Likelihood ratio testing for zero variance components in linear mixed models

Sonja Greven[1,3], Ciprian Crainiceanu[2], Annette Peters[3] and Helmut Küchenhoff[1]

[1] Department of Statistics, LMU Munich University, Munich, Germany, sonja.greven@stat.uni-muenchen.de,
[2] Department of Biostatistics, Johns Hopkins University, Baltimore, USA,
[3] Institute of Epidemiology, GSF-National Research Center for Environment and Health, Neuherberg, Germany

Abstract: We consider the problem of testing for zero variance components in linear mixed models. Typical applications include testing for a random intercept or testing for linearity of a smooth function. We propose two approximations to the finite sample null distribution of the restricted likelihood ratio test statistic. Our approach applies to a wider variety of mixed models than previous results, including those with moderate numbers of clusters, unbalanced designs, or nonparametric smoothing. Extensive simulations show that both proposed approximations outperform the $0.5\chi^2_0 : 0.5\chi^2_1$ approximation and parametric bootstrap currently used.
Our methods are motivated by and applied to the longitudinal epidemiological study Airgene, with the aim of assessing non-linearity of dose-response-functions between ambient air pollution concentrations and inflammation.

Keywords: Boundary; Bootstrap; Penalized splines; Nonparametric smoothing; Air pollution.

1 Introduction

Linear mixed models are widely used to model longitudinal or clustered data and, more recently, to estimate smoothing parameters for penalized splines using REML or ML. We focus on linear mixed models of the form

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z_1 b_1} + \ldots + \boldsymbol{Z_S b_S} + \boldsymbol{\varepsilon}, \tag{1}$$

with random effects $\boldsymbol{b_s} \sim N(\mathbf{0}, \sigma_s^2 \boldsymbol{I_{K_s}})$ pairwise independent and independent of $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{I_n})$, K_s columns in $\boldsymbol{Z_s}$, $\boldsymbol{I_\nu}$ the identity matrix of size ν, and n the sample size.
We are interested in testing one of the variance components

$$H_{0,s} : \sigma_s^2 = 0 \quad \text{versus} \quad H_{A,s} : \sigma_s^2 > 0, \tag{2}$$

corresponding, for example, to testing for a zero random intercept or testing linearity against a general alternative. This problem is non-standard due to the parameter on the boundary of the parameter space. Stram and Lee (1994), using results from Self and Liang (1987), showed that the Likelihood Ratio Test (LRT) statistic for testing (2) has an asymptotic $0.5\chi_0^2 : 0.5\chi_1^2$ null distribution if $\boldsymbol{Y}$ can be divided into independent and identically distributed (i.i.d.) subvectors. However, for penalized spline smoothing responses are not independent at least under the alternative, and longitudinal studies often have unbalanced data or only moderate numbers of subjects. Crainiceanu and Ruppert (2004) derived the finite sample and asymptotic null distribution of the LRT and restricted LRT (RLRT) for testing (2) in models with one variance component ($S = 1$), and showed that it is generally different from $0.5\chi_0^2 : 0.5\chi_1^2$. For $S > 1$, they recommend a parametric bootstrap, which can be computationally very expensive. As the LRT has been seen to have undesirable properties with a high probability mass at zero, we develop two faster approximations to the finite sample null distribution of the RLRT.

2 Two approximations to the RLRT null distribution

2.1 Fast finite sample approximation

Our first approximation is inspired by pseudo-likelihood estimation (Gong and Samaniego, 1981), where nuisance parameters are replaced by consistent estimators. Liang and Self (1996) showed that under certain regularity conditions the asymptotic distribution of the pseudo LRT is the same as that of the LRT if the nuisance parameters are known. For our problem, we could view the $\boldsymbol{b_i}$, $i \neq s$, as nuisance parameters. We assume that under regularity conditions the prediction of $\sum_{i \neq s} \boldsymbol{Z_i b_i}$ is good enough to allow the distribution of the RLRT in model (1) to be closely approximated by the RLRT in the reduced model

$$\widetilde{\boldsymbol{Y}} = \boldsymbol{X\beta} + \boldsymbol{Z_s b_s} + \boldsymbol{\varepsilon}, \tag{3}$$

with $\widetilde{\boldsymbol{Y}} = \boldsymbol{Y} - \sum_{\boldsymbol{i \neq s}} \boldsymbol{Z_i b_i}$ assumed known. As model (3) has only one variance component σ_s^2, the exact null distribution of the RLRT for testing (2) is known (Crainiceanu and Ruppert, 2004) to be

$$RLRT_n \stackrel{d}{=} \sup_{\lambda \geq 0} \left\{ (n-p) \log \left[1 + \frac{N_n(\lambda)}{D_n(\lambda)} \right] - \sum_{l=1}^{K_s} \log(1 + \lambda \mu_{l,n}) \right\}, \tag{4}$$

where $\stackrel{d}{=}$ denotes equality in distribution, p is the number of columns in X,

$$N_n(\lambda) = \sum_{l=1}^{K_s} \frac{\lambda \mu_{l,n}}{1 + \lambda \mu_{l,n}} w_l^2, \quad D_n(\lambda) = \sum_{l=1}^{K_s} \frac{w_l^2}{1 + \lambda \mu_{l,n}} + \sum_{l=K_s+1}^{n-p} w_l^2, \tag{5}$$

$w_l, l = 1, \ldots, n-p$, are independent $N(0,1)$, and $\mu_{l,n}, l = 1, \ldots, K_s$, are the eigenvalues of the $K_s \times K_s$ matrix $\boldsymbol{Z_s'(I_n - X(X'X)^{-1}X')Z_s}$. This distribution can be simulated from very efficiently, as the K_s eigenvalues need to be computed only once, and speed depends only on K_s rather than on the number of observations n.

