Gaussian Variational Approximate Inference and Monte Carlo EM Algorithm for Generalized Linear Mixed Models

STAT 540 Project Presentation

Qianying Zhou

11/29/2018

College of Information Sciences and Technology

Generalized Linear Mixed Models (GLMM)

- A generalized linear mixed model is an extension of the generalized linear model in which the linear predictor contains random effects in addition to the usual fixed effects.
- GLMMs are widely applied to the analysis of grouped data, since the differences among groups (from different distributions) can be modelled as random effects.
- The general form of the model is:

$$y = X\beta + Z\mu + \epsilon, \;\; \mu \sim N(0, G)$$

- Fitting GLMMs via maximum likelihood involves integrating over the random effects. In general, those integrals cannot be expressed in analytical forms.

GLMM

- Consider the exponential family models of the form:

$$\mathbf{y}|\mathbf{u} \sim \exp\{\mathbf{y}^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \mathbf{1}^T b(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) + \mathbf{1}^T c(\mathbf{y})\}, \qquad \mathbf{u} \sim N(\mathbf{0}, \mathbf{G}),$$

- The parameters in the exponential family models are the fixed effects vector \beta and the random effects covariance matrix G. Their loglikelihood is:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \sum_{i=1}^{m} \{\mathbf{y}_i^T \mathbf{X}_i \boldsymbol{\beta} + \mathbf{1}_i^T c(\mathbf{y}_i)\} - \frac{m}{2} \log|\boldsymbol{\Sigma}| - \frac{mK}{2} \log(2\pi)$$
$$+ \sum_{i=1}^{m} \log \int_{\mathbb{R}^K} \exp\left\{\mathbf{y}_i^T \mathbf{Z}_i \mathbf{u} - \mathbf{1}_i^T b(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}) - \frac{1}{2} \mathbf{u}^T \boldsymbol{\Sigma}^{-1} \mathbf{u}\right\} d\mathbf{u}$$

- The K-dimensional integral in the loglikelihood cannot be solved analytically

Gaussian Variational Approximate (GVA)

- GVA introduces a pair of variational parameters μ_i, Λ_i. By Jensen's inequality and concavity of the logarithm function, we can get the lower bound:

$$
\begin{aligned}
\ell(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = & \sum_{i=1}^{m} \{\mathbf{y}_i^T \mathbf{X}_i \boldsymbol{\beta} + \mathbf{1}_i^T c(\mathbf{y}_i)\} - \frac{m}{2} \log|\boldsymbol{\Sigma}| - \frac{mK}{2} \log(2\pi) \\
& + \sum_{i=1}^{m} \log \int_{\mathbb{R}^K} \exp\left\{ \mathbf{y}_i^T \mathbf{Z}_i \mathbf{u} - \mathbf{1}_i^T b(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}) - \frac{1}{2} \mathbf{u}^T \boldsymbol{\Sigma}^{-1} \mathbf{u} \right\} \frac{\phi_{\boldsymbol{\Lambda}_i}(\mathbf{u} - \boldsymbol{\mu}_i)}{\phi_{\boldsymbol{\Lambda}_i}(\mathbf{u} - \boldsymbol{\mu}_i)} d\mathbf{u} \\
\geq & \sum_{i=1}^{m} \{\mathbf{y}_i^T \mathbf{X}_i \boldsymbol{\beta} + \mathbf{1}_i^T c(\mathbf{y}_i)\} - \frac{m}{2} \log|\boldsymbol{\Sigma}| - \frac{mK}{2} \log(2\pi) \\
& + \sum_{i=1}^{m} E_{\mathbf{u} \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)} \Big(\mathbf{y}_i^T \mathbf{Z}_i \mathbf{u} - \mathbf{1}_i^T b(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}) \\
& - \frac{1}{2} \mathbf{u}^T \boldsymbol{\Sigma}^{-1} \mathbf{u} - \log(\phi_{\boldsymbol{\Lambda}_i}(\mathbf{u} - \boldsymbol{\mu}_i)) \Big) \\
\equiv & \underline{\ell}(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\Lambda}),
\end{aligned}
$$

GVA

- The advantage of the lower-bound is that it no longer involves the integrals of size K. Hence, the computational speed is improved.
- We can use Newton-Raphson scheme to get the Gaussian variational approximate maximum likelihood estimators.

Monte Carlo EM algorithm

- Consider the random effects **u** to be the missing data. The complete data here is **W=(Y, u)**
- The monte carlo EM algorithm is as follows:
    1. Choosing starting values $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\sigma}^{(0)}$, set **n=0**
    2. Generate *m* values, $\boldsymbol{u}^{(1)}, \boldsymbol{u}^{(2)}, \boldsymbol{u}^{(3)}, \ldots, \boldsymbol{u}^{(m)}$ from the conditional distribution of **u|Y** using a Metropolis algorithm (use **f(y|u)** as the proposal distribution) and using the current parameter values
    3. Choose:

        (1) $\boldsymbol{\beta}^{(n+1)}$ to maximize a Monte Carlo estimate of **E[ln(f(y|u))]**

        (2) $\boldsymbol{\sigma}^{(n+1)}$ to maximize **E[ln(f(u|$\boldsymbol{\sigma}$))]**

Simulation Study

- Dataset: *Epilepsy* dataset
- Description: 59 epilepsy patients; each of the patients was assigned to a control group (placebo) or a treatment group. The experiment recorded the number of seizures experienced by each patient over 4 two-week periods.
- Structure: 59 x 4 observations on the following 7 variables: count(y), log(base/4), trt, trt*log(base/4), log(age), subject (u), v4
- Each patient can be seen as a group, and we use subject, which is the id of the patient as the random effect
- Consider the Poisson random intercept model:

$$y_{ij} | u_i \sim Poisson(\exp(\beta^T x_{ij} + u_i))$$

Results

- Use adaptive Gauss-Hermite quadrature (AGHQ) here as "gold standard" when the true values of the parameters are not known.

	β_0	β_{base}	β_{trt}	$\beta_{base*trt}$	β_{age}	β_{v4}	σ_0	Time (seconds)
AGHQ	-1.325	0.883	-0.933	0.481	-0.160	0.339	0.251	0.901
GVA (6 ITER)	-1.325	0.883	-0.933	0.481	-0.160	0.339	0.251	0.066
Monte Carlo EM	-1.325	0.883	-0.933	0.481	-0.160	0.339	0.251	108