KVL – Department of Mathematics & Physics – St@tmaster

http://www.matfys.kvl.dk

Module 4: Mixed model theory I

Definition of mixed models

Anders Nielsen

anielsen@dina.kvl.dk

Aim of this module

- Define random effects models in some details
- See wha t's "inside" a mixed model
  - Write down the Likelihood
  - See how model parameters are estimated

Hopefully this will give a deeper understanding of mixed models, and an improved ability to interpret results from these models

Design matrix for a systematic linear model

Consider first the well known fixed effects two way ANOVA:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{ i.i.d. } N(0, \sigma^2), \quad i = 1, 2, \quad j = 1, 2, 3.$$

An expanded view of this model is:

$$\begin{array}{llllllll} y_{11} = \mu & + \alpha_1 & & + \beta_1 & & & + \varepsilon_{11} \\ y_{21} = \mu & & + \alpha_2 & + \beta_1 & & & + \varepsilon_{21} \\ y_{12} = \mu & + \alpha_1 & & & + \beta_2 & & + \varepsilon_{12} \\ y_{22} = \mu & & + \alpha_2 & & + \beta_2 & & + \varepsilon_{22} \\ y_{13} = \mu & + \alpha_1 & & & & + \beta_3 & + \varepsilon_{13} \\ y_{23} = \mu & & + \alpha_2 & & & + \beta_3 & + \varepsilon_{23} \end{array} \tag{1}$$

The exact same in matrix notation:

$$\underbrace{\begin{pmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \\ y_{13} \\ y_{23} \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{13} \\ \varepsilon_{23} \end{pmatrix}}_{\boldsymbol{\varepsilon}} \tag{2}$$

$$\underbrace{\begin{pmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \\ y_{13} \\ y_{23} \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{13} \\ \varepsilon_{23} \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$

- $\mathbf{y}$ is the vector of all observations
- $\mathbf{X}$ is the known as the *design matrix*
- $\boldsymbol{\beta}$ is the vector of fixed effect parameters
- $\boldsymbol{\varepsilon}$ is a vector of independent $N(0, \sigma^2)$ "measurement errors"
  - The vector $\boldsymbol{\varepsilon}$ is said to follow a *multivariate normal distribution*
  - Mean vector $\mathbf{0}$
  - Covariance matrix $\sigma^2\mathbf{I}$
  - Written as: $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$
- $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ specifies the model, and everything can be calculated from $\mathbf{y}$ and $\mathbf{X}$.

Construction of the design matrix

In a general systematic linear model (with both factors and covariates), it is surprisingly easy to construct the design matrix $\mathbf{X}$.

- For each factor: Add one column for each level, with ones in the rows where the corresponding observation is from that level, and zeros otherwise.
- For each covariate: Add one column with the measurements of the covariate.
- Remove linear dependencies (if necessary)

Example: linear regression:

$$y_i = \alpha + \beta \cdot x_i + \varepsilon$$

In matrix notation:

$$\mathbf{y} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \boldsymbol{\varepsilon}$$

The mixed linear model

Consider now the one way ANOVA with random block effect:

$$y_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij}, \quad b_j \sim N(0, \sigma_B^2), \quad \varepsilon_{ij} \sim N(0, \sigma^2), \quad i = 1, 2, \quad j = 1, 2, 3$$

The matrix notation is:

$$\underbrace{\begin{pmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \\ y_{13} \\ y_{23} \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{Z}} \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}}_{\mathbf{u}} + \underbrace{\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{13} \\ \varepsilon_{23} \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$

Notice how this matrix representation is constructed in exactly the same way as for the fixed effects model — but separately for fixed and random effects.

A general linear mixed effects model

A general linear mixed model can be presented in matrix notation by:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \quad \text{where } \mathbf{u} \sim N(0, \mathbf{G}) \text{ and } \boldsymbol{\varepsilon} \sim N(0, \mathbf{R}).$$

- $\mathbf{y}$ is the observation vector
- $\mathbf{X}$ is the design matrix for the fixed effects
- β is the vector containing the fixed effect parameters
- $\mathbf{Z}$ is the design matrix for the random effects
- $\mathbf{u}$ is the vector of random effects
  - It is assumed that $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$
  - cov$(u_i, u_j) = G_{i,j}$ (typically $\mathbf{G}$ has a very simple structure (for instance diagonal))
- $\boldsymbol{\varepsilon}$ is the vector of residual errors
  - It is assumed that $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R})$
  - cov$(\varepsilon_i, \varepsilon_j) = R_{i,j}$ (typically $\mathbf{R}$ is diagonal, but we shall later see some useful exceptions for repeated measurements)

