



Rainer Schwabe

Individuals Differ: Implications on the Design of Experiments

Individuals Differ: Implications on the Design of Experiments

Rainer Schwabe

Otto-von-Guericke University, Magdeburg

rschwabe@ovgu.de



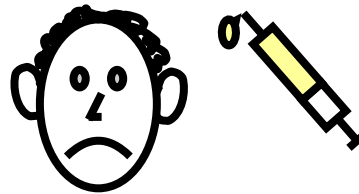
Outline

Prologue: Short introduction

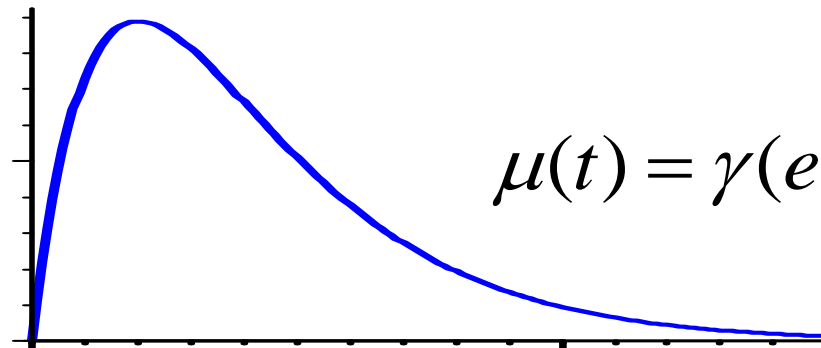
1. Linear models
2. Linear mixed models
3. Non-linear models
4. Non-linear mixed models
5. Miscellaneous

Prologue: Short introduction

➤ Example: Pharmacokinetics

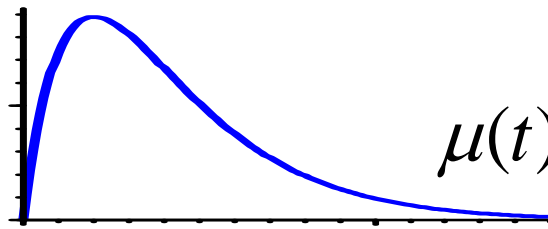


- measure the concentration of a drug in s.o.'s blood over time



$$\mu(t) = \gamma(e^{-\alpha t} - e^{-\beta t})$$

Example: Pharmacokinetics



$$\mu(t) = \gamma(e^{-\alpha t} - e^{-\beta t})$$



➤ aim:

estimate AUC (area under the curve),

c_{\max} (maximal concentration) etc.

$$Y_j(t_j) = \gamma(\exp(-\alpha \cdot t_j) - \exp(-\beta \cdot t_j)) + \varepsilon_j$$

optimal times t_1, \dots, t_n depend on parameters

General model

measurements Y_1, \dots, Y_n at settings t_1, \dots, t_n

$$Y_j(t_j) = \mu(t_j; \boldsymbol{\beta}) + \varepsilon_j$$

resp. mean response

$$E(Y_j(t_j)) = \mu(t_j; \boldsymbol{\beta})$$

➤ response function

μ

➤ parameter

$\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$



General model

measurements Y_1, \dots, Y_n at settings t_1, \dots, t_n

$$Y_j(t_j) = \mu(t_j; \boldsymbol{\beta}) + \varepsilon_j$$

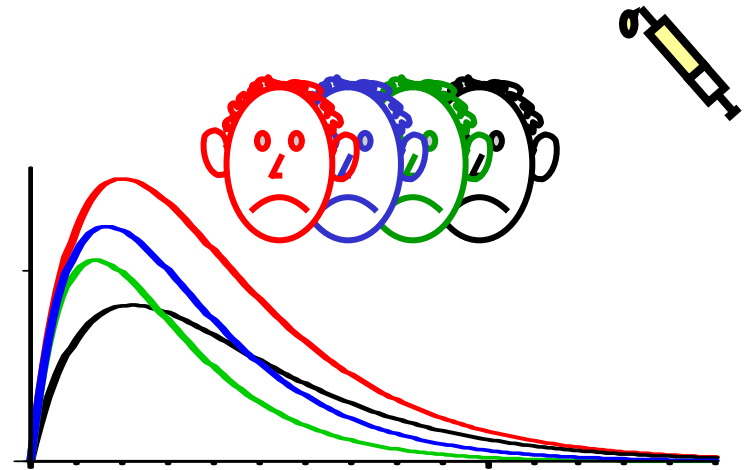
resp. mean response

$$E(Y_j(t_j)) = \mu(t_j; \boldsymbol{\beta})$$

-
- observations uncorrelated
 - + assumptions on $\text{Var}(Y_j)$
or distribution of Y_j

