

Introduction to Mixed Models

Dr Rafael de Andrade Moral
Associate Professor of Statistics, Maynooth University

rafael.deandrademoral@mu.ie
<https://rafamoral.github.io>

Outline

- Random effects models
- Linear mixed models (LMMs)
- Nested and crossed random effects
- Group-level predictors
- Generalized linear mixed models (GLMMs)
- Individual-level random effects
- Bayesian mixed models

Introduction

Mixed Models

- Mixed models are a broad class of models that are applied to data that consist of sub-groups, or *clusters*
- Mixed models have many other names in the literature and in practice
- You may have heard/seen: *multilevel models, hierarchical models, conditional models, random effects models*
- In practice, these are all synonyms, but each name highlights a different part of the construction of a mixed model
- The defining feature of multilevel models is that they are *models of models*¹

¹Andrews, M. (2021). Doing Data Science in R: An Introduction for Social Scientists. SAGE Publishing, London, UK.

Random Effects Models

Random Effects Models

- Example: Incidence of disease in cattle



Random Effects Models

- We have 15 cattle herds, and multiple observations per herd
- The observations within herd are not independent
- We begin with a simple model for herd 4 only

$$\begin{aligned} Y_j &\sim \text{Binomial}(m_j, \pi) \\ \log\left(\frac{\pi}{1-\pi}\right) &= \beta_0 \end{aligned}$$

- j is the index representing observation within herd 4

Random Effects Models

- We can fit a binomial model to each herd separately:

$$\begin{aligned} Y_{ij} &\sim \text{Binomial}(m_{ij}, \pi_{ij}) \\ \log\left(\frac{\pi_i}{1 - \pi_i}\right) &= \beta_{0i} \end{aligned}$$

- i is the index representing herd, and j represents observation within herd
- Note that π is only indexed by i
- There are 15 regression coefficients, one for each herd, representing the log-odds of disease incidence

Random Effects Model

- We have a model of disease incidence for herd 1, another for herd 2, and so on
- But we don't have a model of the distribution of disease incidence across all herds
- We don't have, for instance, a model giving us the mean, standard deviation, or any other information about disease incidence rates across all possible herds, of which our dataset of 15 herds is one possible sample
- To obtain this model, we can use *random effects*

Random Effects Model

- We may write

$$\begin{aligned} Y_{ij} | \beta_{0i} &\sim \text{Binomial}(m_{ij}, \pi_i) \\ \log\left(\frac{\pi_i}{1 - \pi_i}\right) &= \beta_{0i} \\ \beta_{0i} &\sim N(b, \sigma^2) \end{aligned}$$

- Now the log-odds of disease incidence are modelled as normally distributed with mean b and variance σ^2

Random Effects Model

$$\begin{aligned} Y_{ij} | \beta_{0i} &\sim \text{Binomial}(m_{ij}, \pi_i) \\ \log\left(\frac{\pi_i}{1 - \pi_i}\right) &= \beta_{0i} \\ \beta_{0i} &\sim N(b, \sigma^2) \end{aligned}$$

