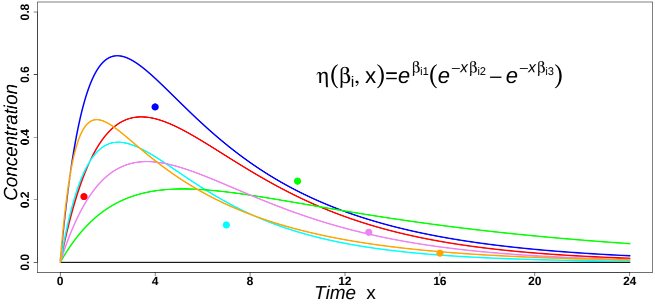
Impact of Information Approximations on Experimental Designs in Nonlinear Mixed Effects Models

Tobias Mielke

PODE 2012
22.03.2012

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

Mixed Effects Models



- Similar functions for different individuals
- Every individual has its own individual parameters
- Vectors of individual parameters are realizations of random vectors

$\rightarrow$ Mixed Effects Models

Mixed Effects Models

Two-stage-model: N individuals with each m observations

- 1. stage (intra-individual variation):

$$Y_{ij} = \eta(\beta_i, x_{ij}) + \epsilon_{ij},\ j = 1, ..., m,\ \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

  - Experimental settings $x_{ij} \in \mathcal{X}$.
  - Known response function η.

- 2. stage (inter-individual variation):

$$\beta_i \sim \mathcal{N}_p(\beta, \sigma^2 D),\ i = 1, ..., N,\ \beta \in \mathbb{R}^p$$

  - β_i and ϵ_{ij} are assumed to be independent.
  - Variance parameters σ^2 and D assumed to be known.

Mixed Effects Models

Individual observation vector:

$$Y_i = \eta(\beta_i, \xi_i) + \epsilon_i,$$

- m - number of observations,
- $\xi_i = (x_{i1}, ..., x_{im}) \in \mathcal{X}^m$ - experimental settings,
- $\eta(\beta_i, \xi_i) := (\eta(\beta_i, x_{i1}), ..., \eta(\beta_i, x_{im}))^T$.

Design matrix:

$$F_{\beta,i} := \left(\frac{\partial \eta(\beta_i, \xi_i)}{\partial \beta_i^T} |_{\beta_i = \beta} \right).$$

Two possible cases:

1.) $\eta(\beta_i, \xi_i)$ linear in β_i or

2.) $\eta(\beta_i, \xi_i)$ nonlinear in β_i.

Estimation

For η nonlinear in β_i:

- No closed form of the probability density $f_{Y_i}(y_i)$:

$$f_{Y_i}(y_i) \quad := \quad \int_{\mathbb{R}^p} f_{Y_i|\beta_i=\mathrm{b}_i}(y_i) f_{\beta_i}(\mathrm{b}_i) d\mathrm{b}_i.$$

- Covariance of estimators in the literature:

$$Cov(\widehat{\beta}) \approx \sigma^2 \left(\sum_{i=1}^{N} F_{\widehat{\beta}_i,i}^T V_{\widehat{\beta}_i,i}^{-1} F_{\widehat{\beta}_i,i} \right)^{-1} \quad \text{for big } m.$$

- Fisher information of interest for small m:

$$\mathfrak{M}(\xi_i) = E\left(\frac{\partial \log(f_{Y_i}(y_i))}{\partial \beta} \frac{\partial \log(f_{Y_i}(y_i))}{\partial \beta^T} \right).$$

Information and Design

- Population designs:

$$\zeta = \begin{pmatrix} \xi_1 & \dots & \xi_k \\ \omega_1 & \dots & \omega_k \end{pmatrix} \text{ with } \sum_{i=1}^{k} \omega_i = 1,$$

  $100 \times \omega_i\%$ observed with settings $\xi_i \in \mathcal{X}^m$.

- Population information:

$$\mathfrak{M}_{pop}(\zeta) := \sum_{i=1}^{k} \omega_i \mathfrak{M}(\xi_i).$$

- For N individuals and population designs ζ holds:

$$\sqrt{N}(\widehat{\beta}_{ML} - \beta) \xrightarrow{\mathfrak{L}} \mathcal{N}_p\left(0, \mathfrak{M}_{pop}(\zeta)^{-1}\right), \ (N \to \infty).$$

- Aim: Minimization of the asymptotic covariance

  $\rightarrow$ D-optimality: $\Phi_D(\zeta) := -\log\det \mathfrak{M}_{pop}(\zeta)$.