Random coefficients

- each individual has its **own** curve



-
- population parameters
 - individual prediction
 - variance components



1. Linear models

- linear response $\mu(t; \boldsymbol{\beta}) = \mathbf{f}(t)^\top \boldsymbol{\beta}$

$$Y_j(t_j) = \mathbf{f}(t_j)^\top \boldsymbol{\beta} + \varepsilon_j$$

random
error

observation
 $j=1, \dots, n$

explanatory
variable

ε_j i.i.d.
 $\text{Var}(\varepsilon_j) = \sigma^2$

-
- regression functions $\mathbf{f} = (f_1, \dots, f_p)$

- parameter

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$$



Vector notation

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\text{Cov}(\mathbf{Y}) =$$

$$\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$$

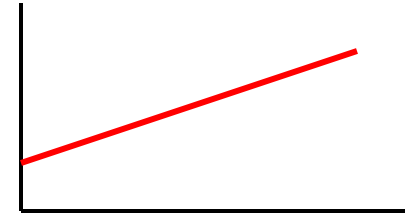
➤ design matrix

$$\mathbf{F} = \begin{pmatrix} \mathbf{f}(t_1)^\top \\ \vdots \\ \mathbf{f}(t_n)^\top \end{pmatrix}$$

$(n \times p)$

➤ Linear regression

$$Y_j(t_j) = \beta_1 + \beta_2 \cdot t_j + \varepsilon_j$$



➤ dimension $p = 2$

➤ regression functions $f_1(t) \equiv 1, f_2(t) = t$

➤ design matrix ($2 \times n$)

$$\mathbf{F} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}$$

Estimation

➤ Gauss Markov Theorem

least squares

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{Y}$$

best linear unbiased estimator for $\boldsymbol{\beta}$

➤ covariance matrix

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{F}^T \mathbf{F})^{-1}$$

➤ “information matrix”

$$\mathbf{M} = \frac{1}{\sigma^2} \mathbf{F}^T \mathbf{F} = \frac{1}{\sigma^2} \sum_{j=1}^n \mathbf{f}(t_j) \mathbf{f}(t_j)^T$$

Estimation

➤ Normal theory

maximum likelihood

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{Y}$$

best unbiased estimator for $\boldsymbol{\beta}$

➤ Fisher information matrix

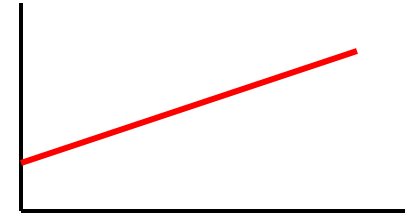
$$\mathbf{M} = \text{Cov}(\text{score fct.}) = \frac{1}{\sigma^2} \sum_{j=1}^n \mathbf{f}(x_j) \mathbf{f}(x_j)^T$$

➤ covariance matrix

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \mathbf{M}^{-1}$$

➤ Linear regression

$$Y_j(t_j) = \beta_1 + \beta_2 \cdot t_j + \varepsilon_j$$



➤ dimension $p = 2$

➤ regression functions $f_1(t) \equiv 1, f_2(t) = t$

➤ information matrix (2×2)

$$\mathbf{F}^T \mathbf{F} = \begin{pmatrix} n & \sum t_j \\ \sum t_j & \sum t_j^2 \end{pmatrix} = \sum \begin{pmatrix} 1 \\ t_j \end{pmatrix} \begin{pmatrix} 1 & t_j \end{pmatrix}$$

Design

➤ “exact design” (t_1, \dots, t_n)

➤ information matrix

$$\mathbf{M}(t_1, \dots, t_n) = \frac{1}{\sigma^2} \sum_{j=1}^n \mathbf{f}(t_j) \mathbf{f}(t_j)^\top$$

➤ aim: choose t_1, \dots, t_n from design region T

to minimise \mathbf{M}^{-1} = Cov($\hat{\boldsymbol{\beta}}$)

resp. $v(t) = \mathbf{f}(t)^\top \mathbf{M}^{-1} \mathbf{f}(t)$ prediction variance

