

一、正交因子分析

(一) 模型

设 p 维向量 $X = (X_1, X_2, \dots, X_p)^T \in \mathbb{R}^p$ $E(X) = \mu = (\mu_1, \mu_2, \dots, \mu_p)^T$
 $\text{Var}(X) = \Sigma$ 。假设 X 可用少数的随机变量 $F_1, F_2, \dots, F_m$ ($m < p$) 比较好地线性近似。写出如下回归模型的因子模型：

$$X_1 = \mu_1 + a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + \varepsilon_1$$

$$X_2 = \mu_2 + a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + \varepsilon_2$$

$\vdots$

$$X_p = \mu_p + a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + \varepsilon_p$$

其中 F_j 称为 **公共因子**， ε_j 称为 **特殊因子**， a_{ij} 称为第 i 个变量在第 j 个因子上的 **载荷**，矩阵 $A = (a_{ij})_{p \times m}$ 称为 **载荷矩阵**。

记 $F = (F_1, F_2, \dots, F_m)^T \in \mathbb{R}^m$ $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)^T \in \mathbb{R}^p$ 因子模型可写作：

$$X = \mu + AF + \varepsilon$$

Tip: A 系数未知， F 为随机的且未知；

相比回归模型， M 系数已知，参数 β 未知但不随机 ($Y = M\beta + \varepsilon$)

(二) 模型假设

$$(1) E(F) = 0 \quad \text{Var}(F) = I$$

$$(2) E(\varepsilon) = 0 \quad \text{Var}(\varepsilon) = \Phi = \text{diag}(\phi_1, \phi_2, \dots, \phi_p)$$

$$(3) \text{Cov}(F, \varepsilon) = E(F\varepsilon^T) = 0_{p \times m}$$

称这样的因子模型为 **正交因子模型**

(三) 正交因子模型的协方差结构

$$\begin{aligned}\Sigma &= \text{Var}(X) = E[(X-\mu)(X-\mu)^T] \\&= E[(AF + \varepsilon)(AF + \varepsilon)^T] \\&= E[AF F^T A^T + \varepsilon F^T A^T + AF \varepsilon^T + \varepsilon \varepsilon^T] \\&= AE(FF^T)A^T + E(\varepsilon F^T)A^T + AE(F \varepsilon^T) + E(\varepsilon \varepsilon^T) \\&= A \text{Var}(F) A^T + 0 + 0 + \text{Var}(\varepsilon) \\&= AA^T + \Phi\end{aligned}$$

故有：

$$\text{Var}(X_i) = \sigma_{ii} = a_{i1}^2 + a_{i2}^2 + \cdots + a_{im}^2 + \phi_i \quad i=1,2,\dots,p$$

记

$h_i^2 = a_{i1}^2 + a_{i2}^2 + \cdots + a_{im}^2$ 表示 m 个公共因子对原始变量 X_i 的方差贡献总和，称为第 i 个 **共同度**（**共性方差**）而其中 ϕ_i 称为 **特殊度**（**特殊方差**）

任意 2 变量 $X_i, X_j \in R$ 的协方差为：

$$\text{Cov}(X_i, X_j) = \text{Cov}\left(\mu_i + \sum_{k=1}^m a_{ik} F_k + \varepsilon_i, \mu_j + \sum_{k=1}^m a_{jk} F_k + \varepsilon_j\right)$$

由 $\text{Var}(F) = I$ $\text{Var}(\varepsilon) = \Phi$ 均对角，故

$$\text{Cov}(X_i, X_j) = \sum_{k=1}^m a_{ik} a_{jk}$$

Tip: $\forall i, j \in \{1, 2, \dots, p\}$ ， X_i, X_j 的协方差仅由公共因子 A 决定，与特殊因子 Φ 无关。

$$\begin{aligned}
\text{Cov}(X, F) &= E[(X - \mu)(F - 0)^T] \\
&= E[(AF + \varepsilon) F^T] \\
&= A E[FF^T] + E(\varepsilon F^T) \\
&= A \text{Var}(F) + 0 \\
&= A
\end{aligned}$$

