

階層的回帰分析について

前川眞一

大学入試センター 研究開発部

1 モデル

1.1 記号

$$\begin{aligned}
 p &= \text{説明変数の数} \\
 m &= \sum_{L=1}^b m_L = \text{総グループ数} \\
 b &= \text{総ブロック数} \\
 m_L &= \text{ブロック } L \text{ のグループ数} \\
 n_j &= \text{グループ } j \text{ の観測値数} \\
 n &= \sum_{j=1}^m n_j = \text{総観測値数} \\
 (\mathbf{y}_j, \mathbf{X}_j) &= \text{第 } j \text{ グループのデータ、それぞれ } n_j \times 1, n_j \times p
 \end{aligned} \tag{1}$$

1.2 2層モデル Simple m-group Regression

$$\mathbf{y}_j \sim N(\mathbf{X}_j \boldsymbol{\beta}_j, \sigma_j^2 \mathbf{I}), \quad j = 1, 2, \dots, m, \quad \text{indep.}, \tag{2}$$

$$\boldsymbol{\beta}_j \stackrel{iid}{\sim} N(\boldsymbol{\gamma}, \boldsymbol{\Psi}) \tag{3}$$

ベイズ推定における事前分布

$$\boldsymbol{\gamma} \sim N(\boldsymbol{\xi}, \boldsymbol{\Phi}) \tag{4}$$

1.3 3層モデル

$$\mathbf{y}_j \sim N(\mathbf{X}_j \boldsymbol{\beta}_j, \sigma_j^2 \mathbf{I}), \quad j = 1, 2, \dots, m, \quad \text{indep.}, \tag{5}$$

$$\boldsymbol{\beta}_j \stackrel{iid}{\sim} N(\boldsymbol{\gamma}_L, \boldsymbol{\Psi}_L), \quad \text{if } j \in L, \tag{6}$$

$$\boldsymbol{\gamma}_L \stackrel{iid}{\sim} N(\boldsymbol{\xi}, \boldsymbol{\Phi}), \quad L = 1, 2, \dots, b \tag{7}$$

ベイズ推定における事前分布

$$\boldsymbol{\xi} \sim N(\boldsymbol{\eta}, \boldsymbol{\Omega}) \tag{8}$$

2層モデルは3層モデルにおいて以下の制約をおいたものと同等となる。

$$\text{MLF: } L = 1, \quad \Phi = \mathbf{O}, \quad \xi = \gamma \quad (9)$$

もしくは

$$\text{MLR: } L = 1, \quad \Phi^{-1} = \mathbf{O} \quad (10)$$

1.4 まとめて書くと

$$\mathbf{y}|\beta, \sigma \sim N(\mathbf{X}\beta, \Sigma_y), \quad \text{この密度を } f_0(\mathbf{y}|\beta, \Sigma_y) \text{ と呼ぶ} \quad (11)$$

$$\beta|\gamma, \Psi \sim N(\mathbf{W}\gamma, \mathbf{C}_2), \quad \text{この密度を } h_2(\beta|\gamma, \mathbf{C}_2) \text{ と呼ぶ} \quad (12)$$

$$\gamma|\xi, \Phi \sim N(\mathbf{G}\xi, \mathbf{C}_3), \quad \text{この密度を } h_3(\gamma|\xi, \mathbf{C}_3) \text{ と呼ぶ} \quad (13)$$

ただし、

$$\mathbf{y} = \mathbf{y}_{(j)} \text{ を縦に並べたもの } \quad n \times 1$$

$$\mathbf{X} = \text{Diag}(\mathbf{X}_j)\mathbf{X}_j \text{ をブロック対角に並べたもの } \quad n \times mp$$

$$\beta = \beta_j \text{ を縦に並べたもの } \quad mp \times 1$$

$$\gamma = \gamma_L \text{ を縦に並べたもの } \quad bp \times 1$$

$$\xi = p \times 1$$

$$\mathbf{W} = \text{Diag}(\mathbf{1}_{m_L} \otimes \mathbf{I}_p), \quad mp \times bp$$

$$\mathbf{G} = \mathbf{1}_b \times \mathbf{I}_p, \quad bp \times p$$

$$\Sigma_y = \text{Diag}(\sigma_j^2 \mathbf{I}_{n_j}), \quad n \times n$$

$$\mathbf{C}_1 = \text{Diag}(\sigma_j^2 (\mathbf{X}'_j \mathbf{X}_j)^{-1}), \quad mp \times mp$$

