_{i,j}dx_{ij}$$

当X是$n\times n$对称矩阵，$(dX)=\wedge_{i\le j}dx_{ij}$

容易验证：$d(XY)=X.dY+dX.Y$

常用变换的Jacobian

定理A2：若$X=BY$, 其中X，Y是$n\times m$矩阵，B是$n\times n$可逆矩阵，则$J(X\to Y)=|B|^m$

证明：$dX=BdY$，$dX=(d\mathbf{x}_1,\dots,d\mathbf{x}_m)$，$dY=(d\mathbf{y}_1,\dots,d\mathbf{y}_m)$, $\mathbf{x}_j=B\mathbf{y}_j$，$d\mathbf{x}_j=Bd\mathbf{y}_j$, 由定理A1，$\wedge_{i=1}^n dx_{ij}=|B|\wedge_{i=1}^n dy_{ij}$，所以

$$(dX)=\wedge_{j=1}^m\wedge_{i=1}^n dx_{ij}=\wedge_{j=1}^m\left(|B|\wedge_{i=1}^n dy_{ij}\right)$$
$$=|B|^m\wedge_{j=1}^m\wedge_{i=1}^n dy_{ij}=|B|^m(dY)$$

推论：若$X=BYC$, 其中X，Y是$n\times m$矩阵，B,C分别是$n\times n$，$m\times m$可逆矩阵，则$J(X\to Y)=|B|^m|C|^n$

定理A3：若$X = BYB^\top$,其中X，Y是$m \times m$对称矩阵，B 是$m \times m$可逆矩阵，则$J(X \to Y) = |B|^{m+1}$

证明：$X = BYB^\top$，$dX = BdYB^\top$，则

$$(dX) = (BdYB^\top) = p(B)(dY)$$

其中$p(B)$是B的元素的多项式。对任何$m \times m$可逆矩阵$A, B,$

$$p(AB)(dY) = (ABdY(AB)^\top) = (ABdYB^\top A^\top) = p(A)(BdYB^\top)$$
$$= p(A)p(B)(dY)$$

所以 $p(AB) = p(A)p(B)$, 则一定存在某个整数 k，使得$p(B) = |B|^k$

取$B = bI_m$, 则$X = BYB^\top = b^2 Y$,

$$(dX) = (b^2 dY) = (b^2)^{m(m+1)/2}(dY) = b^{m(m+1)}(dY) \Rightarrow k = m + 1$$

定理A4：若$X = Y^{-1}$, 其中Y是$m \times m$对称矩阵，则$J(X \to Y) = |Y|^{-(m+1)}$

证明：$YX = I_m$，则$YdX + dYX = 0 \Rightarrow dY = -X^{-1}dXX^{-1}$

$\Rightarrow \ (dY) = (-X^{-1}dXX^{-1}) = |X|^{-(m+1)}(dX)$

与Wishart有关的Jacobian

定理A5：若$A_{m\times m}>0, A=T^{\top}T$, 其中$T$为对角元为正数的上三角矩阵，则$J(A\to T)=2^m\prod_{i=1}^m t_{ii}^{m-i+1}$

证明：

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{12} & a_{22} & \cdots & a_{2m} \\ \vdots & & & \\ a_{1m} & a_{2m} & \cdots & a_{mm} \end{pmatrix} = \begin{pmatrix} t_{11} & 0 & \cdots & 0 \\ t_{12} & t_{22} & \cdots & 0 \\ \vdots & & & \\ t_{1m} & t_{2m} & \cdots & t_{mm} \end{pmatrix} \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1m} \\ 0 & t_{22} & \cdots & t_{2m} \\ \vdots & & & \\ 0 & 0 & \cdots & t_{mm} \end{pmatrix}$$