2.2 Mixture approximation to the Bootstrap

If a parametric bootstrap is preferred but computationally intensive, we propose the following parametric approximation to the RLRT distribution

$$RLRT \overset{d}{\approx} aUD, \tag{6}$$

where $U \sim \text{Bernoulli}(1-p)$, $D \sim \chi^2_1$, and $\overset{d}{\approx}$ denotes approximate equality in distribution. The flexible family of distributions in (6) contains as a particular case the i.i.d. case asymptotic $0.5\chi^2_0 : 0.5\chi^2_1$ distribution with $a = 1$ and $p = 0.5$, and is just as easy to use. p and a can be estimated from a bootstrap sample, while (6) stabilizes estimation of tail quantiles and thus reduces the necessary bootstrap sample size.
Maximum likelihood estimation of p would require the proportion of simulated RLRT values that are exactly zero, and is therefore very sensitive to numerical imprecisions (encountered, for example, with proc `MIXED` in SAS, and the `lme` function in R). We thus propose estimation of p and a using the method of moments, after setting all negative values to zero.

Note that both our proposed approximations are asymptotically identical to the $0.5\chi^2_0 : 0.5\chi^2_1$ approximation when the i.i.d. assumption holds.

3 Simulation Study

We conducted an extensive simulation study, covering a range of important situations with one or two variance components. An overview is given in Table 1. We varied the number of subjects $I = 6, 10$ and observations per subject $J = 5, 25, 50, 100$ for all settings, as well as the value for the respective nuisance variance component $\sigma^2_1 = 0, 0.1, 1, 10, 100$, while all other parameters not restricted to zero under the null were fixed at either 1 or -1. Covariates were sampled from standard normal distributions, with increasing correlation for the case of two smooth functions. Smooth uni- or bivariate functions were modeled using low-rank thin plate splines with smoothing parameters estimated by REML, and with testing of the corresponding variance component translating to testing for linearity $f(x) = \beta_0 + \beta_1 x$ in the univariate, and to testing for additivity and linearity $f(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ in the bivariate case. 10,000 samples each were simulated from the RLRT null distribution and our two approximations compared to a bootstrap and the $0.5\chi^2_0 : 0.5\chi^2_1$ approximation.

	Tested Component	**Null Hypothesis**	**Nuisance Component**
1	Random Intercept	Equality of Means	-
2	Smooth Function	Linearity	-
3	Random Intercept	Equality of Means	Smooth Function
4	Smooth Function	Linearity	Random Intercept
5	Smooth Function	Linearity	Smooth Function
6	Random Slope	Equality of Slopes	Random Intercept
7	Bivariate Smooth	Additivity and Linearity	Random Intercept
8	Random Intercept	Equality of Means	Bivariate Smooth

TABLE 1. Settings for the simulation study.

The fast finite sample approximation produced empirical type I error rates close to the nominal level, comparable to the exact distribution when $S = 1$. The approximation was usually good even for $n = 30$; the necessary sample size increased somewhat when random effects were correlated. The aUD approximation reduced the necessary bootstrap sample size by between 10% and 90%, with the reduction more pronounced for smaller α levels or p values. The $0.5\chi^2_0 : 0.5\chi^2_1$ approximation was always very conservative.

4 Testing smooth dose-response-functions

The Airgene study was conducted in six European cities between May 03 and July 04. One of its aims is to assess association between inflammatory responses and ambient air pollution concentrations in myocardial infarction survivors. 3 inflammatory blood markers (CRP, Fibrinogen, IL-6) were measured every month repeatedly up to 8 times in 1,003 patients. Air pollution and weather variables were measured concurrently in each city. Patients were genotyped and additional information collected at baseline. Analyses had to account for longitudinal data structure and potential nonlinearity of weather and trend variables, with smooth effects estimated in the mixed model framework. As the shape of the air pollution dose-response functions has important policy implications, one aim of the study was to investigate the functional form of the air pollution effects on inflammation. For illustration, we focus on the effect of PM_{10}, particulate matter with diameter less than 10 μm, on Fibrinogen in Barcelona. A total of 1074 valid blood samples and PM_{10} exposures were available for 183 patients.
The model used for the PM_{10}-Fibrinogen dose-response function is

$$FIB_{ij} = u_i + f(PM10_{ij}) + \sum_{l=2}^{L} \beta_l x_{ijl} + \varepsilon_{ij}, \;\; \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2_\varepsilon), \qquad (7)$$

where FIB_{ij} is the jth Fibrinogen value of the ith patient, $u_i \overset{iid}{\sim} N(0, \sigma^2_u)$ is a random patient intercept, and $PM10$ indicates the 5-day-average PM_{10}

