

6. Flexible extensions of the LLMM

Sonja Greven

Summer Term 2015

We have so far made several modeling assumptions in our longitudinal linear mixed model (LLMM) (3.5).

- While model (3.5) assumes a general variance-covariance matrix Σ_i for the errors ε_i , we have typically assumed $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$ in the following.
- We have assumed a linear relationship between \mathbf{Y}_i and the (transformed) covariates in \mathbf{X}_i .
- We have assumed a normal distribution for the random effects \mathbf{b}_i .

What if these assumptions are violated? This chapter discusses more flexible extensions of the LLMM.

Overview Chapter 6 - Flexible extensions of the LLMM

6.1 Smooth models for the mean

6.2 Serial correlation

6.3 Non-normal random effects

Penalized splines in a nutshell

- Consider the **nonparametric regression problem**

$$Y_i = m(x_i) + \epsilon_i, \quad i = 1, \dots, n,$$

$\epsilon_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$, with unknown **smooth** function $m(\cdot)$.

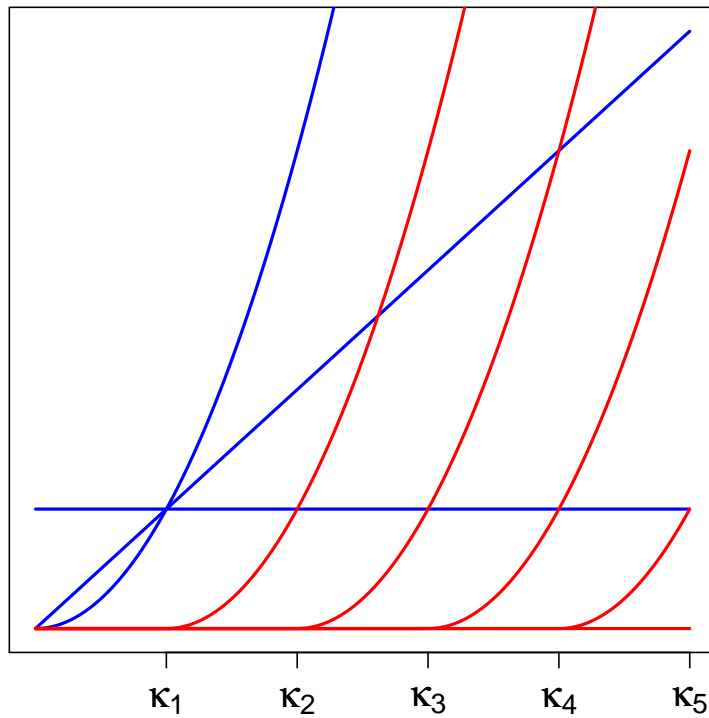
- Approximate $m(\cdot)$ by a linear combination of (many) **spline basis functions**, e.g. truncated polynomials,

$$m(x) \approx \sum_{j=0}^{p-1} \beta_j x^j + \sum_{k=1}^q b_k (x - \kappa_k)_+^{(p-1)},$$

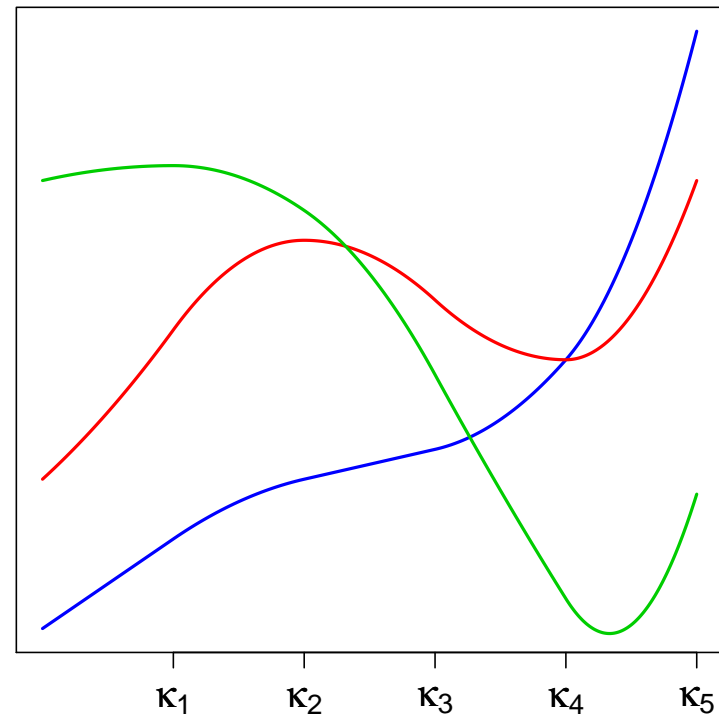
where $\kappa_1, \dots, \kappa_q$ is a sequence of knots, and $u_+^{(p-1)} = (\max\{0, u\})^{(p-1)}$.

