$\boxed{\text{Bayes LME}}$

The Bayesian version of the simple growth curve model results in a posterior that is not analytically tractable. We describe two MCMC approaches to implementation.

1. We could work with the marginal model with first stage

$$\boldsymbol{Y} \mid \boldsymbol{\beta}, \boldsymbol{\alpha} \sim \mathrm{N}_N\{\boldsymbol{x}\boldsymbol{\beta}, \boldsymbol{V}(\boldsymbol{\alpha})\},$$

   with $\boldsymbol{\alpha} = (\sigma_\epsilon^2, \sigma_0^2)$, and second stage

$$\boldsymbol{\beta} \sim \mathrm{N}_{k+1}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0), \quad \sigma_\epsilon^2 \sim \mathrm{Ga}(a_e, b_e), \quad \sigma_0^2 \sim \mathrm{Ga}(a_0, b_0).$$

   Leads to non-known form for $\sigma_\epsilon^2, \sigma_0^2$. Metropolis steps may be used for these parameters.

   Note that we can recover the posterior for $\boldsymbol{b}$ via (1).

2. Conditional model – conditional independencies may be exploited. Gibbs sampling iterates through:
   - $\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{b}, \sigma_\epsilon^2, \sigma_0^2 \propto N_{k+1}(\cdot, \cdot)$
   - $b_i \mid \boldsymbol{y}, \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_0^2 \propto N(\cdot, \cdot)$
   - $\sigma_\epsilon^{-2} \mid \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{b}, \sigma_0^2 \propto \mathrm{Ga}(\cdot, \cdot)$
   - $\sigma_0^{-2} \mid \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{b}, \sigma_\epsilon^2 \propto \mathrm{Ga}(\cdot, \cdot)$

   Note: often convergence is improved by parameterizing in terms of the "centered" set

$$\beta_{01} = \beta_0 + b_1, ..., \beta_{0m} = \beta_0 + b_m$$

   with $\beta_{0i} \mid \beta_0, \sigma_0^2 \sim N(\beta_0, \sigma_0^2)$.

`WinBUGS` *Model*

Priors:

$\beta_0, \beta_1 \propto 1$, $\sigma_\epsilon^2 \propto \sigma_\epsilon^{-2}$, $\sigma_0 \sim U(0, 100)$.

Note: improper priors on $\beta_0, \beta_1, \sigma_\epsilon^2$, but can't be improper on σ_0^2 also as it leads to an improper posterior. So, for example, the prior

$$\pi(\sigma_\epsilon^2, \sigma_0^2) \propto \frac{1}{\sigma_\epsilon^2 \sigma_0^2}$$

should never be used.

```
model
{
for( i in 1 : N ) {
for( j in 1 : T ) {
Y[i , j] ~ dnorm(mu[i , j],eps.tau)
mu[i , j] <- beta0[i] + beta1.mu * x[j]
}
beta0[i] ~ dnorm(beta0.mu,beta0.tau)
}
eps.tau <- exp(logtau)
logtau ~ dflat()
sigma <- 1 / sqrt(eps.tau)
beta0.mu ~ dflat()
beta1.mu ~ dflat()
beta0.sigma ~ dunif(0,100)
beta0.tau <-1/(beta0.sigma*beta0.sigma)
}
```

`WinBUGS` *Data and Initial Estimates*

```
list(x = c(8,10,12,14), N = 11, T = 4,
Y = structure(
.Data =   c(21,20,21.5,23,
21,21.5,24,25.5,
20.5,24,24.5,26,
23.5,24.5,25,26.5,
21.5,23,22.5,23.5,
20,21,21,22.5,
21.5,22.5,23,25,
23,23,23.5,24,
20,21,22,21.5,
16.5,19,19,19.5,
24.5,25,28,28),
.Dim = c(11,4)))

list(beta0 = c(18,18,18,18,18,18,18,18,18,18,18),
beta0.mu = 18, beta1.mu = .5,
logtau = 0, beta0.sigma = 1)
```


Results from iterations 1000–10000

node	mean	sd	MC error	2.5%	median	97.5%
beta0.mu	17.31	0.9844	0.04166	15.37	17.32	19.29
beta1.mu	0.485	0.0548	0.003584	0.3752	0.4851	0.593
beta0.sigma	2.409	0.6861	0.0106	1.465	2.277	4.111
sigma	0.800	0.1046	0.001876	0.6284	0.789	1.035

GEE (working independence):

$\widehat{\beta}_0 = 17.37\ (0.78),\ \widehat{\beta}_1 = 0.480\ (0.067)$.

