Statistical Modelling

Antony Overstall
University of Southampton

(Chapters 1–2 closely based on original notes by
Anthony Davison, Jon Forster & Dave Woods)

3. Nonlinear Models

Overview

1. Basic nonlinear models
2. Extending the nonlinear model
3. Computationally expensive nonlinear models
4. Model discrepancy

Basic nonlinear models

Linear models

- So far we have only considered models where the link function of the mean response is equal to the linear predictor, i.e. in the most general case of the generalised linear mixed model (GLMM)

$$\begin{aligned} \mu_{ij} &= \mathrm{E}(y_{ij}) \\ g(\mu_{ij}) &= \eta_{ij} = x_{ij}^{\mathrm{T}}\beta + z_{ij}^{\mathrm{T}}b_i, \end{aligned}$$

  and where the response distribution for y is from the exponential family of distributions

- The key point is that the linear predictor is a linear function of the parameters.
- The GLMM has the following special cases
  - linear models;
  - generalised linear models (GLMs);
  - linear mixed models (LMMs).
- These "linear" models form the basis of most applied statistical analyses.
- Usually, there is no scientific reason to believe these "linear" models are "true" for a given application. However, they might be "useful".

Nonlinear models

- Begin by assuming that y has a normal distribution and the link function, g, is the identity link and $z_{ij} = 0$, i.e.

$$y_i = x_i^T \beta + \epsilon_i, \tag{1}$$

where $\epsilon_i \sim N(0, \sigma^2)$, independently, where β are the p regression parameters.

- Consider extending this model so that instead of the mean response being the linear predictor $x_i^T \beta$, it is a nonlinear function of parameters, i.e.

$$y_i = \eta(x_i, \beta) + \epsilon_i, \tag{2}$$

where $\epsilon_i \sim N(0, \sigma^2)$, independently, where β are the p nonlinear parameters.

- Obviously, the model specified by (2) has the linear model (1) as a special case when $\eta(x, \beta) = x^{\mathrm{T}} \beta$.

- *Note that, sometimes the term nonlinear model is used to describe any model which is not a linear model (1), which would include GLMs and GLMMs.*

Example - Calcium Data

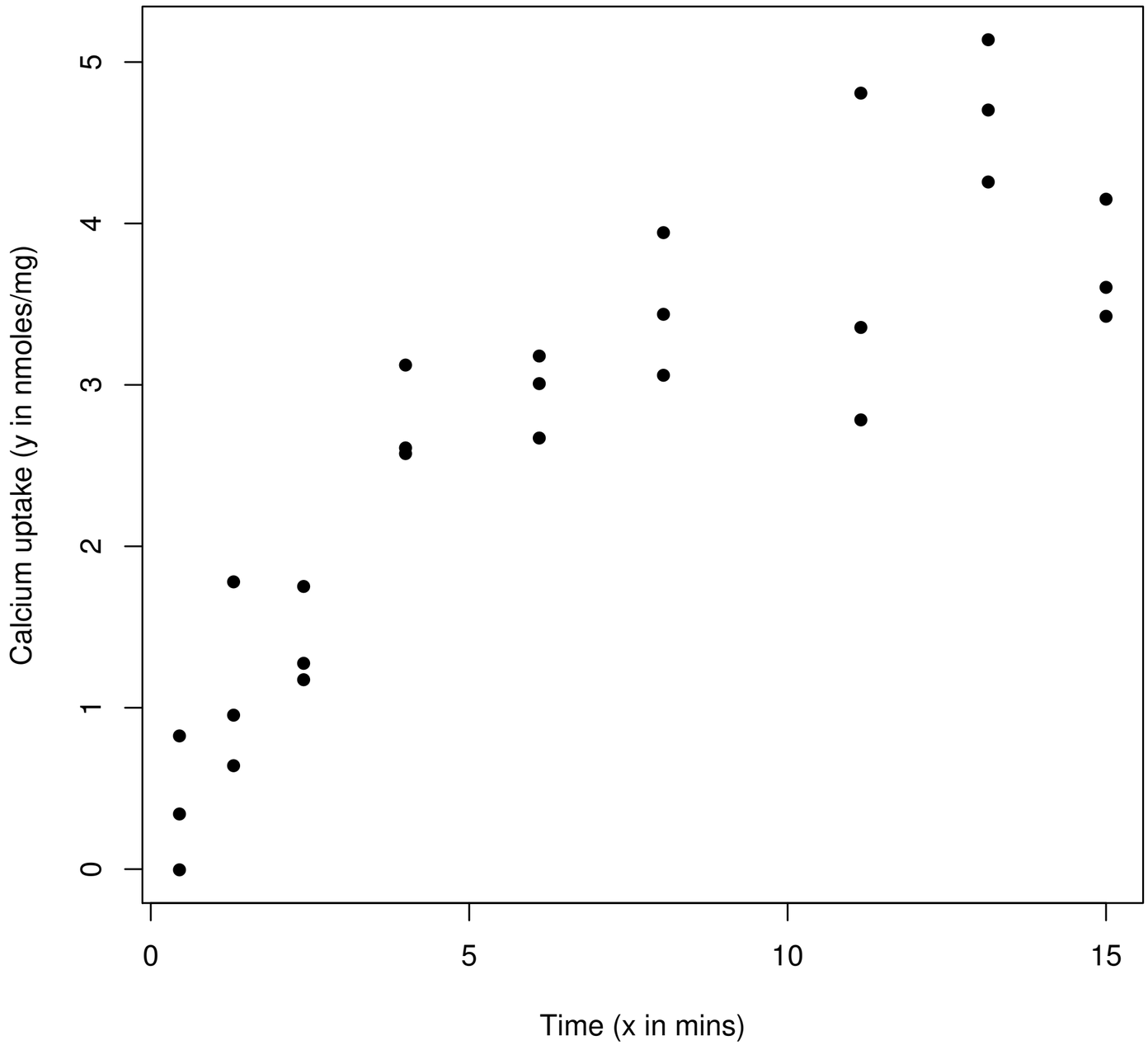
The response, y, is the uptake of calcium (in nmoles per mg) at time x (in minutes) by $n = 27$ cells in "hot" suspension.

```
> calcium
       x       y
1   0.45 0.34170
2   1.30 1.77967
3   2.40 1.75136
4   4.00 3.12273
5   6.10 3.17881
6   8.05 3.05959
7  11.15 4.80735
8  13.15 5.13825
9  15.00 3.60407
10  0.45 -0.00438
11  1.30 0.95384
12  2.40 1.27497
13  4.00 2.60958
```

