第七章 线性混合效应模型

7.1 纵向数据与线性混合效应模型

在前面 1.6 节中虽然已经提到过纵向数据(Longitudinal data)以及线性混合效应模型(LMM)。这里，我们再回忆一下。一般纵向数据是指随时间演进而追踪测得的数据。更确切地说，在一项研究中，我们准备了一定数量K个个体，对每个个体随时间演进作了测量：对个体i在时刻$t_{i1}<t_{i2}<\cdots t_{in_i}$，测量得到$y_{i1},y_{i2},\cdots y_{in_i}$，$i=1,2,\cdots K$。

则$\left\{y_{ij}:1\le j\le n_i,1\le i\le K\right\}$就是纵向数据。对固定的$i$，$y_{i1},y_{i2},\cdots y_{in_i}$构成一个时间序列。因此，纵向数据也可以说由一批长度不等的时间序列构成。而通常的“时间序列分析”是指对一个较长的时间序列数据(例如某特定股市逐日的数据)来作统计分析。这一点把纵向数据与时间序列分别开来。

医学领域是纵向数据分析最主要的应用领域。因为一种药品或治疗方法的效果要在患者身上作时间追踪调查。

此外，社会经济领域也是重要的应用方面。对这方面有另外一个称呼：Panel data。Panel 是分组的意思，每个个体的量测数据成为一组(也许并不牵涉时间)。

纵向数据也是所谓“集团数据”(Clustered data)的一种。后者是指全部数据因某种共性而划分为一些小集团，例如一对父母所生的子女，同一地域或空间所采集的数据等。其共性都表现在看成同一个体上采集的数据。因此，处理纵向数据的统计方法不少也可用于集团数据。

关于纵向数据有一个基本假设：不同个体的测量是相互独立的，即$\left(y_{i1},y_{i2},\cdots y_{in_i}\right),1\le i\le K$是相互独立的。

令$Y_i=\left(y_{i1},y_{i2},\cdots y_{in_i}\right)'$(为$n_i\times 1$列向量)，从而对纵向数据的建模问题就归结为对$Y_i$的分布作假定的问题。一般，若$Y_i$为连续型变量，通常假定其为多维正态分布，此时问题简化为对其期望EY_i以及协方差$Cov(Y_i)$的假定。

由于$y_{i1}, y_{i2}, \cdots y_{in_i}$是同一个个体上的量测，它们之间往往不再独立而是有某种相关性，其形式与具体问题有关。这种相关性有时不易刻画，也是纵向数据统计分析的难点所在。

当Y_i取离散值，或虽取连续值但非多维正态分布时，要确切的假定其联合分布一般不可能。所幸的是，近三十年发展起来的一种叫“估计方程”的方法只要求对EY_i有正确假定就行。

我们采用由 Laird & Ware(1982)发展的混合效应模型来对纵向数据建模，设

$$Y_i = X_i\beta + Z_i b_i + e_i，\ i = 1, 2, \cdots K,$$

其中Y_i为$n_i \times 1$第i个个体的响应变量，X_i为$n_i \times p$固定效应协变量设计矩阵，β为$p \times 1$的未知的总体的固定效应，Z_i为$n_i \times m$的随机效应设计矩阵，b_i为$m \times 1$的随机效应，e_i为$n_i \times 1$的测量误差。

由纵向数据分析的基本假设，假定$\{b_i, e_i, i = 1, 2, \cdots K\}$为零均值相互独立的随机向量。一般来说假设$\{b_i, 1 \le i \le K\}$独立同分布，$Cov(b_i) = D_{m\times m} \ge 0$，且$b_i, e_i$之间也是独立的。这样我们有$Cov(Y_i) = Z_i D Z_i + R_i$，其中$R_i = Cov(e_i) > 0$为$n_i \times n_i$阶矩阵。

称线性混合效应是平衡的，若所有n_i相等且随机效应部分的设计矩阵Z_i也相等。

计$n = \sum_{i=1}^{K} n_i$为所有观测总数，如果写成单一向量的形式：

$$Y = X\beta + Zb + e,$$

这里

$$Y_{n\times 1} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_K \end{pmatrix},\ X_{n\times p} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{pmatrix},\ Z = \begin{pmatrix} Z_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & Z_K \end{pmatrix}$$

为$n \times (mK)$阶矩阵，

$b_{(mK)\times 1}=\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{pmatrix}$，$e_{n\times 1}=\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_K \end{pmatrix}$，

此时$Cov(b)=I_K\otimes D$。

注：$\otimes$为矩阵的 Kronecker product，两矩阵 $A=(a_{ij})_{m\times n}$，$B=(b_{ij})_{p\times q}$)，则 $A\otimes B=\begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}$ 为 $(mp)\times(nq)$阶矩阵。

记 $\eta=\begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_K \end{pmatrix}=\begin{pmatrix} Z_1b_1+e_1 \\ Z_2b_2+e_2 \\ \vdots \\ Z_Kb_K+e_K \end{pmatrix}$，则有 $E\eta=0$，

$$Cov(\eta)=\begin{pmatrix} Z_1DZ_1'+R_1 & 0 & \cdots & 0 \\ 0 & Z_2DZ_2'+R_2 & & \vdots \\ \vdots & \vdots & & 0 \\ 0 & 0 & \cdots & Z_KDZ_K'+R_K \end{pmatrix}$$

得到边际模型(Marginal model)

$$Y=X\beta+\eta。$$

一般情形下，作为量测误差e，通常可以假定$Cov(e)=\sigma^2I_n$(以下不特别声明都作此假定)，这意味着$R_i=\sigma^2I_{n_i}$。若令$\tilde{D}=\dfrac{D}{\sigma^2}$，则

$$Cov(\eta)=\sigma^2\begin{pmatrix} Z_1\tilde{D}Z_1'+I_{n_1} & 0 & \cdots & 0 \\ 0 & Z_2\tilde{D}Z_2'+I_{n_2} & & \vdots \\ \vdots & \vdots & & 0 \\ 0 & 0 & \cdots & Z_K\tilde{D}Z_K'+I_{n_K} \end{pmatrix}$$

此时，对于线性混合效应模型来说，感兴趣的未知参数为$\theta=\left(\beta',\sigma^2,\left(vech(\tilde{D})\right)'\right)'$。

注：这里$vech(\tilde{D})$为$m\times m$阶对称矩阵中独立变化的$\frac{m(m+1)}{2}$个元素构成的列向量。

从而，感兴趣的未知参数θ的维数为$p+1+\frac{m(m+1)}{2}$。

7.2 极大似然估计

本节我们先来研究模型未知参数的估计。假定$e_i \sim N(0,\sigma^2 I_{n_i})$，$b_i \sim N(0,D)$，此时 $Y_i \sim N\left(X_i\beta,\sigma^2(I_{n_i}+Z_i\tilde{D}Z_i')\right)$，$i=1,2,\cdots K$。

略去无关常数项，对数似然为

$$l(\theta)=-\frac{1}{2}\left\{n\ln\sigma^2+\sum_{i=1}^{K}\ln\left|I_{n_i}+Z_i\tilde{D}Z_i'\right| + \frac{\sum_{i=1}^{K}(Y_i-X_i\beta)'\left(I_{n_i}+Z_i\tilde{D}Z'\right)^{-1}(Y_i-X_i\beta)}{\sigma^2}\right\}$$

这里θ由7.1节定义，记$\Theta=\left\{\theta\middle|\beta\in R^p,\sigma^2>0,\tilde{D}_{m\times m}\geq 0(\text{非负定})\right\}$，则参数极大似然估计(MLE)定义为

$$\hat{\theta}=Arg\max_{\theta\in\Theta} l(\theta)。$$

注：参数空间Θ是凸集。

为简化记号，以下记$V_i=V_i(\tilde{D})=I_{n_i}+Z_i\tilde{D}Z_i'$，$e_i(\beta)=Y_i-X_i'\beta$，则

$$l(\theta)=-\frac{1}{2}\left\{n\ln\sigma^2+\sum_{i=1}^{K}\left[\ln|V_i|+\frac{e_i(\beta)'V_i^{-1}e_i(\beta)}{\sigma^2}\right]\right\}。$$

由于

$$V_i^{-1}=\left(I_{n_i}+Z_i\tilde{D}Z_i'\right)^{-1}=I_{n_i}-Z_i\left(I_m+\tilde{D}Z_i'Z_i\right)^{-1}\tilde{D}Z_i'$$
$$=I_{n_i}-Z_i\tilde{D}\left(I_m+Z_i'Z_i\tilde{D}\right)^{-1}Z_i'$$
$$=I_{n_i}-Z_i\left(\tilde{D}^{-1}+Z_i'Z_i\right)^{-1}Z_i'(\text{若}\tilde{D}^{-1}\text{存在})$$
$$|V_i|=\left|I_{n_i}+Z_i\tilde{D}Z'\right|=\left|I_m+\tilde{D}Z_i'Z_i\right|,$$

这样，在对数似然函数中涉及到$n_i\times n_i$阶矩阵求逆就转化为$m\times m$阶矩阵的求逆。此即为dimension-reduction 技巧。

若假设$\tilde{D}^{-1}$存在，则由于

$$\ln|V_i| = \ln\left|I_m + \tilde{D}Z_i'Z_i\right| = \ln\left|\tilde{D}^{-1} + Z_i'Z_i\right| - \ln\left|\tilde{D}^{-1}\right|,$$

将$\tilde{D}^{-1}$看成未知参数，此时对数似然为$\beta, \sigma^2, \tilde{D}^{-1}$的函数。

此外，还可以采用 profile log-likelihood function(子集对数似然函数)技巧来求 MLE。由于在给定β和$\tilde{D}$的条件下，σ^2的 MLE 为

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{K} e_i'(\beta)V_i^{-1}e_i(\beta),$$

将其带入对数似然函数中削去σ^2，从而得到只基于β和$\tilde{D}$的函数，称为β和$\tilde{D}$的 profile log-likelihood function，略去无关常数，记为$l_{profile}(\beta, \tilde{D})$，则

$$l_{profile}(\beta, \tilde{D}) = -\frac{1}{2}\left\{n\ln\sum_{i=1}^{K} e_i'(\beta)V_i^{-1}e_i(\beta) + \sum_{i=1}^{K}\ln|V_i|\right\}。$$

