Optimal Design in Nonlinear Mixed Effects Models

Tobias Mielke

Experiments for Processes with Time and Space Dynamics 2011

Outline

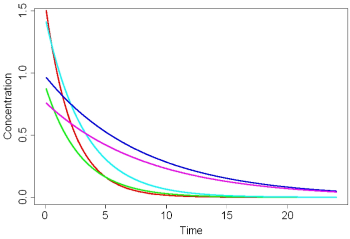
1 Information Approximation

Outline

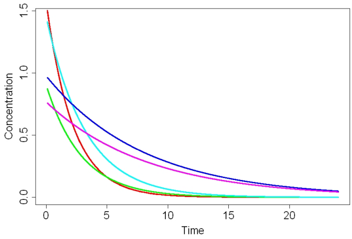
1. Information Approximation

2. Experimental designs

Mixed Effects Models

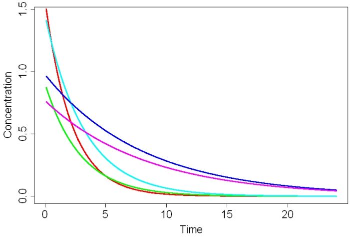


Mixed Effects Models



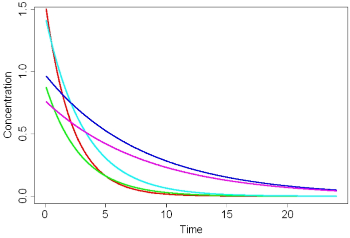
- Similar functions for different individuals

Mixed Effects Models



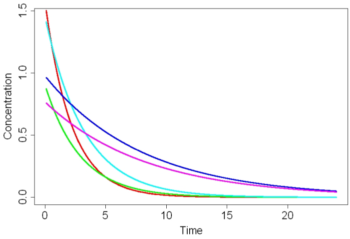
- Similar functions for different individuals
- Every individual has its own individual parameters

Mixed Effects Models



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

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:

- 1. stage (intra-individual variation):

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

Mixed Effects Models

Two-stage-model:

- 1. stage (intra-individual variation):

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

- 2. stage (inter-individual variation):

$$\beta_i = \beta + b_i,\; i = 1, ..., N,\, b_i \sim N_p(0, \sigma^2 D)$$

Mixed Effects Models

Two-stage-model:

- 1. stage (intra-individual variation):

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

- 2. stage (inter-individual variation):

$$\beta_i = \beta + b_i, \ i = 1, ..., N, \ b_i \sim N_p(0, \sigma^2 D)$$

- b_i and ϵ_{ij} are assumed to be independent.

Mixed Effects Models

Two-stage-model:

- 1. stage (intra-individual variation):

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

- 2. stage (inter-individual variation):

$$\beta_i = \beta + b_i,\ i = 1, ..., N,\ b_i \sim N_p(0, \sigma^2 D)$$

- b_i and ϵ_{ij} are assumed to be independent.
- Aim: Estimation of β.

Information Approximation

Individual observation vector:

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

- m_i - number of observations,
- $\xi_i = (x_{i1}, ..., x_{im_i})$ - experimental settings,
- $\eta(\xi_i, \beta_i) := (\eta(x_{i1}, \beta_i), ..., \eta(x_{im_i}, \beta_i))^T$.

Design matrix:

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

Information Approximation

- Optimal designs for estimating $\beta \rightarrow cov(\hat{\beta})$

Information Approximation

- Optimal designs for estimating $\beta \rightarrow cov(\hat{\beta})$
- For Y_i with density $f_{Y_i}(y_i)$ and

$$l(y_i, \beta, \theta) := \log(f_{Y_i}(y_i)) :$$

Maximization of the Fisher information

$$\mathfrak{M}_\beta = E(\frac{\partial l(Y_i, \beta, \theta)}{\partial \beta} \frac{\partial l(Y_i, \beta, \theta)}{\partial \beta^T}).$$

Information Approximation

- Optimal designs for estimating $\beta \to cov(\hat{\beta})$
- For Y_i with density $f_{Y_i}(y_i)$ and

$$l(y_i, \beta, \theta) := \log(f_{Y_i}(y_i)) :$$

Maximization of the Fisher information

$$\mathfrak{M}_\beta = E(\frac{\partial l(Y_i, \beta, \theta)}{\partial \beta} \frac{\partial l(Y_i, \beta, \theta)}{\partial \beta^T}).$$

- Here:

$$f_{Y_i}(y_i) = \int_{\mathbb{R}^p} \phi_{Y_i|\beta_i}(y_i) \phi_{\beta_i}(\beta_i) d\beta_i.$$

Information Approximation

For η nonlinear in β_i:

- Linearization of the model (1):

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

$\rightarrow$ Heteroscedastic nonlinear normal model.

