Adaptive Gauss-Hermite Quadrature for Generalized Linear or Nonlinear Mixed Models

Douglas Bates
Department of Statistics
University of Wisconsin – Madison

May 5, 2009

Abstract

Currently the `lme4` package for `R` provides functions for fitting generalized linear mixed models (GLMMs) or nonlinear mixed models (NLMs) by minimizing the Laplace approximation to the deviance as a function of the fixed-effects parameters, $\boldsymbol{\beta}$, and the parameters $\boldsymbol{\theta}$ that determine the (relative) variance-covariance matrices of the random effects. The Laplace approximation is a special case of a more general class of quadrature methods called Adaptive Gauss-Hermite Quadrature (AGQ). We describe the steps that would be needed to implement AGQ for specific types of GLMMs or NLMs.

1 Numerical quadrature in mixed models

For a generalized linear mixed model (GLMM) the conditional mean,

$$\boldsymbol{\mu}_{\boldsymbol{Y}|\boldsymbol{b}} = \mathrm{E}[\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{B}} = \boldsymbol{b}], \tag{1}$$

of the n-dimensional response random variable, $\boldsymbol{\mathcal{Y}}$, given a value, $\boldsymbol{b}$, of the q-dimensional random effects vector, $\boldsymbol{\mathcal{B}}$, is a function of the linear predictor,

$$\boldsymbol{\eta}(\boldsymbol{\beta}, \boldsymbol{b}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}, \tag{2}$$

where $\boldsymbol{\beta}$ is the p-dimensional fixed-effects parameter and the model matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ are $n \times p$ and $n \times q$, respectively.

The vector-valued "inverse link" function

$$\boldsymbol{g}^{-1} : \boldsymbol{\eta} \mapsto \boldsymbol{\mu}, \tag{3}$$

is defined component-wise from the scalar inverse link $g^{-1} : \eta \mapsto \mu$. (The reasons why this is called the "inverse link" and not the "link" are technical and related to the history of the development of generalized linear models.) Thus the ith element of $\boldsymbol{\mu}$ depends only on the ith element of $\boldsymbol{\eta}$ and the Jacobian matrix, $d\boldsymbol{\mu}/d\boldsymbol{\eta}'$, of this mapping is a diagonal matrix.

Elements of $\boldsymbol{\mathcal{Y}}$ are conditionally independent, given $\boldsymbol{\mathcal{B}} = \boldsymbol{b}$. The conditional distribution of each element of $\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{B}}$ belongs to some well-defined family of distributions such as the binomial family or the Poisson family or the gamma family. For the binomial or Poisson families the conditional distribution of an element of $\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{B}}$ depends only on the conditional mean. Other families, like the gamma, incorporate an additional scale parameter that is common to all the elements. We will write the common scale parameter as σ, even when it does not, by itself, represent the standard deviation of the conditional distribution.

We observe the value, $\boldsymbol{y}$, of $\boldsymbol{\mathcal{Y}}$. We do not observe the value of $\boldsymbol{\mathcal{B}}$. The marginal distribution of the random effects is a multivariate normal distribution with mean $\mathbf{0}$ and a variance-covariance matrix, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, that depends on a parameter vector, $\boldsymbol{\theta}$. (Typically the dimension of $\boldsymbol{\theta}$ is much, much smaller than q.)

Although the conditional distribution, $\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{B}} = \boldsymbol{b}$, may be either discrete or continuous, the distribution $\boldsymbol{\mathcal{B}}|\boldsymbol{\mathcal{Y}} = \boldsymbol{y}$ is always continuous, with density that we will write $f_{\boldsymbol{B}|\boldsymbol{Y}}(\boldsymbol{b}|\boldsymbol{y})$. We can evaluate $f_{\boldsymbol{B}|\boldsymbol{Y}}(\boldsymbol{b}|\boldsymbol{y})$, up to a constant, from the conditional distribution, $\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{B}}$, evaluated at $\boldsymbol{y}$ and $\boldsymbol{b}$, and the marginal density of $\boldsymbol{\mathcal{B}}$, evaluated at $\boldsymbol{b}$. Let us write this unnormalized density as $h(\boldsymbol{b}, \boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma)$. The likelihood of the parameters, given the data, $\boldsymbol{y}$, is

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma|\boldsymbol{y}) = \int_{\mathbb{R}^q} h(\boldsymbol{b}, \boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma)\, d\boldsymbol{b}. \tag{4}$$

The maximum likelihood estimates, $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\theta}}$ and $\widehat{\sigma}$, are the values that jointly maximize $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma|\boldsymbol{y})$ for the observed data $\boldsymbol{y}$. Naturally we must be able to evaluate, or at least approximate, $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma|\boldsymbol{y})$ before we can optimize it with respect to the parameters. Numerical approximations to such integrals is called numerical quadrature.