$$\mathbf{C}_2 = \text{Diag}(\mathbf{I}_{m_L} \otimes \Psi_L), \quad mp \times mp$$

$$\mathbf{C}_3 = \mathbf{I}_b \otimes \Phi, \quad bp \times bp$$

データならびにランダム成分（ベイズではパラメタ） β, γ の同時分布は

$$f(\mathbf{y}, \beta, \gamma|\xi, \sigma, \Psi, \Phi) = f_0(\mathbf{y}|\beta, \Sigma_y) \times h_2(\beta|\gamma, \mathbf{C}_2) \times h_3(\gamma|\xi, \mathbf{C}_3) \quad (14)$$

である。これは線形の正規モデルであるから、周辺分布は

$$\mathbf{y}|\gamma, \sigma, \Psi \sim N(\mathbf{X}\mathbf{W}\gamma, \mathbf{X}\mathbf{C}_2\mathbf{X}' + \Sigma_y) \quad (15)$$

$$\mathbf{y}|\sigma, \Psi, \xi, \Phi \sim N(\mathbf{X}\mathbf{W}\mathbf{G}\xi, \mathbf{X}\mathbf{W}\mathbf{C}_3\mathbf{W}'\mathbf{X}' + \mathbf{X}\mathbf{C}_2\mathbf{X}' + \Sigma_y) \quad (16)$$

$$(17)$$

であることが分かる。したがって、周辺最尤解を求めることが可能となる。

また、ベイズ的には、パラメタの事後分布 h はデータならびにパラメタの同時分布 f に比例するため

$$h(\beta, \gamma|\mathbf{y}, \xi, \sigma, \Psi, \Phi) \propto f(\mathbf{y}, \beta, \gamma|\xi, \sigma, \Psi, \Phi) \quad (18)$$

$$= f_0(\mathbf{y}|\beta, \Sigma_y) \times h_2(\beta|\gamma, \mathbf{C}_2) \times h_3(\gamma|\xi, \mathbf{C}_3) \quad (19)$$

である。

2 尤度関数、あるいはパラメタの事後同時分布

上記の \mathbf{y} の周辺分布は大きすぎるので、これを $\hat{\beta}$ の分布として書き直す。

$$f_0(\mathbf{y}|\beta, \Sigma_y) \propto |\Sigma_y|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)' \Sigma_y^{-1}(\mathbf{y} - \mathbf{X}\beta)\right] \quad (20)$$

$$\propto f_{11}(\hat{\beta}|\mathbf{C}_1) \times f_{12}(\hat{\beta}|\beta, \mathbf{C}_1) \quad (21)$$

ただし、

$$f_{11}(\hat{\beta}|\mathbf{C}_1) \propto \prod_j^m \left[(\sigma_j^2)^{-\frac{1}{2}n_j} \exp\left(-\frac{1}{2} \frac{\text{RSS}_j}{\sigma_j^2}\right) \right] \quad (22)$$

$$\text{RSS}_j = (\mathbf{y}_j - \mathbf{X}_j \hat{\beta}_j)' (\mathbf{y}_j - \mathbf{X}_j \hat{\beta}_j) \quad (23)$$

$$\hat{\beta}_j = (\mathbf{X}_j' \mathbf{X}_j)^{-1} \mathbf{X}_j' \mathbf{y}_j \quad (24)$$

$$f_{12}(\hat{\beta}|\beta, \mathbf{C}_1) \propto \exp\left(-\frac{1}{2}(\hat{\beta} - \beta)' \mathbf{C}_1^{-1}(\hat{\beta} - \beta)\right) \quad (25)$$

$$\hat{\beta} = \hat{\beta}_j \text{を縦に並べたもの } mp \times 1 \quad (26)$$

であるが、 f_{11} の部分には、パラメタ σ のみが含まれ、 β は含まれない。

したがって、

$$h(\beta, \gamma|\hat{\beta}, \xi, \mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3) \propto f(\hat{\beta}, \beta, \gamma|\xi, \mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3) \quad (27)$$

$$\propto f_{11}(\hat{\beta}|\mathbf{C}_1) \times f_{12}(\hat{\beta}|\beta, \mathbf{C}_1) \times h_2(\beta|\gamma, \mathbf{C}_2) \times h_3(\gamma|\xi, \mathbf{C}_3) \quad (28)$$