Penalized splines in a nutshell

Quadratic Spline Basis



Three Example Functions



Penalized splines in a nutshell

- This approximation gives a piecewise polynomial function of degree $(p-1)$ with certain smoothness properties (continuity of $(p-2)$ th derivative. The b_k correspond to jumps in the $(p-1)$ th derivative.)
- Estimation using a **regularization penalty**, to avoid overly wiggly function:

$$\min_{\boldsymbol{\beta}, \mathbf{b}} \{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) + \lambda \mathbf{b}^T \mathbf{b}\}, \quad (6.1)$$

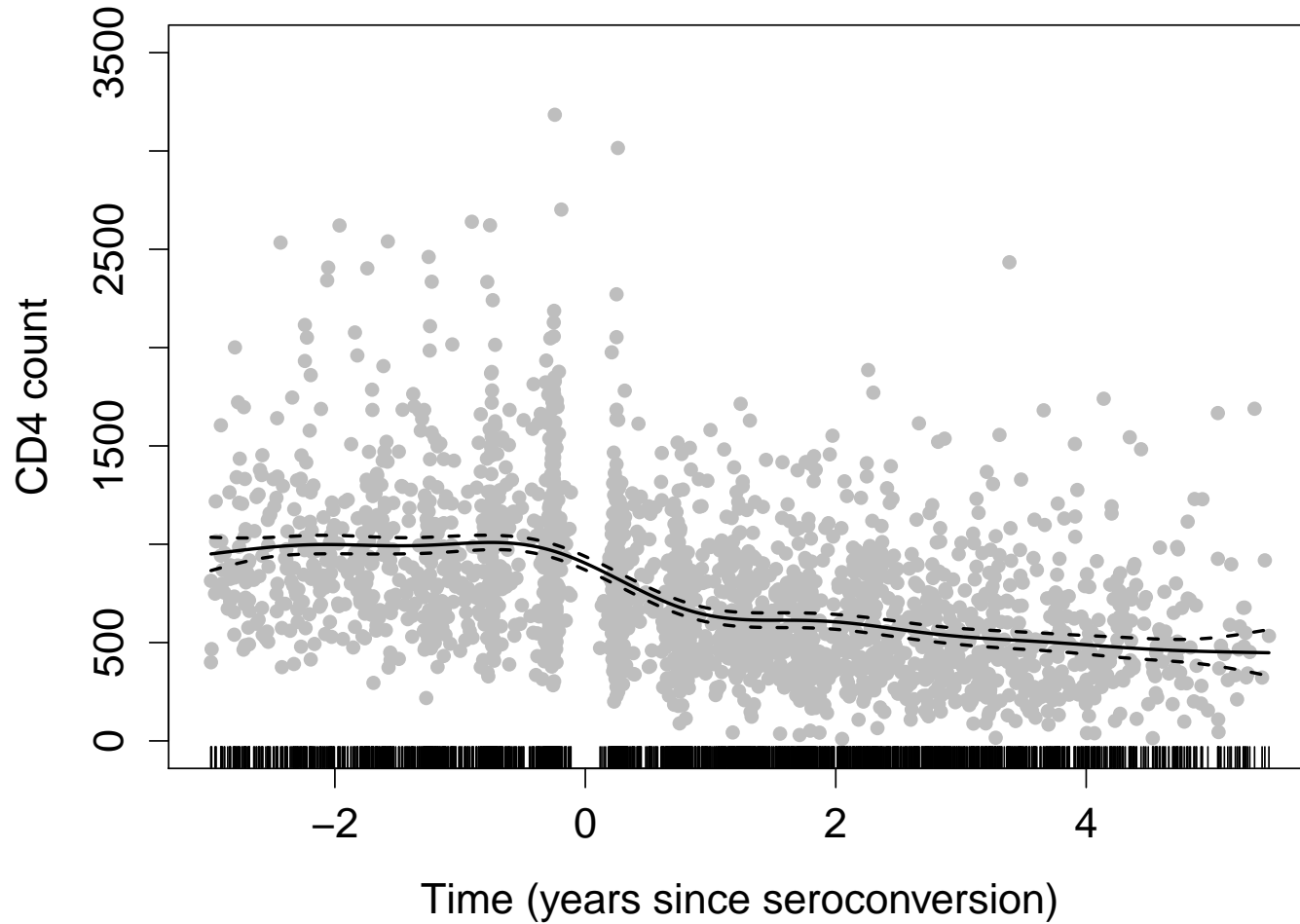
- \mathbf{X} and \mathbf{Z} design matrices for the spline basis functions x^j resp. $(x - \kappa_k)_+^{(p-1)}$
- λ a smoothing parameter.