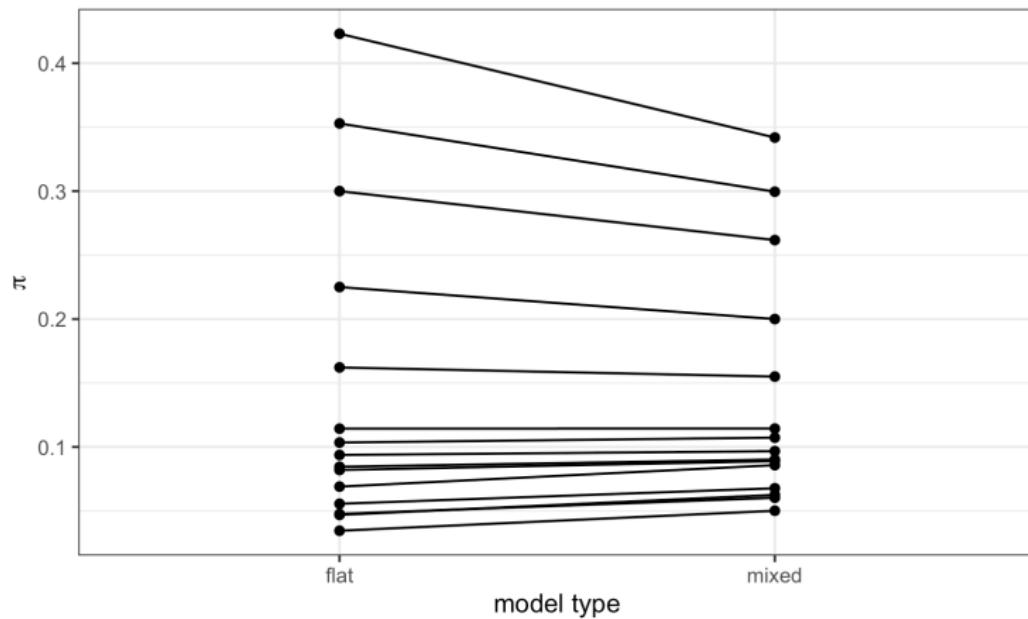
- In this model, $\beta_{01}, \beta_{02}, \dots, \beta_{0,15}$ still represent the same: log-odds of disease incidence rate for herd i
- However, they are drawn from the same normal distribution
- First consequence:
 - We now have a model of the *population* from which $\beta_{01}, \beta_{02}, \dots, \beta_{0,15}$ are a sample
 - Since each β_{0i} defines a model for a herd, the normal distribution from which $\beta_{01}, \beta_{02}, \dots, \beta_{0,15}$ are drawn is a *model of models*
 - The population model for β_{0i} allows us to predict the log-odds, or probability, of disease incidence for any future herd to be sampled

Random Effects Model

$$\begin{aligned} Y_{ij} | \beta_{0i} &\sim \text{Binomial}(m_{ij}, \pi_i) \\ \log\left(\frac{\pi_i}{1 - \pi_i}\right) &= \beta_{0i} \\ \beta_{0i} &\sim N(b, \sigma^2) \end{aligned}$$

- Second consequence:
 - Because we assume $\beta_{01}, \beta_{02}, \dots, \beta_{0,15}$ are all drawn from the same normal distribution, this introduces constraints on the inference of the values of each β_{0i}
 - Now the values of b and σ^2 are also relevant, and because they are unknown, they are also inferred from $\beta_{01}, \beta_{02}, \dots, \beta_{0,15}$
 - This effectively means that the inferences concerning the $\beta_{01}, \beta_{02}, \dots, \beta_{0,15}$ are inter-dependent and mutually constrain one another
 - This phenomenon is referred to as *shrinkage*

Shrinkage



Random Effects Model

- It is common practice to re-parameterise the random effects so that they have a mean of zero, and are deviations from a population mean:

$$\begin{aligned} Y_{ij}|b_i &\sim \text{Binomial}(m_{ij}, \pi_i) \\ \log\left(\frac{\pi_i}{1 - \pi_i}\right) &= \beta_0 + b_i \\ b_i &\sim N(0, \sigma^2) \end{aligned}$$

- Now, β_0 is the population mean, while b_i are the random deviations for each herd, sampled from a normal distribution with a mean of zero
- This is exactly the same model as before, it has just been re-parameterised

Random Effects Model

- This parameterisation for the linear predictor is what gives mixed models their name:

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = \underbrace{\beta_0}_{\text{fixed effect}} + \underbrace{b_i}_{\text{random effect}}$$

- We have a *mix of fixed and random effects*
- Random effects are modelled according to a probability distribution
- Fixed effects are constants

The Normal Random Effects Model

The Normal Random Effects Model

- Example: Alcohol consumption per capita
- We have 189 countries and 22 different years, but not necessarily do we have data from each country in each year
- A possible mixed model for per capita alcohol consumption in country i in year j would be:

$$\begin{aligned}Y_{ij}|b_i &\sim N(\mu_i, \sigma^2) \\ \mu_i &= \beta_0 + b_i \\ b_i &\sim N(0, \sigma_b^2)\end{aligned}$$

