Parameter estimates and confidence intervals for linear mixed-effects models

Hani Nakhoul

22 March 2010

Outline

Outline

1. ML and REML estimation
   - The regression model
   - The mixed effects model
2. Confidence intervals
3. Hypothesis testing

Estimators in the regression model

- Consider

$$\mathcal{Y} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$$

(dim $\mathcal{Y} = n$, dim $\beta = p$)

- Given observed data **y**, the least squares estimator

$$\hat{\beta} = \arg\min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2$$

minimizes $\|\mathbf{y} - \mathbf{X}\beta\|^2$, a sum of n 'residuals.' It satisfies the p linearly independent constraints

$$\mathbf{X}^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\hat{\beta}) = 0.$$

- $\hat{\beta}$ is also the maximum likelihood estimator (MLE) for β.

Estimators in the regression model

- The MLE for σ^2 turns out to be

$$\hat{\sigma}_L^2 = \frac{\|\mathbf{y} - \mathbf{X}\hat{\beta}\|^2}{n},$$

but a more common choice of estimator is

$$\hat{\sigma}_R^2 = \frac{\|\mathbf{y} - \mathbf{X}\hat{\beta}\|^2}{n - p}.$$

- The definition of $\hat{\sigma}_R^2$ indicates that the residuals have $n - p$ degrees of freedom.
- $\hat{\sigma}_R^2$ is also unbiased (i.e. its expected value is σ^2).
- We call $\hat{\sigma}_R^2$ a REML ('residual ' or 'restricted' maximum likelihood) estimator.

Outline

The linear mixed model

- The general form:

$$\begin{aligned} (\mathcal{Y}|\mathcal{B} = \mathbf{b}) &\sim \mathcal{N}(\mathbf{X}\beta + \mathbf{Z}\mathbf{b}, \sigma^2\mathbf{I}_n) \\ \mathcal{B} &\sim \mathcal{N}(\mathbf{0}, \Sigma_\theta) \end{aligned}$$

- $\mathcal{Y}$ response vector of dimension n
- $\mathcal{B}$ random-effects vector of dimension q
- β fixed-effects vector of dimension p
- $\Sigma_\theta = \sigma_1^2 \Lambda_\theta \Lambda_\theta^T$, for $q \times q$ relative covariance factor Λ_θ

The linear mixed model: ML estimation

- Given observed data $\mathbf{y}$, the ML estimators minimize the deviance

$$d(\theta, \beta, \sigma|\mathbf{y}) = -2 \log L(\theta, \beta, \sigma|\mathbf{y}) = n \log(2\pi\sigma^2) + \log(\|\mathbf{L}_\theta\|^2) + \frac{r^2_{\beta,\theta}}{\sigma^2}$$

- $\mathbf{L}_\theta$ is a lower triangular $q \times q$ matrix satisfying

$$\mathbf{L}_\theta \mathbf{L}_\theta^{\mathrm{T}} = \Lambda_\theta^{\mathrm{T}} \mathbf{Z}^{\mathrm{T}} \mathbf{Z} \Lambda_\theta + \mathbf{I}_q$$

- $r^2_{\beta,\theta}$ minimizes the penalized residual sum of squares (PRSS)

$$r^2(\theta, \beta, \mathbf{u}) = \|\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\Lambda_\theta \mathbf{u}\|^2 + \|\mathbf{u}\|^2$$

with respect to $\mathbf{u}$.

The linear mixed model: ML estimation

$$d(\theta, \beta, \sigma|\mathbf{y}) = n\log(2\pi\sigma^2) + \log(\|\mathbf{L}_\theta\|^2) + \frac{r^2_{\beta,\theta}}{\sigma^2}$$

- Conditioned estimators: $\hat{\beta}_\theta = \arg\min_\beta r^2_{\beta,\theta}$ and $\hat{\sigma}^2_\theta = r^2_\theta / n$ (where $r^2_\theta = \min_\beta r^2_{\beta,\theta}$)
- The profiled deviance

  $$\tilde{d}(\theta|\mathbf{y}) = d(\theta, \hat{\beta}_\theta, \hat{\sigma}^2_\theta)$$

  is a function of θ alone; numerical optimization then gives the MLE $\hat{\theta}$ of θ.
- Evaluate $\tilde{d}$ at $\hat{\theta}$ to obtain $\hat{\beta}$ and $\hat{\sigma}$.

The linear mixed model: REML estimation

- We seek estimators $\hat{\theta}_R$ and $\hat{\sigma}^2_R$ that minimize the REML criterion

$$d_R(\theta, \sigma|\mathbf{y}) = -2\log \int_{\mathbb{R}^p} L(\theta, \beta, \sigma|\mathbf{y}) d\beta.$$

- Computing the integral yields

$$d_R(\theta, \sigma|\mathbf{y}) = (n-p)\log(2\pi\sigma^2) + 2\log(\|\mathbf{L}_\theta\| \|\mathbf{R}_X\|) + \frac{r^2_\theta}{\sigma^2}.$$

($\mathbf{R}_X$ is some matrix dependent on $\mathbf{X}$ and θ.)

