

5. Inference in the longitudinal LMM

Sonja Greven

Summer Term 2016

Overview Chapter 5 - Inference in the longitudinal LMM

5.1 Covariance matrix for $\hat{\beta}$

5.2 Inference for the fixed effects

5.3 Inference for the covariance

Covariance matrix for $\hat{\beta}$

Marginal model:

$$\mathbf{Y}_i \sim \mathcal{N}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i(\boldsymbol{\alpha})), \quad \text{or} \quad \mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\alpha})). \quad (5.1)$$

Remember that the ML estimator for $\boldsymbol{\beta}$ is given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{ML}(\boldsymbol{\alpha}) &= \left\{ \sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i(\boldsymbol{\alpha})^{-1} \mathbf{X}_i \right\}^{-1} \left(\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i(\boldsymbol{\alpha})^{-1} \mathbf{Y}_i \right), \\ &= \left\{ \mathbf{X}^T \mathbf{V}(\boldsymbol{\alpha})^{-1} \mathbf{X} \right\}^{-1} \mathbf{X}^T \mathbf{V}(\boldsymbol{\alpha})^{-1} \mathbf{Y}, \end{aligned}$$

where $\boldsymbol{\alpha}$ is replaced by its ML or REML estimate in practice.

$\hat{\beta}_{ML}(\alpha)$ and $\hat{\beta}_{\mathbf{W}}$

For a given α , $\hat{\beta}_{ML}(\alpha)$ is equal to the weighted least squares estimator

$$\hat{\beta}_{ML}(\alpha) = \hat{\beta}_{\mathbf{W}} = \{\mathbf{X}^T \mathbf{W} \mathbf{X}\}^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y} \quad \text{with} \quad \mathbf{W} = \mathbf{V}^{-1}(\alpha).$$

Generally, for any \mathbf{W} , we have

$$\begin{aligned} E(\hat{\beta}_{\mathbf{W}}) &= \{\mathbf{X}^T \mathbf{W} \mathbf{X}\}^{-1} \mathbf{X}^T \mathbf{W} (\mathbf{X} \boldsymbol{\beta}) = \boldsymbol{\beta} \\ \text{Cov}(\hat{\beta}_{\mathbf{W}}) &= \{\mathbf{X}^T \mathbf{W} \mathbf{X}\}^{-1} \{\mathbf{X}^T \mathbf{W} \text{Cov}(\mathbf{Y}) \mathbf{W} \mathbf{X}\} \{\mathbf{X}^T \mathbf{W} \mathbf{X}\}^{-1} \end{aligned}$$

and $\hat{\beta}_{\mathbf{W}}$ is multivariate normal under the normality assumption (5.1). Thus, $\hat{\beta}_{\mathbf{W}}$ is **unbiased** with any \mathbf{W} as long as $E(\mathbf{Y}) = \mathbf{X} \boldsymbol{\beta}$ is correctly specified.

$\hat{\beta}_{ML}(\alpha)$ and $\hat{\beta}_{\mathbf{W}}$

How does the choice of \mathbf{W} affect the covariance of $\hat{\beta}_{\mathbf{W}}$?

If $\text{Cov}(\mathbf{Y}) = \mathbf{V}$, we have that for $\hat{\beta}_{ML} = \hat{\beta}_{\mathbf{V}^{-1}}$

$$\begin{aligned}\text{Cov}(\hat{\beta}_{ML}) &= \{\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\}^{-1} \{\mathbf{X}^T \mathbf{V}^{-1} \text{Cov}(\mathbf{Y}) \mathbf{V}^{-1} \mathbf{X}\} \{\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\}^{-1} \\ &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1},\end{aligned}$$

which can be shown to be smaller (in a matrix sense) than

$$\text{Cov}(\hat{\beta}_{\mathbf{W}}) = \{\mathbf{X}^T \mathbf{W} \mathbf{X}\}^{-1} \{\mathbf{X}^T \mathbf{W} \text{Cov}(\mathbf{Y}) \mathbf{W} \mathbf{X}\} \{\mathbf{X}^T \mathbf{W} \mathbf{X}\}^{-1}$$

for any other \mathbf{W} . Thus, with $\mathbf{W} = \text{Cov}(\mathbf{Y})^{-1}$ the estimator $\hat{\beta}_{\mathbf{W}}$ is most efficient.