Aim: To investigate how calcium is transported across cell membranes

Example - Calcium Data

Plot of calcium uptake against time.



Nonlinear parameters

Nonlinear parameters can be of two different types:

- **Physical parameters** have meaning within the science underlying the model, $\eta(x, \beta)$. Estimating the value of physical parameters contributes to scientific understanding.
- **Tuning parameters** do not have physical meaning. Their presence is often as a simplification of a more complex underlying system. Their estimation is to make the model fit best to reality.

In most cases, in a linear model, the regression parameters are tuning parameters.

Advantages and disadvantages

Advantages

- Can incorporate prior scientific knowledge through the function $\eta(x, \beta)$.
- Can fit simpler models, i.e. less parameters, to adequately describe observed data than by using linear models.
- Can provide extrapolated predictions (typically discouraged for linear models).
- Can directly contribute to scientific understanding through the estimation of physical parameters.

Disadvantages

- Need to specify the function $\eta(x, \beta)$.
- There are computational problems associated with these models.
- *All models are wrong*.

Specifying $\eta(x, \beta)$

How might the function $\eta(x, \beta)$ be specified?

- **Mechanistically** – prior scientific knowledge is incorporated into building a mathematical model for the mean response. This can often be complex and $\eta(x, \beta)$ may not be available in closed form.
- **Phenomenologically (empirically)** – a function $\eta(x, \beta)$ may be posited that appears to capture the non-linear nature of the mean response.

Example - Calcium Data

- Here the calcium uptake "grows" with time.
- There is a large class of phenomenological models for growth curves.
- Consider the non-linear model with

$$\eta(x, \beta) = \beta_0 \left(1 - \exp\left(-x/\beta_1\right)\right). \tag{3}$$

- This is derived by assuming that the rate of growth is proportional to the calcium remaining, i.e.