- or, using an alternative notation,

$$y_{ij} = \underbrace{\beta_0}_{\text{fixed intercept}} + \underbrace{b_i}_{\text{country-level random effect}} + \underbrace{\varepsilon_{ij}}_{\text{random error}}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

- The error ε_{ij} gives the random year by year deviation from the country's average consumption rate

The Normal Random Effects Model

$$\begin{aligned}Y_{ij}|b_i &\sim N(\mu_i, \sigma^2) \\ \mu_i &= \beta_0 + b_i \\ b_i &\sim N(0, \sigma_b^2)\end{aligned}$$

- This model has a *latent variable*: b_i is unobserved
- Also, b_i are not true parameters of the model; in fact, even though the model above estimates a mean response for all 189 countries, it does so while estimating only 3 parameters: β_0 , σ^2 , and σ_b^2
- We *estimate* parameters and *predict* random effects
 - $\hat{\beta}_0$ is an estimated parameter
 - \hat{b}_i is a predicted random effect

The Normal Random Effects Model

- How does this accommodate the fact that observations within the same country are correlated?
- Marginally, it can be shown that

$$\text{Var}(Y_{ij}) = \sigma^2 + \sigma_b^2 \quad \text{and} \quad \text{Cov}(Y_{ij}, Y_{ij'}) = \sigma_b^2$$

- In other words,
 - σ_b^2 is the *between-country* variation
 - σ^2 is the *within-country* variation
 - they both contribute to the total variation

The Normal Random Effects Model

- We have that the *intraclass correlation* (ICC) is

$$\text{Corr}(Y_{ij}, Y_{ij'}) = \frac{\sigma_b^2}{\sigma^2 + \sigma_b^2}$$

- The ICC takes values between 0 and 1
- It tells us how much of the total variance in the data is due to variation between the countries
- If the ICC is close to 1, then most variation in the data is due to variation between countries
 - in this case, there will be *less* shrinkage
- If the ICC is close to 0, then most variation in the data is due to random noise
 - in this case, there will be *more* shrinkage towards an overall mean

Linear Mixed Effects Models

Linear Mixed Effects Models

- A model with random intercepts and slopes: the `sleepstudy` dataset
- The response variable is the average reaction time (in milliseconds) for each of 18 people on each day of a sleep deprivation experiment that lasted 10 days
- Days 0 and 1 were adaptation, day 2 was the baseline and sleep deprivation started after day 2



Linear Mixed Effects Models

- Let's start, like before, with a model for a single participant – subject 334
- The model is

$$\begin{aligned}Y_j &\sim N(\mu_j, \sigma^2) \\ \mu_j &= \beta_0 + \beta_1 \text{day}_j\end{aligned}$$

- We have a mean that changes with day, this is why we index μ with j

Linear Mixed Effects Models

- If we assume that there is a common residual standard deviation term σ for all subjects, rather than one per each of the $J = 18$ subjects, this model is identical to a varying intercept and varying slope linear model:

$$\begin{aligned}Y_{ij} &\sim N(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \beta_{0i} + \beta_{1i}\text{day}_j\end{aligned}$$

- We have a mean that changes according to subject i and day j
- But the same variance σ^2 for all subjects and days within subjects

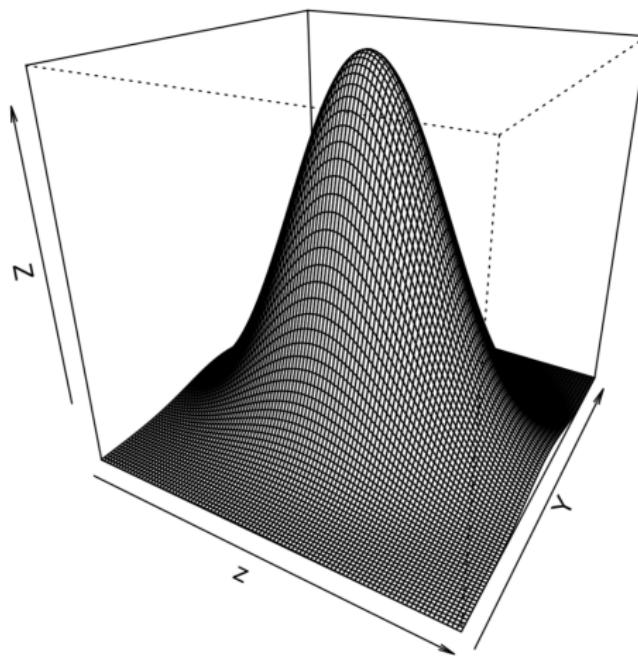
Linear Mixed Effects Models

- The LMM is written as:

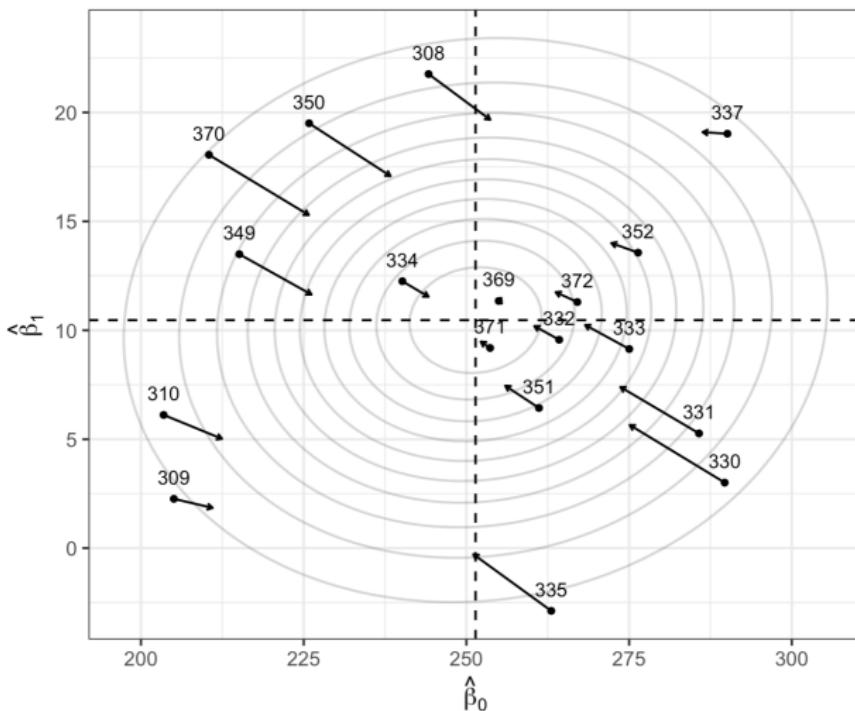
$$\begin{aligned} Y_{ij}|b_{0i}, b_{1i} &\sim N(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})\text{day}_j \\ \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} \right) \end{aligned}$$

- We have now a bivariate normal distribution for the random intercepts and slopes
- $\rho \in (-1, 1)$ is the correlation between random intercepts and slopes

The Bivariate Normal Distribution



Shrinkage



Varying Intercepts Only

- We can also fit a random intercepts model with a fixed slope:

$$\begin{aligned} Y_{ij}|b_{0i} &\sim N(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \beta_0 + b_{0i} + \beta_1 \text{day}_j \\ b_{0i} &\sim N(0, \sigma_0^2) \end{aligned}$$

Varying Slopes Only

- Or a random slopes model with a fixed intercept:

$$\begin{aligned} Y_{ij}|b_{1i} &\sim N(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \beta_0 + (\beta_1 + b_{1i})\text{day}_j \\ b_{1i} &\sim N(0, \sigma_1^2) \end{aligned}$$

