Introduction to General and Generalized Linear Models

Mixed effects models - Part II

Henrik Madsen
Poul Thyregod

Informatics and Mathematical Modelling
Technical University of Denmark
DK-2800 Kgs. Lyngby

January 2011

This lecture

- One-way random effects model, continued
- More examples of hierarchical variation
- General linear mixed effects models

Estimation of parameters

Confidence interval for the variance ratio

In the balanced case one may construct a *confidence interval for the variance ratio* γ. A $1-\alpha$ confidence interval for γ, i.e. an interval (γ_L, γ_U), satisfying

$$P[\gamma_L < \gamma < \gamma_U] = 1 - \alpha$$

is obtained by using

$$\gamma_L = \frac{1}{n}\left(\frac{Z}{F(k-1, N-k)_{1-\alpha/2}} - 1\right)$$
$$\gamma_U = \frac{1}{n}\left(\frac{Z}{F(k-1, N-k)_{\alpha/2}} - 1\right)$$

Estimation of parameters

Theorem (Moment estimates in the random effects model)

Moment estimates for the parameters μ, σ^2 and σ_u^2 are

$$\widetilde{\mu} = \overline{\overline{Y}}_{..}$$

$$\widetilde{\sigma}^2 = \mathrm{SSE}/(N-k)$$

$$\widetilde{\sigma_u}^2 = \frac{\mathrm{SSB}/(k-1) - \mathrm{SSE}/(N-k)}{n_0} = \frac{\mathrm{SSB}/(k-1) - \widetilde{\sigma}^2}{n_0}$$

where the weighted average group size n_0 is given by

$$n_0 = \frac{\sum_1^k n_i - \left(\sum_1^k n_i^2 / \sum_1^k n_i\right)}{k-1} = \frac{N - \sum_i n_i^2/N}{k-1}$$

Estimation of parameters

Distribution of “residual” sum of squares

In the balanced case we have that

$$\mathrm{SSE} \sim \sigma^2 \chi^2(k(n-1))$$
$$\mathrm{SSB} \sim \{\sigma^2 / w(\gamma)\} \chi^2(k-1)$$

and that SSE and SSB are independent.

Estimation of parameters

Unbiased estimates for variance ratio in the balanced case

In the balanced case, $n_1 = n_2 = \cdots = n_k = n$, we can provide explicit unbiased estimators for γ and $w(\gamma) = 1/(1 + n\gamma)$. One has

$$\widetilde{w} = \frac{\mathrm{SSE}\,/\{k(n-1)\}}{\mathrm{SSB}\,/(k-3)}$$

$$\widetilde{\gamma} = \frac{1}{n}\left(\frac{\mathrm{SSB}\,/(k-1)}{\mathrm{SSE}\,/\{k(n-1)-2\}} - 1\right)$$

are *unbiased estimators* for $w(\gamma) = 1/(1 + n\gamma)$ and for $\gamma = \sigma_u^2/\sigma^2$, respectively.

Example - Wool data

Variation	Sum of squares		f	$s^2 = SS/f$	$\mathrm{E}[S^2]$
Between bales	SSB	65.9628	6	10.9938	$\sigma^2 + 4\sigma_u^2$
Within bales	SSE	131.4726	21	6.2606	σ^2

Table: ANOVA table for the baled wool data.

Example - Wool data

The test statistic for the hypothesis $H_0 : \sigma_u^2 = 0$, is

$$z = \frac{10.9938}{6.2606} = 1.76 < F_{0.95}(6, 21) = 2.57$$

The p-value is $P[F(6, 21) \geq 1.76] = 0.16$

Thus, the test fails to reject the hypothesis of no variation between the purity of the bales when testing at a 5% significance level. However, as the purpose is to describe the variation in the shipment, we will estimate the parameters in the random effects model, irrespective of the test result.