Covariance matrix for $\hat{\beta}_{ML}(\alpha)$

$$\begin{aligned}\text{Cov}(\hat{\beta}_{ML}(\alpha)) &= \{\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\}^{-1} \{\mathbf{X}^T \mathbf{V}^{-1} \text{Cov}(\mathbf{Y}) \mathbf{V}^{-1} \mathbf{X}\} \{\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\}^{-1} \\ &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}\end{aligned}$$

- The equality between first and second lines assumes that $\text{Cov}(\mathbf{Y}) = \mathbf{V}$, i.e. that \mathbf{V} is correctly specified.
- In practice, \mathbf{V} is unknown and we have to make assumptions.
- Also, α is replaced by its ML or REML estimate and the resulting additional uncertainty is not accounted for in $\text{Cov}(\hat{\beta}_{ML}(\alpha))$.

Robust covariance for the fixed effects

What happens if our model for $\text{Cov}(\mathbf{Y})$ is wrong? For example, we might assume compound symmetry with constant variance over time, but in fact variances increase over time, or correlations decrease with time distance.

- If the covariance is incorrectly specified, $\text{Cov}(\mathbf{Y}) \neq \mathbf{V}$, the **model-based covariance** $(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$ can be very different from the **true covariance** $\text{Cov}(\hat{\beta}_{ML}(\alpha))$. Then, standard errors, confidence intervals and tests for β based on $(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$ will be incorrect.
- An analysis based on $\text{Cov}(\hat{\beta}_{ML}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$ is thus not **robust** against misspecification of \mathbf{V} .

Robust covariance for the fixed effects

We can write

$$\text{Cov}(\hat{\boldsymbol{\beta}}_{ML}(\boldsymbol{\alpha})) = \left\{ \sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right\}^{-1} \left\{ \sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \text{Cov}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \mathbf{X}_i \right\} \times \left\{ \sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right\}^{-1} \quad (5.2)$$

The **sandwich estimator**, also called **robust** or **empirical variance estimator**, replaces \mathbf{V}_i in (5.2) by $\mathbf{V}_i(\hat{\boldsymbol{\alpha}})$ and $\text{Cov}(\mathbf{Y}_i)$ by

$$\mathbf{r}_i \mathbf{r}_i^T = (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^T.$$

Robust covariance for the fixed effects

The sandwich estimator is

- consistent for $\text{Cov}(\hat{\boldsymbol{\beta}})$ if $N \rightarrow \infty$, as long as $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ is correct.
- a better estimator the higher N is and should be avoided for small N .
- valid even when the covariance is misspecified.
- less efficient than the model-based covariance estimator if the covariance is correctly specified

Correctly specifying the covariance structure is also important when we want to interpret the covariance for our data.