- It depends on our assumptions

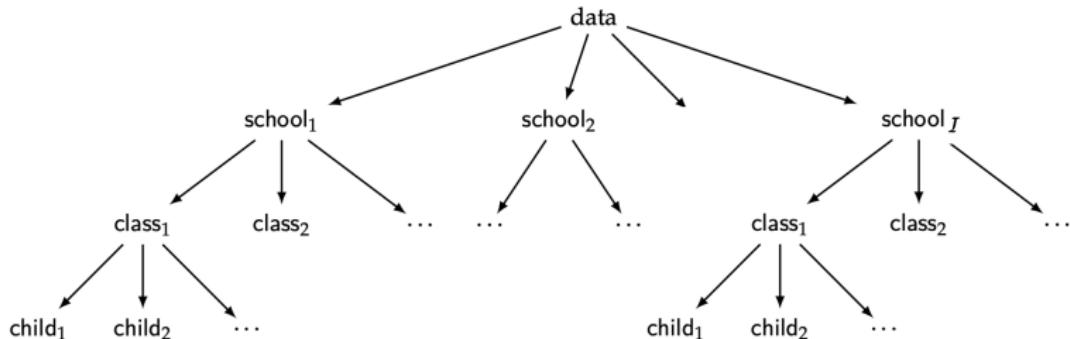
Nested and Crossed Random Effects

Mixed Models for Nested Data

- In all the examples considered thus far, our multilevel models had only two levels, which we can denote level 0 and level 1
 - we had observations in different periods (level 0) within herds (level 1)
 - observations of per capita alcohol consumption in different years (level 0) within different countries (level 1)
 - individual observations of reaction times on different days (level 0) grouped by different experimental subjects (level 1)
- We can easily have groups within groups, and we usually refer to these models as nested multilevel models

Nested Data

- Factors are considered to be *nested* if the levels within a level of a factor are different when compared to another level – easier to visualise with an example
- Example: children's maths scores in different schools
 - child is nested within class, which is nested within school:



Mixed Models for Nested Data

$$\begin{aligned} Y_{ijk} | s_{0i}, s_{1i}, c_{0ij}, c_{1ij} &\sim N(\mu_{ijk}, \sigma^2) \\ \mu_{ijk} &= \beta_0 + s_{0i} + c_{0ij} + (\beta_1 + s_{1i} + c_{1ij})\text{ses}_{ijk} \\ \begin{pmatrix} s_{0i} \\ s_{1i} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^{2(s)} & \rho_s \sigma_0^{(s)} \sigma_1^{(s)} \\ \rho_s \sigma_0^{(s)} \sigma_1^{(s)} & \sigma_1^{2(s)} \end{pmatrix} \right) \\ \begin{pmatrix} c_{0ij} \\ c_{1ij} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^{2(c)} & \rho_c \sigma_0^{(c)} \sigma_1^{(c)} \\ \rho_c \sigma_0^{(c)} \sigma_1^{(c)} & \sigma_1^{2(c)} \end{pmatrix} \right) \end{aligned}$$

- i indexes school, j indexes class within school, and k indexes child within class
- We have 2 sets of bivariate normal random effects
- How many parameters do we have?

Mixed Models for Nested Data

$$\begin{aligned} Y_{ijk} | s_{0i}, s_{1i}, c_{0ij}, c_{1ij} &\sim N(\mu_{ijk}, \sigma^2) \\ \mu_{ijk} &= \beta_0 + s_{0i} + c_{0ij} + (\beta_1 + s_{1i} + c_{1ij})\text{ses}_{ijk} \\ \begin{pmatrix} s_{0i} \\ s_{1i} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^{2(s)} & \rho_s \sigma_0^{(s)} \sigma_1^{(s)} \\ \rho_s \sigma_0^{(s)} \sigma_1^{(s)} & \sigma_1^{2(s)} \end{pmatrix} \right) \\ \begin{pmatrix} c_{0ij} \\ c_{1ij} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^{2(c)} & \rho_c \sigma_0^{(c)} \sigma_1^{(c)} \\ \rho_c \sigma_0^{(c)} \sigma_1^{(c)} & \sigma_1^{2(c)} \end{pmatrix} \right) \end{aligned}$$

- i indexes school, j indexes class within school, and k indexes child within class
- We have 2 sets of bivariate normal random effects
- How many parameters do we have?
 - A: 9 (2 fixed effects, β_0 and β_1 , 5 variance components, and 2 correlations)

