Stat 579: Generalized Linear Models and Extensions

Mixed models

Yan Lu
March, 2018, week 8

Restricted Maximum Likelihood (REML)

REML: uses a likelihood function calculated from the transformed set of data, so that nuisance parameters have no effect.

Neyman-Scott problem

Example (1.7), Jiang's book showed that if the number of parameters increases with the sample size, MLE may not be consistent.

- Suppose two observations are collected from m individuals.

individual 1	y_{11}	y_{12}
individual 2	y_{21}	y_{22}
$\vdots$	$\vdots$	
individual m	y_{m1}	y_{m2}

—-Each individual i has its own (unknown) mean μ_i
—-The observations are independent and normally distributed with variance σ^2, want to estimate σ^2.

$$y_{ij} = \mu_i + \epsilon_{ij}, \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2), i = 1, 2, \cdots, m \qquad (1)$$

- $m + 1$ parameters: $\mu_1, \mu_2, \cdots, \mu_m, \sigma$
number of observations $2m$
$m + 1$ is proportional to $2m$

Let

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{m1} \\ y_{m2} \end{pmatrix}, \mathbf{X} = \mathbf{I}_m \otimes \mathbf{1}_2 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix}_{2m \times m}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{m1} \\ \epsilon_{m2} \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix}.$$

Model (5) can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2}$$

$$
\begin{aligned}
\hat{\sigma}^2_{ML} &= \frac{(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})}{2m} \\
&= \sum_{i=1}^{m}\sum_{j=1}^{2}(y_{ij} - \bar{y}_{i\cdot})^2/2m \\
&= \sum_{i=1}^{m}\frac{1}{2}(y_{i1} - y_{i2})^2/2m
\end{aligned}
$$

Let $Z_i = y_{i1} - y_{i2}$, $Z_i \sim N(0, 2\sigma^2)$, Z_is iid

$$
S^2 = \frac{\sum_{i=1}^{m}(Z_i - \bar{Z})^2}{m-1} \xrightarrow{p} 2\sigma^2
$$

Since $\bar{Z} \xrightarrow{p} 0$,

$$
\frac{\sum_{i=1}^{m} Z_i^2}{m-1} \xrightarrow{p} 2\sigma^2
$$

Hence,

$$
\begin{aligned}
\hat{\sigma}^2_{ML} &= \frac{1}{2} \cdot \frac{\sum_{i=1}^{m} Z_i^2}{2m} \\
&= \frac{1}{2} \cdot \frac{\sum_{i=1}^{m} Z_i^2}{m-1} \cdot \frac{m-1}{2m} \\
&\xrightarrow{p} \frac{1}{2} \cdot 2\sigma^2 \cdot \frac{1}{2} = \sigma^2/2
\end{aligned}
$$

Transformation

Consider transformation $Z_i = y_{i1} - y_{i2}$, $Z_i \overset{iid}{\sim} N(0, 2\sigma^2)$
Let

$$\begin{aligned} \mathbf{Z} &= \mathbf{A}'\mathbf{y} \\ &= \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}_{m \times 2m} \mathbf{y} \end{aligned}$$

Recall that

$$\mathbf{X} = \mathbf{I}_m \otimes \mathbf{1}_2 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix}_{2m \times m}$$

such that $\mathbf{A}'\mathbf{X} = \mathbf{0}$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

transforms to

$$\mathbf{A}'\mathbf{y} = \mathbf{A}'\mathbf{X}\boldsymbol{\beta} + \mathbf{A}'\boldsymbol{\epsilon} \text{ i.e. } \mathbf{Z} = \mathbf{A}'\boldsymbol{\epsilon}$$

- REML is to apply a transformation to the data to eliminate the fixed effects
- Then use the transformed data to estimate the variance component.

$$\hat{\sigma}^2_{REML} \xrightarrow{p} \sigma^2$$

Restricted maximum likelihood (REML)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_{p\times 1} + \mathbf{Z}\boldsymbol{\alpha}_{q\times 1} + \boldsymbol{\epsilon}_{n\times 1} \tag{3}$$

- $\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{G})$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R})$, $\boldsymbol{\alpha}$ and $\boldsymbol{\epsilon}$ are independent.
- $Var(\mathbf{y}) = \mathbf{ZGZ}' + \mathbf{R} = \mathbf{V}$
- $rank(\mathbf{X}) = p$

