Moment-Based Estimation For Hierarchical Models With Three Or More Levels

Ningshan Zhang

August 30, 2016

1 Introduction

In this paper, we extend the moment-based procedure to fit hierarchical GLM models with three or more levels.

We introduce two-level hierarchical generalized linear models and the original moment-based procedure in Section 2. In Section 3 we first describe the nested hierarchical model with three levels and introduce the extended procedure to fit it, then we generalize to any level of hierarchies.

2 Moment-Based Procedure for Hierarchical Generalized Linear Models Of Two Levels

2.1 Hierarchical Generalized Linear Models Of Two Levels

Consider a collection of M groups. In group i, there are n_i random observations denoted by y_{ij} for $j = 1, \ldots, n_i$. Suppose for each observation there are two associated predictor vectors: $\boldsymbol{x}_{ij}$ of dimension p, and $\boldsymbol{z}_{ij}$ of dimension q. Let $\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i$ be the response vector and predictor matrices of dimension $n_i \times 1, n_i \times p, n_i \times q$, where row j equals $y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}$ respectively. In a generalized hierarchical linear model, we assume that conditional on group specific random effects vector $\boldsymbol{u}_i$, the following relationship holds

$$\mathbb{E}\left(\boldsymbol{y}_i|\boldsymbol{u}_i\right) = g_i^{-1}\left(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{u}_i\right),$$

where $\boldsymbol{\beta}$ is a vector of p fixed effects shared across the M group, g_i is some specific link function. Further, assume that within each group i, the response values are independent and

$$\operatorname{Cov}\left(\boldsymbol{y}_i|\boldsymbol{u}_i\right) = \phi I$$

for some (possibly known) dispersion parameter ϕ.

Finally, random effect vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M$ are i.i.d distributed with mean zero and covariance $\operatorname{Cov}\left(\boldsymbol{u}_i\right) = \Sigma$ for some positive-semidefinite matrix Σ.

If we assume that the random effect vector and the response vector are multivariable normal, then we get a prediction for the random coefficient vector $\boldsymbol{u}_i$ by the posterior distribution

$$\boldsymbol{u}_i|\boldsymbol{y}_i \sim \mathcal{N}\left(\hat{\boldsymbol{u}}_i, \phi\boldsymbol{C}i\right)$$

where

$$\hat{\boldsymbol{u}}_i = \boldsymbol{C}_i\boldsymbol{Z}_i^T\left(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)$$
$$\boldsymbol{C}_i = \left(\phi\Sigma^{-1} + \boldsymbol{Z}_i^T\boldsymbol{Z}_i\right)^{-1}$$

2.2 A Moment-Based Procedure for Fitting Two-Levels Hierarchical GLM

In Perry, P. O. (2016), a moment-based estimation procedure is used to fit the two-level hierarchical GLM. Given input $\left(\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\right)_{i=1}^M$, the full procedure works as follows:

1. For each group $i = 1, \ldots, M$:

(a) Construct group specific feature matrix $\boldsymbol{F}_i = [\boldsymbol{X}_i \ \boldsymbol{Z}_i]$; use a singular value decomposition to decompose this matrix as $\boldsymbol{F}_i = \boldsymbol{U}_i \boldsymbol{D}_i \boldsymbol{V}_i^T$, where $\boldsymbol{U}_i$ has full column rank r_i and $\boldsymbol{V}_i = \left[\boldsymbol{V}_{i1}^T \ \boldsymbol{V}_{i2}^T\right]^T$ is a matrix of dimension $(p+q) \times r_i$ with orthogonal columns. Let $\boldsymbol{F}_{0i} = \boldsymbol{U}_i \boldsymbol{D}_i$.

(b) Use Firth's modified score function with data $(\boldsymbol{y}_i, \boldsymbol{F}_{0i})$ to get group-specific effect estimate $\hat{\boldsymbol{\eta}}_{0i}$.

(c) Set $\boldsymbol{D}_i^{-2}$ to be a plug-in estimate of $\phi^{-1}\mathrm{Cov}\left(\hat{\boldsymbol{\eta}}_{0i} | \boldsymbol{u}_i\right)$, set $\hat{\boldsymbol{\eta}}_i = \boldsymbol{V}_i \hat{\boldsymbol{\eta}}_{0i}$.

(d) Compute group-specific dispersion estimate $\hat{\phi}_i$.

