Frequentist inference for linear mixed models

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

March 12, 2012

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
- Inference for general linear mixed models

General objectives

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

The linear model

Suppose $Y \sim N_n(\mu, \tau^2 I)$, $\mu = X\beta$. 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 $\hat{\tau}^2 = \|(I-P)Y\|^2/n$. It follows directly that $\hat{\mu}$ and $\hat{\tau}^2$ 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 $\hat{\tau}^2$ are thus independent too.

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

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

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

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

Confidence intervals can be constructed easily from χ^2 and t distribution (Bo's course).

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

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

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 $LR = L(\hat{\theta}_0) \leq L(\hat{\theta})$ 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 $LR \sim F$ 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.

Back to the linear normal model

Suppose $H_0 : \mu \in L_0$ where $L_0 \subset L$ is a subspace of L of dimension d_0. The maximized likelihood functions under $\mu \in L$ and $\mu \in L_0$ are

$$(\hat{\tau}^2)^{-n/2} \exp(-n/2) \text{ and } (\hat{\tau}_0^2)^{-n/2} \exp(-n/2)$$

where $\hat{\tau}_0^2 = \|(I - P_0)Y\|^2/n$. Thus

$$LR = \Big(\frac{\|(I-P_0)Y\|^2}{\|(I-P)Y\|^2}\Big)^{-n/2}$$

Moreover $\|(I-P_0)Y\|^2 = \|(I-P)Y + (P-P_0)Y\|^2 = \|(I-P)Y\|^2 + \|(P-P_0)Y\|^2$. Thus

$$B = LR^{2/n} = \frac{\|(I-P)Y\|^2}{\|(I-P)Y\|^2 + \|(P-P_0)Y\|^2}$$

is beta $B((n-d)/2, (d-d_0)/2)$-distributed.

Moreover B is in one to one correspondance with

$$F = \frac{\|(P-P_0)Y\|^2/(d-d_0)}{\|(I-P)Y\|^2/(n-d)} = \frac{\|(P-P_0)Y\|^2/(d-d_0)}{\tilde{\tau}^2}$$

which is $F(d-d_0, n-d)$ 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 $\mu \in L_0$. If this is small we tend to believe $\mu \in L_0$.

Suppose L_0 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 (exercise)

Beta and F-distributions

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

$B(\alpha, \alpha')$ 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 $[\chi^2(f_1)/f_1]/[\chi^2(f_2)/f_2]$.

Tests for variance components and fixed effects in balanced two-way ANOVA

Factorization of likelihood function:

$$\begin{aligned}
&|\Sigma|^{-1/2}\exp(-\frac{1}{2}(Y-\xi)^{\mathsf{T}}\Sigma^{-1}(Y-\xi)) = \\
&\qquad \lambda_P^{-m/2}\exp(-\frac{1}{2\lambda_P}\|P_P Y - P_0\xi\|^2)\times \\
&\qquad (\lambda_{P\times T})^{-m(k-1)/2}\exp(-\frac{1}{2\lambda_{P\times T}}\|\tilde{Q}_{P\times T}Y - Q_T\xi\|^2)\times \\
&\qquad (\lambda_I)^{-(n-mk)/2}\exp(-\frac{1}{2\lambda_I}\|Q_I Y\|^2)
\end{aligned}$$

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

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

$$(\lambda_{P\times T})^{-m(k-1)/2}\exp(-\frac{1}{2\lambda_{P\times T}}\|\tilde{Q}_{P\times T}Y - Q_T\xi\|^2)$$

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

$$\hat{\lambda}_{P\times T} = \|Q_{P\times T}Y\|^2/(m(k-1))$$

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

$$\hat{\lambda}_{P\times T,0} = \|\tilde{Q}_{P\times T} Y\|^2/(m(k-1))$$

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

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

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

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

Test for variance components

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

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 $F((m-1)(k-1), n-mk)$ distribution if $\sigma^2_{P\times T} = 0$. Big values critical.

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

Confidence intervals for variance components (Example 2.2 and 2.3 page 67 in Jiang)

Consider one-way ANOVA.

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

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

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

Similarly, if we let $c = (1, 1, \ldots, 1, -(k-1))/k$ then $\|c\|^2 = 1 - 1/k$ and c orthogonal to 1_k. Thereby

$$u_i = \bar{y}_{i\cdot} + \sum_{j=1}^{k} c_j y_{ij} \sim N(\mu, \sigma^2 + \tau^2)$$

and u_i's independent. Thus

$$\frac{\sum_{i=1}^m (u_i - \bar{u})^2}{\sigma^2 + \tau^2} \sim \chi^2(m-1)$$

(as Jiang remarks on page 66, confidence intervals often constructed based on 'pivotal' quantities where subtracting or dividing with the parameter of interest leads to a quantity with known distribution)

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 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, $m = 34$ and $k = 4$.

```
> 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 ' ' 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
```


Exercises

1. Suppose L is spanned by columns $v_1, \ldots, v_d$ in X and L_0 is spanned by X_0 obtained by removing first column v_1 from X. Consider the linear model $Y \sim N(X\beta, \tau I^2)$
   1.1 show that $L = L_0 \oplus L_u$ where L_u is spanned by $u = v_1 - P_0 v_1$.
   1.2 show that the maximum likelihood estimate of β_1 is $\hat{\beta}_1 = Y \cdot u / \|u\|^2$.
   1.3 show that the F-test statistic for the reduction $L \to L_0$ is equal to the squared t-test statistic.
2. Write down all the details of how to obtain the F-test for the fixed factor in the two-way ANOVA.
3. In a one-way ANOVA with one factor A show that the F-test for no fixed effect of A is equal to the F-test for zero variance of the random effects in the mixed ANOVA model with a random effect at each level of A.