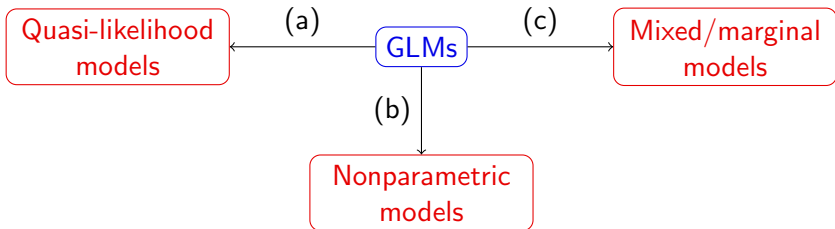# Lecture 16. Mixed Models

Nan Ye

School of Mathematics and Physics
University of Queensland

Recall: Extending GLMs

(a) Relax assumption on the random component.

(b) Relax assumption on the systematic component.

(c) Relax assumption on the data (independence).

Correlated Data

So far...

- We have been working under the assumption that the responses are independent given the covariates.
- This assumption does not hold for many problems.

Examples of correlated responses

- Measurements on clusters of subjects
  - e.g. measurements on patients from the same hospital may be correlated because they are attended by the same set of nurses and doctors, and they are likely to share demographic or socio-economic features.
- Repeated measurements on same subject

This Lecture

Linear mixed model

- Random intercept model
- Modelling consideration: random effects versus fixed effects
- Random intercept and slope model

Generalized linear mixed model

Random Intercept Model

Model definition

- The random intercept model assumes that each cluster/block affect the responses via cluster-specific intercept terms only.

- The model has the form

$$Y_{ij} = \mathbf{x}_{ij}^\top \beta + \alpha_i + \epsilon_{ij},$$
$$\epsilon_{ij} \overset{ind}{\sim} N(0, \sigma^2), \text{independent of}$$
$$\alpha_i \overset{ind}{\sim} N(0, \sigma_A^2),$$

where Y_{ij} and $\mathbf{x}_{ij}$ are the response and covariate vector for the j-th example in cluster i, α_i is a *random* intercept associated with cluster i, and ϵ_{ij} is a Gaussian noise.

As usual, $\mathbf{x}_{ij}$ contains a dummy variable of value 1 corresponding to the intercept term.

Remarks

- The model is called a mixed model because it contains a fixed effect component $\mathbf{x}_{ij}^\top \beta$, and a random effect component α_i.

- When $\sigma_A^2 = 0$, the model reduces to a fixed effects only linear model model with no intra-cluster correlation.

- When $\sigma_A^2 \to \infty$, some people consider this as a fixed effects linear model where each cluster has its own fixed α_i.

Conditional probability $p(\mathbf{Y} \mid \mathbf{X}, \beta, \sigma^2, \sigma_A^2)$

- Assume that there are K clusters, and cluster i has n_j examples.
- Let $\mathbf{Y} = (Y_{11}, \ldots, Y_{1n_1}, \ldots, Y_{K1}, \ldots, Y_{Kn_K})$.
- Let $\mathbf{X}$ be the design matrix with $\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n_1}, \ldots, \mathbf{x}_{K1}, \ldots, \mathbf{x}_{Kn_K}$ as rows.
- The random intercept model defines a conditional distribution of $p(\mathbf{Y} \mid \mathbf{X}, \beta, \sigma^2, \sigma_A^2)$.
- This can be shown to be a multivariate normal distribution $N(\mu, \Sigma)$.

- The mean is given by $\mu = \mathbf{X}\beta$ as

$$\mathbb{E}(Y_{ij}) = \mathbf{x}_{ij}^{\top}\beta,$$

- The covariance matrix Σ is given by

$$\Sigma_{ij,i'j'} = \text{cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} \sigma_A^2 + \sigma^2, & i = i', j = j', \\ \sigma_A^2, & i = i', j \neq j' \\ 0, & \text{otherwise.} \end{cases}$$

Parameter Estimation

- We can choose β by maximizing the likelihood $p(\mathbf{Y} \mid \mathbf{X}, \beta, \sigma^2, \sigma_A^2)$.