2. If ϕ is not given, compute pooled dispersion estimate

$$\hat{\phi} = \frac{\sum_{i=1}^{M} (n_i - r_i)\, \hat{\phi}_i}{\sum_{i=1}^{M} (n_i - r_i)};$$

otherwise set $\hat{\phi} = \phi$.

3. Choose positive semi-definite weight matrices $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_M$. With these weights, compute weighted first and second moments Ω, b, Ω_2, B. (please refer to the original paper for details). The main parameters are estimated by

$$\hat{\boldsymbol{\beta}} = \Omega^{-1} b$$
$$\mathrm{vec}\left(\hat{\Sigma}_u\right) = \Omega_2^{-1} B$$

4. Project $\hat{\Sigma}_u$ onto positive semidefinite cone if necessary.

5. Optionally, use the estimated $\hat{\boldsymbol{\beta}}$ and $\hat{\Sigma}_u$ to compute a new set of weights $\boldsymbol{W}_i, i = 1, \ldots M$ and redo steps 3 and 4.

Throughout this paper, we will refer to this procedure as 'mhglm2L' (moment-based hierarchical GLM two levels).

2.3 Posterior Inference with Plug-in Estimates

We can use normal approximation for the distribution of $\boldsymbol{u}_i$ and $\hat{\boldsymbol{\eta}}_i | \boldsymbol{u}_i$ to compute posterior mean for $\boldsymbol{u}_i$:

$$\hat{\mathbb{E}}\left(\boldsymbol{u}_i | \boldsymbol{y}_i\right) = C_i \boldsymbol{Z}_i^T \left(\left[\boldsymbol{X}_i \boldsymbol{Z}_i\right] \hat{\boldsymbol{\eta}}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}\right),$$
$$\hat{\mathrm{Cov}}\left(\boldsymbol{u}_i | \boldsymbol{y}_i\right) = \hat{\phi} C_i,$$

where $C_i = \left(\hat{\phi} \hat{\Sigma}_u^{-1} + \boldsymbol{Z}_i^T \boldsymbol{Z}_i\right)^{-1}$. Notice that the posterior inference only requires original data as well as plug-in estimates $\hat{\boldsymbol{\beta}}, \hat{\Sigma}, \hat{\phi}$ for $\boldsymbol{\beta}, \Sigma, \phi$ respectively.

2.4 Asymptotic Normality

Under certain assumptions, if $M \to \infty$ and $\sum_{i=1}^{M} \left\|\Omega^{-1} \boldsymbol{V}_{i1} \boldsymbol{W}_i \boldsymbol{V}_{i1}\right\|^2 \to 0$, then $\Omega^{-1/2}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}(0, I)$. For details, see Proposition 6.4 in original paper. We are going to use this asymptotic normality to fit hierarchical models of three levels or more.

2.5 (Optional) Rank-Deficient Situation

In this procedure, fixed effects are identifiable iff $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \ldots \\ \boldsymbol{X}_M \end{bmatrix}$ has full rank. Similarly, random effects' covariance matrix is identifiable iff $\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 \\ \ldots \\ \boldsymbol{Z}_M \end{bmatrix}$ has full rank. We could use pseudo-inverse in the last step of estimating $\hat{\boldsymbol{\beta}}$ and $\hat{\Sigma}_u$ and get a good estimate in the column space. To be more specific, use singular value decomposition to decompose $\boldsymbol{X} = \boldsymbol{X}_0 \boldsymbol{V}_X^T$ and $\boldsymbol{Z} = \boldsymbol{Z}_0 \boldsymbol{V}_Z^T$, where $\boldsymbol{X}_0$ and $\boldsymbol{Z}_0$ are column space with full rank and $\boldsymbol{V}_X$ and $\boldsymbol{V}_Z$ are matrices with orthogonal columns. Then $\boldsymbol{V}_X^T \hat{\boldsymbol{\beta}}$ is a good estimate of identifiable part $\boldsymbol{V}_X^T \boldsymbol{\beta}$, and $\boldsymbol{V}_Z^T \hat{\Sigma}_u \boldsymbol{V}_Z$ is a good estimate of identifiable part $\boldsymbol{V}_Z^T \Sigma_u \boldsymbol{V}_Z$.

