RLRsim: Testing for Random Effects or Nonparametric Regression Functions in Additive Mixed Models

Fabian Scheipl ¹
joint work with Sonja Greven ¹,² and Helmut Küchenhoff ¹

¹Department of Statistics, LMU München, Germany
²Department of Biostatistics, Johns Hopkins University, USA

useR! 2008
August 13, 2008
Outline

Background & Problem Description

Implementation & Application Examples

Simulation Study
Linear Mixed Models

\[
y = X\beta + \sum_{l=1}^{L} Z_l b_l + \varepsilon
\]

\[
b_l \sim \mathcal{N}_{K_l}(0, \lambda_l \sigma^2_{\varepsilon} \Sigma_l), \quad b_l \perp b_s \quad \forall l \neq s
\]

\[
\varepsilon \sim \mathcal{N}_n(0, \sigma^2_{\varepsilon} I_n),
\]

We want to test \( H_0, l \): \( \lambda_l = 0 \) versus \( H_A, l \): \( \lambda_l > 0 \)

\( \iff H_0, l \): \( \text{Var}(b_l) = 0 \) versus \( H_A, l \): \( \text{Var}(b_l) > 0 \)

Application examples:

- testing for equality of means between groups/subjects
- testing for linearity of a smooth function
Linear Mixed Models

\[ y = X\beta + \sum_{l=1}^{L} Z_l b_l + \varepsilon \]

\[ b_l \sim \mathcal{N}_{K_l}(0, \lambda_l \sigma^2_{\varepsilon} \Sigma_l), \quad b_l \perp b_s \quad \forall l \neq s \]

\[ \varepsilon \sim \mathcal{N}_n(0, \sigma^2_{\varepsilon} I_n), \]

We want to test

\[ H_{0,l} : \lambda_l = 0 \text{ versus } H_{A,l} : \lambda_l > 0 \]

\[ \Leftrightarrow H_{0,l} : \text{Var}(b_l) = 0 \text{ versus } H_{A,l} : \text{Var}(b_l) > 0 \]

Application examples:

- testing for equality of means between groups/subjects
- testing for linearity of a smooth function
Additive Models as Linear Mixed Models

Simple additive model:

\[ y = f(x) + \varepsilon \]

\[ f(x_i) \approx \sum_{j=1}^{J} \delta_j B_j(x_i) \]

- fit via PLS: \( \min_\delta (\|y - B\delta\|^2 + \frac{1}{\lambda} \delta'P\delta) \)
- reparametrize s.t. PLS-estimation is equivalent to (RE)ML-estimation

\( (RE)ML \)
Additive Models as Linear Mixed Models

Simple additive model:

\[
\begin{align*}
    y &= f(x) + \varepsilon \\
    f(x_i) &\approx \sum_{j=1}^{J} \delta_j B_j(x_i)
\end{align*}
\]

- fit via PLS: \( \min_{\delta} \|y - B\delta\|^2 + \frac{1}{\lambda} \delta'P\delta \)
- reparametrize s.t. PLS-estimation is equivalent to (RE)ML-estimation given \( \lambda \) in a LMM with
  - fixed effects for the unpenalized part of \( f(x) \)
  - random effects (\( \text{i.i.d. } \mathcal{N}(0, \lambda\sigma^2_\epsilon) \)) for the deviations from the unpenalized part

(Brumback, Ruppert, Wand, 1999; Fahrmeir, Kneib, Lang, 2004)

In R: mgcv::gamm(), lmeSplines
Problem:
Likelihood Ratio Tests for Zero Variance Components

General Case:
- $y_1, \ldots, y_n \overset{i.i.d.}{\sim} f(y|\theta); \ \theta = (\theta_1, \ldots, \theta_p)$
- Test: $H_0: \theta_i = \theta_0^i$ versus $H_A: \theta_i \neq \theta_0^i$
- $LRT = 2 \log L(\hat{\theta}|y) - 2 \log L(\hat{\theta}^0|y) \xrightarrow{n \to \infty} \chi_1^2$
Problem:
Likelihood Ratio Tests for Zero Variance Components

General Case:

- $y_1, \ldots, y_n \overset{i.i.d.}{\sim} f(y|\theta)$; $\theta = (\theta_1, \ldots, \theta_p)$
- Test: $H_0 : \theta_i = \theta_i^0$ versus $H_A : \theta_i \neq \theta_i^0$
- $LRT = 2 \log L(\hat{\theta}|y) - 2 \log L(\hat{\theta}^0|y) \overset{n \to \infty}{\sim} \chi^2_1$

