BIOMETRICS 54, 1560–1568
December 1998

Likelihood Ratio Testing of Variance Components in the Linear Mixed-Effects Model Using Restricted Maximum Likelihood

Christopher H. Morrell

Mathematical Sciences Department, Loyola College, 4501 North Charles Street, Baltimore, Maryland 21210-2699, U.S.A.

SUMMARY

This paper reports the results of an extensive Monte Carlo study of the distribution of the likelihood ratio test statistic using the value of the restricted likelihood for testing random components in the linear mixed-effects model when the number of fixed components remains constant. The distribution of this test statistic is considered when one additional random component is added. The distribution of the likelihood ratio test statistic computed using restricted maximum likelihood is compared to the likelihood ratio test statistic computed from the usual maximum likelihood. The rejection proportion is computed under the null hypothesis using a mixture of chi-square distributions. The restricted likelihood ratio statistic has a reasonable agreement with the maximum likelihood test statistic. For the parameter combinations considered, the rejection proportions are, in most cases, less than the nominal 5% level for both test statistics, though, on average, the rejection proportions for REML are closer to the nominal level than for ML.

1. Introduction

In recent years, the application of the linear mixed-effects (LME) model (Bryk and Raudenbush, 1992; Laird and Ware, 1982; Longford, 1993) to repeated measures data from longitudinal studies has become more frequently used in biomedical, economic, educational, pharmacological, psychological, and sociological studies due to the increasing availability of software that can be used to fit this model. The LME model allows for repeated measurements on each cluster or object under study. Fixed effects are used to model the population average relationship between the response variable and a number of explanatory variables. Heterogeneity among the clusters is accounted for by the inclusion of random effects in the model that permit the associations between the explanatory variables and response variable to vary from cluster to cluster and to deviate from the overall population average association.

The variance components in the LME model may be estimated by maximum likelihood (ML) or restricted maximum likelihood (REML). Maximum likelihood estimates do not take into account the estimation of fixed effects and so are biased downward. Restricted maximum likelihood estimation corrects for the presence of these nuisance parameters by maximizing the likelihood of linearly independent error contrasts to obtain more unbiased estimates (Laird and Ware, 1982; Lindstrom and Bates, 1988).

The random coefficient regression model (Carter and Yang, 1986; Swamy, 1971) is a special case of the linear mixed-effects model where each fixed effect included in the model has a corresponding random effect. A number of authors have considered testing coefficients in this model (Carter and Yang, 1986; Leeper and Chang, 1987; Leeper and Chang, 1992; Morgan, Pantula, and Gumpertz, 1995; Swamy, 1971). Carter and Yang (1986), Leeper and Chang (1987), and Swamy (1971) develop tests of inference about the mean of the random components. This is analogous to tests of the fixed

Email address: chm@Loyola.edu

Key words: Likelihood ratio test; Linear mixed-effects model; Maximum likelihood; Restricted maximum likelihood.

effects parameters in the LME model. Morgan et al. (1995) develop a fixed parameter approach to testing whether the random effects are all zero. This corresponds to the variance component being zero. They compare their method to the likelihood ratio test of the same hypothesis and find that the power of their approach is similar to the power of the likelihood ratio test for a range of values of the variance components.

Öfversten (1993) uses an orthogonal transformation to develop exact tests for variance components in unbalanced linear mixed-effects models. However, the tests are restricted to the situation where the random effects are uncorrelated with each other so that the covariance matrix of the random effects is diagonal. It is suggested that this method may be appropriate for the analysis of designed experiments.

The likelihood ratio (LR) test statistic (Cox and Hinkley, 1974) is an asymptotic test statistic used to test statistical hypotheses in many situations when the null hypothesis involves a restriction on the values of some of the parameters. The parameters of the model are estimated by maximum likelihood under both the null and alternative models. Under a number of regularity conditions, the test statistic $-2 \times \log(\text{likelihood ratio})$ has a χ^2 distribution with degrees of freedom given by the difference in the number of parameters between the null and alternative hypotheses.