故有

$$\text{Cov}(X_i, F_j) = a_{ij} \quad i \in \{1, 2, \dots, p\} \quad j \in \{1, 2, \dots, m\}$$

记 $b_j^2 = a_{1j}^2 + a_{2j}^2 + \dots + a_{pj}^2$ 表示 F_j 对所有原始变量的影响。归因于第 j 个因子的样本总方差所占比例: $\frac{b_j^2}{\sum_{i=1}^p \sigma_{ii}}$

Tip: h_i^2 为 A 第 i 行平方和; b_j^2 为 A 第 j 列平方和。

(四) 模型的伸缩不变性

设 C 为对角阵, 对 X 作伸缩变化 CX :

$$CX = CAF + C\varepsilon \quad (\text{不妨设 } \mu=0)$$

令 $Y = CX$ $\tilde{A} = CA$ $\tilde{\varepsilon} = C\varepsilon$ 此时:

$$Y = \tilde{A}F + \tilde{\varepsilon}$$

模型假设仍然成立, F 因子不变。故问题等价。

取 $C = \text{diag}(1/\sqrt{\sigma_{ii}}) \quad i=1, 2, \dots, p$ 则有

$$\text{Cov}(Y) = \text{Cov}(CX)$$

$$= C \cdot \text{Cov}(X) \cdot C^T$$

$$= C \Sigma C^T$$

$$= R_x$$

即 $\text{Cov}(Y)$ 为原 X 的相关系数阵。

结论：从 X 的协方差 Σ 与从 X 的相关阵 R_x 得到的 F 是相同。与之对比的是，主成分却没有这个性质。

(五) 因子表示的不唯一性

设 T 为正交阵，且 $T \in \mathbb{R}^{m \times m}$ 。 ($TT^T = T^T T = I$)
有

$$\begin{aligned} X &= \mu + AF + \varepsilon \\ &= \mu + ATT^T F + \varepsilon \end{aligned}$$

记 $\tilde{A} = AT$ $\tilde{F} = T^T F$ 有

$$X = \mu + \tilde{A}\tilde{F} + \varepsilon$$

其中仍有 $E(\tilde{F}) = 0$ $\text{Var}(\tilde{F}) = I$ $\text{Cov}(\tilde{F}, \varepsilon) = 0$

仍有

$$\Sigma = AA^T + \Psi = \tilde{A}\tilde{A}^T + \Psi$$

共同度 h_i^2 也不变。

二、因子模型的估计

估计载荷阵 A 和特殊方差 $\Psi = \text{diag}(\phi_i) \quad i=1,2,\dots,p$

(一). 主成分法

设总体协方差阵 Σ 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ 有如下分解:

$$\begin{aligned}\Sigma &= P \Lambda P^T = \sum_{i=1}^p \lambda_i p_i p_i^T \\ &= (\sqrt{\lambda_1} p_1, \sqrt{\lambda_2} p_2, \dots, \sqrt{\lambda_p} p_p) \begin{pmatrix} \sqrt{\lambda_1} p_1 \\ \sqrt{\lambda_2} p_2 \\ \vdots \\ \sqrt{\lambda_p} p_p \end{pmatrix}\end{aligned}$$

$$\begin{aligned}&= (P \Lambda^{\frac{1}{2}}) (P \Lambda^{\frac{1}{2}})^T \\ &\triangleq A A^T\end{aligned}$$

其中 $\Lambda = \text{diag}(\lambda_i)$ P 为正交阵, 即 $A = P \Lambda^{\frac{1}{2}} \in \mathbb{R}^{p \times p}$