Apply a transformation to eliminate the fixed effects $\boldsymbol{\beta}$, choose $\mathbf{A}_{n\times(n-p)}$, $rank(\mathbf{A}) = n - p$, such that

$$\mathbf{A}'_{(n-p)\times n}\mathbf{X}_{n\times p} = \mathbf{0}$$

(3) becomes

$$\begin{aligned} \mathbf{A}'\mathbf{y} &= \mathbf{A}'\mathbf{X}\boldsymbol{\beta} + \mathbf{A}'\mathbf{Z}\boldsymbol{\alpha} + \mathbf{A}'\boldsymbol{\epsilon} \\ &= \mathbf{0} + \mathbf{A}'\mathbf{Z}\boldsymbol{\alpha} + \mathbf{A}'\boldsymbol{\epsilon} \\ &= \mathbf{A}'(\mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}) \end{aligned}$$

Define $\mathbf{z} = \mathbf{A}'\mathbf{y}$, $\mathbf{z} \sim N(\mathbf{0}, \mathbf{A}'\mathbf{V}\mathbf{A})$

$$f_R(\mathbf{z}) = \frac{1}{(2\pi)^{(n-p)/2}|\mathbf{A}'\mathbf{V}\mathbf{A}|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{A}'\mathbf{V}\mathbf{A})^{-1}\mathbf{z}\right\} \qquad (4)$$

Restricted log-likelihood is

$$l_R(\boldsymbol{\theta}) = C - \frac{1}{2}\log(|\mathbf{A}'\mathbf{V}\mathbf{A}|) - \frac{1}{2}\mathbf{z}'(\mathbf{A}'\mathbf{V}\mathbf{A})^{-1}\mathbf{z}$$

REML estimator of $\boldsymbol{\theta}$ is defined as the maximizer of (4).

Remarks

- REML estimator doesn't depend on $\mathbf{A}$
- Choice of $\mathbf{A}$ is not unique, but the results will be the same
- The REML method is a method of estimating $\boldsymbol{\theta}$ (not $\boldsymbol{\beta}$, $\boldsymbol{\beta}$ is eliminated before estimation)
- After $\hat{\boldsymbol{\theta}}_R$ obtained, $\boldsymbol{\beta}$ is usually estimated the same way as the ML, such estimators are referred as the "REML estimator" of $\boldsymbol{\beta}$

Example: one-way balanced random effects model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, i = 1, 2, \cdots, m, j = 1, \cdots, k, n = mk$$

$\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma_\alpha^2 \mathbf{I}_m)$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$
$\boldsymbol{\alpha}$ and $\boldsymbol{\epsilon}$ independent

Source	df	SS	MS
Between groups (SSA)	$m-1$	$\sum_{i=1}^m k(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	SSA/(m-1)
Within groups (SSE)	$n-m$	$\sum_{i,j}(y_{ij} - \bar{y}_{i\cdot})^2$	SSE/(n-m)
total	$n-1$	$\sum_{i,j}(y_{ij} - \bar{y}_{\cdot\cdot})^2$	

Transformation $\mathbf{z} = \mathbf{A}'\mathbf{y}$ for any $\mathbf{A}$ that makes $\mathbf{A}'\mathbf{X} = \mathbf{0}$

$$\begin{aligned} l_R &= -\frac{1}{2}(mk-1)\log 2\pi - \frac{1}{2}\log(mk) - \frac{1}{2}m(k-1)\log\sigma^2 \\ &\quad -\frac{1}{2}(m-1)\log(\sigma^2 + k\sigma_\alpha^2) - \frac{SSE}{2\sigma^2} - \frac{SSA}{2(\sigma^2 + k\sigma_\alpha^2)} \end{aligned}$$

REML equations

$$\frac{\partial l_R}{\partial \sigma^2} = 0, \frac{\partial l_R}{\partial \sigma_\alpha^2} = 0$$

$$\begin{cases} \sigma^2 + k\sigma_\alpha^2 = MSA \\ \sigma^2 = MSE \end{cases}$$

ML and REML estimators for one way random effects model

	ML Estimator	REML Estimator
$\hat{\sigma}^2$	MSE	MSE
$\hat{\sigma}^2_\alpha$	$max\left\{0, k^{-1}\left(\frac{m-1}{m}MSA - MSE\right)\right\}$	$max\left\{0, k^{-1}(MSA - MSE)\right\}$

Comments:

- Both have the same merits of being based on the likelihood principle which leads to useful properties such as consistency, asymptotic normality and efficiency.
- For balanced mixed ANOVA models, the REML estimates for the variance components are identical to classical ANOVA-type estimates obtained from solving the equations which set mean squares equal to their expectations
- Generally REML estimates of variance components are preferred
- REML estimators are not guaranteed to be unbiased, but they are usually less biased than ML estimators
- In general, expect results from ML and REML estimates to differ more as the number of p of fixed effects in the model increase

Analysis of variance estimation (MOM)

Basic idea came from the method of moments, let $\mathbf{Q}$ be the q-dimentional vector whose components are quadratic functions of the data. The ANOVA estimators of the variance components are obtained by solving the system of equation $E(\mathbf{Q}) = \mathbf{Q}$

Source	df	SS	MS
Between groups (SSA)	$m-1$	$\sum_{i=1}^{m} k(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	SSA/(m-1)
Within groups (SSE)	$n-m$	$\sum_{i,j}(y_{ij} - \bar{y}_{i\cdot})^2$	SSE/(n-m)
total	$n-1$	$\sum_{i,j}(y_{ij} - \bar{y}_{\cdot\cdot})^2$	

$$\begin{cases} E(SSA) = (m-1)k\sigma_\alpha^2 + (m-1)\sigma^2 \\ E(SSE) = m(k-1)\sigma^2 \end{cases}$$

ANOVA estimating equations are

$$\begin{cases} SSA = (m-1)k\sigma_\alpha^2 + (m-1)\sigma^2 \\ SSE = m(k-1)\sigma^2 \end{cases}$$

The ANOVA estimators are

$$\hat{\sigma}^2 = MSE$$

$$\hat{\sigma}_\alpha = max\left\{0, k^{-1}(MSA - MSE)\right\}$$

same as REML estimator

Interpretation

Consequence of assuming random

- statistical inferences can be made to the population from which the group effects were drawn
- Random effects induce a correlation among observations with the same group effect
- Using random effects involves making extra assumptions but often results in more precise estimates

Fixed and random effects

- Fixed effects are interpreted as usual. Multiple comparisons of the groups are still of interest.
- Estimates of the random effects α_i, within group, are interpreted similarly. But we are not particularly interested in the specific level that was randomly selected from population.
- Larger estimates of σ_α^2 means importance of the random effects. Commonly, we are interested in the proportion $\hat{\sigma}_\alpha^2/(\hat{\sigma}_\alpha^2 + \hat{\sigma}^2)$

Tests in Gaussian Mixed Models

Example: Animal behavior study (LONSDORF, ET AL (2004) Nature, 428, 715-716)

- Objective: to study the learning of termite fishing by young chimpanzees
- study Design: follow chimpanzee mothers over four field seasons in Gombe National park, Tanzania
- Primary Observation: age at which termite fishing skill has been acquired by the mother's offspring
- Results: female offspring acquired the skill at a younger age (31 v.s 58 months) and spent more time fishing, male offspring preferred to play

Can we apply mixed model to the data?

- Offspring of the same mother are likely to be correlated, need to account for this correlation
- Some mothers termite fish more skillfully than others
  ——offspring of these mothers might acquire the skill at an earlier age
  ——we should account for these "mother effects"
  ——consider as random
- Also account for gender as fixed effects

Mixed model

$$y_{ij} = (age)_{ij} = \beta_0 + \beta_1 sex + \gamma_i + \epsilon_{ij}$$

- i: chimp mother, j: jth offspring
- mother-specific effect: $\gamma_i \sim N(0, \sigma_\gamma^2)$
  $\epsilon_{ij} \sim N(0, \sigma^2)$

Interested Tests

- $$H_0 : \beta_1 = 0 \text{ v.s } \beta_1 \neq 0$$

  If the sex is associated with age of termite fishing

- $$H_0 : \sigma_\gamma^2 = 0 \text{ v.s } \sigma_\gamma^2 > 0$$

  If there is a mother-specific effect

Example: one way random effects model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, i = 1, 2, \cdots, m, j = 1, \cdots, n_i \tag{5}$$

Let $\sum_i n_i = n$, for simplicity, let $n_1 = n_2 = \cdots n_m = k, n = mk$
Let

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1k} \\ y_{21} \\ \vdots \\ y_{2k} \\ \vdots \\ y_{m1} \\ \vdots \\ y_{mk} \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1k} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2k} \\ \vdots \\ \epsilon_{m1} \\ \vdots \\ \epsilon_{mk} \end{pmatrix}, \mathbf{Z} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \end{pmatrix}_{mk \times m}$$