- After dividing by $(-2\sigma^2)$, (6.1) is equivalent to the penalized log-likelihood (4.3) and thus to obtaining the BLUE and BLUP for β and b in the LMM with
 - **fixed effects** $\beta_0, \dots, \beta_{(p-1)}$: subspace of polynomial functions (degree $(p - 1)$),
 - **random effects** $b_1, \dots, b_q \stackrel{iid}{\sim} N(0, d^2)$: deviations from subspace,
 with $d^2 = \sigma^2/\lambda$. We can thus estimate the smoothing parameter $\lambda = \sigma^2/d^2$ data-driven using (RE)ML in this LMM.
- Analogous for other basis choices (e.g. B-Splines with a difference penalty) and for spatial effects, interaction surfaces, varying coefficients, . . . → can combine these with random effects for the longitudinal data in a general LMM ([additive mixed model](#)).

- Examples: smooth mean over time, smooth dose-response functions for covariates, smoothly time-varying effects, spatial surfaces (spatial information on subjects),
- More detailed information: Lecture **Mixed Models** (September 7-11) and books [Ruppert, Wand & Carroll, 2003](#); [Wood, 2006](#).
- In R: packages `mgcv` and `gamm4`.
- Example smooth mean for the CD4 data (and random intercept per subject), fit with the `mgcv` package:

```
gamm(CD4 ~ s(Time), random = list(ID = ~ 1), data = cd4,  
      method = "REML") #ML is default
```


Example: The CD4 data



Overview Chapter 6 - Flexible extensions of the LLMM

6.1 Smooth models for the mean

6.2 Serial correlation

6.3 Non-normal random effects

Introduction serial correlation

Remember the **longitudinal linear mixed model** (3.5):

$$\left\{ \begin{array}{l} \mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \\ \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}), \\ \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i), \\ \mathbf{b}_1, \dots, \mathbf{b}_N, \boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_N \text{ are independent.} \end{array} \right.$$

The implied marginal covariance matrix for \mathbf{Y}_i is

$$\text{Cov}(\mathbf{Y}_i) = \mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i$$

and the implied conditional covariance is

$$\text{Cov}(\mathbf{Y}_i | \mathbf{b}_i) = \boldsymbol{\Sigma}_i.$$

Introduction serial correlation

If $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$, this implies **conditional independence** of the Y_{ij} . When b_i consists of a random intercept, this implies compound symmetry or **equal marginal correlation** between any two measurements on the same subject.

A more general assumption is to decompose

$$\epsilon_i = \epsilon_{(1)i} + \epsilon_{(2)i}, \quad \epsilon_{(1)i} \text{ independent of } \epsilon_{(2)i},$$

with $\epsilon_{(2)i}$ representing serial correlation and $\epsilon_{(1)i}$ an additional i.i.d. error. If $\text{Cov}(\epsilon_{(2)i}) = \tau^2 \mathbf{H}_i$, this implies the covariance decomposition

$$\Sigma_i = \tau^2 \mathbf{H}_i + \sigma^2 \mathbf{I}_{n_i}.$$

The model with i.i.d. error terms ($\tau^2 = 0$) and the model with serial correlation only ($\sigma^2 = 0$) occur as special cases.

Serial correlation

We are now interested in modeling the serial correlation, i.e. the covariance structure $\text{Cov}(\epsilon_{(2)i}) = \tau^2 \mathbf{H}_i$.

One typically assumes a structure that depends only on the temporal distance, with the (j, k) th element of \mathbf{H}_i equal to

$$h_{ijk} = g(|t_{ij} - t_{ik}|)$$

and g a decreasing function with $g(0) = 1$.

The question then is how to choose g .