并没有降维。略去 Σ 的后 $p-m$ 个特征值:

$$\begin{aligned}\Sigma &\approx \sum_{i=1}^m \lambda_i p_i p_i^T \\ &= (\sqrt{\lambda_1} p_1, \sqrt{\lambda_2} p_2, \dots, \sqrt{\lambda_m} p_m) \begin{pmatrix} \sqrt{\lambda_1} p_1 \\ \sqrt{\lambda_2} p_2 \\ \vdots \\ \sqrt{\lambda_m} p_m \end{pmatrix} \\ &\triangleq A A^T\end{aligned}$$

如此 $A = P_1 \Lambda_1^{\frac{1}{2}} \in \mathbb{R}^{p \times m}$ 其中 $P_1 = (p_1, p_2, \dots, p_m) \in \mathbb{R}^{p \times m}$

而 $\Lambda_1 = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathbb{R}^{m \times m}$

然后取 $\Psi = \text{diag}(\Sigma - A A^T)$ 即 $\Sigma - A A^T$ 主对

角元素构成的对角阵 $\Phi \in \mathbb{R}^{p \times p}$

样本估计:

$$\tilde{A} = \tilde{P} \tilde{\Lambda}^{\frac{1}{2}}$$

$$\tilde{\Phi} = \text{diag}(S - \tilde{A}\tilde{A}^T)$$

$$\text{其中 } S = \tilde{P} \tilde{\Lambda} \tilde{P}^T$$

(二). 迭代主因子法

① 对样本 $X \in \mathbb{R}^{n \times p}$ 求相关系数阵 $R \in \mathbb{R}^{p \times p}$

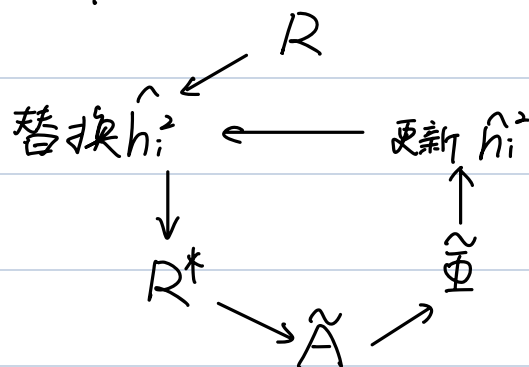
② 利用 $\hat{h}_i^2 = a_{i1}^2 + a_{i2}^2 + \dots + a_{im}^2 \quad i = 1, 2, \dots, p$ 替代 R 的主对角线, 得到 R^*

$$\text{③ } R^* = \tilde{P}_2 \tilde{\Lambda}_2 \tilde{P}_2^T \quad \text{后得到 } \tilde{A}_2 = \tilde{P}_2 \tilde{\Lambda}_2^{\frac{1}{2}}$$

$$\text{④ } \Rightarrow \tilde{\Phi}_2 = \text{diag}(R^* - \tilde{A}_2 \tilde{A}_2^T) \quad \text{后更新 } \hat{h}_i^2$$

$$\text{即: } R^*_{ii} = \hat{h}_i^2 \rightarrow \tilde{\Phi}_{ii}$$

⑤ 回到 ③ 直到收敛



初始 $\hat{h}_i^2$:

$$\hat{h}_{i0}^2 = \sum_{j=1}^m [\text{Cov}(X_i, X_j)]^2$$

或

$$\hat{h}_{i0}^2 = \max_{j=1, 2, \dots, m} [\text{Cov}(X_i, X_j)]^2$$

$i = 1, 2, \dots, p$

Tip: 收敛后 $\hat{\Phi}$ 仍有 $\exists \hat{\Phi}_i < 0$ 则不可取

(三). 极大似然估计法

假设, F, ε 正态, $(F, \varepsilon)^T$ 联合正态, 则有

$$X = \mu + AF + \varepsilon \sim N_p(\mu, AA^T + \Phi)$$

记 $l(\mu, A, \Phi)$ 为 对数似然函数, MLE 为:

$$(\hat{\mu}, \hat{A}, \hat{\Phi}) = \arg \max_{(\mu, A, \Phi)} l(\mu, A, \Phi)$$

$$= -\frac{n}{2} \log |AA^T + \Phi| - \frac{1}{2} (X_i - \mu)^T (AA^T + \Phi)^{-1} (X_i - \mu)$$