Design criteria

➤ minimise

$$D: \quad \det \mathbf{M}^{-1}$$

$$A: \quad \text{trace } \mathbf{M}^{-1}$$

$$\mathbf{c}: \quad \mathbf{c}^{\top} \mathbf{M}^{-1} \mathbf{c}$$

$$\text{IMSE}: \quad \int_T \mathbf{f}(t)^{\top} \mathbf{M}^{-1} \mathbf{f}(t) dt$$

$$G: \quad \max_{t \in T} \mathbf{f}(t)^{\top} \mathbf{M}^{-1} \mathbf{f}(t)$$

Design criteria

➤ interpretation

D : volume of confidence ellipsoid

A : average variance

c : variance of $\mathbf{c}^T \boldsymbol{\beta}$

IMSE: average prediction variance

G : maximal prediction variance

➤ Linear regression

$$Y_j(t_j) = \beta_1 + \beta_2 \cdot t_j + \varepsilon_j \quad 0 \leq t \leq 1$$

$$\mathbf{M} = \frac{1}{\sigma^2} \begin{pmatrix} n & \sum t_j \\ \sum t_j & \sum t_j^2 \end{pmatrix}$$

➤ observations at $t = 0$ and $t = 1$

D-, *G*- and *IMSE*-optimal

50 % at $t = 1$

A-optimal

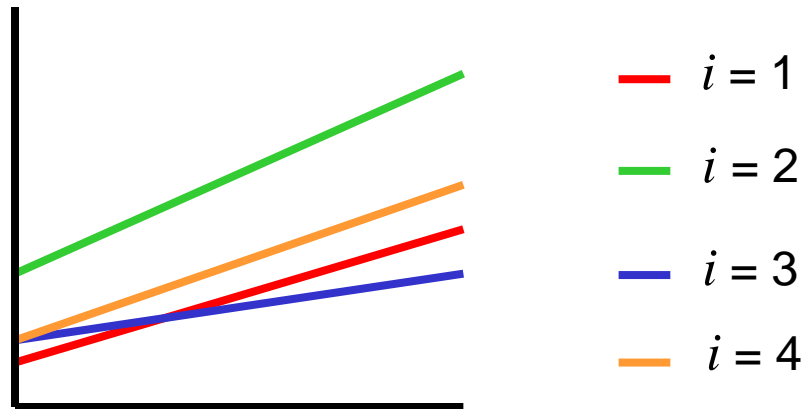
≈ 40 % at $t = 1$

2. Linear mixed models

- each **individual** has its **own** response curve



octodon degus



- **individual** responses follow a common model

Hierarchical model

➤ individual level

$$Y_{ij} = \mathbf{f}(t_{ij})^\top \mathbf{b}_i + \varepsilon_{ij}$$

individual "parameter"

error

individual
 $i=1, \dots, n$

replication
 $j=1, \dots, m_i$

explanatory
variable

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

➤ population level

$$\mathbf{b}_i \sim N_p(\boldsymbol{\beta}, \mathbf{D})$$

independent

variance
components

population parameter



Individual Observational Vector

$$\mathbf{Y}_i = \mathbf{F}_i \boldsymbol{\beta} + \mathbf{F}_i (\mathbf{b}_i - \boldsymbol{\beta}) + \boldsymbol{\varepsilon}_i$$

individual design matrix

➤ individual covariance matrix

$$\mathbf{V}_i = \mathbf{F}_i \mathbf{D} \mathbf{F}_i^\top + \sigma^2 \mathbf{I}_{m_i}$$

observations are **correlated**

➤ e.g. random intercept

$$\mathbf{V}_i = \sigma_\mu^2 \mathbf{1}_{m_i} \mathbf{1}_{m_i}^\top + \sigma^2 \mathbf{I}_{m_i}$$

equal correlations

Estimation of population parameters

- least squares (ML for normal errors)

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^n \mathbf{F}_i^T \mathbf{V}_i^{-1} \mathbf{F}_i \right)^{-1} \sum_{i=1}^n \mathbf{F}_i^T \mathbf{V}_i^{-1} \mathbf{Y}_i$$

- covariance matrix

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^n \mathbf{F}_i^T \mathbf{V}_i^{-1} \mathbf{F}_i \right)^{-1}$$

- note

$$(\mathbf{F}_i^T \mathbf{V}_i \mathbf{F}_i)^{-1} = (\mathbf{F}_i^T \mathbf{F}_i)^{-1} + \mathbf{D}$$

additive
decomposition

Single group designs

- all individuals at the **same** time points

$$m_i \equiv m$$

$$t_{ij} \equiv t_j$$

$$\mathbf{F}_i \equiv \mathbf{F}, \quad \mathbf{V}_i \equiv \mathbf{V}$$

-
- estimation

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^n \mathbf{F}^T \mathbf{F} \right)^{-1} \sum_{i=1}^n \mathbf{F}^T \mathbf{Y}_i$$

does **not** require the knowledge of **D** (WLS=OLS)