Fisher Information

Remember:

$$Y_i = \eta(\beta_i, \xi_i) + \epsilon_i,\ \beta_i \sim \mathcal{N}_p(\beta, \sigma^2 D),\ \epsilon_i \sim \mathcal{N}_m(0, \sigma^2 I_m)$$

log-Likelihood function:

$$l(\beta; y_i) := \log\left(\int_{\mathbb{R}^p} f_{Y_i|\beta_i=\mathrm{b}_i}(y_i) f_{\beta_i}(\mathrm{b}_i)\, d\mathrm{b}_i\right), \text{ and}$$

$$\frac{\partial l(\beta; y_i)}{\partial \beta} = \frac{1}{\sigma^2} D^{-1}(E(\beta_i | Y_i = y_i) - \beta),$$

such that for $\mathfrak{M}(\xi_i)$ follows:

$$\begin{aligned} \mathfrak{M}(\xi_i) &= E\left(\frac{\partial l(\beta; Y_i)}{\partial \beta} \frac{\partial l(\beta; Y_i)}{\partial \beta^T}\right) \\ &= \frac{1}{\sigma^4} D^{-1} Cov(E(\beta_i | Y_i)) D^{-1} \\ &= \frac{1}{\sigma^2} D^{-1} - \frac{1}{\sigma^4} D^{-1} E(Cov(\beta_i | Y_i)) D^{-1} \end{aligned}$$

Fisher Information

Otherwise:

$$Y_i = \eta(\beta + \mathbf{b}_i, \xi_i) + \epsilon_i,\ \mathbf{b}_i \sim \mathcal{N}_p(0, \sigma^2 D),\ \epsilon_i \sim \mathcal{N}_m(0, \sigma^2 I_m)$$

log-Likelihood function:

$$l(\beta; y_i) := \log\left(\int_{\mathbb{R}^p} f_{Y_i|\mathbf{b}_i=b_i}(y_i) f_{\mathbf{b}_i}(b_i) db_i\right), \text{ and}$$

$$\frac{\partial l(\beta; y_i)}{\partial \beta} = \frac{1}{\sigma^2} E\left(F^T_{\beta+\mathbf{b}_i}[y_i - \eta(\beta + \mathbf{b}_i)] | Y_i = y_i\right),$$

such that for $\mathfrak{M}(\xi_i)$ follows:

$$\begin{aligned}\mathfrak{M}(\xi_i) &= E\left(\frac{\partial l(\beta; Y_i)}{\partial \beta}\frac{\partial l(\beta; Y_i)}{\partial \beta^T}\right) \\ &= \frac{1}{\sigma^4} Cov(E(F^T_{\beta+\mathbf{b}_i}[y_i - \eta(\beta + \mathbf{b}_i)] | Y_i)) \\ &= \frac{1}{\sigma^2} E(F^T_{\beta+\mathbf{b}_i} F_{\beta+\mathbf{b}_i}) - \ldots\end{aligned}$$

Fisher Information

Summarizing:

$$Y_i \quad = \quad \eta(\beta_i, \xi_i) + \epsilon_i, \ \beta_i \sim \mathcal{N}_p(\beta, \sigma^2 D), \ \epsilon_i \sim \mathcal{N}_m(0, \sigma^2 I_m)$$

$$\Rightarrow \quad \mathfrak{M}(\xi_i) \leq \min\{\frac{1}{\sigma^2} D^{-1}, \frac{1}{\sigma^2} E(F_{\beta_i}^T F_{\beta_i})\}$$

with respect to the Loewner partial ordering.