$$\begin{aligned} a_{11} &= t_{11}^2 \\ a_{12} &= t_{11}t_{12} \\ &\vdots \\ a_{1m} &= t_{11}t_{1m} \\ a_{22} &= t_{12}^2+t_{22}^2 \\ &\vdots \\ a_{2m} &= t_{12}t_{1m}+t_{22}t_{2m} \\ &\vdots \\ a_{mm} &= t_{1m}^2+\cdots.+t_{mm}^2 \end{aligned} \quad \Rightarrow \quad \begin{aligned} da_{11} &= 2dt_{11} \\ da_{12} &= t_{11}dt_{12}+\cdots \\ &\vdots \\ da_{1m} &= t_{11}dt_{1m}+\cdots \\ da_{22} &= 2t_{22}dt_{22}+\cdots \\ &\vdots \\ da_{2m} &= t_{22}dt_{2m}+\cdots \\ &\vdots \\ da_{mm} &= 2t_{mm}dt_{mm}+\cdots \end{aligned}$$

$$\Downarrow$$

两边同时求外积

定义A1：多元gamma函数：

$$\Gamma_m(a) = \int_{A_{m\times m}>0} \exp(-tr(A))|A|^{a-\frac{m+1}{2}}(dA), a > (m-1)/2$$

定理A6：$a > (m-1)/2$

$$\Gamma_m(a) = \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma\left(a - \frac{(i-1)}{2}\right)$$

证明：$A = T^\top T$, 其中 T为对角元为正数的上三角矩阵, 则

$$\mathrm{t}r(A) = \sum_{i\le j} t_{ij}^2,\ |A| = \prod t_{ii}^2$$

则

$$\Gamma_m(a) = \int_{A_{m\times m}>0} \exp(-tr(A))|A|^{a-\frac{m+1}{2}}(dA)$$

$$= \int \cdots \int \exp(-\sum_{i\le j} t_{ij}^2)\ \prod t_{ii}^{2a-m-1} \left\{2^m \prod_{i=1}^{m} t_{ii}^{m-i+1} \wedge_{i\le j} dt_{ij}\right\}$$

$$= \prod_{i<j} \int \exp(-t_{ij}^2)\, dt_{ij} \times 2^m \prod_{i=1}^{m} \int \exp(-t_{ii}^2)\, t_{ii}^{2a-i} dt_{ii}$$

$$= (\sqrt{\pi})^{m(m-1)/2} \prod_{i=1}^{m} \Gamma\left(a - \frac{(i-1)}{2}\right)$$

定理A7：假设$a > \frac{m-1}{2}, \Sigma_{m\times m} > 0$

$$\int_{A_{m\times m}>0} \exp(-tr(\Sigma^{-1}A)/2)|A|^{a-\frac{m+1}{2}}(dA) = \Gamma_m(a)2^{ma}|\Sigma|^a$$

证明：令$V = \Sigma^{-1/2}A\,\Sigma^{-1/2}/2$

$$(dA) = \left(2\Sigma^{1/2}dV\Sigma^{1/2}\right) = 2^{\frac{m(m+1)}{2}}|\Sigma|^{\frac{m+1}{2}}(dV)$$

$$\int_{A_{m\times m}>0} \exp(-tr(\Sigma^{-1}A)/2)|A|^{a-\frac{m+1}{2}}(dA) =$$

$$= 2^{ma}|\Sigma|^a \int_{V>0} \exp(-tr(V))|V|^{a-\frac{m+1}{2}}(dV) = 2^{ma}|\Sigma|^a\Gamma_m(a)$$

这验证了Wishart分布概率密度积分为1

定理A8. 如果Z是$n \times m$矩阵，秩为m，Z可以唯一分解为(Schmidt正交化)

$$Z = H_1 T,$$

其中$H_1 = (\mathbf{h}_1, \dots, \mathbf{h}_m)$是$n \times m$列正交矩阵，$H_1^\top H_1 = I_m, T$是$m \times m$上三角矩阵（对角元$> 0$），则

$$(dZ) = \prod_{i=1}^m t_{ii}^{n-i} (dT)(H_1^\top dH_1),$$

其中

$$(H_1^\top dH_1) = \bigwedge_{i=1}^m \bigwedge_{j=i+1}^n \mathbf{h}_j^\top d\mathbf{h}_i$$

其中$H = (\mathbf{h}_1, \dots, \mathbf{h}_m, \mathbf{h}_{m+1}, \dots, \mathbf{h}_n)$是$H_1$补全的$n$阶正交矩阵。