REML:

$\widehat{\beta}_0 = 17.37\ (0.86), \widehat{\beta}_1 = 0.479\ (0.053)$

$\widehat{\sigma}_0 = 2.07, \widehat{\sigma}_\epsilon = 0.780$.

ML:

$\widehat{\beta}_0 = 17.37\ (0.85), \widehat{\beta}_1 = 0.480\ (0.053)$

$\widehat{\sigma}_0 = 1.97, \widehat{\sigma}_\epsilon = 0.768$.

Pretty consistent inference!

Linear Mixed Effects Models

For more details see: Hand and Crowder (Chapter 5), DHLZ (Sections 4.4, 4.5), Davison (Section 9.4.2), and Verbeke and Molenberghs.

A mixed effects model is characterized by a combination of fixed effects, $\boldsymbol{\beta}$, a $(k+1)\times 1$ vector, and random effects, $\boldsymbol{b}_i$, a $(q+1)\times 1$ vector.

Notation: Let $\boldsymbol{y}_i = (y_{i1}, ..., y_{in_i})^{\mathrm{T}}$, denote the vector of observations on unit i, $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i})^{\mathrm{T}}$, the design matrix for the fixed effect with $\boldsymbol{x}_{ij} = (1, x_{ij1}, ..., x_{ijk})^{\mathrm{T}}$, and $\boldsymbol{z}_i = (\boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{in_i})^{\mathrm{T}}$, and design matrix for the random effects with $\boldsymbol{z}_{ij} = (1, z_{ij1}, ..., z_{ijq})^{\mathrm{T}}$.

We then have the following (two stage) Linear Mixed Effects Model:

Stage 1: Response model, *conditional* on random effects:

$$\boldsymbol{y}_i = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{6}$$

where $\boldsymbol{\epsilon}_i$ is an $n_i \times 1$ zero mean vector of error terms.

Stage 2: Model for random terms:

$$\begin{aligned} \mathrm{E}[\boldsymbol{\epsilon}_i] &= \mathbf{0}, \ \ \mathrm{var}(\boldsymbol{e}_i) = \boldsymbol{E}_i(\boldsymbol{\alpha}), \\ \mathrm{E}[\boldsymbol{b}_i] &= \mathbf{0}, \ \ \mathrm{var}(\boldsymbol{b}_i) = \boldsymbol{D}(\boldsymbol{\alpha}), \\ \mathrm{cov}(\boldsymbol{b}_i, \boldsymbol{e}_i) &= \mathbf{0}. \end{aligned}$$

From these two stages we have the marginal model:

$$\begin{aligned} \mathrm{E}[\boldsymbol{y}_i] &= \boldsymbol{\mu}_i = \boldsymbol{x}_i\boldsymbol{\beta}, \\ \mathrm{var}(\boldsymbol{y}_i) &= \boldsymbol{V}_i = \boldsymbol{z}_i\boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}} + \boldsymbol{E}_i, \\ \mathrm{cov}(\boldsymbol{y}_i, \boldsymbol{y}_{i'}) &= \mathbf{0}, \quad i \neq i'. \end{aligned}$$

Maximum Likelihood Estimation

To implement MLE we need to specify a distribution for the data, and this follows by specifying distributions for $\boldsymbol{e}_i$ and $\boldsymbol{b}_i$.

Conventional to assume

$$\boldsymbol{\epsilon}_i \sim_{ind} N(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I}_{n_i}), \quad \boldsymbol{b}_i \sim_{iid} N(\mathbf{0}, \boldsymbol{D}),$$

where

$$\boldsymbol{D} = \begin{bmatrix} \sigma_{00}^2 & \sigma_{01}^2 & \ldots & \sigma_{0q}^2 \\ \sigma_{10}^2 & \sigma_{11}^2 & \ldots & \sigma_{1q}^2 \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{q0}^2 & \sigma_{q1}^2 & \ldots & \sigma_{qq}^2 \end{bmatrix}.$$

Growth curve example:

$$\text{cov}(b_{i0}, b_{i1}) = \sigma_{01}^2,$$

covariance between the intercepts and slopes.