进一步，给定$\tilde{D}$，即V_i已知的条件下，β的极大似然估计(这里即广义最小二乘估计)为

$$\hat{\beta} = Arg\min_{\beta}\sum_{i=1}^{K} e_i'(\beta)V_i^{-1}e_i(\beta)。$$

令$h(\tilde{D}) = \min_{\beta}\sum_{i=1}^{K} e_i'(\beta)V_i^{-1}e_i(\beta)$，则可以得到基于$\tilde{D}$的函数，称为$\tilde{D}$的 profile log-likelihood function，略去无关常数，记为$l_{profile}(\tilde{D})$，则

$$l_{profile}(\tilde{D}) = -\frac{1}{2}\left\{n\ln h(\tilde{D}) + \sum_{i=1}^{K}\ln\left|V_i(\tilde{D})\right|\right\}。$$

从而得到$\tilde{D}$的极大似然估计

$$\hat{\tilde{D}} = Arg\max_{\tilde{D}\geq 0} l_{profile}(\tilde{D})。$$

7.3 限制极大似然估计

对于正态分布来说，一般方差的极大似然估计是有偏的。例如对标准正态线性模型$Y \sim N(X\beta, \sigma^2)$，令$ESS$为残差平方和，则$\sigma^2$的极大似然估计为$\frac{ESS}{n}$是有偏的，会低估$\sigma^2$。其无偏估计为$\frac{ESS}{n-p}$，其中$p$为未知参数$\beta$的维数。其原因在于$Y$的作用有一部分用于估计$\beta$了。因此用于估计$\sigma^2$的自由度不是$n$，而是比$n$小，要调整。对于上一节正态分布假定

的线性混合效应模型，要估计协方差阵σ^2, D或者$\sigma^2, \tilde{D}$(这里$\tilde{D} = \frac{D}{\sigma^2}$)，不止一个参数。因此不是单由调整自由度就能解决。

先看一般线性模型限制极大似然估计(Restricted maximum likelihood estimation, RMLE)的做法，然后将其用于线性混合效应模型。设$Y \sim N(X\beta, V)$，这里$X_{n\times p}$列满秩矩阵，$V_{n\times n}$为协方差矩阵依赖于某些未知参数。

考虑Y的两个线性成分：

$$T_{p\times 1} = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

$$U_{(n-p)\times 1} = BY$$

其中B为$(n-p)\times n$阶矩阵满足$B'B = I_n - P_X$，$BB' = I_{n-p}$。

注：这里T即为V已知时β的广义最小二乘估计。由于$I_n - P_X$是幂等矩阵，因此存在$n\times n$正交阵Q使得$I_n - P_X = Q\begin{pmatrix} I_{n-p} & 0 \\ 0 & 0 \end{pmatrix}Q'$。记

$Q = (Q_1, Q_2)$，其中Q_1为$n\times(n-p)$，则有$Q_1Q_1' = I_n - P_X$。由于$QQ' = I_n$，从而有$Q_1'Q_1 = I_{n-p}$。因此只要取$B = Q_1'$即可。且

$$BX = (BB')BX = B(I_n - P_X)X = 0。$$

这样，得到$ET = \beta$，$EU = BX\beta = 0$。由此看出，T这部分用于估计β，于估计协方差阵无益。而U则不然，其变异纯由协方差阵而来，不涉及到β。这部分正是估计协方差时该用上的。

以上就是 RMLE 的想法。为实施这一想法，须求出U的分布。为此，先计算$\begin{pmatrix} T \\ U \end{pmatrix}$的联合分布。由于

$$\begin{pmatrix} T \\ U \end{pmatrix} = \begin{pmatrix} (X'V^{-1}X)^{-1}X'V^{-1} \\ B \end{pmatrix} Y \equiv CY,$$

这里$C = \begin{pmatrix} (X'V^{-1}X)^{-1}X'V^{-1} \\ B \end{pmatrix}$，从而

。

$\begin{pmatrix} T \\ U \end{pmatrix}$的联合密度=

Y的密度×变换的Jacobi行列式的绝对值

=Y的密度×$\left|C^{-1}\right|$的绝对值

由于

$$\begin{aligned} Cov(T,U) &= (X'V^{-1}X)^{-1}X'V^{-1}Cov(Y)B' \\ &= (X'V^{-1}X)^{-1}X'B' = 0 \end{aligned},$$

故T,U独立。

U的密度=$\dfrac{Y\text{的密度}\times\left|C^{-1}\right|\text{的绝对值}}{T\text{的密度}}$

而$Y \sim N(X\beta,V)$，$T \sim N(\beta,(X'V^{-1}X)^{-1})$，

Y的密度为

$$f(Y) = (2\pi)^{-\frac{n}{2}}|V|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(Y-X\beta)'V^{-1}(Y-X\beta)\right],$$

T的密度为

$$g(T) = (2\pi)^{-\frac{p}{2}}\left|X'V^{-1}X\right|^{\frac{1}{2}}\exp\left[-\frac{1}{2}(T-\beta)'(X'V^{-1}X)(T-\beta)\right]$$

而

$$\begin{aligned} &(Y-X\beta)'V^{-1}(Y-X\beta) \\ &= [Y-XT+X(T-\beta)]'V^{-1}[Y-XT+X(T-\beta)] \\ &= (Y-XT)'V^{-1}(Y-XT)+(T-\beta)'(X'V^{-1}X)(T-\beta) \end{aligned},$$

从而

U的密度$=\left|C^{-1}\right|$的绝对值×

$$(2\pi)^{-\frac{n-p}{2}}|V|^{-\frac{1}{2}}\left|X'V^{-1}X\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(Y-XT)'V^{-1}(Y-XT)\right]$$

由于$|C|^2 = |X'X|^{-1}$，与V无关。

注：由C的定义，

$$C'C = V^{-1}X(X'V^{-1}X)^{-2}X'V^{-1} + I_n - P_X。$$

由于$X_{n\times p}$列满秩，其 SVD(奇异值)分解为

$X = H_{n\times n}\begin{pmatrix} \Lambda_p \\ 0 \end{pmatrix}O_{p\times p}$，其中$H,O$为正交矩阵。

令$H'V^{-1}H = \begin{pmatrix} V_1 & V_2 \\ V_2' & V_3 \end{pmatrix}$，则

$$H'C'CH = \begin{pmatrix} V_1 \\ V_2' \end{pmatrix}V_1^{-1}\Lambda_p^{-2}V_1^{-1}\begin{pmatrix} V_1 & V_2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & I_{n-p} \end{pmatrix}。$$

$$\begin{pmatrix} I_p & 0 \\ -V_2'V_1^{-1} & I_{n-p} \end{pmatrix} H'C'CH$$

$$= \begin{pmatrix} V_1 \\ 0 \end{pmatrix} V_1^{-1}\Lambda_p^{-2}V_1^{-1}\begin{pmatrix} V_1 & V_2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & I_{n-p} \end{pmatrix}$$

$$= \begin{pmatrix} \Lambda_p^{-2} & \Lambda_p^{-2}V_1^{-1}V_2 \\ 0 & I_{n-p} \end{pmatrix} ,$$

两边取行列式，得$\left|C\right|^2 = \left|\Lambda_p\right|^{-2} = \left|X'X\right|^{-1}$，

因此$\left|C^{-1}\right|$与V无关。

RMLE 就是从U出发，用极大似然方法来估计V。略去与V无关的量，取对数得到

$$l_R(V) = -\frac{1}{2}\left[\ln\left|V\right| + \ln\left|X'V^{-1}X\right| + (Y - XT)'V^{-1}(Y - XT)\right]$$

(注意：T中含有V)，从而得到V的估计

$$\hat{V} = Arg\max_{V>0} l(V),$$

代入T中，得到β的 RMLE 估计

$$\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y。$$

关于$l_R(V)$的推导还可以从另外一个角度。将β看成随机，$Y\left|\beta \sim N(X\beta, V)\right.$。即

$$f(Y|\beta) = (2\pi)^{-\frac{n}{2}}\left|V\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(Y - X\beta)'V^{-1}(Y - X\beta)\right]$$

取β的先验分布为广义先验，即$\pi(\beta) \propto 1$，则由 Bayes 公式，Y的边际分布为

$$(2\pi)^{-\frac{n-p}{2}}\left|V\right|^{-\frac{1}{2}}\left|X'V^{-1}X\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(Y - XT)'V^{-1}(Y - XT)\right]$$

略去无关常数，取对数后与$l_R(V)$一致。

将 RML 与标准的似然函数比较，发现多了一个因子$\left|X'V^{-1}X\right|^{-\frac{1}{2}}$，对数似然函数多了一项$-\frac{1}{2}\ln\left|X'V^{-1}X\right|$。当$p$相对于$n$很小时，这一项影响不大，此时 RMLE 与 MLE 很接近。当$\frac{p}{n}$不接近 0 时，经验表明使用 RML 会使估计会得到改善。

以上是一般线性模型 RMLE 的做法，对线性混合效应模型来说，其对数限制似然函数为

$$l_R(\theta) = -\frac{1}{2}\left\{(n-p)\ln\sigma^2 + \sum_{i=1}^{K}\ln\left|X_i'V_i^{-1}X_i\right| + \sum_{i=1}^{K}\left[\ln|V_i| + \frac{e_i(\beta)'V_i^{-1}e_i(\beta)}{\sigma^2}\right]\right\}$$

相对于标准的线性混合效应模型对数似然函数作了调整，增加了两项

$$\frac{p}{2}\ln\sigma^2 - \frac{1}{2}\sum_{i=1}^{K}\ln\left|X_i'V_i^{-1}X_i\right|$$

第二项就是刚才说的 RML 与标准的 ML 函数区别。第一项是对σ^2的因子也做了调整。此时在$\theta\in\Theta$上极大化$l_r(\theta)$得到 RMLE 估计。

7.4 Balanced random-coefficient model

本节考虑一类特殊的线性混合效应模型(LMM)。考虑第i个个体的回归模型

$$Y_i = X_i\gamma + e_i,\ 1\le i\le K$$

回归系数不能看成固定效应，而看成随机的

$$Y_i = X_i\gamma_i + e_i,\ 1\le i\le K$$

$\{\gamma_i, 1\le i\le K\}$ 独立同分布，由于$\gamma_i = E\gamma_i + (\gamma_i - E\gamma_i)$，令$\beta = E\gamma_i$，$b_i = \gamma_i - E\gamma_i$，这样化成标准的 LMM

$$Y_i = X_i\beta + X_ib_i + e_i,\quad 1\le i\le K$$

此模型称为 random-coefficient model，是一类特殊的线性混合效应模型，其$Z_i = X_i$。再特殊一点，如果是平衡的，意味着所有$n_i\equiv s$且$Z_i\equiv Z$， $1\le i\le K$。此模型，MLE 有显示表达。此时，模型为

$$Y_i = Z\beta + Zb_i + e_i, 1\le i\le K,$$

其中$Z_{s\times p}$，且$rank(Z) = p < s$。令$\eta_i = Zb_i + e_i$，则$E\eta_i = 0$，$Cov(\eta_i) = \sigma^2(I_s + Z\tilde{D}Z')\equiv\sigma^2 V$。

