Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. *http://www.jstatsoft.org/*

A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models – the R Package pbkrtest

Ulrich Halekoh
Aarhus University

Søren Højsgaard
Aalborg University

Abstract

When testing for reduction of the mean value structure in linear mixed models, it is common to use an asymptotic χ^2 test; for example that minus twice the maximized log–likelihood has an approximate χ^2 distribution under the hypothesis. Such tests can, however, be very poor for small and moderate sample sizes. The **pbkrtest** package implements two alternatives to such approximate χ^2–tests: The package implements a Kenward–Roger approximation for performing F–tests for reduction of the mean structure and also parametric bootstrap methods for achieving the same goal. In addition to describing the methods and aspects of their implementation, the paper also contains several examples and comparison of the various methods.

Keywords: adjusted degree of freedom, denominator degree of freedom, linear mixed model, **lme4**, R, parametric bootstrap, Bartlett correction.

1. Introduction

In this paper we address the question of testing for reduction of the systematic components in mixed effects models. Attention is restricted to models which a linear and where all random effects are Gaussian. The focus in this paper is on the implementation of these models in the **lme4** package, (Bates, Maechler, and Bolker 2011) for R, (R Development Core Team 2012); specifically as implemented in the `lmer()` function.

It is always possible to exploit that the LR test statistic has a limiting χ^2 distribution as the amount of information in the sample goes to infinity. We shall refer to this test as *the asymptotic* χ^2*–test*. However for small and moderate sample sizes the χ^2 approximation can be poor and lead to misleading conclusions. For certain types of studies it is possible to base

the inference on an F–statistic. Such studies generally need to be balanced in some way in the sense the number of observations in each treatment group should be the same and so on. These balance requirements can often not be met in practice. Therefore there is a need for tests which, for a fairly large class of linear mixed models, 1) are better than the asymptotic χ^2–test and 2) which are relatively easy to compute in practice.

The paper is structured as follows: Section 2 describes the problem addressed in more detail and sets the notation of the paper. Section 3 illustrates the problems related to tests in mixed models through several examples. In Section 4 describe the approach taken by Kenward and Roger (1997) to address the inference problem. Section 5 describes an alternative approach based on parametric bootstrap methods. In Section 6 we apply the methods to several data sets. Section 7 contains a discussion and outlines some additional improvements that can be made to the implementation in **pbkrtest**, Halekoh and Højsgaard (2012).

2. Preliminaries and notation

In this paper we focus on linear mixed models which, in the formulation of Laird and Ware (1982), are of the form

$$\mathbf{Y}^N = \mathbf{X}^{N\times p}\boldsymbol{\beta}^p + \mathbf{Z}^{N\times u}\mathbf{b}^u + \boldsymbol{\epsilon}^N \tag{1}$$

where $\mathbf{Y}$ is an N vector of observables with covariance $\mathbb{V}\mathrm{ar}(\mathbf{Y}) = \boldsymbol{\Sigma}^{N\times N}$. The superscripts above refer to the dimension of the quantities. In (1), $\mathbf{X}$ and $\mathbf{Z}$ are the design matrices of the fixed and random effect, $\mathbf{b}$ is the random effects distributed as $\mathbf{b}^u \sim N(\mathbf{0}, \boldsymbol{\Gamma}^{u\times e})$ and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}^{N\times N})$ is the residual errors where $\mathbf{I}^{N\times N}$ is the $N \times N$ identity matrix. It is assumed that $\mathbf{b}$ and $\boldsymbol{\epsilon}$ are independent. This model is a simplification of the more general model proposed in Laird and Ware (1982), who allow the covariance matrix of $\boldsymbol{\epsilon}$ to be a general positive definite matrix.

We are interested testing hypotheses about reductions of the mean value in (1), i.e., testing

$$M_0 : \mathbf{Y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon} \tag{2}$$

where $\mathcal{C}(\mathbf{X}_0) \subset \mathcal{C}(\mathbf{X})$ with $\mathcal{C}(\mathbf{X})$ denoting the column space of $\mathbf{X}$. Let $d = \dim(\mathcal{C}(\mathbf{X})) - \dim(\mathcal{C}(\mathbf{X}_0))$. Notice that the structural forms of the random components of the two models are identical.

In some situations a test for $\mathbb{E}(\mathbf{Y}) = \mathbf{X}_0\boldsymbol{\beta}_0$ under $\mathbb{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ can be made as an F–test; one example is given in Section 3. However, in many practical cases, such an exact F–test is not available and one often resorts to asymptotic tests. One approach is based on the likelihood ratio (LR) test statistic T which is twice the difference of the maximized log–likelihoods

$$T = 2(\log L - \log L_0). \tag{3}$$

Under the hypothesis, T has an asymptotic χ^2_d distribution.

The reduction of the large model to the small model can equivalently be expressed by the equation $\mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ with a non-singular $d \times p$ restriction matrix $\mathbf{L}$. A test of the more general hypothesis $\mathbf{L}(\boldsymbol{\beta} - \boldsymbol{\beta}_H) = \mathbf{0}$ can be based on the Wald test statistic

$$W = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H)^\top\mathbf{L}^\top(\mathbf{L}^\top\hat{\mathbf{V}}\mathbf{L})^{-1}\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H), \tag{4}$$

where $\hat{\boldsymbol{\beta}}$ is an estimate for $\boldsymbol{\beta}$ and $\hat{\mathbf{V}}$ for the covariance matrix of $\hat{\boldsymbol{\beta}}$.

In this paper we focus on the case where $\boldsymbol{\beta}_H = \mathbf{0}$. In Appendix B it is shown how $\mathbf{L}$ can be constructed from $\mathbf{X}$ and $\mathbf{X}_0$. Under the hypothesis, W also has an asymptotic χ^2_d distribution and the Wald and the LR test are hence asymptotically equivalent.

The approximation of the null–distribution of T or W by a χ^2_d distribution can for small samples be quite poor and this can lead to misleading conclusions. Nonetheless, this approximation is often used in practice – mainly because of the lack of attractive alternatives. This paper is aimed at providing some remedies for this.

- Kenward and Roger (1997) provide a modification of W given in (4). They also argue that this modified statistic can be evaluated in an $F_{d,m}$ distribution for which they provide a method for estimating the denominator degrees of freedom m. We have implemented their work in the function `KRmodcomp()` for models of the form (1); notice in particular that attention is restricted to models for which the residuals are independent and have constant variance. Throughout this paper we shall refer to Kenward and Roger (1997) as K&R.
- The second contribution of this paper is to determine either the full null–distribution or moments of the null–distribution of the LR test statistic (3) by a parametric bootstrap approach (Harville and Hinkley 1997, chapter 4).

3. The degree of freedom issue for linear mixed models

In this section we discuss the degree of freedom issue on the basis of the dataset `beets` in the **doBy** package, Højsgaard and Halekoh (2012). The `beets` data come from a split–plot experiment. Although the classical analysis of split plot experiments is described many places in the literature, see e.g., Cochran and Cox (1957, chapter 7), we treat the topic in some detail in order to put the other parts of the article into a context.

3.1. The sugar beets example

The experiment was laid out as follows: The effect of harvest time and sowing time on i) yield (in kg) and ii) sugar percentage of sugar beets is investigated. Five different sowing times and two different harvesting times were used and the experiment was laid out in three blocks. The experimental plan is as follows:

```
Experimental plan for sugar beets experiment

Sowing times:
  1: 4/4, 2: 12/4, 3: 21/4, 4: 29/4, 5: 18/5

Harvest times:
  1: 2/10, 2: 21/10

Plot allocation:
      |  Block 1           |  Block 2           |  Block 3           |
      +--------------------|--------------------|--------------------+
```

```
Plot  | h1  h1  h1  h1  h1 | h2  h2  h2  h2  h2 | h1  h1  h1  h1  h1 | Harvest time
 1-15 | s3  s4  s5  s2  s1 | s3  s2  s4  s5  s1 | s5  s2  s3  s4  s1 | Sowing time
      |--------------------|--------------------|--------------------|
Plot  | h2  h2  h2  h2  h2 | h1  h1  h1  h1  h1 | h2  h2  h2  h2  h2 | Harvest time
16-30 | s2  s1  s5  s4  s3 | s4  s1  s3  s2  s5 | s1  s4  s3  s2  s5 | Sowing time
      +--------------------|--------------------|--------------------+
```

Each block is sub–divided into two plots (called whole–plots in the experimental design literature) which are harvested at two different time points. Each whole–plot is further sub–divided into five plots (called split–plots) and each of the five sowing times were applied to one of these split–plots. All together there are hence 6 whole–plots and 30 split–plots. The harvest time is called the whole–plot treatment and the sowing time is called the split–plot treatment. The area of each split plot was $25m^2$.

In the following i denotes harvest time ($i = 1, 2$), j denotes block ($j = 1, 2, 3$) and k denotes sowing time ($k = 1, \ldots, 5$). Let $I = 2$, $J = 3$ and $K = 5$. For simplicity we assume that there is no interaction between sowing and harvesting times (this assumption is supported by Figure 1). Then a typical model for such an experiment would be:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \delta_k + U_{ij} + \epsilon_{ijk}, \tag{5}$$

where $U_{ij} \sim N(0, \omega^2)$ and $\epsilon_{ijk} \sim N(0, \sigma^2)$. Notice that U_{ij} describes the random variation between whole–plots (within blocks).