We let $\boldsymbol{\alpha}$ denote all variance-covariance parameters, and write $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\alpha})$.

Log-likelihood follows from integrating the random effects from the two-stage model:

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \prod_{i=1}^m \int_{\boldsymbol{b}_i} p(\boldsymbol{y}_i | \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\alpha}) \times p(\boldsymbol{b}_i | \boldsymbol{\alpha}) \; \mathrm{d}\boldsymbol{b}_i,$$

to give

$$\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{\alpha} \sim \prod_{i=1}^m N\{\boldsymbol{x}_i \boldsymbol{\beta}, \boldsymbol{V}_i(\boldsymbol{\alpha})\},$$

and so

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \log L(\boldsymbol{\beta}, \boldsymbol{\alpha}) = -\frac{N}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^m \log |\boldsymbol{V}_i(\boldsymbol{\alpha})| \\ &- \frac{1}{2} \sum_{i=1}^m (\boldsymbol{Y}_i - \boldsymbol{x}_i \boldsymbol{\beta})' \boldsymbol{V}(\boldsymbol{\alpha})_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{x}_i \boldsymbol{\beta}). \qquad (7) \end{aligned}$$

MLE of $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^m \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{V}_i(\widehat{\boldsymbol{\alpha}})^{-1} \boldsymbol{x}_i \right)^{-1} \left(\sum_{i=1}^m \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{V}_i(\widehat{\boldsymbol{\alpha}})^{-1} \boldsymbol{y}_i \right).$$

MLE of $\boldsymbol{\alpha}$ follows from maximization of (7).

Note: if $\boldsymbol{D} = \mathbf{0}$ then $\widehat{\boldsymbol{\beta}}$ is the least squares estimator.

Implementation of MLE and REML

MLE and REML require iteration between $\widehat{\boldsymbol{\beta}}|\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\alpha}}|\widehat{\boldsymbol{\beta}}$.

Originally the *EM algorithm* was used (e.g. Laird and Ware (1982, *Biometrics*).

The fixed and random effect estimates are available in closed form once we know $\boldsymbol{\alpha}$.

Slow convergence has been reported so that now the *Newton-Raphson method* is more frequently used.

Let $\boldsymbol{\theta}$ be a $p \times 1$ parameter vector, $l(\cdot)$ the log-likelihood, $\boldsymbol{G}$ the $p \times 1$ score vector, and $\boldsymbol{I}^\star(\cdot)$ the $p \times p$ observed information matrix. Then a second order Taylor series expansion of l about $\boldsymbol{\theta}^{(t)}$, the estimate at iteration t gives:

$$\boldsymbol{g}^{(t)}(\boldsymbol{\theta}) = l(\boldsymbol{\theta}) + \boldsymbol{G}^{(t)\mathrm{T}}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(t)}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(t)})^{\mathrm{T}}\boldsymbol{I}^{\star(t)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(t)}),$$

differentiating and setting equal to zero:

$$\frac{\partial \boldsymbol{g}^{(t)}}{\partial \boldsymbol{\theta}} = \boldsymbol{G}^{(t)} + \boldsymbol{I}^{\star(t)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(t)}) = \boldsymbol{0},$$

gives the next estimate

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \{\boldsymbol{I}^{\star(t)}\}^{-1}\boldsymbol{G}^{(t)}.$$

The use of the expected information gives *Fisher's scoring method.*

See Lindstrom and Bates (1988, *JASA*) for details.

Lack of convergence of the algorithm/negative estimates, may sometimes indicate that a poor model is being fitted.

The EM Algorithm

We illustrate for MLE and, for example, suppose $\boldsymbol{E}_i = \boldsymbol{I}_{n_i}\sigma^2$. The "missing data" here are the random effects $\boldsymbol{b}_i$ and the errors $\boldsymbol{\epsilon}_i$.

The M-step: Given $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$, obtain estimates $\widehat{\boldsymbol{\alpha}} = (\widehat{\sigma}^2, \widehat{\boldsymbol{D}})$:

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{\sum_{i=1}^m \boldsymbol{\epsilon}_i^{\mathrm{T}}\boldsymbol{\epsilon}_i}{\sum_{i=1}^m n_i} = \frac{t_1}{N} \\ \widehat{\boldsymbol{D}} &= \frac{1}{m}\sum_{i=1}^m \boldsymbol{b}_i\boldsymbol{b}_i^{\mathrm{T}} = \frac{\boldsymbol{t}_2}{m}, \end{aligned}$$

where t_1 and $\boldsymbol{t}_2$ are the sufficient statistics.