- The covariance matrix can be first estimatd using the method of restricted maximum likelihood (REML, a.k.a. residual or reduced maximum likelihood).

  - The idea is to transform the dataset so that the likelihood function of the transformed dataset depends only on Σ, but not on β.

- Once Σ is estimated, we can then estimate β by solving a regularized least squares problem.

  (Details not covered in this course.)

Fixed Effect versus Random Effect

- We can also consider cluster-specific intercepts as fixed effects.
- The model has the form

$$Y_{ij} = \mathbf{x}_{ij}^{\top}\beta + \alpha_i + \epsilon_{ij},$$
$$\epsilon_{ij} \overset{ind}{\sim} N(0, \sigma^2).$$

- This is equivalent to adding the cluster number as a factor covariate.

- If we are interested in the particular clusters in the study, we should treat α_i's as fixed effects.
- If we are not interested in the particular clusters in the study, we should treat α_i's as random effects.
- As a practical consideration, if there are two few samples within each cluster, we treat α_i's as random effects because they cannot be reliably estimated.

Random Intercept and Slope Model

- In general, clusters may affect the responses not only through the cluster-specific intercept terms, but through interactions with certain covariates.
- The general linear mixed model has the following form

$$Y_{ij} = \mathbf{x}_{ij}^\top \beta + \mathbf{z}_{ij}^\top \alpha_i + \epsilon_{ij},$$
$$\epsilon_{ij} \stackrel{ind}{\sim} N(0, \sigma^2), \text{independent of}$$
$$\alpha_i \stackrel{ind}{\sim} N(0, \Sigma_A)$$

$\mathbf{z}_{ij}$ contains a dummy variable of value 1 corresponding to the intercept term.

Remarks

- z_{ij} may contain a subset of covariates in x_{ij}.
- As in the random intercepts model, Y follows a multivariate normal distribution.

Generalized Linear Mixed Model (GLMM)

- Recall: A GLM has the following structure

  (systematic) $\quad \mathbb{E}(Y \mid \mathbf{x}) = h(\beta^\top \mathbf{x})$,

  (random) $\quad Y \mid \mathbf{x}$ follows an exponential family distribution.

- A *generalized linear mixed model* has the following structure

  $$\mathbb{E}(Y_{ij} \mid \mathbf{x}_{ij}, \mathbf{z}_{ij}, \alpha_i) = h(\mathbf{x}_{ij}^\top \beta + \mathbf{z}_{ij}^\top \alpha_i),$$
  $$Y_{ij} \mid \mathbf{x}_{ij}, \mathbf{z}_{ij}, \alpha_i \sim \text{ an exponential family distribution},$$
  $$\alpha_j \overset{ind}{\sim} N(0, \Sigma_A).$$

Example

Data

```
> library(lme4)
> dim(sleepstudy)
[1] 180   3
> head(sleepstudy)
    Reaction Days Subject
1 249.5600   0     308
2 258.7047   1     308
3 250.8006   2     308
4 321.4398   3     308
5 356.8519   4     308
6 414.6901   5     308
```

- 18 subjects (long-distance drivers), normal sleep hours before day 0, but 3 hours sleep for next 10 days.
- Reaction time for a series of test from day 0 to day 9 recorded.