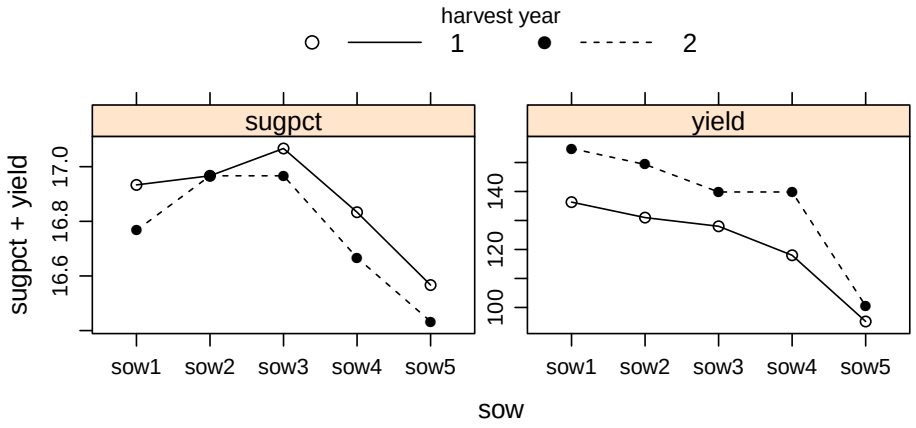

Figure 1: Dependence of sugar percentage and yield [kg] on sowing times and harvest time.

3.2. The asymptotic χ^2–test

Using the `lmer()` function from **lme4**, (Bates *et al.* 2011) we can fit the models and test for no effect of sowing and harvest time as follows:

```
R> library("lme4")
R> data(beets, package='doBy')
R> beet0 <- lmer(sugpct ~ block + sow + harvest + (1|block:harvest),
+                data=beets, REML=FALSE)
R> beet_no.harv <- update(beet0, .~. - harvest)
R> beet_no.sow  <- update(beet0, .~. - sow)
```

We then proceed by testing for no effect of sowing and of harvesting times:

```
R> as.data.frame(anova(beet0, beet_no.sow))

            Df        AIC        BIC    logLik   Chisq Chi Df
beet_no.sow  6  -2.795177   5.612008  7.397588      NA     NA
beet0       10 -79.997378 -65.985404 49.998689 85.2022      4
              Pr(>Chisq)
beet_no.sow           NA
beet0       1.374278e-17

R> as.data.frame(anova(beet0, beet_no.harv))

             Df       AIC       BIC   logLik    Chisq Chi Df
beet_no.harv  9 -69.08356 -56.47278 43.54178       NA     NA
beet0        10 -79.99738 -65.98540 49.99869 12.91382      1
               Pr(>Chisq)
beet_no.harv           NA
beet0        0.0003261646
```

These tests are based on the limiting χ^2 distribution of the LR test statistic and suggest a highly significant effect of both sowing and harvesting times. However the test for no effect of harvesting time is misleading because the hierarchical structure of the data has not been appropriately accounted for. We shall discuss this important issue in detail below.

3.3. The exact F–test

Consider a comparison of two sowing times and of two harvesting times:

$$y_{ij1} - y_{ij2} = \delta_1 - \delta_2 + \epsilon_{ij1} - \epsilon_{ij2} \sim N(\delta_1 - \delta_2, 2\sigma^2) \quad (6)$$

$$y_{1jk} - y_{2jk} = \alpha_1 - \alpha_2 + U_{1j} - U_{2j} + \epsilon_{1jk} - \epsilon_{2jk} \sim N(\alpha_1 - \alpha_2, 2\omega^2 + 2\sigma^2). \quad (7)$$

For the sowing times the whole plot variation cancels out whereas the whole–plot variation prevails for the harvest times. This means that the effect of whole–plot treatments are determined with smaller precision than the effect of split–plot treatments. In some applications (for example if whole–plots are animals and split plots correspond to an application of a treatment at different time points) it is often the case that ω^2 is considerably larger than σ^2. Estimated contrasts for sowing times and harvesting times hence become

$$\frac{1}{IJ}\sum_{ij}(y_{ij1} - y_{ij2}) \sim N(\delta_1 - \delta_2, \frac{2}{IJ}\sigma^2) \quad (8)$$

$$\frac{1}{JK}\sum_{jk}(y_{1jk} - y_{2jk}) \sim N(\alpha_1 - \alpha_2, \frac{2}{J}\omega^2 + \frac{2}{JK}\sigma^2). \quad (9)$$

The variance in (9) is larger than the variance in (8) only if $\omega^2 > \sigma^2(K - I)/(KI)$. For example in the sugar beets example, $(K - I)/(KI) = 3/10$.

Test for no effect of harvest time

Next we consider test statistics. We shall use the notation $y_{i++} = \sum_{jk} y_{ijk}$ and $\bar{y}_{i++} = y_{i++}/(JK)$ etc. Also we let $\tilde{\sigma}^2 = \omega^2 + \sigma^2/K$. The test for no effect of harvest times is based on the marginal model obtained after averaging over the sowing times, i.e.,

$$\bar{y}_{ij+} = \mu + \alpha_i + \beta_j + \bar{\delta}_+ + \bar{U}_{ij} + \bar{\epsilon}_{ij+} \sim N(\mu + \alpha_i + \beta_j + \bar{\delta}_+, \tilde{\sigma}^2). \quad (10)$$

Observe that $\bar{y}_{ij+}$ in (10) has the structure of a model for a balanced two–way layout without replicates. Let $SS_I = \sum_{ijk}(\bar{y}_{i++} - \bar{y}_{+++})^2$ be the sums of squares associated with harvest time. A direct calculation shows that $\mathbb{E}(SS_I) = Q_I + (I-1)[K\omega^2 + \sigma^2]$ where $Q_I = JK\sum_i(\alpha_i - \bar{\alpha}_+)^2$. The corresponding mean squares $MS_I = SS_I/(I-1)$ then has expectation $\mathbb{E}(MS_I) = Q_I/(I-1) + [K\omega^2 + \sigma^2]$. As $Q_I = 0$ iff all α_i are identical, MS_I can be used for constructing a test for no effect of harvesting time. The relevant error sum of squares becomes the residual sum of squares in the marginal model (10), i.e., $SS_{I+J} = \sum_{ijk}(\bar{y}_{ij+} - \bar{y}_{i++} - \bar{y}_{+j+} + \bar{y}_{+++})^2$. A direct calculation shows that $\mathbb{E}(SS_{I+J}) = (I-1)(J-1)[K\omega^2 + \sigma^2]$. Define $MS_{I+J} = SS_{I+J}/[(I-1)(J-1)]$. Then $\mathbb{E}(MS_{I+J}) = [K\omega^2 + \sigma^2]$ and from this we obtain the F–statistic for testing for no effect of harvesting time:

$$F = \frac{MS_I}{MS_{I+J}} \sim F_{(I-1),(I-1)(J-1)} \text{ under the hypothesis.} \tag{11}$$

Test for no effect of sowing time

The test for no effect of sowing time is straight forward. Let $SS_K = \sum_{ijk}(y_{++k} - \bar{y}_{+++})^2 = \sum_{ijk}((\delta_k - \bar{\delta}_+) + (\bar{\epsilon}_{++k} - \bar{\epsilon}_{+++}))^2$ be the sum of squares associated with sowing times and let $MS_K = SS_K/(K-1)$. Following the notation from above a direct calculation shows that $\mathbb{E}(MS_K) = Q_K/(K-1) + \sigma^2$. The corresponding error term becomes $SS_\epsilon = \sum_{ijk}(y_{ijk} - y_{ij+} - y_{++k} + y_{+++})^2$. which is the residual sum of squares for a linear normal model with an effect of sowing time plus an interaction between harvest time and block. Define the mean squares as $MS_\epsilon = SS_\epsilon/(IJ-1)(K-1)$ and a direct calculation shows that $\mathbb{E}(MS_\epsilon) = \sigma^2$ so the F–statistic for no effect of sowing times becomes

$$F = \frac{MS_K}{MS_\epsilon} \sim F_{(K-1),(IJ-1)(K-1)} \text{ under the hypothesis.} \tag{12}$$

Making the relevant F–tests with `aov()`

The `aov()` function makes the tests in (11) and (12) as follows:

```
R> beets$bh <- with(beets, interaction(block, harvest))
R> summary(aov(sugpct ~ block + sow + harvest + Error(bh), data=beets))

Error: bh
          Df  Sum Sq Mean Sq F value Pr(>F)
block      2 0.03267 0.01633   2.579 0.2794
harvest    1 0.09633 0.09633  15.211 0.0599 .
Residuals  2 0.01267 0.00633
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Within
          Df Sum Sq Mean Sq F value   Pr(>F)
sow        4   1.01  0.2525     101 5.74e-13 ***
Residuals 20   0.05  0.0025
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hence, when the hierarchical structure of the experiment has been accounted for, the effect of harvesting time is not significant at the 5% level.

3.4. The Mississippi influents example

The `Mississippi` dataset in the **SASmixed** package, Bates (2011b) contains the nitrogen concentration (in PPM) from several sites at six randomly selected influents of the Mississippi river.