证明参见Muirhead P63，Theorem 2.1.13

记$A = Z^\top Z = T^\top T$, 由定理A5，$J(A \to T) = 2^m \prod_{i=1}^m t_{ii}^{m-i+1}$，即

$$(dA) = 2^m \prod_{i=1}^{m} t_{ii}^{m-i+1} (dT)$$

由定理A8得

定理A9. 如果Z是$n \times m$矩阵，秩为m, Z 具有如下（唯一）分解
$Z = H_1 T$, 记 $A = Z^\top Z$， 则

$$(dZ) = 2^{-m} |A|^{\frac{n-m-1}{2}} (dA)(H_1^\top dH_1),$$

定理A9基本就是我们希望得到的$J(Z \to A, H_1)$ 。

类似于向量的球坐标表示，定理A8-9通过列正交矩阵H_1和上三角矩阵T表示矩阵Z。

$$(H_1^\top dH_1) = \wedge_{i=1}^{m} \wedge_{j=i+1}^{n} \mathbf{h}_j^\top d\mathbf{h}_i$$

列正交的$n \times m$矩阵集合

$$V_{m,n} = \{H_1 \in R^{n\times m}:\ \ H_1^\top H_1 = I_m\} \subset R^{nm}$$

是R^{nm}的$nm - m(m+1)/2$维的子集(surface)，称为*Stiefel manifold,*

两种重要情形：

- $n = m$: $V_{m,m} = \mathcal{O}(m)$， 正交群；
- $m = 1$: $V_{1,n} = S^{n-1}$，单位球面。

考虑$n = m$情形。假设$H \in \mathcal{O}(m)$, $H^\top dH$是反对称矩阵，$(H^\top dH)$是其上三角元素的外积：

$$(H^\top dH) = \wedge_{i<j} \mathbf{h}_j^\top d\mathbf{h}_i$$

它是（左、右）正交不变的，则这种不变的微分形式定义了$\mathcal{O}(m)$上的Harr测度：

$$\mu(\mathcal{D}) = \int_{\mathcal{D}} (H^\top dH), \mathcal{D} \subset \mathcal{O}(m),$$

μ是正交不变测度： $\mu(H\mathcal{D}) = \mu(\mathcal{D}H) = \mu(\mathcal{D})$, 除了一个常数倍数之外，该测度是$\mathcal{O}(m)$上唯一正交不变测度(回忆球面均匀分布)。

一般n，m情形下类似。

$\mathcal{O}(m)$的体积

$$\mu(\mathcal{O}(m)) = \int_{\mathcal{O}(m)} (H^\top dH),$$

例如：$m = 2$

$$H = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} = (\mathbf{h}_1, \mathbf{h}_2)$$

$$(H^\top dH) = \mathbf{h}_2^\top d\mathbf{h}_1 = (-\sin(\theta), \cos(\theta)) \begin{pmatrix} -\sin(\theta)d\theta \\ \cos(\theta)d\theta \end{pmatrix} = d\theta,$$

$$\mu(\mathcal{O}(2)) = \int_{\mathcal{O}(2)} (H^\top dH) = \int_0^{2\pi} d\theta = 2\pi$$

定理A10：$V_{m,n}$的体积

$$\int_{V_{m,n}} (H_1^\top dH_1) = 2^m \pi^{mn/2} / \Gamma_m(n/2)$$

推论A2 (1) $\mu(\mathcal{O}(m)) = \frac{2^m \pi^{\frac{m^2}{2}}}{\Gamma_m\left(\frac{m}{2}\right)}$;

(2) $m = 1$时，$\mu(\mathcal{O}(m)) = \frac{2\pi^{\frac{1}{2}}}{\Gamma\left(\frac{\mathrm{n}}{2}\right)}$，$S^{n-1}$的表面积。

第7讲定理4. $n\ge m$时，$A\sim W_m(n,\Sigma)$的概率密度函数为

$$p_{W_m(n,\Sigma)}(A)=\frac{|\Sigma|^{-n/2}|A|^{(n-p-1)/2}}{2^{nm/2}\Gamma_m(n/2)}\exp\left(-\frac{1}{2}tr(A\Sigma^{-1})\right),$$

