Advanced Data Analysis

Laplace's Method and PQL

Exponential Family GLMM

$$\log[\boldsymbol{y}|\boldsymbol{u}] = \{\boldsymbol{y}^T(\boldsymbol{X\beta}+\boldsymbol{Zu})-\boldsymbol{1}^T b(\boldsymbol{X\beta}+\boldsymbol{Zu})\}/\phi+\boldsymbol{1}^T c(\boldsymbol{y}, \phi$$

$$\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{G})$$

Maximum Likelihood Estimation

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{G}, \phi) &= [\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{G}] \\
&= \int_{\mathbb{R}^q} [\boldsymbol{y}, \boldsymbol{u}]\, d\boldsymbol{u} \\
&= \int_{\mathbb{R}^q} [\boldsymbol{y}|\boldsymbol{u}][\boldsymbol{u}]\, d\boldsymbol{u} \\
&= (2\pi)^{-q/2}|\boldsymbol{G}|^{-1/2} \int_{\mathbb{R}^q} \exp[\{\boldsymbol{y}^T(\boldsymbol{X\beta}+\boldsymbol{Zu}) + \boldsymbol{1}^T c(\boldsymbol{y}) \\
&\qquad -\boldsymbol{1}^T b(\boldsymbol{X\beta}+\boldsymbol{Zu})\}/\phi - \tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{G}^{-1}\boldsymbol{u}\}]\, d\boldsymbol{u}
\end{aligned}$$

GLMMs Big Headache

The likelihood involves an
intractable integral
(often high-dimensional).

Ways Around Intractable Integral Problem

- Analytic integral approximations.
  - Most popular has acronym PQL.
  - However often criticised for being ‘too’ approximate.
- Convert to Bayesianism and use MCMC.
  - Gets around the intractable integral headache.
  - But introduces new MCMC headaches (e.g. implementation, convergence, computing time).
  - However made easier since release of `BRugs` in 2005.

Mathematical Tools for Advanced Data Analysis

- Distribution theory.
- Matrix algebra.
- Vector differential calculus.
- Markov chain Monte Carlo.
- Analytic integral approximation.

Univariate Laplace's Method

$$I = \int_{-\infty}^{\infty} f(x)\,dx, \qquad f(x) > 0.$$

$$I \simeq e^{h(x_0)} \sqrt{\frac{2\pi}{-h''(x_0)}}$$

where $h(x) = \ln\{f(x)\}$ and x_0 satisfies

$$h'(x_0) = 0.$$

Whiteboard Interlude I

This is to derive and illustrate the use of:

Univariate Laplace's Method.

Multivariate Laplace's Method

$$I = \int_{\mathbb{R}^d} f(\boldsymbol{x})\,d\boldsymbol{x}, \qquad f(\boldsymbol{x}) > 0.$$

$$I \simeq e^{h(\boldsymbol{x}_0)} \sqrt{\frac{(2\pi)^d}{-|\mathsf{H}h(\boldsymbol{x}_0)|}}$$

where $h(\boldsymbol{x}) = \ln\{f(\boldsymbol{x})\}$ and $\boldsymbol{x}_0$ satisfies

$$\mathsf{D}h(\boldsymbol{x}_0) = \mathbf{0}.$$

Whiteboard Interlude II

This is to derive:

Penalised Quasi-Likelihood (PQL)

for exponential family GLMMs.

Passages from McCulloch & Searle (2000, p.283)

Breslow and Lin (1995) and Lin and Breslow (1996) show that PQL methods lead to estimators which are asymptotically biased and hence inconsistent......

Unfortunately, for situations like paired binary data the PQL estimator can perform quite badly. Its performance improves as $[\boldsymbol{y}|\boldsymbol{u}]$ gets closer to normal....

However, from a practical point of view, we may prefer to transform such data to make them approximately normal and use LMM methods.

McCulloch & Searle's (2000) Bottom Line

We thus

cannot recommend

the use of simple PQL methods in practice.

A Place for PQL?

- Benchmark for GLMM fitting via PQL.
- PQL can provide reasonable initial values for MCMC $\longrightarrow$ better convergence.
- Some GLMM situations require fast, albeit less accurate, fitting. MCMC can be simply too slow.

(Example: GLMM-based classification; Kauermann, Ormerod & Wand; 2010 ; *Journal of Classification*)

Lecturer's Conclusion

Analytic integral approximation methods such as

Laplace's Method

do have a place in GLMM

(and hence advanced data analysis in general).