Specially for $Cov(\beta_i) = \tau\sigma^2 D$:

$$\mathfrak{M}(\xi_i) \to \quad \frac{1}{\sigma^2} F_\beta^T F_\beta \quad \text{for } \tau \to 0,$$

$$\mathfrak{M}(\xi_i) \to \quad 0 \quad \text{for } \tau \to \infty,$$

$$\mathfrak{M}(\xi_i) \to \quad 0 \quad \text{for } \sigma^2 \to \infty.$$

Information Approximations

Competing approaches in the literature:

- Linearization of the model in some $\beta_0 \in \mathbb{R}^p$:

$$Y_i \approx \eta(\beta_0, \xi_i) + F_{\beta_0}(\beta - \beta_0) + F_{\beta_0}(\beta_i - \beta) + \epsilon_i$$

$$Y_i \overset{app.}{\sim} \mathcal{N}_m\left(\eta(\beta_0, \xi_i) + F_{\beta_0}(\beta - \beta_0), \sigma^2 V_{\beta_0}\right)$$

$\rightarrow$ Linear mixed effects model.

- Linearization of the model in $\beta \in \mathbb{R}^p$:

$$Y_i \approx \eta(\beta, \xi_i) + F_{\beta}(\beta_i - \beta) + \epsilon_i$$

$$Y_i \overset{app.}{\sim} \mathcal{N}_m\left(\eta(\beta, \xi_i), \sigma^2 V_{\beta}\right)$$

$\rightarrow$ Heteroscedastic nonlinear normal model.

Information Approximations

Calculation of the Fisher information:

Assumption: linearized model is the true model.

- Linear Mixed Effects information with $\beta_0 = \beta$

  [Retout et al.(2001), Schmelter(2007)]:

$$\mathbf{M}_1(\xi_i;\beta) := \frac{1}{\sigma^2} F_\beta^T V_\beta^{-1} F_\beta$$

- Nonlinear Heteroscedastic information

  [Retout et al.(2003)]:

$$\mathbf{M}_2(\xi_i;\beta) := \frac{1}{\sigma^2} F_\beta^T V_\beta^{-1} F_\beta + \frac{1}{2} S_\beta,$$

where $S_\beta \geq 0$.

Information Approximations

Remember:

$$Y_i = \eta(\beta_i, \xi_i) + \epsilon_i,\ \beta_i \sim \mathcal{N}_p(\beta, \sigma^2 D),\ \epsilon_i \sim \mathcal{N}_m(0, \sigma^2 I_m)$$

log-Likelihood function:

$$l(\beta; y_i) := \log\left(\int_{\mathbb{R}^p} f_{Y_i|\beta_i=\mathrm{b}_i}(y_i) f_{\beta_i}(\mathrm{b}_i)\, d\mathrm{b}_i\right), \text{ and}$$

$$\frac{\partial l(\beta; y_i)}{\partial \beta} = \frac{1}{\sigma^2} D^{-1}(E(\beta_i | Y_i = y_i) - \beta),$$

such that for $\mathfrak{M}(\xi_i)$ follows:

$$\mathfrak{M}(\xi_i) = E\left(\frac{\partial l(\beta; Y_i)}{\partial \beta} \frac{\partial l(\beta; Y_i)}{\partial \beta^T}\right) = \frac{1}{\sigma^2} D^{-1} - \frac{1}{\sigma^4} D^{-1} E(Cov(\beta_i | Y_i)) D^{-1}$$

$\rightarrow$ Approximation of conditional moments!

Information Approximations

- Approximation of conditional density:

$$\begin{aligned} \beta_i|_{Y_i=y_i} &\overset{app.}{\sim} \mathcal{N}_p(\mu(y_i,\hat{\beta}_i,\beta),\sigma^2 M_{\hat{\beta}_i}^{-1}) \text{ and} \\ E(\beta_i|Y_i=y_i) &\approx \mu(y_i,\hat{\beta}_i,\beta), \\ Cov(\beta_i|Y_i=y_i) &\approx \sigma^2\left(F_{\hat{\beta}_i}^T F_{\hat{\beta}_i} + D^{-1}\right)^{-1}. \end{aligned}$$

- Conditional expectation in $\hat{\beta}_i = \beta$:

$$\begin{aligned} \mathfrak{M}(\xi_i) &= \frac{1}{\sigma^4} D^{-1} Cov(E(\beta_i|Y_i)) D^{-1} \\ &\approx \frac{1}{\sigma^4} F_\beta^T V_\beta^{-1} Cov(Y_i) V_\beta^{-1} F_\beta =: \mathbf{M}_3(\xi_i;\beta). \end{aligned}$$

