Frequentist inference for linear mixed models - continued

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

March 10, 2020

Outline for today

- asymptotic inference for general linear mixed models

Inference for balanced ANOVA - status

Parameters	Estimation	Tests/Confidence intervals
μ/β	Closed form estimates	exact t- or F-tests for mean structure, exact conf. intervals
λ - variances	Closed form estimates	exact χ^2 distributions, conf. intervals
σ^2_F - variances	Closed form estimates	F-tests for zero variance

For other models with sufficient balancedness/orthogonality, exact results can be derived too - e.g. orthodont data (see last part of slides).

The general linear mixed model

We do not have the nice exact results for linear mixed models in general.

Then we need to resort to asymptotic results, approximate F-tests or parametric bootstrap.

Can divide parameter vector into (cf. second lecture)

1. β: regression parameters for mean
2. σ^2: variance of uncorrelated homoscedastic noise
3. ψ: variance/correlation parameters of random effects

Confidence intervals for regression parameter β in general linear model with known correlation structure

Suppose $Y \sim N(X\beta, \sigma^2 V)$ where V known. Equivalently, inference based on $\tilde{Y} \sim N(\tilde{X}\beta, \sigma^2 I)$ obtained by transforming with L^{-1}, $V = LL^\mathsf{T}$.

MLE of $\mu = X\beta$ and β are

$$\hat{\mu} = X(X^\mathsf{T} V^{-1} X)^{-1} X^\mathsf{T} V^{-1} Y$$
$$\hat{\beta} = (X^\mathsf{T} V^{-1} X)^{-1} X^\mathsf{T} V^{-1} Y = (X^\mathsf{T} V^{-1} X)^{-1} X^\mathsf{T} V^{-1} \hat{\mu}$$

Since

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^\mathsf{T} V^{-1} X)^{-1})$$

we can obtain confidence intervals using t statistic.

In practice V typically contains unknown parameter ψ.

Things get more complicated since $\hat{\beta}$ and its distribution then may depend on these unknowns (MLE of $\hat{\beta}$: $V(\psi)$ substituted by $V(\hat{\psi})$ where $\hat{\psi}$ MLE).

If $\hat{\psi}$ consistent then $\hat{\beta}$ will be asymptotically normal and we may use $\hat{\sigma}^2 (X^\mathsf{T} V^{-1}(\hat{\psi}) X)^{-1})$ as approximate covariance matrix.

This gives approximate confidence intervals based on quantiles for normal distribution.

Wald-test

Wald-test: suppose we wish to test $H: K\beta = b$ for some $K: d \times p$ and $b \in \mathbb{R}^d$. Under hypothesis H,

$$T = (K\hat{\sigma}^2 (X^\mathsf{T} V^{-1}(\hat{\psi}) X)^{-1}) K^\mathsf{T})^{-1/2} [K\hat{\beta} - b] \approx N_d(0, I)$$

and

$$\|T\|^2 \approx \chi^2(d)$$

Kenward and Rogers (1997) suggested more accurate $F(d, m)$ approximate distribution of $\frac{\lambda}{d}\|T\|^2$ for some scaling factor $\lambda > 0$ and $m > 0$ - implemented in package `pbkrtest`

Their idea: match mean and variance of $\frac{\lambda}{d}\|T\|^2$ with those of $F(d, m)$ in order to determine scaling factor λ and denominator degrees of freedom m - for more details see Højsgaard and Halekoh (2014).

Example: orthodont

(wrong) Model without random effects - test for no interaction

```
> ort1=lm(distance~age+age:factor(Sex)+factor(Sex))
> ort2=lm(distance~age+factor(Sex))
> anova(ort2,ort1)
Analysis of Variance Table

Model 1: distance ~ age + factor(Sex)
Model 2: distance ~ age + age:factor(Sex) + factor(Sex)
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    105 541.87
2    104 529.76  1    12.114 2.3782 0.1261
```

F-test with $F(1, 104)$ distribution. p-value 0.1261.

t-test with 104 degrees of freedom gives same p-value.

orthodont - continued

More appropriate model with random effects:

```
> ort4=lmer(distance~age*Sex+(1|Subject))
> ort4.1=lmer(distance~age+Sex+(1|Subject))#remove interact
> KRmodcomp(ort4,ort4.1)
F-test with Kenward-Roger approximation; computing time: 0.
large : distance ~ age * Sex + (1 | Subject)
small : distance ~ age + Sex + (1 | Subject)
         stat     ndf     ddf F.scaling p.value
Ftest  6.3027  1.0000 79.0000         1  0.0141 *
```

F-test with $F(1,79)$ distribution.

Now p-value is 0.0141 (due to more appropriate modeling of variance structure).