また、周辺尤度は

$$f(\hat{\beta}|\xi, \mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3) \propto \int_{\Gamma} \int_{\beta} f(\hat{\beta}, \beta, \gamma|\xi, \mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3) d\beta d\gamma \quad (29)$$

となるが、これを最大とするパラメタ $(\xi, \mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3)$ の値をもとめる。 $\hat{\beta}$ の周辺分布は

$$\hat{\beta} \sim N(\mathbf{W}\mathbf{G}\xi, \mathbf{C}_1 + \mathbf{C}_2 + \mathbf{W}\mathbf{C}_3\mathbf{W}') \quad (30)$$

となる。

3 EMアルゴリズムによる推定

3.1 E-step: データと分散成分を既知とした場合のパラメタの条件付き分布

$$\begin{pmatrix} \beta \\ \gamma \end{pmatrix} \sim N\left(\begin{pmatrix} \beta^* \\ \gamma^* \end{pmatrix}, \begin{bmatrix} D(\beta) & \text{Cov}(\beta, \gamma) \\ \text{Cov}(\gamma, \beta) & D(\gamma) \end{bmatrix}\right) \quad (31)$$

$$\text{Cov}(\beta, \gamma) = \mathbf{C}_1(\mathbf{C}_1 + \mathbf{C}_2)^{-1} \mathbf{W}\mathbf{D}(\gamma) \quad (32)$$

3.1.1 C_3 を用いた表現

$$\begin{aligned}\mathcal{D}(\beta) &= \mathbf{W}\mathbf{C}_3\mathbf{W}' + \mathbf{C}_2 - (\mathbf{W}\mathbf{C}_3\mathbf{W}' + \mathbf{C}_2)(\mathbf{W}\mathbf{C}_3\mathbf{W}' + \mathbf{C}_2 + \mathbf{C}_1)^{-1}(\mathbf{W}\mathbf{C}_3\mathbf{W}' + \mathbf{C}_2) \\ &= (\mathbf{C}_1^{-1} + (\mathbf{C}_2 + \mathbf{W}\mathbf{C}_3\mathbf{W}')^{-1})^{-1}\end{aligned}\quad (34)$$

$$\beta^* = \mathbf{W}\mathbf{G}\xi + (\mathbf{W}\mathbf{C}_3\mathbf{W}' + \mathbf{C}_2)(\mathbf{W}\mathbf{C}_3\mathbf{W}' + \mathbf{C}_2 + \mathbf{C}_1)^{-1}(\hat{\beta} - \mathbf{W}\mathbf{G}\xi) \quad (35)$$

$$= \mathcal{D}(\beta) (\mathbf{C}_1^{-1}\hat{\beta} + (\mathbf{C}_2 + \mathbf{W}\mathbf{C}_3\mathbf{W}')^{-1}\mathbf{W}\mathbf{G}\xi) \quad (36)$$

$$\mathcal{D}(\gamma) = \mathbf{C}_3 - \mathbf{C}_3\mathbf{W}'(\mathbf{W}\mathbf{C}_3\mathbf{W}' + \mathbf{C}_2 + \mathbf{C}_1)^{-1}\mathbf{W}\mathbf{C}_3 \quad (37)$$

$$\gamma^* = \mathcal{D}(\gamma)\mathbf{W}'(\mathbf{C}_1 + \mathbf{C}_2)^{-1}\hat{\beta} + (\mathbf{I} - \mathcal{D}(\gamma)\mathbf{W}'(\mathbf{C}_1 + \mathbf{C}_2)^{-1}\mathbf{W})\mathbf{G}\xi \quad (38)$$

