Computation of the likelihood function for GLMMs

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

April 30, 2012

Outline for today

- Computation of likelihood function
  1. Gaussian quadrature
  2. Monte Carlo methods
- Newton-Raphson
- EM-algorithm
- case study of non-linear mixed effects model

Generalized linear mixed effects models

Consider stochastic variable $\mathbf{Y} = (Y_1, \dots, Y_n)$ 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 $f_i(\mathbf{y}|\mathbf{u})$ 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.

Adaptive Gaussian Quadrature - one random effect

Gauss-Hermite quadrature (numerical integration) is

$$\int_{\mathbb{R}} f(x)\phi(x)\mathrm{d}x \approx \sum_{i=1}^{n} w_i f(x_i)$$

where ϕ is the standard normal density and $(x_i, w_i), i = 1, n$ are certain arguments and weights which can be looked up in a table.

We can replace $\approx$ with $=$ whenever f is a polynomial of degree $2n-1$ or less.

Adaptive Gauss-Hermite quadrature:

$$\int f(y|u)f(u)\mathrm{d}u \approx \int \frac{f(y|u)f(u)}{\phi(u;\mu_{LP},\sigma^2_{LP})}\phi(u;\mu_{LP},\sigma^2_{LP})\mathrm{d}u = \int \frac{f(y|\sigma_{LP}x+\mu_{LP})f(\sigma_{LP}x+\mu_{LP})}{\phi(x)}\sigma_{LP}\phi(x)\mathrm{d}x$$

(change of variable: $x = (u-\mu_{LP})/\sigma_{LP}$)

Advantage

$$\frac{f(y|u)f(u)}{\phi(u;\mu_{LP},\sigma^2_{LP})} = \frac{f(y|\sigma_{LP}x+\mu_{LP})f(\sigma_{LP}x+\mu_{LP})}{\phi(x)} \quad x = (u-\mu_{LP})/\sigma_{LP}$$

close to constant $(f(y))$ – hence adaptive G-H quadrature very accurate.

GH scheme with $n = 5$:

x	2.020	0.959	0.0000000	-0.959	-2.020
w	0.011	0.222	0.533	0.222	0.011

(x's are roots of Hermite polynomial computed using `ghq` in library `glmmML`).

(GH schmes for $n = 5$ and $n = 10$ available on web page)

Prediction of random effects for GLMM

Conditional mean

$$\mathbb{E}[U|Y=y] = \int uf(u|y)\mathrm{d}u$$

is minimum mean square error predictor, i.e.

$$\mathbb{E}(U-\hat{U})^2$$

is minimal with $\hat{U} = H(Y)$ where $H(y) = \mathbb{E}[U|Y=y]$

Difficult to analytically evaluate

$$\mathbb{E}[U|Y=y] = \int uf(y|u)f(u)/f(y)\mathrm{d}u$$

Computation of conditional expectations (prediction)

$$\mathbb{E}[U|Y=y] = \int u\frac{f(y|u)f(u)}{f(y)}\mathrm{d}u = \frac{1}{f(y)}\int (\sigma_{LP}x+\mu_{LP})\frac{f(y|\sigma_{LP}x+\mu_{LP})f(\sigma_{LP}x+\mu_{LP})}{\phi(x)}\sigma_{LP}\phi(x)\mathrm{d}x$$

Note:

$$(\sigma_{LP}x+\mu_{LP})\frac{f(y|\sigma_{LP}x+\mu_{LP})f(\sigma_{LP}x+\mu_{LP})}{\phi(x)}\sigma_{LP}$$

behaves like a first order polynomial in x - hence GH still accurate.

Difficult cases for numerical integration - dimension $m > 1$

- correlated random effects
- crossed random effects
- nested random effects

Not possible to factorize likelihood into low-dimensional integrals

Number of quadrature points $\approx k^m$ where k is number of quadrature points for $1D$ – hence numerical quadrature may not be feasible.

Alternatives: PQL and Laplace-approximation or *Monte Carlo computation*.

Monte Carlo computation of likelihood for GLMM

Likelihood function is an expectation:

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

Use Monte Carlo simulations to approximate expectation.

NB: also applicable in high dimensions

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 $\mathbb{R}^m$.