The linear mixed model: REML estimation

- Conditioned estimator: $\hat{\sigma}^2_{R,\theta} = r^2_\theta/(n-p)$
- The profiled REML criterion

$$\tilde{d}_R(\theta|\mathbf{y}) = d_R(\theta, \hat{\sigma}^2_{R,\theta})$$

  is a function of θ alone.
- The REML estimators are therefore $\hat{\theta}_R = \arg\min_\theta \tilde{d}_R(\theta|\mathbf{y})$ and $\hat{\sigma}^2_R = \hat{\sigma}^2_{R,\hat{\theta}_R}$.
- Although the REML criterion d_R does not depend on β, by custom $\hat{\beta}_R = \hat{\beta}_{\theta_R}$.

ML vs. REML: which to choose?

- REML estimation is less biased than ML estimation, although it is not true in general that the REML variance estimator is unbiased.
- The linear mixed-effects REML estimator takes the regression estimator as a special case.
- ML estimation produces model-fit statistics (e.g. AIC, BIC) that can be used to compare models fitted to the same data.
- ML estimators are required to find confidence intervals.

Confidence intervals: the general method

- Obtain ML estimators of
  - σ_1, standard deviation of random effects
  - σ, standard deviation of residual $\|\mathbf{y} - \mathbf{X}\beta\|^2$
  - β, the fixed-effects parameter

  which together give a globally optimal fit.
- Fix one parameter at a specific value, find best possible fit, and compute change in deviance from globally optimal fit.
- The change in deviance is the likelihood ratio test (LRT) statistic ($d = -2 \log L$).
- Take signed square root ζ as value of $\mathcal{Z} \sim \mathcal{N}(0, 1)$.

The profile zeta plot

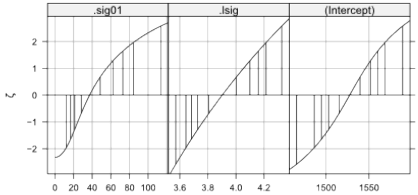

Fig. 1.5 Signed square root, ζ, of the likelihood ratio test statistic for each of the parameters in model fm1ML. The vertical lines are the endpoints of 50%, 80%, 90%, 95% and 99% confidence intervals derived from this test statistic.

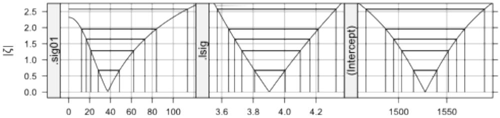

Fig. 1.6 Profiled deviance, on the scale |ζ|, the square root of the change in the deviance, for each of the parameters in model fm1ML. The intervals shown are 50%, 80%, 90%, 95% and 99% confidence intervals based on the profile likelihood.

The profile zeta plot

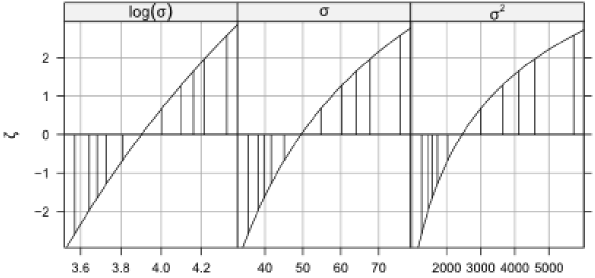

Fig. 1.7 Signed square root, ζ, of the likelihood ratio test statistic as a function of $\log(\sigma)$, of σ and of σ^2. The vertical lines are the endpoints of 50%, 80%, 90%, 95% and 99% confidence intervals.

The profile pairs plot

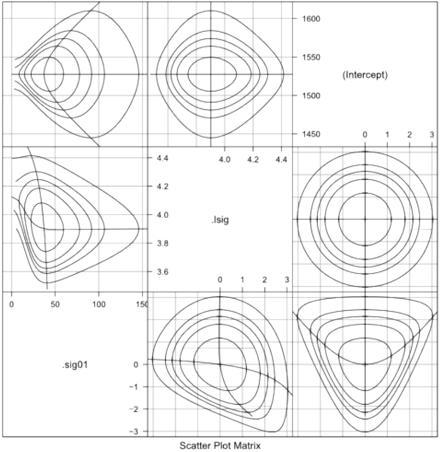

Fig. 1.9 Profile pairs plot for the parameters in model fm1. The contour lines correspond to two-dimensional 50%, 80%, 90%, 95% and 99% marginal confidence regions based on the likelihood ratio. Panels below the diagonal represent the (ζ_i, ζ_j) parameters; those above the diagonal represent the original parameters.

Hypotheses and models

- Given models A and B, we say A is nested in B if A is a 'special case' of B.
- Example: regression model nested in a linear mixed-effects model.
- Given null (H_0) and alternative (H_A) hypotheses about parameters, construct:
  - A reference model incorporating both H_0 and H_A
  - A nested (null hypothesis) model satisfying only H_0

Hypothesis testing

- Take estimates for parameters in each model.
- Evaluate change in deviance (the LRT statistic):

$$d_{nested} - d_{ref} = -2 \log \frac{L_{nested}}{L_{ref}} \sim \chi^2_{df},$$

  where *df* (degrees of freedom) is the difference in number of parameters between the reference and nested models.
- Refer to the χ^2_{df} distribution to determine the significance of the choice of model.

The `ergoStool` data

- Nine subjects (A-I) assess the difficulty of getting up from one of four stools (T1-T4).
- Difficulty is measured on the Borg scale (6-20): the higher the value, the greater the difficulty.
- What is a suitable model?