Covariance in single group designs

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \frac{1}{n} \left(\sigma^2 (\mathbf{F}_1^\top \mathbf{F}_1)^{-1} + \mathbf{D} \right)$$

-
- covariance decomposes additively into
 - » covariance in the model without random effects ($\mathbf{D} = \mathbf{0}$)
 - » dispersion \mathbf{D} of the parameter
-
- single group designs are optimal within the class of weighted approximate designs (equivalence theorems)

Linear criteria

- minimise

A, IMSE, c

$$\text{trace}\left(\mathbf{L}\left(\sigma^2(\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{D}\right)\mathbf{L}^\top\right)$$

$$= \sigma^2 \text{trace}\left(\mathbf{L}(\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{L}^\top\right) + \overbrace{\text{trace}(\mathbf{LDL}^\top)}^{\text{constant !}}$$

- result:

optimal designs in the model

without random effects

remain optimal with random effects



D - and G -criteria

➤ minimise

D -criterion

$$\det\left(\sigma^2(\mathbf{F}^\top \mathbf{F})^{-1} + \mathbf{D}\right) > \det\left(\sigma^2(\mathbf{F}^\top \mathbf{F})^{-1}\right) + \det(\mathbf{D})$$

➤ resp.

G -criterion

$$\sup_t \left(\mathbf{f}(t)^\top \left(\sigma^2(\mathbf{F}^\top \mathbf{F})^{-1} + \mathbf{D} \right) \mathbf{f}(t) \right)$$

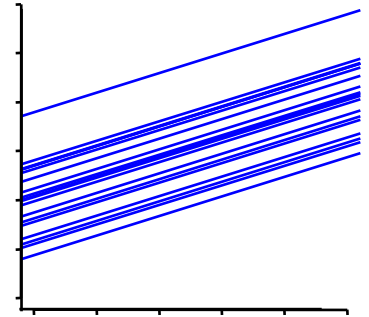
$$< \sigma^2 \sup_t \left(\mathbf{f}(t)^\top (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{f}(t) \right) + \sup_t \left(\mathbf{f}(t)^\top \mathbf{D} \mathbf{f}(t) \right)$$

➤ optimal designs may differ ! \implies examples

Random intercept (random blocks)

- parallel individual curves

$$Y_{ij} = a_i + \mathbf{f}(t_{ij})^\top \boldsymbol{\beta} + \varepsilon_{ij}$$



-
- G - and D -optimal designs

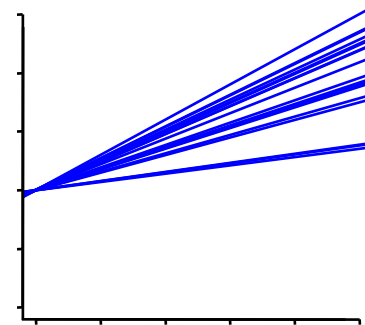
do not depend on σ_a^2



Random treatment effect

- common baseline
(quite unrealistic)

$$Y_{ij} = \mu + b_i t_{ij} + \varepsilon_{ij}$$

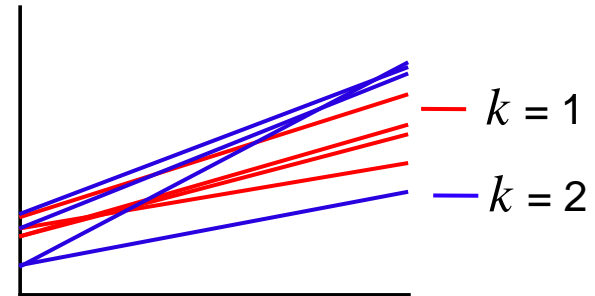


- *D*-and *G*-optimal design differ

	observations at baseline	
	σ_b^2 small	σ_b^2 large
<i>D</i> -criterion	$m/2$	$m-1$
<i>G</i> -criterion	$m/2$	1
<i>IMSE</i> -criterion	$m/2$	$m/2$