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

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

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

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

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:

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

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

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

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)

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

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

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

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

This suffices for finding MLE:

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

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

Simulation of random variables

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

Suppose f is a non-standard density but

$$f(z) \leq Kg(z)$$

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 $W \leq \frac{f(Y)}{Kg(Y)}$ return Y (accept); otherwise go to 1 (reject).

Note probability of accept is $1/K$.

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 (later).

Proof of rejection sampling:

compute $P(Y \leq y|accept) = P(Y \leq y|W \leq \frac{f(Y)}{Kg(Y)})$

Prediction of U using conditional simulation

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

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

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

$$f(y|u; \beta_0)f(u; \tau_0^2)/f(y; \theta_0) \leq K\, t_\nu(u; \mu_{LP}, \sigma_{LP}^2)$$

for some constant K.

Rejection sampling:

1. Generate $V \sim t_\nu(\mu_{LP}, \sigma_{LP}^2)$ and $W \sim \text{Unif}(]0, 1[)$
2. Return V if $W \leq f(V|y; \theta_0)/(K\, t(V; \mu_{LP}, \sigma_{LP}^2))$; 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

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

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) $\theta_{l+1} = \mathrm{argmax}_\theta\, Q(\theta_l,\theta)$.

For LNMM E-step can be computed explicitly (Jiang page 165) 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.

Overview of Jiang page 165-171

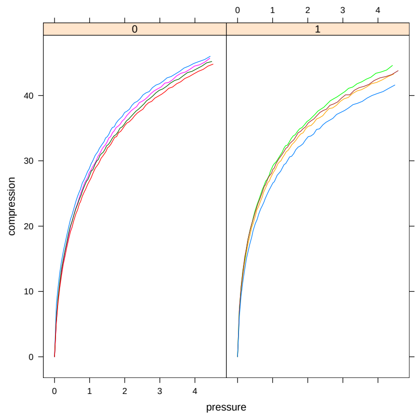
Page 165-167 (Monte Carlo) EM for linear mixed model and threshold model.

Page 168-169: Newton-Rahpson and importance sampling.

Page 170-171: Monte Carlo EM with adaptive importance sampling (t-distribution + Laplace approximation)

Case study: non-linear mixed effects model for cow mats data

Compression vs. pressure for two brands of mats



Non-linear relation

$$y = \mathrm{mmf}(x) = \frac{ab + cx^d}{b + x^d},$$

Random variation between mats of same brand, small measurement noise.

Estimation of non-linear model with fixed effects:

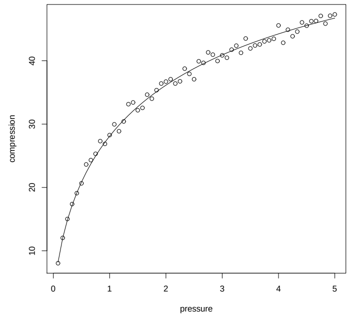
```
nlsfit=nls(nedtryk~mmf(tryk,a,b,c,d),start=
   c(a=0.1,b=1.670,c=80,d=0.6),data=mattressdata1)
```

Estimation of non-linear model with a, b, c as random effects:

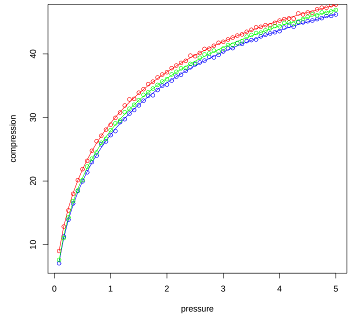
```
nlmerfit=nlmer(nedtryk~mmfnlmer(tryk,a,b,c,d)~(a|matno)+
(b|matno)+(c|matno),mattressdata1,start=c(a=0.04,b=1.64,c=74,d=0.64))
```


Simulated data from the two models:

Fixed effects: residual standard error 0.72



With random effects: residual standard error 0.17



Std. err. for a, b, c are $0.64, 0.05$ and 0.14

Random effects model gives much better representation of variability in data.

NB: to assess influence of variability of different parameters we need to look at partial derivatives (sensitivities) wrt. these parameters.

Exercises

1. Exercise 4.3 on page 229 in Jiang.
2. R exercises on exercise-sheets exercises6.pdf and exercises7.pdf.
3. Check formulas for $u(\theta)$ and $j(\theta)$. How can these expression be computed using importance sampling ?