```
R> data(Mississippi, package="SASmixed")
R> Mississippi$influent <- factor(Mississippi$influent)
R> Mississippi$Type <- factor(Mississippi$Type)
R> head(Mississippi)
```

```
  influent  y Type
1        1 21    2
2        1 27    2
3        1 29    2
4        1 17    2
5        1 19    2
6        1 12    2
```

The influents were characterized according to watersheds as follows. Type=1: No farmland in watershed (influents no. 3 and 5); Type=2: Less than 50% farmland in watershed (influents no. 1,2 and 4); Type=3: More than 50% farmland in watershed (influent no. 6). Measurements from the same influent are expected to be similar and there is no particular interest in the individual influents. It is more interesting to investigate the effect of the watershed type on the nitrogen concentration.

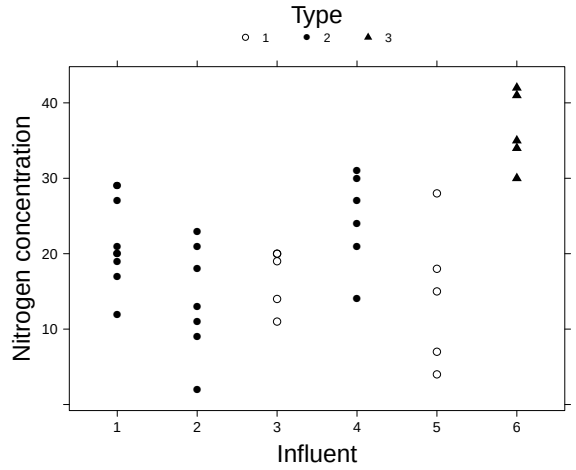

Figure 2: Nitrogen concentration in PPM at six different influents of the Mississippi differentiated for three types of watershed.

A typical model for such data would be

$$y_i = \alpha_{Type(i)} + U_{influent(i)} + \epsilon_i$$

where $U_l \sim N(0, \omega^2)$ and $\epsilon_i \sim N(0, \sigma^2)$. The χ^2–test suggests that the effect of `Type` is highly significant:

```
R> miss1 <- lmer(y ~ Type + (1|influent), data=Mississippi, REML=FALSE)
R> miss0 <- update(miss1, .~. - Type)
R> anova(miss1, miss0)
```

```
Data: Mississippi
Models:
miss0: y ~ (1 | influent)
miss1: y ~ Type + (1 | influent)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
miss0  3 262.56 267.39 -128.28
miss1  5 256.57 264.63 -123.29 9.9834      2   0.006794 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Trusting large sample asymptotic results is questionable. If data had been balanced such there were the same number of influents for each watershed type and the same number of recordings for each influent, then we could have made a proper F–test along the lines of Section 3.1.
An alternative is clearly to analyze the means for each influent and this yields a much less clear indication of an effect of watershed type:

```
R> Miss.mean <- summaryBy(y ~ influent + Type, data=Mississippi, FUN=mean)
R> miss1_lm <- lm(y.mean~Type, data=Miss.mean)
R> anova(miss1_lm)

Analysis of Variance Table

Response: y.mean
          Df  Sum Sq Mean Sq F value  Pr(>F)
Type       2 298.276 149.138  7.0702 0.07322 .
Residuals  3  63.282  21.094
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4. Approximate F-statistic and the KR–approximation

In this section we describe first the K&R approach of testing the hypothesis $\mathbf{L}(\boldsymbol{\beta} - \boldsymbol{\beta}_H) = \mathbf{0}$ for a more general model than (1). We describe then the class of linear mixed models fitted with `lmer()` for which the `KRmodcomp()` of the package **pbkrtest** provides the K&R approach.

4.1. A multivariate normal model

K&R consider for $\mathbf{Y}$ the multivariate normal model

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}).$$

The covariance-matrix $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$ is assumed to be a function of M parameters collected in the vector $\boldsymbol{\gamma}$. We denote the REML estimates of these parameters with $\hat{\boldsymbol{\gamma}}$. The unbiased (Kackar and Harville 1984) REML estimate of $\boldsymbol{\beta}$ is then

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Phi}(\hat{\boldsymbol{\gamma}})\mathbf{X}^\top\boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}})^{-1}\mathbf{Y} \text{ with } \boldsymbol{\Phi}(\hat{\boldsymbol{\gamma}}) = \left(\mathbf{X}^\top\boldsymbol{\Sigma}(\hat{\boldsymbol{\gamma}})^{-1}\mathbf{X}\right)^{-1}. \tag{13}$$

Here, $\boldsymbol{\Phi}$ is the covariance matrix of the asymptotic distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}(\hat{\boldsymbol{\gamma}})$ is a consistent estimate of this covariance matrix.
A scaled Wald-type statistics of testing the hypothesis $\mathbf{L}(\boldsymbol{\beta} - \boldsymbol{\beta}_H) = \mathbf{0}$ is

$$F = \frac{1}{d}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H)^\top\mathbf{L}^\top(\mathbf{L}^\top\hat{\mathbf{V}}\mathbf{L})^{-1}\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H) \tag{14}$$

where $\hat{\mathbf{V}}$ is some positive definite symmetric matrix.

Notice that the usual Wald test statistic uses $\hat{\mathbf{V}} = \boldsymbol{\Phi}(\hat{\boldsymbol{\gamma}})$. In this case F has asymptotically a $\frac{1}{d}\chi^2_d$ distribution (which can be thought of as the limiting distribution of an $F_{d,m}$–distribution when $m \to \infty$.) For some models, F has an exact F–distribution under the hypothesis. One example of this is a balanced one–way analysis of variance.

4.2. The approach of Kenward and Roger

K&R modify the statistic F in (14) to improve the small sample properties by approximating the distribution of F by an F–distribution, and they also provide a method for calculating the denominator degrees of freedom. The fundamental idea is to calculate the approximate mean and variance of their statistic and then match moments with an F–distribution to obtain the denominator degrees of freedom. K&R left out some detail in the derivation of their method. Alnosaier (2007) provides more details, weakens some of the assumptions for the approach, and extends the list of models for which it is known that the approach yields exact F-tests.

K&R take two steps to improve the small sample distributional properties of F. Firstly, Kackar and Harville (1984) showed that the covariance matrix of $\hat{\boldsymbol{\beta}}$ can be written as the sum $\mathbb{V}\mathrm{ar}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Phi} + \boldsymbol{\Lambda}$ where $\boldsymbol{\Lambda}$ expresses the bias by which the asymptotic covariance matrix $\boldsymbol{\Phi}$ underestimates $\mathbb{V}\mathrm{ar}(\hat{\boldsymbol{\beta}})$. K&R combine a Taylor approximation to $\boldsymbol{\Lambda}$ with a biased corrected modification of $\boldsymbol{\Phi}(\hat{\boldsymbol{\gamma}})$ using second order Taylor expansion to derive a new estimate $\boldsymbol{\Phi}_A(\hat{\boldsymbol{\gamma}})$. In the statistic F in (14), K&R replace the matrix $\hat{\mathbf{V}}$ with $\hat{\mathbf{V}} = \boldsymbol{\Phi}_A(\hat{\boldsymbol{\gamma}})$.

Secondly, K&R derive a scaling factor λ (such that the statistic they consider is λF) and a denominator degree of freedom m by matching approximations of the expectation and variance of λF with the moments of a $F_{d,m}$ distribution. In more detail, K&R derive an approximation for the expectation $E^\star$ and variance $V^\star$ based on a first order Taylor expansion of F. Then they solve the system of equations

$$\mathbb{E}(F) \approx \lambda E^\star = \mathbb{E}(F_{d,m}) = \frac{m}{m-2}, \tag{15}$$

$$\mathbb{V}\mathrm{ar}(F) \approx \lambda^2 V^\star = \mathbb{V}\mathrm{ar}(F_{d,m}) = \frac{2m^2(d+m-2)}{d(m-2)^2(m-4)} = \{\mathbb{E}(F_{d,m})\}^2 \frac{2(d+m-2)}{d(m-4)}, \tag{16}$$

where $\mathbb{E}(F_{d,m})$ and $\mathbb{V}\mathrm{ar}(F_{d,m})$ denote expectation and variance of a $F_{d,m}$–distributed random variable. The $E^\star$ and $V^\star$ are slightly modified without changing the order of approximation such that for the balanced one-way anova model and the Hoteling's T^2 model the exact F-tests are reproduced (Alnosaier 2007, Chapters 4.1, 4.2). We shall refer to these two steps as *the Kenward–Roger approximation* (or K&R–approximation in short). The details of the computations are provided in Appendix A.1. In particular, the solution to the equations above is given in (27). Recall that the mean of a $F_{d,m}$ distribution exists provided that $m > 2$ and the variance exists provided that $m > 4$. The moment matching method does however not prevent estimates of m that are smaller or equal to 2. K&R did not address this problem and we did neither in our implementation.

4.3. Models for which `KRmodcomp()` provides tests

The `KRmodcomp()` function of the **pbkrtest** package provides the KR–approximation for linear

mixed models of the form (1) where $\mathbf{\Sigma}$ is a sum of know matrices

$$\mathbf{\Sigma} = \sum_r \gamma_r \mathbf{G}_r^{N\times N} + \sigma^2 \mathbf{I}^{N\times N}. \tag{17}$$

The matrices $\mathbf{G}_r$ are usually very sparse matrices. Variance component models and random coefficient models are models which have this simplified covariance structure. For details we refer to Appendix A.1.

5. Parametric bootstrap

An alternative approach is based on parametric bootstrap, and this is also implemented in **pbkrtest**. The setting is the LR test statistic T for which we have an observed value t_{obs}. The question is in which reference distribution t_{obs} should be evaluated; i.e., what is the null–distribution of T. Instead of relying on the approximation of the null–distribution by a χ^2_d distribution one can use parametric bootstrap:

First, create B (e.g., $B = 1000$) bootstrap samples $y^1, \ldots, y^B$ by simulating from $\hat{f}_0(y)$ (where $\hat{f}_0$ denotes the fitted distribution under the hypothesis). Next, calculate the corresponding values $T^* = \{t^1, \ldots, t^B\}$ of the LR test statistic. For what follows, let E^*_T and V^*_T denote sample mean and sample variance of T^*. These simulated values can then be regarded as samples from the null–distribution and these values can be used in different ways which are implemented in the `PBmodcomp()` function. The labels below refer to the output from `PBmodcomp()`, see Section 6:

`PBtest:` Direct calculation of tail probabilities: The values T^* provide an empirical null–distribution in which t_{obs} can be evaluated. Let $I(x)$ is an indicator function which is 1 if x is true and 0 otherwise. The p–value then becomes the tail probability in T^*, i.e.,

$$p = \frac{1}{B}\sum_{k=1}^{B} I(t^k \geq t_{obs}), \tag{18}$$

`PBkd:` Approximate null–distribution by a kernel density estimate. The p–value is then calculated from the kernel density estimate.

`Gamma:` Approximate the null–distribution by a gamma distribution with mean E^*_T and variance V^*_T.

`Bartlett:` Improving the LR test statistic by a Bartlett type correction: The LR test statistic T can be scaled to better match the χ^2_d distribution as

$$T_B = \frac{d}{E^*_T} T.$$

`F:` Approximate the null–distribution of T/d by an $F_{d,m}$ distribution with mean E^*/d. This yields a single equation for deriving m, namely $m = 2E^*_T/(E^*_T - d)$.

We shall make the following remarks to the quantities mentioned in the listing above (in Section 6 we also provide a graphical illustrations of these approaches):

1) Regarding `PBtest` and `PBkd` recall that the definition of a p–value for a composite hypothesis is (see e.g., Casella and Berger (2002), p. 397)

$$p = \sup_{\boldsymbol{\theta}} P_{\boldsymbol{\theta}}(T > t_{obs})$$

where the supremum is taken over all possible values $\boldsymbol{\theta} = (\boldsymbol{\beta}_0, \boldsymbol{\gamma})$ under the hypothesis. When this supremum can not be evaluated in practice it is often exploited that for large samples $P_{\boldsymbol{\theta}}$ is approximately the distribution function for a χ^2_d distribution which is independent of $\boldsymbol{\theta}$. Implicit in (18) is therefore a definition of a bootstrapped p–value to be $p = P_{\hat{\boldsymbol{\theta}}}(T > t_{obs})$ and then (18) is used for the calculation. Determining the tail of a distribution as in (18) by sampling requires a large number of samples B (but how large B must be depends in practice obviously on the size of t_{obs}). An alternative is to provide a smooth estimate of the distribution of the null–distribution by, for example, a kernel density estimate (see e.g., Silverman and Young (1987))

2) The quantities, `Gamma`, `Bartlett` and `F` are based on assuming a parametric form of the null distribution such that the null distribution can be determined from at most the first two sample moments of T^*. It requires in general fewer samples to obtain credible estimates for these moments than for obtaining the tail probabilities in (18). We have no compelling mathematics argument why T^* should be well approximated by a Gamma distribution but simulations suggests this to be the case. Moreover, since a χ^2_d distribution is also a Gamma distribution, it is appealing to approximate T^* by a Gamma distribution where we match the first two moments. In practice this means that we obtain a distribution with heavier tail than the χ^2_d distribution. The idea behind adjusting the LR test statistic by a Bartlett type correction as in $T_B = \frac{T}{E^*_T/d}$ is to obtain a a statistic whose distribution becomes closer to a χ^2_d distribution, cfr. Cox (2006), p. 130. See also e.g., Jensen (1993) for a more comprehensive treatment of Bartlett corrections. Approximating the distribution of T/d by an $F_{d,m}$ distribution can be motivated as follows: Under the hypothesis, T is in the limit χ^2_d distributed so T/d has in the limit a χ^2_d/d distribution with expectation 1 and variance $2/d$. This is, loosely speaking, the same as an $F_{d,m}$ distribution with an infinite number of denominator degrees of freedom m. By estimating m as $m = 2E^*_T/(E^*_T - d)$ we obtain the increased flexibility of an F distribution with a larger variance than $2/d$, i.e., a distribution with a heavier tail than that of a χ^2_d/d distribution.

3) Lastly, a general problem with the parametric bootstrap approach is that it is computationally intensive (so is the K&R–approximation, too). However the **pbkrtest** package allows for the samples to be drawn in parallel by utilizing several processors on the computer.

6. Applications of the methods

This section contains applications of the methods described in Section 4 and Section 5 to the examples in Section 3. This section also contains additional examples.

6.1. The sugar beets example

For the sugar beets example of Section 3.1, the K&R–approximation provides the following results. The test for the harvest time yields