对此模型，β的最小二乘估计

$$\hat{\beta}_{OLS} = (Z'Z)^{-1}Z'\bar{Y},$$

这里$\bar{Y}_{s\times 1} = \frac{1}{K}\sum_{i=1}^{K} Y_i$。

当$\tilde{D}$已知时，可得β的广义最小二乘估计$\hat{\beta}_{GLS}$，定义为

$$\hat{\beta}_{GLS} = Arg\min_{\beta}\sum_{i=1}^{K}(Y_i - Z\beta)'V^{-1}(Y_i - Z\beta)。$$

注：通常$\hat{\beta}_{GLS}$并不是β的估计，因为含有未知的$\tilde{D}$。

定理 7.4.1：对此模型有$\hat{\beta}_{GLS} = \hat{\beta}_{OLS}$。

证明：由于

$$\sum_{i=1}^{K}(Y_i - Z\beta)'V^{-1}(Y_i - Z\beta) = \sum_{i=1}^{K}(Y_i - \bar{Y})'V^{-1}(Y_i - \bar{Y}) + K(\bar{Y} - Z\beta)'V^{-1}(\bar{Y} - Z\beta)。$$

从而

$$\hat{\beta}_{GLS} = Arg\min_{\beta}\sum_{i=1}^{K}(Y_i - Z\beta)'V^{-1}(Y_i - Z\beta) = Arg\min_{\beta}(\bar{Y} - Z\beta)'V^{-1}(\bar{Y} - Z\beta)$$

又由于$V^{-1} = \left(I_s + Z\tilde{D}Z'\right)^{-1} = I_s - Z\left(I_m + \tilde{D}Z'Z\right)^{-1}\tilde{D}Z'$及$(\bar{Y} - Z\hat{\beta}_{OLS})'Z = 0$，有

$$(\bar{Y} - Z\beta)'V^{-1}(\bar{Y} - Z\beta) = (\bar{Y} - Z\hat{\beta}_{OLS})'V^{-1}(\bar{Y} - Z\hat{\beta}_{OLS}) + (Z\hat{\beta}_{OLS} - Z\beta)'V^{-1}(Z\hat{\beta}_{OLS} - Z\beta)。$$

因此，当$\beta = \hat{\beta}_{OLS}$时达到最小值。

由此定理，对此模型有$\hat{\beta}_{GLS} = \hat{\beta}_{OLS} = \hat{\beta}_{ML}$，这里$\hat{\beta}_{ML}$表示$\beta$的极大似然估计。

定理 7.4.2：对此模型，其方差参数的极大似然(ML)估计及限制极大似然(RML)估计为

$$\hat{\sigma}^2_{ML} = \hat{\sigma}^2_{RML} = \frac{1}{K(s-p)}\sum_{i=1}^{K} Y_i'(I_s - Z(Z'Z)^{-1}Z')Y_i',$$

$$\tilde{D}_{ML} = \frac{1}{K\hat{\sigma}^2_{ML}}(Z'Z)^{-1}Z'\hat{E}\hat{E}'Z(Z'Z)^{-1} - (Z'Z)^{-1},$$

$$\tilde{D}_{RML} = \frac{1}{(K-1)\hat{\sigma}^2_{ML}}(Z'Z)^{-1}Z'\hat{E}\hat{E}'Z(Z'Z)^{-1} - (Z'Z)^{-1},$$

这里$\hat{E}\hat{E}'$为$s\times s$矩阵

$$\hat{E}\hat{E}' = \sum_{i=1}^{K}(Y_i - Z\hat{\beta}_{ML})(Y_i - Z\hat{\beta}_{ML})'。$$

7.5 LM Model with random intercepts

本节考虑另一类特殊的线性混合效应模型(LMM)，截距项具有随机效应，具体为：

$y_{ij}=a_i+w_{ij}'\gamma+e_{ij}$， $j=1,2\cdots n_i, i=1,2\cdots K$，

其中$\{a_i, 1\le i\le K\}$独立同分布，为 random intercepts。由于$a_i=Ea_i+(a_i-Ea_i)$，令$\alpha=Ea_i$，$b_i=a_i-Ea_i$，则$a_i=\alpha+b_i$。则模型为

$y_{ij}=a+w_{ij}'\gamma+b_i+e_{ij}$， $j=1,2\cdots n_i, i=1,2\cdots K$

这里随机效应b_i是一维的，假设$b_i\sim N(0,\sigma_b^2)$，由于通常假设$e_{ij}\sim N(0,\sigma^2)$，令$d=\dfrac{\sigma_b^2}{\sigma^2}\ge 0$，则$b_i\sim N(0,\sigma^2 d)$。写成向量形式

$$Y_i=X_i\beta+E_{n_i}b_i+e_i，1\le i\le K，$$

其中X_i为$n_i\times p$阶满秩矩阵，其第j行$x_{ij}'=(1,w_{ij}')$，$j=1,2\cdots n_i$；$\beta=(\alpha,\gamma')'$为$p\times 1$固定效应，E_{n_i}为$n_i\times 1$元素全为 1 的向量。

注：相对于标准形式$m=1, Z_i=E_{n_i}, \tilde{D}=d$。

由 7.2 节，可以得到基于β,d的 profile 对数似然

$$l_{profile}(\beta,d)=-\frac{1}{2}\left\{n\ln\sum_{i=1}^{K}e_i'(\beta)(I_{n_i}+d\cdot E_{n_i}E_{n_i}')^{-1}e_i(\beta)+\sum_{i=1}^{K}\ln\left|I_{n_i}+d\cdot E_{n_i}E_{n_i}'\right|\right\},$$

这里$e_i(\beta)=Y_i-X_i\beta$为$n_i\times 1$向量。

由于

$$\left|I_{n_i}+d\cdot E_{n_i}E_{n_i}'\right|=1+n_i d,$$

$$\left(I_{n_i}+d\cdot E_{n_i}E_{n_i}'\right)^{-1}=I_{n_i}-\frac{d}{1+n_i d}E_{n_i}E_{n_i}',$$

从而

$$l_{profile}(\beta,d)=-\frac{1}{2}\left\{n\ln\left[S(\beta)-\sum_{i=1}^{K}\frac{n_i^2 d h_i^2(\beta)}{1+n_i d}\right]+\sum_{i=1}^{K}\ln(1+n_i d)\right\}$$

这里

$$S(\beta)=\sum_{i=1}^{K}\left\|Y_i-X_i\beta\right\|^2,$$

$$h_i(\beta)=\frac{1}{n_i}\sum_{j=1}^{n_i}(y_{ij}-x_{ij}'\beta)\equiv\overline{y}_i-\overline{x}_i'\beta。$$

其限制对数似然为

$$l_r(\beta,d)=-\frac{1}{2}\left\{(n-p)\ln\left[S(\beta)-\sum_{i=1}^{K}\frac{n_i^2 d h_i^2(\beta)}{1+n_i d}\right]\right.$$

$$\left.+\sum_{i=1}^{K}\ln(1+n_i d)+\ln\left|\sum_{i=1}^{K}\left(X_i'X_i-\frac{n_i^2 d}{1+n_i d}\bar{x}_i\bar{x}_i'\right)\right|\right\}\text{。}$$

对于平衡观测(即$n_i\equiv s$)，profile 对数似然有简单表达

$$l_{profile}(\beta,d)=-\frac{1}{2}\left\{Ks\ln\left[S(\beta)-\frac{s^2 d}{1+sd}\sum_{i=1}^{K}h_i^2(\beta)\right]\right.$$

$$\left.+K\ln(1+sd)\right\}$$

由$\dfrac{\partial l_{profile}(\beta,d)}{\partial d}=0$，得到$d$的 MLE(在给定$\beta$条件下)

$$d=\frac{s^2\sum_{i=1}^{K}h_i^2(\beta)-S(\beta)}{s\left[S(\beta)-s\sum_{i=1}^{K}h_i^2(\beta)\right]}$$

假设对平衡数据，这意味着$n_i\equiv s$，$X_i=X$，此时截距项具有随机效应的模型参数极大似然(ML)估计及限制极大似然(RML)估计有显示表示。

引理 7.5.1：记$X=\begin{pmatrix}E_s & U\end{pmatrix}$为$s\times p$阶列满秩矩阵，$\bar{x}=\dfrac{X'E_s}{s}$为$p\times 1$平均值向量，则有

1.$(X'X)^{-1}X'E_s=\left(1,0_{1\times(p-1)}\right)'$；

2.$X(X'X)^{-1}X'E_s=E_s$；3.$\bar{x}'(X'X)^{-1}\bar{x}=\dfrac{1}{s}$。

对此模型，β的最小二乘估计

$$\hat{\beta}_{OLS}=(X'X)^{-1}X'\bar{Y},$$

这里$\bar{Y}_{s\times 1}=\dfrac{1}{K}\sum_{i=1}^{K}Y_i$。

当d已知时，可得β的广义最小二乘估计$\hat{\beta}_{GLS}$，定义为

$$\hat{\beta}_{GLS}=Arg\min_{\beta}\sum_{i=1}^{K}(Y_i-X\beta)'\left(I_s+d\cdot E_sE_s'\right)^{-1}(Y_i-X\beta)\text{。}$$

注：通常$\hat{\beta}_{GLS}$并不是β的估计，因为含有未知的d。

定理 7.5.1：对此模型有$\hat{\beta}_{GLS}=\hat{\beta}_{OLS}$。

证明：由于

$$\begin{aligned}&\sum_{i=1}^{K}(Y_i-X\beta)'\left(I_s+d\cdot E_sE_s'\right)^{-1}(Y_i-X\beta)\\&=\sum_{i=1}^{K}(Y_i-\bar{Y})'\left(I_s+d\cdot E_sE_s'\right)^{-1}(Y_i-\bar{Y})\\&+K(\bar{Y}-X\beta)'\left(I_s+d\cdot E_sE_s'\right)^{-1}(\bar{Y}-X\beta)\end{aligned}$$

。

从而

$$\begin{aligned}\hat{\beta}_{GLS}&=Arg\min_{\beta}\sum_{i=1}^{K}(Y_i-Z\beta)'V^{-1}(Y_i-Z\beta)\\&=Arg\min_{\beta}(\bar{Y}-Z\beta)'V^{-1}(\bar{Y}-Z\beta)\end{aligned}$$