Stram and Lee (1994) show that the likelihood ratio test of variance components in the LME model does not satisfy the usual regularity conditions of the likelihood ratio test. They apply the results of Self and Liang (1987) to determine the correct asymptotic distribution of $-2 \times \log(\text{likelihood ratio})$ for the LME model when the number of fixed effects remains constant. In particular, the asymptotic distribution of this test statistic for removing one random component is a 50:50 mixture of χ^2 distributions. Stram and Lee (1994) point out that, if a single χ^2 distribution is employed, the degree of bias is typically small.

Pierce and Peters (1992) explore the use of higher order asymptotics to provide better approximations to the likelihood to construct confidence intervals and perform tests of one parameter. They discuss an adjusted profile log-likelihood that corrects the likelihood for nuisance parameters. This is the same correction made by REML in the LME model. They also present a test statistic that corrects for nuisance parameters and provides an information adjustment. For discrete distributions and in the single parameter case, this test statistic provides confidence intervals and tail probabilities that are better than those based on the uncorrected log-likelihood in that they more closely approximate the exact results.

The likelihood ratio test uses the likelihood evaluated at the maximum likelihood estimated under both the null and alternative models to compute the test statistic. However, the distribution of the likelihood ratio test statistic for random components from the LME model when the restricted likelihood is used has not been investigated.

In this paper, results of an extensive Monte Carlo study that explores the distribution of this test statistic are presented. The simulation investigates varying numbers of clusters, average number of observations per cluster, differing numbers of fixed and random effects, various values of the ratio of the between-cluster variation to within-cluster variation, and various correlation structures for the covariance matrix of the random effects. For each sample generated, both ML and REML are used to fit the null and alternative models. The likelihood ratio test statistic is computed from the ML log-likelihood when ML estimation is used and from the REML log-likelihood when REML estimation is used.

2. The Linear Mixed-Effects Model

The linear mixed-effects model has the form (Laird and Ware, 1982; Lindstrom and Bates, 1988)

$$y_i = X_i\beta + Z_i b_i + e_i, \qquad i = 1, \dots, m, \tag{1}$$

where y_i is the $n_i \times 1$ vector of observations for cluster i, X_i is an $n_i \times p$ design matrix of independent variables for the fixed effects, β is a $p \times 1$ vector of fixed effects parameters, Z_i is an $n_i \times q$ design matrix of independent variables for the random effects, the b_i are independent $q \times 1$ vectors of random effects with $\mathrm{N}(0, D)$ distribution, and the e_i are independent $n_i \times 1$ vectors of random errors with $\mathrm{N}(0, \sigma^2 I)$ distributions. The b_i are independent of the e_i. The total number of observations is $N = \Sigma\, n_i$.

If estimates of σ^2 and the random effects covariance matrix D are available, then the estimator of the fixed-effects parameter vector β is the generalized least squares estimator (Laird and Ware, 1982)

$$\hat{\beta} = \left(\sum_{i=1}^{m} X_i^{\mathrm{T}} V_i^{-1} X_i\right)^{-1} \sum_{i=1}^{m} X_i^{\mathrm{T}} V_i^{-1} y_i, \tag{2}$$

where $V_i = \sigma^2 I + Z_i D Z_i^{\mathrm{T}}, i = 1, \ldots, m$. The covariance matrix of these estimates is given by

$$\mathrm{cov}(\hat{\beta}) = \left(\sum_{i=1}^{m} X_i^{\mathrm{T}} V_i^{-1} X_i\right)^{-1}.$$

The variance components σ^2 and D are estimated either using ML or REML. The marginal log-likelihood for computing maximum likelihood estimates is given by (Laird, Lange, and Stram, 1987) as

$$L_{\mathrm{ML}}(\beta, \theta) = -\sum_{i=1}^{m} \left(\ln|V_i| - (y_i - X_i\beta)^{\mathrm{T}} V_i^{-1} (y_i - X_i\beta)\right)/2, \tag{3}$$

where the vector θ contains the unique elements of σ^2 and D. To compute REML estimates of the variance components, the log-likelihood becomes

$$L_{\mathrm{REML}}(\beta, \theta) = -\ln\left(\left|\sum_{i=1}^{m} X_i^{\mathrm{T}} V_i^{-1} X_i\right|\right) \Big/ 2 + L_{\mathrm{ML}}(\beta, \theta). \tag{4}$$