Problem for testing $H_0 : \text{Var}(b_1) = 0$

Underlying assumptions for asymptotics violated:

- data in LMM not independent
- $\theta^0$ not an interior point of the parameter space $\Theta$
Previous Results:

- **Stram, Lee (1994); Self, Liang (1987):** For i.i.d. observations/subvectors, testing on the boundary of \( \Theta \):
  \[ LRT \overset{as}{\sim} 0.5\delta_0 : 0.5\chi^2_1 \]
Previous Results:

- Stram, Lee (1994); Self, Liang (1987): for i. i. d. observations/subvectors, testing on the boundary of $\Theta$:
  \[ LRT \overset{as}{\sim} 0.5\delta_0 : 0.5\chi_1^2 \]

- Crainiceanu, Ruppert (2004):
  - Stram/Lee mixture very conservative for non-i. i. d. data, small samples
  - $LRT$ often with large point mass at zero, restricted $LRT$ (RLRT) more useful
  - derive exact finite sample distributions of $LRT$ and $RLRT$ in LMMs with one variance component
Previous Results:

- **Stram, Lee (1994); Self, Liang (1987):** for i.i.d. observations/subvectors, testing on the boundary of $\Theta$:
  \[ LRT \sim 0.5\delta_0 : 0.5\chi^2_1 \]

- **Crainiceanu, Ruppert (2004):**
  - Stram/Lee mixture very conservative for non-i.i.d. data, small samples
  - $LRT$ often with large point mass at zero, restricted $LRT$ ($RLRT$) more useful
  - derive exact finite sample distributions of $LRT$ and $RLRT$ in LMMs with one variance component

- **Greven et al. (2007):**
  pseudo-ML arguments to justify application of results in Crainiceanu, Ruppert (2004) to models with multiple variance components
**RLRsim: Algorithm**

\[ RLR_{T_n} \sim \sup_{\lambda \geq 0} \left( (n - p) \log \left( 1 + \frac{N_n(\lambda)}{D_n(\lambda)} \right) - \sum_{k=1}^{K} \log \left( 1 + \lambda \mu_k,n \right) \right), \]

\[ N_n(\lambda) = \sum_{k=1}^{K} \frac{\lambda \mu_{k,n}}{1 + \lambda \mu_{k,n}} w_k^2; \quad D_n(\lambda) = \sum_{k=1}^{K} \frac{w_k^2}{1 + \lambda \mu_{k,n}} + \sum_{k=K+1}^{n-p} w_k^2 \]

\[ w_k \sim \mathcal{N}(0, 1); \quad \mu: \text{eigenvalues of } \Sigma^{1/2}Z'(I_n - X(X'X)^{-1}X)Z\Sigma^{1/2} \]
**RLRsim: Algorithm**

\[
RLRT_n \sim \sup_{\lambda \geq 0} \left( (n - p) \log \left( 1 + \frac{N_n(\lambda)}{D_n(\lambda)} \right) - \sum_{k=1}^{K} \log \left( 1 + \lambda \mu_{k,n} \right) \right),
\]

\[
N_n(\lambda) = \sum_{k=1}^{K} \frac{\lambda \mu_{k,n}}{1 + \lambda \mu_{k,n}} w_k^2; \quad D_n(\lambda) = \sum_{k=1}^{K} \frac{w_k^2}{1 + \lambda \mu_{k,n}} + \sum_{k=K+1}^{n-p} w_k^2
\]

\[w_k \sim \mathcal{N}(0, 1); \quad \mu: \text{eigenvalues of } \Sigma^{1/2}Z'(I_n - X(X'X)^{-1}X)Z\Sigma^{1/2}\]

Rapid simulation from this distribution:

- do eigenvalue decomposition to get \(\mu\)
- repeat:
  - draw \((K + 1)\) \(\chi^2\) variates
  - one-dimensional maximization in \(\lambda\) (via grid search)
RLRsim: Algorithm

\[ RLRT_n \sim \sup_{\lambda \geq 0} \left( (n - p) \log \left( 1 + \frac{N_n(\lambda)}{D_n(\lambda)} \right) - \sum_{k=1}^{K} \log (1 + \lambda \mu_{k,n}) \right), \]

\[ N_n(\lambda) = \sum_{k=1}^{K} \frac{\lambda \mu_{k,n}}{1 + \lambda \mu_{k,n}} w_k^2; \quad D_n(\lambda) = \sum_{k=1}^{K} \frac{w_k^2}{1 + \lambda \mu_{k,n}} + \sum_{k=K+1}^{n-p} w_k^2 \]

\[ w_k \sim \mathcal{N}(0, 1); \quad \mu: \text{eigenvalues of } \Sigma^{1/2}Z'(I_n - X(X'X)^{-1}X)Z\Sigma^{1/2} \]

