Computation of the likelihood function for GLMMs - Monte Carlo methods

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

November 17, 2023

Outline for today

- Monte Carlo methods
- Computation of the likelihood function using importance sampling
- Newton-Raphson and the EM-algorithm

Generalized linear mixed effects models

Consider stochastic variable $\mathbf{Y} = (Y_1, \ldots, Y_n)$ and random effects $\mathbf{U}$.

Two step formulation of GLMM:

- $\mathbf{U} \sim N(0, \Sigma)$.
- Given realization $\mathbf{u}$ of $\mathbf{U}$, Y_i independent and each follows density $f_i(y_i|\mathbf{u})$ with mean $\mu_i = g^{-1}(\eta_i)$ and linear predictor $\eta = X\beta + Z\mathbf{u}$.

I.e. conditional on $\mathbf{U}$, Y_i follows a generalized linear model.

NB: GLMM specified in terms of marginal density of $\mathbf{U}$ and conditional density of $\mathbf{Y}$ given $\mathbf{U}$. But the likelihood is the marginal density of $f(\mathbf{y})$ which can be hard to evaluate !

Likelihood for generalized linear mixed model

For normal linear mixed models we could compute the marginal distribution of $\mathbf{Y}$ directly as a multivariate normal. That is, $f(\mathbf{y})$ is a density of a multivariate normal distribution.

For a generalized linear mixed model it is difficult to evaluate the integral:

$$f(\mathbf{y}) = \int_{\mathbb{R}^m} f(\mathbf{y}, \mathbf{u}) \mathrm{d}\mathbf{u} = \int_{\mathbb{R}^m} f(\mathbf{y}|\mathbf{u}) f(\mathbf{u}) \mathrm{d}\mathbf{u}$$

since $f(\mathbf{y}|\mathbf{u})f(\mathbf{u})$ is a very complex function.

Monte Carlo computation of likelihood for GLMM

Likelihood function is an expectation:

$$L(\theta) = f(y;\theta) = \int_{\mathbb{R}} f(y|u;\beta) f(u;\tau^2) \mathrm{d}u = \mathbb{E}_{\tau^2} f(y|U;\beta)$$

Use Monte Carlo simulations to approximate expectation.

NB: also applicable in high dimensions

However, naive methods may fail !

Simple Monte Carlo: $g(u) = f(u;\tau^2)$

$$L(\theta) = \int_{\mathbb{R}} f(y|u;\beta) f(u;\tau^2)\mathrm{d}u = \mathbb{E}_{\tau^2} f(y|U;\beta) \approx L_{SMC}(\theta) =$$

$$\frac{1}{M}\sum_{l=1}^{M} f(y|U^l;\beta) \quad \text{where } U^l \sim N(0,\tau^2) \text{ independent}$$

Monte Carlo variance:

$$\mathbb{V}\mathrm{ar}(L_{SMC}(\theta)) = \frac{1}{M}\mathbb{V}\mathrm{ar} f(y|U^1;\beta)$$

NB: variance is of order $1/M$ regardless of dimension !

NB: large M is needed to get large reduction of standard error which is of order $1/\sqrt{M}$

Estimate $\mathbb{V}\mathrm{ar} f(y|U^1;\beta)$ using empirical variance estimate based on $f(y|U^l;\beta)$, $l = 1,\dots,M$:

$$\frac{1}{M-1}\sum_{l=1}^{M}(f(y|U^l;\beta) - L_{SMC}(\theta))^2$$

Often $\mathbb{V}\mathrm{ar} f(y|U^1;\beta)$ is large so large M is needed.

Increasing dimension leads to worse performance (useless in high dimensions)

Importance sampling

Consider $Z \sim f$ and suppose we wish to evaluate $\mathbb{E}h(Z)$ where $h(Z)$ has large variance.

Suppose we can find density g so that

$$\frac{h(z)f(z)}{g(z)} \approx \text{const} \quad \text{and } h(z)f(z) > 0 \Rightarrow g(z) > 0$$

Then

$$\mathbb{E}h(Z) = \int \frac{h(z)f(z)}{g(z)} g(z)\mathrm{d}z = \mathbb{E}\frac{h(Y)f(Y)}{g(Y)}$$

where $Y \sim g$.

Note variance of $h(Y)f(Y)/g(Y)$ small so estimate

$$\mathbb{E}h(Z) \approx \frac{1}{n}\sum_{i=1}^{n} \frac{h(Y_i)f(Y_i)}{g(Y_i)}$$

has small Monte Carlo error.