Example - Wool data

Now lets find a 95% confidence interval for the ratio $\gamma = \sigma_u^2/\sigma^2$. As $F(6,21)_{0.025} = 1/F(21,6)_{0.975}$, one finds the interval

$$\gamma_L = \frac{1}{4}\left(\frac{1.76}{F(6,21)_{0.975}} - 1\right) = 0.25 \times \left(\frac{1.76}{3.09} - 1\right) = -0.11$$

$$\gamma_U = \frac{1}{4}\left(\frac{1.76}{F(6,21)_{0.025}} - 1\right) = 0.25 \times (1.76 \times 5.15 - 1) = 2.02$$

Maximum likelihood estimates

Theorem (Maximum likelihood estimates for the parameters under the random effects model)

The maximum likelihood estimates for μ, σ^2 *and* $\sigma_u^2 = \sigma^2\gamma$ *are determined by*

- *For* $\sum_i n_i^2(\overline{y}_{i\cdot} - \overline{\overline{y}}_{\cdot\cdot})^2 < \mathrm{SSE} + \mathrm{SSB}$ *one obtains*

$$\widehat{\mu} = \overline{\overline{y}}_{\cdot\cdot} = \frac{1}{N}\sum_i n_i \overline{y}_{i\cdot}$$

$$\widehat{\sigma}^2 = \frac{1}{N}(\mathrm{SSE} + \mathrm{SSB})$$

$$\widehat{\gamma} = 0$$

Maximum likelihood estimates

Theorem (Maximum likelihood estimates for the parameters under the random effects model continued)

- *For* $\sum_i n_i^2(\overline{y}_{i\cdot} - \overline{\overline{y}}_{\cdot\cdot})^2 > \mathrm{SSE} + \mathrm{SSB}$ *the estimates are determined as solution to*

$$\widehat{\mu} = \frac{1}{W(\widehat{\gamma})} \sum_{i=1}^{k} n_i w_i(\widehat{\gamma}) \overline{y}_{i\cdot}$$

$$\widehat{\sigma}^2 = \frac{1}{N} \left\{ \mathrm{SSE} + \sum_{i=1}^{k} n_i w_i(\widehat{\gamma})(\overline{y}_{i\cdot} - \widehat{\mu})^2 \right\}$$

$$\frac{1}{W(\widehat{\gamma})} \sum_{i=1}^{k} n_i^2 w_i(\widehat{\gamma})^2 (\overline{y}_{i\cdot} - \widehat{\mu})^2 = \frac{1}{N} \left\{ \mathrm{SSE} + \sum_{i=1}^{k} n_i w_i(\widehat{\gamma})(\overline{y}_{i\cdot} - \widehat{\mu})^2 \right\}$$

where

$$W(\gamma) = \sum_{i=1}^{k} n_i w_i(\gamma).$$

Maximum likelihood estimates

The maximum likelihood estimate $\widehat{\mu}$ is a weighted average of the group averages

$\widehat{\mu}$ is a weighted average of the group averages, $\overline{y}_{i\cdot}$, with the estimates for the marginal precisions

$$\sigma^2 n_i w_i(\gamma) = \frac{\sigma^2}{\mathrm{Var}[\overline{Y}_{i\cdot}]}$$

as weights. We have the marginal variances

$$\mathrm{Var}[\overline{Y}_{i\cdot}] = \sigma_u^2 + \frac{\sigma^2}{n_i} = \frac{\sigma^2}{n_i}(1 + n_i\gamma) = \frac{\sigma^2}{n_i w_i(\gamma)}$$

When the experiment is *balanced*, i.e. when $n_1 = n_2 = \cdots = n_k$, then all weights are equal, and one obtains the simple result that $\widehat{\mu}$ is the crude average of the group averages.