Models for the serial correlation

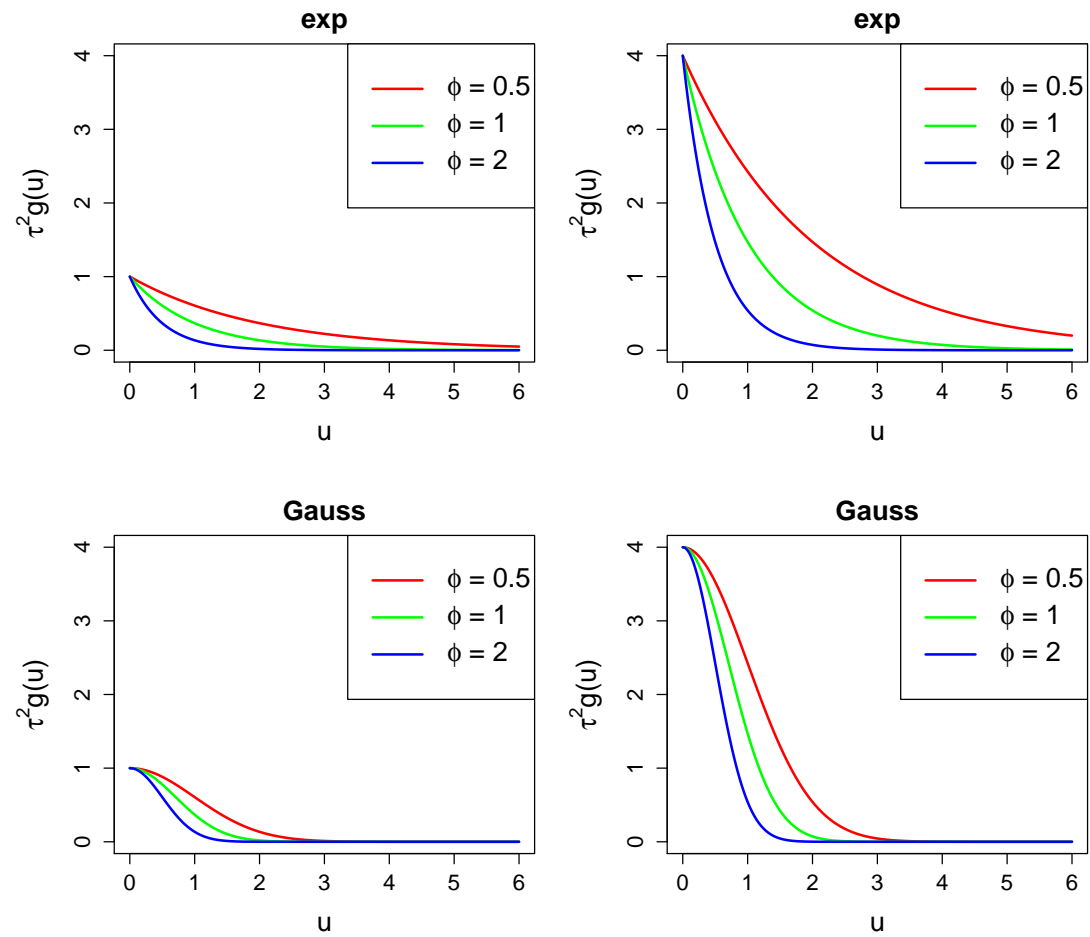
Two commonly used covariance structures are

$$\tau^2 g(u) = \tau^2 \exp(-\phi u) \quad (\text{exponential})$$

$$\tau^2 g(u) = \tau^2 \exp(-\phi u^2) \quad (\text{Gaussian})$$

- Large differences for small u
- Only two parameters: τ and ϕ
- Criterion for choosing between the two g : Maximum likelihood