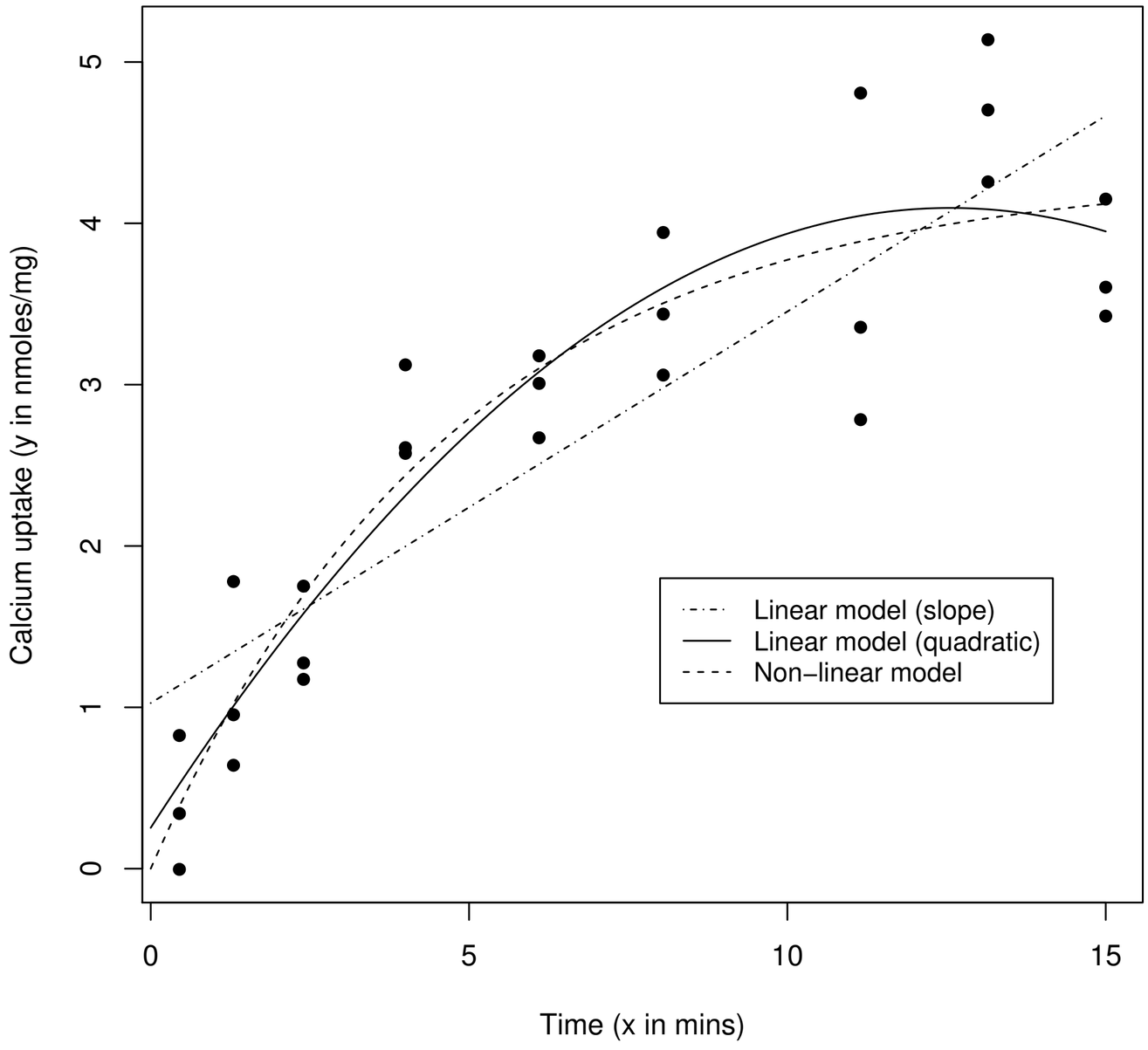
~~init~~ initial condition

$$\frac{\mathrm{d}\eta}{\mathrm{d}x} = (\beta_0 - \eta)/\beta_1. \qquad \eta(0, \beta) = 0$$

- The solution to this differential equation is (3).
- Interpretation of parameters:
  - β_0 - final size of population;
  - β_1 - (inversely) controls growth rate.

Example - Calcium Data

- Plot of calcium uptake against time.
- Includes fitted lines for three different models



Example - Calcium Data

- A comparison of the goodness-of-fit for the three models:

Model	Parameters (p)	$l(\widehat{\theta})$	AIC
Linear model (slope)	2	-28.70	63.40
Linear model (quadratic)	3	-20.95	49.91
Non-linear model	2	-20.95	47.91

- The goodness-of-fit for the quadratic and nonlinear models is identical (to 2 decimal places).
- Since the nonlinear model is simpler (less parameters), it is the preferred model.

Extending the nonlinear model

Introduction

- Nonlinear models can be extended to
  1. non-normal responses;
  2. clustered responses;

  in the same way as linear models.
- Here, we consider clustered responses and briefly discuss the nonlinear mixed model.

Example - Theophylline

- Theophylline is an anti-asthmatic drug.
- An experiment was performed on $n = 12$ individuals to investigate the way in which the drug leaves the body.
- The study of drug concentrations inside organisms is called *pharmacokinetics*.
- An oral dose, D_i, was given to the ith individual at time $t = 0$, for $i = 1, \ldots, n$.
- The concentration of theophylline in the blood was then measured at 11 time points in the next 25 hours.
- Let y_{ij} be the theophylline concentration (mg/L) for individual i at time t_{ij}.

