Computation of the likelihood function for GLMMs

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

November 5, 2007

Outline for today

- likelihood for GLMM
- penalized quasi-likelihood estimation
- Laplace approximation
- Gaussian quadrature
- case study of non-linear mixed effects model

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 $f_i(\mathbf{y}|\mathbf{u})$ 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 models.

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.

Today: numerical methods for evaluating likelihood of GLMM.

Non-normal example: logistic regression with random intercepts

$$U_j \sim N(0,\tau^2),\ j=1,\dots,m$$
$$Y_j|U_j=u_j \sim \text{binomial}(n_j,p_j)$$
$$\log(p_j/(1-p_j))=\eta_j=\beta+U_j$$
$$p_j=\exp(\eta_j)/(1+\exp(\eta_j))$$

Conditional density:

$$f(y|u;\beta)=\prod_j p_j^{y_j}(1-p_j)^{1-y_j}=\prod_j \frac{\exp(\beta+u_j)^{y_j}}{(1+\exp(\beta+u_j))^{n_j}}$$

Likelihood function ($u=(u_1,\dots,u_m)$)

$$\int_{\mathbb{R}^m} f(y|u;\beta)f(u;\tau^2)\mathrm{d}u=\prod_j\int_{\mathbb{R}}\frac{\exp(\beta+u_j)^{y_j}}{(1+\exp(\beta+u_j))^{n_j}}\frac{\exp(-u_j^2/(2\tau^2))}{\sqrt{2\pi\tau^2}}\mathrm{d}u_j$$

Integrals can not be evaluated in closed form.

One-dimensional case

Compute

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

Possibilities:

- Laplace approximation.
- Numerical integration/quadrature (e.g. Gaussian quadrature as in `PROC NLMIXED` (SAS) or `GLLAM` (STATA)) (one level of random effects, dimensions one or two).

Laplace approximation

Let $g(u)=\log(f(y|u)f(u))$ and choose $\hat{u}$ so $g'(\hat{u})=0$ ($\hat{u}=\arg\max g(u)$).

Taylor expansion around $\hat{u}$:

$$g(u)\approx\tilde{g}(u)=$$
$$g(\hat{u})+(u-\hat{u})g'(\hat{u})+\frac{1}{2}(u-\hat{u})^2g''(\hat{u})=g(\hat{u})-\frac{1}{2}(u-\hat{u})^2(-g''(\hat{u}))$$

I.e. $\exp(\tilde{g}(u))$ proportional to normal density $N(\mu_{LP},\sigma^2_{LP})$, $\mu_{LP}=\hat{u}\ \sigma^2_{LP}=-1/g''(\hat{u})$.

$$L(\theta)=\int_{\mathbb{R}}\exp(g(u))\mathrm{d}u\approx\int_{\mathbb{R}}\exp(\tilde{g}(u))\mathrm{d}u$$
$$=\exp(g(\hat{u}))\int_{\mathbb{R}}\exp\big(-\frac{1}{2\sigma^2_{LP}}(u-\mu_{LP})^2\big)\mathrm{d}u=\exp(g(\hat{u}))\sqrt{2\pi\sigma^2_{LP}}$$

Laplace approximation also possible for higher dimensions (multivariate Taylor expansion).

NB:

$$f(u|y)=f(y|u)f(u)/f(y)\propto\exp(g(u))\approx \mathit{const}\exp\big(-\frac{1}{2\sigma^2_{LP}}(u-\mu_{LP})^2\big)$$

where $\mu_{LP}=\hat{u}\ \sigma^2_{LP}=,-1/g''(\hat{u})$.

Hence

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

Note: μ_{LP} is mode of conditional distribution - used for prediction of random effects in `lmer` (`ranef()`).

Adaptive Gaussian Quadrature

Gauss-Hermite quadrature (numerical integration) is

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

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 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

(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

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

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

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

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.

Penalized quasi-likelihood

One solution: do not use likelihood function but something simpler.

$\theta = (\beta, \tau^2)$

PQL estimates $\hat{\theta}$ and $\hat{u}$ maximize joint density

$$f(y, u; \theta) = f(y|u; \beta) f(u; \tau^2).$$

PQL estimates less accurate than ML.

Asymptotic results require increasing number of observations for each random effect.

Implemented in `lmer` and SAS macro `glimmix`.

Difficult cases

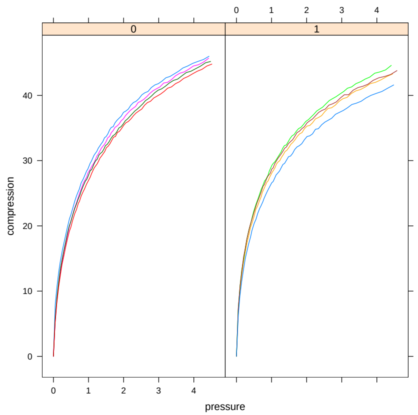
- correlated random effects
- crossed random effects
- nested random effects

Not possible to factorize likelihood into low-dimensional integrals – hence numerical quadrature not applicable.

PQL and Laplace-approximation still applicable (`lmer()`).

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:

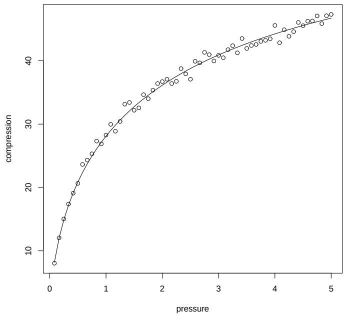
```
nlsfit=nls(nedtryk~mmf(tryk,a,b,c,d),start=
   c(a=0.1,b=1.670,c=80,d=0.6),data=mattressdata1)
```

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

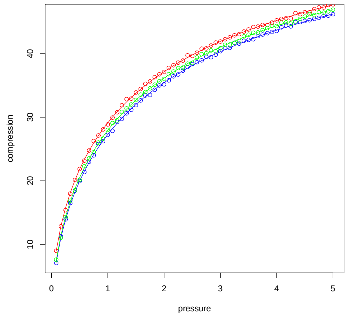
```
nlmerfit=nlmer(nedtryk~mmfnlmer(tryk,a,b,c,d)~(a|matno)+
(b|matno)+(c|matno),mattressdata1,start=c(a=0.04,b=1.64,c=74,d=0.64))
```

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

See exercises sheet on webpage.