

A Bayesian approach to non-ignorable missing response in generalized linear mixed models

Christian Heumann, LMU Muenchen, Department of Statistics

Email: Christian.Heumann@stat.uni-muenchen.de

Abstract

Missing values are an important special case of coarsened data (Heitjan and Rubin, 1991). Most often, cases with missing values in some of the studied variables are simply ignored in the analysis (in practice). As a rule of thumb we can say that whenever the studied outcome or response variable is missing and the probability of missing is dependent on that outcome variable, simple methods like complete case analysis, last observation carried forward (LOCF) etc. can lead to an inefficient and/or a biased estimation. In this talk we treat the problem of non-ignorable missing response in generalized linear mixed models (GLMM), i.e. models with random effects, for non-normal, especially binary data. This type of models is often used for longitudinal or clustered binary data. The concepts can of course also be applied to normal or multicategorical data. Recently, a connection of mixed models with smoothing has been formulated and thus the applied methods may be of interest beyond what is studied in the talk. Direct Maximum Likelihood estimation is complicated for GLMMs and therefore a Monte Carlo EM algorithm has been proposed in the literature (Ibrahim et al., 2001). While point estimates from such an algorithm are usually reliable when the log-likelihood behaves well, variance estimates are hard to get. Missing values are complicating the estimation process additionally so that a Bayesian approach may be a valuable and attractive alternative. In the talk we develop such an approach based on selection models and shared parameter models and show how modern Markov Chain Monte Carlo methods can be applied. In a number of simulation studies we examined the performance of the so called hybrid Monte Carlo (HMC) algorithm which avoids the random walk behaviour of simple MCMC algorithms by using auxiliary variables and the derivative of the posterior distribution. The mixing behaviour is observed to be good which reduces the needed length of the Markov Chain considerably and thus the runtime. Some cautionary notes for applying this MCMC algorithm are given in the talk. A data example will also be presented.

Keywords: Missing data, generalized linear mixed models (GLMM), non-ignorable, Bayesian estimation, selection model, hybrid Monte Carlo