Note: in fact exact test (see slides in the end)

test of Sex effect

Suppose we omit interaction between age and sex. Test for sex effect:

```
> ort4.1=lmer(distance~age+Sex+(1|Subject),data=Orthodont)
> ort4.2=lmer(distance~age+(1|Subject),data=Orthodont)
> KRmodcomp(ort4.1,ort4.2)
F-test with Kenward-Roger approximation; computing time: 0.
large : distance ~ age + Sex + (1 | Subject)
small : distance ~ age + (1 | Subject)
         stat     ndf     ddf F.scaling  p.value
Ftest  9.2921  1.0000 25.0000         1 0.005375 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

F-test with $F(1,25)$ distribution.

Note: in fact exact test (see slides in the end)

Genes data revisited

```
> fit1=lmer(intensity~treatment*factor(exon)+(1|patient)+(
> fit2=lmer(intensity~treatment+factor(exon)+(1|patient)+(
> KRmodcomp(fit1,fit2)
F-test with Kenward-Roger approximation; computing time: 0.
large : intensity ~ treatment + factor(exon) + (1 | patient
    treatment:factor(exon)
small : intensity ~ treatment + factor(exon) + (1 | patient
          stat      ndf      ddf F.scaling p.value
Ftest   1.4278   7.0000 126.0000         1  0.1998
```

Note: in this case KR-approximation coincides with exact $F(7,126)$-distribution !

In balanced ANOVA models using `lmer` we can use `KRmodcomp` to compute F-tests.

Noise parameter σ^2

Guess: distribution of REML $\hat{\sigma}^2$ close to $\sigma^2\chi^2(n-d)/(n-d)$ (where d dimension of mean space)

At least true if ψ known.

Testing hypotheses regarding σ^2 not relevant (there is always noise/measurement error)

Parameter ψ

Need to resort to asymptotic results...

Asymptotic inference

Let $l_n(\theta) = \log L_n(\theta)$ denote the log likelihood function and let

$$s_n(\theta) = \frac{\mathrm{d}l_n(\theta)}{\mathrm{d}\theta} \quad j_n(\theta) = -\frac{\mathrm{d}s_n(\theta)}{\mathrm{d}\theta^{\mathsf{T}}}$$

denote the score function and observed information. n is 'number of observations'. Recall $\mathbb{E}s_n(\theta) = 0$ and $\mathbb{V}\mathrm{ar}s_n(\theta) = i_n(\theta)$ where $i_n(\theta)$ is the Fisher information.

Suppose there is normalizing sequence c_n so

1. $c_n^{1/2}(\hat{\theta}_n - \theta)$ is bounded (in probability) and $i_n(\theta)/c_n$ is bounded
2. $c_n^{-1}[j_n(\theta) - i_n(\theta)] \to 0$ (in probability)
3. $c_n^{-1}(j_n(\hat{\theta}) - j_n(\theta)) \to 0$ (in probability)
4. $i_n(\theta)^{-1/2}s_n(\theta) \approx N(0, I)$ (CLT)

Then

$$(\hat{\theta}_n - \theta) \approx N(0, i_n(\theta)^{-1})$$

Asymptotic normality

By a first order Taylor expansion around $\hat{\theta}_n$,

$$s_n(\theta) \approx j_n(\hat{\theta}_n)(\hat{\theta}_n - \theta)$$

Thereby

$$\begin{aligned} i_n(\theta)^{-1/2}s_n(\theta) \approx &(i_n(\theta)/c_n)^{-1/2}\frac{1}{c_n}[j_n(\hat{\theta}_n) - j_n(\theta)]c_n^{1/2}(\hat{\theta}_n - \theta) + \\ &(i_n(\theta)/c_n)^{-1/2}\frac{1}{c_n}j_n(\theta)c_n^{1/2}(\hat{\theta}_n - \theta) \end{aligned}$$

Using assumptions 1.-3. on previous slide

$$i_n(\theta)^{-1/2}s_n(\theta) \approx i_n(\theta)^{1/2}(\hat{\theta}_n - \theta)$$

which by assumption 4. gives

$$i_n(\theta)^{1/2}(\hat{\theta}_n - \theta) \approx N(0, I) \Rightarrow (\hat{\theta}_n - \theta) \approx N(0, i_n(\theta)^{-1})$$

Wald-test (again)

Wald-test: suppose we wish to test $H : K\theta = c$ for some $K : d \times p$ and $c \in \mathbb{R}^d$. Under hypothesis H,

$$T = (Ki_n(\theta)^{-1}K^{\mathsf{T}})^{-1/2}[K\hat{\theta}_n - c] \approx N_d(0, I)$$

and

$$\|T\|^2 \approx \chi^2(d)$$

Asymptotic distribution of likelihood ratio

Suppose $H_0 : \theta \in \Theta_0$ with alternative hypothesis $\theta \in \Theta$. Then under 'regularity' conditions

$$-2 \log Q = -2[l(\hat{\theta}_{0,n}) - l(\hat{\theta}_n)] \approx \chi^2(d - d_0)$$

where d_0 and d number of 'free' parameters under H_0 and alternative, respectively.

Limitations of asymptotic results

- Usual 'regularity' conditions require that parameters do not fall on the boundary under H_0 ($\hat{\theta}_n - \theta_0$ can not be normal under restriction $\hat{\theta}_n \geq \theta_0$). Thus problematic if we want to test whether a variance is zero.
- Under $H : \tau^2 = 0$ for variance component τ^2 (or if true τ^2 close to zero), distribution of $\hat{\tau}^2$ skew (not normal).
- Need asymptotic normality of $s_n(\theta)$. Not always obvious how to use CLT for general linear mixed models - what should tend to infinity ? - and observations not independent (for independent observations we assume number of observations $n = c_n$ tend to infinity and use CLT)

Regarding last item: e.g. in one-way ANOVA we might require k tending to infinity rather than just $n = mk$ tending to infinity.

Faraway (2006), section 8.2 recommends parametric bootstrap for testing variance components (when exact results not applicable):

1. Simulate *iid* data $Y_1^*, \ldots, Y_B^*$ from model under null hypothesis.
2. Recompute likelihood ratio test for each simulated data set.
3. Compare observed LR with simulated distribution.

Likelihood ratio tests: ML vs REML

Note: we can only consider ratios between likelihoods evaluated for the *same* dataset.

For variance components ψ we may use REML likelihoods but *not* for mean parameters β since REML transformed data depends on model for the mean.

Following slides exemplifies inference based on asymptotic results.

Never do this in practice for balanced ANOVA where exact results are available !!!

Gene-expression data using `lmer`