Tip: $(\hat{\mu}, \hat{A}, \hat{\Phi})$ 无解析解, 可使用数值优化

(四) 样本估计

常常根据样本相关阵 R , 作谱分解: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ p_i 为 λ_i 对应的单元特征向量, 且 $P = (p_1, p_2, \dots, p_p) \in \mathbb{R}^{p \times p}$, 有估计

$$\begin{aligned} \text{载荷矩阵 } \tilde{A} &= (\sqrt{\lambda_1} p_1, \sqrt{\lambda_2} p_2, \dots, \sqrt{\lambda_m} p_m) \\ &\triangleq (\hat{a}_{ij})_{p \times m} \quad (p > m) \end{aligned}$$

特殊方差 $\tilde{\phi}_i = 1 - \sum_{j=1}^m \hat{a}_{ij}^2 \quad i=1, 2, \dots, p$

共同度 $\tilde{h}_i^2 = \sum_{j=1}^m \hat{a}_{ij}^2 \quad i=1, 2, \dots, p$

F_j 对所有 p 个原始变量的方差贡献估计:

$$\tilde{b}_j^2 = \sum_{i=1}^p \hat{a}_{ij}^2 = \|\sqrt{\lambda_j} p_j\|^2 = \lambda_j \quad j=1, 2, \dots, m$$

Tip: 因子个数 m 的选取

例如:
$$\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j} \geq \alpha$$

三、因子旋转

满足因子模型的方差结构

$$\Sigma = AA^T + \Psi$$

的载荷矩阵 A 不唯一

相应的, 因子模型

$$X - \mu = AF + \varepsilon$$

的公共因子 F 不唯一

eg. $X - \mu = AF + \varepsilon = ATT^T F + \varepsilon$ 当 T 为正交阵成立, 此时

$$A' = AT \quad F' = T^T F \text{ 也满足 } X - \mu = A'F' + \varepsilon$$

eg.

要求 $G = AA^T$ 为对角阵, 且对角元素从大到小排列.

为对角阵, 且各对角元素从大到小排列, 这使得产生的第一公共因子对所有原始变量的共同方差贡献最大, 第二公共因子在与第一公共因子不相关的前提下贡献次大, 等等, 类似于主成分分析中的主成分排序。这样可以唯一确定载荷阵 A (各列可以有正负号的变化), 从而确定公共因子。

因子正交旋转: T 为正交阵, 否则为因子斜交旋转

常见的正交旋转:

① Varimax: 每个因子 F_j 仅有少数大的 a_{ij} ($i=1, 2, \dots, p$)
即 A 中列差别大。

$$\begin{aligned} \max V &= \frac{1}{p} \sum_{j=1}^m \left(\sum_{i=1}^p \tilde{a}_{ij}^{*4} - \frac{1}{p} \left(\sum_{i=1}^p \tilde{a}_{ij}^{*2} \right)^2 \right) \\ &= \sum_{j=1}^m \left(\frac{1}{p} \sum_{i=1}^p \left(\tilde{a}_{ij}^{*2} - \frac{1}{p} \sum_{i=1}^p \tilde{a}_{ij}^{*2} \right)^2 \right) \end{aligned}$$

$$\left(\tilde{a}_{ij}^* = \hat{a}_{ij} / \hat{h}_i \right)$$

② Quartimax: 每个原始变量 X_i 仅与一个 F_j 强相关, 与其他近似不相关。即 A 中行差别大
将上式 i, j 互换即可。

四、因子得分 估计 F

模型 $X - \mu = LF + \varepsilon$

协方差结构 $\Sigma = LL^T + \psi$

公共因子 F 的估计称为 **因子得分**

(一) 加权最小二乘法

$$\min_F \sum_{i=1}^p \frac{\varepsilon_i^2}{\psi_{ii}} = \varepsilon^T \psi^{-1} \varepsilon \quad \text{证明见后}$$
$$= (X - \mu - LF)^T \psi^{-1} (X - \mu - LF)$$