Models for the serial correlation



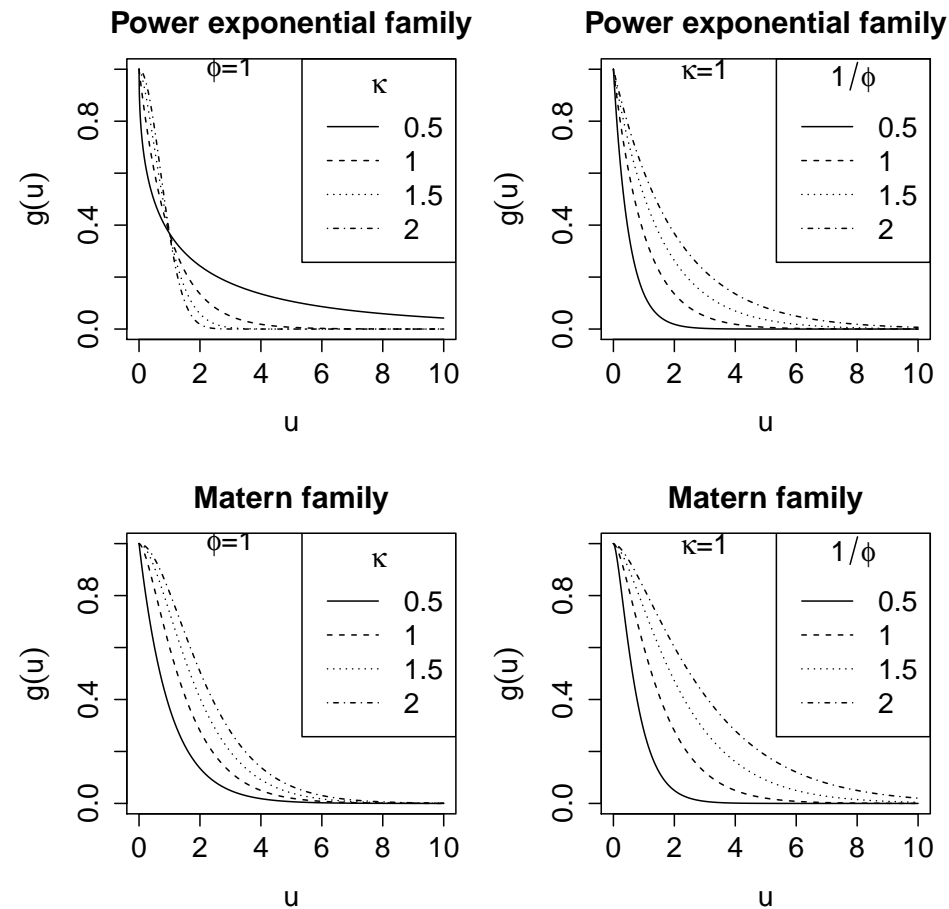
Models for the serial correlation

There are several extensions to model the correlation flexibly. Exponential and Gaussian correlation functions are special cases of the two parameter **power exponential family**

$$g(u) = \exp(-\phi|u|^\kappa), \quad \phi > 0, 0 < \kappa \leq 2$$

for $\kappa = 1, 2$. Another very flexible two-parameter family is the **Matérn family**. In both cases, ϕ is a scale parameter and κ a shape parameter.

Models for the serial correlation



Models for the serial correlation

A third approach (Royston & Altman, 1994; Lesaffre, Asefa & Verbeke (1999)) uses so-called **fractional polynomials**

$$\tau^2 g(u) = \exp\left\{\phi_0 + \sum_{k=1}^m \phi_k u^{(p_k)}\right\}$$

with $\tau^2 = \exp(\phi_0)$ and

$$x^{(p_j)} = \begin{cases} x^{p_j} & \text{if } p_j > 0 \\ \ln(x + 1) & \text{if } p_j = 0 \\ (x + 1)^{p_j} - 1 & \text{if } p_j < 0 \end{cases} .$$

Correlation structures in R

lme allows general covariances Σ_i through the argument `correlation`. The `corStruct` object assumes $\Sigma_i = \tau^2 \mathbf{H}_i$, i.e. $\sigma^2 = 0$, no additional i.i.d. error term. Some of the available `corClasses`:

- **corCompSymm** and **corSymm**: constant respectively general unstructured correlation
- **corAR1** and **corARMA**: AR(1)- resp. ARMA(p,q)-correlation, time variable must be an integer
- **corCAR1**: “continuous time AR(1)”
- **corExp** and **corGaus**: Exponential and Gaussian correlation functions

From the `nlme` manual: “Users may define their own `corStruct` classes by specifying a constructor function [...]”.

A note of caution

In applications, the effect of serial correlation is often dominated by the combination of random effects and measurement error (also causing convergence problems when all are present in the model).

Keep in mind that the total number of covariance parameters in α should be relatively small unless there are many repeated observations per subject.

Example orthodont data: a model with random intercept and slope, i.i.d. error plus serial correlation yields at least $3 + 1 + 2 = 6$ parameters for the 10 disjunct entries in V_i . For $n_i \equiv 4$ observations per subject it is difficult to distinguish between different serial correlations.

Example orthodont data

```
mycorr1 <- corExp(form = ~ 1|Subject)
mycorr2 <- corGaus(form = ~ 1|Subject)

mymodel0 <- lme(distance ~ age + Sex, data = Orthodont,
               random=~1|Subject)

mymodel1 <- lme(distance ~ age + Sex, data = Orthodont,
               random=~1|Subject, corr=mycorr1)

mymodel2 <- lme(distance ~ age + Sex, data = Orthodont,
               random=~1|Subject, corr=mycorr2)

anova(mymodel0, mymodel1, mymodel2)
```

	Model	df	logLik	
yields:	mymodel0	0	5	-218.7562
	mymodel1	1	6	-218.6984
	mymodel2	2	6	-218.7156

Do we need serial correlation? The semi-variogram

We can assess serial correlation using the **semi-variogram**, also sometimes called variogram.

We can view the $Y_{ij} = Y_i(t_{ij}), j = 1, \dots, n_i$, as samples from underlying continuous-time **stochastic processes** $\{Y_i(t), t \in \mathbb{R}\}$.