Let $\beta = \mu$,

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{mk \times 1}$$

One way random effects model (5) can be written as

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$$

- $\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma_\alpha^2 \mathbf{I}_m)$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{mk})$
  $\boldsymbol{\alpha}$ and $\boldsymbol{\epsilon}$ independent
- Fixed: μ
  Random: α_i related to σ_α^2 and ϵ_{ij} related to σ^2

F test

F-test

$$H_0 : \sigma_\alpha^2 = 0 \text{ v.s } \sigma_\alpha^2 > 0$$

ANOVA table

Source	df	SS	MS
Between groups (SSB)	$m-1$	$\sum_{i=1}^{m} n_i(\bar{y}_{i.} - \bar{y}_{..})^2$	SSB/(m-1)
Within groups (SSE)	$n-m$	$\sum_{i,j}(y_{ij} - \bar{y}_{i.})^2$	SSE/(m(k-1))
total	$n-1$	$\sum_{i,j}(y_{ij} - \bar{y}_{..})^2$	

- $\dfrac{MSB/(\sigma^2 + k\sigma_\alpha^2)}{MSE/\sigma^2}$ has an F distribution with degrees of freedom $(m-1, m(k-1))$.
- Under H_0, MSB/MSE has an F distribution with degrees of freedom $(m-1, m(k-1))$.

Some notations

Projection matrix: $\mathbf{P_X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

$\mathbf{Z} \ominus \mathbf{X} = \mathbf{P_{X^\perp}}\mathbf{Z}$

$\mathbf{P_{X^\perp}} = \mathbf{I} - \mathbf{P_X}$

F test for mixed effects model

Suppose for the general mixed model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$$

Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2', \cdots, \boldsymbol{\alpha}_s')'$ with $\boldsymbol{\alpha}_i \sim N(\mathbf{0}, \sigma_i^2\mathbf{I})$,
$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_s)$ with $\mathbf{Z}_i$ associated with random vector $\boldsymbol{\alpha}_i$.
One wishes to test the hypothesis

$$H_0 : \sigma_1^2 = 0 \text{ v.s } \sigma_1^2 > 0$$

Write the mixed model as the following

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{\alpha}_1 + \mathbf{Z}_{-1}\boldsymbol{\alpha}_{-1} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\alpha}_{-1} = (\boldsymbol{\alpha}_2', \cdots, \boldsymbol{\alpha}_s')'$, $\mathbf{Z}_{-1} = (\mathbf{Z}_2, \cdots, \mathbf{Z}_s)$

Consider the quadratic form (Jiang, chapter 2)

$$\begin{aligned} q_1 &= \sigma^{-2}\mathbf{y}' P_{\mathbf{Z}_1 \ominus (\mathbf{X},\mathbf{Z}_{-1})}\mathbf{y} \\ &= \mathbf{y}' \left\{ P_{\mathbf{Z}_1 \ominus (\mathbf{X},\mathbf{Z}_{-1})}/\sigma^2 \right\} \mathbf{y} \end{aligned}$$

$$\begin{aligned} q_2 &= \sigma^{-2}\mathbf{y}' P_{(\mathbf{X},\mathbf{Z})^\perp}\mathbf{y} \\ &= \mathbf{y}' \left\{ P_{(\mathbf{X},\mathbf{Z})^\perp}/\sigma^2 \right\} \mathbf{y} \end{aligned}$$

$$q_1 \sim \chi^2_{r1}, r_1 = rank\left\{(\mathbf{X},\mathbf{Z})\right\} - rank\left\{(\mathbf{X},\mathbf{Z}_{-1})\right\}$$

$$q_2 \sim \chi^2_{r2}, r_2 = n - rank\left\{(\mathbf{X},\mathbf{Z})\right\}$$

q_1 and q_2 are independent,

$$F = \frac{q_1/r_1}{q_2/r_2} \sim F_{r_1,r_2}$$

Likelihood ratio test

- based on the likelihood ratio, which expresses how many times more likely the data under one model than the other
- used to compare the fit of two models

Let $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_n)$, where $Y_1, Y_2, \cdots, Y_n$ have joint pdf $f(\mathbf{y}, \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Omega}$, and consider the hypothesis

$$H_0 : \boldsymbol{\theta} \in \boldsymbol{\Omega}_0 \text{ v.s. } H_\alpha : \boldsymbol{\theta} \in \boldsymbol{\Omega} - \boldsymbol{\Omega}_0$$