3 Hierarchical Generalized Linear Models of Three Levels Or More

3.1 Three Levels Model

Consider a collection of M groups. In group i, there are m_i subgroups nested within it, denote as group ij for $j = 1, \dots m_i$. In group ij, there are n_{ij} random observations denoted by y_{ijk} for $k = 1, \dots, n_{ij}$. Suppose for each observation there are three associated predictor vectors: $\boldsymbol{x}_{ijk}$ of dimension p, $\boldsymbol{z}_{ijk}^1$ of dimension q_1 and $\boldsymbol{z}_{ijk}^2$ of dimension q_2. Let $\boldsymbol{y}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}^1, \boldsymbol{Z}_{ij}^2$ be the response vector and predictor matrices of dimension $n_{ijk} \times 1, n_{ijk} \times p, n_{ijk} \times q_1, n_{ijk} \times q_2$. In a three-level generalized linear model, we assume that conditional on group specific random effects vectors $\boldsymbol{u}_i^1$ and $\boldsymbol{u}_{ij}^2$, the following holds

$$\mathbb{E}\left(\boldsymbol{y}_{ij} | \boldsymbol{u}_i^1, \boldsymbol{u}_{ij}^2\right) = g_{ij}^{-1}\left(\boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^1 \boldsymbol{u}_i^1 + \boldsymbol{Z}_{ij}^2 \boldsymbol{u}_{ij}^2\right),$$

where $\boldsymbol{\beta}$ is a vector of p fixed effects shared across all hierarchies, g_{ij} is some specific link function; $\boldsymbol{u}_i^1$ is a vector of q_1 random effects shared across subgroups under group i; $\boldsymbol{u}_{ij}^2$ is a vector of q_2 random effects that is specific to group ij. Further, assume that within each group ij, the response values are independent and

$$\mathrm{Cov}\left(\boldsymbol{y}_{ij} | \boldsymbol{u}_i^1, \boldsymbol{u}_{ij}^2\right) = \phi I$$

for dispersion parameter ϕ.

Finally, random effects vectors $\boldsymbol{u}_i^1$ are i.i.d distributed with mean zero and covariance $\mathrm{Cov}\left(\boldsymbol{u}_i^1\right) = \Sigma^1$ for some positive semidefinite matrix Σ^1; random effects vectors $\boldsymbol{u}_{ij}^2$ are i.i.d. distributed with mean zero and covariance $\mathrm{Cov}\left(\boldsymbol{u}_{ij}^2\right) = \Sigma^2$ for positive semidefinite matrix Σ^2.

3.2 Fitting Three Levels Model

Denote the input data as $\left(\boldsymbol{y}_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}^1, \boldsymbol{Z}_{ij}^2\right)_{ij \in \{11, \dots m_0 m_{m_0}\}}$. To gain intuition, it is helpful to first consider data under group 1: $\left(\boldsymbol{y}_{1j}, \boldsymbol{X}_{1j}, \boldsymbol{Z}_{1j}^1, \boldsymbol{Z}_{1j}^2\right)_{j \in \{1 \dots m_1\}}$. We can apply mhglm2L to $\left(\boldsymbol{y}_{1j}, \left[\boldsymbol{X}_{1j}\ \boldsymbol{Z}_{1j}^1\right], \boldsymbol{Z}_{1j}^2\right)_{j \in \{1 \dots m_1\}}$, since by assumption

$$\mathbb{E}\left(\boldsymbol{y}_{1j} | \boldsymbol{u}_1^1, \boldsymbol{u}_{1j}^2\right) = g_{1j}^{-1}\left(\left[\boldsymbol{X}_{1j}\ \boldsymbol{Z}_{1j}^1\right]\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u}_1^1 \end{bmatrix} + \boldsymbol{Z}_{1j}^2 \boldsymbol{u}_{1j}^2\right).$$

Note that $\boldsymbol{u}_1^1$ is viewed as part of fixed effects within group 1. The mhglm2L procedure returns estimates $\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}}_1^1 \end{bmatrix}$ and $\hat{\Sigma}$, together with first moment matrix Ω. Here $\hat{\Sigma}$ could be used directly as an estimate for $\mathrm{Cov}\left(\boldsymbol{u}_{1j}^2\right) = \Sigma^2$.