3.1.2 C_3^{-1} を用いた表現

$$\mathcal{D}(\beta) = (\mathbf{C}_1^{-1} + \mathbf{C}_2^{-1} - \mathbf{C}_2^{-1}\mathbf{W}(\mathbf{W}'\mathbf{C}_2^{-1}\mathbf{W} + \mathbf{C}_3^{-1})^{-1}\mathbf{W}'\mathbf{C}_2^{-1})^{-1} \quad (39)$$

$$= \mathbf{U}^{-1} + \mathbf{U}^{-1}\mathbf{C}_2^{-1}\mathbf{W}\mathbf{\Delta}^{-1}\mathbf{W}'\mathbf{C}_2^{-1}\mathbf{U}^{-1} \quad (40)$$

where

$$\mathbf{U} = \mathbf{C}_1^{-1} + \mathbf{C}_2^{-1} \quad \text{and} \quad \mathbf{\Delta} = \mathbf{C}_3^{-1} + \mathbf{W}'(\mathbf{C}_1 + \mathbf{C}_2)^{-1}\mathbf{W} \quad (41)$$

$$\beta^* = \mathcal{D}(\beta)\mathbf{C}_1^{-1}\hat{\beta} + (\mathbf{I} - \mathcal{D}(\beta)\mathbf{C}_1^{-1})\mathbf{W}\mathbf{G}\xi \quad (42)$$

$$\mathcal{D}(\gamma) = \mathbf{\Delta}^{-1} = (\mathbf{C}_3^{-1} + \mathbf{W}'(\mathbf{C}_1 + \mathbf{C}_2)^{-1}\mathbf{W})^{-1} \quad (43)$$

$$\gamma^* = \mathcal{D}(\gamma) (\mathbf{W}'(\mathbf{C}_1 + \mathbf{C}_2)^{-1}\hat{\beta} + \mathbf{C}_3^{-1}\mathbf{G}\xi) \quad (44)$$

3.2 M-step

Note that $\mathcal{D}(\beta)$ is the $mp \times mp$ block diagonal matrix consisting of b ($m_{LP} \times m_{LP}$) matrices, $\mathcal{D}(\gamma)$ is the $bp \times bp$ block diagonal matrix consisting of b ($p \times p$) matrices, and $\mathbf{\Delta}$ is the $(bp \times bp)$ block diagonal matrix with b ($p \times p$) matrices as diagonal elements.

Since minus twice the log complete data likelihood can be written as

$$\begin{aligned}-2 \ln(L) &= -2 \ln[f(\hat{\beta}, \beta, \gamma | \xi, \sigma^2, \Psi, \Phi)] \\ &= \ln[f_1(\hat{\beta} | \beta, \sigma^2) f_2(\beta | \gamma, \mathbf{C}_2) f_3(\gamma | \xi, \mathbf{C}_3)] \\ &= \sum_j [n_j \ln(\sigma_j^2)] + \sum_j \left[\frac{RSS_j}{\sigma_j^2} \right] \\ &\quad + \sum_j [(\hat{\beta}_j - \beta_j)' (\mathbf{X}_j' \mathbf{X}_j) (\hat{\beta}_j - \beta_j) / \sigma_j^2] \\ &\quad + \ln |\mathbf{C}_2| + (\beta - \mathbf{W}\gamma)' \mathbf{C}_2^{-1} (\beta - \mathbf{W}\gamma) + \ln |\mathbf{C}_3| + (\gamma - \mathbf{G}\xi)' \mathbf{C}_3^{-1} (\gamma - \mathbf{G}\xi)\end{aligned}\quad (45)$$

its posterior expectation becomes

$$-2\mathcal{E} \ln L(\pi | \hat{\beta}, \pi^{(0)}) = \sum_j n_j \ln(\sigma_j^2) + \sum_j \frac{RSS_j}{\sigma_j^2}$$

$$\begin{aligned}
& + \sum_j \{(\widehat{\beta}_j - \beta_j^*)'(\mathbf{X}_j' \mathbf{X}_j)(\widehat{\beta}_j - \beta_j^*)/\sigma_j^2 \\
& \quad + \text{tr}[(\mathbf{X}_j' \mathbf{X}_j)\mathcal{D}(\beta_j)]/\sigma_j^2\} \\
& + \ln |\mathbf{C}_2| + (\beta^* - \mathbf{W}\gamma^*)' \mathbf{C}_2^{-1} (\beta^* - \mathbf{W}\gamma^*) \\
& + \text{tr}[\mathbf{C}_2^{-1} \{\mathcal{D}(\beta) - \mathbf{W}\text{Cov}(\beta, \gamma)' - \text{Cov}(\beta, \gamma)\mathbf{W}' + \mathbf{W}\mathcal{D}(\gamma)\mathbf{W}'\}] \\
& + \ln |\mathbf{C}_3| + (\gamma^* - \mathbf{G}\xi)' \mathbf{C}_3^{-1} (\gamma^* - \mathbf{G}\xi) \\
& \quad + \text{tr}[\mathbf{C}_3^{-1} \mathcal{D}(\gamma)]
\end{aligned} \tag{46}$$