A small simulation study

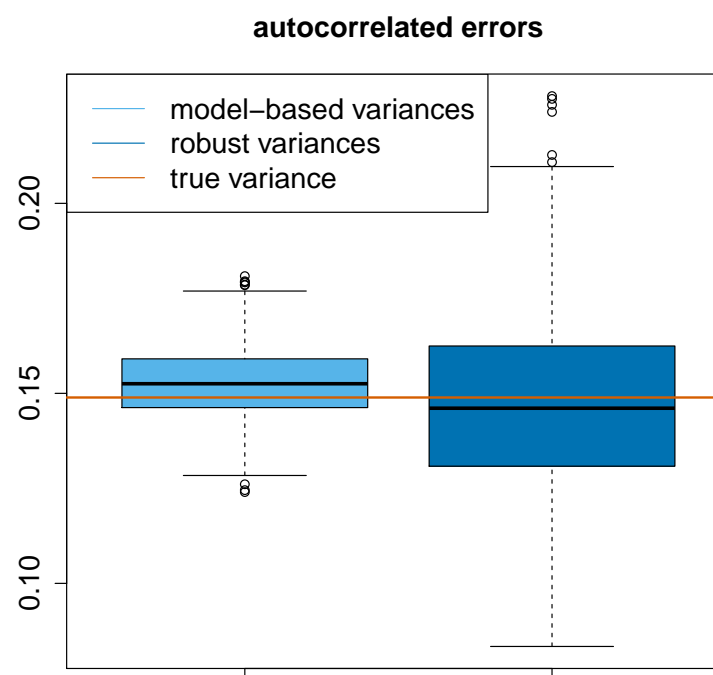
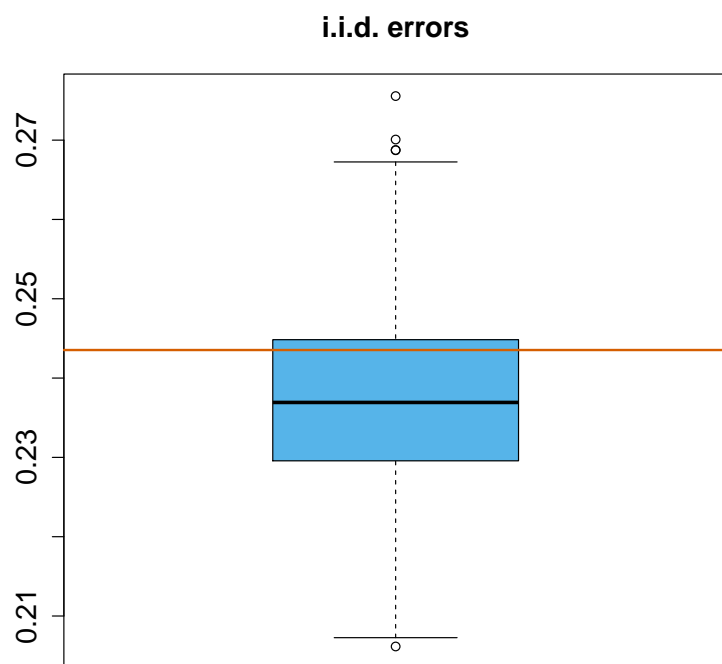
Simulate $R = 500$ data sets from the model

$$Y_{ij} = \beta x_{ij} + b_i + \epsilon_{ij}, \quad i = 1, \dots, 100; \quad j = 1, \dots, 10,$$

with b_i i.i.d. $\mathcal{N}(0, 15^2)$ and ϵ_i i.i.d. $\mathcal{N}(\mathbf{0}_{n_i}, \Sigma_i)$, where a) $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$ or b) Σ_i equal to τ^2 times an AR(1) correlation matrix ($\rho = 0.7, \tau^2 = 15^2$). Obtain estimates $\hat{\beta}$ for $\beta = 1$ and model-based estimates for the variance of $\hat{\beta}$ under the assumption of random intercept plus i.i.d. errors. We can

- approximate the true variance of $\hat{\beta}$ by the variance of the R estimates
- look at the spread of the model-based variances in the R runs
- compare this to the spread of the robust variances in the R runs.

A small simulation study - Results



Covariance matrix of $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}})$

It can be useful to construct confidence intervals (e.g. for the longitudinal trajectory for subject i) or prediction intervals (e.g. for a new observation) for expressions involving \mathbf{b} . We then need the covariance matrix for

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{pmatrix} = (\mathbf{C}^T \mathbf{R}^{-1} \mathbf{C} + \text{diag}(\mathbf{0}, \mathbf{G}^{-1}))^{-1} \mathbf{C}^T \mathbf{R}^{-1} \mathbf{Y},$$

with $\mathbf{C} = (\mathbf{X} | \mathbf{Z})$, which can be seen to be (model-based)

$$\text{Cov} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} - \mathbf{b} \end{pmatrix} = (\mathbf{C}^T \mathbf{R}^{-1} \mathbf{C} + \text{diag}(\mathbf{0}, \mathbf{G}^{-1}))^{-1}.$$

Note that $\text{Cov}(\hat{\mathbf{b}} - \mathbf{b})$ is used, as $\text{Cov}(\hat{\mathbf{b}})$ does not recognize the random variation in \mathbf{b} .