The E step: Estimate the sufficient statistics given the current values $\widehat{\boldsymbol{\alpha}}$, via their expected values:

$$\begin{aligned} \widehat{t}_1 &= \mathrm{E}\left[\sum_{i=1}^m \boldsymbol{\epsilon}_i^{\mathrm{T}}\boldsymbol{\epsilon}_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right] \\ \widehat{\boldsymbol{t}}_2 &= \mathrm{E}\left[\sum_{i=1}^m \boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{b}_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}}), \widehat{\boldsymbol{\alpha}}\right]. \end{aligned}$$

Estimation of Random Effects

See Verbeke and Molenbergs (Chapter 7) and Robinson (1991, "That BLUP is a good thing: the estimation of random effects", *Statistical Science*.

Preparation: Suppose $\boldsymbol{U}$ is an $n \times 1$ vector of random variables, and $\boldsymbol{V}$ is an $m \times 1$ vector of random variables. Then $\text{cov}(\boldsymbol{U}, \boldsymbol{V}) = \boldsymbol{C}$ is an $n \times m$ matrix with (i, j)-th element $\text{cov}(U_i, V_j)$, $i = 1, ..., n; j = 1, ..., m$. Also $\text{cov}(\boldsymbol{V}, \boldsymbol{U}) = \boldsymbol{C}^{\mathrm{T}}$.

Now suppose $\boldsymbol{V} = \boldsymbol{AU}$ where $\boldsymbol{A}$ is an $m \times n$ matrix. Then $\text{cov}(\boldsymbol{U}, \boldsymbol{AU}) = \boldsymbol{W}\boldsymbol{A}^{\mathrm{T}}$ where $\boldsymbol{W} = \text{cov}(\boldsymbol{U})$, and $\text{cov}(\boldsymbol{AU}, \boldsymbol{U}) = \boldsymbol{AW}$.

Empirical Bayes

Since $\boldsymbol{b}_i$ are random effects, it is natural (though not essential) to use Bayesian methods for estimation.

We suppose first that $\boldsymbol{\beta}, \boldsymbol{\alpha}$ are known. Then we have seen that the estimator that minimizes the mean squared error is

$$\widehat{\boldsymbol{b}}_i = \mathrm{E}[\boldsymbol{b}_i | \boldsymbol{y}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}].$$

With $\boldsymbol{\epsilon}_i \sim_{ind} \mathrm{N}(\boldsymbol{0}, \boldsymbol{E}_i)$ and $\boldsymbol{b}_i \sim_{ind} \mathrm{N}(\boldsymbol{0}, \boldsymbol{D})$ and $\text{cov}(\boldsymbol{\epsilon}_i, \boldsymbol{b}_i) = \boldsymbol{0}$, we have

$$\begin{bmatrix} \boldsymbol{b}_i \\ \boldsymbol{y}_i \end{bmatrix} \sim \mathrm{N}_{q+1+n_i} \left(\begin{bmatrix} 0 \\ \boldsymbol{x}_i\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \boldsymbol{D} & \boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}} \\ \boldsymbol{z}_i\boldsymbol{D} & \boldsymbol{V}_i \end{bmatrix} \right),$$

since

$$\text{cov}(\boldsymbol{b}_i, \boldsymbol{y}_i) = \text{cov}(\boldsymbol{b}_i, \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i) = \text{cov}(\boldsymbol{b}_i, \boldsymbol{z}_i\boldsymbol{b}_i) = \boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}},$$

and similarly $\text{cov}(\boldsymbol{y}_i, \boldsymbol{b}_i) = \boldsymbol{z}_i\boldsymbol{D}$.