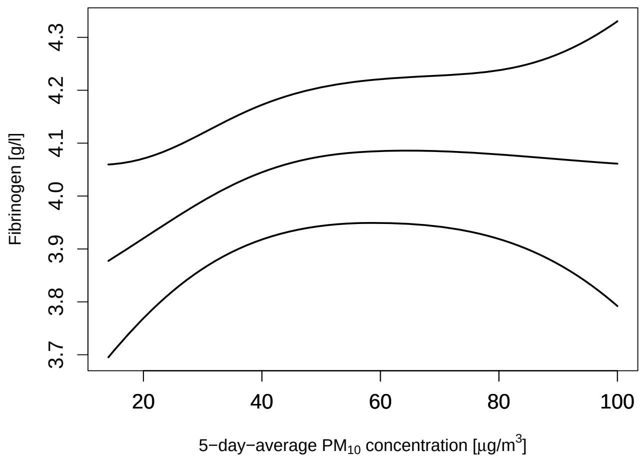

FIGURE 1. Estimated smooth PM_{10}-Fibrinogen dose-response function in Barcelona.

exposure before blood withdrawal. $f(.)$ is a smooth, unspecified, function estimated using penalized cubic B-Splines and penalizing deviations from linearity (Greven et al., 2006). Linear covariates x_l are patient's age, asthma diagnosis, time trend, weekday and air temperature (cubic polynomial).
Figure 1 shows the estimated smooth PM_{10} effect on Fibrinogen in Barcelona. An important scientific question is whether the dose-response function is linear. This is equivalent to testing (2) in (7), where σ_s^2 is a variance component controlling the smoothness of $f(.)$. Note that the i.i.d. assumption is violated and that the model includes two variance components.
The test statistic for testing linearity of $f(\cdot)$ against a general alternative takes the value $RLRT = 2.9$. Test results for all four approximations are given in Table 2. The fast finite sample approximation reduces computation time by 4 orders of magnitude, while results are similar to a bootstrap. The aUD approximation gives results close to the bootstrap even for sample sizes 100 times lower. The $0.5\chi_0^2 : 0.5\chi_1^2$ approximation is clearly conservative. In all cases, results indicate a significant difference from linearity.

5 Summary

We have discussed testing for zero variance components in linear mixed models. Possible applications include, but are not limited to, testing for zero random intercepts or slopes and testing for linearity of a smooth function against a general alternative. For models with one variance component, we recommend directly using the exact null distribution of the RLRT

Approximation	Samples	Time	p-value
Fast finite sample	100,000	~35sec	0.023
aUD	100	~4min	0.026
aUD	1,000	~40min	0.031
aUD	10,000	~8h	0.029
$0.5\chi_0^2 : 0.5\chi_1^2$	-	-	0.044
Bootstrap	10,000	~8h	0.025

TABLE 2. Testing the PM_{10} effect on Fibrinogen in Barcelona for linearity. Computation time was measured on a standard PC using Matlab (f.f.s.) / SAS.

statistic derived in Crainiceanu and Ruppert (2004), which can be simulated efficiently. For models with more than one variance component, we have proposed two approximations to the finite sample null distribution of the RLRT. Extensive simulations showed superiority of both approximations over the $0.5\chi_0^2 : 0.5\chi_1^2$ approximation and parametric bootstrap currently used. Our results extend existing methodology to linear mixed models with more than one variance component and lacking independence assumption. We have illustrated the use in testing for linearity of dose-response-functions for longitudinal data on air pollution health effects.

Acknowledgments: The first author was supported by the German Academic Exchange Service (DAAD). We would like to thank Yu-Jen Cheng and Fabian Scheipl for discussions related to the topic of this paper. Special thanks to the Airgene study group. The Airgene study was funded by EU contract QLK4-CT-2002-O2236.

References

Crainiceanu, C., and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B*, **66**(1), 165-185.

Gong, G. and Samaniego, F. (1981). Pseudo Maximum Likelihood Estimation: Theory and Applications. *The Annals of Statistics*, **9**(4), 861-869.

Greven, S., Küchenhoff, H., and Peters, A. (2006). Additive mixed models with P-Splines. *Proceedings of the 21st International Workshop on Statistical Modelling*, 201-207. Eds.: Hinde J, Einbeck J and Newell J.

Liang, K.-Y. and Self, S.G. (1996). On the Asymptotic Behaviour of the Pseudolikelihood Ratio Test Statistic. *Journal of the Royal Statistical Society: Series B*, **58**(4), 785-796.

Self, S.G. and Liang, K.-Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association*, **82**(398), 605-610.

Stram, D.O. and Lee, J.-W. (1994). Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics*, **50**(3), 1171-1177.