Maximum likelihood estimation for linear mixed models

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

February 13, 2012

Outline for today

- linear mixed models
- the likelihood function
- maximum likelihood estimation
- restricted maximum likelihood estimation and ANOVA

Some further useful results

- suppose Y is partitioned into Y_1 and Y_2 with corresponding partitioning of Σ into Σ_{ij}, $i = 1, 2$. Then Y_1 and Y_2 are independent $\Leftrightarrow \Sigma_{12} = \Sigma_{21} = 0$.
- AY and BY independent $\Leftrightarrow A\Sigma B^\mathsf{T} = 0$ (Jiang, C.1).
- Matrix identities:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

$$(C^{-1} + DA^{-1}B)^{-1}DA^{-1} = CD(BCD + A)^{-1}$$

$$(C^{-1} + B^tA^{-1}B)^{-1}B^tA^{-1} = CB^t(BCB^t + A)^{-1}$$

Linear mixed models

Consider mixed model:

$$Y_{ij} = \beta_1 + U_i + \beta_2 x_{ij} + \epsilon_{ij}$$

May be written in matrix vector form as

$$Y = X\beta + ZU + \epsilon$$

where $\beta = (\beta_1, \beta_2)^\mathsf{T}$, $U = (U_1, \dots, U_m)^\mathsf{T}$ and $\epsilon = (\epsilon_{11}, \epsilon_{12}, \dots, \epsilon_{mk})^\mathsf{T}$.

Linear mixed model: general form

Consider model

$$Y = X\beta + ZU + \epsilon$$

where $U \sim N(0, \Psi)$ and $\epsilon \sim N(0, D)$ are independent.

All previous models special cases of this.

Then Y has multivariate normal distribution

$$Y \sim N(X\beta, Z\Psi Z^\mathsf{T} + D)$$

Inverse of covariance matrix

Assume D positive definite (e.g. scaled identity matrix).

Then $Z\Psi Z^\mathsf{T} + D$ guaranteed to be positive definite and

$$(Z\Psi Z^\mathsf{T} + D)^{-1} = D^{-1} - DZ(\Psi^{-1} + Z^\mathsf{T} D^{-1} Z)^{-1} Z^\mathsf{T} D^{-1}$$

Right hand side may be easier to evaluate if Ψ^{-1} and $Z^\mathsf{T} D^{-1} Z$ sparse (e.g. AR(1) random effects - next slide)

Example AR(1) - covariance and inverse covariance

Consider $U_1 = \epsilon_0$ and

$$U_i = aU_{i-1} + \epsilon_i, \quad i = 2, \dots, m$$

where ϵ_i independent zero-mean normal with variances $\mathbb{V}\mathrm{ar}\epsilon_0 = \sigma_0^2$ and $\mathbb{V}\mathrm{ar}\epsilon_i = \sigma^2$, $i > 1$.

Then $U = B\epsilon$ so $U \sim N_n(0, BCB^\mathsf{T})$ where $C = \mathrm{diag}(\sigma_0^2, \sigma^2, \dots, \sigma^2)$. Hence $\Psi = BCB^\mathsf{T}$ and $\Psi^{-1} = (B^{-1})^\mathsf{T} C^{-1} B^{-1}$.

Expressions for covariances simplify in the stationary case $|a| < 1$ and $\sigma_0^2 = \sigma^2/(1 - a^2)$.

Marginal models

If assumption of normality is not tenable, *marginal models* are sometimes used.

That is, just specify $\mathbb{E}Y = X\beta$ and $\mathbb{C}\mathrm{ov}\,Y = V(\theta)$ but drop normality assumption.

Use methods for estimation of β and θ which only depends on specified mean and covariance (quasi-likelihood/generalized estimating equations).

Inference for random effects not possible.

Balanced mixed models

Balanced mixed models arise when model specified using cross-combinations of balanced factors/grouping variables or nested factors.

Example: balanced two-way analysis of variance.

Example: nested model for reflectance measurements.

For such models very nice inference results are available (later).

Can be written as linear mixed models with Z-matrix of the form given below on page 5 in Jiang.