Overview Chapter 5 - Inference in the longitudinal LMM

5.1 Covariance matrix for $\hat{\beta}$

5.2 Inference for the fixed effects

5.3 Inference for the covariance

Hypothesis testing for β - The TLC trial

Consider again the model from 4.1 for lead blood levels in the TLC trial

$$Y_{ij} = \beta_0 + \beta_1 g_i + \beta_2 I(t_j = 1) + \beta_3 I(t_j = 4) + \beta_4 I(t_j = 6) \\ \beta_5 g_i I(t_j = 1) + \beta_6 g_i I(t_j = 4) + \beta_7 g_i I(t_j = 6) + b_i + \epsilon_{ij}.$$

We might be interested in testing the following hypotheses:

- $H_{0,1}: \beta_1 = 0$
- $H_{0,2}: \beta_5 = \beta_6 = \beta_7 = 0$.

Interpretation?

We can formulate these hypotheses as linear hypotheses on β ,

$$H_0 : \mathbf{L}\beta = \mathbf{0} \quad \text{versus} \quad H_1 : \mathbf{L}\beta \neq \mathbf{0}, \quad (5.3)$$

where \mathbf{L} is a matrix with p columns. For example, $H_{0,1}$ corresponds to

$$\mathbf{L} = (0, 1, 0, 0, 0, 0, 0, 0)$$

and $H_{0,2}$ to

$$\mathbf{L} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

We can use an (approximate) Wald-Test, an (approximate) t - or F -Test or a likelihood ratio test (LRT) to test for the fixed effects.

Approximative Wald test

The Wald test uses that $\hat{\boldsymbol{\beta}}$ is approximately normally distributed:

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \stackrel{a}{\sim} \mathcal{N}(\mathbf{0}, (\mathbf{X}^T \mathbf{V}(\hat{\boldsymbol{\alpha}})^{-1} \mathbf{X})^{-1}).$$

Test statistic:

$$T = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{L}^T \left[\mathbf{L} \left(\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i(\hat{\boldsymbol{\alpha}})^{-1} \mathbf{X}_i \right)^{-1} \mathbf{L}^T \right]^{-1} \mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

Null distribution: Under $H_0: \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$, T is approximately χ^2 distributed with $\text{rank}(\mathbf{L})$ degrees of freedom. (Note that the asymptotics here are for $N \rightarrow \infty$.)

Approximative t - or F -test

The Wald test statistic uses the estimated covariance which ignores estimation uncertainty in $\hat{\alpha}$ and underestimates the variability in $\hat{\beta}$.

This problem is often alleviated by using approximate t - or F -statistics:

- For a single β_k , the distribution of $(\hat{\beta}_k - \beta_k) / \widehat{SE}(\hat{\beta}_k)$ is approximated by a t -distribution.
- For general hypotheses (5.3), an F -approximation to the distribution of $F = T / \text{rank}(\mathbf{L})$ is used, with $\text{rank}(\mathbf{L})$ as numerator degrees of freedom.

The t degrees of freedom or the denominator degrees of freedom for the F -distribution are estimated from the data.

- A Satterthwaite-type approximation is commonly used (it basically matches the moments of the distributions)
- The Kenward-Rogers method gives better results for small samples.

Estimating the degrees of freedom can be computationally expensive.

Note: The approximate Wald, t - and F -tests can also be ‘robustified’ by using the robust covariance matrix estimate instead of $\left(\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i (\hat{\boldsymbol{\alpha}})^{-1} \mathbf{X}_i\right)^{-1}$.

Likelihood ratio test (LRT)

An alternative to the Wald test is the likelihood ratio test (LRT). For small samples, the LRT tends to be more reliable than the Wald test.

The LRT can be used for general nested hypotheses

$$H_0 : \beta \in \Theta_{\beta,0} \quad \text{versus} \quad H_1 : \beta \in \Theta_{\beta},$$

where $\Theta_{\beta,0}$ is a subspace of the parameter space Θ_{β} for the fixed effects.

Test statistic:

$$\lambda_N = -2 \ln \left[\frac{L_{ML}(\widehat{\theta}_{ML,0})}{L_{ML}(\widehat{\theta}_{ML})} \right],$$

with $\widehat{\theta}_{ML,0}$ and $\widehat{\theta}_{ML}$ the ML estimators over $\Theta_{\beta,0}$ and Θ_{β} , respectively.