由于$(\bar{Y}-X\hat{\beta}_{OLS})'X=0$，$(\bar{Y}-X\hat{\beta}_{OLS})'E_s=0$，

$$\begin{aligned}&(\bar{Y}-X\beta)'\left(I_s-\frac{d}{1+sd}E_sE_s'\right)(\bar{Y}-X\beta)\\&=(\bar{Y}-X\hat{\beta}_{OLS})'\left(I_s-\frac{d}{1+sd}E_sE_s'\right)(\bar{Y}-X\hat{\beta}_{OLS})\\&+(X\beta-X\hat{\beta}_{OLS})'\left(I_s-\frac{d}{1+sd}E_sE_s'\right)(X\beta-X\hat{\beta}_{OLS})\end{aligned}$$

，

因此，当$\beta=\hat{\beta}_{OLS}$时达到最小值。

由此定理，对此模型有$\hat{\beta}_{GLS}=\hat{\beta}_{OLS}=\hat{\beta}_{ML}$，这里$\hat{\beta}_{ML}$表示$\beta$的极大似然估计。

前面已经推导过，在给定β,d时σ^2的极大似然(ML)估计满足

$$\begin{aligned}\sigma^2&=\frac{1}{Ks}\sum_{i=1}^{K}(Y_i-X\beta)'\left(I_s+d\cdot E_sE_s'\right)^{-1}(Y_i-X\beta)\\&=\frac{1}{Ks}\left[S(\beta)-\frac{s^2d}{1+sd}\sum_{i=1}^{K}h_i^2(\beta)\right]\end{aligned}$$

由于给定β，d的极大似然(ML)估计满足

$$d=\frac{s^2\sum_{i=1}^{K}h_i^2(\beta)-S(\beta)}{s\left[S(\beta)-s\sum_{i=1}^{K}h_i^2(\beta)\right]}=\frac{(s-1)\sum_{i=1}^{K}h_i^2(\beta)}{\left[S(\beta)-s\sum_{i=1}^{K}h_i^2(\beta)\right]}-\frac{1}{s}$$

将d代入，再将β用$\hat{\beta}_{ML}=\hat{\beta}_{OLS}$代入，得到

$$\hat{\sigma}_{ML}^2=\frac{S(\hat{\beta}_{ML})-s\sum_{i=1}^{K}h_i^2(\hat{\beta}_{ML})}{K(s-1)},$$

$$\hat{d}_{ML}=\frac{(s-1)\sum_{i=1}^{K}h_i^2(\hat{\beta}_{ML})}{\left[S(\hat{\beta}_{ML})-s\sum_{i=1}^{K}h_i^2(\hat{\beta}_{ML})\right]}-\frac{1}{s}=\frac{\sum_{i=1}^{K}h_i^2(\hat{\beta}_{ML})}{K\hat{\sigma}_{ML}^2}-\frac{1}{s}。$$

注：此特殊情形$h_i(\hat{\beta}_{ML})=\bar{y}_i-\bar{\bar{y}}$，这里$\bar{y}_i=\frac{E_s'Y_i}{s}$，$\bar{\bar{y}}=\frac{E_s'\bar{Y}}{s}$。

下面考察其限制极大似然(RML)估计。在此特殊情形下，$X_i \equiv X$

$$\begin{aligned}
&\left|\sum_{i=1}^{K}\left(X_i'X_i - \frac{s^2 d}{1+sd}\bar{x}_i\bar{x}_i'\right)\right| \\
&= \left|K(X'X - \frac{d}{1+sd}X'E_sE_s'X)\right| \\
&= \left|K(X'X)\right|\left|I_s - \frac{d}{1+sd}(X'X)^{-1}X'E_sE_s'X\right| \\
&= \left|K(X'X)\right|\left|1 - \frac{d}{1+sd}E_s'X(X'X)^{-1}X'E_s\right| \\
&= \left|K(X'X)\right|(1+sd)^{-1}
\end{aligned}$$

其限制对数似然(略去无关常数)也有简单表示

$$\begin{aligned}
l_r(\beta,d) = -\frac{1}{2}\Big\{&(Ks-p)\ln\left[S(\beta) - \frac{s^2 d}{1+sd}\sum_{i=1}^{K}h_i^2(\beta)\right] \\
&+ K\ln(1+sd) - \ln(1+sd)\Big\}
\end{aligned}$$ 。

由$\dfrac{\partial l_r(\beta,d)}{\partial d} = 0$，得到$d$的 RMLE(在给定$\beta$条件下)

$$d = \frac{[K(s-1)-p+1]\sum_{i=1}^{K}h_i^2(\beta)}{(K-1)\left[S(\beta) - s\sum_{i=1}^{K}h_i^2(\beta)\right]} - \frac{1}{s}$$ 。

由定理 7.5.1，对此模型有$\hat{\beta}_{GLS} = \hat{\beta}_{OLS} = \hat{\beta}_{ML} = \hat{\beta}_{RML}$，这里$\hat{\beta}_{RML}$表示$\beta$的限制极大似然估计。类似前面的推导，可以得到此时$\sigma^2$和$d$的限制极大似然(RML)估计如下：

$$\hat{\sigma}_{RML}^2 = \frac{S(\hat{\beta}_{RML}) - s\sum_{i=1}^{K}h_i^2(\hat{\beta}_{RML})}{K(s-1)-p+1},$$

$$\begin{aligned}
\hat{d}_{RML} &= \frac{[K(s-1)-p+1]\sum_{i=1}^{K}h_i^2(\hat{\beta}_{RML})}{(K-1)\left[S(\hat{\beta}_{RML}) - s\sum_{i=1}^{K}h_i^2(\hat{\beta}_{RML})\right]} - \frac{1}{s} \\
&= \frac{\sum_{i=1}^{K}h_i^2(\hat{\beta}_{RML})}{(K-1)\hat{\sigma}_{RML}^2} - \frac{1}{s}
\end{aligned}$$ 。

7.6 MINQUE for variance parameters

前面提到的线性混合效应模型(LMM)的ML 估计及 RML 估计，都要假定随机效应和误差的分布(通常假定正态分布)。换句话说这些估计依赖于分布假定。此外，方差参数在 LMM 中是关键，例如前面提到若$\tilde{D}$已知，则固定效应β的估计可以用广义最小二乘方法来估计。因此本节研究方差参数的估计问题。由于只涉及到模型的二阶矩，因此可以用观测的某些二次函数来估计而不涉及分布问题。

Rao(1973)提出了一种现在称为最小范数二次无偏估计(Minimum Norm Quadratic Unbiased Estimator, MINQUE)方法来估计方差参数。为简述其方法，先看标准线性模型：$Y_{n\times 1}=X_{n\times p}\beta_{p\times 1}+e_{n\times 1}$，$Ee=0$，$Cov(e)=\sigma^2 I_n$。众所周知，方差参数$\sigma^2$的无偏估计为

$$\hat{\sigma}^2=\frac{Y'\left[I_n-X(X'X)^- X'\right]Y}{n-r},$$

其中$r=rank(X)$。

假设用Y的某二次函数作为σ^2的估计

$$\hat{\sigma}^2=Y'AY,$$

这里$A_{n\times n}$为某对称矩阵。由于要作为σ^2的估计，因此要求对所有Y，$Y'AY\ge 0$，因此要求$A\ge 0$。其次，我们希望此估计是无偏的。由于$E(Y'AY)=\beta'X'AX\beta+\sigma^2 tr(A)$，要满足无偏要求则必须有

$$X'AX=0\text{且}tr(A)=1。$$

按照最小范数要求，在上面的约束下求A使得$tr(AA')=\min$。

采用 Lagrange 乘子法，定义 Lagrange 函数：

$$L(A,\Lambda_1,\lambda_2)=\frac{1}{2}tr(AA')+tr(X'AX\Lambda_1')+\lambda_2\left[1-tr(A)\right],$$

这里Λ_1为$p\times p$，λ_2为常数。

由以下公式

$$\frac{\partial tr(A)}{\partial A}=I,\quad \frac{\partial tr(AA')}{\partial A}=2A,\quad \frac{\partial tr(CAB)}{\partial A}=C'B',$$

有

$$\frac{\partial L(A,\Lambda_1,\lambda_2)}{\partial A}=A+X\Lambda_1X'-\lambda_2 I_n=0。$$

从而得$A=\lambda_2 I_n - X\Lambda_1 X'$，再由$X'AX=0$得$0=X'AX=\lambda_2 X'X - X'X\Lambda_1 X'X$，因此有$\Lambda_1=\lambda_2(X'X)^-$。代入得到$A=\lambda_2\left[I_n - X(X'X)^- X'\right]$。再由$tr(A)=1$，得到$\lambda_2=\dfrac{1}{n-rank(X)}$。故$A=\dfrac{I_n - X(X'X)^- X'}{n-rank(X)}$。可见，常见的$\hat{\sigma}^2$即为 MINQUE。

若误差为正态分布，则$Var(Y'AY)=2\sigma^4 tr(AA')$，因此最小化$tr(AA')$也意味着最小化估计$\hat{\sigma}^2$的方差。此时$Var(\hat{\sigma}^2)=\dfrac{2\sigma^4}{n-r}$。如果误差非正态分布，则$Var(Y'AY)$依赖于一些三阶和四阶矩。但直观上若$A$的元素增加$\rho$倍，其方差$Var(Y'AY)$也增加$\rho^2$倍，因此要使得$tr(AA')=\min$也是有道理的。

将上述 MINQUE 方法用于 LMM。考虑 LMM 的边际模型(见 7.1 节)

$$Y = X\beta + \eta，E\eta = 0，$$

这里$\eta = Zb + e$，$Z=\begin{pmatrix} Z_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & Z_K \end{pmatrix}$为$n\times(mK)$阶矩阵，

$$Cov(\eta)=\begin{pmatrix} Z_1DZ_1' & 0 & \cdots & 0 \\ 0 & Z_2DZ_2' & & \vdots \\ \vdots & \vdots & & 0 \\ 0 & 0 & \cdots & Z_KDZ_K' \end{pmatrix}+\sigma^2 I_n$$

$$= Z(I_K\otimes D)Z' + \sigma^2 I_n$$

先考察σ^2的 MINQUE 问题。考察二次型$Y'AY$，$A\ge 0$。此时

$$E(Y'AY)=\beta'X'AX\beta + tr\left\{A\left[Z(I_K\otimes D)Z'+\sigma^2 I_n\right]\right\}。$$