Maximum likelihood estimates

The estimate for $\boldsymbol{\sigma}^2$ utilizes also the variation between groups

We observe that the estimate for σ^2 is not only based upon the variation within groups, SSE, but the estimate does also utilize the knowledge of the variation between groups, as

$$\mathrm{E}[(\overline{Y}_{i\cdot} - \mu)^2] = \mathrm{Var}[\overline{Y}_{i\cdot}] = \frac{\sigma^2}{n_i w_i(\gamma)}$$

and therefore, the terms $(\overline{y}_{i\cdot} - \mu)^2$ contain information about σ^2 as well as γ.

Maximum likelihood estimates

The estimate for σ^2 is not necessarily unbiased

We observe further that – as usual with ML-estimates of variance – the estimate for σ^2 is not necessarily unbiased.

Instead of the maximum likelihood estimate above, it is common practice to adjust the estimate. Later we shall introduce the so-called *residual maximum likelihood* (REML) estimates for σ^2 and σ_u^2, obtained by considering the distribution of the residuals.

Maximum likelihood estimates

Maximum-likelihood-estimates in the balanced case

In the balanced case, $n_1 = n_2 = \cdots = n_k$ the weights

$$w_i(\gamma) = \frac{1}{1 + n\gamma}$$

do not depend on i, and then

$$\widehat{\mu} = \frac{1}{k} \sum_{i=1}^{k} \overline{y}_{i+} = \overline{\overline{y}}_{++},$$

which is the same as the moment estimate.
When $(n-1)\,\mathrm{SSB} > \mathrm{SSE}$ then the maximum likelihood estimate corresponds to an inner point.

Maximum likelihood estimates

Maximum-likelihood-estimates in the balanced case, continued

$$N\sigma^2 = \mathrm{SSE} + \frac{\mathrm{SSB}}{1+n\gamma}$$

$$N\frac{n}{1+n\gamma}\;\frac{\mathrm{SSB}}{k} = \mathrm{SSE} + \frac{\mathrm{SSB}}{1+n\gamma}$$

with the solution

$$\widehat{\sigma}^2 = \frac{\mathrm{SSE}}{N-k}$$

$$\widehat{\gamma} = \frac{1}{n}\left[\frac{\mathrm{SSB}}{k\widehat{\sigma}^2} - 1\right]$$

$$\widehat{\sigma}_b^2 = \frac{\mathrm{SSB}/k - \widehat{\sigma}^2}{n}$$

Estimation of random effects, BLUP-estimation

- In a mixed effects model, it is not clear what fitted values, and residuals are.
- Our best prediction for subject i is not given by the mean relationship, μ.
- It may sometimes be of interest to estimate the random effects.
- The *best linear unbiased predictor* (BLUP) in the one-way case is

$$\mu_i = \Big(1 - w_i(\gamma)\Big)\overline{y}_i + w_i(\gamma)\mu$$

- Thus, the estimate for μ_i is a weighted average between the individual bale averages, $\overline{y}_i$ and the overall average $\widehat{\mu}$ with weights $(1 - w_i(\gamma))$ and $w_i(\gamma)$, where

$$w_i(\gamma) = \frac{1}{1 + n_i\gamma}$$

General linear mixed effects models

Definition (Linear mixed effects model)

The model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{U} + \boldsymbol{\epsilon}$$

with $\boldsymbol{X}$ and $\boldsymbol{Z}$ denoting known matrices, and where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{U} \sim N(\mathbf{0}, \boldsymbol{\Psi})$ are independent is called a *mixed general linear model*. In the general case may the covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ depend on some unknown parameters, ψ, that also need to be estimated.
The parameters $\boldsymbol{\beta}$ are called *fixed effects* or *systematic effects*, while the quantities $\boldsymbol{U}$ are called *random effects*.

