Gauss-Hermite approximation for GLMM

Wentao Li
Wentao.Li@uth.tmc.edu

November 9, 2020

Introduction

See the loglikelihood function

$$\log\{\mathcal{L}_i(\theta_0)\} = \sum_{i=1}^{n_i} \log\left\{\int_{\mu_{i0}}\left[\prod_{j=1}^{n_i}\mathbb{P}(y_{ij}|\pi_{ij})\right]\phi(\mu_{i0};\theta_0)\mathrm{d}\mu_{i0}\right\}$$

where π_{ij} is a function of μ_{i0} and defined as

$$\pi_{ij} = \frac{\exp(X_{ij}^{\top}\beta_0 + \mu_{i0})}{1 + \exp(X_{ij}^{\top}\beta_0 + \mu_{i0})}$$

and the distribution $\mathbb{P}$ follows density of logit and ϕ is a univariate normal, see that

$$\prod_{j=1}^{n_i}\mathbb{P}(y_{ij}|\pi_{ij}) = \prod_{j=1}^{n_i}\pi_{ij}^{y_{ij}}(1-\pi_{ij})^{(1-y_{ij})}$$

$$\phi(\mu_{i0};\theta_0) = \frac{1}{\sqrt{2\pi}\tau_0}\exp(-\mu_{i0}^2/2\tau_0^2)$$

Note that $f_{\theta_0}(\mu_{i0}) := \prod_{j=1}^{n_i}\mathbb{P}(y_{ij}|\pi_{ij})\phi(\mu_{i0};\theta_0)$, and see the log term inside the loglikelihood function that

$$\int_{\mu_{i0}} f_{\theta_0}(\mu_{i0})\mathrm{d}\mu_{i0} = \int_{\mu_{i0}} e^{\log f_{\theta_0}(\mu_{i0})}\mathrm{d}\mu_{i0} = \int_{\mu_{i0}} e^{g(\mu_{i0})}\mathrm{d}\mu_{i0} \tag{1}$$

Here, instead of using Laplace approximation, we consider using Gauss-Hermite approximation. And as the definition of it, physicists' version of Hermite polynomial $H_k(x)$ and weight h_k is defined as followings,

$$H_k(x) = (-1)^k e^{x^2}\frac{d^k}{dx^k}e^{-x^2}$$

$$h_k = \frac{2^{k-1}k!\sqrt{\pi}}{k^2\left[H_{k-1}(x_k)\right]^2}$$

where x_k are the roots of $H_k(x) = 0$.

Thus, with the Gauss-Hermite approximation, (1) can be approximate by

$$\int_{\mu_{i0}} e^{g(\mu_{i0})}\mathrm{d}\mu_{i0} \approx \sqrt{2\pi}\hat{\omega}\sum_{k=1}^{l} h_k \exp\left\{g(\hat{\mu}_{i0} + \sqrt{2\pi}\hat{\omega}x_k) + x_k^2\right\}, \quad \hat{\omega} = \sqrt{-\frac{1}{g''(\hat{\mu}_{i0})}}$$

notice that when $k = 1$, it is a Laplace approximation.

Expansion on the problem

Goal

$$\arg\max \log \mathcal{L} = \arg\max \sum_{i=1}^{m} \log \mathcal{L}_i$$

Notation

- i: Index of sites
- j: Index of patients in a specific site
- k: Index of Hermite polynomial
- l: Number of Hermite polynomial
- m: Number of sites
- n_i: Number of patients in site i
- $\mathcal{L}_i$: Log-likelihood function for site i
- $\theta_0 := (\beta_0, \tau_0)$
- $\phi := (\sqrt{2\pi}\tau_0)^{-1} \exp(-\mu_{i0}^2 / 2\tau_0^2)$

