MIXED-EFFECTS MODELS

ARTHUR ALLIGNOL

INTRODUCTION

THE NORMAL LINEAR MODEL

For the i-th observation, the linear model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_q X_{qi} + \varepsilon_i,$$

with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

Written in a matrix form

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

with $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \mathbf{I}\sigma^2)$ and

- $\mathbf{y} = (y_1, \ldots, y_n)$ the response vector
- $\mathbf{X}$ the design matrix
- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ the regression coefficients

INTRODUCTION

Hierarchical data are collected when the sampling is performed at two or more levels, one *nested* within the other. E.g.,

- Students within schools (2 levels)
- Students within classrooms within schools (3 levels)
- Individuals within communities within nations (3 levels)

  (Non-nested data are also possible. E.g., high-school students who each have multiple teachers)

INTRODUCTION

Longitudinal data are collected when individuals are followed over time and several measurements are performed.

- Annual data on employement and income for a sample of adults
- Headache score at several visits following treatment

⇒ In these examples it is generally not reasonable to assume that observations within the same unit (e.g., school) or measurements within the same individual, are independent of each other

INTRODUCTION

Mixed-effect models allow to take into account dependencies on hierarchical, longitudinal and other dependent data.

- Unlike the standard linear model, mixed-effect models include more than one source of random variations, i.e., more than one random effect
- ANOVAs could accomodate these kind of dependencies but mixed models are more general. They can deal with irregular and missing observations

LINEAR MIXED-EFFECTS MODEL

THE LINEAR MIXED-EFFECTS MODEL

The *Laird-Ware form* of the linear mixed model

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \cdots + \beta_q x_{qij} + b_{1i} z_{1ij} + \cdots + b_{ri} z_{rij} + \varepsilon_{ij}$$

where

- y_{ij} is the value of the response variable for the j-th of n_i observations in the i-th of M groups or clusters
- $\beta_0, \ldots, \beta_q$ are the fixed-effects coefficients, which are identical for all groups
- $x_{1ij}, \ldots, x_{qij}$ are the fixed-effect regressors for observation j in group i

THE LINEAR MIXED-EFFECTS MODEL

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \cdots + \beta_q x_{qij} + b_{1i} z_{1ij} + \cdots + b_{ri} z_{rij} + \varepsilon_{ij}$$

- $b_{1i}, \ldots, b_{ri}$ are the random-effect coefficients for group i.
  - We assume $b_{ki} \sim \mathcal{N}(0, \psi_k^2)$, $\operatorname{cov}(b_{ki}, b_{k'i}) = \psi_{kk'}$, and $b_{ki}, b_{ki'}$ independent for $i \neq i'$
  - The b_{ki} are thought as random variables, not as parameters. Therefor similar to the errors ε_{ij}
- $z_{1ij}, \ldots, z_{rij}$ are the random effects regressors.
  - The zs are almost always a subset of the xs
  - When there is a random intercept, $z_{1ij} = 1$
- ε_{ij} is the error for observation j in group j
  - $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_{ijj}^2)$. We assume $\varepsilon_{ij}, \varepsilon_{i'j'}$ are independent for $i \neq i'$

THE LINEAR MIXED-EFFECTS MODEL

The Laird-Ware model in matrix form

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

where

- $\mathbf{y}_i$ is the $n_i \times 1$ response vector for observations in group i
- $\mathbf{X}_i$ is the $n_i \times p$ model matrix for the fixed effects for observations in group i
- $\mathbf{b}$ is the $p \times 1$ vector of fixed effect coefficients
- $\mathbf{Z}_i$ if the $n_i \times r$ model matrix for the random effects for observations in group i
- $\mathbf{b}_i$ is the $r \times 1$ vector of random effect coefficients for group i
- $\boldsymbol{\varepsilon}_i$ id the $n_i \times 1$ vector of errors for observations in group i

We assume that $\mathbf{b}_i \sim \mathcal{N}(0, \boldsymbol{\psi})$ and

$\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma^2 \boldsymbol{\Lambda})$. $\mathbf{I}_{n_i}\sigma^2$ are the within-group errors that are homoscedastic and independent.

INFERENCE

INFERENCE

Linear mixed-effects models can be estimated by maximum likelihood. However, this method tends to underestimate the variance components. A modified version of maximum likelihood, known as *restricted maximum likelihood* is therefore often recommended; this provides consistent estimates of the variance components.