- Conditional variance in $\hat{\beta}_i = \beta_i$:

$$\mathfrak{M}(\xi_i) \approx \frac{1}{\sigma^2} E(F_{\beta_i}^T V_{\beta_i}^{-1} F_{\beta_i}) =: \mathbf{M}_4(\xi_i;\beta).$$

Information Approximations

Alternative motivation of $\mathbf{M}_4(\xi_i;\beta)$:

- Linearization of the model in some estimate $\widehat{\beta}_i \in \mathbb{R}^p$:

$$Y_i \quad \approx \quad \eta(\widehat{\beta}_i,\xi_i) + F_{\widehat{\beta}_i}(\beta - \widehat{\beta}_i) + F_{\widehat{\beta}_i}(\beta_i - \beta) + \epsilon_i$$

$$Y_i \quad \overset{app.}{\sim} \quad \mathcal{N}_m\left(\eta(\widehat{\beta}_i,\xi_i) + F_{\widehat{\beta}_i}(\beta - \widehat{\beta}_i), \sigma^2 V_{\widehat{\beta}_i}\right)$$

$\rightarrow$ Linear mixed effects model

$\rightarrow$ Individual i provides Information:

$$\frac{1}{\sigma^2} F^T_{\widehat{\beta}_i} V_{\widehat{\beta}_i} F_{\widehat{\beta}_i}$$

- Population provides the information:

$$\mathbf{M}_4(\xi_i;\beta) := \frac{1}{\sigma^2} E(F^T_{\widehat{\beta}_i} V^{-1}_{\widehat{\beta}_i} F_{\widehat{\beta}_i})$$

with according distribution of $\widehat{\beta}_i$.

- Example given by *Two-Stage* and *LB*-Estimators.

Information Approximations

Other possible approximation:

- Unbiased estimator for β provided by:

$$\sum_{i=1}^{N} (y_i - E(Y_i))^T \, Cov(Y_i)^{-1} \, (y_i - E(Y_i)) \to \min_{\beta}$$

- Quasi-Information:

$$\mathbf{M}_5(\xi_i;\beta) \quad := \quad \frac{\partial E(Y_i)^T}{\partial \beta} Cov(Y_i)^{-1} \frac{\partial E(Y_i)}{\partial \beta^T}$$

- Known:

$$\mathbf{M}_5(\xi_i;\beta) \leq \mathfrak{M}(\xi_i).$$

Example 1

Observations as

$$Y_i \quad = \quad \exp(\beta_i) + \epsilon_i, \;\; \beta_i \sim \mathcal{N}(\beta, d), \;\; \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

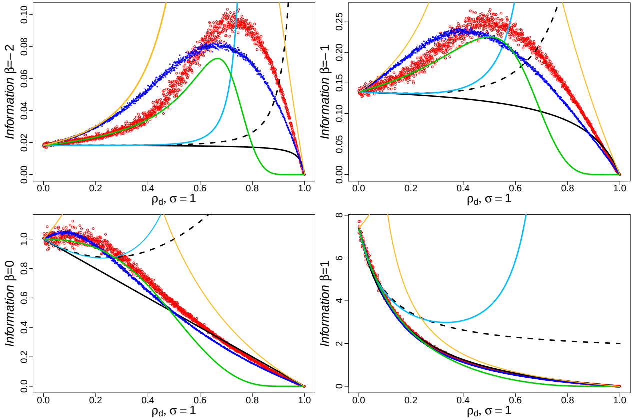
- Closed form approximations for:

  $\mathbf{M}_1$, $\mathbf{M}_2$, $\mathbf{M}_3$ and $\mathbf{M}_5$.

- Approximation of $\mathbf{M}_4$ using Monte-Carlo.
- Comparison with a simulated Fisher information for
  - 1250 values of d and given $\sigma^2 = 1$
  - 1250 values of σ^2 and given $d = 1$
  - 10000 observations per setting $d, \; \sigma^2 \in \mathbb{R}^+$.