```
R> (kr.h <- KRmodcomp(beet0, beet_no.harv))

F-test with Kenward-Roger approximation; computing time: 0.06 sec.
Large : sugpct ~ block + sow + harvest + (1 | block:harvest)
small : sugpct ~ block + sow + (1 | block:harvest)
    Fstat df1      df2   p.value F.scaling
 15.21048   1 2.000003 0.0598979         1
```

and for the effect of sow time one gets

```
R> (kr.s <- KRmodcomp(beet0, beet_no.sow))

F-test with Kenward-Roger approximation; computing time: 0.06 sec.
Large : sugpct ~ block + sow + harvest + (1 | block:harvest)
small : sugpct ~ block + harvest + (1 | block:harvest)
 Fstat df1 df2 p.value F.scaling
   101   4  20       0         1
```

Similarly, for parametric bootstrap we obtain:

```
R> library("pbkrtest")
R> NSIM <- 200
R> (pb.h <- PBmodcomp(beet0, beet_no.harv, nsim=NSIM))

Parametric bootstrap test; bootstrap samples: 138 computing time: 7.24 sec.
large : sugpct ~ block + sow + harvest + (1 | block:harvest)
small : sugpct ~ block + sow + (1 | block:harvest)
              stat df   p.value   ddf
LRT      12.913822  1 0.0003262    NA
PBtest   12.913822 NA 0.0289855    NA
PBkd     12.913822 NA 0.0284624    NA
Gamma    12.913822 NA 0.0299234    NA
Bartlett  3.344713  1 0.0674212    NA
F        12.913822  1 0.0437875 2.699

R> (pb.s <- PBmodcomp(beet0, beet_no.sow, nsim=NSIM))

Parametric bootstrap test; bootstrap samples: 200 computing time: 2.70 sec.
large : sugpct ~ block + sow + harvest + (1 | block:harvest)
small : sugpct ~ block + harvest + (1 | block:harvest)
             stat df   p.value   ddf
LRT      85.20220  4 0.0000000    NA
PBtest   85.20220 NA 0.0000000    NA
PBkd     85.20220 NA 0.0000000    NA
Gamma    85.20220 NA 0.0000000    NA
Bartlett 62.24384  4 0.0000000    NA
F        21.30055  4 0.0003773 7.422
```

First, it is noted that the p–values reported from both `KRmodcomp()` and `PBmodcomp()` generally are 1) within the same order of magnitude and 2) close to the results of the exact F–test of Section 3.1. Hence the results would all suggest the same qualitative conclusion, namely that there is little (if any) evidence for an effect of harvest time and strong evidence for an effect of sowing time. Secondly, it is noticed that `KRmodcomp()` is much faster than `PBmodcomp()` in these examples. However the difference in computing time is much smaller for other types of models / datasets; for example for certain random regression models (not reported in this paper).

It is illustrative to look at a graphical representation of the results `PBmodcomp()` for a large number of bootstrap samples. We draw 5000 bootstrap samples for testing of no effect of harvest time for the sugar beets example as follows:

```
R> pb.h <- PBmodcomp(beet0, beet_no.harv, nsim=5000)

Parametric bootstrap test; bootstrap samples: 2664 computing time: 121.26 sec.
large : sugpct ~ block + sow + harvest + (1 | block:harvest)
small : sugpct ~ block + sow + (1 | block:harvest)
              stat df   p.value   ddf
LRT      12.913822  1 0.0003262    NA
PBtest   12.913822 NA 0.0581832    NA
PBkd     12.913822 NA 0.0614923    NA
Gamma    12.913822 NA 0.0543317    NA
Bartlett  2.905237  1 0.0882923    NA
F        12.913822  1 0.0470218 2.581
```

The output shows that much fewer samples than 5000 are available. This is because `lmer()` often fails to refit models based on simulated data. Figure 3 shows the histogram of the simulated values for the null-distribution. The solid vertical line to the right is the observed value of the LR test statistic (which is also the F–statistic as there is only $d = 1$ degree of freedom). The dashed vertical line to the left is the Bartlett corrected statistic. The various estimated densities are overlaid the histogram. Of the parametric densities, the best overall fit to the simulated null-distribution is provided by the Gamma distribution. The χ^2 distribution approximates the simulated null-distribution very poorly. The kernel density, the Gamma and the F distributions all provide estimates of the tail probabilities that are similar to each other and to the p–value of the proper F–test provided in Section 3.

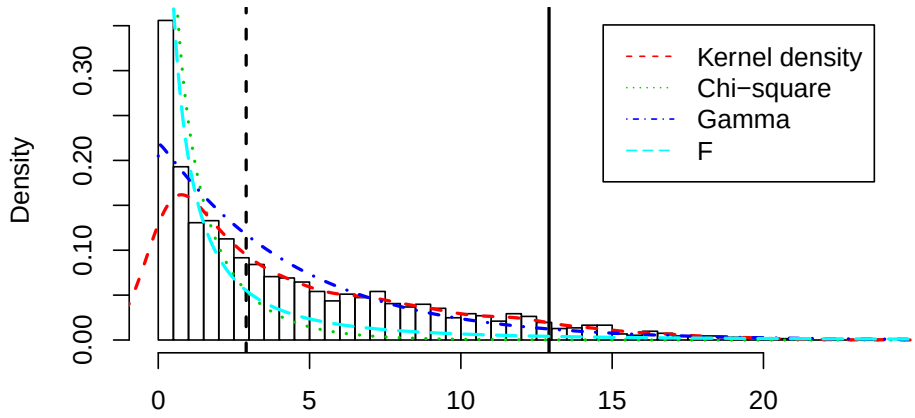

Figure 3: Histogram of simulated null-distribution for testing of no effect of harvest time for the sugar beets example with approximately 2500 samples. The solid vertical line to the right is the observed value of the LR test statistic. The dashed vertical line to the left is the Bartlett corrected statistic. The various estimated densities are overlaid the histogram.

Figure 4 shows the same as Figure 3 but with much fewer samples. It is much less clear which of the parametric densities fit best to the samples. However the p–values are similar to those found above with many samples and they certainly suggest the same qualitative conclusion: there is no effect of harvesting time.

