Frequentist inference for linear mixed models

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

March 6, 2020

Outline for today

- Likelihood ratio test
- Inference for the linear normal model
- Balanced one- and two-way ANOVA - test for fixed effects and variance components

General objectives

- Determine distributions of parameter estimates (confidence intervals)
- Perform tests for hypotheses of interest (e.g. likelihood ratio tests)

The linear model

Suppose $Y \sim N_n(\mu, \sigma^2 I)$, $\mu = X\beta$. Let P be orthogonal projection on $L = \text{span}(X)$ of dimension d (assuming X full rank, $P = X(X^\mathsf{T} X)^{-1} X^\mathsf{T}$).

Then $\hat{\mu} = PY$ and $\hat{\sigma}^2 = \|(I-P)Y\|^2/n$. It follows directly that $\hat{\mu}$ and $\hat{\sigma}^2$ are independent. Moreover $\hat{\beta} = (X^\mathsf{T} X)^{-1} X^\mathsf{T} Y$ is the unique solution to $X\hat{\beta} = \hat{\mu}$ and $\hat{\beta}$ and $\hat{\sigma}^2$ are thus independent too.

$\hat{\mu} \sim N(\mu, \sigma^2 P)$, $\hat{\beta} \sim N(\beta, \sigma^2 (X^\mathsf{T} X)^{-1})$ and $\hat{\sigma}^2 \sim \sigma^2 \chi^2(n-d)/n$.

Issue: distribution of $\hat{\beta}$ involves unknown σ^2. Let v_i the i'th diagonal element in $(X^\mathsf{T} X)^{-1}$. Then $\hat{\beta}_i \sim N(\beta_i, \sigma^2 v_i)$ and

$$t = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\tilde{\sigma}^2 v_i}} \sim \frac{N(0,1)}{\sqrt{\chi^2(n-d)/(n-d)}} \sim t(n-d)$$

where $\tilde{\sigma}^2 = n\hat{\sigma}^2/(n-d)$ is REML estimate.

Confidence intervals can be constructed easily from χ^2 and t distribution (Mat5).

We can also use t distribution to test $H_0 : \beta_i = b_0$. Small and large values of

$$t = \frac{\hat{\beta}_i - b_0}{\sqrt{\tilde{\sigma}^2 v_i}}$$

are critical for this hypothesis (note $t \sim t(n-d)$ under H_0).

p-value is the probability of observing larger value of $|t|$ in repeated experiments than the one actually observed.

Likelihood ratio tests

Consider a statistical model with parameter space Θ and a hypothesis $H_0 : \theta \in \Theta_0$ where $\Theta_0 \subset \Theta$.

Let $\hat{\theta} = \text{argmax}_{\Theta} L(\theta)$ and $\hat{\theta}_0 = \text{argmax}_{\Theta_0} L(\theta)$.

Then $LR = L(\hat{\theta}_0)/L(\hat{\theta}) \leq 1$ and the smaller ratio, the less we believe in H_0 (the less data are likely under H_0 than under the alternative $\theta \in \Theta \setminus \Theta_0$).

To judge how small LR is we compare LR with its distribution under H_0 - say $LR \sim F_{LR}$ under H_0.

The p-value is the probability (under H_0 and repeated sampling) of observing a smaller value of LR than the one, lr, actually observed: $p = F_{LR}(lr)$.

Often $Q = -2 \ln LR$ is used - in which case large values of Q are critical and $p = 1 - F_Q(q)$ ($q = -2\log(lr)$).

The problem is to determine F_{LR} (or F_Q). For certain models the exact distributions are known but in general we need to rely on asymptotic arguments.

Recap: Beta and F-distributions

$\chi^2(\nu) = \Gamma(\nu/2, 2)$ ($\beta = 2$ scale parameter)

$B(\alpha, \alpha')$ distribution of $\Gamma(\alpha, \beta)/[\Gamma(\alpha, \beta) + \Gamma(\alpha', \beta)]$ where $\Gamma(\alpha, \beta)$ and $\Gamma(\alpha', \beta)$ independent.

$F(f_1, f_2)$ distribution of $[\chi^2(f_1)/f_1]/[\chi^2(f_2)/f_2]$.