Likelihood for linear mixed model

log likelihood for linear mixed model:

$$-\frac{1}{2}\log(|\Sigma(\psi)|) - \frac{1}{2}(y - X\beta)^{\mathsf{T}}\Sigma(\psi)^{-1}(y - X\beta))$$

ψ: parameters for $\Sigma(\psi)$ (e.g. variance components)

MLE and weighted least squares

Assume ψ known. MLE for β is weighted least squares estimate

$$\hat{\beta}(\psi) = \arg\min_{\beta}(y - X\beta)^{\mathsf{T}}\Sigma(\psi)^{-1}(y - X\beta)$$

Differentiate and equate to zero:

$$X^{\mathsf{T}}\Sigma(\psi)^{-1}(y - X\beta) = 0 \Leftrightarrow \hat{\beta}(\psi) = (X^{\mathsf{T}}\Sigma^{-1}X)^{-1}X^{\mathsf{T}}\Sigma(\psi)^{-1}y$$

(provided relevant inverses exist)

Covariance parameters for ψ: often numerical optimization is needed to maximize profile likelihood

$$-\frac{1}{2}\log(|\Sigma(\psi)|) - \frac{1}{2}(y - X\hat{\beta}(\psi))^{\mathsf{T}}\Sigma(\psi)^{-1}(y - X\hat{\beta}(\psi))$$

A geometric approach to estimation - orthogonal projections

Suppose L subspace of $\mathbb{R}^n$. Let $L^{\perp} = \{w \in \mathbb{R}^n | w \bullet v \text{ for all } v \in L\}$. P *orthogonal projection* on L if $P^2 = P$, $P^{\mathsf{T}} = P$ and $L = \text{span}P$.

Then $x = v + w$ where $v = Px \in L$ and $w = (I - P)x \in L^{\perp}$. By Pythagoras, Px minimizes $\|x - v\|$ over $v \in L$. This implies

1. decomposition $x = v + w$ where $v \in L$ and $w \in L^{\perp}$ is unique.
2. if Q and P are both orthogonal projections on L then $P = Q$.

Estimation using orthogonal projections

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

Then by Pythagoras, $\|Y - X\beta\|^2 = \|Y - PY\|^2 + \|PY - X\beta\|^2$. Hence $\hat{\mu} = Py$ and $\hat{\beta} = (X^\top X)^{-1}X^\top y$.

Moreover $\hat{\tau}^2 = \|Y - PY\|^2/n = \|Y - X\hat{\beta}\|^2/n$.

Suppose now $Y \sim N_n(\mu, \tau^2 V)$ where $V = LL^\top$ fixed. Then MLE based on Y and $\tilde{Y} = L^{-1}Y$ equivalent. Note $\mathbb{C}\text{ov}(\tilde{Y}) = \tau^2 I$ and $\mathbb{E}Y = L^{-1}X\beta = \tilde{X}\beta$. Hence by the above,

$$\hat{\beta} = (\tilde{X}^\top \tilde{X})^{-1}\tilde{X}^\top \tilde{y} = (X^\top V^{-1}X)^{-1}X^\top V^{-1}y$$

and

$$\hat{\tau}^2 = (y - X\hat{\beta})V^{-1}(y - X\hat{\beta})/n$$

Section 1.8 in Jiang - alternative derivation

Suppose $\mathbb{C}\text{ov}\,\epsilon = \tau^2 I$ and $\mathbb{C}\text{ov}\,U = \Psi = \sigma^2 L(\theta)L(\theta)^\top$

Then
$$\Sigma = \tau^2(I + \phi Z L(\theta)L(\theta)^\top Z^\top) = \tau^2 V(\theta, \phi)$$
where $\phi = \sigma^2/\tau^2$.

Given ϕ and θ,

$$\hat{\beta}(\phi, \theta) = (X^\top V^{-1}(\phi, \theta)X)^{-1}X^\top V(\psi, \theta)^{-1}y$$

and

$$\hat{\tau}^2(\phi, \theta) = \frac{1}{n}(y - X\hat{\beta}(\phi, \theta))^\top V(\psi, \theta)^{-1}(y - X\hat{\beta}(\phi, \theta))$$

Then use

$$V(\psi, \theta)^{-1} = (I + \phi Z L(\theta)L(\theta)^\top Z^\top)^{-1} = \\ I - Z(\phi^{-1}(L(\theta)^\top)^{-1}L(\theta) + Z^\top Z)^{-1}Z^\top$$

and
$$|I_n + \phi Z L(\theta)L(\theta)^\top Z^\top| = |I_m + \phi L(\theta)L(\theta)^\top Z^\top Z|$$
to arrive at the results in Jiang, section 1.8.