```
R> pb.h <- PBmodcomp(beet0, beet_no.harv, nsim=100)

Parametric bootstrap test; bootstrap samples: 48 computing time: 2.64 sec.
large : sugpct ~ block + sow + harvest + (1 | block:harvest)
```

```
small : sugpct ~ block + sow + (1 | block:harvest)
              stat df   p.value   ddf
LRT      12.913822  1 0.0003262    NA
PBtest   12.913822 NA 0.0625000    NA
PBkd     12.913822 NA 0.0902998    NA
Gamma    12.913822 NA 0.0689132    NA
Bartlett  2.807526  1 0.0938230    NA
F        12.913822  1 0.0477480 2.556
```

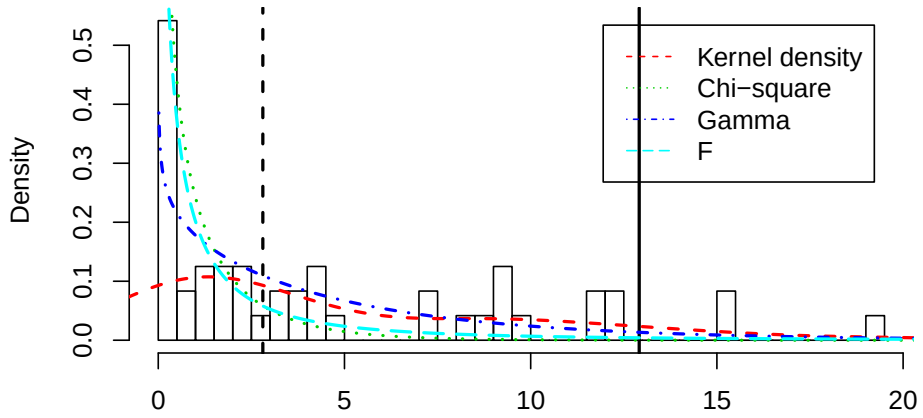

Figure 4: Histogram of simulated null-distribution for testing of no effect of harvest time for the sugar beets example with approximately 50 samples. The solid vertical line to the right is the observed value of the LR test statistic. The dashed vertical line to the left is the Bartlett corrected statistic. The various estimated densities are overlaid the histogram.

6.2. The Mississippi influents example

For the `Mississippi` data of Section 3.4 our methods provide the following results:

```
R> KRmodcomp(miss1, miss0)

F-test with Kenward-Roger approximation; computing time: 0.06 sec.
Large : y ~ Type + (1 | influent)
small : y ~ (1 | influent)
    Fstat df1      df2   p.value F.scaling
 6.368942   2 3.320734 0.0730335 0.9996716

R> PBmodcomp(miss1, miss0)

Parametric bootstrap test; bootstrap samples: 200 computing time: 6.90 sec.
large : y ~ Type + (1 | influent)
small : y ~ (1 | influent)
              stat df   p.value   ddf
LRT      9.983400  2 0.0067941    NA
PBtest   9.983400 NA 0.1250000    NA
PBkd     9.983400 NA 0.1171705    NA
Gamma    9.983400 NA 0.0797071    NA
Bartlett 4.893715  2 0.0865652    NA
F        4.991700  2 0.0835540 3.923
```

Hence we obtain p–values which are in the order of 10 times the p–value provided by the χ^2 approximation. The p–values we obtain are in good accordance with the p–value obtained when analyzing the means as done in Section 3.4.

6.3. Random regression – A simulation

K&R perform a small simulation study on a simple random regression model. We made a simulation using the same model set–up and use it to compare the results between the different tests we provide and to the K&R approach as implemented by the MIXED procedure of the SAS software system,(SAS Institute Inc. 2008).

Kenward and Roger (1997, Table 4) consider the following random coefficient model

$$y_{jt} = \beta_0 + \beta_1 \cdot t_j + A_j + B_j \cdot t_j + \epsilon_{jt}$$

with

$$\mathbb{C}\text{ov}(A_j, B_j) = \begin{bmatrix} 0.250 & -0.133 \\ -0.133 & 0.250 \end{bmatrix} \text{ and } \mathbb{V}\text{ar}(\epsilon_{jt}) = 0.25. \tag{19}$$

There are observed $j = 1, \ldots, 24$ subjects divided into three groups of eight subjects. For each group observations are made at the non overlapping times $t = 0, 1, 2; t = 3, 4, 5$ and $t = 6, 7, 8$. The data for the simulation were generated under the assumption that $\beta_0 = \beta_1 = 0$, the (A_j, B_j) and ϵ_{jt} are normally distributed with zero expectation, (A_j, B_j) are independent from ϵ_{tj} and observations from different subjects are independent.

The full model and the reduced models are fitted by