正式估计值由最小二乘估计 (回归: $X = \mu + LF + \varepsilon$) 知

$$F^* = \hat{f} = (L^T \psi^{-1} L)^{-1} L^T \psi^{-1} (X - \mu)$$

代入估计值: (极大似然估计 $\hat{L} \hat{\psi}$)

$$\hat{f}_j^{LS} = (\hat{L}^T \hat{\psi}^{-1} \hat{L})^{-1} \hat{L}^T \hat{\psi}^{-1} (X_j - \bar{x}) \quad j = 1, 2, \dots, n$$

若 X 标准化后

$$\hat{f}_j = (\hat{L}_z^T \hat{\psi}_z^{-1} \hat{L}_z)^{-1} \hat{L}_z^T \hat{\psi}_z^{-1} z_j \quad j = 1, 2, \dots, n$$

若使用主成分估计 ($\tilde{L}$, 估计不出 ψ 或得说 $\psi = I$)

$$\tilde{f}_j = (\tilde{L}^T \tilde{L})^{-1} \tilde{L}^T (X_j - \bar{x})$$

标准化后

$$\tilde{f}_j = (\tilde{L}_z^T \tilde{L}_z)^{-1} \tilde{L}_z^T z_j$$

$$\tilde{L} = P_1 \Lambda_1^{\frac{1}{2}}$$

(一) 主成分估计:

$$\tilde{L} = [\sqrt{\lambda_1} \hat{e}_1 \quad \sqrt{\lambda_2} \hat{e}_2 \quad \dots \quad \sqrt{\lambda_m} \hat{e}_m]^T$$

此时有因子得分

$$\tilde{f}_j = (\Lambda_1^{\frac{1}{2}} P_1^T P_1 \Lambda_1^{\frac{1}{2}})^{-1} \Lambda_1^{\frac{1}{2}} P_1^T (X_j - \bar{x})$$

$$\hat{f}_j = \left[\frac{1}{\sqrt{\lambda_i}} \hat{e}_i^T (x_j - \bar{x}) \right]_{i=1,2,\dots,m}^T \in \mathbb{R}^m = \left[\frac{\sqrt{\lambda_i}}{\lambda_i} e_i^T (x_j - \bar{x}) \right]_{i=1,2,\dots,m}$$

对比 PCA 的主成分

$$\hat{y}_j = \left[\hat{e}_i^T (x_j - \bar{x}) \right]_{i=1,2,\dots,m}^T \in \mathbb{R}^m$$

(二). 回归估计法

$$\begin{cases} \frac{1}{n} \sum_{j=1}^n \hat{f}_j = 0 \\ \frac{1}{n-1} \sum_{j=1}^n \hat{f}_j \hat{f}_j^T = I \end{cases}$$

Assumption:

① 假设 $X - \mu = LF + \varepsilon \sim N_p(0, LL^T + \psi)$

② 假设 $(X - \mu, F)_{p+m}^T \sim N_{p+m}(0, \Sigma^*)$ 联合正态分布假设. $\Sigma^* \in \mathbb{R}^{(p+m) \times (p+m)}$

$$\Sigma^* = \text{Var} \begin{pmatrix} X - \mu \\ F \end{pmatrix} = \text{Var} \begin{pmatrix} LF + \varepsilon \\ F \end{pmatrix}$$

$$= \begin{pmatrix} LL^T + \psi & L \\ L^T & I_m \end{pmatrix} \quad \begin{pmatrix} X - \mu \\ F \end{pmatrix} \sim N_{p+m}(0, \Sigma^*)$$