Profile log likelihood for θ:

$$l(\theta) = -\frac{1}{2}\log|\tau^2(\phi, \theta)V(\psi, \theta)| - \frac{n}{2} = \\ -\frac{n}{2}\log\tau^2(\phi, \theta) - \frac{1}{2}\log|I_m + \phi L(\theta)L(\theta)^\top Z^\top Z|$$

Some further useful matrix results

Consider
$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

Suppose A_{11} is invertible. Then $|A| = |A_{11}||A_{22} - A_{21}A_{11}^{-1}A_{12}|$.
Proof: use that

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix} = \begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} A$$

Moreover, if $A : n \times k$ and $B : k \times n$ then

$$|I_n + AB| = |I_k + BA|$$

Proof: use above result on

$$\begin{bmatrix} I_n & -A \\ B & I_k \end{bmatrix}$$

MLE's of variances biased or inconsistent

Simple normal sample $E\hat{\sigma}^2 = \sigma^2(n-1)/n$ (later consider balanced one-way anova too, Jiang, page 12)

Bias arise from estimation of μ ($\sum_i (y_i - \mu)^2$ vs $\sum_i (y_i - \bar{y}.)^2$).

Neyman-Scott example: $y_{ij} = \mu_i + \epsilon_{ij}$, $i = 1, \ldots, m$ and $j = 1, 2$. MLE of τ^2 not consistent as n tends to infinity.

REML (restricted/residual maximum likelihood)

Idea: linear transform of data which eliminates mean. Suppose design matrix $X : n \times p$ and let $A : n \times (n-p)$ have columns spanning the orthogonal complement $L^\perp$ of $L = \mathrm{span}X$. Then $A^\mathsf{T} X = 0$.

Transformed data ($(n-p) \times 1$)

$$\tilde{Y} = A^\mathsf{T} Y = A^\mathsf{T} Z\alpha + A^\mathsf{T}\epsilon$$

has mean 0 and covariance matrix $A^\mathsf{T}\Sigma(\psi)A$. Then proceed as for MLE.

NB: suppose A and B both span $L^\perp$. Then the same REML estimate of ψ is obtained (proof $B = AC$ for an invertible matrix C, write out likelihoods for $\tilde{Y}$ using A and AC).

REML

Given REML estimate $\hat{\psi}$ we use weighted least squares estimate of β:

$$\hat{\beta} = (X^\mathsf{T}\Sigma(\hat{\psi})^{-1}X)^{-1}X^\mathsf{T}\Sigma^{-1}(\hat{\psi})y$$

REML examples

Simple normal sample: A has columns $e_i - 1_n/n$, $i = 1, \ldots, n-1$ where 1_n is the n-vector of 1's and e_i is the ith standard basis vector.

Alternative: use columns $e_i - e_n$, $i = 1, \ldots, n-1$.

Neyman-Scott problem: see page 14 in Jiang.

Implementation of REML (Section 1.8 in Jiang)

Suppose $\mathbb{C}\mathrm{ov}\epsilon = \tau^2 I$ and $\mathbb{C}\mathrm{ov}U = \Psi = \sigma^2 L(\theta)L(\theta)^\mathsf{T}$

Then

$$\Sigma = \tau^2(I + \phi Z L(\theta)L(\theta)^\mathsf{T} Z^\mathsf{T}) = \tau^2 V(\theta,\phi)$$

where $\phi = \sigma^2/\tau^2$.

Choose A so that columns form an orthogonal basis for $L^\perp$ where $L = \mathrm{span}X$. Then $A^\mathsf{T} A = I$ and $AA^\mathsf{T} = I - X(X^\mathsf{T} X)^{-1}X^\mathsf{T}$ (since AA^T is a projection matrix).

$$\mathbb{C}\mathrm{ov}A^\mathsf{T} Y = A^\mathsf{T}\Sigma A = \tau^2(I + \phi A^\mathsf{T} Z L(\theta)L(\theta)^\mathsf{T} Z^\mathsf{T} A)$$

Hence given (ϕ,θ) estimate of τ^2 is

$$\hat{\tau}^2(\phi,\theta) = \tilde{Y}^\mathsf{T}\tilde{Y} - \tilde{Y}^\mathsf{T} A^\mathsf{T} Z[\phi^{-1}(L(\theta)L(\theta)^\mathsf{T})^{-1} + Z^\mathsf{T} AA^\mathsf{T} Z]^{-1} Z^\mathsf{T} A\tilde{Y}$$

Finally insert explicit expressions for AA^T and $A^\mathsf{T} A$.