Note that (4) has just one additional term from (3) and the log-likelihood can be written as

$$L_{\mathrm{REML}}(\beta, \theta) = \ln\left(|\mathrm{cov}(\hat{\beta})|\right)/2 + L_{\mathrm{ML}}(\beta, \theta).$$

The form of this log-likelihood is the same as the adjusted log-likelihood in Pierce and Peters (1992). If the fixed effects β in (3) and (4) are replaced by their generalized least squares estimates in (2), both the log likelihoods are free of the fixed effects and are functions of the random effects alone. However, (4) contains an additional term as compared to (3). Consequently, this paper investigates the validity of using (4) in tests of random effects in the LME model when the number of fixed effects remains constant.

In the simulations, the LME models are fit using a FORTRAN routine developed by Mary Lindstrom (see Lindstrom and Bates, 1988). This program was the precursor to the SPlus linear mixed-effects function, lme (Pinheiro and Bates, 1995). It implements the Newton–Raphson algorithm to estimate the random effects in the model using either maximum likelihood or restricted maximum likelihood and applies the orthogonality convergence criterion (Bates and Watts, 1981) to determine convergence. The estimation procedure ensures that the estimate of D remains positive definite and reaches convergence in cases where the SAS Proc Mixed procedure is unable to converge. This algorithm "appears to be very effective at finding the maximum likelihood value of a positive semidefinite D lying on the constraint surface" (Stram and Lee, 1994, p. 1176). Under the alternative model, a positive-definite starting estimate of D is provided by appending a small value to the diagonal of the true D under the null model. The program then iterates until convergence is reached. Lindstrom and Bates (1988) give an example where unnecessary additional random effects are included and convergence is still obtained. Even though they expected the likelihood surface to be quite flat, the Newton–Raphson method converged quickly. They point out that "it is often necessary to overfit a data set before settling on a reduced model" (Lindstrom and Bates, 1988, p. 1020). Their algorithm performs this task well.

3. The Test Statistics

For a given set of fixed effects in the model, we wish to test the null hypothesis that there are q random effects in the model versus the alternative that there are $q + 1$ random effects, i.e., the fixed-effects design matrices X_i have the same set of columns under both the null and alternative hypotheses while the random effects design matrices Z_i contain one additional column under the alternative hypothesis. Under the null hypothesis, the random effects covariance matrix is a $q \times q$ positive-definite matrix. Under the alternative hypothesis, the random effects covariance matrix is a $(q + 1) \times (q + 1)$ positive-semidefinite matrix (Stram and Lee, 1994).

Let L_{ML0} be the log-likelihood from the maximum likelihood estimation computed under the null model and L_{ML1} be the log-likelihood from the maximum likelihood estimation computed under the alternative model. Then the ML log-likelihood ratio test statistic is

$$LR_{\mathrm{ML}} = 2(L_{\mathrm{ML1}} - L_{\mathrm{ML0}}).$$

Similarly, using the REML log-likelihood, the test statistic is

$$LR_{\mathrm{REML}} = 2(L_{\mathrm{REML1}} - L_{\mathrm{REML0}}).$$

The distributions of these test statistics will be compared to each other under a variety of conditions. For data generated from the null model, the false rejection rate is computed using critical values from the χ^2_{q+1} distribution as well as for critical values from the 50:50 mixture of χ^2_{q+1} and χ^2_q distributions (Stram and Lee, 1994).

4. Null and Alternative Models for Simulation Study

For each set of parameters for the null model, three designs will be simulated. All three designs will have the same mean number of observations per cluster. The first will have $n_i = n$ observations for each cluster with the observations at the same design points for each cluster (a balanced uniform design). The second will be a symmetric triangular design. Some clusters will have more observations than the average while others will have fewer observations than the average. These cluster sample sizes decrease symmetrically away from the mean cluster sample size. The final design will be a decreasing design. There are many clusters with few observations per cluster decreasing to a few clusters with many observations per cluster.