Example - Theophylline

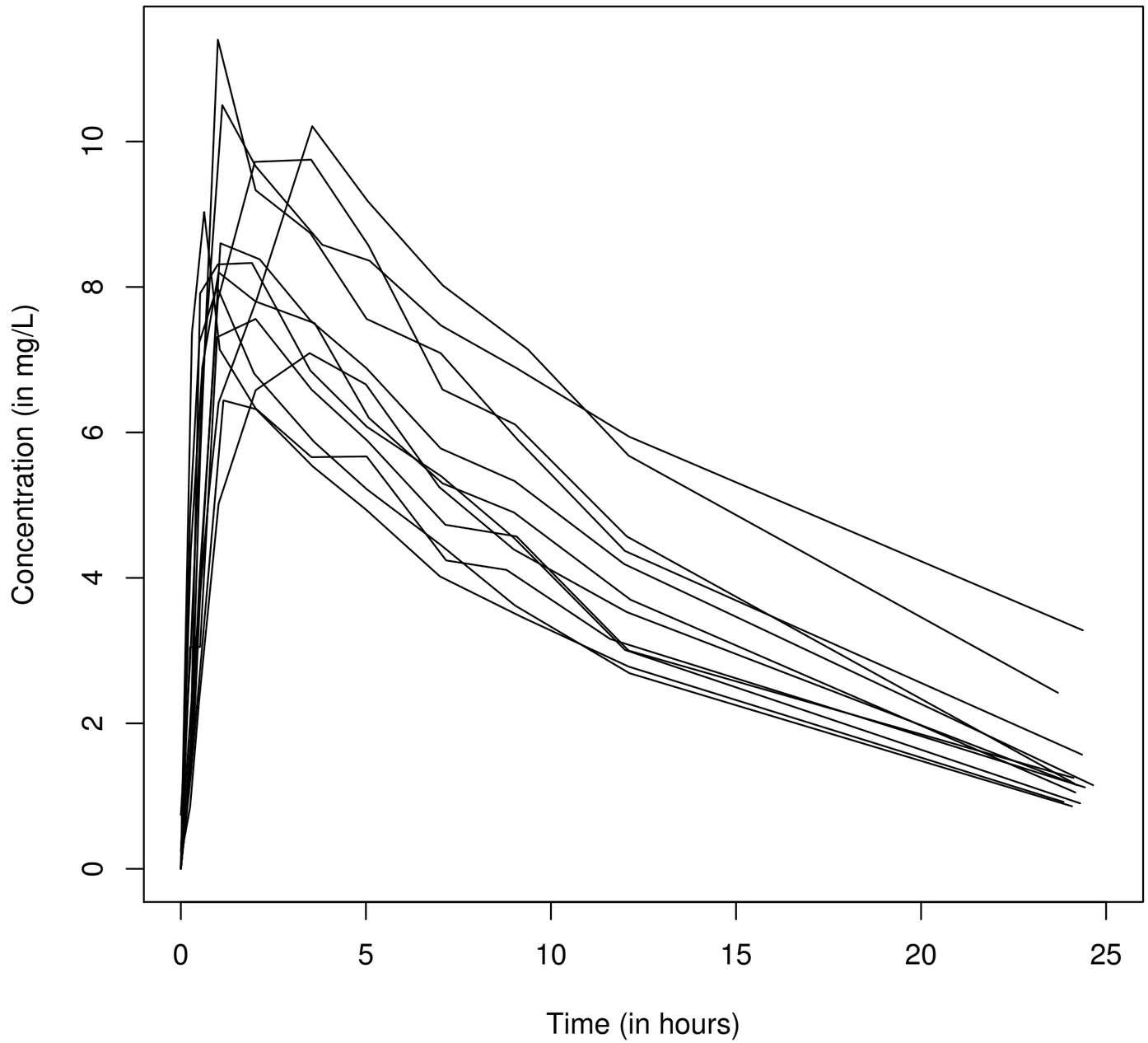
i Di tij yij

```
> Theoph
Grouped Data: conc ~ Time | Subject
    Subject   Wt Dose  Time  conc
1         1 79.6 4.02  0.00  0.74
2         1 79.6 4.02  0.25  2.84
3         1 79.6 4.02  0.57  6.57
4         1 79.6 4.02  1.12 10.50
5         1 79.6 4.02  2.02  9.66
6         1 79.6 4.02  3.82  8.58
7         1 79.6 4.02  5.10  8.36
8         1 79.6 4.02  7.03  7.47
9         1 79.6 4.02  9.05  6.89
10        1 79.6 4.02 12.12  5.94
11        1 79.6 4.02 24.37  3.28
12        2 72.4 4.40  0.00  0.00
13        2 72.4 4.40  0.27  1.72
```

Weight - not used in the model

Example - Theophylline

Plot of concentration of theophylline against time for each of the individuals.



There is a sharp increase in concentration followed by a steady decrease.