Information Approximation

For η nonlinear in β_i:

- Linearization of the model (1):

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

$\rightarrow$ Heteroscedastic nonlinear normal model.

- Linearization of the model (2):

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

$\rightarrow$ Linear mixed effects model.

Information Approximation

For η nonlinear in β_i:

- Linearization of the model (1):

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

$\rightarrow$ Heteroscedastic nonlinear normal model.

- Linearization of the model (2):

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

$\rightarrow$ Linear mixed effects model.

- Fisher information: Assume the new model is true.

Information Approximation

Alternative:

- With $E_{y_i}(\beta_i) := E(\beta_i | Y_i = y_i)$:

$$\frac{\partial l(y_i, \beta, \theta)}{\partial \beta} = \frac{1}{\sigma^2} D^{-1} (E_{y_i}(\beta_i) - \beta).$$

Information Approximation

Alternative:

- With $E_{y_i}(\beta_i) := E(\beta_i | Y_i = y_i)$:

$$\frac{\partial l(y_i, \beta, \theta)}{\partial \beta} = \frac{1}{\sigma^2} D^{-1} (E_{y_i}(\beta_i) - \beta).$$

- For $Var_{y_i}(\beta_i) = Var(\beta_i | Y_i = y_i)$ follows:

$$\begin{aligned} \mathfrak{M}_\beta &= \frac{1}{\sigma^2} D^{-1} Var(E_{Y_i}(\beta_i)) D^{-1} \frac{1}{\sigma^2} \\ &= \frac{1}{\sigma^2} D^{-1} - \frac{1}{\sigma^2} D^{-1} E(Var_{Y_i}(\beta_i)) D^{-1} \frac{1}{\sigma^2}. \end{aligned}$$

Information Approximation

Alternative:

- With $E_{y_i}(\beta_i) := E(\beta_i | Y_i = y_i)$:

$$\frac{\partial l(y_i, \beta, \theta)}{\partial \beta} = \frac{1}{\sigma^2} D^{-1} (E_{y_i}(\beta_i) - \beta).$$

- For $Var_{y_i}(\beta_i) = Var(\beta_i | Y_i = y_i)$ follows:

$$\begin{aligned} \mathfrak{M}_\beta &= \frac{1}{\sigma^2} D^{-1} Var(E_{Y_i}(\beta_i)) D^{-1} \frac{1}{\sigma^2} \\ &= \frac{1}{\sigma^2} D^{-1} - \frac{1}{\sigma^2} D^{-1} E(Var_{Y_i}(\beta_i)) D^{-1} \frac{1}{\sigma^2}. \end{aligned}$$

$\rightarrow$ Approximation of conditional moments.

Information Approximation

Laplace approximation:

- Idea: Approximation of

$$f_{Y_i}(y_i) \quad = \quad \int c \cdot exp(\frac{-h(\beta, D, y_i, \beta_i)}{2\sigma^2}) d\beta_i$$

by a normal density.

Information Approximation

Laplace approximation:

- Idea: Approximation of

$$f_{Y_i}(y_i) = \int c \cdot exp(\frac{-h(\beta, D, y_i, \beta_i)}{2\sigma^2}) d\beta_i$$

by a normal density.

- Taylor expansion around $\hat{\beta}_i$ with

$$\hat{\beta}_i := \underset{\beta_i \in \mathbb{R}^p}{argmin}\{h(\beta, D, y_i, \beta_i)\}$$

Information Approximation

Laplace approximation:

- Idea: Approximation of

$$f_{Y_i}(y_i) = \int c \cdot exp(\frac{-h(\beta, D, y_i, \beta_i)}{2\sigma^2}) d\beta_i$$

  by a normal density.

- Taylor expansion around $\hat{\beta}_i$ with

$$\hat{\beta}_i := \underset{\beta_i \in \mathbb{R}^p}{argmin}\{h(\beta, D, y_i, \beta_i)\}$$

- Problem: $\hat{\beta}_i$ depends on $y_i \rightarrow$ not normal.

