Notes on inference and model selection with mixed effect models

Marco D. Visser[*1,2]

[1]Departments of Experimental Plant Ecology and Animal Ecology
& Ecophysiology, Radboud University Nijmegen, the Netherlands
[2]Smithsonian Tropical Research Institute, Panama

November 13, 2013

Contents

1 Introduction

Mixed models (MM) are a popular tool in ecology today, one that is rapidly gaining popularity as MM have become easy implement in most software packages (SAS, SPSS, R:lme4). The popularity of MM is likely largely due to the fact that ecological datasets often violate the assumptions of classical statistical tests. There is also increased interest in directly estimating variance (between individuals, or in space and time) as theoretical studies emphasize the effects of variability on e.g. population dynamics (Pfister & Stevens, 2003). However, recent studies have shown the majority of studies in ecology (58 - 95%) used these tools inappropriately (Bolker *et al.*, 2009). Much is still unknown about mixed models, this document aims to summarize some of the key issues when

---

[*]m.visser@science.ru.nl

trying to apply model selection and inference to MM, without going into great detail.

2 HOW MAXIMUM LIKELIHOOD ESTIMATE ARE APPROXIMATED. A KEY CONSIDERATION.

To obtain Maximum Likelihood estimated for mixed models with random effects one must integrate likelihoods overall possible values of the random effects. For instance, if we are studying a system of organisms (e.g. seedlings, owlets, daphnia) and we were interested in the variation in survival over time, as well as the "classical" mean survival. The system could be described by:

$$S_t \sim bin(p, N_{t-1}) \tag{2.1}$$

Where S_t are the amount of surviving individuals at time t, from an original population of N_{t-1} and p is a random variable distributed as $p \sim beta(\alpha, \beta)$. The likelihood of observing a set of S survivors, from N individuals over T years, given the parameters *alpha* and *beta* would be:

$$L(\alpha, \beta \mid S, N) = \prod_{t=1}^{T} \left[\int_0^1 beta(p \mid \alpha, \beta) bin(S_t, N_{t-1} \mid p) dp \right] \tag{2.2}$$

The example of integrating over all values of the the "random effect" p to obtain the MLE for *alpha* and *beta* is one of the few cases where an analytical solution exists (called the beta-binomial). However, in most cases, no analytical solution exists and integration must be done numerically. Even for simple problems this quickly becomes infeasible. For these reasons statisticians have come up ways to approximate the MLE of model parameters including random effects. These techniques include:

1. Pseudo and penalized quasi-likelihoods [PQL]

2. Laplace Approximations [LA]

3. Guasse-Hermite quadrature [GHQ]

4. Monte Carlo Markov Chain methods [MCMC]

All of the above can again be distinguished between standard ML estimation, in which the fixed effect parameters are assumed to be precisely correct when estimating the random effects (as above in 2.2), or restricted maximum likelihood (REML) which averages over uncertainty in the fixed effects (Pinheiro & Bates, 2000). It is good to consider the precise method used in approximating the MLE, as this has serious consequences for model inference and selection.

3   A GUIDE TO MIXED-MODEL SELECTION

Each of the above mentioned approximation methods, have key benefits and key disadvantages. I list some considerations below, for each method:

1. PQL is fast, yet yields biased estimated when variances in random effects (the sd's) are large. PQL also gives quasi-likelihood which many statisticians feel cannot be used in inference (Wald, Z and T statistics) or selection (e.g. AIC, DIC). Basically all inferences based on the likelihood are invalid (Joe, 2008). P-values and CI are complicated to calculate [1].

2. LA, is less biased, and approaches the real ML and can therefore be used in likelihood based inference test - however it assumes the likelihood distribution is approximately normal. It is also slower and less flexible than PQL. P-values and CI are complicated as in PQL.

3. GHQ, are even more precise than LA and also approach the real ML. It is much slower than LA and fitting models with more than 2–3 random effects is not considered feasible. P-values and CI are complicated as in PQL.

4. MCMC methods are highly flexible, can handle many random-effects, and are theoretically well founded with the "Bayasian Framework". CI intervals and "p-values" (i.e. quantiles) are simple to calculate from the posterior distribution samples. They are notoriously slow however, and technically challenging to implement. MCMC methods give very similar answers to the above methods when datasets are informative and priors weak.

4   PROBLEMS ASSOCIATED WITH RANDOM EFFECTS

Statistics as Z, t, χ^2 and F are poor for models containing random effects as standard deviations are strictly positive (≥ 0), and thus violate the null hypothesis assumption ($\sigma = 0$). Likelihood ratio tests are also problematic and highly unsuited for PQL estimates. Although when using LA and GHQ, likelihood ratio test can be used on random effects in some cases, however not if using REML. Guidelines on these issues are sparse, and I will update this section if more information arises.

Another considerable problem in model selection and inference, is how to decide how many parameters a models has (in AIC) r df (in t, χ^2 and F) when including random effects. How many parameters to you effectively have? With random effects included there is no straight answer, and simply counting the random effect sds is considered wrong. I have found no satisfactory answer to this, barring the use of DIC with MCMCs. DIC uses pD ('the effective number

---

[1]see Douglas Bates rant on the matter: https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html

of parameters') instead of the number of parameters (Spiegelhalter *et al.*, 2002). The idea behind pD is that it is a more appropriate measure of model complexity than parameters alone which may say little of how complex a model is to fit.

One strategy that can be used in combination with model selection tools as AIC in combination with random effects, is to select among model with the same random effects fit to the same data. This still leaves the question on how to select among models with different random effects.

5   Please contribute

If you have any comments or suggestion to improve this document, and make it a more comprehensive guide to inference with Mixed Effect models. Please add your suggestion through [2] or fork this repository on github [3].

References

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.S.S. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, **24**, 127–135.

Joe, H. (2008) Accuracy of laplace approximation for discrete response mixed models. *Computational Statistics & Data Analysis*, **52**, 5066–5074.

Pfister, C.A. & Stevens, F.R. (2003) Individual variation and environmental stochasticity: implications for matrix model predictions. *Ecology*, **84**, 496–510.

Pinheiro, J.C. & Bates, D.M. (2000) *Linear mixed-effects models: basic concepts and examples.* Springer.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van Der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 583–639.

---

[2]`http://github.com/MarcoDVisser/mmnotes/issues`
[3]`http://github.com/MarcoDVisser/mmnotes`