Null distribution: Under H_0 , λ_N is asymptotically ($N \rightarrow \infty$) chi-square distributed with degrees of freedom (df) equal to the difference between the dimensions of Θ_{β} and $\Theta_{\beta,0}$.

Important:

Remember that for REML estimation, the likelihood is based on the error contrasts and these depend on \mathbf{X} .

⇒ Restricted likelihood values are not comparable between H_0 and H_1 when the fixed effects differ (“comparing apples and oranges”).

⇒ **Then, the likelihood ratio test cannot be used with REML estimation!**

Example: TLC trial

Test for $H_{0,1}: \beta_1 = 0$. Extract from the output for `summary(lmeREML)`:

Fixed effects: `lead ~ group * week`

	Value	Std.Error	DF	t-value	p-value
(Intercept)	26.272	0.9370175	294	28.037898	0.0000
groupS	0.268	1.3251428	98	0.202242	0.8401
week1	-1.612	0.8428574	294	-1.912542	0.0568
week4	-2.202	0.8428574	294	-2.612542	0.0094
week6	-2.626	0.8428574	294	-3.115592	0.0020
groupS:week1	-11.406	1.1919804	294	-9.568950	0.0000
groupS:week4	-8.824	1.1919804	294	-7.402807	0.0000
groupS:week6	-3.152	1.1919804	294	-2.644339	0.0086

Interpretation? Which test does R use?

Example: TLC trial

Likelihood ratio test for $H_{0,2}: \beta_5 = \beta_6 = \beta_7 = 0$:

	ML	REML
Under $\beta \in \mathbf{R}^8$	$\ell_{ML} = -1235.411$	$\ell_{REML} = -1230.311$
Under $\beta_5 = \beta_6 = \beta_7 = 0$	$\ell_{ML} = -1284.682$	$\ell_{REML} = -1282.019$
λ_N	98.543	-
df	3	-
p-value	< 0.0001	-

```
> lmeML <- lme(lead ~ group * week, random = ~ 1 | id,
              data = lead, method = "ML")
> lmeML2 <- lme(lead ~ group + week, random = ~ 1 | id,
               data = lead, method = "ML")
> anova(lmeML2, lmeML)
```

Overview Chapter 5 - Inference in the longitudinal LMM

5.1 Covariance matrix for $\hat{\beta}$

5.2 Inference for the fixed effects

5.3 Inference for the covariance

Inference for the variance components

While inference for β is often of primary interest, inference for the variance components is also of importance:

- For interpreting the random effects and variance structure in the data.
- As a correctly specified variance structure is necessary for valid inference for the fixed effects (unless using robust inference)
- As many methods to analyze data with missings or drop-out need a correctly specified variance structure (see Chapter 11).

Inference for the variance components - Example rats

Consider as an example the rat data and the model

$$Y_{ij} = \beta_0 + b_{1i} + (\beta_{g_i} + b_{2i})t_j + \epsilon_{ij}$$

- Y_{ij} skull height for animal i at time t_j
- t_j time variable, $t = \log(1 + (age - 45)/10)$
- g_i treatment group of rat i , $g_i \in \{1, 2, 3\}$

with assumptions

$$\mathbf{b}_i = (b_{1i}, b_{2i}) \sim \mathcal{N}(\mathbf{0}, \mathbf{D}), \quad \text{ind. of} \quad \epsilon_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}).$$

Denote

$$\text{Cov}(\mathbf{b}_i) = \mathbf{D} = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}.$$

Consider three possible models:

- M_0 : no random effects ($\mathbf{b}_i \equiv \mathbf{0}$), $d_{11} = d_{12} = d_{22} = 0$
- M_1 : only a random intercept ($b_{2i} \equiv 0$), $d_{12} = d_{22} = 0$
- M_2 : (correlated) random intercept and slope.

We can compare M_2 and M_1 by testing for $H_{0,1} : d_{12} = d_{22} = 0$

and M_1 and M_0 by testing for $H_{0,2} : d_{11} = 0$.

