A Tutorial on Fitting Linear Mixed Models Using JAGS and Stan

Tanner Sorensen[1], Shravan Vasishth[1,2]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

[1] University of Potsdam, Department of Linguistics, Potsdam, Germany
`tsorense@uni-potsdam.de`

[2] University of Sheffield, School of Mathematics and Statistics, Sheffield, UK
`vasishth@uni-potsdam.de`

Abstract

Linear mixed modeling using the package `lme4` is a standard statistical tool in psychology and linguistics research. Recent developments in statistical computing—most recently, the arrival of Stan—have made it relatively easy to fit complex linear mixed models in a fully Bayesian framework. However, it may not be obvious to users of frequentist tools such as `lme4` how such Bayesian models can be defined. We provide a tutorial showing how the most common linear mixed models (repeated measures designs with a full variance-covariance structure in the random effects) can be fit and evaluated using JAGS and Stan.

Keywords: Hierarchical linear modeling, repeated measures factorial designs

1 Introduction

Ever since the arrival of the package `lme4` [1], the use of linear mixed models in psychology and linguistics has increased dramatically. In the present tutorial, we show how standard models in psychology and linguistics can be fit easily using Bayesian tools such as JAGS [4] and Stan [5]. Our presentation focuses on practical details, in order to allow the reader to quickly start writing their own models. For simplicity, we focus on two basic designs: a two-condition repeated measures study, and a 2×2 repeated measures factorial design. All code and data are provided here: http://www.ling.uni-potsdam.de/∼vasishth/BayesLMMs.html.

There are no prerequisites apart from having some exposure to fitting linear mixed models using `lme4`, and having the relevant software installed: JAGS, Stan, `rjags`, `rstan` in R, and any associated software; see the JAGS (http://mcmc-jags.sourceforge.net) and Stan (mc-stan.org) websites for details. Stan is the more general programming language for psycholinguistic research because it allows the researcher to flexibly fit fairly complex models. However, we chose to introduce JAGS first because it uses BUGS syntax, which is currently widely used in textbooks; anyone learning to do Bayesian analysis would need to understand BUGS syntax in order to read introductory books.

2 A typical repeated measures design

We show how to fit the following two types of models using JAGS and Stan, for a two-condition experiment, and for a 2×2 factorial design. More complex designs are extensions of these two basic types. In this abstract, we only discuss the two-condition case, using simulated data.

Suppose we have a two-condition experiment, where the dependent variable is reading time (RT) and the predictor is a sum-contrast coded categorical predictor, condition; assume that we have several participants $j = 1, \dots, J$, and each participant has been exposed to multiple items $k = 1, \dots, K$. The data are indexed by $i = 1, \dots, I$. Using `lme4`, a standard model defines a crossed varying intercepts and varying slopes structure:

$$\text{RT}_{ijk} = \beta_0 + u_{0j} + w_{0k} + (\beta_1 + u_{1j} + w_{1k})\text{condition}_i + \epsilon_i \tag{1}$$

The `lme4` syntax for this model is:

```
m1<-lmer(rt~condition+(1+condition|subj)+(1+condition|item),data)
```

The varying intercepts u_{0j} and w_{0k} are adjustments to the fixed intercept β_0, and the varying slopes u_{1j} and w_{1k} are adjustments to the fixed slope β_1. We assume that these adjustments are normally distributed with mean 0 and some unknown variance. Importantly, we also assume that the varying intercepts and slopes for participants and for items are correlated. Model 1 is useful in psycholinguistic research because it faithfully reflects all the sources of variance in the experimental design. The experiment design suggests a natural partitioning of RT into groups associated with a given subject or a given item. These groups are specified by considering RT_{ijk} with either the subject index j or the item index k held constant. Groups defined along these lines display systematically different patterns of variance. For example, by-subject variance in language comprehension tasks has been attributed to factors such as individual differences in processing speed [3]. Cognitive models of performance in such tasks often make predictions about the relationship between a subject's random intercept and slope. For instance, a fast reader might be expected to take less time resolving an ungrammaticality on the rationale that processing speed in normal reading reflects the processing speed for resolving ungrammaticalities. Such a prediction can be evaluated experimentally by estimating the correlation between u_0 and u_1. A positive correlation between the varying intercepts and varying slopes would support such a prediction.