Mixed Models for Nested Data

- Removing the correlation between random intercepts and slopes for class within school:

$$\begin{aligned} Y_{ijk} | s_{0i}, s_{1i}, c_{0ij}, c_{1ij} &\sim N(\mu_{ijk}, \sigma^2) \\ \mu_{ijk} &= \beta_0 + s_{0i} + c_{0ij} + (\beta_1 + s_{1i} + c_{1ij})\text{ses}_{ijk} \\ \begin{pmatrix} s_{0i} \\ s_{1i} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^{2(s)} & \rho_s \sigma_0^{(s)} \sigma_1^{(s)} \\ \rho_s \sigma_0^{(s)} \sigma_1^{(s)} & \sigma_1^{2(s)} \end{pmatrix} \right) \\ \begin{pmatrix} c_{0ij} \\ c_{1ij} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^{2(c)} & 0 \\ 0 & \sigma_1^{2(c)} \end{pmatrix} \right) \end{aligned}$$

- This is equivalent to setting $\rho_c = 0$; therefore we now have 8 parameters

Mixed Models for Nested Data

- Removing the random slope for class within school:

$$\begin{aligned}
 Y_{ijk} | s_{0i}, s_{1i}, c_{0ij}, c_{1ij} &\sim N(\mu_{ijk}, \sigma^2) \\
 \mu_{ijk} &= \beta_0 + s_{0i} + c_{0ij} + (\beta_1 + s_{1i} + c_{1ij})\text{ses}_{ijk} \\
 \begin{pmatrix} s_{0i} \\ s_{1i} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^{2(s)} & \rho_s \sigma_0^{(s)} \sigma_1^{(s)} \\ \rho_s \sigma_0^{(s)} \sigma_1^{(s)} & \sigma_1^{2(s)} \end{pmatrix} \right) \\
 c_{0ij} &\sim N(0, \sigma^{2(c)})
 \end{aligned}$$

- This is equivalent to setting $\sigma_1^{2(c)} = 0$; therefore we now have 7 parameters

Mixed Models for Nested Data

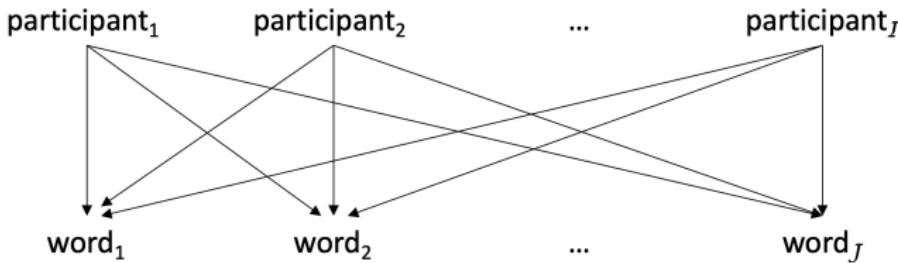
- Removing the random slope for school:

$$\begin{aligned}Y_{ijk} | s_{0i}, s_{1i}, c_{0ij}, c_{1ij} &\sim N(\mu_{ijk}, \sigma^2) \\ \mu_{ijk} &= \beta_0 + s_{0i} + c_{0ij} + (\beta_1 + s_{1i} + c_{1ij})\text{ses}_{ijk} \\ s_{0i} &\sim N(0, \sigma^{2(s)}) \\ c_{0ij} &\sim N(0, \sigma^{2(c)})\end{aligned}$$

- This is equivalent to setting $\sigma_1^{2(s)} = 0$ and $\rho_s = 0$; therefore we now have 5 parameters

Crossed Data

- If two factors are not nested, then they are *crossed*
- The levels within a factor are the same across all (or some) levels of another factor
- Example: reaction time in the British Lexicon Project
 - participants are shown text strings and they have to decide whether they are words or not



Mixed Models for Crossed Data

$$\begin{aligned} Y_{ij}|p_i, w_j &\sim \mathcal{N}(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \beta_0 + p_i + w_j \\ p_i &\sim \mathcal{N}(0, \sigma_p^2) \\ w_j &\sim \mathcal{N}(0, \sigma_w^2) \end{aligned}$$