Conclusion: 条件分布

$$E(F | X = x) = \frac{L^T (LL^T + \psi)^{-1} (x - \mu)}{I_m - L^T (LL^T + \psi)^{-1} L}$$

$$\text{Cov}(F | X = x) = I_m - L^T (LL^T + \psi)^{-1} L$$

于是因子得分: (代入 MLE 估计 $\hat{L}, \hat{\psi}$)

$$\hat{f}_j^R = \hat{L}^T \underbrace{(\hat{L} \hat{L}^T + \hat{\psi})^{-1}}_{SS \quad S} (x_j - \bar{x}) \quad j=1,2,\dots,m$$

(三). 近似等价性

$$\hat{f}_j^{LS} = (\hat{L}^T \hat{\Psi}^{-1} \hat{L})^{-1} \hat{L}^T \hat{\Psi}^{-1} (x_j - \bar{x})$$

$$\hat{f}_j^R = \hat{L}^T (\hat{L} \hat{L}^T + \hat{\Psi})^{-1} (x_j - \bar{x})$$

Proof: 代数变换

$$\begin{aligned} \hat{f}_j^{LS} &= (\hat{L}^T \hat{\Psi}^{-1} \hat{L})^{-1} \\ &\quad \cdot (I + \hat{L}^T \hat{\Psi}^{-1} \hat{L}) (I + \hat{L}^T \hat{\Psi}^{-1} \hat{L})^{-1} \\ &\quad \cdot \hat{L}^T \hat{\Psi}^{-1} (x_j - \bar{x}) \\ &= (\hat{L}^T \hat{\Psi}^{-1} \hat{L})^{-1} (I + \hat{L}^T \hat{\Psi}^{-1} \hat{L}) \\ &\quad \cdot \hat{L}^T (\hat{L} \hat{L}^T + \hat{\Psi})^{-1} \cdot (x_j - \bar{x}) \end{aligned}$$

证明

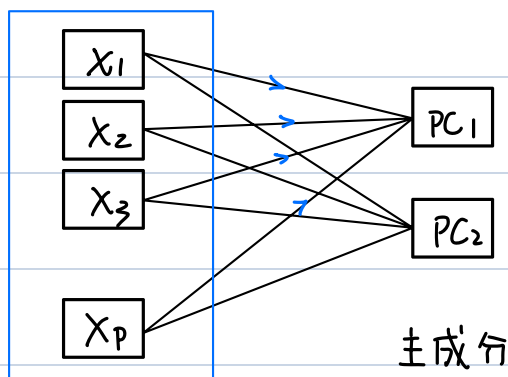
$$= (\hat{L}^T \hat{\Psi}^{-1} \hat{L})^{-1} (I + \hat{L}^T \hat{\Psi}^{-1} \hat{L}) \cdot \hat{f}_j^R$$

$$= (\Delta^{-1} + I) \hat{f}_j^R$$

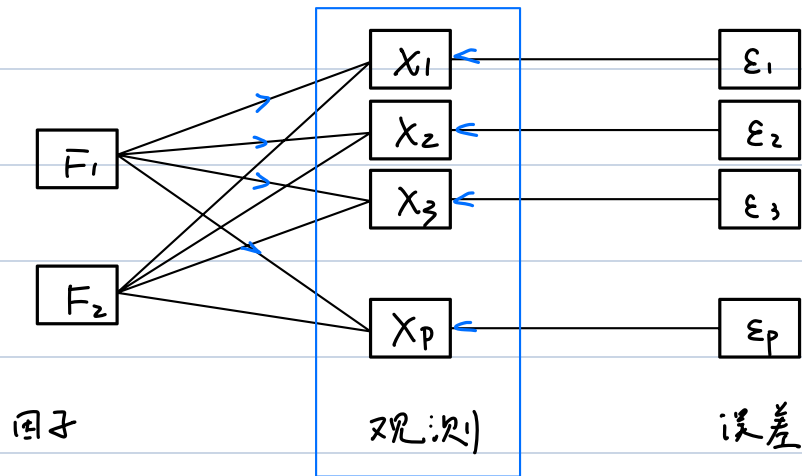
其中 $\Delta = \hat{L}^T \hat{\Psi}^{-1} \hat{L}$