Rapid simulation from this distribution:

- do eigenvalue decomposition to get \( \mu \)
- repeat:
  - draw \((K + 1)\) \(\chi^2\) variates
  - one-dimensional maximization in \(\lambda\) (via grid search)

→ computational cost depends on \(K\), not \(n\)
→ implemented in C ⇒ quasi-instantaneous
→ easy extension to models with \(L > 1\)
Example: One Variance Component

Test for random intercept (nlme::lme):

```r
> m0 <- lme(distance ~ age + Sex, data = Orthodont, random = ~ 1)
> system.time(print( exactRLRT(m0) ), gcFirst=T)

simulated finite sample distribution of RLRT.
(p-value based on 10000 simulated values)
RLRT = 47.0114, p-value < 2.2e-16

user     system    elapsed
 0.42      0.00      0.42
```

```r
> system.time(simulate.lme(m0,nsim=10000,method='REML'), gcFirst=T)

user     system    elapsed
55.00     0.03      55.48
```
Example: Two Variance Components

Test for random slope with nuisance random intercept
(lme4::lmer):

```r
> m0 <- lmer(distance ~ age + Sex + (1|Subject), data = Orthodont)
> mA <- update(m0, .~. + (0 + age|Subject))
> mSlope <- update(mA, .~. - (1|Subject))
> exactRLRT(mSlope, mA, m0)
```

simulated finite sample distribution of RLRT.
(p-value based on 10000 simulated values)

RLRT = 0.8672, p-value = 0.1603
Example: Testing for Linearity of a Smooth Function

```r
> library(mgcv); data(trees)
> m1 <- gamm(I(log(Volume)) ~ Height + s(Girth, m = 2),
+     data = trees)$lme
```

Significant deviations from linearity?
Example: Testing for Linearity of a Smooth Function

```r
> library(mgcv); data(trees)
> m1 <- gamm(I(log(Volume)) ~ Height + s(Girth, m = 2),
+       data = trees)$lme
>
> exactRLRT(ml)

simulated finite sample distribution of RLRT. 
(p-value based on 10000 simulated values) 
RLRT = 5.4561, p-value = 0.0052
## Simulation Study: Settings

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>tested VC</th>
<th>nuisance VCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>equality of group means</td>
<td>random intercept</td>
<td>- random slope uni- / bivariate smooth</td>
</tr>
<tr>
<td>equality of group trends</td>
<td>random slope</td>
<td>random intercept</td>
</tr>
<tr>
<td>no effect / linearity</td>
<td>univariate smooth</td>
<td>- random intercept uni- / bivariate smooth</td>
</tr>
<tr>
<td>additivity</td>
<td>bivariate smooth</td>
<td>2 univariate smooths</td>
</tr>
</tbody>
</table>

Goal: compare size & power of tests for zero variance components

- sample sizes $n = 50, 100, 500$
- mildly unbalanced group sizes for $K = 5, 20$
- details: Scheipl, Greven, Küchenhoff (2007)
Simulation study

Compared Tests:

- **RLR-type tests:**
  RLRsim, parametric bootstrap, $0.5\delta_0 : 0.5\chi_1^2$

- **F-type tests:**
  bootstrap $F$-type statistics, mgcv’s approximate $F$-test, SAS-implementations of generalized $F$-test etc..
Simulation study

Compared Tests:

- **RLR-type tests:**
  - RLRsim, parametric bootstrap, $0.5\delta_0 : 0.5\chi_1^2$

- **F-type tests:**
  - bootstrap $F$-type statistics, mgcv’s approximate $F$-test,
  - SAS-implementations of generalized $F$-test etc..

Main Results:

- **RLRsim:** equivalent performance to bootstrap RLRT, but practically instantaneous
- $\chi^2$-mixture approximation for RLRT: always conservative, lower than nominal size & reduced power
- bootstrap RLRT, bootstrap $F$-type statistics similar
- $F$-test from mgcv: similar power as $\chi^2$-mixture, occasionally seriously anti-conservative
Conclusion

- conventional RLRTs for $\text{Var(Random Effect)} = 0$ are broken, but not beyond repair.

  $\Rightarrow$ RLRsim

  - is a rapid, more powerful alternative that performs as well as a parametric bootstrap.
  - has a convenient interface for models fit with `nlme::lme` or `lme4::lmer`.
  - Current limitations: no correlated random effects, no serial correlation, only Gaussian responses.
Further Reading:

