Frequentist inference for linear mixed models - continued

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

March 12, 2012

Exact tests for variance components in general mixed ANOVA model (Jiang page 51-53)

Consider

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

where $\alpha_i \sim N_{d_i}(0, \sigma_i^2 I)$'s and $\epsilon \sim N_n(0, \tau^2 I)$ independent.

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

$$Q_1 Y \sim N(0, \tau^2 Q_1 + \sigma_1^2 Q_1 Z_1 Z_1^\mathsf{T} Q_1) \quad Q_I Y \sim N(0, \tau^2 Q_I)$$

Under $H_1 : \sigma_1^2 = 0$,

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

Application to two-way balanced ANOVA (Jiang page 52-53)

Suppose random effects factors A, B and $A \times B$.

$$\mathbb{R}^n = L_0 \oplus V_A \oplus V_B \oplus V_{A\times B} \oplus V_I$$

If $\sigma_{A\times B}^2 = 0$ then preceeding results leads to consideration of

$$F = \frac{\|Q_A Y\|^2/(a-1)}{\|(Q_{A\times B} + Q_I)Y\|^2/(n-(a+b-1)}$$

for testing $\sigma_A^2 = 0$.

If $\sigma_{A\times B}^2 \neq 0$ we can consider

$$\frac{\|Q_A Y\|^2/(a-1)}{\|Q_{A\times B} Y\|^2/((a-1)(b-1))}$$

for testing $\sigma_A^2 = 0$.

Confidence intervals for regression parameters (Jiang page 70-72)

Suppose $Y \sim N(X\beta, \tau^2 V)$ where V known. Inference is equivalent for $\tilde{Y} \sim N(\tilde{X}\beta, \tau^2 I)$ 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
$$\hat{\beta} \sim N(\beta, \tau^2 (X^\mathsf{T} V^{-1} X)^{-1})$$
we can obtain confidence intervals using pivotal t or F statistics.

If V contains unknown parameters θ things in general more complicated since $\hat{\beta}$ and its distribution then may depend on these unknowns (MLE of $\hat{\beta}$: $V(\theta)$ substituted by $V(\hat{\theta})$ where $\hat{\theta}$ MLE).

If $\hat{\theta}$ consistent then $\hat{\beta}$ will be asymptotically normal and we may use $\hat{\tau}^2 (X^\mathsf{T} V^{-1}(\hat{\theta}) X)^{-1})$ as approximate covariance matrix. This gives approximate confidence intervals as (2.29) in Jiang.

WLS and BLUE

Suppose that Y has mean $X\beta$ and known covariance matrix Σ (but Y may not be normal). Then

$$(X^\mathsf{T} \Sigma^{-1} X)^{-1} X^\mathsf{T} \Sigma^{-1} Y$$

is a weighted least squares estimate since it minimizes

$$(Y - X\beta)^\mathsf{T} \Sigma^{-1} (Y - X\beta).$$

It is also the best linear unbiased estimate (BLUE) - that is the unbiased estimate with smallest variance in the sense that

$$\mathbb{V}\mathrm{ar}\tilde{\beta} - \mathbb{V}\mathrm{ar}\hat{\beta}$$

is positive semi-definite for any other linear unbiased estimate $\tilde{\beta}$.

Proof of BLUE: Suppose that $\hat{\mu} = PY$ is the weighted least squares estimate and that $\tilde{\mu} = BY$ is another unbiased linear estimate of μ. Unbiasedness means that $\mathbb{E}\hat{\mu} = \mathbb{E}\tilde{\mu} = \mu$ for all μ. Thus $\mathbb{E}(P - B)Y = (P - B)\mu = 0$ for all μ and $(P - B)Px = 0$ for any $x \in \mathbb{R}^n$. It now follows that $x^\mathsf{T}(\mathbb{V}\mathrm{ar}\tilde{\mu} - \mathbb{V}\mathrm{ar}\hat{\mu})x = x^\mathsf{T}\mathbb{V}\mathrm{ar}(\tilde{\mu} - \hat{\mu})x \geq 0$

Concerning $\hat{\beta}$ and $\tilde{\beta}$ we may write these as $C\hat{\mu}$ and $C\tilde{\mu}$ for some matrix C provided X has full rank (C then left-inverse of X).

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'

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

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

Recall $\mathbb{V}\mathrm{ar} s_n(\theta) = i_n(\theta)$ where $i_n(\theta)$ is the Fisher information.

Suppose there is normalizing sequence c_n so $c_n^{-1} i_n(\theta) \to i(\theta)$, $c_n^{-1}(j_n(\theta) - i(\theta)) \to 0$, and

$$\sqrt{c_n^{-1}} s_n(\theta) \approx N(0, i(\theta))$$

(CLT). Then

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

Wald-test

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,

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

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 and d_0 number of 'free' parameters under H and alternative, respectively.

Limitations of asymptotic results:

- based on Taylor expansions around θ. Problematic if θ on boundary of parameter space under H_0 (e.g. when testing variances equal to zero).
- Need asymptotic normality of $s_n(\theta)$. Not always obvious how to use CLT for general linear mixed models - what should tend to infinity ? (for independent observations we assume number of observations $n = c_n$ tend to infinity and use CLT)

Faraway (2006), section 8.2 recommends parametric bootstrap for testing variance components:

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.

Exercises

1. Excs. 2.1, 2.3, 2.14 og 2.15 i Jiang.
2. Show that $\mathbb{V}\mathrm{ar}\tilde{\beta} - \mathbb{V}\mathrm{ar}\hat{\beta}$ positive definite implies $\mathbb{V}\mathrm{ar}c^{\mathsf{T}}\tilde{\beta} - \mathbb{V}\mathrm{ar}c^{T}\hat{\beta} > 0$ for any vector c (of same dimension as β).
3. Sketch arguments of asymptotic normality of parameter estimates in the case of *iid* observations.