四. 主成分 & 因子分析

PCA: 原始变量的几个线性组合



探索性因子分析 EFA



补记

9.5 证明不等式(9-19)

提示: 因为 $S - \tilde{L}\tilde{L}' - \tilde{\Psi}$ 在对角线上为 0,

$(S - \tilde{L}\tilde{L}' - \tilde{\Psi})$ 元素的平方和 $\leq (S - \tilde{L}\tilde{L}')$ 元素的平方和

现在, $S - \tilde{L}\tilde{L}' = \hat{\lambda}_{m+1} \hat{e}_{m+1} \hat{e}_{m+1}' + \dots + \hat{\lambda}_p \hat{e}_p \hat{e}_p' = \hat{P}_{(2)} \tilde{\Lambda}_{(2)} \hat{P}_{(2)}'$, 其中 $\hat{P}_{(2)} = [\hat{e}_{m+1} \dots \hat{e}_p]$ 且 $\tilde{\Lambda}_{(2)}$ 是对角元为 $\hat{\lambda}_{m+1}, \dots, \hat{\lambda}_p$ 的对角阵.

利用 (A) 元素的平方和 $= \text{tr} AA'$ 及 $\text{tr}[\hat{P}_{(2)} \tilde{\Lambda}_{(2)} \tilde{\Lambda}_{(2)} \hat{P}_{(2)}'] = \text{tr}[\tilde{\Lambda}_{(2)} \tilde{\Lambda}_{(2)}]$.

证明: $S - (\tilde{L}\tilde{L}' + \tilde{\Psi})$ 的元素平方和 $\leq \lambda_{m+1}^2 + \lambda_{m+2}^2 + \dots + \lambda_p^2$

Proof: 记 $S = P \Lambda P^T = [P_1 P_2] \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix} \begin{bmatrix} P_1^T \\ P_2^T \end{bmatrix} = P_1 \Lambda_1 P_1^T + P_2 \Lambda_2 P_2^T$

而 $\tilde{L}\tilde{L}' = P_1 \Lambda_1 P_1^T$ 而 $\tilde{\Psi}$ 为 $S - \tilde{L}\tilde{L}' = P_2 \Lambda_2 P_2^T$ 的对角元素

故 $S - \tilde{L}\tilde{L}' - \tilde{\Psi} = P_2 \Lambda_2 P_2^T - \text{diag}(P_2 \Lambda_2 P_2^T)$

其元素平方和 $\leq P_2 \Lambda_2 P_2^T$ 的元素平方和

而由 A 元素平方和 $= \text{tr}(AA^T)$ 知

$$\text{tr}(P_2 \Lambda_2 P_1^T \cdot P_2 \Lambda_2 P_2^T) = \text{tr}(P_2 \Lambda_2^2 P_2^T)$$

$$= \text{tr}(P_2 P_2^T \Lambda_2^2)$$

$$= \sum_{i=m+1}^p \lambda_i^2$$

故 $S - \tilde{L}\tilde{L}' - \tilde{\Psi}$ 的元素平方和 $\leq \lambda_{m+1}^2 + \lambda_{m+2}^2 + \dots + \lambda_p^2 \quad \square$

补证:

$$(I + L^T \psi^{-1} L)^{-1} L^T \psi^{-1} = L^T (\psi + L L^T)^{-1}$$

其中 $L \in \mathbb{R}^{p \times m}$, $\psi \in \mathbb{R}^{p \times p}$ 且对角, $I \in \mathbb{R}^{p \times p}$