Back to the linear normal model

Suppose $H_0 : \mu \in L'$ where $L' \subset L$ is a subspace of L of dimension d'. The maximized likelihood functions under $\mu \in L$ and $\mu \in L'$ are

$$(\hat{\sigma}^2)^{-n/2} \exp(-n/2) \text{ and } ((\hat{\sigma}^2)')^{-n/2} \exp(-n/2)$$

where $(\hat{\sigma}^2)' = \|(I-P')Y\|^2/n$. Thus

$$LR = \Big(\frac{\|(I-P')Y\|^2}{\|(I-P)Y\|^2}\Big)^{-n/2}$$

Moreover $\|(I-P')Y\|^2 = \|(I-P)Y + (P-P')Y\|^2 = \|(I-P)Y\|^2 + \|(P-P')Y\|^2$. Thus

$$B = LR^{2/n} = \frac{\|(I-P)Y\|^2}{\|(I-P)Y\|^2 + \|(P-P')Y\|^2}$$

is beta $B((n-d)/2, (d-d')/2)$-distributed.

Moreover B is in one to one correspondence with

$$F = \frac{\|(P-P')Y\|^2/(d-d')}{\|(I-P)Y\|^2/(n-d)} = \frac{\|(P-P')Y\|^2/(d-d')}{\tilde{\sigma}^2}$$
$$= \frac{(\|(I-P')Y\|^2 - \|(I-P)Y\|^2)/(d-d')}{\tilde{\sigma}^2}$$

which is $F(d-d', n-d)$ distributed. Note large values of F and small values of B are critical.

Note: numerator in F measures differences in estimates of μ under respectively $\mu \in L$ and $\mu \in L'$. If this is small we tend to believe $\mu \in L'$.

Suppose L' is obtained from L by removing ith column in X - this corresponds to $H_0 : \beta_i = 0$. Then F is the squared t statistic for β_i.

Example F and t tests in linear model

```
#fit model with sex specific intercepts and slopes
> ort1=lm(distance~age+age:factor(Sex)+factor(Sex))
> summary(ort1)
...
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            16.3406     1.4162  11.538  < 2e-16 ***
age                     0.7844     0.1262   6.217 1.07e-08 ***
factor(Sex)Female       1.0321     2.2188   0.465    0.643
age:factor(Sex)Female  -0.3048     0.1977  -1.542    0.126
...
> #compute F-tests respecting hierarchical principle
> drop1(ort1,test="F")
Single term deletions
....
                Df Sum of Sq    RSS    AIC F value  Pr(F)
<none>                       529.76 179.75
age:factor(Sex)  1     12.11 541.87 180.19  2.3782 0.1261
```

age:Sex not significant ! (but recall, model is wrong)

Tests continued

Using anova() to test reduction from ort1 to ort2

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

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

Tests continued

```
> ort2=lm(distance~age+factor(Sex))
> drop1(ort2,test="F")
Single term deletions

Model:
distance ~ age + factor(Sex)
            Df Sum of Sq    RSS    AIC F value     Pr(F)
<none>                   541.87 180.19
age          1    235.36 777.23 217.15  45.606 8.253e-10 *
factor(Sex)  1    140.46 682.34 203.09  27.218 9.198e-07 *
```

both age and sex significant (but model still wrong)

Test for fixed effects in balanced two-way ANOVA

Factorization of likelihood function:

$$
\begin{aligned}
|\Sigma|^{-1/2}\exp(-\frac{1}{2}(Y-\mu)^\mathsf{T}\Sigma^{-1}(Y-\mu)) = \\
\lambda_P^{-d_P/2}\exp(-\frac{1}{2\lambda_P}\|\tilde{Q}_P Y - Q_0\mu\|^2)\times \\
(\lambda_{P\times T})^{-(d_{P\times T}-d_P)/2}\exp(-\frac{1}{2\lambda_{P\times T}}\|\tilde{Q}_{P\times T}Y - Q_T\mu\|^2)\times \\
(\lambda_I)^{-(n-d_{P\times T})/2}\exp(-\frac{1}{2\lambda_I}\|Q_I Y\|^2)
\end{aligned}
$$

Formally equivalent to product of likelihoods for three linear normal models.

($d_{P\times T}, d_P, d_T$ dimensions of $L_{P\times T}$, L_P and L_T)

Suppose we want to test hypothesis of no treatment effect $H_0 : \beta_t = 0,\ t = 1,\ldots,d_T$. Note that the only likelihood-factor which differs under H_0 is the second one:

$$(\lambda_{P\times T})^{-(d_{P\times T}-d_P)/2}\exp(-\frac{1}{2\lambda_{P\times T}}\|\tilde{Q}_{P\times T}Y - Q_T\mu\|^2)$$