- i indexes participant and j indexes word

Group-Level Predictors

Group-Level Predictors

- In the examples thus far, the predictor variables were at the bottom level, level 0
 - at level 0, we had children, and the values of the predictor variable `ses` were specific to each child
- Consider the following two related data sets, which are based on the `MathAchieve` and `MathAchSchool` data sets, respectively, available in the `nlme` package
 - In one of the datasets we have variables measured for each *pupil in a school* (level 0: ethnic minority status, sex, home socioeconomic background, and standardised maths scores)
 - For the other, we have variables measured for each *school* (level 1: school size, whether it is public or catholic, percentage of students on academic track, a measure of the discrimination climate, whether there is a high proportion of minority students, and mean socioeconomic status of the school)

A Model Ignoring the Group-Level Predictors

$$\begin{aligned} Y_{ij} | b_{0i}, b_{1i}, c_{0i}, c_{1i}, d_{0i}, d_{1i} &\sim N(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \beta_0 + b_{0i} + c_{0i} + d_{0i} + (\beta_1 + b_{1i})\text{ses}_{ij} \\ &\quad + (\beta_2 + c_{1i})\text{sexMale}_{ij} \\ &\quad + (\beta_3 + d_{1i})\text{minorityYes}_{ij} \\ \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^{2(b)} & \rho_b \sigma_0^{(b)} \sigma_1^{(b)} \\ \rho_b \sigma_0^{(b)} \sigma_1^{(b)} & \sigma_1^{2(b)} \end{pmatrix} \right) \\ \begin{pmatrix} c_{0i} \\ c_{1i} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^{2(c)} & \rho_c \sigma_0^{(c)} \sigma_1^{(c)} \\ \rho_c \sigma_0^{(c)} \sigma_1^{(c)} & \sigma_1^{2(c)} \end{pmatrix} \right) \\ \begin{pmatrix} d_{0i} \\ d_{1i} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^{2(d)} & \rho_d \sigma_0^{(d)} \sigma_1^{(d)} \\ \rho_d \sigma_0^{(d)} \sigma_1^{(d)} & \sigma_1^{2(d)} \end{pmatrix} \right) \end{aligned}$$

- 4 fixed effects, 5 variance components, 3 correlations = 12 parameters
- Convergence problems

A Model Considering Group-Level Predictors

$$\begin{aligned} Y_{ij}|b_{0i}, b_{1i} &\sim N(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})\text{ses}_{ij} + \beta_2\text{sexMale}_{ij} \\ &\quad + \beta_3\text{himinty}_i + \beta_4\text{sectorPublic}_i + \beta_5\text{pracad}_i + \beta_6\text{disclim}_i \\ \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} &\sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^{2(b)} & \rho_b \sigma_0^{(b)} \sigma_1^{(b)} \\ \rho_b \sigma_0^{(b)} \sigma_1^{(b)} & \sigma_1^{2(b)} \end{pmatrix} \right) \end{aligned}$$

- Predictors indexed only by i (the school index) are group-level predictors (at level 1)
- Predictors indexed by ij (j -th pupil within school i) are at level 0

GLMMs

Generalized Linear Mixed Models

- The LMM is a particular case of the GLMM
- We can change the distributional assumption at the top level of the hierarchy according to the nature of our response variable and our assumptions
 - count data: $Y \sim \text{Poisson}(\mu)$, or $Y \sim \text{NegBin}(\mu, \phi)$
 - binary data: $Y \sim \text{Bernoulli}(\pi)$
 - discrete proportions: $Y \sim \text{Binomial}(m, \pi)$
 - strictly positive and skewed data: $Y \sim \text{Gamma}(\mu, \phi)$, or $Y \sim \text{Inverse Gaussian}(\mu, \phi)$
 - ...
- The appropriate link function must be used to model μ

Generalized Linear Mixed Models

- Estimation is more complex when our conditional assumption for Y is a distribution other than the normal
- This is because when we estimate a mixed model via maximum likelihood (or REML), we are optimising the (restricted) likelihood function of the *marginal* distribution of Y rather than the conditional
- In other words, we have the distributions of $Y|b$ and of b , but we want to find the distribution of Y
- The joint distribution is obtained by multiplying the two:

$$f(y, b) = f(y|b)f(b)$$

- The marginal is obtained by integrating over the random effects:

$$f(y) = \int f(y|b)f(b)db$$

- When $f(y|b)$ and $f(b)$ are both normal densities, there is a closed-form result for $f(y)$ (which is, in fact, another normal density)
- Otherwise, the integral may not have an analytic solution and we must resort to *numerical methods*