Information Approximation

Laplace approximation:

- Idea: Approximation of

$$f_{Y_i}(y_i) = \int c \cdot exp(\frac{-h(\beta, D, y_i, \beta_i)}{2\sigma^2}) d\beta_i$$

  by a normal density.

- Taylor expansion around $\hat{\beta}_i$ with

$$\hat{\beta}_i := \underset{\beta_i \in \mathbb{R}^p}{argmin} \{h(\beta, D, y_i, \beta_i)\}$$

- Problem: $\hat{\beta}_i$ depends on $y_i \rightarrow$ not normal.
- Alternative: linearization of $\eta(\xi_i, \beta_i)$ around $\hat{\beta}_i$.

Information Approximation

Laplace approximation:

- Note:

$$f_{\beta_i|Y_i=y_i} = \frac{\phi_{Y_i|\beta_i}(y_i)\phi_{\beta_i}(\beta_i)}{f_{Y_i}(y_i)}$$

- Similar approximation to the numerator:

$$\beta_i|_{Y_i=y_i} \overset{app.}{\sim} N(\hat{\beta}_i, \sigma^2 M_{\hat{\beta}_i}^{-1})$$

with $M_{\hat{\beta}_i} = F_{\hat{\beta}_i}^T F_{\hat{\beta}_i} + D^{-1}$.

Information Approximation

Laplace approximation:

- Note:

$$f_{\beta_i|Y_i=y_i} = \frac{\phi_{Y_i|\beta_i}(y_i)\phi_{\beta_i}(\beta_i)}{f_{Y_i}(y_i)}$$

- Similar approximation to the numerator:

$$\beta_i|_{Y_i=y_i} \overset{app.}{\sim} N(\hat{\beta}_i, \sigma^2 M_{\hat{\beta}_i}^{-1})$$

with $M_{\hat{\beta}_i} = F_{\hat{\beta}_i}^T F_{\hat{\beta}_i} + D^{-1}$.

- Try β:

$$\beta_i|_{Y_i=y_i} \overset{app.}{\sim} N(\beta + M_\beta^{-1} F_\beta^T (y_i - \eta(\xi_i, \beta)), \sigma^2 M_\beta^{-1})$$

Information Approximation

Fisher information approximations

- With $V_\beta := \sigma^2(I_{m_i} + F_\beta D F_\beta^T)$:

Information Approximation

Fisher information approximations

- With $V_\beta := \sigma^2(I_{m_i} + F_\beta D F_\beta^T)$:
- Conditional expectation:

$$\mathfrak{M}_{1,\beta} := \frac{1}{\sigma^2} D^{-1} Var(E_{Y_i}(\beta_i)) D^{-1} \frac{1}{\sigma^2}$$

$$= F_\beta^T V_\beta^{-1} Var(Y_i) V_\beta^{-1} F_\beta.$$

Information Approximation

Fisher information approximations

- With $V_\beta := \sigma^2(I_{m_i} + F_\beta D F_\beta^T)$:
- Conditional expectation:

$$\begin{aligned}\mathfrak{M}_{1,\beta} &:= \frac{1}{\sigma^2} D^{-1} Var(E_{Y_i}(\beta_i)) D^{-1} \frac{1}{\sigma^2} \\ &= F_\beta^T V_\beta^{-1} Var(Y_i) V_\beta^{-1} F_\beta.\end{aligned}$$

- Conditional variance:

$$\begin{aligned}\mathfrak{M}_{2,\beta} &:= \frac{1}{\sigma^2} D^{-1} - \frac{1}{\sigma^2} D^{-1} E(Var_{Y_i}(\beta_i)) D^{-1} \frac{1}{\sigma^2} \\ &= F_\beta^T V_\beta^{-1} F_\beta.\end{aligned}$$

Information Approximation

Other possible approximations:

- For known moments $E(Y_i)$ and $Var(Y_i)$:

$$\mathfrak{M}_{3,\beta} := F_{\beta,QL}^T Var(Y_i)^{-1} F_{\beta,QL}$$

Information Approximation

Other possible approximations:

- For known moments $E(Y_i)$ and $Var(Y_i)$:

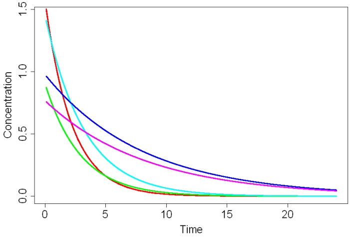
$$\mathfrak{M}_{3,\beta} := F_{\beta,QL}^T Var(Y_i)^{-1} F_{\beta,QL}$$

