A fixed effects approach to GLMs with clustered data

Göran Broström

Department of Statistics
Umeå University
SE–901 87 Umeå, Sweden

1 Introduction

In situations where a large data set is partitioned into many relatively small groups, and you want to test for group differences, the number of parameters tend to increase with sample size. This fact causes the standard assumptions underlying asymptotic results to be violated. There are (at least) two possible solutions to the problem, first, a random intercepts model, and second, a fixed effects model, where asymptotics are replaced by a simple form of bootstrapping.

In the `glmML` package, both these approaches are implemented. In this paper, only the fixed effects approach is considered.

2 The fixed effects model

In the fixed effects model, testing is performed via a simple bootstrap. Under the null hypothesis of no grouping effect, the grouping factor can be randomly permuted without changing the probability distribution. This is one basic idea in the estimation of the p-value by simulation. The direct parametric aproach is to draw bootstrap samples from the estimated probability distribution.

We first show how to write down the log-likelihood function and all the first and second partial derivatives. Then we introduce the profiling approach which reduces an optimizing problem in high dimensions to a problem consisting of numerically solving several one-variable equations and optimization in low dimensions. The profiling cannot be done explicitely, but it is possible via *inplicit differentiation*.

The procedure is implemented in `glmmML` for the *Binomial* and *Poisson* families of distributions. Some comparisons with alternative approaches are made by simulation.