The first design reflects the situation where there is a designed experiment with no attrition. The second design has some subjects who do not have the required number of observations. An equal number of subjects is assigned additional observations to maintain the correct mean number of observations per subject. The last design occurs commonly in prospective longitudinal studies (e.g., the Baltimore Longitudinal Study of Aging; Shock et al., 1984). There are many subjects with few repeated observations decreasing to few subjects with many repeated observations.

Table 1 gives the combinations of the numbers of fixed and random effects in the model, average number of observations per cluster, and the number of Monte Carlo replications. Each of these sets of parameters is run with m = 20, 100, and 500 clusters for each of the three designs. For the first five parameter combinations, 1000 iterations were performed. However, since the computing time increases dramatically with the number of random effects and the sample sizes, for the last three parameter combinations, only 500 replications were performed.

The fixed effects for models with $p = 2$ correspond to an intercept term and a longitudinal follow-up time term. When $p = 10$, additional terms involving time^2, age, and age^2, as well as cross-product interaction terms, are included. There are also two dummy variables in the model, one varying within clusters (alternating zeros and ones) while the other is fixed within clusters but varies between clusters (75% zeros and 25% ones).

Table 2 gives the choice of variance components for the simulations for the three values of q under the null hypothesis. For all simulations, the error variance $\sigma^2 = 10$. For models with one random effect, the variance of the random effect D_{11} is taken to be equal to σ^2, $0.1\sigma^2$, and $10\sigma^2$.

Table 1

Number of fixed effects (p), random effects under the null hypothesis (q), average number of observations per cluster (n), and number of Monte Carlo replications ($nrep$)

p	2	2	10	10	10	10	10	10
q	1	1	1	1	2	2	3	3
n	2	5	4	10	4	10	4	10
$nrep$	1000	1000	1000	1000	1000	500	500	500

Table 2

Variance component combinations for the three choices of q, the number of random effects in the null model. When q = 3, the correlations are ρ_{12}, ρ_{13}, and ρ_{23}.

q	1	1	1	2	2	2	3	3	3	3	3
σ^2	10	10	10	10	10	10	10	10	10	10	10
diag(D)	10	1	100	10, 1	10, 1	10, 1	10, 1, 4	10, 1, 4	10, 1, 4	10, 1, 4	10, 1, 4
Correlations				0	.5	−.5	0, 0, 0	.5, 0, 0	−.5, 0, 0	.5, .5, .5	−.5, −.5, .5

Table 3

Rejection proportion of the likelihood ratio statistic based on maximum likelihood (ML) and restricted maximum likelihood (REML) at the 5% level based on a 50:50 mixture of χ^2 distributions. $q = 1$; 1000 Monte Carlo replications. $\sigma^2 = 10$.

				Balanced		Triangular		Decreasing	
p	n	D	m	ML	REML	ML	REML	ML	REML
2	2	1	20	.040	.035	.046	.045	.048	.054
			100	.026	.026	.032	.032	.036	.039
			500	.022	.022	.042	.041	.041	.044
		10	20	.032	.027	.045	.047	.042	.049
			100	.034	.033	.043	.042	.046	.044
			500	.016	.016	.058	.058	.046	.046
		100	20	.034	.027	.052	.053	.055	.058
			100	.028	.027	.046	.047	.052	.058
			500	.015	.015	.054	.055	.052	.054
2	5	1	20	.029	.038	.018	.021	.016	.026
			100	.036	.038	.040	.046	.026	.027
			500	.043	.043	.046	.047	.044	.050
		10	20	.029	.031	.034	.034	.031	.037
			100	.047	.049	.042	.049	.024	.027
			500	.043	.045	.053	.055	.035	.038
		100	20	.034	.038	.041	.046	.034	.037
			100	.055	.058	.040	.041	.031	.036
			500	.043	.044	.051	.053	.036	.037
10	4	1	20	.023	.020	.027	.033	.027	.039
			100	.021	.023	.032	.037	.043	.047
			500	.041	.044	.039	.041	.051	.056
		10	20	.045	.030	.041	.038	.028	.026
			100	.032	.033	.036	.033	.043	.045
			500	.047	.050	.046	.050	.045	.052
		100	20	.052	.040	.049	.045	.040	.033
			100	.035	.034	.039	.040	.042	.047
			500	.039	.041	.043	.044	.050	.054
10	10	1	20	.025	.039	.020	.031	.010	.022
			100	.029	.034	.027	.033	.039	.046
			500	.037	.040	.033	.037	.050	.053
		10	20	.037	.045	.036	.037	.030	.036
			100	.031	.037	.043	.048	.034	.048
			500	.039	.045	.029	.038	.045	.050
		100	20	.045	.047	.032	.040	.033	.031
			100	.034	.038	.040	.046	.043	.056
			500	.042	.048	.030	.036	.042	.048