要使其为σ^2的无偏估计要求满足

$$X'AX=0，Z'AZ=0，tr(A)=1。$$

定义$n\times(p+mK)$阶矩阵W如下

$$W=\begin{pmatrix}X & Z\end{pmatrix}=\begin{pmatrix}X_1 & Z_1 & 0 & 0\\ \vdots & 0 & \ddots & 0\\ X_K & 0 & 0 & Z_k\end{pmatrix}。$$

由于假设$A\geq 0$，故

$$X'AX=0，Z'AZ=0\Leftrightarrow W'AW=0。$$

类似上面的推导，此时σ^2的 MINQUE 为

$$\hat{\sigma}^2_{MINQUE}=\frac{Y'\left[I_n-W(W'W)^-W'\right]Y}{n-rank(W)}。$$

注：分子

$$Y'\left[I_n-W(W'W)^-W'\right]Y=\min_{v\in R^{p+mK}}\|Y-Wv\|^2$$

$$=\min_{\substack{\beta\in R^p\\ \gamma_1,\gamma_2\cdots\gamma_K\in R^m}}\sum_{i=1}^{K}\|Y_i-X_i\beta-Z_i\gamma_i\|^2,$$

在正态分布假定下有

$$Var(\hat{\sigma}^2_{MINQUE})=\frac{2\sigma^4}{n-rank(W)}。$$

现在来考察D的 MINQUE 问题。由于D为$m\times m$阶矩阵，因此将D按列拉成列向量，记作$vec(D)$，$vec(D)$为$m^2\times 1$向量。构造Y的二次函数$A(Y\otimes Y)$，这里A为$m^2\times n^2$矩阵。

注：相当于把D的所有m^2元素看成未知的，实际上D的未知只有$\frac{m(m+1)}{2}$。此外并没有考虑$D\geq 0$，所以对矩阵A没有加限制。因此所得 MINQUE 不一定能保证其非负定。

利用$\otimes$对加法有分配率与结合律，有

$$E\left[A(Y\otimes Y)\right]=A(X\beta\otimes X\beta)+AE(Zb\otimes Zb)+AE(e\otimes e)。$$

再利用$\otimes$的性质

$$X\beta\otimes X\beta=(X\otimes X)(\beta\otimes\beta),$$

$$E(e\otimes e)=Evec(ee')=\sigma^2 vec(I_n),$$

$$\begin{aligned}E(Zb\otimes Zb)&=E(Z\otimes Z)(b\otimes b)\\&=(Z\otimes Z)Evec(bb')\\&=(Z\otimes Z)vec(I_n\otimes D)\end{aligned}$$

注意到$vec(I_n \otimes D)$为$vec(D)$的已知线性函数，即

$$vec(I_n \otimes D) = J \cdot vec(D),$$

这里J为$(nm)^2 \times m^2$的已知矩阵。因此要使得$E\left[A(Y \otimes Y)\right]$为$vec(D)$的无偏估计，则$A$满足

$$A(X \otimes X) = 0，\quad Avec(I_n) = 0,$$

$$A(Z \otimes Z)J = I_{m^2}。$$

因此在以上约束下，求A使得$tr(AA') = \min$。

分别定义$n^2 \times (p^2+1)$阶矩阵F及$(nm)^2 \times m^2$阶矩阵H如下：

$$F = \left[vec(I_n) \quad X \otimes X\right],$$

$$H = (Z \otimes Z)J$$

则上优化问题变为，求A使得$tr(AA') = \min$，满足以下约束

$$AF = 0,$$

$$AH = I_{m^2}。$$

构造 Lagrange 函数：

$$L(A, \Lambda_1, \Lambda_2) = \frac{1}{2} tr(AA') + tr(AF\Lambda_1') + tr\left[(I_{m_2} - AH)\Lambda_2'\right]。$$

则 $\frac{\partial L}{\partial A} = A + \Lambda_1 F' - \Lambda_2 H' = 0$，从而得到$A = \Lambda_2 H' - \Lambda_1 F'$。由$0 = AF = \Lambda_2 H'F - \Lambda_1 F'F$，得 $\Lambda_1 = \Lambda_2 H'F(F'F)^{-1}$，代入得$A = \Lambda_2 H'\left[I_{n^2m^2} - F(F'F)^{-1}F'\right]$。再由$AH = I_{m^2}$，有$\Lambda_2 = \left(H'\left[I_{n^2m^2} - F(F'F)^{-1}F'\right]H\right)^{-1}$。从而

$$A = \left(H'\left[I_{n^2m^2} - F(F'F)^{-1}F'\right]H\right)^{-1} H'\left[I_{n^2m^2} - F(F'F)^{-1}F'\right]。$$

注意到

$$F'F = \begin{pmatrix} n & vec'(X'X) \\ vec(X'X) & X'X \otimes X'X \end{pmatrix},$$

$$(F'F)^{-1} = \frac{1}{n-p} \begin{pmatrix} 1 & -vec'\left[(X'X)^{-1}\right] \\ -vec\left[(X'X)^{-1}\right] & (n-p)(X'X)^{-1} \otimes (X'X)^{-1} + vec\left[(X'X)^{-1}\right]vec'\left[(X'X)^{-1}\right] \end{pmatrix},$$

经过一系列繁琐推导$\hat{D}_{MINQUE}$有显示表达，且对于7.4,7.5小节的特殊情形有更简单表示。可以证明对 7.4 节的平衡数据其$\hat{D}_{MINQUE} = \hat{D}_{RML}$。

7.7 矩估计(Method of Moments)

上节介绍了估计方差参数D的 MINQUE，此方法并不需要对模型分布作何假定。本节来研究D的矩估计方法，同样也不需要对模型的分布作何种假定。矩估计方法(MM)的好处是，通常这样得到的估计都是无偏的且是一致的(consistent)。这里我们不妨假定σ^2的无偏估计已经给出，例如可以用上节的$\hat{\sigma}^2_{MINQUE}$。此方法最早由 Henderson(1953)提出，用于一种特殊且常见的线性混合效应模型。

称为方差分量模型(Variance components models，后面将提到)。Henderson 对此模型提出了三种方法I,II,III。由于 Henderson 方法III包括前两种。所以，本节的方法可以看成 Henderson 方法III的推广，见 Searle et.al(1992)。

矩估计方法一般有这样三个步骤。第一步，先用传统的最小二乘得到β的估计，即

$$\hat{\beta}_{OLS}=\left(\sum_{i=1}^{K}X_i'X_i\right)^{-1}\sum_{i=1}^{K}X_i'Y_i,$$

这样得到残差向量$\hat{e}_i=Y_i-X_i\hat{\beta}_{OLS},1\le i\le K$。然后利用残差向量$\hat{e}_i$来回归随机效应部分的$Z_i$，得到随机效应$b_i$的某种估计：

$$\hat{b}_i=\left(Z_i'Z_i\right)^{-}Z_i'\hat{e}_i，1\le i\le K。$$

此处可能涉及到$Z_i'Z_i$不可逆，用广义逆代替。

第二步，计算一些交叉乘积和的期望值，例如$\sum_{i=1}^{K}\hat{b}_i\hat{b}_i'$的期望。

注意到$Y_i=X_i\beta+\eta_i$，这里$\eta_i=Z_ib_i+e_i$，且$E\eta_i=0,Cov(\eta_i)=\sigma^2I_{n_i}+Z_iDZ_i'$，从而有

$$\hat{e}_i=Y_i-X_i\hat{\beta}_{OLS}=\eta_i-\left(\sum_{j=1}^{K}X_j'X_j\right)^{-1}\sum_{j=1}^{K}X_j'\eta_j。$$

为简单起见，记$V_i=\sigma^2I_{n_i}+Z_iDZ_i'$，$N=\left(\sum_{j=1}^{K}X_j'X_j\right)^{-1}$，则$\hat{e}_i=\eta_i-N\sum_{j=1}^{K}X_j'\eta_j$。

$$E\left(\hat{e}_i\hat{e}_i'\right)=E\left[\eta_i-N\sum_{j=1}^{K}X_j'\eta_j\right]\left[\eta_i-N\sum_{t=1}^{K}X_t'\eta_t\right]'$$

$$=V_i-V_iX_iNX_i'-X_iNX_i'V_i+X_iN\sum_{j=1}^{K}X_j'V_jX_jNX_i'$$

。

将$V_i=\sigma^2 I_{n_i}+Z_iDZ_i'$代入，合并整理得

$$E\left(\hat{e}_i\hat{e}_i'\right)=\sigma^2(I_{n_i}-X_iNX_i') +Z_iDZ_i'-Z_iDZ_i'X_iNX_i'-X_iNX_i'Z_iDZ_i' +X_iN\sum_{j=1}^{K}X_j'Z_jDZ_j'X_jNX_i'$$

。

记$Z_i^+=(Z_i'Z_i)^-Z_i'$，$J_i=Z_i^+Z_i=(Z_i'Z_i)^-Z_i'Z_i$（为对称矩阵，如果满秩，则$J_i=I_m$），则得到

$$E\sum_{i=1}^{K}\left(\hat{b}_i\hat{b}_i'\right)=\sigma^2\sum_{i=1}^{K}Z_i^+(I_{n_i}-X_iNX_i')Z_i'^+ +\sum_{i=1}^{K}J_iDJ_i -\sum_{i=1}^{K}J_iDZ_i'X_iNX_i'Z_i'^+ -\sum_{i=1}^{n}Z_i^+X_iNX_i'Z_iDJ_i +\sum_{i=1}^{K}Z_i^+X_iN\sum_{j=1}^{K}X_j'Z_jDZ_j'X_jNX_i'Z_i'^+$$

。

第三步，采用矩估计方法(MM)

令

$$L=\sum_{i=1}^{K}\left(\hat{b}_i\hat{b}_i'\right)-\hat{\sigma}^2\sum_{i=1}^{K}Z_i^+(I_{n_i}-X_iNX_i')Z_i'^+,$$

为统计量，这里$\hat{\sigma}^2$为σ^2的无偏估计，则

$$EL=\sum_{i=1}^{K}J_iDJ_i-\sum_{i=1}^{K}J_iDZ_i'X_iNX_i'Z_i'^+-\sum_{i=1}^{n}Z_i^+X_iNX_i'Z_iDJ_i +\sum_{i=1}^{K}Z_i^+X_iN\sum_{j=1}^{K}X_j'Z_jDZ_j'X_jNX_i'Z_i'^+$$