Note that the number of the second and the third term of (4.15) can be written as

$$RSS_j^* = (\mathbf{y}_j - \mathbf{X}_j \beta_j^*)' (\mathbf{y}_j - \mathbf{X}_j \beta_j^*) \tag{47}$$

Using these forms, the updating formulae in the M-step (i. e. the solutions of the equations obtained from the expected log likelihood conditional on $\boldsymbol{\pi}^{(0)}$) may be summarized as

$$\begin{aligned}
\xi &= [\sum_{L=1}^b \gamma_L^*] / b \\
&= (\mathbf{G}' \mathbf{G})^{-1} \mathbf{G}' \gamma^*
\end{aligned} \tag{48}$$

$$\sigma_j^2 = [RSS_j^* + \text{tr}\{(\mathbf{X}_j' \mathbf{X}_j)\mathcal{D}(\beta_j)\}] / n_j \tag{49}$$

$$\Psi_L = [\sum_{j \in L} (\beta_j^* - \gamma_L^*)(\beta_j^* - \gamma_L^*)' + \mathbf{Q}_j] / m_L$$

where,

$$\mathbf{Q}_j = \mathcal{D}(\beta_j) + \mathcal{D}(\gamma_L) - \text{Cov}(\beta_j, \gamma_L) - \text{Cov}(\beta_j, \gamma_L)' \tag{50}$$

and,

$$\Phi = [\sum_{L=1}^b [(\gamma_L^* - \xi)(\gamma_L^* - \xi)' + \mathcal{D}(\gamma_L)]] / b \tag{51}$$

4 Extended m-group regression model

4.1 Model

Let us extend the original m-group regression model as follows:

$$\mathbf{y}_k = \mathbf{Z}_k \boldsymbol{\alpha} + \mathbf{X}_k \boldsymbol{\beta}_k + \mathbf{F}_k \boldsymbol{\xi}_k + \boldsymbol{\varepsilon}_k, \quad k = 1, 2, \dots, m, \quad (52)$$

where \mathbf{Z}_k and $\boldsymbol{\alpha}$ are, respectively, the $N_k \times p_1$ regressor matrix and the associated regression coefficient vector which is assumed to be common to all the m groups, \mathbf{X}_k and $\boldsymbol{\beta}_k$ are, respectively, the $N_k \times p_2$ regressor matrix and the associated regression coefficient vector which is assumed to be exchangeable across all the m groups, and \mathbf{F}_k and $\boldsymbol{\xi}_k$ are, respectively, the $N_k \times p_3$ regressor matrix and the associated regression coefficient vectors which is not exchangeable across groups. The distributional assumptions are the same as before. That is,

$$\boldsymbol{\varepsilon}_k \sim N(\mathbf{0}, \boldsymbol{\sigma}_k^2 \mathbf{I}), \quad k = 1, 2, \dots, m, \quad (53)$$

and

$$\boldsymbol{\beta}_k \sim N(\boldsymbol{\gamma}, \boldsymbol{\Psi}), \quad k = 1, 2, \dots, m. \quad (54)$$

where the size of the $\boldsymbol{\Psi}$ matrix is now $p_2 \times p_2$.

This implies that, marginally,

$$\mathbf{y}_k \sim N(\mathbf{Z}_k \boldsymbol{\alpha} + \mathbf{X}_k \boldsymbol{\gamma} + \mathbf{F}_k \boldsymbol{\xi}_k, \boldsymbol{\sigma}_k^2 \mathbf{I} + \mathbf{X}_k \boldsymbol{\Psi} \mathbf{X}_k'), \quad k = 1, 2, \dots, m. \quad (55)$$

and \mathbf{y}_k and $\mathbf{y}_{k'}$ $k' \neq k$, are independent. Therefore, the marginal likelihood to be maximized can be written as

$$\ell = -\frac{1}{2} \sum_{k=1}^m \left[\ln |\boldsymbol{\Sigma}_k| + (\mathbf{y}_k - \mathbf{G}_k \boldsymbol{\theta}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_k - \mathbf{G}_k \boldsymbol{\theta}_k) \right], \quad (56)$$

where

$$\mathbf{G}_k = [\mathbf{Z}_k, \mathbf{X}_k, \mathbf{F}_k], \quad (57)$$

and

$$\boldsymbol{\theta}_k = [\boldsymbol{\alpha}', \boldsymbol{\beta}_k', \boldsymbol{\xi}_k']'. \quad (58)$$

Note that the definition of the $\boldsymbol{\Sigma}_k$ matrix is the same as before.