1.1 Terms and grouping factors

When a GLMM is fit using `lmer`, the model matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ and the form of the variance-covariance matrix, $\Sigma(\boldsymbol{\theta})$ are defined by the model formula as it is applied to the data. Consider the `Contraception` data set from the `mlmRev` package

```
> data(Contraception, package = "mlmRev")
> str(Contraception <- Contraception[, -1])
```

```
'data.frame':	1934 obs. of  5 variables:
 $ district: Factor w/ 60 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ use     : Factor w/ 2 levels "N","Y": 1 1 1 1 1 1 1 1 1 1 ...
 $ livch   : Factor w/ 4 levels "0","1","2","3+": 4 1 3 4 1 1 4 4 2 4 ...
 $ age     : num  18.44 -5.56 1.44 8.44 -13.56 ...
 $ urban   : Factor w/ 2 levels "N","Y": 2 2 2 2 2 2 2 2 2 2 ...
```

```
> summary(Contraception)
```

```
    district     use       livch          age             urban   
 14     : 118   N:1175   0 :530   Min.   :-13.5600   N:1372  
 1      : 117   Y: 759   1 :356   1st Qu.: -7.5599   Y: 562  
 46     :  86            2 :305   Median : -1.5599           
 25     :  67            3+:743   Mean   :  0.0022           
 6      :  65                     3rd Qu.:  6.4400           
 30     :  61                     Max.   : 19.4400           
 (Other):1420                                                
```

In the model

```
> print(fm1 <- glmer(use ~ urban + age + I(age^2) +
+     livch + (1 | district), Contraception, binomial),
+     corr = FALSE)
```

```
Generalized linear mixed model fit by the Laplace approximation
Formula: use ~ urban + age + I(age^2) + livch + (1 | district)
   Data: Contraception
  AIC  BIC logLik deviance
 2389 2433  -1186     2373
Random effects:
 Groups   Name        Variance Std.Dev.
 district (Intercept) 0.226    0.475   
Number of obs: 1934, groups: district, 60

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.035075   0.174361   -5.94  2.9e-09
```

```
urbanY       0.697269   0.119879    5.82  6.0e-09
age          0.003533   0.009231    0.38      0.7
I(age^2)    -0.004562   0.000725   -6.29  3.2e-10
livch1       0.815054   0.162190    5.03  5.0e-07
livch2       0.916496   0.185100    4.95  7.4e-07
livch3+      0.915019   0.185769    4.93  8.4e-07
```

the binary response variable, `use` (whether the woman uses artifical contraception), is modeled with a conditional binomial distribution according to the formula

```
use ~ urban + age + I(age^2) + livch + (1 | district)
```

Random effects terms in a formula are distinguished by the presence of the vertical bar character (‘`|`’) in the term. There is one random effects term, `(1|district)`, in this formula. The expression on the right hand side of the `|` is evaluated as a factor, called the *grouping factor* for the term. The expression on the left hand side of the `|` is evaluated as a linear model formula in the model frame, creating a model matrix. For this term, `(1|district)`, the right hand expression, `district`, indicates that the grouping factor is the district and the left hand expression, `1`, denotes the trivial model matrix with a single column, all of whose entries are unity.

Let k be the number of random-effects terms in the formula. For the ith term we write n_i for the number of levels of the grouping factor and q_i for the number of columns in the model matrix formed from the left hand side. For this model $k = 1$, $n_1 = 60$ and $q_1 = 1$. The dimension of the random effects vector is $q = \sum_{i=1}^{k} n_i q_i$.

The $n \times q$ model matrix, $\boldsymbol{Z}$, consists of k groups of columns where the ith group has $n_i q_i, i = 1, \ldots, k$ columns. Each of these groups is further divided into q_i subgroups, each with n_i columns. The pattern of nonzeros in each subgroup is that of the indicators of the ith grouping factor. The values of the nonzero elements in the jth subgroup are the jth column of the model matrix from the left hand side of the term.