The generalized likelihood ratio (GLR) is defined by

$$\lambda(\mathbf{y}) = \frac{\max_{\boldsymbol{\theta} \in \boldsymbol{\Omega}_0} f(\mathbf{y}; \boldsymbol{\theta})}{\max_{\boldsymbol{\theta} \in \boldsymbol{\Omega}} f(\mathbf{y}; \boldsymbol{\theta})} = \frac{f(\mathbf{y}; \hat{\boldsymbol{\theta}}_0)}{f(\mathbf{y}; \hat{\boldsymbol{\theta}})}$$

- $\hat{\boldsymbol{\theta}}$ denote the usual MLE of $\boldsymbol{\theta}$
- $\hat{\boldsymbol{\theta}}_0$ denotes the MLE under the restriction that H_0 is true.
- if $\mathbf{y} \sim f(\mathbf{y}; \theta_1, \cdots, \theta_k)$, then under $H_0 : (\theta_1, \theta_2, \cdots, \theta_r) = (\theta_{10}, \theta_{20}, \cdots, \theta_{r0}), r < k$. Approximately, for large n, $-2\log\lambda(\mathbf{y}) \sim \chi^2(r)$ (Hartley and Rao (1967), Jiang (2005c) provided rigorous proof)
- An appropriate size α test is to reject H_0 if

$$-2\log\lambda(\mathbf{y}) \geq \chi^2_{1-\alpha}(r)$$

One way Random Effects Model

$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, i = 1, \cdots, m, j = 1, \cdots, k.$
$\alpha_i \sim N(0, \sigma_\alpha^2)$, $\epsilon_{ij} \sim N(0, \sigma^2)$

$$\begin{aligned} l(\mu, \sigma_\alpha^2, \sigma^2) &= c - \frac{1}{2}(n-m)\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{m}\log(\sigma^2 + k\sigma_\alpha^2) \\ &\quad - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{k}(y_{ij} - \mu)^2 + \frac{\sigma_\alpha^2}{2\sigma^2}\sum_{i=1}^{m}\frac{k^2}{\sigma^2 + k\sigma_\alpha^2}(\bar{y}_{i\cdot} - \mu)^2 \end{aligned}$$

Under

$$H_0 : \sigma_\alpha^2 = 0 \text{ v.s } \sigma_\alpha^2 > 0$$

$$l(\mu, 0, \sigma^2) = c - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{k}(y_{ij} - \mu)^2$$

Set partial derivatives equal to 0

$$\frac{\partial l}{\partial \mu} = \frac{1}{2\sigma^2} \sum_{i=1}^{m} \sum_{j=1}^{k} 2(y_{ij} - \mu) = 0$$

$$\begin{aligned} \frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2} \cdot \frac{1}{\sigma^2} - \frac{-\sum_{i=1}^{m} \sum_{j=1}^{k} 2(y_{ij} - \mu)^2}{4\sigma^4} \\ &= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{m} \sum_{j=1}^{k} (y_{ij} - \mu)^2}{2\sigma^4} \end{aligned}$$

$$\hat{\mu} = \bar{y}_{..}$$

$$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^{m} \sum_{j=1}^{k} (y_{ij} - \bar{y}_{..})^2}{n}$$

Without any restriction

$$\hat{\mu} = \bar{y}_{..}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{m}\sum_{j=1}^{k}(y_{ij} - \bar{y}_{.})^2}{m(k-1)} = MSE$$

$$\hat{\sigma}_\alpha^2 = \frac{1}{k}\left[\frac{k\sum_{i=1}^{m}(\bar{y}_{i.} - \bar{y}_{..})^2}{m} - \hat{\sigma}^2\right]$$

- $-2\log\lambda(\mathbf{y})$ can be calculated
- However, under $H_0 : \sigma_\alpha^2 = 0$ v.s $\sigma_\alpha^2 > 0$, $-2\log\lambda(\mathbf{y})$ is not distributed as χ_1^2.

  It could be approximated by a mixture of $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$
- If there are no boundry issue, say $H_0 : \sigma_\alpha^2 = 2$ v.s $\sigma_\alpha^2 \neq 2$ $-2\log\lambda(\mathbf{y})$ is distributed as χ_1^2.