```
fit1=lmer(intensity~treatment*factor(exon)
                 +(1|patient)+(1|patient:treatment),data=gene1,R
fit2=lmer(intensity~treatment+factor(exon)
                 +(1|patient)+(1|patient:treatment),data=gene1,R
anova(fit1,fit2)
     Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
fit2 12 266.80 303.70 -121.40
fit1 19 270.11 328.54 -116.06 10.686      7     0.1529

fit3=lmer(intensity~factor(exon)+
       (1|patient)+(1|patient:treatment),data=gene1,REML=F)
anova(fit2,fit3)

     Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
fit3 11 268.16 301.99 -123.08
fit2 12 266.80 303.70 -121.40 3.3627      1    0.06669
```

Note REML=F. Qualitatively same conclusions as before.

Estimates of fixed effects parameters and "t"-tests

```
                         Estimate Std. Error t value
(Intercept)                2.8776     0.1558  18.474
treatmentT                -0.2847     0.1431  -1.990
factor(exon)2316222  -1.4461     0.1475  -9.806
factor(exon)2316227  -0.3440     0.1475  -2.333
factor(exon)2316230  -0.2567     0.1475  -1.741
factor(exon)2316231  -0.2757     0.1475  -1.870
factor(exon)2316232   1.5414     0.1475  10.452
factor(exon)2316233   2.9420     0.1475  19.949
factor(exon)2316234   0.2695     0.1475   1.828

#pvalue for treatment:
> 2*(1-pnorm(1.99))
0.04659
```

p-value based on asymptotic normality bit smaller than for exact F-test.

REML-test of zero chip variance:

```
fit1=lmer(intensity~treatment+factor(exon)+
                       (1|patient)+(1|patient:treatment),dat
fit2=lmer(intensity~treatment+factor(exon)+(1|patient),data=gene
anova(fit1,fit2)