Since there is no marked difference in the results for these three choices, for models with more than one random effect, the variance of the intercept random effect (a column of ones in Z_i) D_{11} is chosen to be equal to σ^2 and various correlation structures are selected. For two random effects, the variance of the second random effect D_{22} is 1 (for the slope of longitudinal change) and the correlation between the two random effects is 0 (independent random effects), .5, and $-.5$. When the null model contains three random effects, the diagonal elements of D (corresponding to intercept, slope for longitudinal change, and a dummy variable with alternating zeros and ones) are 10, 1, and 4. Five correlation structures, independent effects, and correlations ρ_{12}, ρ_{13}, and ρ_{23} of (.5, 0, 0), ($-.5$, 0, 0), (.5, .5, .5), and ($-.5$, $-.5$, .5) are used.

Initially, two fixed effects and one random effect are investigated. Since REML corrects the likelihood for the number of fixed effects in the model, the remaining models consider a larger number of fixed effects in the models ($p = 10$). A total of 84 combinations of input parameters is used. Within each combination, all three designs are investigated.

For models with one random effect (intercept) under the null hypothesis, the alternative model contains an additional random effect for slope of longitudinal change (time). For two random effects under the null hypothesis, the Z_i matrix for the null model contains random intercept and follow-up time (for slope of longitudinal change) terms, and the alternative model appends (follow-up time)^2 to allow for quadratic growth curves. Finally, for three random effects under the null hypothesis (intercept, follow-up time, and a dummy variable alternating between zero and one within the cluster), (follow-up time)^2 is again added under the alternative model.

The data generation proceeds as follows. The sample sizes, design points, and number of repeated observations are input. The design matrices for the fixed and random effects are set up (columns of X_i and Z_i). The mean model is computed for each design point for each cluster by multiplying the jth column of X_i by the corresponding fixed effect (β_j). The random effects b_i are generated from an $\mathrm{N}_q(\mathbf{0}, D)$ distribution for each cluster. These multivariate normal data are generated using the marginal distribution of the first random effect and then the conditional distributions of subsequent random effects given the previously generated random effects. The normal deviates are generated using the NAG routine G05FDE (NAG, 1991). These random effects are multiplied by the appropriate column of Z_i and added to the mean model to obtain the cluster-specific data. Finally, the random errors e_{ij} are generated from the $\mathrm{N}(0, \sigma^2)$ distribution and added to obtain the final data set.

5. Results

Tables 3, 4, and 5 give the rejection percents at the 5% significance level by comparing the test statistic to the critical value from a 50:50 mixture of χ^2_q and χ^2_{q+1} distributions. The rejection proportions for the χ^2_{q+1} distribution are not presented as these rejection proportions will be even smaller than the rejection proportions based on the mixture of χ^2 distributions. The rejection proportions tend to be less than the nominal 5% significance level for the combinations of parameters in this simulation.

For two fixed effects (and one random effect), the simulations show that the rejection proportions for REML are the closest to .05 for the triangular design while the balanced design has rejection proportions that are the furthest from the nominal 5% level. In addition, for 500 clusters, the rejection proportions are also closest to 5%.

For 10 fixed effects, the balanced design has the rejection proportions that are the closest to 5% while the decreasing design is the furthest from 5%. For $n = 4$, the rejection proportions are closer

Table 4

Rejection proportion of the likelihood ratio statistic based on maximum likelihood (ML) and restricted maximum likelihood (REML) at the 5% *level based on a* 50:50 *mixture of* χ^2 *distributions.* $q = 2$; 1000 *Monte Carlo replications for* $n = 4$; 500 *Monte Carlo replications for* $n = 10$. $\sigma^2 = 10$, $\mathrm{diag}(D) = (10, 1)$.