General linear mixed effects models

It follows from the independence of $\boldsymbol{U}$ and $\boldsymbol{\epsilon}$ that

$$\mathrm{D}\left[\begin{pmatrix}\boldsymbol{\epsilon}\\ \boldsymbol{U}\end{pmatrix}\right] = \begin{pmatrix}\boldsymbol{\Sigma} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Psi}\end{pmatrix}$$

The model may also be interpreted as a *hierarchical model*

$$\boldsymbol{U} \sim N(\boldsymbol{0}, \boldsymbol{\Psi})$$
$$\boldsymbol{Y}|\boldsymbol{U} = \boldsymbol{u} \sim N(\boldsymbol{X\beta} + \boldsymbol{Zu}, \boldsymbol{\Sigma})$$

General linear mixed effects models

The *marginal distribution* of $\boldsymbol{Y}$ is a normal distribution with

$$\mathrm{E}[\boldsymbol{Y}] = \boldsymbol{X}\boldsymbol{\beta}$$
$$\mathrm{D}[\boldsymbol{Y}] = \boldsymbol{\Sigma} + \boldsymbol{Z}\boldsymbol{\Psi}\boldsymbol{Z}^T$$

We shall introduce the symbol $\boldsymbol{V}$ for the dispersion matrix in the marginal distribution of $\boldsymbol{Y}$, i.e.

$$\boldsymbol{V} = \boldsymbol{\Sigma} + \boldsymbol{Z}\boldsymbol{\Psi}\boldsymbol{Z}^T$$

The matrix $\boldsymbol{V}$ may grow rather large and cumbersome to handle.

One-way model with random effects - example

The one-way model with random effects

$$Y_{ij} = \mu + U_i + e_{ij}$$

We can formulate this as

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{U} + \boldsymbol{\epsilon}$$

with

$$\begin{aligned}
\boldsymbol{X} &= \mathbf{1}_N \\
\beta &= \mu \\
\boldsymbol{U} &= (U_1, U_2, \dots, U_k)^T \\
\boldsymbol{\Sigma} &= \sigma^2 \boldsymbol{I}_N \\
\boldsymbol{\Psi} &= \sigma_u^2 \boldsymbol{I}_k
\end{aligned}$$

where $\mathbf{1}_N$ is a column of 1's. The i, j'th element in the $N \times k$ dimensional matrix $\boldsymbol{Z}$ is 1, if y_{ij} belongs to the i'th group, otherwise it is zero.

Estimation of fixed effects and variance parameters

- The fixed effect parameters $\boldsymbol{\beta}$ and the variance parameters $\boldsymbol{\psi}$ are estimated from the marginal distribution of $\boldsymbol{Y}$.
- For fixed $\boldsymbol{\psi}$ the estimate of $\boldsymbol{\beta}$ is found as the solution of

$$(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})\boldsymbol{\beta} = \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}$$

- This is the well-known *weighted least squares (WLS)* formula.
- In some software systems the solution is called the generalised least squares (GLS).
- Note, however, that the solution may depend on the unknown variance parameters $\boldsymbol{\psi}$ as we saw in the case of the unbalanced one-way random effect model.

Estimation of fixed effects and variance parameters

- The observed Fisher information for $\boldsymbol{\beta}$ is

$$I(\widehat{\boldsymbol{\beta}}) = \boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X}$$

- An estimate for the dispersion matrix for $\widehat{\boldsymbol{\beta}}$ is determined as

$$\text{Var}[\widehat{\beta}] = (\boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X})^{-1}$$

Estimation of fixed effects and variance parameters

- In order to determine estimates for the variance parameters $\boldsymbol{\psi}$ we shall modify the profile likelihood for $\boldsymbol{\psi}$ in order to compensate for the estimation of $\boldsymbol{\beta}$

- The *modified profile log-likelihood* is

$$\ell_m(\boldsymbol{\psi}) = -\frac{1}{2}\log|\boldsymbol{V}| - \frac{1}{2}\log|\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}| - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$$

- When $\widehat{\boldsymbol{\beta}}$ depends on $\boldsymbol{\psi}$ it is necessary to determine the solution by iteration.