```
R> mod1 <- lmer(y ~ A + t + (1+t|subject))
R> mod_no.int   <- lmer(y~0 + t + (1+t|subject))
R> mod_no.slope <- lmer(y~A  + (1+t|subject))
```

Parm	$\alpha \times 100$	LR	KR(R)	KR(SAS)	PBtest	PBkd	Bartlett	Gamma	F
β_1	1	1.6	0.7	1.4	1.4	1.4	1.3	1.5	0.8
β_2	1	1.8	1.2	1.0	1.3	1.4	1.2	1.4	0.8
β_1	5	7.0	4.3	5.2	6.1	6.0	5.8	6.2	5.2
β_2	5	6.9	5.5	5.1	5.5	5.5	5.3	5.6	5.1
β_1	10	13.5	9.4	10.0	11.7	11.6	11.6	11.9	11.3
β_2	10	12.8	10.7	10.0	10.5	10.4	10.4	10.5	10.6

Table 1: Observed test sizes ($\times 100$) for three test levels $\alpha = 0.01, 0.05, 0.1$ for $H_0 : \beta_k = 0$ from the random coefficient model. The results are based on 20000 simulations, for the bootstrapped p-values 500 subsamples were taken. KR(R) and KR(SAS) are the K&R approximations as implemented in `KRmodcomp()` and in SAS, the other results refer to the null-distribution of the log-likelihood ratio test-statistic, either the χ^2 appoximation (LR) or bootstrapped values. `PBtest` relates to the raw parametric bootstrap p-value. The other p-values are based on approximations to the bootstrap distribution either via a kernel-density, a Bartlett correction, a Gamma or a F distribution.

The likelihood-ratio test is for both parameters and for all α's anti-conservative as expected. For all other approaches the observed test-levels are closer to the nominal levels and in most

cases anti-conservative. The Kenward-Roger approach from our implementation yields conservative results for the tests on the intercept parameter β_1 and the test in columns `F` yields conservative results for the lowest nominal level. The difference of the results of the Kenward-Roger approach between our implementation and that of SAS may lie in the different treatment of cases in which the covariance matrix $\mathbf{\Gamma}$ is singular.

7. Discussion

In this paper we have presented our implementation of a K&R–approximation for tests in linear mixed models. In the implementation, there are several matrices of the order $N \times N$ where N is the number of observations. We have exploited that several of the matrices involved in the computations will in many cases will be sparse via the facilities in the **Matrix** package. Nonetheless, the current implementation of the K&R–approximation does not always scale to large datasets. As an example, consider a repeated measurement problem in which repeated measurements are made on a collection of subjects. It there are many subjects and the time series for each subject is short then there is a sparseness to be exploited. On the other hand, if there are a few long time series then the matrices involved will have a non–negligible number on non–zero elements. One approach to speed up the computations is to compute the average of the observed and expected information matrices rather than the expected information matrix. This can lead to substantial improvements in computing time because some of the computationally most intractable terms vanish in the average information. See Gilmour, Thompson, and Cullis (1995) and Jensen, Mantysaari, Madsen, and Thompson (1996) for details. This may become available in later versions of **pbkrtest**. A very specific issue which we have no clear answer to is how the K&R–approximation should be modified in case of a singular estimate of the covariance matrix.

Contrary to the K&R–approximation, the parametric bootstrap approach has the advantage that it is easy to implement; all that is required is a way of sampling data from fitted model under the hypothesis. Furthermore the parametric bootstrap approach is straight forward to implement for other types of problems, for example for logistic regression and other types of generalized linear models and for generalized linear mixed models. A problem with the parametric bootstrap approaches is the randomness of the results; repeated applications to the same dataset does not give entirely identical results. Moreover, calculating the reference distribution by sampling is computationally demanding. However, **pbkrtest** implements the possibility of parallel computing of the reference distribution using multiple processors via the **snow** package. There are various possibilities for speeding up the parametric bootstrap computations: 1) The Bartlett type correction we implemented is such a possibility because the correction depends only on the mean of the simulated null–distribution and the Gamma approximation depends only on the mean and variance of the simulated null–distribution. Estimating these two moments will in general require fewer simulations than estimating the tail of the the null–distribution. Hence, if one chooses to focus on these two distributions then one may get credible results with few samples. 2) It may also be possible to devise a sequential sampling scheme such that sampling stops when the estimates of the first or the first two moments have stabilized. 3) Instead of fixing the number of parametric bootstrap samples B in advance, on may instead continue to draw samples until h (e.g., $h = 20$) values of the test statistic which are more extreme than the observed test statistic have been obtained. If this takes B' samples then the p–value to report is h/B'. If there is little evidence against

the hypothesis then only a small number B' of simulations would be needed. This idea is the parametric bootstrap version of the approach of Besag and Clifford (1991) for calculating sequential Monte Carlo p–values.

An important final comment is that we do not in any way claim that we have found an omnibus panacea solution to a difficult problem. Instead we have provided two practically applicable alternatives to relying on large sample asymptotics when testing for the reduction of the mean value in general linear mixed models.

8. Acknowledgements

This work was supported by the Danish National Advanced Technology Foundation through the ILSORM project.

References

Alnosaier WS (2007). "Kenward-Roger Approximate F Test for Fixed Effects in Mixed Linear Models." *Phd dissertation*, Oregon State University. Pages 132, URL http://ir.library.oregonstate.edu/xmlui/bitstream/handle/1957/5262/mydissertation.pdf?sequence=1.

Bates D (2011a). "Linear Mixed Model Implementation in **lme4**." *Vignette called lme4::implementation of the **lme4** package of R*, University of Wisconsin - Madison.

Bates D (2011b). ***SASmixed**: Data sets from "SAS System for Mixed Models"*. R package version 1.0-1, URL http://CRAN.R-project.org/package=SASmixed.

Bates D, Maechler M, Bolker B (2011). ***lme4**: Linear mixed-effects models using S4 classes*. R package version 0.999375-42, URL http://CRAN.R-project.org/package=lme4.

Besag J, Clifford P (1991). "Sequential Monte Carlo p–values." *Biometrika*, **78**(2), 301–304.

Casella G, Berger RL (2002). *Statistical Inference*. 2 edition. Duxbury.

Cochran WG, Cox GM (1957). *Experimental Design*. 2nd edition. Chapman and Hall.

Cox DR (2006). *Principles of Statistical Inference*. Cambridge.

Gilmour AR, Thompson R, Cullis BR (1995). "An Efficient Algorithm for REML Estimation in Linear Mixed Models." *Biometrics*, **51**, 1440–1450.

Halekoh U, Højsgaard S (2012). ***pbkrtest**: Parametric bootstrap and Kenward Roger based methods for mixed model comparison*. R package version 0.3.0.

Harville DA (1997). *Matrix Algebra from a Statistician's Perspective*. Springer.

Harville DA, Hinkley DV (1997). *Bootstrap Methods and their Applications*. Cambridge University Press.

Højsgaard S, Halekoh U (2012). ***doBy:*** *doBy - Groupwise summary statistics, general linear contrasts, population means (least-squares-means), and other utilities.* R package version 4.5.0, URL http://CRAN.R-project.org/package=doBy.

Jensen J, Mantysaari EA, Madsen P, Thompson R (1996). "Residual Maximum Likelihood Estimation of (Co)Variance Components in Multivariate Mixed Linear Models using Average Information." *Jour. Ind. Soc. Ag. Statistics*, **49**, 215–236.

Jensen JL (1993). "A Historical Sketch and Some New Results on the Improved Log Likelihood Ratio Statistic." *Scandinavian Journal of Statistics*, **20**(1), 1–15.

Kackar RN, Harville DA (1984). "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models." *Journal of the American Statistical Association*, **79**, 853–862.

Kenward MG, Roger JH (1997). "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood." *Biometrics*, **53**(3), 983–997.

Laird N, Ware J (1982). "Random-Effects Models for Longitudinal Data." *Biometrics*, **38**, 963–974.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

SAS Institute Inc (2008). *SAS/STAT Software, Version 9.2.* Cary, NC. URL http://www.sas.com/.

Silverman BW, Young GA (1987). "The bootstrap: To Smooth or not to Smooth?" *Biometrika*, **74**(3), 469–479.

A. Technical details for the KR–approximation

A.1. Computations related to the KR–approximation

In this appendix more details of the implementation of the approach of K&R in `KRmodcomp()` is given. First we describe the structure of the design matrix of the random effects $\mathbf{Z}$ and the related structure of the covariance matrix $\mathbf{\Gamma}$. Secondly, the sequence of computations with the matrices available from a fitted model object from `lmer()` and the derived matrices are given.