From the asymptotic normality property of mhglm2L,

$$\Omega^{-1/2}\left(\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}}_1^1 \end{bmatrix} - \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u}_1^1 \end{bmatrix}\right) \xrightarrow{d} \mathcal{N}(0, I).$$

We can rewrite this as regression formulas:

$$\mathbb{E}\left(\Omega^{-1/2}\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}}_1^1 \end{bmatrix} | \boldsymbol{u}_i^1\right) \approx \left(\Omega^{-1/2}\right)_{1:p}\boldsymbol{\beta} + \left(\Omega^{-1/2}\right)_{(p+1):(p+q_1)}\boldsymbol{u}_i^1,$$
$$\mathrm{Cov}\left(\Omega^{-1/2}\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}}_1^1 \end{bmatrix} | \boldsymbol{u}_i^1\right) \approx I,$$

where $\left(\Omega^{-1/2}\right)_{a:b}$ indicates the submatrix containing only columns $\{a, a+1, \ldots, b\}$;

We can repeat the above procedure for every group i. That is, we can apply mhglm2L to $\left(\boldsymbol{y}_{ij}, \left[\boldsymbol{X}_{ij}\ \boldsymbol{Z}_{ij}^1\right], \boldsymbol{Z}_{ij}^2\right)_{j\in\{1\ldots m_i\}}$ and get estimate $\left[\hat{\boldsymbol{\beta}}_i^T\ \hat{\boldsymbol{u}}_i^{1T}\right]^T, \hat{\Sigma}_i^2$ as well as Ω_i for group $i, i=1,\ldots,M$. By the same asymptotic normality, we have

$$\begin{aligned} \mathbb{E}\left(\tilde{\boldsymbol{y}}_i|\boldsymbol{u}_i^1\right) &\approx \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} + \tilde{\boldsymbol{Z}}_i\boldsymbol{u}_i^1, \\ \mathrm{Cov}\left(\tilde{\boldsymbol{y}}_i|\boldsymbol{u}_i^1\right) &\approx I, \end{aligned}$$

with $\tilde{\boldsymbol{y}}_i = \Omega_i^{-1/2}\left[\hat{\boldsymbol{\beta}}_i^T\ \hat{\boldsymbol{u}}_i^{1T}\right]^T, \tilde{\boldsymbol{X}}_i = \left(\Omega^{-1/2}\right)_{1:p}, \tilde{\boldsymbol{Z}}_i = \left(\Omega^{-1/2}\right)_{(p+1):(p+q_1)}$. Therefore data $\left(\tilde{\boldsymbol{y}}_i, \tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{Z}}_i\right)_{i\in\{1,\ldots,M\}}$ asymptotically meets the model assumptions for two-levels GLM, and we could apply mhglm2L to it.

Now we are ready to write down the full procedure for estimating parameters of a three-level HGLM model.

1. For each group $i=1,\ldots,M$: apply mhglm2L to data $\left(\boldsymbol{y}_{ij}, \left[\boldsymbol{X}_{ij}\ \boldsymbol{Z}_{ij}^1\right], \boldsymbol{Z}_{ij}^2\right)_{j\in\{1,\ldots,m_i\}}$. Denote the estimated fixed effects, covariance matrix and dispersion parameter as $\left[\hat{\boldsymbol{\beta}}_i^T\ \hat{\boldsymbol{u}}_i^{1T}\right]^T, \hat{\Sigma}_i^2, \hat{\phi}_i$. Weighted first moment Ω_i will also be used later.

2. If ϕ is not given, compute pooled dispersion estimate

$$\hat{\phi} = \frac{\sum_{i=1}^M \sum_{j=1}^{m_i}\left(n_{ij}-r_{ij}\right)\hat{\phi}_i}{\sum_{i=1}^M \sum_{j=1}^{m_i}\left(n_{ij}-r_{ij}\right)};$$

   otherwise set $\hat{\phi} = \phi$.

3. Estimate Σ^2 by taking weighted average over all groups

$$\hat{\Sigma}^2 = \frac{\sum_{i=1}^M \sum_{j=1}^{m_i} n_{ij}\hat{\Sigma}_i^2}{\sum_{i=1}^M \sum_{j=1}^{m_i} n_{ij}}.$$

4. For each group i, let

$$\begin{aligned} \tilde{\boldsymbol{y}}_i &= \Omega_i^{-1/2}\left[\hat{\boldsymbol{\beta}}_i^T\ \hat{\boldsymbol{u}}_i^{1T}\right]^T, \\ \tilde{\boldsymbol{X}}_i &= \left(\Omega_i^{-1/2}\right)_{1:p}, \\ \tilde{\boldsymbol{Z}}_i &= \left(\Omega^{-1/2}\right)_{(p+1):(p+q_1)}. \end{aligned}$$

   Apply mhglm2L algorithm to $\left(\tilde{\boldsymbol{y}}_i, \tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{Z}}_i\right)_{i\in\{1,\ldots,M\}}$ with gaussian as link function. The returned estimates are $\hat{\boldsymbol{\beta}}, \hat{\Sigma}^1$ respectively.