LRT for the variance components

- We can test these hypotheses using a likelihood ratio test.
- Different from testing for the fixed effects, the LRT can also be used with REML estimation if the fixed effects do not differ between H_0 and H_1 .
- The test statistic is again two times the difference in log-likelihoods between alternative and null model.
- What is the null distribution? For variances, a common assumption is violated: Under the null, the parameter does not lie in the interior but on the boundary of the parameter space, as 0 is on the boundary of $[0, \infty)$.

LRT for the variance components

What happens intuitively under $H_{0,2} : d_{11} = 0$ for large N ?

- In (approximately) half of the cases, d_{11} is estimated to be positive.
- In (approximately) half of the cases, d_{11} would be estimated to be negative, but \hat{d}_{11} is instead set to 0 due the restriction $d_{11} \geq 0$. (If the numerical estimation routine properly incorporates these constraints.)

Thus, \hat{d}_{11} is not approximately normal for large N , but follows a 0.5:0.5 mixture of a point mass at zero and a half-normal distribution.

Likewise, the LRT statistic is not asymptotically χ_1^2 distributed, but follows a 0.5:0.5 mixture of a point mass at zero and a χ_1^2 distribution under H_0 .

LRT for the variance components

More generally, for

$$\mathbf{D} = \begin{pmatrix} d_{11} & \cdots & d_{1q} \\ \vdots & \ddots & \vdots \\ d_{1q} & \cdots & d_{qq} \end{pmatrix}$$

and testing

$$H_0 : d_{1q} = \cdots = d_{qq} = 0 \quad \text{vs.} \quad H_1 : d_{qq} > 0,$$

the null distribution of the LRT is asymptotically ($N \rightarrow \infty$) a 0.5:0.5 mixture between a χ_{q-1}^2 and a χ_q^2 distribution. (χ_0^2 denotes a point mass at zero.) p-values are averages between those that would be obtained using a χ_{q-1}^2 and a χ_q^2 distribution, respectively.

LRT for the variance components - Some notes

- Note that this mixture is not valid if a nuisance parameter is on the boundary (e.g. if testing the random slope, the random intercept is ≈ 0).
- For $q = 1$ and $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$, an exact distribution is available ([Crainiceanu and Ruppert, 2004](#)) and implemented in the R-package `RLRsim`.
- For testing a covariance parameter, the asymptotic null distribution is χ_1^2 .
- For more complex hypotheses for D , more complex chi-square mixtures can arise (see [Self & Liang, 1987](#); [Stram & Lee, 1994, 1995](#)).
- Score or Wald tests have the same asymptotic distribution as the LRT, but are more difficult to compute here ([Molenberghs & Verbeke, 2007](#)).

Example: The rat data, testing $H_{0,1} : d_{12} = d_{22} = 0$

```
> lme1 <- lme(RESPONSE ~ group * t - group,
              random = ~ 1 | SUBJECT, data = rats)
> lme2 <- lme(RESPONSE ~ group * t - group,
              random = ~ t | SUBJECT, data = rats)
> anova(lme1, lme2)
      Model df      logLik  Test      L.Ratio p-value
lme1      1  6   -466.2016
lme2      2  8   -466.2016 1 vs 2 1.857238e-07      1
> test.statistic <- 1.857238e-07
> p.value <- 0.5 * (1-pchisq(test.statistic,1)) +
             0.5 * (1-pchisq(test.statistic,2))
> p.value #p-value from equal mixture chi_1^2:chi_2^2
[1] 0.999828
```

Interpretation?

Example: The rat data, testing $H_{0,2} : d_{11} = 0$

```
> lme0 <- lm(RESPONSE ~ group * t - group, data = rats)
> library(RLRsim)
> exactRLRT(m = lme1, m0 = lme0, mA = lme1) # exact p-value
simulated finite sample distribution of RLRT.
(p-value based on 10000 simulated values)
data:
RLRT = 175.421, p-value < 2.2e-16

> test.statistic <- 175.421
> p.value <- 0.5 * (1-pchisq(test.statistic,df = 1))
> p.value #p-value from equal mixture  $\chi_0^2:\chi_1^2$ 
[1] 0
```

Interpretation?