Competing linear mixed-effects models can be compared using a likelihood ratio test. If however the models have been estimated by restricted maximum likelihood this test can only be used if both models have the same set of fixed effects.

INFERENCE

Inference for the βs follow from maximum likelihood theory

Hypothesis testing and confidence intervals less obvious, e.g.,

- *Testing the random effect*: $H_0 : \sigma^2 = 0$ → at the boundary of the parameter space
- *F-tests*: degrees of freedom need to be estimated in some ways (except for simple experimental designs)

MODEL DIAGNOSTIC

The normality of the random effects as well as the normality of the residuals need to be checked.

ILLUSTRATION

- Data from the 1982 "High School and Beyond" survey, and pertain to 7185 U.S. high-school students from 160 schools — about 45 students on average per school.
  - 70 of the high schools are Catholic schools and 90 are public schools.
- The object of the data analysis is to determine how students' math achievement scores are related to their family socioeconomic status.
  - We will entertain the possibility that the level of math achievement and the relationship between achievement and SES vary among schools.
  - If there is evidence of variation among schools, we will examine whether this variation is related to school characteristics - in particular, whether the school is a public school or a Catholic school and the average SES of students in the school.

LONGITUDINAL STUDIES

RANDOM INTERCEPT MODEL

Let y_{ij} represent the observation made at time t_j on individual i. A possible model for the observation y_{ij} might be

$$y_{ij} = \beta_0 + \beta_1 t_j + b_i + \varepsilon_{ij}.$$

Here the total residual that would be present in the usual linear regression model has been partitioned into a subject-specific random component b_i which is constant over time plus a residual ε_{ij} which varies randomly over time.

- $\mathrm{E}(b_i) = 0$ and $\mathrm{var}(b) = \sigma_b^2$
- $\mathrm{E}(\varepsilon_{ij}) = 0$ with $\mathrm{var}(\varepsilon_{ij}) = \sigma^2$
- b_i and ε_{ij} independent of each other and of time t_j

$$\mathrm{var}(y_{ij}) = \mathrm{var}(u_i + \varepsilon_{ij}) = \sigma_b^2 + \sigma^2$$

RANDOM INTERCEPT

The covariance between the total residuals at two time points j and k in the same individual is $\mathrm{cov}(b_i + \varepsilon_{ij}, b_i + \varepsilon_{ik}) = \sigma_b^2$.

Note that these covariances are induced by the shared random intercept; for individuals with $b_i > 0$, the total residuals will tend to be greater than the mean, for individuals with $b_i < 0$ they will tend to be less than the mean.

$$\mathrm{cor}(b_i + \varepsilon_{ij}, b_i + \varepsilon_{ik}) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2}.$$

This is an *intra-class correlation* interpreted as the proportion of the total residual variance that is due to residual variability between subjects.

RANDOM INTERCEPT AND SLOPE MODEL

In this model there are two types of random effects, the first modelling heterogeneity in intercepts, b_i, and the second modelling heterogeneity in slopes, v_i:

$$y_{ij} = \beta_0 + \beta_1 t_j + b_i + v_i t_j + \varepsilon_{ij}$$

The two random effects are assumed to have a bivariate normal distribution with zero means for both variables and variances σ_b^2 and σ_v^2 with covariance σ_{uv}:

$$\mathrm{var}(b_i + v_i t_j + \varepsilon_{ij}) = \sigma_b^2 + 2\sigma_{bv} t_j + \sigma_v^2 t_j^2 + \sigma^2$$

which is no longer constant for different values of t_j.

RANDOM INTERCEPT AND SLOPE MODEL

$$\operatorname{cov}(b_i + v_i t_j + \varepsilon_{ij}, b_i + v_i t_k + \varepsilon_{ik}) = \sigma_b^2 + \sigma_{bv}(t_j - t_k) + \sigma_v^2 t_j t_k$$

is not constrained to be the same for all pairs t_j and t_k.

ILLUSTRATION

Beat the blues

Depression is a major public health problem across the world. Antidepressants are the front line treatment, but many patients either do not respond to them, or do not like taking them. The main alternative is psychotherapy, and the modern 'talking treatments' such as *cognitive behavioural therapy* (CBT) have been shown to be as effective as drugs, and probably more so when it comes to relapse.