- Nonlinear heteroscedastic normal model:

$$\mathfrak{M}_{4,\beta} := F_\beta^T V_\beta^{-1} F_\beta + \frac{1}{2} S, \text{ with}$$

$$S_{jk} = Tr\ V_\beta^{-1} \frac{\partial V_\beta}{\partial \beta_j} V_\beta^{-1} \frac{\partial V_\beta}{\partial \beta_k}.$$

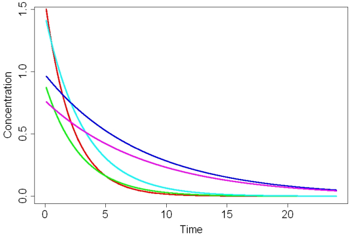
Example



- Log-Concentration:

$$Y_i = -\beta_{2i} - x_i exp(\beta_{1i} - \beta_{2i}) + \epsilon_i, \ i = 1, ..., N$$

Example



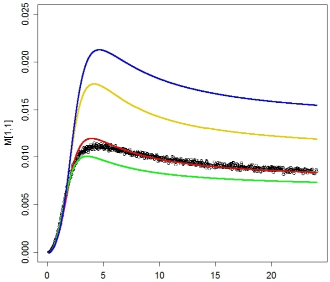
- Log-Concentration:

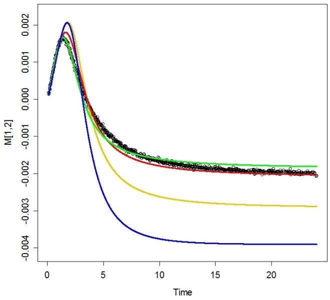
$$Y_i = -\beta_{2i} - x_i exp(\beta_{1i} - \beta_{2i}) + \epsilon_i, \ i = 1, ..., N$$

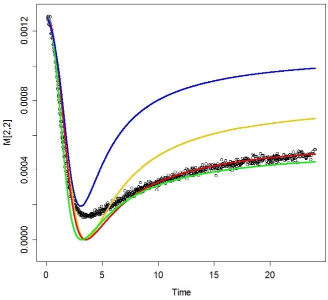
- - $\epsilon_i \sim N(0, \sigma^2)$, $\xi_i = (x_i) \in [t_{min}, t_{max}]$,
  - $\beta_i = (\beta_{1i}, \beta_{2i})^T \sim N(\beta, \sigma^2 D)$,
  - $\beta = (\beta_1, \beta_2)^T$, $D = diag(d_1, d_2)$.

Example

- $\mathfrak{M}_{1,\beta}$: Orange
- $\mathfrak{M}_{2,\beta}$: Red
- $\mathfrak{M}_{3,\beta}$: Green
- $\mathfrak{M}_{4,\beta}$: Blue







Design-Problem

- Experiment settings ξ_i for individuals: $x_{ij} \in \mathcal{X}$ with m_{ij} observations:

$$\xi_i = \begin{pmatrix} x_{i1} & \dots & x_{im_i} \end{pmatrix} \in \mathcal{X}^{m_i}.$$

Design-Problem

- Experiment settings ξ_i for individuals: $x_{ij} \in \mathcal{X}$ with m_{ij} observations:

$$\xi_i = \begin{pmatrix} x_{i1} & \dots & x_{im_i} \end{pmatrix} \in \mathcal{X}^{m_i}.$$

- Population design ζ:

$$\zeta = \begin{pmatrix} \xi_1 & \dots & \xi_k \\ \omega_1 & \dots & \omega_k \end{pmatrix}, \sum_{j=1}^{k} \omega_j = 1,$$

ω_j: weight of indivdual design ξ_j in the population.

- Population design $\hat{=}$ approximate design on $\mathcal{X}^m$

Design-Problem

- Choose design, such that:

$$Cov(\hat{\beta}) \approx \mathfrak{M}_{pop}(\zeta)^{-1}$$

is minimal.

Design-Problem

- Choose design, such that:

$$Cov(\hat{\beta}) \approx \mathfrak{M}_{pop}(\zeta)^{-1}$$

  is minimal. $\rightarrow$ Optimality-criteria

  - c-optimality: $\Phi_c(\zeta)$
  - A-optimality: $\Phi_A(\zeta)$
  - D-optimality: $\Phi_D(\zeta)$
  - ...

