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]

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

useR! 2008
August 13, 2008

Outline

Linear Mixed Models

$$
\begin{aligned}
\mathbf{y} &= \mathbf{X}\beta + \sum_{l=1}^{L} \mathbf{Z}_l \mathbf{b}_l + \varepsilon \\
\mathbf{b_l} &\sim \mathcal{N}_{K_l}(\mathbf{0}, \lambda_l \sigma_\varepsilon^2 \mathbf{\Sigma}_l),\ \mathbf{b_l} \perp \mathbf{b_s}\, \forall l \neq s \\
\varepsilon &\sim \mathcal{N}_n(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_n),
\end{aligned}
$$

Linear Mixed Models

$$\mathbf{y} = \mathbf{X}\beta + \sum_{l=1}^{L} \mathbf{Z}_l \mathbf{b}_l + \varepsilon$$
$$\mathbf{b_l} \sim \mathcal{N}_{K_l}(\mathbf{0}, \lambda_l \sigma_\varepsilon^2 \mathbf{\Sigma}_l),\ \mathbf{b_l} \perp \mathbf{b_s}\ \forall l \neq s$$
$$\varepsilon \sim \mathcal{N}_n(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_n),$$

We want to test

$$H_{0,l} : \lambda_l = 0 \text{ versus } \quad H_{A,l} : \lambda_l > 0$$
$$\Leftrightarrow H_{0,l} : \text{Var}(\mathbf{b_l}) = 0 \text{ versus } \quad H_{A,l} : \text{Var}(\mathbf{b_l}) > 0$$

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:

$$\mathbf{y} = f(\mathbf{x}) + \varepsilon$$

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

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

Additive Models as Linear Mixed Models

Simple additive model:

$$\mathbf{y} = f(\mathbf{x}) + \varepsilon$$

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

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

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

Problem:
Likelihood Ratio Tests for Zero Variance Components

General Case:

- $y_1, \ldots, y_n \overset{\text{i.i.d.}}{\sim} f(y|\boldsymbol{\theta})$; $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$
- Test: $H_0 : \theta_i = \theta_i^0$ versus $H_A : \theta_i \neq \theta_i^0$
- $LRT = 2 \log L(\hat{\boldsymbol{\theta}}|\mathbf{y}) - 2 \log L(\hat{\boldsymbol{\theta}}^{\mathbf{0}}|\mathbf{y}) \overset{n \to \infty}{\sim} \chi_1^2$

Problem for testing $H_0 : \text{Var}(\mathbf{b_l}) = 0$
Underlying assumptions for asymptotics violated:

- data in LMM not independent
- $\boldsymbol{\theta}^0$ not an interior point of the parameter space Θ

Previous Results:

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

Previous Results:

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

Previous Results:

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

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

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

$$w_k \sim \mathcal{N}(0,1); \;\; \boldsymbol{\mu}\text{: eigenvalues of } \mathbf{\Sigma}^{1/2}\mathbf{Z}'(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X})\mathbf{Z}\mathbf{\Sigma}^{1/2}$$

RLRsim: Algorithm

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

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

$$w_k \sim \mathcal{N}(0,1); \;\; \boldsymbol{\mu}\text{: eigenvalues of } \boldsymbol{\Sigma}^{1/2}\mathbf{Z}'(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X})\mathbf{Z}\boldsymbol{\Sigma}^{1/2}$$

Rapid simulation from this distribution:

- do eigenvalue decomposition to get $\boldsymbol{\mu}$
- repeat:
  - draw $(K+1)$ χ^2 variates
  - one-dimensional maximization in λ (via grid search)

RLRsim: Algorithm

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

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

$$w_k \sim \mathcal{N}(0,1); \;\; \boldsymbol{\mu}\text{: eigenvalues of } \boldsymbol{\Sigma}^{1/2}\mathbf{Z}'(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X})\mathbf{Z}\boldsymbol{\Sigma}^{1/2}$$

Rapid simulation from this distribution:

- do eigenvalue decomposition to get $\boldsymbol{\mu}$
- repeat:
  - draw $(K+1)$ χ^2 variates
  - one-dimensional maximization in λ (via grid search)

$\rightarrow$ computational cost depends on K, not n

$\rightarrow$ implemented in C $\Rightarrow$ quasi-instantaneous

$\rightarrow$ easy extension to models with $L > 1$

Example: One Variance Component

Test for random intercept (`nlme::lme`):

```
> 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

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

       user     system    elapsed
      55.00       0.03      55.48
```

Example: Two Variance Components

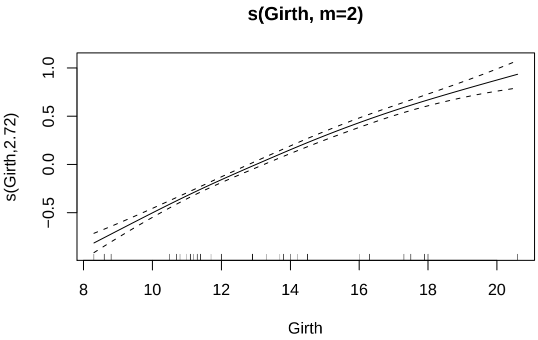
Test for random slope with nuisance random intercept (`lme4::lmer`):

```
> 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

```
> 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

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

> exactRLRT(m1)

        simulated finite sample distribution of RLRT.
        (p-value based on 10000 simulated values)
RLRT = 5.4561, p-value = 0.0052
```

Simulation Study: Settings

H_0	tested VC	nuisance VCs
equality of group means	random intercept	-
		random slope
		uni-/bivariate smooth
equality of group trends	random slope	random intercept
no effect / linearity	univariate smooth	-
		random intercept
		uni-/bivariate smooth
additivity	bivariate smooth	2 univariate smooths

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, $0.5\delta_0 : 0.5\chi^2_1$
- *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, $0.5\delta_0 : 0.5\chi^2_1$
- *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
- χ^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 χ^2-mixture, occasionally seriously anti-conservative

Conclusion

- conventional RLRTs for 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:

- Crainiceanu, C. and Ruppert, D. (2004) *Likelihood ratio tests in linear mixed models with one variance component*, JRSS-B, **66**(1), 165–185.
- Greven, S., Crainiceanu, C. M., Küchenhoff, H. and Peters, A. (2008) *Restricted Likelihood Ratio Testing for Zero Variance Components in Linear Mixed Models*, JCGS, to appear.
- Scheipl, F., Greven, S., and Küchenhoff, H. (2008) *Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models*, CSDA, **52**(7), 3283–3299.