其中$\Gamma_m(x)=\pi^{m(m-1)/4}\prod_{i=1}^{m}\Gamma\left(x-\frac{i-1}{2}\right),\ x>(p-1)/2.$

证明：$A=\mathrm{Z}^{\top}\mathrm{Z},\mathrm{Z}_{\mathrm{n}\times\mathrm{m}}$的概率密度

$$p(\mathrm{Z})=\frac{1}{(2\pi)^{nm/2}|\Sigma|^{n/2}}\exp\left(-\frac{1}{2}tr\{\Sigma^{-1}\mathrm{Z}^{\top}Z)\}\right)(dZ),$$

$\mathrm{Z}=H_1T,H_1^{\top}H_1=I_m,H_1$是$n\times m,W=\mathrm{Z}^{\top}\mathrm{Z}=T^{\top}T$

由定理A9，$(dZ)=2^{-m}|A|^{\frac{n-m-1}{2}}(dA)(H_1^{\top}dH_1)$,

所以A,H_1的联合密度

$$p(A,H_1)=\frac{1}{(2\pi)^{nm/2}|\Sigma|^{n/2}}\exp(-\frac{1}{2}tr(\Sigma^{-1}A))$$
$$\times 2^{-m}|A|^{\frac{n-m-1}{2}}(dA)(H_1^{\top}dH_1),$$

对H_1积分，利用定理A10的体积公式，即得A的密度。

Deemer and Olkin (1951) The Jacobians of certain matrix transformations useful in multivariate analysis: based on lectures of P.L.Hsu at the University of North Carolina,1947. Biometika,38.

THE JACOBIANS OF CERTAIN MATRIX TRANSFORMATIONS USEFUL IN MULTIVARIATE ANALYSIS

BASED ON LECTURES OF P. L. HSU AT THE UNIVERSITY OF NORTH CAROLINA, 1947

BY WALTER L. DEEMER AND INGRAM OLKIN, *University of North Carolina*

Editorial Note. The following paper was submitted by Prof. Hotelling in the summer of 1950 for publication in *Biometrika* with the accompanying Note from Col. Deemer and Mr Olkin:

'In 1947 Prof. P. L. Hsu gave courses in multivariate analysis at the University of North Carolina in which he developed new techniques for finding Jacobians of certain matrix transformations. Hsu returned to China at the end of that academic year, leaving as a record of this material only the notes of students in his classes.

'We (Deemer and Olkin) became interested in these matrix transformations in the course of our work in multivariate analysis. Since we did not take Hsu's courses, we used Ralph Bradley's notes as a basis for our studies.

'In the spring of 1948 there was a seminar in multivariate analysis under the direction of Prof. Harold Hotelling. At his request we prepared some lectures on matrix transformations. At the completion of the seminar, Prof. Hotelling and Prof. R. C. Bose suggested that in view of the importance of these techniques and their non-availability, we should prepare an expository paper giving a systematic development with all proofs given in detail.

'All the new ideas of importance in this paper are due to Hsu. Our contribution has been to organize the material in logical form, making all proofs complete with the necessary lemmas explicitly stated and proved.

'Efforts were made to communicate with Prof. Hsu in order that he could review this material before it was circulated. To date such efforts have failed.'

Since this contribution was received contact has been made with Prof. Hsu and the paper is now published with his approval. A suggestion which he made for improvement has been added as a Note on p. 361. However, in view of the liberty they have taken in the exposition of Hsu's ideas, the American authors would like it to be clear that they are to be held responsible for any errors.

FOREWORD

BY HAROLD HOTELLING

We are apparently at the beginning of a major development in the use of statistical procedures for joint treatment of a multiplicity of correlated variates. Many of the new methods depend ultimately on the distribution of the roots of certain determinantal equations. This distribution was published simultaneously in 1939 by P. L. Hsu (1939) and R. A. Fisher (1939) in the *Annals of Eugenics*, and, excepting for a constant multiplier, by S. N. Roy (1938–40) in *Sankhyā*. Hsu's derivation, which, like Roy's, is otherwise complete, demonstrates the correctness of his formula for this constant multiplier only for the case of three variates. Proof that the formula is correct for the general case has turned out to be