     Df    AIC    BIC  logLik Chisq Chi Df Pr(>Chisq)
fit2 11 276.93 310.76 -127.47
fit1 12 266.82 303.73 -121.41 12.11      1  0.0005016 ***
```

Same qualitative conclusion as before: variance is non-zero.

But χ^2 approximation could be very poor.

Summary inference for general linear mixed model

Parameters	Estimation	Tests/Confidence intervals
β	Closed form estimates given ψ	Approximate F-test and conf. int. based on approx. normality
σ^2	Closed form estimates given ψ	Approximate χ^2 distribution
τ^2, θ	Numerical approximation	Asymptotic results (?!) or parametric bootstrap

Some further examples of exact results

- Orthogonal decomposition and exact F-tests for orthodont data
- Test for variance components in general variance components model

Back to orthodont data

Model in vector form:

$$Y = X\beta + Z_S U + \epsilon$$

Here X is design matrix for intercept, age, sex and age:sex effects while Z_S is design matrix for subject factor S (balanced).

Decomposition of covariance matrix:

$$\mathbb{C}\mathrm{ov}\, Y = 4\tau^2 P_S + \sigma^2 I = \lambda_S P_S + \sigma^2 \tilde{Q}_I \quad \lambda_S = 4\tau^2 + \sigma^2 \quad \tilde{Q}_I = I - P_S$$

Regarding X: we make age orthogonal to 1_n and Z_S by subtracting mean age. After removing redundant columns, X is four-dimensional.

Decomposition of data vector:

$$P_S Y \sim N(1_n \xi + \mathrm{sex}\beta_{\mathrm{sex}}, \lambda_S P_S)$$

where $\mathrm{sex}_i = 1$ if ith observation female and 0 otherwise.

$$\tilde{Q}_I Y \sim N(\mathrm{age}\beta_{\mathrm{age}} + \mathrm{age{:}sex}\beta_{\mathrm{age:sex}}, \sigma^2 \tilde{Q}_I)$$

where $\mathrm{age{:}sex}_i = \mathrm{age}_i$ if ith observation female and 0 otherwise.

Note $\tilde{Q}_I$ is projection on $108 - 27 = 81$ dimensional subspace. Mean vector is 2-dimensional. Thus F-test for no sex-age interaction:

$$\frac{\|P_{age*sex}\tilde{Q}_I Y - P_{age}\tilde{Q}_I Y\|^2}{\tilde{\sigma}^2} = \frac{\|P_{age*sex} Y - P_{age} Y\|^2}{\tilde{\sigma}^2} \sim F(1, 81-2)$$

where $P_{age*sex}$ projection on span{age,age:sex} while P_{age} projection on span{age}.

P_S projection on 27 dimensional subspace. Mean space for $P_S Y$ is 2 dimensional. F-test for no sex effect:

$$\frac{\|P_{sex}P_S Y - P_0 P_S Y\|^2}{\tilde{\lambda}_S} = \frac{\|P_{sex}Y - P_0 Y\|^2}{\tilde{\lambda}_S} \sim F(1, 27-2)$$

```
> anova(lm(distance~age*Sex+Subject,data=Orthodont))

          Df Sum Sq Mean Sq  F value    Pr(>F)
age        1 235.36 235.356 122.4502 < 2.2e-16 ***
Sex        1 140.46 140.465  73.0806 7.407e-13 ***
Subject   25 377.91  15.117   7.8648 7.484e-13 ***
age:Sex    1  12.11  12.114   6.3027    0.0141 *
Residuals 79 151.84   1.922
```

F-test for Sex: 140.465/15.117= 9.2919 - compare with previous results !

NB: hierarchical principle not pertinent here since age is not a factor !

Exact tests for variance components in general variance components model

Consider

$$Y = X\beta + \sum_{i=1}^{K} Z_i U_i + \epsilon$$

where $U_i \sim N_{d_i}(0, \sigma_i^2 I)$'s and $\epsilon \sim N_n(0, \sigma^2 I)$ independent (not necessarily balanced model).

Let $L = \text{span}\{X, Z_1, \ldots, Z_K\}$ and $L_{-1} = \text{span}\{X, Z_2, \ldots, Z_K\}$. Assume $L \neq L_{-1}$. Then

$$\mathbb{R}^n = L_{-1} \oplus V_1 \oplus V_I$$

where $V_1 = L \ominus L_{-1}$.

Let Q_1 orthogonal projection on V_1 and Q_I orthogonal projection on V_I. Then

$$Q_1 Y \sim N(0, \sigma^2 Q_1 + \sigma_1^2 Q_1 Z_1 Z_1^\top Q_1) \quad Q_I Y \sim N(0, \sigma^2 Q_I)$$

and independent.

Under $H_1 : \sigma_1^2 = 0$, $\|Q_1 Y\|^2$ and $\|Q_I Y\|^2$ independent scaled χ^2 and

$$\frac{\|Q_1 Y\|^2/d_1}{\|Q_I Y\|^2/d_I} \sim F(d_1, d_I)$$

Large values critical.

Exercises

1. Show that T on slide 9 has approximate $N_d(0, I)$ distribution.
2. Consider code on webpage with simulation study and parametric bootstrap for one-way anova. Simulate data set with number of groups $k = 5$, number of observations for each group $m = 50$, true value of $\sigma^2 = 1$ and true value of τ^2 equal to 0.5.
   - 2.1 Test the hypothesis $\tau^2 = 0$ using 1) exact anova test 2) likelihood ratio test and asymptotic p-value 3) likelihood ratio test with bootstrap p-value. Compare the results.
   - 2.2 Repeat 2.1 with $k = 50$ and $m = 5$.
   - 2.3 Repeat 2.1 and 2.2 with true $\tau^2 = 0$.
   - 2.4 Using the bootstrap results, assess how well $-2 \log \text{LR}$ is approximated by a $\chi^2(1)$-distribution