This corresponds to working with a linear normal model with data $\tilde{Y} = \tilde{Q}_{P\times T}Y$, mean vector $\tilde{\mu} = Q_T\mu$ and variance $\lambda_{P\times T}$.
Therefore

$$\hat{\lambda}_{P\times T} = \|Q_{P\times T}Y\|^2/(d_{P\times T}-d_P)$$

Under H_0, $\mu = \xi 1_n$ whereby $\tilde{\mu} = Q_T\mu = 0$. The maximum likelihood estimate of $\lambda_{P\times T}$ under H_0 is therefore

$$\hat{\lambda}_{P\times T,0} = \|\tilde{Q}_{P\times T}Y\|^2/(d_{P\times T}-d_P)$$

According to the results for the linear normal model, the likelihood ratio becomes equivalent with the F-statistic

$$\frac{\|Q_T Y\|^2/(d_T-1)}{\|Q_{P\times T}Y\|^2/((d_P-1)(d_T-1))}$$

(recall $\|\tilde{Q}_{P\times T}Y\|^2 = \|Q_{P\times T}Y\|^2 + \|Q_T Y\|^2$)

Note $Q_T Y = P_T Y - P_0 Y$ hence $\|Q_T Y\|^2 = \sum_{ptr}(\bar{Y}_{\cdot t\cdot} - \bar{Y}_{\cdots})^2$ (measures how much treatment group means differ from total mean)

Tests of fixed effects in balanced ANOVA with random effects

Likelihood ratio tests equivalent to F-tests within the appropriate strata.

I.e. need to identify the appropriate random effect whose mean square (λ-estimate) becomes the denominator of the F-test. This random effect is the coarsest random effect which is finer than the fixed effect under investigation.

Anova table:

```
> fit1=lm(intensity~treat*factor(exon)+factor(patient)+
                       factor(patient):treat,data=gene1)
> anova(fit1)
Analysis of Variance Table

Response: intensity
                       Df  Sum Sq Mean Sq  F value    Pr(>F)
treat                  1    3.242   3.242  14.4796 0.0002199 ***
factor(exon)           7  254.343  36.335 162.2852 < 2.2e-16 ***
factor(patient)        9  15.405    1.712   7.6449 6.703e-09 ***
treat:factor(exon)     7    2.238   0.320   1.4278 0.1998234
treat:factor(patient)  9    8.190   0.910   4.0643 0.0001345 ***
Residuals              126 28.211   0.224
```

F-test for TxE: 1.4278 with p-value 0.1998.

ANOVA with treatment*exon removed:

```
> fit2=lm(intensity~treatment+factor(exon)+
        factor(patient)+factor(patient):treatment,data=gene1)
> anova(fit2)
Analysis of Variance Table

Response: intensity
                           Df  Sum Sq Mean Sq  F value    Pr(>F)
treatment                   1   3.242   3.242  14.1608 0.0002508 ***
factor(exon)              7 254.343  36.335 158.7121 < 2.2e-16 ***
factor(patient)             9  15.405   1.712   7.4766 8.472e-09 ***
treatment:factor(patient  9   8.190   0.910   3.9749 0.0001636 ***
Residuals                 133  30.448   0.229
```

treatment:factor(patient) (TxP) effect significant as systematic effect.

Due to hierarchic principle we would not pursue test of treatment main effect in a model with TxP as systematic effect.

How do we obtain test for treatment in model with TxP as random effect ?