Several treatments

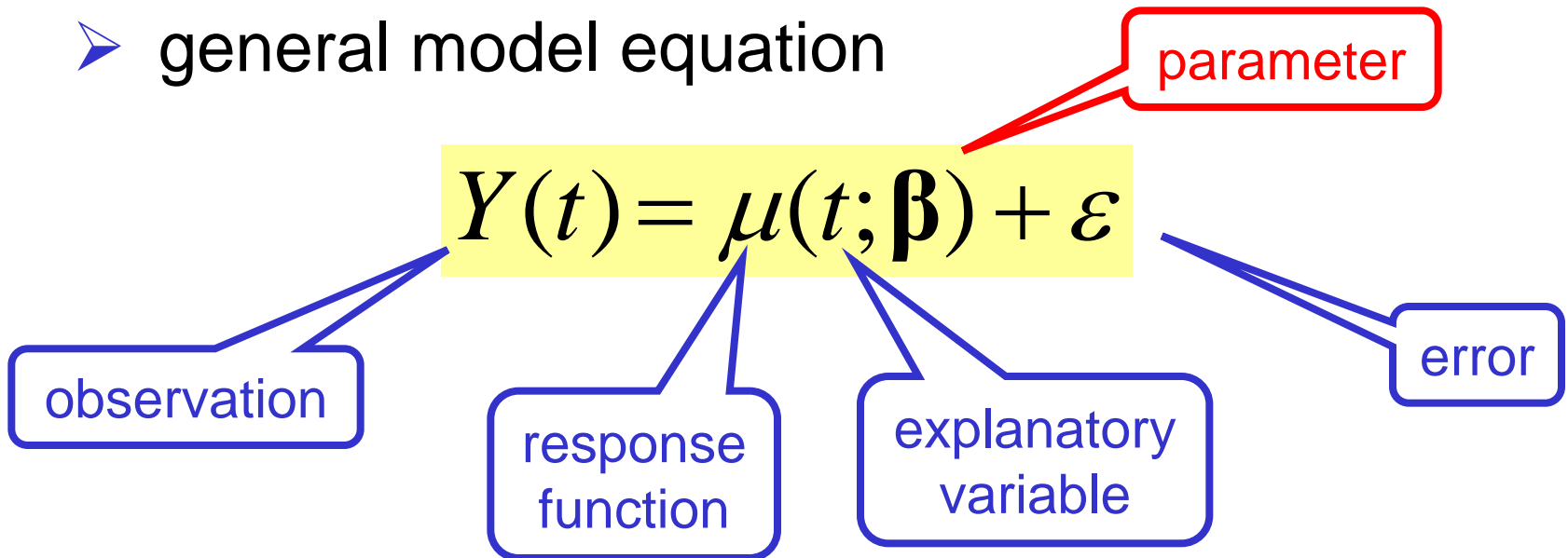
- individuals nested in groups



-
- random intercept (baseline)
 - » information may percolate through from one group to the other:
more observations at baseline
 - » optimal designs depends on σ_a^2
 - » standard designs are quite efficient

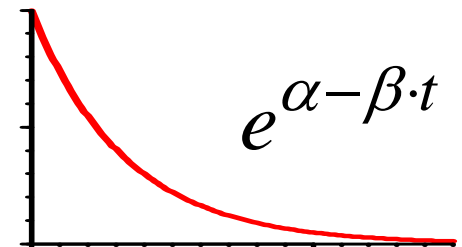
3. Non-linear models

- general model equation



- resp. mean response

$$E(Y(t)) = \mu(t; \beta)$$



Linearised model

$$Y_{\beta_0}(t) = \mathbf{f}_{\beta_0}(t)^\top \boldsymbol{\beta} + \varepsilon$$

➤ linearised regression function

$$\mathbf{f}_{\beta_0}(t) = \frac{\partial \mu(t; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}$$

at $\boldsymbol{\beta}_0 = \boldsymbol{\beta}$

Homoscedastic errors

$$Y_{\boldsymbol{\beta}}(t) = \mathbf{f}_{\boldsymbol{\beta}}(t)^{\top} \boldsymbol{\beta} + \varepsilon$$

- equal variances

$$\text{Var}(Y_{\boldsymbol{\beta}}(t)) = \sigma^2$$

- information matrix

$$\mathbf{M}_{\boldsymbol{\beta}}(t_1, \dots, t_n) = \frac{1}{\sigma^2} \sum \mathbf{f}_{\boldsymbol{\beta}}(t_j) \mathbf{f}_{\boldsymbol{\beta}}(t_j)^{\top}$$

depends on $\boldsymbol{\beta}$

Local optimality

➤ information matrix depends on β

⇒ optimal design may depend on β

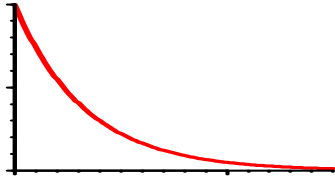
➤ local approach

design **locally** D -optimal at β

⇔ design minimises $\det \mathbf{M}_\beta^{-1}$

➤ alternatives  weighted, minimax criteria

Example: Exponential decay



$$Y(t) = e^{\alpha - \beta \cdot t} + \varepsilon$$

$$t \geq 0$$

$$\boldsymbol{\beta} = (\alpha, \beta)^\top$$

- linearisation

$$Y(t) \approx e^{\alpha_0 - \beta_0 \cdot t} \cdot (\alpha - \beta t) + \varepsilon$$