。

从而D的矩估计$\hat{D}_{MM}$，可以解如下矩阵方程：

$$L=\sum_{i=1}^{K}J_i\hat{D}_{MM}J_i-\sum_{i=1}^{K}J_i\hat{D}_{MM}Z_i'X_iNX_i'Z_i'^+-\sum_{i=1}^{n}Z_i^+X_iNX_i'Z_i\hat{D}_{MM}J_i +\sum_{i=1}^{K}Z_i^+X_iN\sum_{j=1}^{K}X_j'Z_j\hat{D}_{MM}Z_j'X_jNX_i'Z_i'^+$$

。

记$R_{ij}=Z_i^+X_iNX_j'Z_j$，则上方程变为

$$L=\sum_{i=1}^{K}J_i\hat{D}_{MM}J_i-\sum_{i=1}^{K}J_i\hat{D}_{MM}R_{ii}'-\sum_{i=1}^{n}R_{ii}\hat{D}_{MM}J_i+\sum_{i=1}^{K}\sum_{j=1}^{k}R_{ij}\hat{D}_{MM}R_{ij}'$$

。

两边取vec，从而得到$vec(\hat{D}_{MM})=F^{-1}vec(L)$，这里

$$F=\sum_{i=1}^{K}J_i\otimes J_i-\sum_{i=1}^{K}J_i\otimes R_{ii}-\sum_{i=1}^{n}R_{ii}\otimes J_i+\sum_{i=1}^{K}\sum_{j=1}^{k}R_{ij}\otimes R_{ij}$$

。

如果对所以i，$Z_i'Z_i \ge aI_m$，则

$$tr(\sum_{i=1}^{K} Z_i^+ X_i N X_i' Z_i'^+) = tr\sum_{i=1}^{K} N X_i' Z_i'^+ Z_i^+ X_i$$
$$= tr\sum_{i=1}^{K} N X_i' (Z_i'Z_i)^{-1} X_i \le \frac{p}{a},$$

可以证明$\frac{1}{K}\sum_{i=1}^{K} N X_i' Z_i'^+ Z_i^+ X_i = o_p(1)$。同理可以证明$\sum_{i=1}^{K} R_{ii}, \sum_{i=1}^{K}\sum_{j=1}^{K} R_{ij}$有界，从而

$$K^{-1}F - I_m \otimes I_m = o_p(1)$$

$$K^{-1}L = K^{-1}\sum_{i=1}^{K}\left(\hat{b}_i\hat{b}_i'\right) - \hat{\sigma}^2 K^{-1}\sum_{i=1}^{K}(Z_i'Z_i)^{-1} + o_p(1),$$

故

$$\hat{D}_{MM} = K^{-1}\sum_{i=1}^{K}\left(\hat{b}_i\hat{b}_i'\right) - \hat{\sigma}^2 K^{-1}\sum_{i=1}^{K}(Z_i'Z_i)^{-1} + o_p(1)。$$

这也很好解释，因为当$K \to \infty$时，$\hat{\beta}_{OLS}$趋于真实的β，从而对$1 \le i \le K$

$$E\left(\hat{b}_i\hat{b}_i'\right) \approx E\left(Z_i^+ \eta_i \eta_i' Z_i'^+\right) = \sigma^2 (Z_i'Z_i)^{-1} + D。$$

7.8 随机效应的估计

先假定模型参数β, σ^2, D已知，来对第i个个体的随机效应b_i来进行估计。考虑条件期望$E\left(b_i \middle| Y_i\right)$，在正态分布假定下，由于

$$\begin{pmatrix} b_i \\ Y_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ X_i\beta \end{pmatrix} \begin{pmatrix} D & DZ_i' \\ Z_i D & \sigma^2 I_{n_i} + Z_i D Z_i' \end{pmatrix}\right),$$

从而

$$E\left(b_i \middle| Y_i\right) = DZ_i'(\sigma^2 I_{n_i} + Z_i D Z_i')^{-1}(Y_i - X_i\beta)$$

因此，假定方差参数$\tilde{D} = \frac{D}{\sigma^2}$已知，则$\beta$可用广义最小二乘$\hat{\beta}$来估计，从而有$b_i$的估计：

$$\hat{b}_i = \tilde{D}Z_i'(I_{n_i} + Z_i\tilde{D}Z_i')^{-1}(Y_i - X_i\hat{\beta})。$$

由于$Z_i'(I_{n_i} + Z_i\tilde{D}Z_i')^{-1} = (I_m + Z_i'Z_i\tilde{D})^{-1}Z_i'$，因此

$$\hat{b}_i = \tilde{D}(I_m + Z_i'Z_i\tilde{D})^{-1}Z_i'(Y_i - X_i\hat{\beta})。$$

如果$D = 0$，没有随机效应，估计$\hat{b}_i = 0$。上式也是最优线性无偏预测(Best Linear Unbiased Predictor, BLUP)，见 Henderson(1963)。

考虑混合效应模型

$$Y = X\beta + Zb + e,$$

这里

$Cov(e) = \sigma^2 I_n$，$Cov(b) = \sigma^2 (I_K \otimes \tilde{D}) \equiv \sigma^2 H$，$Cov(Y) = \sigma^2 (I_n + ZHZ') \equiv \sigma^2 \Sigma$。

考察b的线性无偏预测$\hat{b} = CY$，由无偏性有$CX = 0$。此外希望$E(\hat{b}-b)'(\hat{b}-b) = \min$。由于

$$\hat{b} - b = CY - b = (CZ - I)b + Ce,$$

$$E(\hat{b}-b)'(\hat{b}-b) = trCov(\hat{b}-b) = \sigma^2 tr\left[CC' + (CZ-I)H(CZ-I)'\right]。$$

即在约束$CX = 0$下，使得

$$tr\left[CC' + (CZ-I)H(CZ-I)'\right] = \min。$$

构造 Lagrange 函数

$$L(C,\Lambda) = \frac{1}{2} tr\left[CC' + (CZ-I)H(CZ-I)'\right] - tr(CX\Lambda'),$$

由 $\frac{\partial L}{\partial C} = C + (CZ - I)HZ' - \Lambda X' = 0$，得 $C = (HZ' + \Lambda X')(I + ZH'Z)^{-1} = (HZ' + \Lambda X')\Sigma^{-1}$。

再由$CX = 0$得

$$\Lambda = -HZ'\Sigma^{-1} X (X'\Sigma^{-1} X)^{-1},$$

因此，

$$\begin{aligned} C &= \left[HZ' - HZ'\Sigma^{-1} X (X'\Sigma^{-1} X)^{-1} X'\right]\Sigma^{-1} \\ &= HZ'\Sigma^{-1}\left[I - X (X'\Sigma^{-1} X)^{-1} X'\Sigma^{-1}\right] \end{aligned}。$$

从而

$$\begin{aligned} \hat{b} &= HZ'\Sigma^{-1}\left[I - X (X'\Sigma^{-1} X)^{-1} X'\Sigma^{-1}\right]Y \\ &= HZ'\Sigma^{-1}(Y - X\hat{\beta}) \end{aligned}$$

这里$\hat{\beta}$为(H，即Σ已知)广义最小二乘估计。

将$H = I_K \otimes \tilde{D}$代入，考察第$i$块有

$$\hat{b}_i = \tilde{D} Z_i' (I_{n_i} + Z_i \tilde{D} Z_i')^{-1} (Y_i - X_i \hat{\beta}),$$

此式与前面假定正态分布时推导一样。

此外，还可以采用下面的优化问题方式同时得到固定效应β及随机效应b_i的估计：

$$\min_{\beta, b_1, b_2, \cdots b_K} \sum_{i=1}^{K} \left(\|Y_i - X_i\beta - Z_i b_i\|^2 + b_i' \tilde{D} b_i \right)。$$

这是因为，给定β，当$b_i = (Z_i' Z_i + \tilde{D}^{-1})^{-1} Z_i' (Y_i - X_i\beta)$时达到最小。

再将b_i代入，得到关于β的优化问题

$$\min_{\beta}\sum_{i=1}^{K}(Y_i-X_i\beta)'G_i(Y_i-X_i\beta),$$

这里

$$G_i=\left[I_{n_i}-(Z_i'Z_i+\tilde{D}^{-1})^{-1}Z_i'\right]\left[I_{n_i}-(Z_i'Z_i+\tilde{D}^{-1})^{-1}Z_i'\right]+Z_i(Z_i'Z_i+\tilde{D}^{-1})^{-1}\tilde{D}^{-1}(Z_i'Z_i+\tilde{D}^{-1})^{-1}Z_i'。$$

经化简有$G_i=(I_{n_i}+Z_i\tilde{D}Z_i')^{-1}$，这意味着当$\beta$为($\tilde{D}$已知)广义最小二乘估计时达到最小。

7.9 方差分量模型(Variance Component Models)

前面已经提到过，线性混合效应模型的一般形式可以写成

$$Y=X\beta+Zb+e,$$

Y为$n\times1$观测向量，$X_{n\times p}$为已知设计矩阵，$\beta_{p\times1}$未知为固定效应，$Z_{n\times q}$为已知设计矩阵，$b_{q\times1}$为随机效应，且设$Eb=0$，$Cov(b)=G$非负定，e为随机误差且与b独立，$Ee=0$，$Cov(e)=R$为正定矩阵。$Cov(Y)=ZGZ'+R$。，

前面几节，我们主要关注由 Laird & Ware(1982)发展起来的一类混合效应模型。这类混合效应模型针对每个个体建模，相当于$Z=diag(Z_1,Z_2,\cdots Z_K)$是对角块，其中$Z_i$为$n_i\times m$阶矩阵，$q=mK$，随机效应协方差阵$Cov(b)$也是对角块$diag(D_{m\times m},\underbrace{D_{m\times m}\cdots D_{m\times m}}_{K个})$，即$G=I_K\otimes D_{m\times m}$。为简单起见假定$Cov(e)=R=\sigma^2I_n$，这里$n=\sum_{i=1}^{K}n_i$。

比这个更早的一类线性混合效应模型，其形式表现为随机效应部分Zb可以表示成r个互不相关(通常假定独立)的随机因子的叠加，即

$$Zb=\sum_{j=1}^{r}Z_jb_j,$$

这里每个Z_j为$n\times q_j$的已知矩阵，b_j为$q_j\times1$阶矩阵。记$q=\sum_{j=1}^{r}q_j$，通常假定$Eb_j=0$，$Cov(b_j)=\sigma_j^2I_{q_j}$。从而有：

$$Cov(Y)=\sum_{j=1}^{r}\sigma_j^2 Z_j Z_j' + R。$$

此模型将Y的波动(方差)分解为r个随机因子的波动和误差波动的叠加，因此称为方差分量模型(Variance Component Models)。相对于标准的线性混合效应模型来说，相当于