Design-Problem

- Choose design, such that:

$$Cov(\hat{\beta}) \approx \mathfrak{M}_{pop}(\zeta)^{-1}$$

  is minimal. $\rightarrow$ Optimality-criteria

  - c-optimality: $\Phi_c(\zeta)$
  - A-optimality: $\Phi_A(\zeta)$
  - D-optimality: $\Phi_D(\zeta)$
  - ...

- How to determine optimal designs?

  $\rightarrow$ Use of equivalence theorems

  $\rightarrow$ Use of algorithms

Design-Problem

Theorem

A design ζ^* is Φ_L-optimal for estimating $\beta \in \mathbb{R}^p$ if and only if the function $g_{\zeta_*}(\xi)$ is for all $\xi \in \mathcal{X}^m$ nonpositive:

$$Tr\left[\mathfrak{M}_{pop}(\zeta^*)^{-1} L \mathfrak{M}_{pop}(\zeta^*)^{-1}(\mathfrak{M}_\beta(\xi) - \mathfrak{M}_{pop}(\zeta^*))\right] \leq 0.$$

Design-Problem

Theorem

A design ζ^* is Φ_L-optimal for estimating $\beta \in \mathbb{R}^p$ if and only if the function $g_{\zeta_*}(\xi)$ is for all $\xi \in \mathcal{X}^m$ nonpositive:

$$Tr\left[\mathfrak{M}_{pop}(\zeta^*)^{-1} L \mathfrak{M}_{pop}(\zeta^*)^{-1}(\mathfrak{M}_\beta(\xi) - \mathfrak{M}_{pop}(\zeta^*))\right] \leq 0.$$

- For D-optimality: $L = \mathfrak{M}_{pop}(\zeta^*)$

Design-Problem

Theorem

A design ζ^* is Φ_L-optimal for estimating $\beta \in \mathbb{R}^p$ if and only if the function $g_{\zeta_*}(\xi)$ is for all $\xi \in \mathcal{X}^m$ nonpositive:

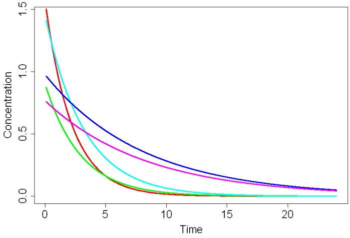
$$Tr\ [\mathfrak{M}_{pop}(\zeta^*)^{-1} L \mathfrak{M}_{pop}(\zeta^*)^{-1} (\mathfrak{M}_\beta(\xi) - \mathfrak{M}_{pop}(\zeta^*))] \leq 0.$$

- For D-optimality: $L = \mathfrak{M}_{pop}(\zeta^*)$

→ In our situation:

- Individual information: $\mathfrak{M}_\beta(\xi)$
- Population information: $\mathfrak{M}_{pop}(\zeta) := \sum_{i=1}^{k} \omega_i \mathfrak{M}_\beta(\xi_i)$.

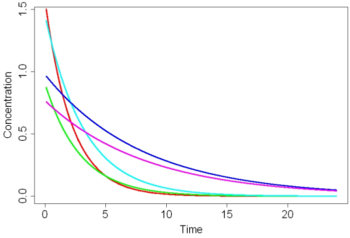
Example



- Log-Concentration:

$$Y_i = -\beta_{2i} - x_i exp(\beta_{1i} - \beta_{2i}) + \epsilon_i, \; i = 1, ..., N$$

Example



- Log-Concentration:

$$Y_i = -\beta_{2i} - x_i exp(\beta_{1i} - \beta_{2i}) + \epsilon_i, \; i = 1, ..., N$$

- - $\epsilon_i \sim N(0, \sigma^2)$, $\xi_i = (x_i) \in [t_{min}, t_{max}]$,
  - $\beta_i = (\beta_{1i}, \beta_{2i})^T \sim N(\beta, \sigma^2 D)$,
  - $\beta = (\beta_1, \beta_2)^T$, $D = diag(d_1, d_2)$.

Optimal Design - Analytically

Simplest situation: One observation per individual: $\xi = (x)$.

- Information matrix $F_\beta(x)^T V_\beta(x)^{-1} F_\beta(x)$:

$$\tilde{g}_\zeta(x) := F_\beta(x) \begin{pmatrix} c & -b \\ -b & a \end{pmatrix} F_\beta(x)^T - 2(ac - b^2) V_\beta(x) \leq 0$$

Optimal Design - Analytically

Simplest situation: One observation per individual: $\xi = (x)$.