Example - Theophylline

- Compartmental models are a common class of model used in pharmacokinetics studies.
- If the initial dosage is D, then a two-compartment open pharmacokinetic model is

$$\eta(\beta, D, t) = \frac{D\beta_1\beta_2}{\beta_3(\beta_2-\beta_1)}\left(\exp\left(-\beta_1 t\right) - \exp\left(-\beta_2 t\right)\right),$$

  where the (positive) nonlinear parameters are
  - β_1 is the elimination rate and controls the rate at which the drug leaves the organism;
  - β_2 is the absorption rate and controls the rate at which the drug enters the blood;
  - β_3 is the clearance and controls the volume of blood for which a drug is completely removed per time unit.

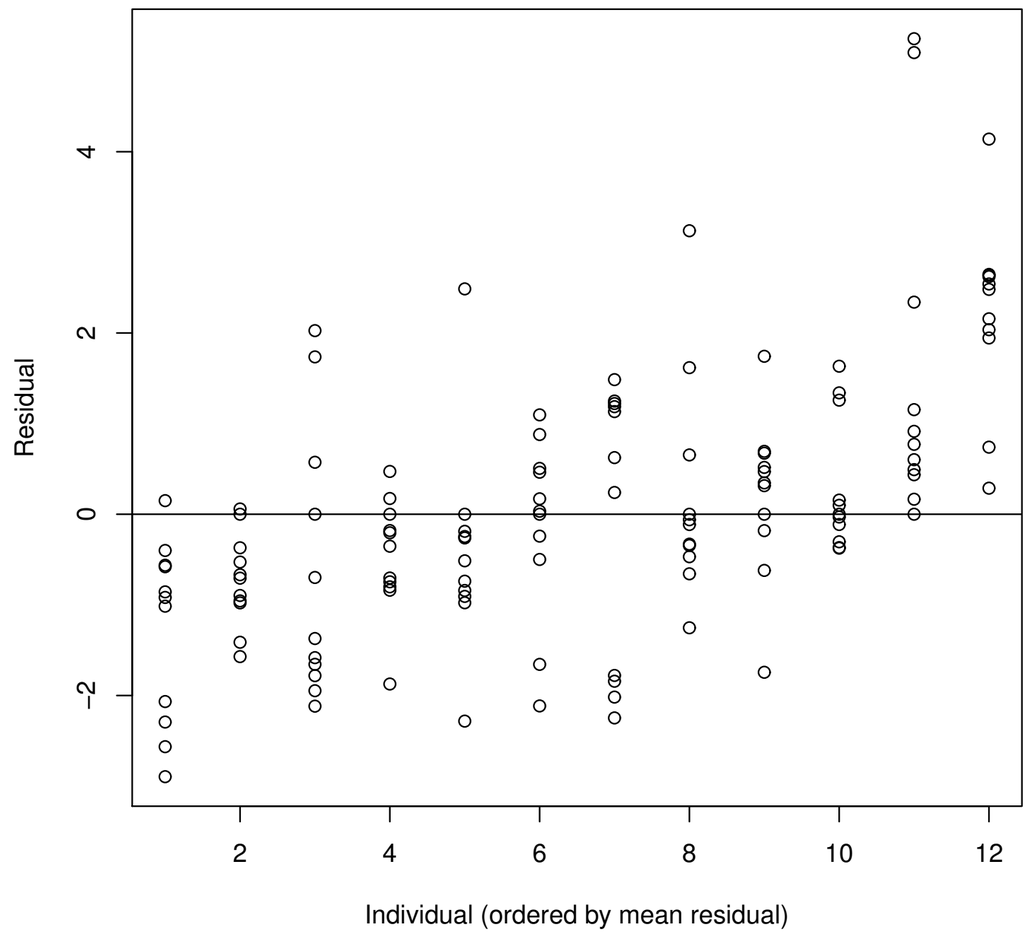
Example - Theophylline

- Initially ignore the dependence induced from repeated measurements on individuals and assume the following basic nonlinear model

$$y_{ij} = \eta(\beta, D_i, t_{ij}) + \epsilon_{ij},$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$.

- Residuals show evidence of an unexplained difference between individuals.



Nonlinear mixed effects models

- A nonlinear mixed model is

fixed effects

random effects

$$y_{ij} = \eta(\beta + b_i, x_{ij}) + \epsilon_{ij},$$

where

$$\begin{aligned} \epsilon_{ij} &\overset{iid}{\sim} N(0, \sigma^2), \\ b_i &\overset{ind}{\sim} N(0, \Sigma_b), \end{aligned}$$

and Σ_b is a $q \times q$ covariance matrix.