$$\mathbf{f}_\beta(t) = e^{\alpha - \beta \cdot t} (1, -t)^\top$$

- information matrix

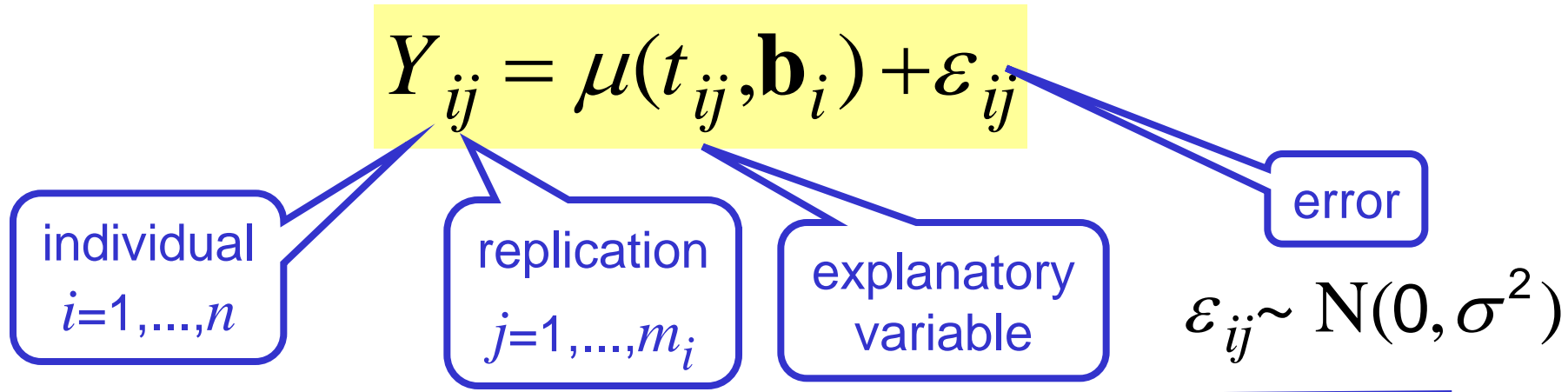
$$\mathbf{M}_\beta = \frac{e^{2\alpha}}{\sigma^2} \sum e^{-2\beta \cdot t_j} \begin{pmatrix} 1 & -t_j \\ -t_j & t_j^2 \end{pmatrix}$$

-
- D -optimal design

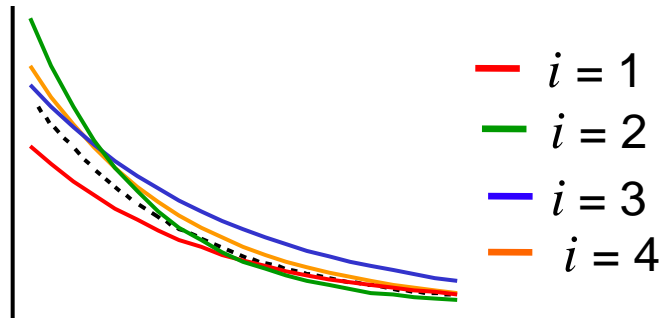
$$t_1^* = 0, \quad t_2^* = 1/\beta$$

4. Non-linear mixed models

- individual response specified by common model



- individual curves



4. Non-linear mixed models

- hierarchical modelling

individual "parameter"

$$Y_{ij} = \mu(t_{ij}, \mathbf{b}_i) + \varepsilon_{ij}$$

individual
 $i=1, \dots, n$

replication
 $j=1, \dots, m_i$

explanatory
variable

error

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

- individual parameters:

$$\mathbf{b}_i \sim N_p(\boldsymbol{\beta}, \mathbf{D})$$

independent

variance
components

population parameter



Linearised model

$$\mathbf{Y}_{i,\beta_0} = \mathbf{F}_{i,\beta_0} \boldsymbol{\beta} + \mathbf{F}_{i,\beta_0} (\mathbf{b}_i - \boldsymbol{\beta}) + \boldsymbol{\varepsilon}_i$$

- linearised design matrix

$$\mathbf{F}_i = \mathbf{F}_{i,\beta} = \left(\frac{\partial \mu(t_{ij}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top} \right)_{j=1, \dots, m_i}$$