Profile REML log likelihood for (ϕ,θ):

$$l(\phi,\theta) = -\frac{n-p}{2}\log\hat{\tau}^2(\phi,\theta) - \frac{1}{2}\log|(I+\phi Z^\mathsf{T} AA^\mathsf{T} Z L(\theta)L(\theta)^\mathsf{T}| - \frac{n-p}{2}$$

Balanced one-way ANOVA

The traditional approach: analysis of variance

Decomposition of variance ($i = 1,\dots,m$, $j = 1,\dots,k$):

$$\sum_{ij}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2 = \sum_{ij}(Y_{ij} - \bar{Y}_{i\cdot})^2 + k\sum_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 = SSE + SSA$$

Expected sums of squares:

$$\mathbb{E}SSE = m(k-1)\tau^2$$

$$\mathbb{E}SSA = k(m-1)\sigma^2 + (m-1)\tau^2$$

Moment-based estimates:

$$\hat{\tau}^2 = \frac{SSE}{m(k-1)}$$

$$\hat{\sigma}^2 = \frac{SSA/(m-1) - \hat{\tau}^2}{k}$$

NB: difficulties if unbalanced design of experiment.

NB: $\hat{\sigma}^2$ may be negative.

MLE for balanced one-way ANOVA

Let $\lambda = \tau^2 + k\sigma^2$. Solving equations on Jiang page 12 in the balanced case $k_i = k$ leads to ($\lambda = k\sigma^2 + \tau^2$):

$$\hat{\mu} = \bar{y}_{\cdot\cdot}, \hat{\tau}^2 = SSE/m(k-1), \hat{\lambda} = SSA/m$$

Note: $\hat{\lambda}$ biased $\Rightarrow \hat{\sigma}^2 = (\hat{\lambda} - \hat{\tau}^2)/k$ biased.

$$\mathbb{E}\hat{\lambda} = (k(m-1)\sigma^2 + (m-1)\tau^2)/m = k\frac{m-1}{m}\sigma^2 + \frac{m-1}{m}\tau^2$$

Asymptotically unbiased as m tends to infinity (then $\mathbb{V}\mathrm{ar}(\hat{\mu})$ tends to zero)

REML for balanced one-way ANOVA

E.g. A as for simple normal sample, i.e. $\tilde{y}_{ij} = y_{ij} - \bar{y}_{\cdot\cdot}$. Then REML equations for estimating τ^2 and σ^2 coincide with the moment equations (Exercise 1.16).

However for REML the estimates always restricted to be positive (i.e. if ANOVA estimate is negative then REML is on the boundary of the parameter space).

Exercises

1. Verify 'further useful results' and 'further useful matrix results'.
2. Exercises 1.10 in Jiang.
3. formulate random intercept and slope model for Orthodont data (day 1) as multivariate normal $N(\boldsymbol{\mu}, \Sigma)$. What are the design matrices X and Z ?
4. Compute Σ^{-1} when $Y_i = \mu + U_i + \epsilon_i$ and random effects follow stationary AR(1).
5. Show that the REML variance estimate for a simple normal sample coincides with s^2.
6. Compute variance of MLE $\hat{\sigma}^2$ and s^2 given that $\sum_{i=1}^n (x_i - \bar{x})^2$ is $\sigma^2\chi^2(n-1)$ (hint: $\mathbb{V}\mathrm{ar}\chi^2(f) = 2f$). What happens with the difference between the two estimates when n tends to infinity ?

7. Check the formulas for the moment-based estimates in the one-way ANOVA example.
8. Compute MLE and REML estimates for the Neyman-Scott example. Compute mean and variance for the estimates of τ^2 (Excs 1.8 in Jiang).
9. Show that if A and B both span the orthogonal complement of span$X$ then the same REML estimates are obtained from AY and BY (Excs 1.9 in Jiang).
10. Compute the projection a vector y on a vector v. Compute the projection of a vector y on span$X$ when columns in X are orthogonal.