- This model specifies that $\beta_i = \beta + b_i$ are the nonlinear parameters for the ith cluster, i.e. the cluster-specific nonlinear parameters.
- In the case of the Theophylline example, each individual would have unique elimination rate, absorption rate and clearance.
- Obviously, $\beta_i \sim N(\beta, \Sigma_b)$. The mean, β, of the cluster-specific nonlinear parameters across all individuals are the population nonlinear parameters.

Non-linear mixed effects models

- We might like to specify the model in a way such that only a subset of the nonlinear parameters can be different for each individual, and the remainder fixed for all individuals.
- Suppose $q \leq p$ nonlinear parameters are can be different for each individual, then a more general way of writing the nonlinear mixed model is

$$y_{ij} = \eta(\beta + Ab_i, x) + \epsilon_{ij},$$

where

$$\begin{aligned} \epsilon_{ij} &\sim N(0, \sigma^2) \\ b_i &\sim N(0, \Sigma_b), \end{aligned}$$

where Σ_b is a $q \times q$ covariance matrix and A is a $p \times q$ binary matrix.

- A allows the specification of the fixed and varying nonlinear parameters.

E.g. Suppose $p=3$, and we want only the 2nd element of $\beta_i = (\beta_{i1}, \beta_{i2}, \beta_{i3})$ to vary across individuals, then $q=1$ and $A = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ $Ab_i = \begin{pmatrix} 0 \\ b_i \\ 0 \end{pmatrix}$

Special case of linear mixed models

- The linear mixed model is a special case of the nonlinear mixed model where

$$\eta(\beta, x) = x^{\mathrm{T}}\beta.$$

- Then

$$\begin{aligned} \eta(\beta + Abx) &= x^{\mathrm{T}}(\beta + Ab) \\ &= x^{\mathrm{T}}\beta + x^{\mathrm{T}}Ab, \end{aligned}$$

so $z = A^{\mathrm{T}}x$.

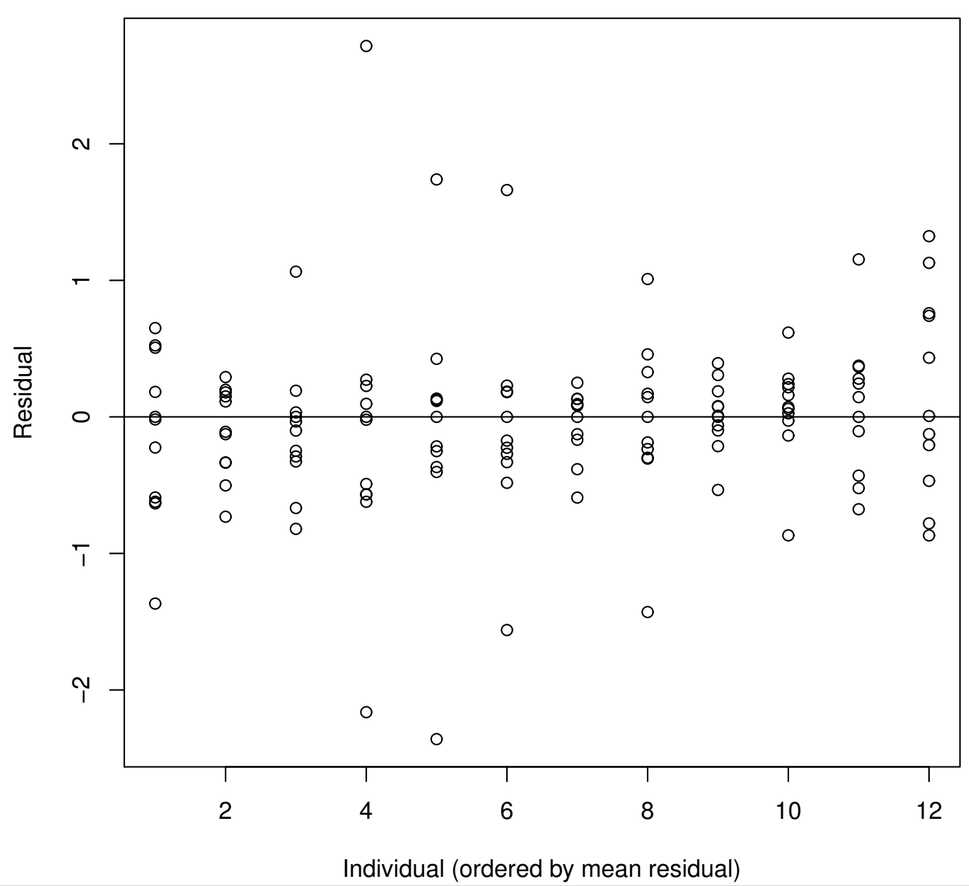
- For a random intercept model, where $q = 1$, $A^{\mathrm{T}} = (1, 0, \ldots, 0)$.