$Z_{n\times q}=\begin{pmatrix} Z_1 & Z_2 & \cdots & Z_r \end{pmatrix}$，$b_{q\times 1}=\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_r \end{pmatrix}$，

$Cov(b)=diag\begin{pmatrix} \sigma_1^2 I_{q_1} & \sigma_2^2 I_{q_2} & \cdots & \sigma_r^2 I_{q_r} \end{pmatrix}$。

为简单起见，以下不特别申明考虑$R=\sigma^2 I_n$。

为表示简单，记$\sigma_0^2=\sigma^2$，$Z_0=I_n$，则此时

$$\begin{aligned} Cov(Y) &= \sum_{j=1}^{r}\sigma_j^2 Z_j Z_j' + \sigma^2 I_n \\ &= \sum_{j=0}^{r}\sigma_j^2 Z_j Z_j' \equiv V(\theta) \end{aligned},$$

这里$\theta=\left(\sigma_0^2,\sigma_1^2,\sigma_2^2,\cdots\sigma_r^2\right)'$。

例 7.9.1：单向分类随机效应模型

考虑模型 $y_{ij}=\mu+\alpha_i+e_{ij}$，$j=1,2\cdots n_i$，$i=1,2,\cdots a$，这里$\mu$为固定效应，$\alpha_i$为随机效应。设$\alpha_i\ i.i.d\sim N(0,\sigma_\alpha^2)$，$e_{ij}\ i.i.d\sim N(0,\sigma^2)$。

记 $n=\sum_{i=1}^{a}n_i$，$Y_{n\times 1}=(y_{11},\cdots y_{1n_1},\cdots y_{a1},\cdots y_{an_a})'$，$b_{a\times 1}=(\alpha_1,\alpha_2,\cdots\alpha_a)'$，$e_{n\times 1}=(e_{11},\cdots e_{1n_1},\cdots e_{a1},\cdots e_{an_a})'$，则写成标准形式(相当于$r=1$)

$$Y=E_n\mu+Zb+e,$$

这里，$Z_{n\times a}=\begin{pmatrix} E_{n_1} & & & \\ & E_{n_2} & & \\ & & \ddots & \\ & & & E_{n_a} \end{pmatrix}$。

$Cov(Y)=\sigma_a^2 diag(E_{n_1}E_{n_1}',\cdots E_{n_a}E_{n_a}')+\sigma^2 I_n$。

$V(\theta)=\theta_0 I_n+\theta_1 diag(E_{n_1}E_{n_1}',\cdots E_{n_a}E_{n_a}')$，其中$\theta_0=\sigma^2$，$\theta_1=\sigma_\alpha^2$且$X=E_n$。

对方差分量这类线性混合效应模型来说，前面关于参数估计的方法也都可以用于此模型。具体地，假定随机效应及误差为正态分布。此时对数似然函数为：

$$l=-\frac{n}{2}\ln(2\pi)-\frac{1}{2}\log\left|V(\theta)\right| -\frac{1}{2}(Y-X\beta)'V^{-1}(\theta)(Y-X\beta)$$ 。

$$l_\beta=\frac{\partial l}{\partial\beta}=X'V^{-1}(\theta)Y-X'V^{-1}(\theta)X\beta,$$

对 $j=0,1\cdots r$，有

$$l_{\theta_j}=\frac{\partial l}{\partial\theta_j}=-\frac{1}{2}tr\left[V^{-1}(\theta)Z_jZ_j'\right] +\frac{1}{2}(Y-X\beta)'V^{-1}(\theta)Z_jZ_j'V^{-1}(\theta)(Y-X\beta)$$

令上面偏导数为零就得到似然方程组。

此外，还有另外一种等价的表示。

令

$$P(\theta)=V^{-1}(\theta)-V^{-1}(\theta)X\left[X'V^{-1}(\theta)X\right]^{-}X'V^{-1}(\theta),$$

由 β 的方程得

$$X'V^{-1}(\theta)X\beta=X'V^{-1}(\theta)Y,$$

也可写成

$$V^{-1}(\theta)(Y-X\beta)=P(\theta)Y。$$

关于 θ 的方程，对 $j=0,1\cdots r$，有

$$tr\left[V^{-1}(\theta)Z_jZ_j'\right]=Y'P(\theta)Z_jZ_j'P(\theta)Y。$$

可见，后面关于方差参数的 $r+1$ 个方程与 β 无关，可以解出 θ，然后得到 β 的估计。

以上我们得到参数的极大似然(ML)估计。同前面的，也可以采用限制极大似然(RML)方法得到参数的 RMLE。设 $rank(X_{n\times p})=r$，则存在 $B_{n\times(n-r)}$，$rank(B)=n-r$，使得 $B'X=0$。由 $Y\sim N(X\beta,V(\theta))$，则 $B'Y\sim N(0,B'V(\theta)B)$，这样由 BY 的似然函数(限制似然函数)，可以得到其似然方程，这只需要在前面的关于 θ 的似然方程里 $Y\to B'Y$，$Z\to B'Z$，$X\to B'X=0$，$V(\theta)\to B'V(\theta)B$，$P(\theta)\to\left(B'V(\theta)B\right)^{-1}$ 代替即可。

对 $j=0,1\cdots r$，有

$$tr\left[\left(B'V(\theta)B\right)^{-1}B'Z_jZ_j'B\right] = Y'B\left(B'V(\theta)B\right)^{-1}B'Z_jZ_j'B\left(B'V(\theta)B\right)^{-1}B'Y \text{。}$$

令 $M(\theta)=B\left(B'V(\theta)B\right)^{-1}B'$，则上方程组写为对 $j=0,1\cdots r$，

$$tr\left[M(\theta)Z_jZ_j'\right]=Y'M(\theta)Z_jZ_j'M(\theta)Y \text{。}$$

以下说明，$M(\theta)$ 并不依赖于 B 的选择，其恒等于前面的 $P(\theta)$。即，限制极大似然方程为 $tr\left[P(\theta)Z_jZ_j'\right]=Y'P(\theta)Z_jZ_j'P(\theta)Y$，$j=0,1\cdots r$。

注： 由于 $\left[V^{-\frac{1}{2}}(\theta)X\right]'V^{\frac{1}{2}}(\theta)B=0$， 因此 $rank\left(V^{-\frac{1}{2}}(\theta)X \quad V^{\frac{1}{2}}(\theta)B\right)=rank(X)+rank(B)=n$，因此

$$\begin{aligned} I_n &= P_{V^{-\frac{1}{2}}(\theta)X}+P_{V^{\frac{1}{2}}(\theta)B} \\ &= V^{-\frac{1}{2}}(\theta)X\left[X'V^{-1}(\theta)X\right]^{-}X'V^{-\frac{1}{2}}(\theta) \\ &\quad +V^{\frac{1}{2}}(\theta)B\left[B'V^{-1}(\theta)B\right]^{-1}B'V^{\frac{1}{2}}(\theta) \end{aligned}$$

即有 $M(\theta)\equiv P(\theta)$。

例 7.9.2：平衡数据单向分类随机效应模型

考虑模型 $y_{ij}=\mu+\alpha_i+e_{ij}$， $j=1,2\cdots b$，$i=1,2,\cdots a$，这里 μ 为固定效应，α_i 为随机效应。设 $\alpha_i\ i.i.d \sim N(0,\sigma_\alpha^2)$，$e_{ij}\ i.i.d \sim N(0,\sigma^2)$。记 $n=ab$，$Y_{n\times1}=(y_{11},\cdots y_{1b},\cdots y_{a1},\cdots y_{ab})'$，$\alpha_{a\times1}=(\alpha_1,\alpha_2,\cdots\alpha_a)'$，$e_{n\times1}=(e_{11},\cdots e_{1b},\cdots e_{a1},\cdots e_{ab})'$，则写成标准形式(相当于 $r=1$)

$$Y=X\mu+Z\alpha+e,$$

这里，$X_{n\times1}=E_a\otimes E_b=E_{ab}$，$Z_{n\times a}=I_a\otimes E_b$。

$Cov(Y)=\sigma_\alpha^2I_a\otimes E_bE_b'+\sigma^2I_{ab}$。

$V(\theta)=\theta_0I_{ab}+\theta_1I_a\otimes E_bE_b'$，其中 $\theta_0=\sigma^2$，$\theta_1=\sigma_\alpha^2$。

$$V^{-1}(\theta)=\theta_0^{-1}I_{ab}-\frac{\theta_0^{-1}\theta_1}{\theta_0+\theta_1 b}I_a\otimes E_bE_b',$$

$$V^{-1}(\theta)X=\frac{E_{ab}}{\theta_0+\theta_1 b},\quad X'V^{-1}(\theta)X=\frac{ab}{\theta_0+\theta_1 b}\text{。}$$

由 μ 的似然方程得

$$\mu=\bar{y}_{..}\text{。}$$

$$V^{-2}(\theta)=\theta_0^{-2}I_{ab}-\frac{2\theta_0^{-1}\theta_1+\theta_0^{-2}\theta_1^2 b}{(\theta_0+\theta_1 b)^2}I_a\otimes E_b E_b'$$

$$trV^{-1}(\theta)=ab\theta_0^{-1}(1-\frac{\theta_1}{\theta_0+\theta_1 b})$$

$$V^{-1}(\theta)\,\mathrm{Z}=\frac{1}{\theta_0+\theta_1 b}I_a\otimes E_b$$

$$V^{-1}(\theta)\,\mathrm{ZZ}'=\frac{1}{\theta_0+\theta_1 b}I_a\otimes E_b E_b'$$

$$V^{-1}(\theta)\,\mathrm{ZZ}'V^{-1}(\theta)=\frac{1}{(\theta_0+\theta_1 b)^2}I_a\otimes E_b E_b'$$

θ_0,θ_1的似然方程组为：

$$ab\theta_0^{-1}(1-\frac{\theta_1}{\theta_0+\theta_1 b})=\theta_0^{-2}(Y-X\mu)'(Y-X\mu)-\frac{2\theta_0^{-1}\theta_1+\theta_0^{-2}\theta_1^2 b}{(\theta_0+\theta_1 b)^2}(Y-X\mu)'(I_a\otimes E_b E_b')(Y-X\mu),$$

$$\frac{ab}{\theta_0+\theta_1 b}=(Y-X\mu)'\frac{(I_a\otimes E_b E_b')}{(\theta_0+\theta_1 b)^2}(Y-X\mu)。$$