Derivatives

Denote that

$$\mathcal{L}_i = \mathcal{L}_i(\beta_0, \mu_{i0}; X_{i\cdot}, y_{i\cdot}) = \sqrt{2\pi}\hat{\omega} \sum_{k=1}^{l} h_k \exp\left\{ g(\hat{\mu}_{i0} + \sqrt{2\pi}\hat{\omega} x_k; \beta_0) + x_k^2 \right\}$$

h_k and $\hat{\omega}$ are defined as above. And

$$\begin{aligned} g(\mu_{i0}; \beta_0) &= \log \prod_{j=1}^{n_i} \mathbb{P}(y_{ij} | \pi_{ij}) \phi(\mu_{i0}; \theta_0) \\ &= \sum_{j=1}^{n_i} [\log \mathbb{P}(y_{ij} | \pi_{ij})] + \log \phi(\mu_{i0}; \theta_0) \\ &= \sum_{j=1}^{n_i} [y_{ij} \log \pi_{ij} + (1 - y_{ij}) \log(1 - \pi_{ij})] + \log \phi(\mu_{i0}; \theta_0) \end{aligned}$$

where

$$\pi_{ij} = \frac{\exp\left(X_{ij}^\top \beta_0 + \mu_{i0} \right)}{1 + \exp\left(X_{ij}^\top \beta_0 + \mu_{i0} \right)}$$

Step 1: Maximize $g(\mu_{i0})$

To maximize $g(\mu_{i0})$, we need to get the derivatives

$$\begin{aligned}\frac{\partial g}{\partial \mu_{i0}} &= \sum_{j=1}^{n_i}\left[y_{ij}\frac{1}{\pi_{ij}}\frac{\partial \pi_{ij}}{\partial \mu_{i0}} - (1-y_{ij})\frac{1}{1-\pi_{ij}}\frac{\partial \pi_{ij}}{\partial \mu_{i0}}\right] + \frac{1}{\phi}\frac{\partial \phi}{\partial \mu_{i0}} \\ &= \sum_{j=1}^{n_i}(y_{ij}-\pi_{ij}) - \frac{\mu_{i0}}{\tau_0^2}\end{aligned}$$

where

$$\frac{\partial \phi}{\partial \mu_{i0}} = (\sqrt{2\pi}\tau_0)^{-1}\exp(-\mu_{i0}^2/2\tau_0^2)\cdot(-\mu_{i0}/\tau_0^2)$$

and

$$\frac{\partial^2 g}{\partial \mu_{i0}^2} = -\sum_{j=1}^{n_i}\frac{\partial \pi_{ij}}{\partial \mu_{i0}} - \frac{1}{\tau_0^2} < 0$$

where

$$\frac{\partial \pi_{ij}}{\partial \mu_{i0}} = \frac{\exp\left(X_{ij}^\top \beta_0 + \mu_{i0}\right)}{\left[1+\exp\left(X_{ij}^\top \beta_0 + \mu_{i0}\right)\right]^2}$$

see that it is a convex problem, using newton's method, we can derive $\hat{\mu}_{i0} = \arg\max_{\mu_{i0}} g(\mu_{i0})$ that is global optimum.

Step 2: Maximization preparation of β_0 in LOCAL

See the derivative

$$\frac{\partial \pi_{ij}}{\partial \beta_0} = \frac{X_{ij}\exp\left(X_{ij}^\top \beta_0 + \mu_{i0}\right)}{\left[1+\exp\left(X_{ij}^\top \beta_0 + \mu_{i0}\right)\right]^2}$$

and denote $f_k(\hat{\mu}_{i0};\beta_0) := h_k \exp\left\{g(\hat{\mu}_{i0}+\sqrt{2\pi}\hat{\omega}x_k;\beta_0) + x_k^2\right\}$, then

$$\frac{\partial \mathcal{L}_i}{\partial \beta_0} = \sqrt{2\pi}\hat{\omega}\sum_{k=1}^{l}\left\{f_k(\hat{\mu}_{i0};\beta_0)\cdot \left.\frac{\partial g(\mu_{i0};\beta_0)}{\partial \beta_0}\right|_{\mu_{i0}=\hat{\mu}_{i0}+\sqrt{2\pi}\hat{\omega}x_k}\right\} \tag{2}$$

$$= \sqrt{2\pi}\hat{\omega}\sum_{k=1}^{l}\left\{f_k(\hat{\mu}_{i0};\beta_0)\sum_{j=1}^{n_i}(X_{ij}y_{ij} - X_{ij}\pi_{ij})\right\} \tag{3}$$

and the second derivative

$$\frac{\partial^2 \mathcal{L}_i}{\partial \beta_0^2} = \sqrt{2\pi}\hat{\omega}\sum_{k=1}^{l}\Bigg\{f_k(\hat{\mu}_{i0};\beta_0)\sum_{j=1}^{n_i}(X_{ij}y_{ij} - X_{ij}\pi_{ij})\left[\sum_{j=1}^{n_i}(X_{ij}y_{ij} - X_{ij}\pi_{ij})\right]^\top \tag{4}$$

$$+ f_k(\hat{\mu}_{i0};\beta_0)\sum_{j=1}^{n_i}\left(-X_{ij}\frac{\partial \pi_{ij}}{\partial \beta_0}\right)\Bigg\} \tag{5}$$