Example 1

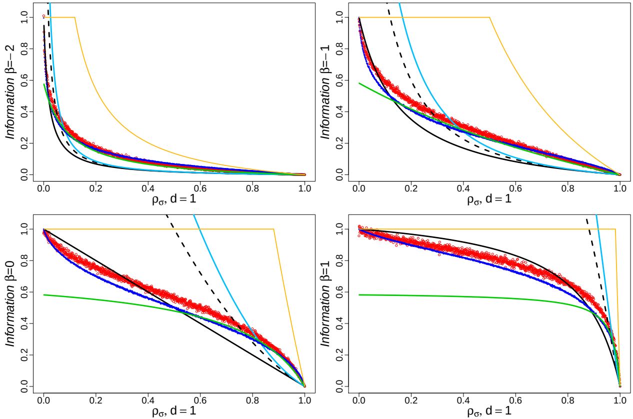
$$Y_i = \exp(\beta_i) + \epsilon_i,\ \beta_i \sim \mathcal{N}(\beta, d = \tfrac{\rho_d}{1-\rho_d}),\ \epsilon_i \sim \mathcal{N}(0,1)$$



- Red: Simulated Fisher information; Dark-blue: $\mathbf{M}_4$
- $\mathbf{M}_1$: solid, $\mathbf{M}_2$: dashed, $\mathbf{M}_3$: Light-blue, $\mathbf{M}_5$: green

Example 1

$$Y_i = \exp(\beta_i) + \epsilon_i,\ \beta_i \sim \mathcal{N}(\beta, 1),\ \epsilon_i \sim \mathcal{N}(0, \sigma^2 = \frac{\rho_\sigma}{1-\rho_\sigma})$$



- Red: Simulated Fisher information; Dark-blue: $\mathbf{M}_4$
- $\mathbf{M}_1$: solid, $\mathbf{M}_2$: dashed, $\mathbf{M}_3$: Light-blue, $\mathbf{M}_5$: green

Conclusions

In the example:

- $\mathbf{M}_1$: doesn't work entirely well
- $\mathbf{M}_2$: limits for $d \to \infty$ or $\sigma^2 \to 0$ don't coincide
- $\mathbf{M}_3$: diverges for $d \to \infty$.
- $\mathbf{M}_4$: no closed form
- $\mathbf{M}_5$: limit for $\sigma^2 \to 0$ doesn't coincide

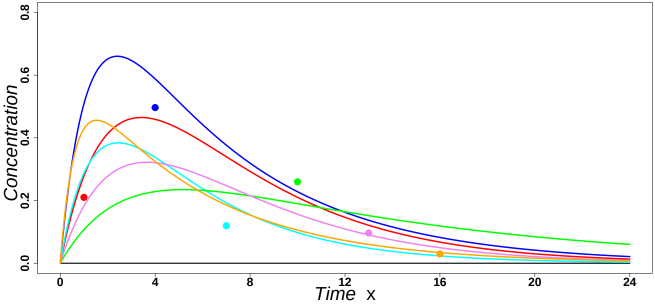
$\to$ Don't trust any approximation!

Generally:

- $m \to \infty$ corresponds to $\sigma^2 \to 0$

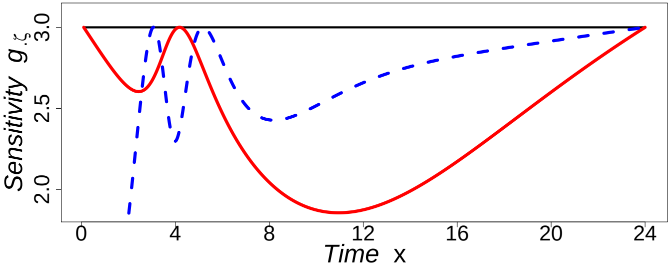
  $\to$ Only $\mathbf{M}_1$ and $\mathbf{M}_4$ reliable

One-Compartment-Model



Here: one observation per individual: $m = 1$

- $Y_i = \eta(\beta_i, x_i) + \epsilon_i$ and
- $\eta(\beta_i, x_i) := \log\left(\exp(\beta_{i1})\left[\exp(-x_i\beta_{i2}) - \exp(-x_i\beta_{i3})\right]\right)$
- Experimental settings: $\xi_i = x_i \in [0.1, 24]$.