Collectively, the above model can be written as

$$\mathbf{y} = \mathbf{Z} \boldsymbol{\alpha} + \mathbf{X} \boldsymbol{\beta} + \mathbf{F} \boldsymbol{\xi} + \boldsymbol{\varepsilon}, \quad (59)$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon), \quad (60)$$

and

$$\boldsymbol{\beta} \sim N(\mathbf{W} \boldsymbol{\gamma}, \mathbf{C}_2), \quad (61)$$

where

$$\boldsymbol{\beta} = [\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \dots, \boldsymbol{\beta}_m']', \quad (62)$$

$$\boldsymbol{\xi} = [\boldsymbol{\xi}_1', \boldsymbol{\xi}_2', \dots, \boldsymbol{\xi}_m']', \quad (63)$$

$$\mathbf{Z} = [\mathbf{Z}_1', \mathbf{Z}_2', \dots, \mathbf{Z}_m']', \quad (64)$$

$$\mathbf{X} = \text{Diag}(\mathbf{X}_k), \quad (65)$$

$$\mathbf{F} = \text{Diag}(\mathbf{F}_k), \quad (66)$$

and

$$\mathbf{W} = \mathbf{1}_m \otimes \mathbf{I}_{p_2}. \quad (67)$$

Note that the dispersion matrices are the same as the ones in the original m-group regression model except for the sizes of Ψ and \mathbf{C}_2 matrices.

Therefore, marginal distribution of \mathbf{y} is

$$\mathbf{y} \sim N(\mathbf{Z}\boldsymbol{\alpha} + \mathbf{X}\mathbf{W}\boldsymbol{\gamma} + \mathbf{F}\boldsymbol{\xi}, \Sigma_\varepsilon + \mathbf{X}\mathbf{C}_2\mathbf{X}'), \quad (68)$$

4.2 Estimation of parameters via EM algorithm

In our model, the parameters of the model are: $\sigma_k^2, k = 1, 2, \dots, m, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \Psi$, and $\boldsymbol{\xi}_k, k = 1, 2, \dots, m$. In order to estimate the parameters we divide the whole set of parameters into three subsets, namely, 1) $\sigma_k^2, k = 1, 2, \dots, m, \boldsymbol{\gamma}$, and Ψ , 2) $\boldsymbol{\alpha}$, and 3) $\boldsymbol{\xi}_k, k = 1, 2, \dots, m$, and alternate the maximization of the marginal log likelihood with respect to each subset.

For the first subset, by noticing the fact that the conditional distribution of $\boldsymbol{\beta}_k$ given \mathbf{y}_k is

$$\boldsymbol{\beta}_k \sim N(\boldsymbol{\beta}_k^*, \mathbf{Q}_k), \quad (69)$$

where

$$\mathbf{Q}_k = \sigma_k^2 (\mathbf{X}_k' \mathbf{X}_k + \sigma_k^2 \Psi^{-1})^{-1} \quad (70)$$

$$= \sigma_k^2 (\mathbf{X}_k' \mathbf{X}_k)^{-1} (\sigma_k^2 (\mathbf{X}_k' \mathbf{X}_k)^{-1} + \Psi)^{-1} \Psi \quad (71)$$

and

$$\boldsymbol{\beta}_k^* = \mathbf{Q}_k \left(\frac{1}{\sigma_k^2} \mathbf{X}_k' (\mathbf{y}_k - \mathbf{Z}_k \boldsymbol{\alpha} - \mathbf{F}_k \boldsymbol{\xi}_k) + \Psi^{-1} \boldsymbol{\gamma} \right) \quad (72)$$

$$= \sigma_k^2 (\mathbf{X}_k' \mathbf{X}_k)^{-1} (\sigma_k^2 (\mathbf{X}_k' \mathbf{X}_k)^{-1} + \Psi)^{-1} \left(\frac{1}{\sigma_k^2} \Psi \mathbf{X}_k' (\mathbf{y}_k - \mathbf{Z}_k \boldsymbol{\alpha} - \mathbf{F}_k \boldsymbol{\xi}_k) + \boldsymbol{\gamma} \right), \quad (73)$$

we have the following EM steps.