The distribution of $\mathbf{y}$

From the model description:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \quad \text{where } \mathbf{u} \sim N(0, \mathbf{G}) \text{ and } \boldsymbol{\varepsilon} \sim N(0, \mathbf{R}).$$

We can compute the mean vector $\boldsymbol{\mu} = E(\mathbf{y})$ and covariance matrix $\mathbf{V} = \text{var}(\mathbf{y})$:

$$\begin{aligned} \boldsymbol{\mu} &= E(\mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}) = \mathbf{X}\beta \quad \text{[All other terms have mean zero]} \\ \mathbf{V} &= \text{var}(\mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}) \quad \text{[from model]} \\ &= \text{var}(\mathbf{X}\beta) + \text{var}(\mathbf{Z}\mathbf{u}) + \text{var}(\boldsymbol{\varepsilon}) \quad \text{[all terms are independent]} \\ &= \text{var}(\mathbf{Z}\mathbf{u}) + \text{var}(\boldsymbol{\varepsilon}) \quad \text{[variance of fixed effects is zero]} \\ &= \mathbf{Z}\text{var}(\mathbf{u})\mathbf{Z}' + \text{var}(\boldsymbol{\varepsilon}) \quad \text{[}\mathbf{Z}\text{ is constant]} \\ &= \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} \quad \text{[from model]} \end{aligned}$$

So $\mathbf{y}$ follows a multivariate normal distribution:

$$\mathbf{y} \sim N(\mathbf{X}\beta, \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})$$

One way ANOVA with random block effect

Consider again the model:

$$y_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij}, \quad b_j \sim N(0, \sigma_B^2), \quad \varepsilon_{ij} \sim N(0, \sigma^2), \quad i = 1, 2, \quad j = 1, 2, 3$$

Calculation of $\boldsymbol{\mu}$ and $\mathbf{V}$ gives:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu + \alpha_1 \\ \mu + \alpha_2 \end{pmatrix} \quad \& \quad \mathbf{V} = \begin{pmatrix} \sigma^2+\sigma_B^2 & \sigma_B^2 & 0 & 0 & 0 & 0 \\ \sigma_B^2 & \sigma^2+\sigma_B^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2+\sigma_B^2 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_B^2 & \sigma^2+\sigma_B^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2+\sigma_B^2 & \sigma_B^2 \\ 0 & 0 & 0 & 0 & \sigma_B^2 & \sigma^2+\sigma_B^2 \end{pmatrix}$$

Notice that two observations from the same block are correlated.

The likelihood function

- The *likelihood L* is a function of model parameters and observations
- For given parameter values L returns a measure of the probability of observing $\mathbf{y}$
- The *negative log likelihood ℓ* for a mixed linear model is:

$$\ell(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \propto \frac{1}{2} \left\{ \log |\mathbf{V}(\boldsymbol{\gamma})| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{V}(\boldsymbol{\gamma}))^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}$$

- Here $\boldsymbol{\gamma}$ is the variance parameters (σ^2 and σ_B^2 in our example)
- A natural estimate is to choose the parameters that make our observations most likely:

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \underset{(\boldsymbol{\beta}, \boldsymbol{\gamma})}{\text{argmin}} \, \ell(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma})$$

- This is the *maximum likelihood (ML)* method

The restricted/residual maximum likelihood method

- The maximum likelihood method tends to give (slightly) too low estimates of the random effects parameters. We say it is *biased downwards*
- The simplest example is:

  $(x_1, \ldots, x_N) \sim$ i.i.d. $N(\mu, \sigma^2)$

  $\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \overline{x})^2$ is the maximum likelihood estimate, but

  $\hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \overline{x})^2$ is generally preferred, because it is *unbiased*
- The *restricted/residual maximum likelihood (REML)* method modifies the maximum likelihood method by minimizing:

$$\ell_{re}(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \propto \frac{1}{2} \left\{ \log |\mathbf{V}(\boldsymbol{\gamma})| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{V}(\boldsymbol{\gamma}))^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \log |\mathbf{X}'(\mathbf{V}(\boldsymbol{\gamma}))^{-1}\mathbf{X}| \right\}$$

  which gives unbiased estimates (at least in balanced cases)
- The REML method is generally preferred in mixed models