Estimation of fixed effects and variance parameters

- The modification to the profile likelihood equals the so-called *residual maximum likelihood* (REML)-method using the marginal distribution of the *residual* $(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\psi})$.
- In REML the problem of biased variance components is solved by setting the fixed effects estimates equal to the WLS solution above in the likelihood function and then maximising it to find the variance component terms only.
- The reasoning is that the fixed effects cannot contribute with information on random effects leading to a justification of not estimating these parameters in the same likelihood.
- The method is also termed *restricted maximum likelihood* method because the model may be embedded in a more general model for the group observation vector $\boldsymbol{Y}_i$ where the random effects model *restricts* the correlation coefficient in the general model.

Estimation of fixed effects and variance parameters

- It is observed that the REML-estimates are obtained by minimising

$$(\boldsymbol{Y}-\boldsymbol{X\beta})^T\boldsymbol{V}^{-1}(\boldsymbol{\psi})(\boldsymbol{Y}-\boldsymbol{X\beta})+\log|\boldsymbol{V}^{-1}(\boldsymbol{\psi})|+\log|\boldsymbol{X}^T\boldsymbol{V}^{-1}(\boldsymbol{\psi})\boldsymbol{X}|$$

- A comparison with the full likelihood function in shows that it is the variance term $\log|\boldsymbol{X}^T\boldsymbol{V}^{-1}(\boldsymbol{\psi})\boldsymbol{X}|$ which is associated with the estimation of $\boldsymbol{\beta}$ that causes the REML estimated variance components to be unbiased.
- If accuracy of estimates of the variance terms are of greater importance than bias then the full maximum likelihood should be considered instead.
- An optimal weighting between bias and variance of estimators is obtained by the estimators optimising the so-called *Godambe Information*
- In balanced designs REML gives the classical moment estimates of variance components (constrained to be non-negative).

Estimation of random effects

- Formally, the random effects, $\boldsymbol{U}$ are not parameters in the model, and the usual likelihood approach does not make much sense for "estimating" these random quantities.
- It is, however, often of interest to assess these "latent", or "state" variables.
- We formulate a so-called *hierarchical likelihood* by writing the joint density for observable as well as unobservable random quantities.
- By putting the derivative of the hierarchical likelihood equal to zero and solving with respect to $\boldsymbol{u}$ one finds that the estimate $\widehat{\boldsymbol{u}}$ is solution to

$$(\boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} + \boldsymbol{\Psi}^{-1})\boldsymbol{u} = \boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

  where the estimate $\widehat{\boldsymbol{\beta}}$ is inserted in place of $\boldsymbol{\beta}$.
- The solution is termed the *best linear unbiased predictor*

Simultaneous estimation of $\boldsymbol{\beta}$ and $\boldsymbol{u}$

- The estimates for $\boldsymbol{\beta}$ and for $\boldsymbol{u}$ are those values that simultaneously maximize $\ell(\boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{u})$ for a fixed value of $\boldsymbol{\psi}$.
- The *mixed model equations* are

$$\begin{pmatrix} \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} + \boldsymbol{\Psi}^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \boldsymbol{u} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y} \end{pmatrix}$$

- The equations facilitate the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{u}$ without calculation of the marginal variance $\boldsymbol{V}$, or its inverse.
- The estimation may be performed by an *iterative back-fitting algorithm*.

Interpretation as empirical Bayes estimate

It is seen from

$$(\boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}+\boldsymbol{\Psi}^{-1})\boldsymbol{u}=\boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})$$

that the BLUP-estimate $\widehat{\boldsymbol{u}}$ for the random effects has been “shrunk” towards zero, as it is a weighted average of the direct estimate, $(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})$, and the prior mean, $\mathrm{E}[\boldsymbol{U}]=\boldsymbol{0}$, where the weights are the precision $\boldsymbol{\Psi}^{-1}$ in the distribution of $\boldsymbol{U}$.