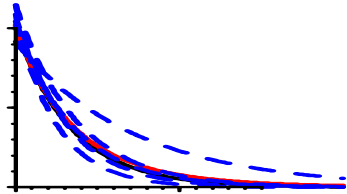
at $\beta_0 = \beta$

- linearised covariance matrix $\mathbf{V}_i = \mathbf{F}_i \mathbf{D} \mathbf{F}_i^\top + \sigma^2 \mathbf{I}_{m_i}$

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{ML}} - \boldsymbol{\beta}) \stackrel{?}{\rightarrow} N_p \left(\mathbf{0}, \lim \left(\frac{1}{n} \sum_{i=1}^n \mathbf{F}_i^\top \mathbf{V}_i^{-1} \mathbf{F}_i \right)^{-1} \right)$$

\mathbf{M}_β

Example: Exponential decay



$$Y_{ij}(t_j) = e^{a_i - b_i \cdot t_j} + \varepsilon_{ij} \quad t \geq 0$$

- linearisation

$$Y_{ij}(t) \approx e^{\alpha_0 - \beta_0 \cdot t} \cdot (\alpha - \beta t) + e^{\alpha_0 - \beta_0 \cdot t} \cdot ((a_i - \alpha) - (b_i - \beta)t) + \varepsilon$$

- information matrix for estimation of β

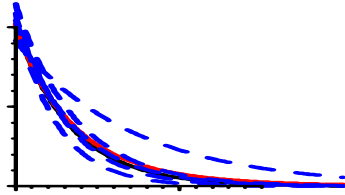
$$\mathbf{M}_\beta = \left(\left(\frac{e^{2\alpha}}{\sigma^2} \sum e^{-2\beta \cdot t_j} \begin{pmatrix} 1 & -t_j \\ -t_j & t_j^2 \end{pmatrix} \right)^{-1} + \mathbf{D} \right)^{-1}$$

- D -optimal design

$$t_1^* = 0, \quad t_2^* > 1/\beta$$

depends
on \mathbf{D}

Example: Exponential decay, prediction



$$Y_{ij}(t_j) = e^{a_i - b_i \cdot t_j} + \varepsilon_{ij} \quad t \geq 0$$

- linearisation

$$Y_{ij}(t) \approx e^{\alpha_0 - \beta_0 \cdot t} \cdot (\alpha - \beta t) + e^{\alpha_0 - \beta_0 \cdot t} \cdot ((a_i - \alpha) - (b_i - \beta)t) + \varepsilon$$

- information for individual prediction of \mathbf{b}_i

$$\ln \det \mathbf{M}_\beta \propto \ln \det \left(\frac{e^{2\alpha}}{\sigma^2} \sum e^{-2\beta \cdot t_j} \begin{pmatrix} 1 & -t_j \\ -t_j & t_j^2 \end{pmatrix} + \mathbf{D}^{-1} \right) + \dots$$