Importance sampling for GLMM

$g(\cdot)$ probability density on $\mathbb{R}^m$.

$$L(\theta) = \int_{\mathbb{R}} f(y|u;\beta) f(u;\tau^2) \mathrm{d}u = \int_{\mathbb{R}} \frac{f(y|u;\beta)f(u;\tau^2)}{g(u)} g(u) \mathrm{d}u = \mathbb{E}\frac{f(y|V;\beta)f(V;\tau^2)}{g(V)} \text{ where } V \sim g(\cdot).$$

$$L(\theta) \approx L_{IS}(\theta) = \frac{1}{M} \sum_{l=1}^{M} \frac{f(y|V^l;\beta)f(V^l;\tau^2)}{g(V^l)} \text{ where } V^l \sim g(\cdot), l = 1, \ldots,$$

Find g so $\mathbb{V}\mathrm{ar} \frac{f(y|V;\beta)f(V;\tau^2)}{g(V)}$ small.

$\mathbb{V}\mathrm{ar} L_{IS}(\theta) < \infty$ if $f(y|v;\theta)f(v;\theta)/g(v)$ bounded (i.e. use $g(\cdot)$ with heavy tails).

Possibility: Note

$$\frac{f(y|u,\beta)f(u;\tau^2)}{f(u|y;\theta)} = f(y;\theta) = L(\theta) = \text{const}$$

Laplace: $U|Y = y \approx N(\mu_{LP}, \sigma^2_{LP})$

Use $g(\cdot)$ density for $N(\mu_{LP}, \sigma^2_{LP})$ or $t_\nu(\mu_{LP}, \sigma^2_{LP})$-distribution:

$$\frac{f(y|u,\beta)f(u;\tau^2)}{g(u)} \approx \text{const}$$

so small variance.

Simulation straightforward.

Note: ‘Monte Carlo version’ of adaptive Gaussian quadrature.

Possibility: Consider fixed θ_0:

$$g(u) = f(u|y,\theta_0) = f(y|u;\theta_0)f(u;\theta_0)/L(\theta_0)$$

Then

$$L(\theta) = \int_{\mathbb{R}} \frac{f(y|u;\beta)f(u;\tau^2)}{g(u)} g(u)\mathrm{d}u = L(\theta_0)\int_{\mathbb{R}} \frac{f(y|u;\beta)f(u;\tau^2)}{f(y|u;\beta_0)f(u;\tau_0^2)} f(u|y,\theta_0)\mathrm{d}u =$$

$$L(\theta_0)\mathbb{E}_{\theta_0}\Big[\frac{f(y|U;\beta)f(U;\tau^2)}{f(y|U;\beta_0)f(U;\tau_0^2)}|Y=y\Big] \Leftrightarrow \frac{L(\theta)}{L(\theta_0)} = \mathbb{E}_{\theta_0}\Big[\frac{f(y|U;\beta)f(U;\tau^2)}{f(y|U;\beta_0)f(U;\tau_0^2)}|Y = y$$

So we can estimate ratio $L(\theta)/L(\theta_0)$ where $L(\theta_0)$ is unknown constant.

This suffices for finding MLE:

$$\arg\max_\theta L(\theta) = \arg\max_\theta \frac{L(\theta)}{L(\theta_0)} \approx \arg\max_\theta \frac{1}{M}\sum_{l=1}^{M} \frac{f(y|U^l;\beta)f(U^l;\tau^2)}{f(y|U^l;\beta_0)f(U^l;\tau_0^2)}$$

where $U_l \sim f(u|y;\theta_0)$

Simulation from $U|Y = y$

Problems:

- $$f(u|y) = \frac{f(y|u)f(u)}{f(y)}$$

  is a non-standard density.
- We know numerator $f(y|u)f(u)$ but we do not know normalizing constant

  $$f(y) = \int f(y|u)f(u)\mathrm{d}u$$

  which is the likelihood

How to simulate a non-standard density $f(z) = h(z)/c$ where we only know $h(z)$ - i.e. only know

$$f(z) \propto h(z)$$

Simulation of random variables

Direct methods exists for many standard distributions (normal, binomial, t, etc.: `rnorm()`, `rbinom()`, `rt()` etc.)