Hence from properties of a multivariate normal distribution $\boldsymbol{b}_i|\boldsymbol{y}_i$ is normal with

$$\begin{aligned} \mathrm{E}[\boldsymbol{b}_i|\boldsymbol{y}_i] &= \boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta}), \\ \mathrm{var}(\boldsymbol{b}_i|\boldsymbol{y}_i) &= \boldsymbol{D} - \boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{V}_i^{-1}\boldsymbol{z}_i\boldsymbol{D}. \end{aligned}$$

A matrix identity that is often used in the context of estimation of $\boldsymbol{b}_i$ is

$$\begin{aligned} \boldsymbol{V}_i^{-1} &= (\boldsymbol{E}_i^{-1} + \boldsymbol{z}_i\boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}})^{-1} \\ &= \boldsymbol{E}_i^{-1} - \boldsymbol{E}_i^{-1}\boldsymbol{z}_i(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{E}_i^{-1}\boldsymbol{z}_i + \boldsymbol{D}^{-1})^{-1}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{E}_i^{-1}. \end{aligned}$$

see Searle, Casella and McCulloch (1991, p. 453).

From this identity we may derive

$$(\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{E}_i^{-1}\boldsymbol{z}_i + \boldsymbol{D}^{-1})^{-1}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{E}_i^{-1} = \boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}}(\boldsymbol{E}_i^{-1} + \boldsymbol{z}_i\boldsymbol{D}\boldsymbol{z}_i^{\mathrm{T}})^{-1},$$

so that another expression for the estimate of $\boldsymbol{b}_i$ is

$$\mathrm{E}[\boldsymbol{b}_i|\boldsymbol{y}_i] = (\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{E}_i^{-1}\boldsymbol{z}_i + \boldsymbol{D}^{-1})^{-1}\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{E}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta}).$$

In practice, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are replaced by estimates, to give

$$\mathrm{E}[\boldsymbol{b}_i|\boldsymbol{y}_i] = \boldsymbol{D}(\widehat{\boldsymbol{\alpha}})\boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{V}_i(\widehat{\boldsymbol{\alpha}})^{-1}(\boldsymbol{y}_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}}).$$

No easy way of accounting for extra uncertainty in estimation of $\boldsymbol{\beta}, \boldsymbol{\alpha}$ – so interval estimates for $\widehat{\boldsymbol{b}}_i$ will be too short.

Random effects estimates may be used to assess model assumptions such as normality, and constant variance – but don't forget that these are estimates (not observed).

Fixed or Random Effects?

The most natural way of thinking about random effects is as parameters that are associated with units which may be thought of as drawn from some hypothetical infinite population.

Hence it may be natural to think of parameters associated with tomato plants, workers in a factory, clinics in a city, as random effects which are drawn from some common distribution.

However, if we are interested in (say) the effects of four treatment arms, then it may or may not be reasonable to think of them as being drawn from a population of treatment effects. If, for example, one of them is a control then we may feel that the effect of this arm is not "similar" to the other treatments – this notion is formalized through the Bayesian notion of *exchangeability* which we discuss later.

Advantages of Random Effects

In the way we have developed the random effects formulation as a way of modeling dependencies – the inclusion of random effects induced dependencies on responses on the same unit. We could have allowed for the dependencies by allowing each unit to have its own set of fixed effects, however.

There are a number of reasons why we may want to consider a random effects formulation:

- We are interested in making inference about the *population* from which the individual effects were drawn.
- We wish to make inference about a particular unit and wish to make use of information from the other units (which recall are viewed as similar) – this is particularly true when the data on a unit of interest is sparse.

Random effects models provide an *economical* way of modeling dependencies. For example, consider the simple one-way ANOVA model:

Stage 1: $Y_{ij} \sim_{ind} \text{N}(\mu + b_i, \sigma_\epsilon^2)$, $i = 1, ..., m$, $j = 1, ..., n$.

Stage 2: $b_i \sim_{iid} \text{N}(0, \sigma_0^2)$, $i = 1, ..., m$.

Does this model have $m + 3$ parameters, or 3?

A fixed effects model with an effect for each unit would have $m + 2$ parameters (can think of this as the above with $\sigma_\alpha^2 = \infty$).

By assuming a common distribution we have "tied" the m random effect parameters together.

More Flexible Covariance Structures

We discuss covariance models in the context of longitudinal data (though aspects of the discussion are relevant to other types of clustered data).

Whether we take a GEE or LME approach (with inference from the likelihood or from the posterior) we require flexible yet parsimonious covariance models.