The variogram can be used to characterize the covariance of a stochastic process $\{Y(t), t \in \mathbb{R}\}$:

$$v(u) = \frac{1}{2} \mathbf{E}[\{Y(t) - Y(t - u)\}^2], \quad u \geq 0.$$

It is well-defined for stationary processes (i.e. the joint probability distribution does not change when shifted in time) and some non-stationary processes (with stationary increments $Y(t) - Y(t - u)$).

The semi-variogram for the random intercept model

In the random intercept model

$$Y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + b_i + \epsilon_{(1)ij} + \epsilon_{(2)ij}$$

with i.i.d. $\epsilon_{(1)ij}$ and serially correlated $\epsilon_{(2)ij}$, consider the centered $Y_{ij}^c = Y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}$. We have

$$\begin{aligned} \text{Var}(Y_{ij}^c) &= d^2 + \sigma^2 + \tau^2 \\ \text{Cov}(Y_{ij}^c, Y_{ik}^c) &= d^2 + \tau^2 g(|t_{ij} - t_{ik}|), \quad j \neq k. \end{aligned}$$

We thus have a stationary process, as the mean is 0 (constant), the variance is constant, the covariance only depends on the temporal distance, and as these together determine the joint distributions for the Gaussian case.

We then obtain

$$\begin{aligned}v(|t_{ij} - t_{ik}|) &= \frac{1}{2}\mathbf{E}[(Y_{ij}^c - Y_{ik}^c)^2] \\&= \frac{1}{2}[\text{Var}(Y_{ij}^c) + \text{Var}(Y_{ik}^c) - 2\text{Cov}(Y_{ij}^c, Y_{ik}^c)] \\&= d^2 + \sigma^2 + \tau^2 - [d^2 + \tau^2 g(|t_{ij} - t_{ik}|)] \\&= \sigma^2 + \tau^2(1 - g(|t_{ij} - t_{ik}|))\end{aligned}$$

for $i = 1, \dots, N$ and $j \neq k$ and thus

$$v(u) = \sigma^2 + \tau^2(1 - g(u)).$$

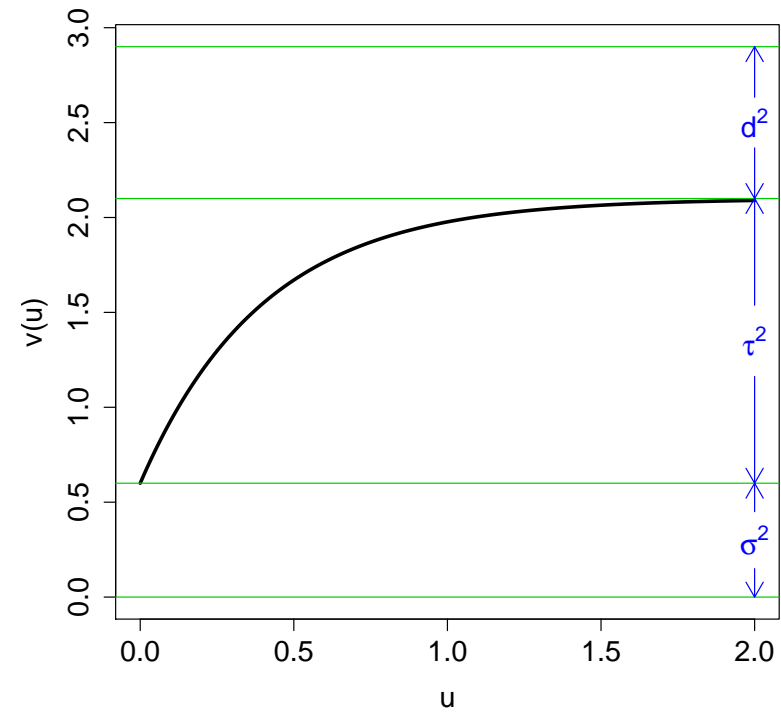
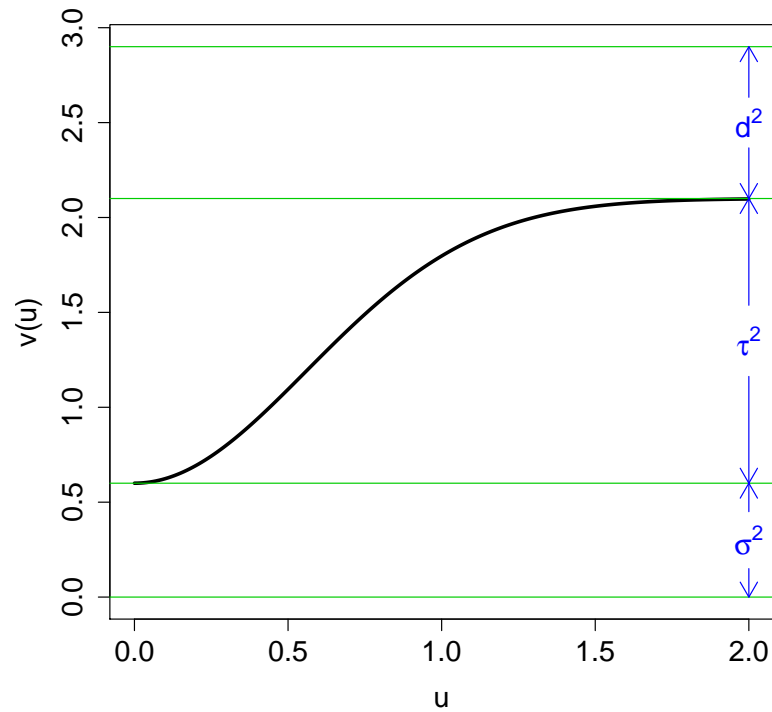
Properties:

- A decreasing g results in an increasing semi-variogram.
- $v(0) = \sigma^2$ (“nugget effect”)
- $\lim_{u \rightarrow +\infty} v(u) = \sigma^2 + \tau^2 \leq \text{Var}(Y_{ij}^c) = d^2 + \sigma^2 + \tau^2$

The variogram plus the variance thus give us information on all unknown components d^2 , σ^2 , τ^2 and $g(\cdot)$.

How would the semi-variogram look for purely serial correlation (i.e. only $\epsilon_{(2)i}$)?

The semi-variogram



Which shape does $g(\cdot)$ have in these two plots?

The empirical semi-variogram

The function $v(\cdot)$ is unknown in practice and has to be estimated.

- Consider the $\sum_{i=1}^N n_i(n_i - 1)/2$ half-squared differences

$$v_{ijk} = (r_{ij} - r_{ik})^2/2 \quad \text{for } i = 1, \dots, N$$

between pairs of residuals $r_{ij} = y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}$ and r_{ik} on the same subjects.

- Plot v_{ijk} against $|t_{ij} - t_{ik}|$.
- Smooth $(|t_{ij} - t_{ik}|, v_{ijk})$. The resulting smooth function $\hat{v}(\cdot)$ is the **empirical semi-variogram**.

The empirical semi-variogram

As for $i \neq l$

$$\frac{1}{2}E(r_{ij} - r_{lk})^2 = \sigma^2 + \tau^2 + d^2,$$

the total variance can be estimated as

$$\hat{\sigma}^2 + \hat{\tau}^2 + \hat{d}^2 = \frac{1}{2N^*} \sum_{i \neq l} \sum_{j=1}^{n_i} \sum_{k=1}^{n_l} (r_{ij} - r_{lk})^2$$

with $N^* = n^2 - \sum_{i=1}^N n_i^2$.

The empirical semi-variogram

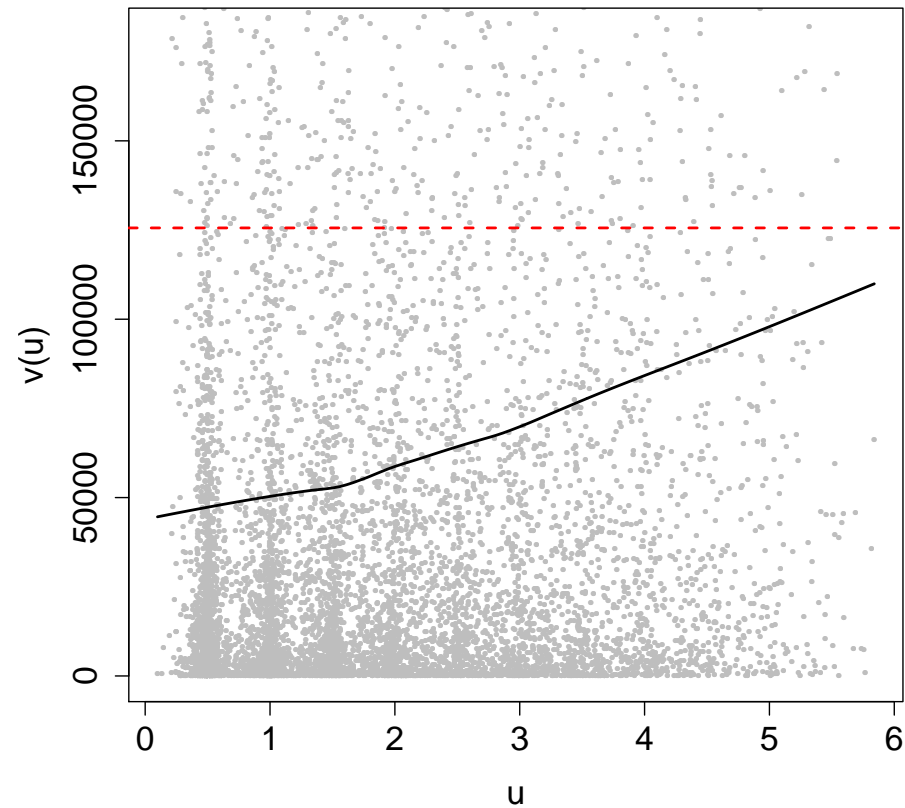
- The semi-variogram gives starting values for τ , σ and d useful in numerical optimization schemes.
- The comparison of the empirical semi-variogram with a fitted variogram gives an informal plausibility check for the assumed covariance structure.
- The extrapolation of the variance σ^2 from the semi-variogram is only feasible if there are enough pairs of observations with small time distance.
- In R: Function Variogram in package nlme or own code.