3.3 Posterior Inference for Three Levels Model

Given the top-level mhglm2L results, we can directly perform posterior inference and get estimates $\hat{\boldsymbol{u}}_i^1, i=1,\ldots,M$. To get estimates for $\boldsymbol{u}_{ij}^2$, for every group i perform posterior inference with plug-in $\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}}_i^1 \end{bmatrix}, \hat{\Sigma}^2, \hat{\phi}$; the results are $\hat{\boldsymbol{u}}_{ij}^2, j=1,\ldots,m_i$.

3.4 More Levels and Notations

In general, the hierarchical GLM can have any number of levels. All the assumptions can be extended naturally. We will use the following notations throughout the paper.

Assume there are L levels of hierarchies (including the root). Let $g_1g_2\ldots g_{L-1}$ denote the path from top to bottom level: it belongs to group g_1 at the second level, and under g_1 it belongs to group g_2 at the third level, etc. At $(l+1)$ -level, for every group $g_1\ldots g_l$ there are $m_{g_1\ldots g_l}$ subgroups, $l \le L-2$.

Let m_0 be the number of groups at second level. For example, to apply this notation to the three-levels model above, we have $L = 3$, group ij is equivalent to group $g_1 g_2$ with $g_1 = i, g_2 = j$, and $m_0 = M$.

At $(l + 1)$-th level, for every path $g_1 \dots g_l$ there is a vector of $q_l$random effects shared across all subgroups under current group, denote as $\boldsymbol{u}^l_{g_1 \dots g_l}$. For every group $g_1 \dots g_{L-1}$ at bottom level, let $\boldsymbol{y}_{g_1 \dots g_{L-1}}$ denote the response vector, $\boldsymbol{X}_{g_1 \dots g_{L-1}}$ denote the predictor matrices for fixed effects. Moreover, let $\boldsymbol{Z}^l_{g_1 \dots g_{L-1}}, l = 1, \dots, L - 1$ denote the predictor matrices for random effects $u^l_{g_1 \dots g_l}, l = 1, \dots L - 1$ respectively.

For every group $g_1 \dots g_{L-1}$, conditional on random effects vectors $u^l_{g_1 \dots g_l}, l = 1, \dots L-1$, the following holds

$$\mathbb{E}\left(\boldsymbol{y}_{g_1 \dots g_{L-1}} | \boldsymbol{u}^1_{g_1}, \dots \boldsymbol{u}^{L-1}_{g_1 \dots g_{L-1}}\right) = f^{-1}_{g_1 \dots g_{L-1}} \left(\boldsymbol{X}_{g_1 \dots g_{L-1}} \boldsymbol{\beta} + \sum_{l=1}^{L-1} \boldsymbol{Z}^l_{g_1 \dots g_{L-1}} \boldsymbol{u}^l_{g_1 \dots g_l}\right)$$

$$\mathrm{Cov}\left(\boldsymbol{y}_{g_1 \dots g_{L-1}} | \boldsymbol{u}^1_{g_1}, \dots \boldsymbol{u}^{L-1}_{g_1 \dots g_{L-1}}\right) = \phi I$$

where $f_{g_1 \dots g_{L-1}}$ is some link function.

For level $(l + 1), l = 1, \dots L-1$, all the random effects vectors $u^l_g, g \in \{\text{all paths down to level } l + 1\}$ are i.i.d. distributed with mean zero and covariance $\mathrm{Cov}\left(u^l_g\right) = \Sigma^l$.

Finally, for convenience we define the pa(parent) and ch(children) mapping functions:

$$\begin{aligned}
\mathrm{pa}\left(g_1 \dots g_l\right) &= g_1 \dots g_{l-1} \\
\mathrm{pa}\left(g_1\right) &= \emptyset \\
\mathrm{ch}\left(g_1 \dots g_l\right) &= \{g_1 \dots g_{l+1} : \mathrm{pa}\left(g_1 \dots g_{l+1}\right) = g_1 \dots g_l\} \\
\mathrm{ch}\left(g_1 \dots g_{L-1}\right) &= \emptyset
\end{aligned}$$

3.5 Fitting Any Levels Model

We could generalize the fitting procedure to any number of levels. That is we could recursively apply mhglm2L from bottom to the top levels.