Example - Theophyline

- Returning to the Theophyline example, we fit the nonlinear mixed model, allowing all of the nonlinear parameters to vary across individuals, i.e. $A = I_3$.
- Estimates:

$$\begin{array}{rclrcl} \widehat{\beta}_1 & = & 0.0864 & \widehat{\Sigma}_{b11} & = & 0.0166 \\ \widehat{\beta}_2 & = & 1.6067 & \widehat{\Sigma}_{b22} & = & 0.9349 \\ \widehat{\beta}_3 & = & 0.0399 & \widehat{\Sigma}_{b33} & = & 0.0491 \end{array}$$

- $\text{AIC} = 372.6$



Example - Theophyline

- The estimated value of Σ_{b11} is "small" so we fit the nonlinear mixed model, allowing absorption rate and clearance to vary across individuals, i.e.

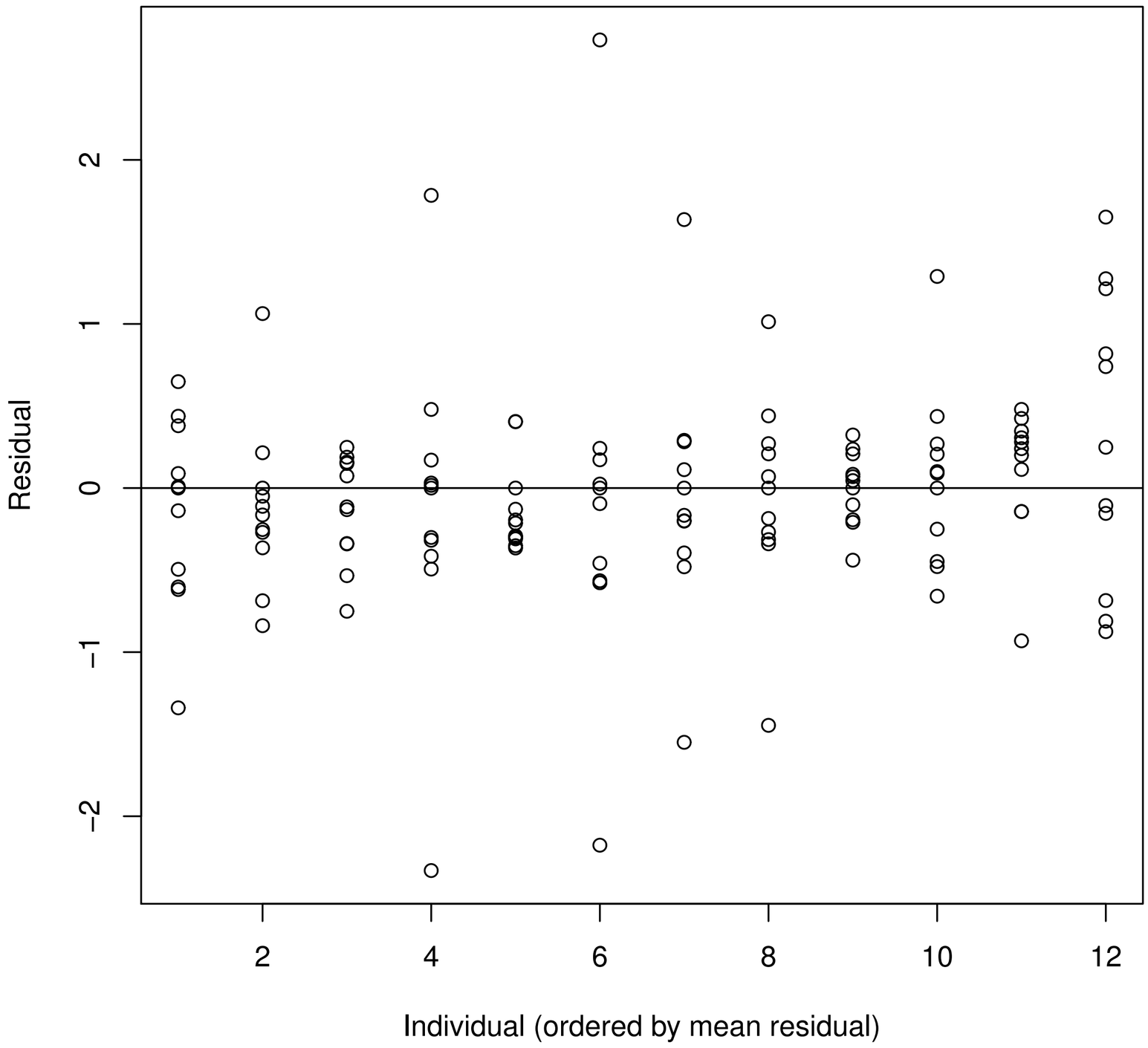
$$A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

- Estimates:

$$\begin{array}{lllll} \widehat{\beta}_1 & = & 0.0859 & & \\ \widehat{\beta}_2 & = & 1.6032 & \widehat{\Sigma}_{b22} & = & 0.6147 \\ \widehat{\beta}_3 & = & 0.0397 & \widehat{\Sigma}_{b33} & = & 0.0284 \end{array}$$

- $\mathrm{AIC} = 368.6$
- No further model simplifications reduce the AIC.