The empirical semi-variogram for the CD4 data

Empirical semi-variogram with lowess smoother for the CD4 cell count data:

```
> mygamm <- gamm(CD4 ~ s(Time), data = cd4, method = "REML")
> r = resid(mygamm$lme); t = cd4$Time; id = cd4$ID
> half.squared<- outer(r, r, FUN = function(r1, r2) (r1 - r2)^2/2)
> distances    <- outer(t, t, FUN = function(t1, t2) abs(t1 - t2))
> samei        <- outer(id, id, FUN = function(i1, i2) (i1 == i2))
> nondiag      <- 1 - diag(length(r))
> u <- as.vector(distances[samei & nondiag])
> v <- as.vector(half.squared[samei & nondiag])
> totvar <- mean(half.squared[!samei])
> smooth_mean <- lowess(u,v, iter = 0)
> plot(u, v, pch = 16, cex = 0.5, col = 8, ylim = c(0, 180000))
> lines(smooth_mean, lwd = 2)
> abline(h = totvar, col = 2, lwd = 2, lty = 2)
```

The empirical semi-variogram for the CD4 data



The semi-variogram for models with more random effects

If there are additional random effects in the model, Y_{ij}^c is no longer stationary (e.g. the variance increases with t_{ij} in the presence of a random slope in t_{ij}).

We could consider the subject-specific residuals $y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} - \mathbf{z}_{ij}^T \hat{\mathbf{b}}_i$ to estimate the variogram $v(\cdot)$ for ϵ_{ij} . However, $\hat{\mathbf{b}}_i$ is sensitive to both the normality assumption and the assumed \mathbf{V}_i in fitting.

See [Verbeke & Molenberghs, 2000](#), Section 10.4.4., for a discussion of alternatives using transformed residuals $\mathbf{A}_i^T \mathbf{r}_i$ with the $n_i \times (n_i - q)$ matrix \mathbf{A}_i chosen such that $\mathbf{A}_i^T \mathbf{A}_i = \mathbf{I}_{n_i - q}$ and $\mathbf{A}_i^T \mathbf{Z}_i = \mathbf{0}_{(n_i - q) \times q}$.

Overview Chapter 6 - Flexible extensions of the LLMM

6.1 Smooth models for the mean

6.2 Serial correlation

6.3 Non-normal random effects

Non-normal random effects

To allow for non-normally distributed random effects, approaches exist that assume instead e.g. t-distributions or mixture distributions

$$f(\mathbf{b}_i | \boldsymbol{\pi}, \boldsymbol{\phi}) = \sum_{k=1}^K \pi_k f(\mathbf{b}_i | \boldsymbol{\phi}_k),$$

with

- weights $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$, $\sum_{k=1}^K \pi_k = 1$,
- parametric families of distributions $f(\mathbf{b}_i | \boldsymbol{\phi}_k)$ with parameters $\boldsymbol{\phi}_k$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_K)$, (often assumed as multivariate normal $f(\mathbf{b}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma})$ with mean $\boldsymbol{\mu}_k$ and homogeneous covariances $\boldsymbol{\Sigma}$).

Non-normal random effects

Estimation is usually done using Bayesian approaches or an EM-algorithm, see e.g. Chapter 12 in [Verbeke & Molenberghs \(2000\)](#) for a discussion of the latter approach.

[Verbeke & Molenberghs \(2000\)](#) call the mixture model the **heterogeneity model**, as it assumes heterogeneous subgroups in the population.