

# 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*<sup>1</sup>

---

<sup>1</sup> Andrews, 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  $\pi$  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  $\sigma^2$

# Random Effects Model

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

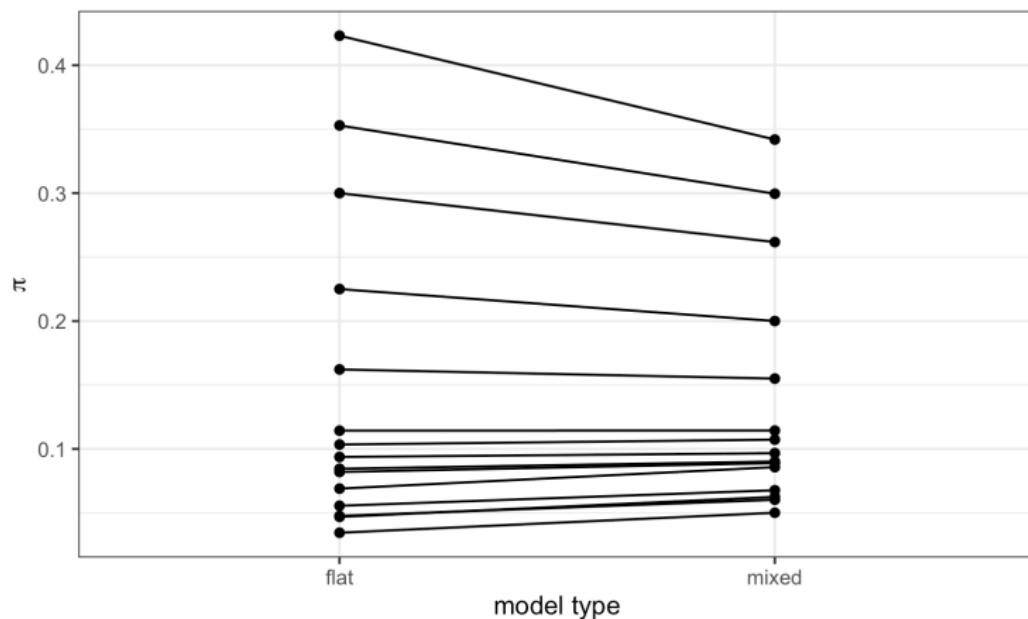
- In this model,  $\beta_{01}, \beta_{02}, \dots, \beta_{0,15}$  still represent the same: log-odds of disease incidence rate for herd  $i$
- However, they are drawn from the same normal distribution
- First consequence:
  - We now have a model of the *population* from which  $\beta_{01}, \beta_{02}, \dots, \beta_{0,15}$  are a sample
  - Since each  $\beta_{0i}$  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  $\beta_{0i}$  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  $\beta_{0i}$
  - Now the values of  $b$  and  $\sigma^2$  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,  $\beta_0$  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  $\varepsilon_{ij}$  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:  $\beta_0$ ,  $\sigma^2$ , and  $\sigma_b^2$
- 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,
  - $\sigma_b^2$  is the *between-country* variation
  - $\sigma^2$  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

## 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)
- 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: septoria data
  - split-plot design with four blocks
  - within each block, the preceding crop is the main plot; wheat cultivar, dose splitting and fertilisation level are randomised within the preceding crop
  - subplots are nested within previous crop, which are nested within blocks

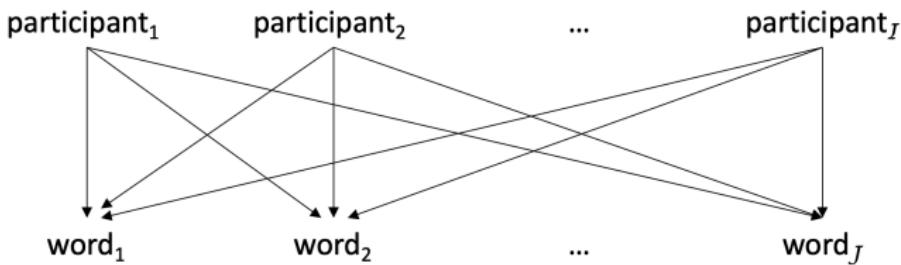
# Mixed Models for Nested Data

$$\begin{aligned} Y_{ijk} | \text{block}_i, \text{plot}_{ij} &\sim N(\mu_{ijk}, \sigma^2) \\ \mu_{ijk} &= \beta_0 + \text{block}_i + \text{plot}_{ij} + \text{fixed effects} \\ \text{block}_i &\sim N(0, \sigma_b^2) \\ \text{plot}_{ij} &\sim N(0, \sigma_p^2) \end{aligned}$$

- $i$  indexes blocks,  $j$  indexes plots within blocks, and  $k$  indexes subplots within plots within blocks
- We have 2 sets of normal random effects

# 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

# 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  $\mu$

# 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

## ■ Examples

- Septoria data (beta family)
- LAI yield (gamma family and smooth terms)
- Marker data (double gamma GLMM)
- Contamination data (negative binomial GLMM)

# 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