E Step:

Calculate $\boldsymbol{\beta}_k^*$ and \mathbf{Q}_k above.

M Step:

Update $\sigma_k^2, \boldsymbol{\gamma}$, and Ψ as

$$\sigma_k^2 = (\text{RSS}_k^* + \text{tr}(\mathbf{X}_k' \mathbf{X}_k \mathbf{Q}_k)) / N_k, \quad (74)$$

where

$$\text{RSS}_k^* = (\mathbf{y}_k - \mathbf{G}_k \boldsymbol{\theta}_k^*)' (\mathbf{y}_k - \mathbf{G}_k \boldsymbol{\theta}_k^*), \quad (75)$$

and

$$\boldsymbol{\theta}_k^* = [\boldsymbol{\alpha}', \boldsymbol{\beta}_k^{*'}, \boldsymbol{\xi}_k^{*'}]', \quad (76)$$

$$\boldsymbol{\gamma} = \sum_{k=1}^m \boldsymbol{\beta}_k^*/m, \quad (77)$$

and

$$\boldsymbol{\Psi} = (\sum_{k=1}^m \boldsymbol{\beta}_k^* \boldsymbol{\beta}_k^{*'} - m \boldsymbol{\gamma} \boldsymbol{\gamma}' + \sum_{k=1}^m \mathbf{Q}_k)/m. \quad (78)$$

For the second and the third subsets, the weighted least squares solutions are

$$\boldsymbol{\alpha} = (\mathbf{Z}'(\boldsymbol{\Sigma}_\varepsilon + \mathbf{X} \mathbf{C}_2 \mathbf{X}')^{-1} \mathbf{Z})^{-1} \mathbf{Z}'(\boldsymbol{\Sigma}_\varepsilon + \mathbf{X} \mathbf{C}_2 \mathbf{X}')^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\gamma} - \mathbf{F} \boldsymbol{\xi}) \quad (79)$$

and

$$\boldsymbol{\xi} = (\mathbf{F}'(\boldsymbol{\Sigma}_\varepsilon + \mathbf{X} \mathbf{C}_2 \mathbf{X}')^{-1} \mathbf{F})^{-1} \mathbf{F}'(\boldsymbol{\Sigma}_\varepsilon + \mathbf{X} \mathbf{C}_2 \mathbf{X}')^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\gamma} - \mathbf{Z} \boldsymbol{\alpha}). \quad (80)$$

Repeated applications of the EM steps, namely, Eq.(74) through Eq.(80), will maximize the log marginal likelihood.

*

References

- Boscardim, W. J. and Gelman, A. (1996) Bayesian Regression with Parametric Models for Heteroscedasticity. —sl *Advances in Econometrics* 11A, 87-109.
- Braun, H. I., Jones, D. H., Rubin, D. B. and Thayer, D. T. (1983). Empirical Bayes Estimation of Coefficients in the General Linear Model from Data of Deficient Rank, *Psychometrika*, **48**(2), 171-181.
- Bryk, A.S. and Raudenbush S. W. (1992) *Hierarchical Linear Models : Applications and Data Analysis Methods* (Advanced Qualitative Techniques in the Social Sciences, No. 1) Sage Pubns; ISBN: 0803946279
- DeLeeuw, J. and Kreft, I. (1986). Random Coefficient Models for Multilevel Analysis, *Journal of Educational Statistics*, **11**(19), 57-85.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm (with Discussion), *Journal of the Royal Statistical Society Series B*, **39**, 1-18.
- Dempster, A. P., Rubin, D. B. and Tsutakawa, R. K. (1981). Estimation in Covariance Components Models, *JASA*, **76**, (374), 341-535.
- Dunbar, S. B., Mayekawa, S. and Novick, M. R. (1986) Simultaneous estimation of regression functions for marine corps technical training specialties. *Journal of Educational Statistics*. Vol 11, No.4, 275-292

- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995) *Bayesian Data Analysis* London: Chapman & Hall ISBN 0 412 03991 5
- Goldstein, H. (1986). Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares, *Biometrika*, **73**(1), 43-56.
- Goldstein, H. (1995). *Multilevel Models in Educational and Social Research*(2nd ed.) London: Arnold.
- Lindley, D. V.(1971) The Estimation of Many Parameters. In *Foundations of Statistical Science*, ed. V. P. Godambe and D. A. Sprott. Toronto: Holt, Rinehart, and Winston.
- Lindley, D. V. and Novick, M. R.(1981) The Role of Exchangeability in Inference. *Annals of Statistics*, 9, 45-58.
- Lindley, D. V. and Smith, A. F. M. (1972). Bayes Estimates for the Linear Model, *Journal of the Royal Statistical Society, Series B*, **34**, 1-41.
- Longford, N. T. (1993) *Random Coefficient Models* (Oxford Statistical Science, No 11) Clarendon Pr; ISBN: 0198522649
- Novick, M. R., Jackson, P. H., Thayer, D. T. and Cole, N. S. (1972). Estimating Multiple Regressions in M Groups-A Cross Validation Study, *British Journal of Mathematical and Statistical Psychology*, **25**, 33-50.
- O'Hagan, A. (1976). On Posterior Joint and Marginal Modes, *Biometrika*, **63**, 329-333.
- Ramsay, J.O. (1975) Solving implicit equations in psychometric data analysis. *Psychometrika*, **40**, 337-360.
- Raudenbush, S. W. (1988). Educational Applications of Hierarchical Linear Models: A Review. *Journal of Educational Statistics*, **13**, 85-116.
- Shigemasu, K. (1976). Development and Validation of a Simplified M-Group Regression Model, *Journal of Educational Statistics*, **1**(2), 157-180.
- Shigemasu K. and Mayekawa S. (1996). A Bayesian Hierarchical Linear Model with Educational Applications, *J of Japan Statist. Soc.*, Vol.26, No.1, 1-23.
- Smith, A. F. M. (1973). A General Bayesian Linear Model. *Journal of the Royal Statistical Society, Series B*, **35**, 61-75.
- Strenio, J. L. F., Weisberg, H. I. and Bryk, A. S. (1983). Empirical Bayes Estimation of Individual Growth Curve Parameters and their Relationship to Covariates, *Biometrics*, **39**, 71-86.

Matrix Inversion Formulae

1. $(A + B)^{-1} = A^{-1}(A^{-1} + B^{-1})^{-1}B^{-1}$
2. $(A + B)^{-1} = B^{-1}(A^{-1} + B^{-1})^{-1}A^{-1}$
3. $(A + B)^{-1} = A^{-1} - A^{-1}(A^{-1} + B^{-1})^{-1}A^{-1}$
4. $(A + B)^{-1} = B^{-1} - B^{-1}(A^{-1} + B^{-1})^{-1}B^{-1}$
5. $(A + B)^{-1} = A^{-1} - A^{-1}B(A + B)^{-1}$
6. $(A + B)^{-1} = A^{-1} - (A + B)^{-1}BA^{-1}$
7. $(A + B)^{-1} = B^{-1} - B^{-1}A(A + B)^{-1}$
8. $(A + B)^{-1} = B^{-1} - (A + B)^{-1}AB^{-1}$
9. $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$
10. $(A + BCD)^{-1}B = A^{-1}B(C^{-1} + DA^{-1}B)^{-1}C^{-1}$
11. $D(A + BCD)^{-1} = C^{-1}(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$
12. $D(A + BCD)^{-1}B = C^{-1} - C^{-1}(C^{-1} + DA^{-1}B)^{-1}C^{-1}$

- Formulae 1-4 replace A and B by A^{-1} and B^{-1} .
- Formulae 5-8 leave $(A + B)^{-1}$ as it is.
- Formula 9-12 replace A and C by A^{-1} and C^{-1} .
- Formulae 1-4 are useful to avoid taking inverses of each term when calculating something like $(A^{-1} + B^{-1})^{-1}$. That is, by formula 1 it is equal to $A(A + B)^{-1}B$.
- Formulae 5-8 are useful to simplify something like $A(A + B)^{-1}$.

For example:

$$A(A + B)^{-1} = I - B(A + B)^{-1} \text{ by formula 5}$$

$$(A + B)^{-1}A = I - (A + B)^{-1}B \text{ by formula 6}$$

$$B(A + B)^{-1} = I - A(A + B)^{-1} \text{ by formula 7}$$

$$(A + B)^{-1}B = I - (A + B)^{-1}A \text{ by formula 8}$$