A design ζ is *D*-optimal if and only if:

$$g_\zeta(\xi) := \operatorname{tr} \mathfrak{M}_{pop}(\zeta)^{-1}\mathfrak{M}(\xi) \leq p, \ \forall \xi \in \mathcal{X}^m.$$

$\mathbf{M}_1$-Information:

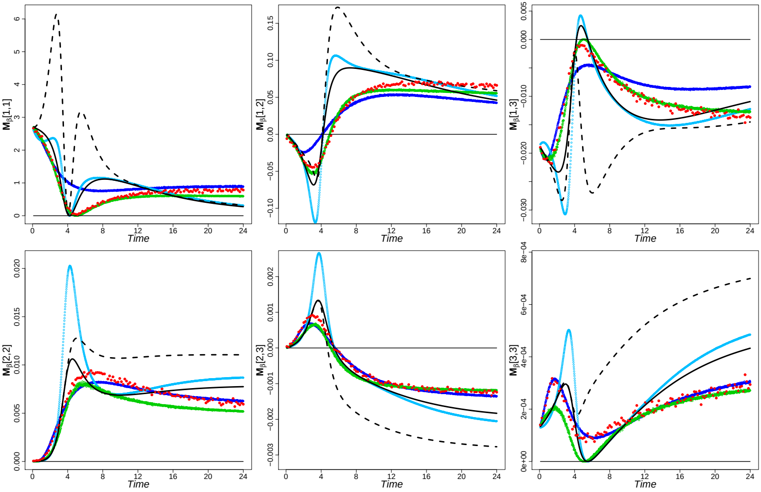
$$\rightarrow \zeta_1 = \begin{pmatrix} 0.10 & 4.18 & 24.00 \\ 0.33 & 0.33 & 0.33 \end{pmatrix}, \ eff_1(\zeta_2) = 0.66$$

$\mathbf{M}_2$-Information:

$$\rightarrow \zeta_2 = \begin{pmatrix} 3.10 & 5.18 & 24.00 \\ 0.61 & 0.08 & 0.31 \end{pmatrix}, \ eff_2(\zeta_1) = 0.55$$

$\rightarrow$ Information matters!

Example 2



- Red: Simulated Fisher information; Dark-blue: $\mathbf{M}_4$
- $\mathbf{M}_1$: solid, $\mathbf{M}_2$: dashed, $\mathbf{M}_3$: Light-blue, $\mathbf{M}_5$: green

One-Compartment-Model

D-optimal Designs for the proposed approximations:

$\mathbf{M}_j$	ζ_j^*	*eff*
$\mathbf{M}_1$	$\begin{pmatrix} 0.10 & 4.18 & 24.00 \\ 0.33 & 0.33 & 0.33 \end{pmatrix}$	0.83
$\mathbf{M}_2$	$\begin{pmatrix} 3.10 & 5.18 & 24.00 \\ 0.61 & 0.09 & 0.30 \end{pmatrix}$	0.88
$\mathbf{M}_3$	$\begin{pmatrix} 0.10 & 3.28 & 4.47 & 24.00 \\ 0.21 & 0.20 & 0.28 & 0.31 \end{pmatrix}$	0.87
$\mathbf{M}_4$	$\begin{pmatrix} 2.85 & 24.00 \\ 0.41 & 0.59 \end{pmatrix}$	0.95
$\mathbf{M}_5$	$\begin{pmatrix} 0.10 & 4.47 & 22.20 \\ 0.32 & 0.35 & 0.33 \end{pmatrix}$	0.82
$\mathfrak{M}$	$\begin{pmatrix} 2.32 & 6.40 & 24.00 \\ 0.61 & 0.01 & 0.38 \end{pmatrix}$	1.00

Conclusions/Outlook

- Conclusions:
  - No overall satisfying approximation
  - Different information $\rightarrow$ different design
  - Compute designs and compare their efficiency!
- Outlook:
  - More insight needed on:
    - approximations for $\mathbf{M}_4$ and $\mathbf{M}_5$
    - appropriateness of sparse sampling designs
  - Information for the variance parameters?
  - Locally optimal designs

Thank you for your attention!