Reaction times vs. days of sleep deprivation for 18 subjects
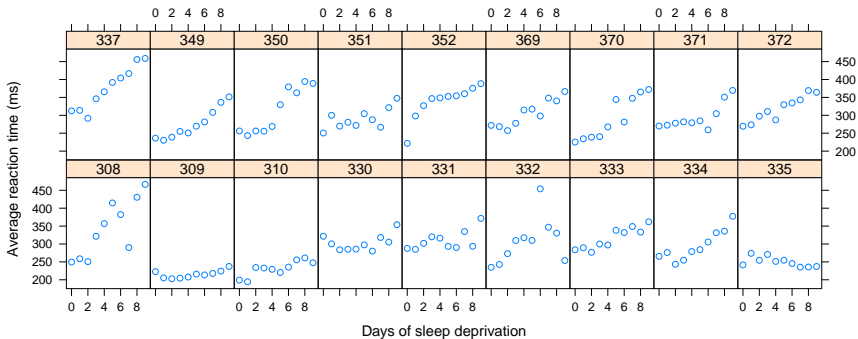
We consider the following linear mixed model with a random intercept and a random slope

$$Y_{ij} = \beta_0 + \beta_1 * \mathrm{day}_{ij} + \alpha_{i0} + \alpha_{i1} * \mathrm{day}_{ij} + \epsilon_{ij},$$
$$\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2), \text{ independent of}$$
$$\begin{pmatrix} \alpha_{i0} \\ \alpha_{i1} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{A0}^2 & \rho\sigma_{A0}\sigma_{A1} \\ \rho\sigma_{A0}\sigma_{A1} & \sigma_{A1}^2 \end{pmatrix} \right)$$

```
fit.lmm = lmer(Reaction ~ Days + (Days | Subject),
 data=sleepstudy)
```

- The term (Days | Subject) is a random effect term.
- It introduces a term $\mathbf{z}_{ij}^{\top}\alpha_i$ in the linear mixed model.
- The cluster index i is the Subject value.
- $\mathbf{z}_{ij}$ contains the Days covariate, and an dummy variable of value 1.

```
Random effects:
 Groups    Name          Variance Std.Dev. Corr
 Subject   (Intercept) 612.09    24.740
           Days          35.07     5.922   0.07
 Residual              654.94    25.592
Number of obs: 180, groups:  Subject, 18

Fixed effects:
            Estimate Std. Error t value
(Intercept)  251.405      6.825  36.838
Days          10.467      1.546   6.771

Correlation of Fixed Effects:
     (Intr)
Days -0.138
```

Estimated fixed effects parameters

$$\hat{\beta}_0 = 251.405 \text{ms},$$
$$\hat{\beta}_1 = 10.467 \text{ms/day}.$$

Estimated variance parameters

$$\hat{\sigma}^2_{A0} = 612.09,$$
$$\hat{\sigma}^2_{A1} = 35.07,$$
$$\hat{\rho} = 0.07.$$

- Baseline reaction times: normally distributed with mean estimated to be 251.405ms and standard deviation estimated to be $\sqrt{612.09} = 24.74$ ms.
- Increase in reaction times for each additional day of sleep derivation: normally distributed with mean estimated to be 10.467ms/day and standard deviation estimated to be $\sqrt{35.07} = 5.92$ms/day.
- Correlation between a subject's intercept and slope is estimated to be 0.07. It appears that a subject's response to sleep deprivation is not related much at all to their inherent reaction ability.

Simplified model?

```
   > fit0 = lmer(Reaction ~ Days + (1 | Subject), data=sleepstudy)
   > anova(fit0, fit.lmm)
refitting model(s) with ML (instead of REML)
Data: sleepstudy
Models:
fit0: Reaction ~ Days + (1 | Subject)
fit.lmm: Reaction ~ Days + (Days | Subject)
        Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
fit0     4 1802.1 1814.8 -897.04   1794.1
fit.lmm  6 1763.9 1783.1 -875.97   1751.9 42.139      2  7.072e-10
    ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

- The χ^2 test is approximate but the computed *p*-value is generally conservative (bigger than correct *p*-value).
- Thus we cannot drop the random slope to simplify the model to a random intercept model.

What You Need to Know

- In many occasions, responses are correlated due to some form of clustering.
- Random intercept model models the effect of clustering using cluster-specific intercepts.
- Random intercept and slope model extends random intercept model by allowing interaction between clusters and some covariates.
- Generalized linear mixed model generalizes linear mixed model by allowing the response to follow an exponential family distribution.