Lecture 4: Gaussian Model - Restricted Maximum Likelihood Estimation (REML) (09-06-2007)

1 Introduction

We introduce restricted maximum likelihood (REML) estimation for Gaussian random effects model.

2 Balanced One-Way Classification with Random Effects

2.1 Estimation Based on Expectations

$$y_{kj} = \mu + a_k + e_{kj},\ (k = 1, \ldots, K;\ j = 1, \ldots, J).$$

where a_k and e_{ij} are normal random variables with variances σ_a^2 and σ^2.

Source	d.f.	Mean Square	Expected Mean Square
Treatments	K-1	s_a^2	$\sigma^2 + J\sigma_a^2$
Error	$K(J-1)$	s^2	σ^2

where, $s_a^2 = J\sum_k(\bar{y}_{k.} - \bar{y}_{..})^2/(K-1)$ and $s^2 = \sum_j\sum_k(y_{kj} - \bar{y}_{k.})^2/(K(J-1))$. If we let

$$s_a^2{=}\sigma^2 + J\sigma_a^2 \text{ and } s^2{=}\sigma^2,$$

$\sigma_a^2{=}(s_a^2 - s^2)/J$ may be negative.

2.2 True Maximum likelihood estimator (Herbach, 1959)

$\hat{\sigma}^2 = s^2$, $J\hat{\sigma}_a^2 = (1 - K^{-1})s_a^2 - s^2$, when $(1 - K^{-1})s_a^2 \geq s^2$

$\hat{\sigma}_a^2 = 0$, $\hat{\sigma}^2 = [(K-1)s_a^2 + K(J-1)s^2]/KJ$, when $(1 - K^{-1})s_a^2 < s^2$

Remark: **REML** refers to a slightly different method from true maximum likelihood method.

3 Motivating Example: Neyman-Scott Problem

3.1 Data and Interest of Variance Components

We have m pairs of observation (m means: μ_1,μ_2,...,μ_m):

$$y_{i,1}, y_{i,2}, (i = 1, \ldots, n) \tag{1}$$

all are i.i.d. as N(0,σ^2). We take the m means as nuisance parameters and are interested in estimating σ^2.

3.2 Estimation

1) Make transformation

$$z_i = y_{i,1} - y_{i,2}, (i = 1, 2, \ldots, m) \tag{2}$$

2) Now m means are gone and $z_i \sim$N(0,2σ^2), for which MLE of σ^2 is straightforward.

4 General REML Procedure (Patterson and Thompson, 1971)

4.1 Mixed Model

$$y = X\beta + Zb + \epsilon \tag{3}$$

where,

1) $X_{n,p}$ is fixed effects (p treatments) design matrix, rank($X_{n,p}$)=p.

2) β is fixed effects.

3) $Z_{n,r}$ is random effects (blocks) design matrix.

4) b is random effects, variance-covariance matrix $V(b) = D$.

5) ϵ is errors, variance-covariance matrix $V(\epsilon) = \Sigma$.

6) **Objectives**: estimate β,D and Σ.

4.2 Estimation

1) Find matrix $A_{n,n-p}$ such that rank(A)=$n - p$, $A'X = 0$.

2) Define $z = A'y$. And $z \sim N(0, A'VA)$, i.e.,

$$f_R(z) = \frac{1}{(2\pi)^{(n-p)/2}|A'VA|^{0.5}} \exp\{-\frac{1}{2}z'(A'VA)^{-1}z\} \tag{4}$$

3) Now that only residual z is considered, so restricted maximum likelihood is also called residual maximum likelihood (REML).

4) The restricted log-likelihood (for residuals $z = A'y$) is

$$l_R(\theta) = c - \frac{1}{2}\log(A'VA) - \frac{1}{2}z'(A'VA)^{-1}z \tag{5}$$

5) The variance components estimation (variance matrix for normal density) is derived from previous class (ML estimation for normal density) as

$$\frac{\partial l_R}{\partial \theta_i} = \frac{1}{2}\left\{(y)'P\frac{\partial V}{\partial \theta_i}Py - \mathrm{tr}(P\frac{\partial V}{\partial \theta_i})\right\}, i = 1, 2, \dots, q. \tag{6}$$

where,

$$P = A(A'VA)^{-1}A' \tag{7}$$

6) Proof idea: in textbook (1.10), replace original components by updated ones after transformation, i.e.,

- ⋄ y by $y^* = z = A'y$
- ⋄ V by $V^* = A'VA$
- ⋄ X by $X^* = A'X = 0$ (crucial!)
- ⋄ **Note** P by $(A'VA)^{-1}$ only (Why X is gone?, since $X^* = 0$)

4.3 A' Selection

(Patterson and Thompson, Section 3) We could always choose

$$A' = I - X(X'X)^{-1}X' \tag{8}$$

5 Example 1: Neyman Scott Problem

$$\begin{bmatrix} y_{1,1} \\ y_{1,2} \\ y_{2,1} \\ y_{2,2} \\ \vdots \\ y_{m,1} \\ y_{m,2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix} + \begin{bmatrix} \epsilon_{1,1} \\ \epsilon_{1,2} \\ \epsilon_{2,1} \\ \epsilon_{2,2} \\ \vdots \\ \epsilon_{m,1} \\ \epsilon_{m,2} \end{bmatrix} \tag{9}$$

Where,

$$X = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{10}$$

Thus A could be

$$A' = \begin{bmatrix} 1 & -1 & 0 & & \dots \\ 0 & 0 & 1 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{11}$$

and

$$A'VA = 2\sigma^2 I_m \tag{12}$$

6 Example 2: One-Way Random Effects ANOVA Model

Data

$$y_{k,j} = \mu + a_k + e_{k,j},\ (k = 1, \dots, K;\ j = 1, \dots, J).$$

We have

$$\begin{bmatrix} y_{1,1} \\ y_{1,2} \\ \vdots \\ y_{1,J} \\ y_{2,1} \\ y_{2,2} \\ \vdots \\ y_{2,J} \\ \vdots \\ y_{K,1} \\ y_{K,2} \\ \vdots \\ y_{K,J} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mu + \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & 0 \\ 1 & \vdots & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & 0 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{bmatrix} + \begin{bmatrix} \epsilon_{1,1} \\ \vdots \\ \epsilon_{1,J} \\ \vdots \\ \epsilon_{K,1} \\ \vdots \\ \epsilon_{K,J} \end{bmatrix} \tag{13}$$

Where,

$$X = ?Z = ? \tag{14}$$

Thus A could be ? (Excercise 1.12 without asymptotic variance-covariance, **Homework 1-2**).

7 References

◇ Herbach, L.H. (1959), Properties of model II-type analysis of variance tests. Annals of Mathematical Statistics, 30, 939-959.

◇ Jiang, J. (2006), Linear and Generalized Linear Mixed Models and Their Applications. Springer.

◇ Patterson, HD and Thompson,R (1972), Recovery of inter-block Information when block sizes are unequal. Biometrika, 58(3), 545-554.

◇ Thompson, W.A. (1962), The problem of negative estimates of variance components. Annals of Mathematical Statistics, 33, 273-289.