Proof:

引理: Woodbury 矩阵恒等式

$$(A + U C V)^{-1} = A^{-1} - A^{-1} U (C^{-1} + V A^{-1} U)^{-1} V A^{-1}$$

其中 $A \in \mathbb{R}^{n \times n}$ 可逆, $C \in \mathbb{R}^{k \times k}$ 可逆, $U, V^T \in \mathbb{R}^{n \times k}$

回到原题: $A = I$ $C = \psi^{-1}$ $L = V$ $L^T = U$

$$\begin{aligned} \text{有 } (I + L^T \psi^{-1} L)^{-1} &= I - I L^T (\psi + L I L^T)^{-1} L I \\ &= I - L^T (\psi + L L^T)^{-1} L \end{aligned}$$

$$\begin{aligned} \text{故 LHS} &= (I - L^T (\psi + L L^T)^{-1} L) \cdot L^T \psi^{-1} \\ &= L^T \psi^{-1} - L^T (\psi + L L^T)^{-1} L L^T \psi^{-1} \\ &= L^T (\psi^{-1} - (\psi + L L^T)^{-1} L L^T \psi^{-1}) \end{aligned}$$

$$\begin{aligned} \text{其中 } \psi^{-1} - (\psi + L L^T)^{-1} L L^T \psi^{-1} &= (\psi + L L^T)^{-1} (\psi + L L^T) \psi^{-1} - (\psi + L L^T)^{-1} L L^T \psi^{-1} \\ &= (\psi + L L^T)^{-1} (I + \psi^{-1} L L^T) - (\psi + L L^T)^{-1} L L^T \psi^{-1} \\ &= (\psi + L L^T)^{-1} (I + (L L^T \psi^{-1})^T - L L^T \psi^{-1}) \\ &= (\psi + L L^T)^{-1} \end{aligned}$$

$$\text{故 LHS} = L^T (\psi + L L^T)^{-1} = \text{RHS} \quad \square$$

补证:

$$\begin{aligned}\min_F \sum_{i=1}^p \frac{\varepsilon_i^2}{\psi_{ii}} &= \varepsilon^T \Psi^{-1} \varepsilon \\ &= (X - \mu - LF)^T \Psi^{-1} (X - \mu - LF)\end{aligned}$$

矩阵求导:

$$\frac{\partial a^T X}{\partial X} = a$$

$$\frac{\partial X^T A X}{\partial X} = (A^T + A)X$$

$$\begin{aligned}\text{记原式 } G(F) &= (X - \mu - LF)^T \Psi^{-1} (X - \mu - LF) \\ &= (X - \mu)^T \Psi^{-1} (X - \mu) + F^T L^T \Psi^{-1} L F \\ &\quad - F^T L^T \Psi^{-1} (X - \mu) - (X - \mu)^T \Psi^{-1} L F\end{aligned}$$

$$\begin{aligned}\text{故 } \frac{\partial G}{\partial F} &= 0 + 2(L^T \Psi^{-1} L)F \\ &\quad - 2L^T \Psi^{-1} (X - \mu) \\ &= 0\end{aligned}$$

$$\text{有 } (L^T \Psi^{-1} L) F^* = L^T \Psi^{-1} (X - \mu)$$

$$\text{故 } F^* = (L^T \Psi^{-1} L)^{-1} L^T \Psi^{-1} (X - \mu)$$

条件分布

⑤ (条件分布 conditional distribution) 已知 X 是 X_1, X_2 的联合分布 (joint distribution) 服从如下联合正态分布:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

且 Σ_{22} 可逆, 则

$$X_1 | X_2 \sim N(\mu_1 + \Sigma_{12} \cdot \Sigma_{22}^{-1} \cdot (X_2 - \mu_2), \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})$$

即给定 X_2 时 X_1 的分布.