Dimension Reduction of Random Effects for Generalized Linear Mixed Models

Christina Knudson

July 21, 2020

Abstract

something

1 Theory

Let $y = (y_1, \ldots, y_n)^T$ be a vector of observed data. Let $u = (u_1, \ldots, u_q)'$ be a vector of unobserved random effects. Let β be a vector of p fixed effect parameters and let ν be a vector of T variance components for the random effects. Let θ be a vector of length $p + T$ containing all unknown parameters. Then the data y are distributed conditionally on the random effects according to $f_\theta(y|u)$ and the random effects are distributed according to $f_\theta(u)$. Although $f_\theta(u)$ does not actually depend on β and $f_\theta(y|u)$ does not depend on ν, we write the density like this to keep notation simple in future equations.

Since u is unobservable, the log likelihood must be expressed by integrating out the random effects:

$$l(\theta) = \log \int f_\theta(y|u) f_\theta(u) \, du \tag{1}$$

2 Model fitting function

This will be the main function the user will use. The users will need to specify the response and the predictors using the R formula mini-language as interpreted by model.matrix. They'll need to specify the family (either binomial or Poisson, though really any exponential family would work and I could add more later). The user will specify the random effects in the same way as for the R function `reaster` in the R package `aster` (Geyer, 2014). That is, random effects will be expressed using the R formula mini-language. Thus, a sample command with fixed predictors x_1 and x_2 and with random effects *school* and *classroom* (in data set as categorical variables) would look like

```
glmm(y ~ x1+ x2, list(~0+school,~0+classroom),  family.glmm="binomial.glmm",
data=schooldat,varcomps.names=c("school","classroom"),varcomps.equal=c(1,2),
debug=FALSE )
```

Section 3 contains more information on the family.

3 Families

This function will be hidden from the user. These functions (along with the distribution of random effects) are necessary to approximate the log likelihood.

3.1 Distribution of random effects is normal (`distRand`)

In this subsection, I discuss both the assumed distribution of the unobserved random effects ($N(0, D)$) and the distribution used to generate the simulated random effects. The equations in this section will provide $\tilde{f}(u)$, $\log f_\theta(u)$, $\nabla \log f_\theta(u)$, and $\nabla^2 \log f_\theta(u)$ for equation **??**.

References

Geyer, C. J. (2014). R package `aster` (aster models), version .8-30. *http://cran.r-project.org/package=aster*.