由第二式，第一式化简为

$$\theta_0+\theta_1=\frac{(Y-X\mu)'(Y-X\mu)}{ab}。$$

将$\mu=\overline{y}_{..}$带入，得

$$\theta_0+\theta_1=\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\overline{y}_{..})^2,$$

$$\theta_0+\theta_1 b=\frac{b}{a}\sum_{i=1}^{a}(\overline{y}_{i.}-\overline{y}_{..})^2。$$

故得

$$\hat{\mu}=\overline{y}_{..},$$

$$\hat{\theta}_0=\frac{1}{a(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\overline{y}_{i.})^2,$$

$$\hat{\theta}_1=\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\overline{y}_{..})^2-\hat{\theta}_0$$

$$=\frac{1}{a}\sum_{i=1}^{a}(\overline{y}_{i.}-\overline{y}_{..})^2-\frac{\hat{\theta}_0}{b}$$

注：显然，似然方程组的解$\hat{\theta}_1$有可能取负值。因此，似然方程组的解不总是 ML 估计。当$\hat{\theta}_1$取负值时，方程组的解没有落到参数空间里。此时，似然函数的最大值在其边界$\theta_1=0$达到。即θ_1的 ML 估计为$\max\{\hat{\theta}_1,0\}$。

此外，对方差参数部分限制极大似然方程

$$tr\left[P(\theta)Z_jZ_j'\right]=Y'P(\theta)Z_jZ_j'P(\theta)Y，\quad j=0,1\cdots r。$$

还有另外一种表示。注意到由$P(\theta)$的定义有

$$P(\theta)V(\theta)P(\theta)=P(\theta)。$$

从而

$$\begin{aligned}tr\left[P(\theta)Z_jZ_j'\right]&=tr\left[P(\theta)V(\theta)P(\theta)Z_jZ_j'\right]\\&=tr\left[P(\theta)Z_jZ_j'P(\theta)V(\theta)\right]\\&=\sum_{i=0}^{r}tr\left[P(\theta)Z_jZ_j'P(\theta)Z_iZ_i'\right]\theta_i\end{aligned}。$$

这样，极大似然方程可以写为：

$$\left(tr\left[P(\theta)Z_iZ_i'P(\theta)Z_jZ_j'\right]\right)_{i,j=0}^{r}\theta=\begin{pmatrix}Y'P(\theta)Z_0Z_0'P(\theta)Y\\\vdots\\Y'P(\theta)Z_rZ_r'P(\theta)Y\end{pmatrix}。$$

注：上面这种形式可以在求解时构造θ的迭代求解——Anderson 迭代方法。

方差分量模型参数的 ML 估计及 RML 估计都要假定随机效应和误差的分布(通常假定正态分布)。换句话说这些估计依赖于分布假定。

与前面所说的混合效应模型一样，方差分量模型方差参数θ在模型中是关键。若θ已知，则固定效应β的估计可以用广义最小二乘方法来估计。接下来对于方差分量模型的方差参数介绍两种估计方法：ANOVA 估计方法及 MINQUE 方法。这些方法只涉及到模型的矩，而不涉及到分布问题。

设方差分量模型为：

$$Y=X\beta+\sum_{i=1}^{r}Z_ib_i+e。$$

为表示简单，记$\sigma_0^2=\sigma^2$，$Z_0=I_n$，$b_0=e$则此时模型为$Y=X\beta+\sum_{i=0}^{r}Z_ib_i$

$$Cov(Y)=\sum_{j=0}^{r}\sigma_j^2Z_jZ_j'\equiv V(\theta)，$$

这里$\theta=\left(\sigma_0^2,\sigma_1^2,\sigma_2^2,\cdots\sigma_r^2\right)'$。

假设有 $r+1$ 个二次型 $s_i = Y'A_iY$（假定 $A_i \ge 0$），$i=0,1,\cdots,r$。则

$$Es_i = EY'A_iY$$
$$= tr\left[A_iV(\theta)\right] + \beta' X'A_iX\beta 。$$

若 A_i 满足 $X'A_iX = 0$。这样有

$$Es_i = EY'A_iY = tr\left[A_iV(\theta)\right]$$
$$= tr\left[A_i\sum_{j=0}^{r}\theta_j Z_jZ_j'\right] = \sum_{j=0}^{r} tr(Z_j'A_iZ_j)\theta_j 。$$

记 $S=(s_0,s_1,\cdots s_r)'$，$C=\left(C_{ij}=tr(Z_j'A_iZ_j)\right)$，$0\le i,j\le r$，则写成矩阵形式有：

$$ES = C\theta 。$$

从而由矩估计方法得到 θ 的估计 $\hat\theta$

$$S = C\hat\theta ,$$

即

$$\hat\theta = C^{-1}S ,$$

若 C 可逆。

这样，一个关键的问题是：满足 $X'A_iX=0$

$i=0,1,\cdots,r$ 的这 $r+1$ 个 $\{A_i\}$ 是否存在？Henderson 指出，可以选取

$$A_0 = I_n - P_{(X,Z_1,\cdots Z_r)} ,$$

对 $i=1,2,\cdots,r$

$$A_i = \left[I_n - P_{(X,Z_1,\cdots Z_{i-1})}\right] - \left[I_n - P_{(X,Z_1,\cdots Z_i)}\right]$$
$$= P_{(X,Z_1,\cdots Z_i)} - P_{(X,Z_1,\cdots Z_{i-1})} 。$$

注：之所以称为 ANOVA 方法是因为也是利用 Y 的一些平方和来估计。

上面的方法是刚好选取 $r+1$ 个，C 是方阵。事实上也可以超过 $r+1$ 个，此时 C 不再是方阵，则 θ 的估计 $\hat\theta$ 为最小二乘解：

$$\hat\theta = (C'C)^{-1}C'S 。$$

再转到 MINQUE 方法。考察方差分量的已知线性函数 $c'\theta$，考虑用某二次型 $Y'AY$ 来估计 $c'\theta$，要求 A 对称且满足 $AX=0$。此时

$$EY'AY = tr\left[AV(\theta)\right] = \sum_{j=0}^{r} tr(Z_j'AZ_j)\theta_j ,$$

要满足无偏性，所以对 $j=0,1,\cdots,r$

$$tr(Z_j'AZ_j)=c_j。$$

另一方面，若对 $j=0,1,\cdots,r$，b_j都已知(为$q_j\times 1$向量)，则θ_j的估计应该用$\frac{b_j'b_j}{q_j}$，从而$c'\theta$的估计为 $\sum_{j=0}^{r}c_j\frac{b_j'b_j}{q_j}=b'\Delta b$，这里 $\Delta=diag\left(\frac{c_0}{q_0}I_{q_0},\cdots,\frac{c_r}{q_r}I_{q_r}\right)$，$b=\left(b_0',b_1',\cdots,b_r'\right)'$。

记$Z=\left(Z_0,Z_1,\cdots,Z_r\right)$，则$Y=X\beta+Zb$，从而有$Y'AY=b'Z'AZb$（假设$AX=0$）。从而欲使得$Y'AY$是好的估计，对一切$b$有$b'Z'AZb$与$b'\Delta b$很接近。若用某种范数来度量，则选择$A$使得$\|Z'AZ-\Delta\|=\min$。注意到$A$满足以下条件：

$$AX=0$$

$$j=0,1,\cdots,r，tr(Z_j'AZ_j)=c_j。$$

对已知线性函数$c'\theta$，若估计$Y'AY$满足上约束条件且使得$\|Z'AZ-\Delta\|=\min$，则称$Y'AY$为$c'\theta$的最小范数二次无偏估计(MINQUE)。

C.R.Rao 选择加权欧式范数。设权矩阵(已知)

$$W=diag(w_0^2I_{q_0},w_1^2I_{q_1},\cdots w_r^2I_{q_r})，$$

令

$$F=W^{\frac{1}{2}}(Z'AZ-\Delta)W^{\frac{1}{2}}$$

则加权范数$\|Z'AZ-\Delta\|^2=tr(F'F)$。利用矩阵$A$满足的约束条件，可得

$$tr(F'F)=tr(AV(\theta_W))^2-tr(\Delta W)^2，$$

这里$V(\theta_W)=\sum_{j=0}^{r}w_j^2Z_jZ_j'>0$，$\theta_W=\left(w_0^2,w_1^2,w_2^2,\cdots w_r^2\right)'$。

从而，MINQUE 估计问题就归结为求下述极值问题：

$$\min tr(AV(\theta_W))^2$$

$$\begin{cases}AX=0,\\ tr(AZ_iZ_i')=c_i,i=0,1,\cdots,r\end{cases}。$$

定理 7.9.1：上极值问题的解为

$$A^*=\sum_{j=0}^{r}\lambda_iP(\theta_W)Z_iZ_i'P(\theta_W)，$$

其中

$P(\theta_W)=V^{-1}(\theta_W)-V^{-1}(\theta_W)X\left[X'V^{-1}(\theta_W)X\right]^{-}X'V^{-1}(\theta_W)$，

$\lambda=(\lambda_0,\lambda_1,\cdots,\lambda_r)'$满足矩阵方程

$$\left(tr\left[P(\theta_W)Z_iZ_i'P(\theta_W)Z_jZ_j'\right]\right)_{i,j=0}^{r}\lambda=c\text{。}$$

这样，由上定理 $c'\theta$ 的 MINQUE 为 $c'\hat{\theta}=Y'A^{*}Y$。事实上这里$\hat{\theta}$为下方程组的解：

$$\left(tr\left[P(\theta_W)Z_iZ_i'P(\theta_W)Z_jZ_j'\right]\right)_{i,j=0}^{r}\theta=\begin{pmatrix}Y'P(\theta_W)Z_0Z_0'P(\theta_W)Y\\ \vdots\\ Y'P(\theta_W)Z_rZ_r'P(\theta_W)Y\end{pmatrix}$$

注：由$c'\hat{\theta}=Y'A^{*}Y$，有

$$c'\hat{\theta}=\lambda'\left(tr\left[P(\theta_W)Z_iZ_i'P(\theta_W)Z_jZ_j'\right]\right)_{i,j=0}^{r}\hat{\theta},$$

$$Y'A^{*}Y=\sum_{j=0}^{r}\lambda_iY'P(\theta_W)Z_iZ_i'P(\theta_W)Y$$

$$=\lambda'\begin{pmatrix}Y'P(\theta_W)Z_0Z_0'P(\theta_W)Y\\ \vdots\\ Y'P(\theta_W)Z_rZ_r'P(\theta_W)Y\end{pmatrix},$$

再由c的任意性，可得。