Generalized Linear Mixed Models

- From *fastest* and *less precise* to *slowest* and *most precise*:
 - Penalised Quasi-Likelihood (PQL): joint optimisation of parameters and random effects
 - Laplace Approximation: approximates the integrand to a normal density function
 - Gauss-Hermite Quadrature: approximates the integral with a sum
 - Monte Carlo Integration: creates many samples from a known distribution
- `lme4::glmer` implements PQL through `nAGQ = 0`, (adaptive) Laplace approximation through `nAGQ = 1` (the default), and GH-quadrature with q quadrature points with `nAGQ = q`

Generalized Linear Mixed Models

- Example: *Arabidopsis* fruit production

$$Y_{ijkl} | r_i, p_{ij}, g_k \sim \text{Poisson}(\mu_{ijkl})$$

$$\begin{aligned}\log \mu_{ijkl} = & \beta_0 + r_i + p_{ij} + g_k + \beta_1 \text{rack2}_{ijkl} + \beta_2 \text{statusPetri.Plate}_{ijkl} \\ & + \beta_3 \text{statusTransplant}_{ijkl} + \beta_4 \text{nutrient8}_{ijkl} \\ & + \beta_5 \text{amdUnclipped}_{ijkl} + \beta_6 \text{nutrient8:amdUnclipped}_{ijkl}\end{aligned}$$

$$r_i \sim N(0, \sigma_r^2)$$

$$p_{ij} \sim N(0, \sigma_p^2)$$

$$g_k \sim N(0, \sigma_g^2)$$

- We have 10 parameters: 7 regression coefficients and 3 variance components

Fixed vs. Random Factor

- We typically use fixed effects when we assume there is a *systematic difference* between levels of a factor
- We typically use random effects when we assume that what we observe is a *random sample from a population*

Individual-Level Random Effects

Individual-Level Random Effects

- Recall that:
 - $Y \sim \text{Binomial}(m, \pi)$, $E[Y] = m\pi$, $\text{Var}(Y) = m\pi(1 - \pi)$
 - $Y \sim \text{Poisson}(\mu)$, $E[Y] = \mu$, $\text{Var}(Y) = \mu$
- When the variability is larger than assumed by these models, we refer to this phenomenon as *overdispersion*
- There are many different causes
 - variability of the experimental material
 - correlation between individual responses
 - cluster and multistage sampling leading to complex dependencies
 - aggregation
 - omitted unobserved variables

Individual-Level Random Effects

- There are many alternative models that accommodate extra-variability
- One particular way of doing so is by adding an *individual-level random effect* to the linear predictor
- For a **Poisson-Normal** GLMM:

$$Y_i|z_i \sim \text{Poisson}(\mu_i)$$

$$\log \mu_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + z_i$$

$$z_i \sim N(0, \sigma_z^2)$$

- For a **Binomial-Normal** GLMM:

$$Y_i|z_i \sim \text{Binomial}(m, \pi_i)$$

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + z_i$$

$$z_i \sim N(0, \sigma_z^2)$$

- Example: Coffee berry borer trapping data

Bayesian Mixed Models

Bayesian Mixed Models

- So far we have fitted all models using the classical, or frequentist paradigm
- The parameters are estimated via maximum likelihood, or restricted maximum likelihood
- We can fit all models presented thus far using the Bayesian paradigm as well
- In the Bayesian paradigm, we assign prior distributions to the model parameters, and combine them with the likelihood to obtain posterior distributions, which are used for inference
- Sometimes we can obtain these posterior distributions analytically, but other times we resort to Markov Chain Monte Carlo (MCMC) methods
- There are many MCMC methods available in the literature; STAN implements Hamiltonian MCMC
- Short tutorial using `brms`