Efficient MCMC Estimation of Mixed Effects Binary Logit Models

Master's thesis by Markus Hainy †

May 2011

Abstract

Mixed effects binary logit models are an important family of regression models for the analysis of panel data when the dependent variable is dichotomous. The model may include *fixed* as well as *random* (unit-specific) effects.

The standard way of estimating mixed effects logit models in a Bayesian context is to construct a Markov chain for the parameters by sampling from the conditional marginal posterior distributions (*Gibbs* sampling), complemented by *Metropolis-Hastings (MH)* steps in order to sample from analytically intractable marginal posteriors. The problem-specific optimal choice of the proposal distribution for the MH step is critical for the performance of the sampler.

Newly developed data augmentation schemes introduce latent utility variables which linearize the model. The linearized model is called *random utility model (RUM)*. It is often preferable to work with the differences of the utilities (*difference RUM* or *dRUM*). The *data augmentation, frequentist estimation (DAFE)* algorithm uses the mean and variance of the non-normal errors of the linearized model to automatically construct an appropriate proposal distribution for the MH step. *Auxiliary mixture samplers* approximate the non-normal error by a mixture of normal distributions. These algorithms involve the sampling of additional latent mixture component indicators, but no MH step is needed anymore.

The standard methods (using random walk and independence MH implementations) and the data augmentation methods are adapted for mixed effects binary logit models and implemented in `R`. The methods are applied to two data sets which could not be fully estimated using classical techniques due to data sparsity. The performances of the algorithms are compared based on measures such as integrated autocorrelation time and effective sampling rate.

Keywords: Binomial data; data augmentation; Markov chain Monte Carlo; logit model; random utility model; mixed effects models

† Johannes Kepler University Linz