Suppose $f(z) \propto h(z)$ is a non-standard density but

$$h(z) \leq Kg(z)$$

for some constant K and standard density g.

Then we may apply rejection sampling:

1. Generate $Y \sim g$ and $W \sim \text{unif}[0,1]$.
2. If $W \leq \frac{h(Y)}{Kg(Y)}$ return Y (accept); otherwise go to 1 (reject).

Note probability of accept is c/K where $c = \int h(z)\mathrm{d}z$.

If f is high-dimensional density it may be hard to find g with small K so rejection sampling mainly useful in small dimensions.

MCMC is then useful alternative (we'll briefly consider this in MRF part of course)

Proof of rejection sampling:

Show

$$P(Y \leq y | accept) = P\big(Y \leq y | W \leq \frac{h(Y)}{Kg(Y)}\big) = \int_{-\infty}^{y} f(v)\mathrm{d}v \quad (1)$$

Hint: write out $P(Y \leq y, W \leq \frac{h(Y)}{Kg(Y)})$ as integral in terms of the densities of Y and W.

Also recall $c = \int h(v)\mathrm{d}v$ and $f(v) = h(v)/c$.

Prediction of U using conditional simulation

Compute Monte Carlo estimate of $\mathbb{E}(U|Y=y)$ using importance sampling or conditional simulations of $U|Y=y$:

$$\mathbb{E}(U|Y=y) \approx \frac{1}{M}\sum_{m=1}^{M} U^m, \quad U^m \sim f(u|y)$$

We can also evaluate e.g. $P(U_i > c|y)$ or $P(U_i > U_l, l \neq i|Y)$ etc.

Conditional simulation of $U|Y = y$ using rejection sampling

Note

$$f(u|y;\theta) \propto f(y|u;\beta)f(u;\tau^2) \leq K\, t_\nu(u;\mu_{LP},\sigma^2_{LP})$$

for some constant K.

Rejection sampling:

1. Generate $V \sim t_\nu(\mu_{LP},\sigma^2_{LP})$ and $W \sim \text{Unif}(]0,1[)$
2. Return V if $W \leq f(y|V;\beta)f(V;\tau^2)/(K\, t(V;\mu_{LP},\sigma^2_{LP}))$; otherwise go to 1.

Maximization of likelihood using Newton-Raphson

Let

$$V_\theta(y,u) = \frac{\mathrm{d}}{\mathrm{d}\theta} \log f(y,u|\theta)$$

Then

$$u(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} \log L(\theta) = \mathbb{E}_\theta[V_\theta(y,U)|Y=y]$$

and

$$\begin{aligned} j(\theta) &= -\frac{\mathrm{d}^2}{\mathrm{d}\theta^\mathsf{T}\mathrm{d}\theta} \log L(\theta) \\ &= -\big(\mathbb{E}_\theta[\mathrm{d}V_\theta(y,U)/\mathrm{d}\theta^\mathsf{T}|Y=y] + \mathbb{V}\mathrm{ar}_\theta[V_\theta(y,U)|Y=y]\big) \end{aligned}$$

Newton-Raphson:

$$\theta_{l+1} = \theta_l + j(\theta_l)^{-1} u(\theta_l)$$

All unknown expectations and variances can be estimated using the previous numerical integration or Monte Carlo methods !

EM-algorithm

Given current estimate θ_l:

1. (E) compute $Q(\theta_l, \theta) = \mathbb{E}_{\theta_l}[\log f(y, U|\theta)|Y = y]$
2. (M) $\theta_{l+1} = \text{argmax}_\theta\, Q(\theta_l, \theta)$.

For LNMM E-step can be computed explicitly but seems pointless as likelihood is available in closed form.

For GLMMs (E) step needs numerical integration or Monte Carlo.

Convergence of EM-algorithm can be quite slow. Maximization of likelihood using Newton-Raphson seems better alternative.

Exercises

1. why heavy-tailed importance sampling density ? (show that variance of Monte Carlo estimate is finite when importance sampling ratio is bounded)
2. `R` exercises on exercise-sheet `exercises_imp.pdf`. Note code (unfortunately with a few mistakes) available on website.
3. Show that the rejection sampler works - i.e. equation (1)
4. Simulate a binomial distribution ($n = 10, p = 0.2$) using simulations of a Poisson distribution (mean 2) and rejection sampling. What is the acceptance rate ? Can you simulate a Poisson using simulations of a binomial ?