- Information matrix $F_\beta(x)^T V_\beta(x)^{-1} F_\beta(x)$:

$$\tilde{g}_\zeta(x) := F_\beta(x) \begin{pmatrix} c & -b \\ -b & a \end{pmatrix} F_\beta(x)^T - 2(ac - b^2) V_\beta(x) \leq 0$$

- Optimal population designs:

$$\zeta^* = \begin{pmatrix} (x_1) & (x_2) \\ 0.5 & 0.5 \end{pmatrix},$$

with $x_1, x_2 \in [t_{min}, t_{max}]$, such that

$$\sigma^2 + cov(Y(x_1), Y(x_2)) = 0.$$

First Order Kinetics

$\mathfrak{M}_{1,\beta}$-Information:

$$\rightarrow \zeta_1 = \begin{pmatrix} (1.70) & (24.00) \\ 0.5 & 0.5 \end{pmatrix},\ \delta_2(\zeta_1) = 0.9794$$

$\mathfrak{M}_{2,\beta}$-Information:

$$\rightarrow \zeta_2 = \begin{pmatrix} (1.13) & (11.98) \\ 0.5 & 0.5 \end{pmatrix},\ \delta_2(\zeta_2) = 1.0000$$

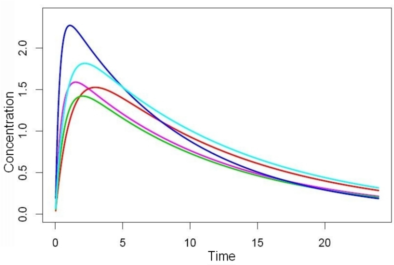
$\mathfrak{M}_{3,\beta}$-Information:

$$\rightarrow \zeta_3 = \begin{pmatrix} (0.93) & (11.99) \\ 0.5 & 0.5 \end{pmatrix},\ \delta_2(\zeta_3) = 0.9949$$

$\mathfrak{M}_{4,\beta}$-Information:

$$\rightarrow \zeta_4 = \begin{pmatrix} (0.96) & (\ 6.00) \\ 0.47 & 0.53 \end{pmatrix},\ \delta_2(\zeta_4) = 0.9895$$

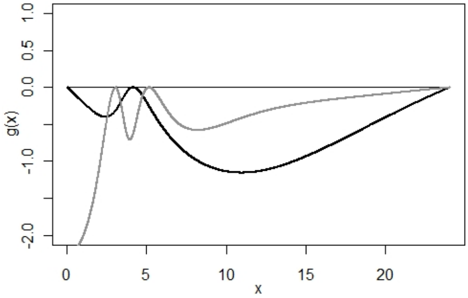
First Order Kinetics



$$Y_i = \frac{ke_i}{V_i(ke_i - \frac{Cl_i}{V_i})}(\exp(-x_i\frac{Cl_i}{V_i}) - \exp(-x_i ke_i))exp(\epsilon_i)$$

- $\epsilon_i \sim N(0, \sigma^2)$, $\xi_i = (x_i) \in [t_{min}, t_{max}]$,
- $ke_i = ke \exp(b_1)$, $Cl_i = Cl \exp(b_2)$, $V_i = V \exp(b_3)$
- $(b_1, b_2, b_3) \sim N(0, \sigma^2 D)$, $D = diag(d_1, d_2, d_3)$.

First Order Kinetics



$\mathfrak{M}_{2,\beta}$-Information:

$$\rightarrow \zeta_2 = \begin{pmatrix} (0.10) & (4.18) & (24.00) \\ 0.33 & 0.33 & 0.33 \end{pmatrix}, \delta_4(\zeta_2) = 0.55$$

$\mathfrak{M}_{4,\beta}$-Information:

$$\rightarrow \zeta_4 = \begin{pmatrix} (3.10) & (5.18) & (24.00) \\ 0.61 & 0.08 & 0.31 \end{pmatrix}, \delta_2(\zeta_4) = 0.66$$

Conclusions/Outlook

- Conclusions:
  - Motivation of model approximation
  - Different information $\rightarrow$ different design

Conclusions/Outlook

- Conclusions:
  - Motivation of model approximation
  - Different information $\rightarrow$ different design
- Outlook:
  - $\mathfrak{M}_{2,\beta}$-Information realistic?
  - Laplace approximation in β_i?
  - Efficiency of algorithms?

Thank you for your attention!