Structure for $\mathbf{Z}$ *and* $\mathbf{\Gamma}$

The description of the structure of $\mathbf{Z}$ and $\mathbf{\Gamma}$ draws heavily on the description given in a vignette of the **lme4** package (Bates 2011a).

For a linear mixed model fitted with `lmer()` it is assumed that we have $i = 1, \ldots f$ grouping factors denoted by $\mathbf{f}_i$. It is allowed that $\mathbf{f}_i = \mathbf{f}_{i'}$ for $i \neq i'$. The ith grouping factor $\mathbf{f}_i$ has g_i levels and there are q_i random effects for each level. The random effects for group level j are collected in the vector $\mathbf{b}_{ij} = (b_{ij1}, \ldots b_{ijq_i})^\top$ and the random effects of $\mathbf{f}_i$ are $\mathbf{b}_i^\top = (\mathbf{b}_{ij}^\top)$.

It is assumed that the random effects from different grouping factors and from different levels of a grouping factor are independent, i.e.

$$\mathbb{C}\mathrm{ov}(\mathbf{b}_i, \mathbf{b}_{i'}) = 0 \text{ for } i \neq i' \text{ and } \mathbb{C}\mathrm{ov}(\mathbf{b}_{ij}, \mathbf{b}_{ij'}) = 0 \text{ for } j \neq j'.$$

The covariance matrix of the random effects for grouping level j of factor $\mathbf{f}_i$ is independent of the grouping level and is denoted by

$$\mathbb{V}\mathrm{ar}(\mathbf{b}_{ij}) = \mathbf{\Gamma}_i^{q_i \times q_i} = (\gamma_{i;rr'}).$$

We assume that all of the elements of $\mathbf{\Gamma}_i$ are parameters that vary freely except that $\mathbf{\Gamma}_i$ must be positive definite. Hence $\mathbb{V}\mathrm{ar}(\mathbf{b}_i) = \mathbf{I}^{g_i \times g_i} \otimes \mathbf{\Gamma}_i$ where $\mathbf{I}^{g_i \times g_i}$ the identity matrix of dimension g_i.

For the sugar beet example there is one factor, the interaction $U_{i'j'}$ between block and harvest. In the present notation $f = 1, g_1 = 6, q_1 = 1$, $\mathbf{b}_1 = (b_{1,1}, \ldots b_{1,6})^\top$ and $\mathbf{Z}_1 = \mathbf{I}^6 \otimes \mathbf{1}^5$ where $\mathbf{1}^5$ is a vector of one's.

For the random coefficient model of the simulation example there is one grouping factor, subject, with 24 levels, hence $f = 1, g_1 = 24$ and $q_1 = 2$ random effects (A_j, B_j) for subject j such that $\mathbf{b}_1 = (A_1, B_1, \ldots A_{24}, B_{24})$,

$$\mathbf{Z}_1^{72\times 48} = \begin{bmatrix} 1 & 0 & & & & \\ 1 & 1 & & & & \\ 1 & 2 & & & & \\ & & \cdots & & & \\ & & & & 1 & 6 \\ & & & & 1 & 7 \\ & & & & 1 & 8 \end{bmatrix} \tag{20}$$

and $\mathbf{\Gamma}_1$ is the matrix in (19). If in the simulation example the two random effects A_j and B_j were assumed to be uncorrelated the model would be specified with `lmer()` as `y~lmer(y~A + t + (1|subject) + (0+t|subject))`. Now there are two grouping factors, both are equal to subject, hence $f = 2, g_1 = g_2 = 24, q_1 = q_2 = 1$, $\mathbf{b}_1 = (A_1, \ldots A_{24})^\top$, $\mathbf{b}_2 = (B_1, \ldots B_{24})^\top$ and

$$\mathbf{Z}_1^{72\times 24} = \begin{bmatrix} 1 & & \\ 1 & & \\ 1 & & \\ & \cdots & \\ & & 1 \\ & & 1 \\ & & 1 \end{bmatrix}, \quad \mathbf{Z}_2^{72\times 24} = \begin{bmatrix} 0 & & \\ 1 & & \\ 2 & & \\ & \cdots & \\ & & 6 \\ & & 7 \\ & & 8 \end{bmatrix}. \tag{21}$$

Let $\boldsymbol{\gamma}_i = (\gamma_{i;11}, \gamma_{i;2,1}, \ldots, \gamma_{i;q_i1}, \gamma_{i;22}, \ldots, \gamma_{i;q_iq_i})^\top$ denote the $s_i = q_i(q_i+1)/2$ vector of the elements of the lower triangular $\mathbf{\Gamma}_i$. For the kth element $\gamma_{i;k}$ of $\boldsymbol{\gamma}_i$ it holds that $\gamma_{i;k} = \gamma_{i;rr'}$ where $k = (r-1) * (q_i - r/2) + r'$. Then we may write

$$\mathbf{\Gamma}_i = \sum_{k=1}^{s_i} \gamma_{i;k} \mathbf{E}_{i;k}.$$

The $\mathbf{E}_{i;k}$ are the $q_i \times q_i$ symmetric incidence matrices with ones at the position (r, r') and (r', r). Now,

$$\begin{aligned}\mathbb{V}\mathrm{ar}(\mathbf{Z}_i\mathbf{b}_i) &= \mathbf{Z}_i\,\mathbb{V}\mathrm{ar}(\mathbf{b}_i)\mathbf{Z}_i^\top = \mathbf{Z}_i(\mathbf{I}^{g_i\times g_i}\otimes\boldsymbol{\Gamma}_i)\mathbf{Z}_i^\top \\ &= \mathbf{Z}_i(\mathbf{I}^{g_i\times g_i}\otimes\sum_{k=1}^{s_i}\gamma_{i;k}\mathbf{E}_{i;k})\mathbf{Z}_i^\top = \sum_{k=1}^{s_i}\gamma_{i;k}\mathbf{Z}_i(\mathbf{I}^{g_i\times g_i}\otimes\mathbf{E}_{i;k})\mathbf{Z}_i^\top.\end{aligned}$$

With $\mathbf{D}_i = \sum_{k=1}^{s_i}\gamma_{i:k}\mathbf{Z}_i(\mathbf{I}^{g_i\times g_i}\otimes\mathbf{E}_{i;k})\mathbf{Z}_i^\top$ the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{Y}$ is

$$\boldsymbol{\Sigma} = \sum_{i=1}^{f}\mathbb{V}\mathrm{ar}(\mathbf{Z}_i\mathbf{b}_i) + \mathbb{V}\mathrm{ar}\,(\boldsymbol{\epsilon}) = \sum_{i=1}^{f}\mathbf{D}_i + \sigma^2\mathbf{I}^{N\times N} \tag{22}$$

where f is the number of grouping factors. Let $\boldsymbol{\gamma}$ denote the vector of length M made by concatenation of the vectors $\boldsymbol{\gamma}_i$ and, as the last element, the σ^2. Let $\mathbf{G}_r = \mathbf{Z}_i(\mathbf{I}^{g_i\times g_i}\otimes\mathbf{E}_{i;r})\mathbf{Z}_i^\top$ where r refers to the rth element in $\boldsymbol{\gamma}$ and i is the group factor $\mathbf{f}_i$ related to the covariance parameter γ_r. Note that $\mathbf{G}_M = \mathbf{I}^{N\times N}$.

Then $\boldsymbol{\Sigma}$ can be written as a linear combination of known matrices

$$\boldsymbol{\Sigma} = \sum_{r=1}^{M}\gamma_r\mathbf{G}_r. \tag{23}$$

For the sugar beets example $\mathbf{G}_1 = \mathbf{I}^{6\times6}\otimes\mathbf{J}^{5\times5}$ where $\mathbf{J} = \mathbf{1}^5\mathbf{1}^{5\top}$. For the simulation example $\mathbf{G}_1 = \mathbf{I}^{24\times24}\otimes\mathbf{J}^3$ is related to $\gamma_1 = 0.25$ and $\mathbf{G}_2$ is related to the covariance $\gamma_2 = -0.133$, with $\mathbf{G}_2 = \mathrm{diag}(\mathbf{I}^{3\times3}\otimes\mathbf{A}, \mathbf{I}^{3\times3}\otimes\mathbf{B}, \mathbf{I}^{3\times3}\otimes\mathbf{C})$ and

$$\mathbf{A} = \begin{bmatrix}0&1&2\\1&2&3\\2&3&4\end{bmatrix},\quad \mathbf{B} = \begin{bmatrix}6&7&8\\7&8&9\\8&9&10\end{bmatrix},\quad \mathbf{C} = \begin{bmatrix}12&13&14\\13&14&15\\14&15&16\end{bmatrix}. \tag{24}$$

The representation (23) has two simplifying consequences. Firstly, the derivative of $\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\gamma}$ is (see e.g. Harville (1997, equation (8.15)))

$$\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\gamma_r} = -\boldsymbol{\Sigma}^{-1}\mathbf{G}_r\boldsymbol{\Sigma}^{-1}.$$

Secondly, the estimate of the covariance matrix of $\hat{\boldsymbol{\beta}}$ can be expressed without using higher derivatives of $\boldsymbol{\Sigma}^{-1}$ (cf. Kenward and Roger (1997, equation (5)).

Implementation of the K&R approach in the `KRmodcomp()` *function*

The following estimates are directly provided by `lmer()`: 1) The parameter estimate $\hat{\boldsymbol{\beta}}$, 2) The vector $\hat{\boldsymbol{\gamma}}$ of the REML estimated covariance parameters and 3) The estimate $\boldsymbol{\Phi}(\hat{\boldsymbol{\gamma}})$ of the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$.

The estimate of the covariance matrix for $\hat{\boldsymbol{\gamma}}$

$$\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\gamma}}) = \mathbf{W}^{M\times M}$$

is not directly available from `lmer()`, but is estimated in (26) from the inverse information matrix, (cf. also Kenward and Roger (1997, equations (4) and (5))).

The implementation of the K&R–approximation in **pbkrtest** is based on the following quantities.

1. For each covariance parameter γ_r in $\boldsymbol{\gamma}$ we use

$$\mathbf{G}_r^{N\times N} = \mathbf{Z}_i(\mathbf{I}^{g_i\times g_i} \otimes \mathbf{E}_r)\mathbf{Z}_i^\top, \tag{25}$$

where i refers to the group for the covariance parameter γ_r.

2. Then the estimated covariance matrix for $\mathbf{Y}$ becomes $\hat{\boldsymbol{\Sigma}} = \sum_r^M \hat{\gamma}_r \mathbf{G}_r$.

3. For the computations to follow, we define the following auxiliary matrices:

- $\mathbf{T}^{N\times p} = \boldsymbol{\Sigma}^{-1}\mathbf{X}$
- $\mathbf{H}_r^{N\times N} = \mathbf{G}_r\boldsymbol{\Sigma}^{-1}$, $r = 1, \ldots, M$
- $\mathbf{O}_r^{N\times p} = \mathbf{G}_r\boldsymbol{\Sigma}^{-1}\mathbf{X} = \mathbf{H}_r\mathbf{X}$, $r = 1, \ldots, M$
- $\boldsymbol{\Omega}_r^{N\times N} = \frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\gamma_r} = -\boldsymbol{\Sigma}^{-1}\mathbf{G}_r\boldsymbol{\Sigma}^{-1}$, $r = 1, \ldots, M$. Notice that $\boldsymbol{\Omega}_r$ is not used in any computation in the implementation in **pbkrtest** but $\boldsymbol{\Omega}_r$ appears in the derivations below.

4. For each covariance parameter γ_r let

$$\mathbf{P}_r^{p\times p} = \mathbf{X}^\top\boldsymbol{\Omega}_r\mathbf{X} = -\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{G}_r\boldsymbol{\Sigma}^{-1}\mathbf{X} = -\mathbf{T}^\top\mathbf{G}_r\mathbf{T} = -\mathbf{T}^\top\mathbf{O}_r.$$

5. For each pair (γ_r, γ_s) of covariance parameters let

$$\begin{aligned}\mathbf{Q}_{rs}^{p\times p} &= \mathbf{X}^\top\boldsymbol{\Omega}_r\hat{\boldsymbol{\Sigma}}\boldsymbol{\Omega}_s\mathbf{X} = \mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{G}_r\boldsymbol{\Sigma}^{-1}\mathbf{G}_s\boldsymbol{\Sigma}^{-1}\mathbf{X}\\ &= \mathbf{T}^\top\mathbf{G}_r\boldsymbol{\Sigma}^{-1}\mathbf{G}_s\mathbf{T} = \mathbf{O}_r^\top\boldsymbol{\Sigma}^{-1}\mathbf{O}_s.\end{aligned}$$