				Balanced		Triangular		Decreasing	
p	n	Correlation	m	ML	REML	ML	REML	ML	REML
10	4	0	20	.075	.055	.056	.040	.045	.034
			100	.048	.046	.042	.042	.050	.044
			500	.051	.050	.049	.048	.049	.049
		.5	20	.071	.053	.049	.042	.032	.025
			100	.050	.049	.040	.040	.043	.041
			500	.047	.047	.045	.047	.048	.048
		−.5	20	.074	.054	.060	.043	.046	.032
			100	.052	.048	.043	.041	.052	.051
			500	.054	.053	.047	.047	.052	.054
10	10	0	20	.036	.028	.040	.034	.028	.030
			100	.030	.028	.034	.028	.040	.038
			500	.044	.044	.046	.044	.042	.042
		.5	20	.032	.030	.032	.030	.036	.028
			100	.026	.028	.032	.030	.036	.038
			500	.032	.032	.044	.046	.042	.046
		−.5	20	.038	.030	.050	.036	.028	.030
			100	.034	.032	.036	.038	.042	.036
			500	.042	.046	.052	.052	.044	.048

Table 5

Rejection proportion of the likelihood ratio statistic based on maximum likelihood (ML) and restricted maximum likelihood (REML) at the 5% level based on a 50:50 mixture of χ^2 distributions. $q = 3$; 500 Monte Carlo replications. $\sigma^2 = 10$, $\mathrm{diag}(D) = (10, 1, 4)$.

				Balanced		Triangular		Decreasing	
p	n	Correlation	m	ML	REML	ML	REML	ML	REML
10	4	0	20	.118	.058	.062	.050	.036	.028
		0	100	.042	.038	.034	.032	.018	.022
		0	500	.034	.034	.042	.044	.060	.060
		.5	20	.128	.062	.072	.060	.044	.028
		0	100	.038	.034	.030	.032	.022	.020
		0	500	.040	.038	.040	.040	.052	.048
		−.5	20	.126	.072	.064	.044	.034	.026
		0	100	.048	.042	.034	.034	.028	.030
		0	500	.028	.028	.036	.038	.056	.058
		.5	20	.178	.092	.078	.056	.032	.020
		.5	100	.058	.050	.038	.036	.022	.026
		.5	500	.036	.034	.044	.042	.040	.040
		−.5	20	.162	.092	.054	.044	.020	.016
		−.5	100	.056	.046	.030	.032	.032	.038
		.5	500	.042	.044	.048	.050	.044	.044
10	10	0	20	.028	.024	.030	.030	.010	.008
		0	100	.050	.052	.030	.032	.034	.036
		0	500	.058	.058	.038	.040	.042	.044
		.5	20	.028	.024	.036	.040	.010	.014
		0	100	.048	.048	.030	.032	.034	.036
		0	500	.056	.058	.046	.048	.044	.044
		−.5	20	.030	.026	.026	.024	.010	.008
		0	100	.050	.052	.040	.038	.040	.038
		0	500	.046	.046	.040	.042	.044	.042
		.5	20	.028	.028	.034	.034	.010	.012
		.5	100	.042	.048	.024	.032	.022	.024
		.5	500	.052	.052	.040	.040	.050	.052
		−.5	20	.022	.022	.032	.034	.018	.018
		−.5	100	.034	.038	.034	.034	.032	.032
		.5	500	.046	.046	.048	.050	.038	.038

to 5%. The rejection proportions do not differ for the REML test statistics for the three levels of q. However, the rejection proportions for ML are closer to 5% for $q = 2$ and $q = 3$ than for $q = 1$.

Comparing how close the REML vs. ML proportions are to the nominal 5% level, the REML method performed worse than the ML method on only 34.1% of the tests. The improvement in the REML proportions over ML is greatest for small numbers of clusters ($m = 20$), and the improvement declines with increasing m. This is to be expected as, asymptotically, the tests will be identical. With respect to the number of random effects in the model, the improvement of REML over ML is largest for $q = 3$. The improvement is largest for the balanced design and smallest for the decreasing design where the REML and ML proportions show close agreement. Finally, the further the ML statistic deviates from the nominal level, the greater the improvement in the rejection proportions provided by REML.