If the correlation between the subjects' random intercepts u_{0j} and random slopes u_{1j} is to be estimated, the model must specify that u_{0j} and u_{1j} covary. This simply means that the different variances are not mutually independent. The assumption is that u_0 and u_1 are normally distributed with mean 0 and with variance and covariance given by the matrix Σ_u, given below. The parameters in Σ_u are unknown, and so we must estimate them. The parameter ρ_u indicates the correlation between u_0 and u_1; it is our estimate of ρ_u by which we evaluate predictions about individual differences in experimental conditions. A variance-covariance matrix Σ_w can likewise be defined for by-item random effects.

$$\Sigma_u = \begin{bmatrix} \sigma^2_{\text{u0}} & \rho_u\, \sigma_{u0}\sigma_{u1} \\ \rho_u\, \sigma_{u0}\sigma_{u1} & \sigma^2_{u1} \end{bmatrix} \tag{2}$$

$$\Sigma_w = \begin{bmatrix} \sigma^2_{\text{w0}} & \rho_w\, \sigma_{w0}\sigma_{w1} \\ \rho_w\, \sigma_{w0}\sigma_{w1} & \sigma^2_{w1} \end{bmatrix} \tag{3}$$

To state all of this more formally, the linear mixed model for this particular example can be specified by assuming that the varying intercepts and slopes by participants and by items have the following bivariate distribution:

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Sigma}_u\right) \quad \begin{pmatrix} w_{0k} \\ w_{1k} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Sigma}_w\right) \tag{4}$$

The variance components associated with participants and items are generally treated as nuisance parameters in psycholinguistic work. They are generally included in the model only to discount the possibility that the fixed effects estimates depend on the particular subjects and items used in the experiment. However, as mentioned above, there are situations where these variance components can be of intrinsic theoretical interest; an example is where we have a theory about how reading speed of a participant affects the magnitude of their ungrammaticality effect.

In addition to explaining how these models are specified in JAGS and Stan, we also show how to evaluate model fit; here, we follow the recommendations in [2]. Gelman and colleagues recommend posterior predictive checks: generating samples from the posterior distribution and comparing the properties of these samples with the data.

Finally, we discuss why it is worth going through all the effort of learning to fit Bayesian linear mixed models. The most important reason is that, even though the experimental design demands that a full variance-covariance structure be specified, we often do not have enough data to estimate all the variance components. `lme4` will often fail to estimate the parameters in such a situation, or fail to converge. In the Bayesian framework, in such situations the priors will dominate in determining the posterior distributions. A related point is that `lme4` can often deliver estimates of correlation parameters that bear little relation to the true value; this becomes a serious issue when these parameters are of theoretical interest. In the Bayesian setting, these estimates will tend to be quite conservative. A further advantage of the Bayesian framework is that models can be flexibly altered to reflect our assumptions about how the data were generated. This allows for unprecedented flexibility in fitting linear mixed models that better reflect the underlying generating process.

References

[1] Douglas Bates and Deepayan Sarkar. *lme4: Linear mixed-effects models using S4 classes*, 2007. R package version 0.9975-11.

[2] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, third edition, 2014.

[3] Reinhold Kliegl, Michael E.J. Masson, and Eike M. Richter. A linear mixed model analysis of masked repetition priming. *Visual Cognition*, 18(5):655–681, 2010.

[4] Martyn Plummer. Jags version 3.3.0 manual. *International Agency for Research on Cancer. Lyon, France*, 2012.

[5] Stan Development Team. Stan: A C++ library for probability and sampling, version 2.1, 2013.