The data to be used in this chapter arise from a clinical trial of an interactive, multimedia program known as 'Beat the Blues' designed to deliver cognitive behavioural therapy to depressed patients via a computer terminal.

In a randomised controlled trial of the program, patients with depression recruited in primary care were randomised to either the Beating the Blues program, or to 'Treatment as Usual' (TAU).

ILLUSTRATION

Here, we concentrate on the *Beck Depression Inventory II* (BDI). Measurements on this variable were made on the following five occasions:

- Prior to treatment,
- Two months after treatment began and
- At one, three and six months follow-up, i.e., at three, five and eight months after treatment.

There is interest here in assessing the effect of taking antidepressant drugs (`drug`, yes or no) and length of the current episode of depression (`length`, less or more than six months).

GENERALIZED MIXED-EFFECTS MODELS

GENERALIZED MIXED-EFFECTS MODELS

The Generalized linear mixed model is a straighforward extension of the generalized linear model, adding random effects to the linear predictors, and expressing the expected value of the response conditional of the random effects:

$$
\begin{aligned}
g(\mu_{ij}) = g[\mathrm{E}(y_{ij})|b_{1i},\ldots,b_{1r}] = \eta_{ij} \\
\eta_{ij} = \beta_0 + \beta_1 x_{1ij} + \cdots + \beta_q x_{qij} + b_{1i} z_{1ij} + \cdots + b_{ri} z_{rij}
\end{aligned}
$$

- The conditional distribution of y_{ij} given the random effects is a member of the exponential family
- The variance of y_{ij} is a function of μ_{ij} and a dispersion parameter ϕ
- We further assume that the random effects are normally distributed with mean 0 and covariance matrix $\mathbf{\Psi}$
- The estimation of generalized linear mixed models by ML is not straightforward, because the likelihood function includes integrals that are analytically intractable.

GENERALISED ESTIMATING EQUATIONS

INTRODUCTION

- The assumption of the independence of the repeated measurements in an GLM will lead to estimated standard errors that are too small for the between-subjects covariates (at least when the correlation between the repeated measurements are positive) as a result of assuming that there are more independent data points than are justified.

Robust variance estimates can help to obtain reasonably satisfactory results on longitudinal data with a non-normal response

But perhaps more satisfactory than these methods to simply 'fix-up' the standard errors given by the independence model, would be an approach that fully utilises information on the data's structure, including dependencies over time: *GEE*.

GENERALISED ESTIMATING EQUATIONS (GEE)

Let Y_ij be a vector of random variables representing the responses on a given individual and let $\mathrm{E}Y_{ij} = \mu_{ij}$ which is linked to the predictors in some appropriate way

$$g(\mu_{ij}) = \mathbf{X}_{ij}\boldsymbol{\beta}.$$

and let

$$\mathrm{var}Y_{ij} = \mathrm{var}(Y_{ij}; \alpha, \beta)$$

where α represents parameters that model the correlation structure within individuals.

Estimates for β may be obtained based on the score equation

$$\sum_i \frac{\partial \mu_i^T}{\partial \beta} \mathrm{var}(Y_i)^{-1}(Y_i - \mu_i) = 0$$

These can be seen as the multivariate analogue of those used for the quasi-likelihood.

GENERALISED ESTIMATING EQUATIONS

Estimates of the parameters of most interest, i.e., those that determine the average responses over time, are still valid even when the correlation structure is incorrectly specified

But their standard errors might remain poorly estimated if the working correlation matrix is far from the truth.

Possibilities for the working correlation matrix that are most frequently used in practice are:

GENERALISED ESTIMATING EQUATIONS

- An identity matrix: no correlation at all.
- An exchangeable correlation matrix: with a single parameter which gives the correlation of each pair of repeated measures.
- An autoregressive correlation matrix: also with a single parameter but in which $\operatorname{corr}(y_i, y_k) = \vartheta^{|k-i|}, i \neq k$. With ϑ less than one this gives a pattern in which repeated measures further apart in time are less correlated than those that are closer to one another.
- An unstructured correlation matrix: with $K(K-1)/2$ parameters in which $\operatorname{corr}(y_i, y_k) = \vartheta_{ik}$ and where K is the number of repeated measures.

IMPORTANT: That's a marginal model whereas GLMM is conditional on the random effects