Figure 1 presents the estimated probability density function for $m = 500$ clusters, $p = 10$ fixed effects, and $n = 10$ observations per cluster for the balanced design for 1, 2, and 3 random effects. For one random effect, $D_{11} = 10$, and for $q = 2$ and 3, the random effects under the null model are assumed to be independent. The graphs show that the estimated distributions for the statistics are almost identical and lie between the χ^2_q and χ^2_{q+1} curves.

6. Conclusions

The simulations have shown that, under the null hypothesis, the distributions for the REML likelihood ratio test statistics are similar to the distribution of the ML likelihood ratio test statistic for one extra random effect in the model. In fact, the REML test statistic performs slightly better

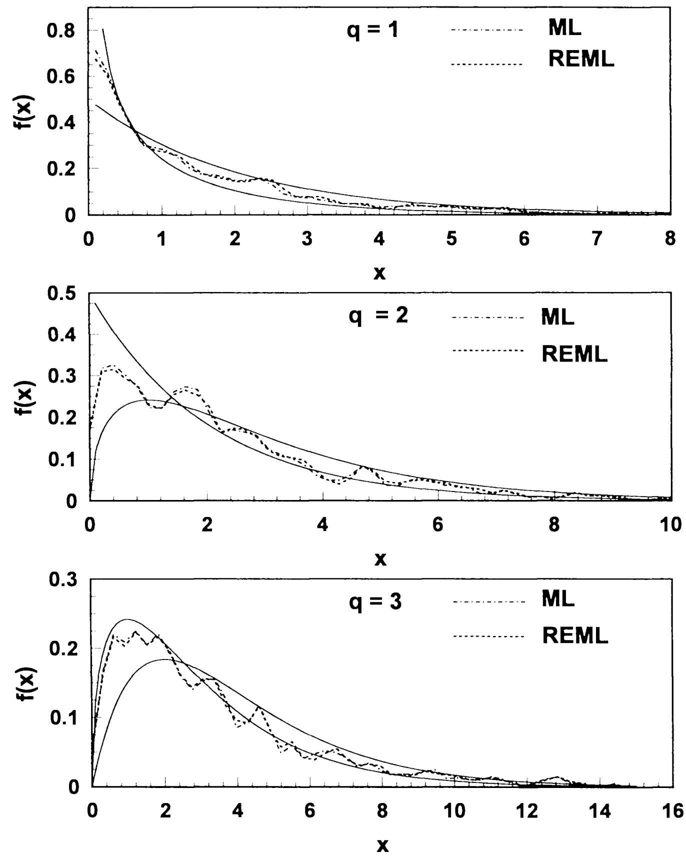

Figure 1. Density estimates of the ML and REML likelihood ratio test statistics. The solid lines are for the χ^2_q and χ^2_{q+1} distributions. $m = 500$ clusters, $n = 10$ observations per cluster, $p = 10$ fixed effects for the balanced design. For $q = 1$, $D_{11} = 10$, and for $q = 2$ and 3, the correlations are zero. The density estimates for $q = 1$ are based on 1000 replications, while those for $q = 2$ and 3 are for 500 replications.

than the ML test statistic in that, on average, the rejection proportions are closer to the nominal level for the REML test statistic than for the ML test statistic. Thus, if one chooses to estimate the variance components in an LME model using restricted maximum likelihood, the likelihood ratio test may be applied using the value of the restricted likelihood to obtain valid inferences about the number of variance components in the model. In most analyses, the researcher should carefully consider the number of clusters, observations per cluster, and the degree to which clusters vary from each other in choosing a reasonable beginning number of random effects. Then the appropriate number of random effects may be chosen by a backward elimination of random components one at a time (Morrell, Pearson, and Brant, 1997).

