Lecture 3: Gaussian Model - Maximum Likelihood Estimation (09-04-2007)

1 Introduction

We study the maximum likelihood for Gaussian models with mixed effects.

2 Maximum Likelihood (ML) Estimation: Gaussian Models

For statistical inference, we make the simple assumption of normal distribution within general linear regression framework.

1 $E(\tilde{y}_n) = \sum_{j=1}^{p} \tilde{x}_j \beta_j$, where $\tilde{x}_j$ is of size $n \times 1$.

2 $\tilde{y}_n = \sum_{j=1}^{p} \tilde{x}_j \beta_j + \tilde{\epsilon}_n$, where error vector $\tilde{\epsilon}_n$ follows a distribution

3 $\tilde{\epsilon}_n \sim$ normal distribution

If y is a random vector of length p, with expected value $E(y) = \mu$ and covariance matrix $Var[y] = V$, the multivariate normal density function is given by

$$f(y) = (2\pi)^{-p/2}|V|^{-1/2}exp(-\frac{1}{2}(y-\mu)^T V^{-1}(y-\mu)) \tag{1}$$

i.e.,

$$l(\beta, \theta) = c - \frac{1}{2}log(|V|) - \frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta) \tag{2}$$

From (1.7) and (1.8) in Text (Jiang, 2006), Y has dimension n, then we have

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= X'V^{-1}y - X'V^{-1}X\beta \\ \frac{\partial l}{\partial \theta_r} &= \frac{1}{2}\left\{(y - X\beta)^T V^{-1}\frac{\partial V}{\partial \theta_r}V^{-1}(y - X\beta) - \mathrm{tr}(V^{-1}\frac{\partial V}{\partial \theta_r})\right\} \end{aligned} \tag{3}$$

where θ_r (r=1,...,q) is the r-th component of V.

3 Asymptotic Covariance Matrix

p11, textbook. (1.11-1.13).

4 Example 1: $V = \sigma^2 I_n$

1) θ=θ_r=σ^2, $r \equiv 1$.

2)

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= X'(\sigma^2)^{-1}y - X'(\sigma^2)^{-1}X\beta \\ \frac{\partial l}{\partial \sigma^2} &= \frac{1}{2}\left\{(y - X\beta)^T(\sigma^2)^{-1}\frac{\partial \sigma^2 I_n}{\partial \sigma^2}(\sigma^2)^{-1}(y - X\beta) - \mathrm{tr}((\sigma^2)^{-1}\frac{\partial(\sigma^2 I_n)}{\partial \sigma^2})\right\} \end{aligned} \tag{4}$$

Let to be zero, i.e.,

3)

$$\begin{aligned} \frac{\partial l}{\partial \beta} &\propto X'y - X'X\beta = 0 \\ \frac{\partial l}{\partial \sigma^2} &\propto (y - X\beta)^T(\sigma^2)^{-2}(y - X\beta) - n(\sigma^2)^{-1} = 0 \end{aligned} \tag{5}$$

4) i.e.,

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y \\ \hat{\sigma^2} &= (y - X\beta)^T(y - X\beta)/n \end{aligned} \tag{6}$$

5) Remark. $\hat{\sigma^2}$ is biased.

6) Remark. This special model does not have random effects.

5 Example 2: Standard Variance Components Model

(1.2) p5, textbook.

$$Y = X\beta + Z_1\alpha_1 + Z_2\alpha_2 + \cdots + Z_q\alpha_q + \epsilon \tag{7}$$

i.e., we have multiple (q) random effects. Where,

1) α_i ~N(0,$\sigma_i^2 I_{m_i}$),($i = 1, ..., q$)

2) ϵ ~N(0,$\sigma_0^2 I_N$). N is number of total observations.

3) α_is and ϵ are all independent.

Then we have

1) $V = \sum_{i=0}^{q} \sigma_i^2 Z_i Z_i'$ ($Z_0 = I_N$)

2) $\theta = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_q^2)'$

3) $\frac{\partial V}{\partial \sigma_i^2} = Z_i Z_i'$, $(i = 0, 1, ..., q)$

The ML solution is from (♠)

$$\begin{aligned} &\beta = (X'V^{-1}X)^{-1}X'V^{-1}y \\ &(y - X\beta)^T V^{-1} Z_i Z_i' V^{-1}(y - X\beta) = \mathrm{tr}(Z_i' V^{-1} Z_i), 0 \le i \le q. \end{aligned} \tag{8}$$

Remark. There may be multiple ML solutions (roots).

6 Example 3: One-way ANOVA with Mixed Effects

6.1 Textbook Part

1) Model: Example 1.1 (Textbook, p4).

2) Solution: p.12 (textbook).

6.2 Solution by Referring to Example 2

Homework 1-1 For **Example 1.1** in textbook, please present (♠) in terms of model parameters. You do not need to derive the exact result as textbook p12.

1) Hint: $(aI_k + bJ_k)^{-1} = I_k/a - bJ_k/[a(a + bk)]$

7 MLE in Multivariate Normal Distribution

y_1,y_2,. . .,y_n are i.i.d. random variable (vector) from multivariate normal distribution $N_p(\mu, V)$. We assume $n \ge p$, $V > 0$. Then

1) Joint density is

$$L(Y; \mu, V) = (2\pi)^{-np/2}|V|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{k=1}^{n}(y_k - \mu)^T V^{-1}(y_k - \mu)\right\} \tag{9}$$

2) We denote

   1) $\bar{y} = \frac{1}{n}\sum_{k=1}^{n} y_k$

   2) $A = \sum_{k=1}^{n}(y_k - \mu)(y_k - \mu)'$,

   3) $S = \sum_{k=1}^{n}(y_k - \bar{y})'(y_k - \bar{y})$

   Remark. $A = S + n(\bar{y} - \mu)(\bar{y} - \mu)'$

3) We have maximum likelihood estimation (MLE) to be

   1) $\hat{\mu} = \bar{y}$

   2) If μ is known, then $\hat{V} = \frac{1}{n}A$

   3) If μ is unknown, then $\hat{V} = \frac{1}{n}S$

   Remark.

4) We have maximum likelihood estimation (MLE) to be

   1) $E(S) = (n-1)V$

   2) $E(\hat{V}) = \frac{n-1}{n}V$ (biased)

8 Appendix: Matrix Algebra: Differentiation

A is a matrix with elements to be functions of θ.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \frac{\partial A}{\partial \theta} = \begin{pmatrix} \partial a_{11}/\partial\theta & \partial a_{12}/\partial\theta \\ \partial a_{21}/\partial\theta & \partial a_{22}/\partial\theta \end{pmatrix}, \quad (10)$$

For vectors a,b,θ and matrix A, we have the following results:

1) (Inner-product): $\frac{\partial(a'b)}{\partial\theta} = (\frac{\partial a'}{\partial\theta})b + (\frac{\partial b'}{\partial\theta})a$

2) (Quadratic form): $\frac{\partial}{\partial x}x'Ax = 2Ax$

3) (Inverse): $\frac{\partial A^{-1}}{\partial\theta_i} = -A^{-1}(\frac{\partial A}{\partial\theta_i})A^{-1}$

4) (Log-linear): A is positive definite: $\frac{\partial}{\partial\theta_i}log(|A|) = tr(A^{-1}\frac{\partial A}{\partial\theta_i})$

9 References

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

2) Zhang, Y.-T. and Fang, K.-T. (1997), Multivariate Analysis, Science Press, Beijing.