- D -optimal design

$$t_1^* = 0, \quad t_2^* < 1/\beta$$

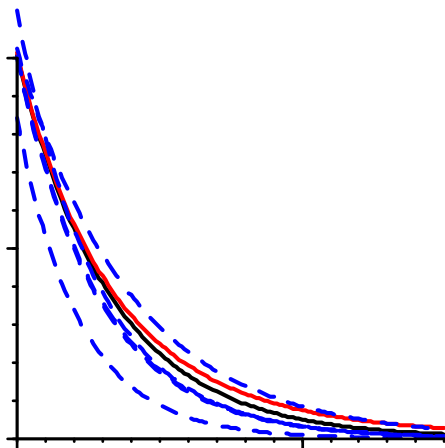
depends
on \mathbf{D}

Example: Exponential decay

- parameterisation matters

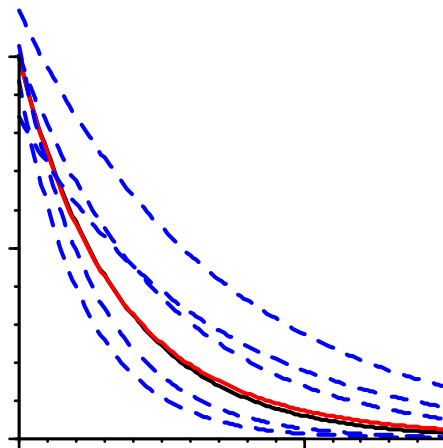
normal

$$Y_{ij}(t_j) = e^{a_i - b_i t_j} + \varepsilon_{ij}$$



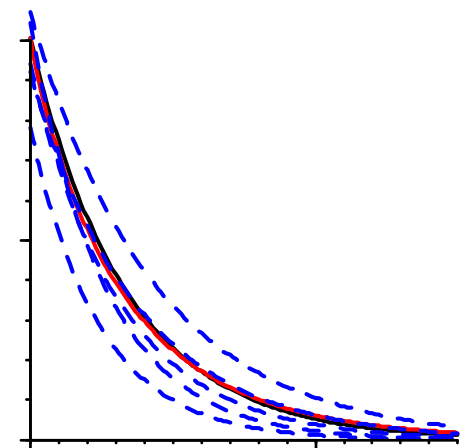
lognormal

$$Y_{ij}(t_j) = e^{a_i - e^{b_i} t_j} + \varepsilon_{ij}$$



inverse normal

$$Y_{ij}(t_j) = a_i e^{-t_j/b_i} + \varepsilon_{ij}$$



individual curves: mean curve \neq typical curve

- impact on optimal designs ?

5. Miscellaneous

- locally optimal designs
 - » depend on parameter β (unknown!) and/or \mathbf{D}
 - » may badly perform,
if parameter β (or \mathbf{D}) is misspecified
- possible solutions
 - » averaged criteria (“Bayes”, pseudo-Bayes)
according to a weight (“prior”) on β (or \mathbf{D})
 - » minimax criteria (maximin efficiency)
guarding against least favourable values of β (or \mathbf{D})

Weighted (averaged) criteria

- criterion is weighted by a distribution π on β
(or \mathbf{D})  (“prior”)

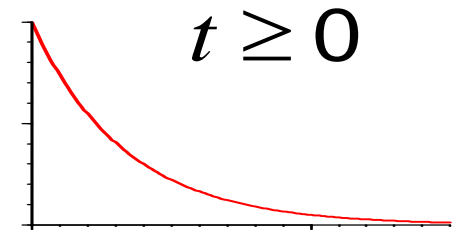
e.g. weighted D -criterion: $-\int \ln(\det(\mathbf{M}_\beta)) \pi(d\beta)$

- optimal design depends on the “prior”

- analytical evaluation often not possible
 - » numerical integration
 - » need for good algorithms

Example: Exponential decay

$$Y(t) = e^{\alpha - \beta \cdot t} + \varepsilon$$



- homoscedastic:

$$\text{Var}(Y(t)) = \sigma^2$$

-
- D -optimal design w.r.t. π

$$t_1^* = 0, \quad t_2^* = 1 / \int \beta \pi(d\beta)$$

mean of β

if the dispersion of π is small

Minimax criteria

- minimise the criterion for the worst value of β (or \mathbf{D}) on a parameter region B (region of interest)

e.g. maximin D -efficiency

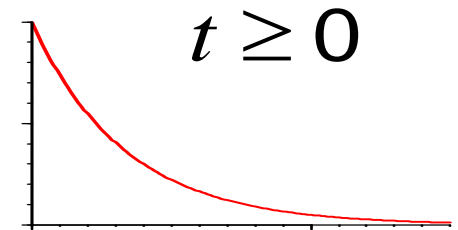
$$- \min_{\beta \in B} \left(\frac{\det(\mathbf{M}_{\beta})}{\det(\mathbf{M}_{\beta}^*)} \right)^{1/p}$$

- » standardisation of the criterion with respect to the locally optimal design

-
- optimal design depends on the region

Example: Exponential decay

$$Y(t) = e^{\alpha - \beta \cdot t} + \varepsilon$$



- homoscedastic:

$$\text{Var}(Y(t)) = \sigma^2$$

-
- maximin D -efficient design on $B = [\beta_1, \beta_2]$

$$t_1^* = 0, \quad t_2^* = 1 / \beta_2$$

worst case

Further problems

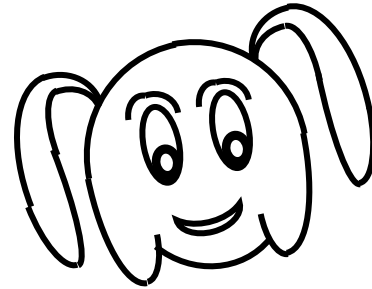
- model specification
 - » parameterisation
 - » distribution of random coefficients
 - » (conditional) distribution of observations
 - » correlation structure **within** individuals
 - » sparse sampling

- **reliable** asymptotic theory
 - » alternative estimation methods (quasi likelihood ...)

- **reliable** software for data analysis



OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG



Thank you !

Some kind of conclusion

**At the beginning we were confused
about the problem**

-

Now we are confused on a higher level

Disclaimer

By purpose, there has not been given any reference to the literature, because, due to the vast amount, mentioning some would have meant omitting others, which were equally or even more important.

Some (incomplete) list of literature may be available from the author.