Acknowledgements

The work of Christopher Morrell was supported by research grant 1 R03 DC 02566-01 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. The simulations were performed while on sabbatical in the Statistical Sciences Department at the University of Cape Town, South Africa. The author thanks the associate editor and two anonymous referees for comments that improved the presentation.

Résumé

Cet article présente les résultats d'une étude étendue de Monte Carlo de la distribution de la statistique du test du rapport de vraisemblance utilisant la valeur de la vraisemblance restreinte pour effectuer le test des composantes aléatoires dans un modèle linéaire à effets mixtes, quand le nombre de composantes à effets fixes est constant. Nous considérons la distribution de cette statistique de test lorsque l'on ajoute une composante aléatoire. La distribution de la statistique du test du rapport de vraisemblance employant le maximum de vraisemblance restreint est comparée à celle calculée avec le maximum de vraisemblance habituel. La proportion de rejet est calculée, sous l'hypothèse nulle, en utilisant un mélange de distributions du χ^2. Il existe une concordance raisonnable entre les deux tests. Pour les combinaisons de paramètres considérées, les proportions de rejet sont, dans la majorité des cas, inférieure au niveau nominal de 5% pour les deux statistiques bien que, en moyenne, la proportion de rejet pour la première (REML) soit plus proche du niveau nominal de la seconde (ML).

REFERENCES

Bates, D. M. and Watts, D. G. (1981). A relative offset orthogonality convergence criterion for nonlinear least squares. *Technometrics* **23,** 179–183.

Bryk, A. S. and Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods.* London: Sage Publications.

Carter, R. L. and Yang, M. C. K. (1986). Large sample inference in random coefficient regression models. *Communications in Statistics—Theory and Methods* **15,** 2507–2535.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* London: Chapman Hall.

Laird, N., Lange, N., and Stram, D. (1987). Maximum likelihood computations with repeated measures: Application of the EM algorithm. *Journal of the American Statistical Association* **82,** 97–105.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38,** 963–974.

Leeper, J. D. and Chang, H. L. (1987). Hypothesis testing for two-group incomplete-data growth curve problems. *Communications in Statistics—Theory and Methods* **16,** 3715–3728.

Leeper, J. D. and Chang, S. W. (1992). Comparison of general multivariate and random-effects models for growth curve analysis: Incomplete-data small-sample situations. *Journal of Statistical Computation and Simulation* **44,** 93–104.

Lindstrom, M. J. and Bates, D. M. (1988). Newton–Raphson and EM algorithms for linear mixed effects models for repeated measures data. *Journal of the American Statistical Association* **83,** 1014–1022.

Longford, N. T. (1993). *Random Coefficient Models.* Oxford: Clarendon Press.

Morgan, B., Pantula, S., and Gumpertz, M. (1995). Testing whether coefficients are random in a linear random coefficient regression model. *Proceedings of the Biometrics Section of the American Statistical Association* 15–24.

Morrell, C. H., Pearson, J. D., and Brant, L. J. (1997). Linear transformations of linear mixed-effects models. *The American Statistician* **51,** 338–343.

NAG. (1991). *Fortran Library Manual.* Oxford: Numerical Algorithms Group.

Öfversten, J. (1993). Exact tests for variance components in unbalanced mixed linear models. *Biometrics* **49,** 45–57.

Pierce, D. A. and Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families. *Journal of the Royal Statistical Society, Series B* **54,** 701–737.

Pinheiro, J. C. and Bates, D. M. (1995). *Mixed effects models and classes for S and SPlus.* Technical Report 89, Department of Biostatistics, University of Wisconsin, Madison.

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,** 605–610.

Shock, N. W., Greulich, R. C., Andres, R., Lakatta, E. G., Arenberg, D., and Tobin, J. D. (1984). *Normal human aging: The Baltimore Longitudinal Study of Aging.* NIH Publication 84-2450, U.S. Government Printing Office, Washington, D.C.

Stram, D. O. and Lee, J. W. (1994). Variance component testing in the longitudinal mixed effects model. *Biometrics* **50,** 1171–1177.

Swamy, P. A. V. B. (1971). *Statistical Inference in Random Coefficient Regression Models.* Berlin: Springer Verlag.

Received November 1996; *revised October* 1997; *accepted November* 1997.