With aov()

```
> fit1=aov(intensity~treatment*factor(exon)+Error(factor(patient)+factor(patient):treatment),data=gene1)
> summary(fit1)

Error: factor(patient)
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9   15.4   1.712

Error: factor(patient):treatment
          Df Sum Sq Mean Sq F value Pr(>F)
treatment  1  3.242   3.242   3.563 0.0917 .
Residuals  9  8.190   0.910
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Error: Within
                         Df Sum Sq Mean Sq F value Pr(>F)
factor(exon)            7 254.34   36.33 162.285 <2e-16 ***
treatment:factor(exon)  7   2.24    0.32   1.428    0.2
Residuals                 126  28.21    0.22
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Suppose we remove TxE from model of mean vector. Then we can use F-test for T in PxT stratum.

Test for zero between group variance in one-way ANOVA

Want to test $H: \tau^2 = 0$.

Recall $\lambda = m\tau^2 + \sigma^2$. Hence $\tau^2 = 0$ equivalent to $\lambda = \sigma^2$.

Natural statistic (though not LR):

$$F = \frac{\tilde{\lambda}}{\hat{\sigma}^2}$$

which is $F(k-1, k(m-1))$ under H.

Large values critical.

NB: F-test is identical to F-test for no systematic effect of the factor defining the one-way ANOVA !

Test for variance components in two-way ANOVA

Recall $\lambda_I = \sigma^2$, $\lambda_{P\times T} = \sigma^2 + n_{P\times T}\sigma^2_{P\times T}$ and $\lambda_P = \sigma^2 + n_{P\times T}\sigma^2_{P\times T} + n_P\sigma^2_P$.

Hence e.g. $\sigma^2_{P\times T} = 0 \Leftrightarrow \lambda_I = \lambda_{P\times T}$.

Natural statistic (but not LR) for testing $\sigma^2_{P\times T} = 0$ is statistic

$$F = \frac{\tilde{\lambda}_{P\times T}}{\tilde{\lambda}_I}$$

which has $F((d_P - 1)(d_T - 1), n - d_{P\times T})$ distribution if $\sigma^2_{P\times T} = 0$. Big values critical.

Note $\tilde{\lambda}_{P\times T} = \|Q_{P\times T}Y\|^2/((d_P-1)(d_T-1))$ so F is identical to statistic for testing fixed effect of factor $P \times T$ in a linear normal model without random effects.

Confidence intervals for variance components

Confidence intervals for 'λ' variance parameters straightforward due to the exact χ^2 distributions of their estimates.

Except for $\sigma^2 = \lambda_I$, confidence intervals for original variance parameters more complicated (estimates distributed as differences of scaled χ^2 distributions).

Confidence interval for signal to noise ratio (M & T Thm 5.2 and Remark 5.10)

Consider one-way ANOVA.

$$F = \frac{SSB/(k-1)}{SSE/(k(m-1))} = \frac{\tilde{\lambda}}{\hat{\sigma}^2} \sim \frac{\sigma^2 + m\tau^2}{\sigma^2}F(k-1, k(m-1)) = (1+m\gamma)F(k-1, k(m-1))$$

Thus with q_L and q_U e.g. 2.5% and 97.5% quantiles for $F(k-1, k(m-1))$ we have

$$P(q_L \le F/(1+m\gamma) \le q_U) = 95\% \Leftrightarrow$$
$$P((F/q_U - 1)/m \le \gamma \le (F/q_L - 1)/m) = 95\%$$

Test of card board variance components

One-way anova: test of no card board heterogeneity.

F-test:

$$F = \frac{\tilde{\lambda}_P}{\tilde{\lambda}_I} = \frac{0.0273}{0.00006} = 450$$

which is $F(33, 102)$ distributed.

p-value

```
> 1-pf(450,33,102)
[1] 0
```


Implementation in R

For cardboard/reflectance data, $k = 34$ and $m = 4$.

```
> anova(lm(Reflektans~factor(Pap.nr.)))
Analysis of Variance Table

Response: Reflektans
                 Df  Sum Sq Mean Sq F value    Pr(>F)
factor(Pap.nr.)  33 0.90088 0.02730   470.7 < 2.2e-16 ***
Residuals       102 0.00592 0.00006
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```


Using anova to test reduction

```
> m1=lm(Reflektans~factor(Pap.nr.))
> m2=lm(Reflektans~1)
> anova(m2,m1)
Analysis of Variance Table

Model 1: Reflektans ~ 1
Model 2: Reflektans ~ factor(Pap.nr.)
  Res.Df     RSS Df Sum of Sq     F    Pr(>F)
1    135 0.90679
2    102 0.00592 33   0.90088 470.7 < 2.2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```


Using lmer to test TxE

```
>fit1=lmer(intensity~treatment*factor(exon)+(1|patient)+(1
>fit2=lmer(intensity~treatment+factor(exon)+(1|patient)+(1
>anova(fit1,fit2)
Data: gene1
Models:
fit2: intensity ~ treatment + factor(exon) + (1 | patient)
fit1: intensity ~ treatment * factor(exon) + (1 | patient)
     Df AIC    BIC    logLik   deviance Chisq Df Pr(>Chisq)
fit2 12 266.80 303.70 -121.40   242.80
fit1 19 270.11 328.54 -116.06   232.11 10.686 7 0.1529
```

`anova()` applied to `lmer` effects computes p-values based on approximate χ^2 distribution of $-2logLR$ rather than on exact F-distribution :(

Next time we will see how the `KRmodcomp()` procedure can be used to compute F-tests based on `lmer`-objects.

Exercises

1. (one-way ANOVA) Show that the test for zero between group variance is equivalent to the test for no fixed effect of the factor defining the groups.
2. (gene example) Show that the F-test for a systematic treatment:patient effect is equivalent to the F-test for zero chip variance.
3. Write down all the details of how to obtain the F-test for the fixed factor in the two-way ANOVA.
4. Compute estimate and F-test for σ_P^2 for the genes data example.