Illustrative formulae

Simplest model

$$y_i \sim \alpha + \beta x_i + (\epsilon_i)$$

Multilevel (hierarchical) models

(linear or generalised, with discrete or continuous variables)

May include cross-correlations, grouping dependencies, dynamical aspects (e.g. auto-correlations)

Most general formulation: $y = F(t, x, z, ...)$ and $y \sim \mathcal{P}$ or in *levels*, e.g.

$$y = F(t|A, B) + e$$
$$A = f_A(x, z|a^1)$$
$$B = f_B(x, z|b^1)$$

$a^1 = ... + a^2$ and $b^1 = ...$, where $e, a^1, a^2, ...$ are distributed according to:

- (multi-) variate Normal (when frequentist) distributions
- (multi-) variate Prior (when Bayesian) distributions

Example: simple MLM

Each subject is observed many times.

The response (y) of each subject is a linear function of time (at time points i).

The parameters (intercept and slope) of these functions have a normal distribution with higher level parameters $\mu_\alpha, \mu_\beta, ...$.

$$y_i \sim \alpha_{ji} + \beta_{1ji} t + \epsilon_{ij}(\sigma))$$

or

$$\begin{aligned} \mathrm{y}_i &\sim N\left(\alpha_{j[i]} + \beta_{1j[i]}(\mathrm{t}), \sigma^2\right) \\ \begin{pmatrix} \alpha_j \\ \beta_{1j} \end{pmatrix} &\sim N\left(\begin{pmatrix} \mu_{\alpha_j} \\ \mu_{\beta_{1j}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha_j} & \rho_{\alpha_j \beta_{1j}} \\ \rho_{\beta_{1j}\alpha_j} & \sigma^2_{\beta_{1j}} \end{pmatrix}\right), \text{ for Subject j} = 1, \ldots, \text{J} \end{aligned}$$

and

$$\mu \sim ...$$

Note on time-varying and time-constant predictors, while groups/clusters present

$$y \sim t + x_{tc} + x_{tv}^{B} + x_{tv}^{W} + z$$

$$+interact(t, x, z) + (1 + t|g) + (1 + x_{tv}^{W}|g)$$

Why Bayesian

The advantages of using Bayesian approach

- interpretation of results (while in frequentist approach is difficult, unless the model is linear, with no inverse link function and no interaction terms OR! unless we do *simulations*):
  - by inspecting the posterior distribution at different levels of predictors
  - being able to make probabilistic statements about a scientific hypothesis
- combining *all possible models*, according to:
  - posterior probability of models, given the data and
  - posterior probability of parameters, given all models and data, which gives
  - posterior mean and standard deviation of parameter of interest <-> point estimate and uncertainty