Notice that μ_{i0} in (3), (4) and (5) are replaced by $\hat{\mu}_{i0}+\sqrt{2\pi}\hat{\omega}x_k$ where $\hat{\mu}_{i0}$ is the maximand of function $g(\cdot)$ with respect to μ_{i0}.

Step 3: Maximization of β_0 in GLOBAL

Reminds that $\mathcal{L} = \sum_{i=1}^{m} \log \mathcal{L}_i$, then another Newton's method is applied in global log-likelihood function,

$$\frac{\partial \mathcal{L}}{\partial \beta_0} = \sum_{i=1}^{m} \frac{\mathcal{L}_i'(\beta_0)}{\mathcal{L}_i(\beta_0)} \qquad \frac{\partial^2 \mathcal{L}}{\partial \beta_0^2} = \sum_{i=1}^{m} \left[\frac{\mathcal{L}_i''(\beta_0)}{\mathcal{L}_i(\beta_0)} - \left(\frac{\mathcal{L}_i'(\beta_0)}{\mathcal{L}_i(\beta_0)} \right)^2 \right]$$

where

$$\frac{\mathcal{L}_i'(\beta_0)}{\mathcal{L}_i(\beta_0)} = \frac{\sum_{k=1}^{l} \left\{ f_k(\hat{\mu}_{i0}; \beta_0) \sum_{j=1}^{n_i} (X_{ij} y_{ij} - X_{ij} \pi_{ij}) \right\}}{\sum_{k=1}^{l} f_k(\hat{\mu}_{i0}; \beta_0)}$$

and

$$\frac{\mathcal{L}_i''(\beta_0)}{\mathcal{L}_i(\beta_0)} =$$

$$\frac{\sum_{k=1}^{l} \left\{ f_k(\hat{\mu}_{i0}; \beta_0) \left[\sum_{j=1}^{n_i} (X_{ij} y_{ij} - X_{ij} \pi_{ij}) \left[\sum_{j=1}^{n_i} (X_{ij} y_{ij} - X_{ij} \pi_{ij}) \right]^\top - \sum_{j=1}^{n_i} (X_{ij} y_{ij} - X_{ij} \pi_{ij}) \left[\sum_{j=1}^{n_i} \left(X_{ij} \frac{\partial \pi_{ij}}{\partial \beta_0} \right) \right]^\top \right] \right\}}{\sum_{k=1}^{l} f_k(\hat{\mu}_{i0}; \beta_0)}$$

Now, focus on $\beta_0^{(n+1)} = \beta_0^{(n)} - \dfrac{\mathcal{L}'(\beta_0^{(n)})}{\mathcal{L}''(\beta_0^{(n)})}$, deduce that

$$\frac{\mathcal{L}'(\beta_0^{(n)})}{\mathcal{L}''(\beta_0^{(n)})} = \frac{\sum_{i=1}^{m} \frac{\mathcal{L}_i'(\beta_0)}{\mathcal{L}_i(\beta_0)}}{\sum_{i=1}^{m} \frac{\mathcal{L}_i''(\beta_0)}{\mathcal{L}_i(\beta_0)} - \sum_{i=1}^{m} \left(\frac{\mathcal{L}_i'(\beta_0)}{\mathcal{L}_i(\beta_0)} \right)^2}$$

Summary

With a more rigorous math works, I fix the confusion of index and actually the number of polynomial can affect the direction in Newton's method. And I also notice that under such model, the GLMM is inherited separable, given the local log-likelihood $\mathcal{L}_i$ and sum them up to global log-likelihood $\mathcal{L} = \sum_{i}^{m} \mathcal{L}_i$ for further computation.