Notice that $\mathbf{Q}_{rs}$ is generally not symmetric but $\mathbf{Q}_{rs} = \mathbf{Q}_{sr}^\top$ and hence $\mathbf{Q}_{rs} + \mathbf{Q}_{sr}$ is symmetric. This symmetry property is exploited below. Moreover, $\mathrm{tr}(\mathbf{Q}_{rs}) = \mathrm{tr}(\mathbf{Q}_{sr})$.

6. For each pair (γ_r, γ_s) of covariance parameters let

$$\begin{aligned}K_{rs} &= \mathrm{tr}(\boldsymbol{\Omega}_r\boldsymbol{\Sigma}\boldsymbol{\Omega}_s\boldsymbol{\Sigma})\\ &= \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{G}_r\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{G}_s\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}) = \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{G}_r\boldsymbol{\Sigma}^{-1}\mathbf{G}_s).\end{aligned}$$

7. Twice the expected information matrix for $\hat{\boldsymbol{\gamma}}$ then becomes:

$$2\cdot\{\mathbf{I}_E\}_{rs} = K_{rs} - 2\cdot\mathrm{tr}(\boldsymbol{\Phi}\mathbf{Q}_{rs}) + \mathrm{tr}(\boldsymbol{\Phi}\mathbf{P}_r\boldsymbol{\Phi}\mathbf{P}_s).$$

Notice that $\mathrm{tr}(\boldsymbol{\Phi}\mathbf{Q}_{rs}) = \mathrm{tr}(\boldsymbol{\Phi}\mathbf{Q}_{sr})$ and $\mathrm{tr}(\boldsymbol{\Phi}\mathbf{P}_r\boldsymbol{\Phi}\mathbf{P}_s) = \mathrm{tr}(\boldsymbol{\Phi}\mathbf{P}_s\boldsymbol{\Phi}\mathbf{P}_r)$.

8. The asymptotic covariance matrix of the random effects parameters becomes

$$\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\gamma}}) = \mathbf{W}^{M\times M} = 2\cdot\mathbf{I}_E^{-1}. \tag{26}$$

9. Define

$$\begin{aligned}\mathbf{U}^{p\times p} &= \sum_{r=1}^M\sum_{s=1}^M W_{rs}(\mathbf{Q}_{rs} - \mathbf{P}_r\boldsymbol{\Phi}\mathbf{P}_s)\\ &= \sum_{1\le r<s\le M} W_{rs}(\mathbf{Q}_{rs} + \mathbf{Q}_{rs}^\top - \mathbf{P}_r\boldsymbol{\Phi}\mathbf{P}_s - \mathbf{P}_s\boldsymbol{\Phi}\mathbf{P}_r) + \sum_{r=1}^M W_{rr}(\mathbf{Q}_{rr} - \mathbf{P}_r\boldsymbol{\Phi}\mathbf{P}_r).\end{aligned}$$

Notice that the last equation holds because of $\mathbf{Q}_{sr} = \mathbf{Q}_{rs}^\top$. Letting $\tilde{\mathbf{U}} = \sum_{1\le r<s\le M} W_{rs}(\mathbf{Q}_{rs} - \mathbf{P}_r\boldsymbol{\Phi}\mathbf{P}_s)$, one can write alternatively

$$\mathbf{U} = \tilde{\mathbf{U}} + \tilde{\mathbf{U}}^\top + \sum_{r=1}^{M} W_{rr}(\mathbf{Q}_{rr} - \mathbf{P}_r\boldsymbol{\Phi}\mathbf{P}_r).$$

10. The adjusted estimate of $\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\beta}})$ is then

$$\hat{\boldsymbol{\Phi}}_A = \boldsymbol{\Phi}(\hat{\boldsymbol{\gamma}}) + 2 \cdot \hat{\boldsymbol{\Lambda}} \text{ where } \hat{\boldsymbol{\Lambda}}^{p\times p} = \hat{\boldsymbol{\Phi}}\mathbf{U}\hat{\boldsymbol{\Phi}},$$

and the adjusted test statistic is (where d is the rank of $\mathbf{L}$)

$$F = \frac{1}{d}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H)^\top\mathbf{L}^\top(\mathbf{L}\hat{\boldsymbol{\Phi}}_A\mathbf{L}^\top)^{-1}\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_H).$$

11. K&R derive a scaling factor λ for the F statistic given above (such that the statistic they finally propose is λF) and a denominator degrees of freedom m by matching approximate first and second moments of the λF statistic with the moments of a $F_{d,m}$ distribution. In this connection K&R use the following quantities:

    (a) $\boldsymbol{\Theta} = \mathbf{L}^\top(\mathbf{L}\boldsymbol{\Phi}\mathbf{L}^\top)^{-1}\mathbf{L}$

    (b) $A_1 = \sum_{r=1}^{M}\sum_{s=1}^{M} W_{rs}\,\mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi})\,\mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi})$ (where W_{rs} are the elements of the covariance matrix $\mathbf{W}$ from equation (26)).

    (c) With $\circ$ denoting the Hadamard product,

$$\begin{aligned} A_2 &= \sum_{s}^{M}\sum_{s}^{M} W_{rs}\,\mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi}) \\ &= \sum_{r}^{M}\sum_{s}^{M} W_{rs}\mathbf{1}^\top\Big[(\boldsymbol{\Phi}\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_i)\circ(\boldsymbol{\Phi}\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_j)\Big]\mathbf{1}. \end{aligned}$$

    (d) $B = \frac{1}{2d}(A_1 + 6A_2)$

    (e) $E^* = 1/(1 - \frac{A_2}{d})$

    (f) $V^\star = \frac{2}{d}\left(\frac{1+c_1B}{(1-c_2B)^2(1-c_3B)}\right)$. The c_is are simple functions of A_1, A_2 and d.

    (g) $\rho = V^*/(2[E^*]^2)$

    $E^\star$ and $V^\star$ are approximate expectation and variance of F based on the first order Taylor expansion of F.

12. Then K&R end up with the following values for m and λ:

$$m = 4 + \frac{d+2}{d\rho - 1} \text{ and } \lambda = \frac{m}{E^*(m-2)}. \tag{27}$$

A.2. Some numerical issues

In the computation of ρ we encountered numerical problems in the calculation of ρ for some models where the division of two numbers both equal to zero are encountered. One can write the ρ as

$$\rho = \frac{1}{2}\left(\frac{D}{V_1}\right)^2 \cdot \frac{V_0}{V_2}$$

where $V_0 = 1 + c_1 B$, $V_1 = 1 - c_2 B$, $V_2 = 1 - c_3 B$ and $D = 1 - A_2/d$. The V_1 and D can become simultaneously very small yielding an unreliable ratio D/V_1. We resolve this problem by setting the ratio to 1 if $\max(|D|, |V_1|) < 10^{-11}$.

For example, for a simple block design,

$$Y_{bt} = \mu + \alpha_t + \epsilon_b + \epsilon_{bt}, \quad b = 1, \ldots, n_b, \quad , t = 1, \ldots, n_t \tag{28}$$

one has for $n_t = 2$ and $n_b = 3$ or for $n_t = 3$ and $n_b = 2$ an exact F–test with $m = 2$ denominator degrees of freedom. We have for a specific application of this model found that D and V_1 were very close to zero. If we define the ratio to be 1 in this case then we end up with the the correct answer, i.e. with $m = 2$. For the same design but for $n_t = 2$ and $n_b = 5$ or $n_t = 3$ and $n_b = 3$ we have for a specific application found that $V_2 = 0$ which lead to that $\rho = \infty$. This caused no problem since the correct $m = 4$ degrees of freedom are obtained from equation (27).

B. Construction of a restriction matrix

Let $\mathbf{A}^{k \times n_A}$ and $\mathbf{B}^{k \times n}$ be two real matrices with $\mathcal{C}(\mathbf{A}) \subset \mathcal{C}(\mathbf{B})$ and $n_A < n$. Then one choice of a $k \times n$ restriction matrix $\mathbf{L}$ is $\mathbf{L} = (\mathbf{I} - \mathbf{A}(\mathbf{A}^\top \mathbf{A})^- \mathbf{A}^\top)\mathbf{B}$ which satisfies

$$\mathcal{C}(\mathbf{A}) = \{\mathbf{Bb} | \mathbf{b} \in R^n, \mathbf{Lb} = \mathbf{0}\} \tag{29}$$

where $(\mathbf{A}^\top \mathbf{A})^-$ is a generalized inverse of $(\mathbf{A}^\top \mathbf{A})$.

Proof: Let $U = \{\mathbf{Bb} | \mathbf{b} \in R^n, \mathbf{Lb} = \mathbf{0}\}$. $\mathcal{C}(\mathbf{A}) \subset U$ is obvious because of the assumption that $\mathcal{C}(\mathbf{A}) \subset \mathcal{C}(\mathbf{B})$. Conversely, $U \subset \mathcal{C}(\mathbf{A})$ can be seen as follows: $\mathbf{P_A} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^- \mathbf{A}^\top$ is the orthogonal projection onto $\mathcal{C}(\mathbf{A})$. Any vector $\mathbf{Bb} \in \mathcal{C}(\mathbf{B})$ can be written as

$$\mathbf{Bb} = \mathbf{P_A Bb} + (\mathbf{I} - \mathbf{P_A})\mathbf{Bb} = \mathbf{P_A Bb} + \mathbf{LBb} = \mathbf{P_A Bb} \tag{30}$$

and therefore $\mathbf{Bb} \in \mathcal{C}(\mathbf{A})$.

Notice: Another choice of $\mathbf{L}$ is the orthogonal projection onto the orthogonal complement of $\mathcal{C}(\mathbf{A})$ in $\mathcal{C}(\mathbf{B})$ which is $\mathbf{L} = (\mathbf{I} - \mathbf{P_A})\mathbf{P_B} = \mathbf{P_B} - \mathbf{P_A}$. The proof follows along the lines of (30) replacing $\mathbf{P_A}$ with $\mathbf{P_B} - \mathbf{P_A}$.

Regarding computations: The computation is done via the QR decomposition of the concatenated matrix $\mathbf{D} = (\mathbf{A}|\mathbf{B})$. Let r_A and r_B the ranks of $\mathbf{A}$ and $\mathbf{B}$. The QR-decomposition $\mathbf{D} = \mathbf{QR}$ provides via pivoting a matrix $\mathbf{Q}$ such that for the matrix $\mathbf{Q}_1$ of the first r_A columns of $\mathbf{Q}$ one has $\mathcal{C}(\mathbf{Q}_1) = \mathcal{C}(\mathbf{A})$. The matrix $\mathbf{Q}_2$ of the following $r_B - r_A$ columns has $\mathcal{C}(\mathbf{Q}_2)$ which is the orthogonal complement of $\mathcal{C}(\mathbf{A})$ in $\mathcal{C}(\mathbf{B})$. Then $\mathbf{Q}_2\mathbf{Q}_2^\top$ is the orthogonal projection onto this complement. With the note above we have a solution $\mathbf{L} = \mathbf{Q}_2\mathbf{Q}_2^\top\mathbf{B}$. Because

the nullspaces of $\mathbf{Q}_2\mathbf{Q}_2^\top$ and $\mathbf{Q}_2^\top$ are the same a $\mathbf{L}$ with row-rank equal to its rank is obtained by $\mathbf{Q}_2^\top\mathbf{B}$.

Affiliation:

Ulrich Halekoh
Department of Animal Health and Bioscience
Aarhus University
Blichers Allé, 8830 Tjele, Denmark
E-mail: ulrich.halekoh@agrsci.dk
Søren Højsgaard
Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7G, 9220 Aalborg Ø, Denmark
E-mail: sorenh@math.aau.dk
URL: http://people.math.aau.dk/~sorenh/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume VV, Issue II *Submitted:* yyyy-mm-dd
MMMMMM YYYY *Accepted:* yyyy-mm-dd