In GEE we require a working covariance model

$$\text{cov}(\boldsymbol{Y}_i) = \boldsymbol{W}_i,$$

$i = 1, ..., m$.

With LME we have so far assumed the model

$$\boldsymbol{y}_i = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{8}$$

with $\boldsymbol{b}_i \sim_{ind} \text{N}(\boldsymbol{0}, \boldsymbol{D})$ and $\boldsymbol{\epsilon}_i \sim_{ind} \text{N}(\boldsymbol{0}, \boldsymbol{E}_i)$, with $\boldsymbol{E}_i = \boldsymbol{I}_{n_i}\sigma^2$.

With $\boldsymbol{z}_i\boldsymbol{b}_i = \boldsymbol{1}_{n_i}b_i$ we obtained an *exchangeable* correlation structure.

An obvious extension for longitudinal data is to assume

$$\boldsymbol{y}_i = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{b}_i + \boldsymbol{\delta}_i + \boldsymbol{\epsilon}_i,$$

with:

- Random effects $\boldsymbol{b}_i \sim_{ind} \text{N}(\boldsymbol{0}, \boldsymbol{D})$.
- Serial correlation $\boldsymbol{\delta}_i \sim_{ind} \text{N}(\boldsymbol{0}, \boldsymbol{R}_i\sigma_\delta^2)$, with $\boldsymbol{R}_i$ an $n_i \times n_i$ correlation matrix with elements
$$R_{ijj'} = \text{corr}(Y_{ij}, Y_{ij'}|\boldsymbol{b}_i),$$
$j, j' = 1, ..., n_i$.
- Measurement error $\boldsymbol{\epsilon}_i \sim_{ind} \text{N}(0, \boldsymbol{I}_{n_i}\sigma_\epsilon^2)$.

In general it is difficult to identify all three sources of variability – but the above provides a useful conceptual model.

Examination of Covariance Structure

Consider a stochastic process $Y(t)$ and let

$$\gamma(t,s) = \text{cov}\{Y(t), Y(s)\} = \text{E}[\{Y(t) - \mu(t)\}\{Y(s) - \mu(s)\}],$$

denote the *autocovariance function* of $Y(t)$. The term *serial dependence* says that there is dependence between $Y(t)$ and $Y(s)$ for at least some pairs (s,t) with $s \neq t$.

We write

$$Y(t) = \mu(t) + r(t),$$

where $\mu(t)$ is the trend and $r(t)$ is a residual process.

Such a process is second-order stationary if $\text{E}[r(t)]$ is equal to a constant (which we take to be zero, any intercept being absorbed into $\mu(t)$), for all t, and $\gamma(t,s)$ depends only on $|t-s|$.

Example: The simplest example of a stationary random sequence is *white noise* which consists of a sequence of mutually independent random variables, each with mean 0 and finite variance σ^2.

There is a fundamental difficulty with trying to decompose $Y(t)$ into the trend and the stochastic component in a single series because the two are unidentifiable without further assumptions.

Is it serial dependence in the residuals, or a high-order polynomial trend for example?

The Autocorrelation Function

For a second-order stationary random process, the autocovariance function is

$$\gamma(u) = \text{cov}\{Y(t), Y(t+u)\},$$

so that $\gamma(0)$ is the variance of $Y(t)$ for all t.

The autocorrelation function is defined as

$$\rho(u) = \frac{\gamma(u)}{\gamma(0)}.$$

For equally-spaced data we could fit a model and then examine the autocrrelation function (ACF) of the residuals,

$$r_t = \frac{y_t - \widehat{y}_t}{\widehat{\text{var}}(Y_t)^{1/2}}.$$

Consider a stochastic process $r(t)$, and realizatons r_t, $t = 1, ..., n$. The *emprical* autocorrelation is defined as

$$\widehat{\rho}(u) = \widehat{\text{corr}}\{r(t), r(t+u)\} = \frac{\sum_{t=1}^{n-u} r_t r_{t+u}/(n-u)}{\sum_{t=1}^{n} r_t^2/n},$$

$u = 0, 1, ...$

A *correlogram* plot is $\widehat{\rho}(u)$ versus u. If the residuals are a white noise process, we have the asymptotic result

$$\sqrt{n}\, r_t \ \to_d \ \text{N}(0,1),$$

to give confidence bands $\pm 2/\sqrt{n}$.