Extensions to nonnormal responses



Extensions to nonnormal responses

- Nonlinear models can be extended to nonnormal responses in the same way as linear models.
- The most general model is the generalised nonlinear mixed model (GNLMM).
- y_{ij} is from exponential family.
- $\mathrm{E}(y_{ij}) = \mu_{ij}$.
- $g(\mu_{ij}) = \eta(\beta + Ab_i, x_{ij})$.
- This model has the following special cases:

linear model	nonlinear model
linear mixed model	nonlinear mixed model
generalised linear model	generalised nonlinear model
generalised linear mixed model	

Issues

There are various technical and practical issues related to fitting nonlinear models (some of these are common to GLMs and GLMMs).

- ☐ Approximation of likelihood function (random effects are integrated out)
- ☐ Convergence of optimisation routines to find estimates
- ☐ Existence of estimates
- ☐ Reliability of asymptotic inference
- ☐ Computational expense of evaluating $\eta(\beta, x)$.
- ☐ *All models are wrong*.

These are all areas of current research.

Computationally expensive nonlinear models

Computationally expensive nonlinear models

- It typically requires many evaluations of $\eta(\beta, x)$ to fit a nonlinear model, either to find the estimate of β or to generate an MCMC sample.
- What happens if the non-linear model $\eta(\beta, x)$ is computationally expensive?
- For example, $\eta(\beta, x)$ could be the numerical solution to a system of differential equations where the exact solution is not available in closed form.
- The numerical solution to $\eta(\beta, x)$, implemented in computer code, is computationally expensive to evaluate.
- This can render model fitting to be infeasible.

Computer experiments and emulators

- One approach is to develop an approximation, $\widehat{\eta}(\beta, x)$, to the non-linear model $\eta(\beta, x)$.
- Evaluation of $\widehat{\eta}(\beta, x)$ replaces evaluation of $\eta(\beta, x)$ in all model fitting procedures.
- The approximation is typically called an *emulator* or *surrogate*.
- How is such an emulator constructed?
- Answer: via a computer experiment. This topic is briefly discussed here and will be covered in much more detail on the APTS Week 4 module: **Design of Experiments and Studies**

Computer experiments and emulators

- Let $z = (\beta, x)$ be the inputs to the nonlinear model such that $\eta(\beta, x) = \eta(z)$. Let d be the dimension of z.
- The nonlinear model is evaluated at a "small" number, m, of inputs

$$\zeta = \{z^1, \ldots, z^m\},$$ ← "design"

where $z^i = (\beta^i, x^i)$, for $i = 1, \ldots, m$.

- Finally, let $\eta^i = \eta(z^i)$, for $i = 1, \ldots, m$ and $\eta = \left(\eta^1, \ldots, \eta^m\right)$.

Gaussian Process Emulators

- The most commonly-used emulator is a Gaussian Process (GP) emulator.
- Here any finite collection of evaluations of $\eta(z)$ is assumed to have a multivariate normal distribution.
- Suppose $\eta_0 = \eta(z^0) = \eta(\beta^0, x^0)$ is the value of the nonlinear model we wish to predict.
- Assumption:

correlation matrix

$$\begin{pmatrix} \eta \\ \eta^0 \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \vdots \\ \theta \end{pmatrix}, \tau^2 \begin{pmatrix} C & c^{\mathrm{T}} \\ c & 1 \end{pmatrix} \right),$$

  i.e. a multivariate normal with (marginally) common mean θ and variance τ^2.

- Note that

$$C_{ij} \quad - \quad \text{correlation between } \eta(z^i) \text{ and } \eta(z^j)$$
$$c_i \quad - \quad \text{correlation between } \eta(z^i) \text{ and } \eta(z^0)$$

Gaussian Process Emulators

- By the properties of the multivariate normal

$$\eta^0|\eta \sim N\left(\theta + c^{\mathrm{T}}C^{-1}(\eta - \theta 1_m), \tau^2\left(1 - c^{\mathrm{T}}C^{-1}c\right)\right),$$

where 1_m is a vector of m ones.

- Structure is imposed on the elements of C and c as follows

$$\begin{aligned} C_{ij} &= \kappa(z^i, z^j; \rho) \\ c_i &= \kappa(z^i, z^0; \rho) \end{aligned}$$

where $\kappa(\cdot, \cdot; \rho)$ is a correlation function depending on ρ.

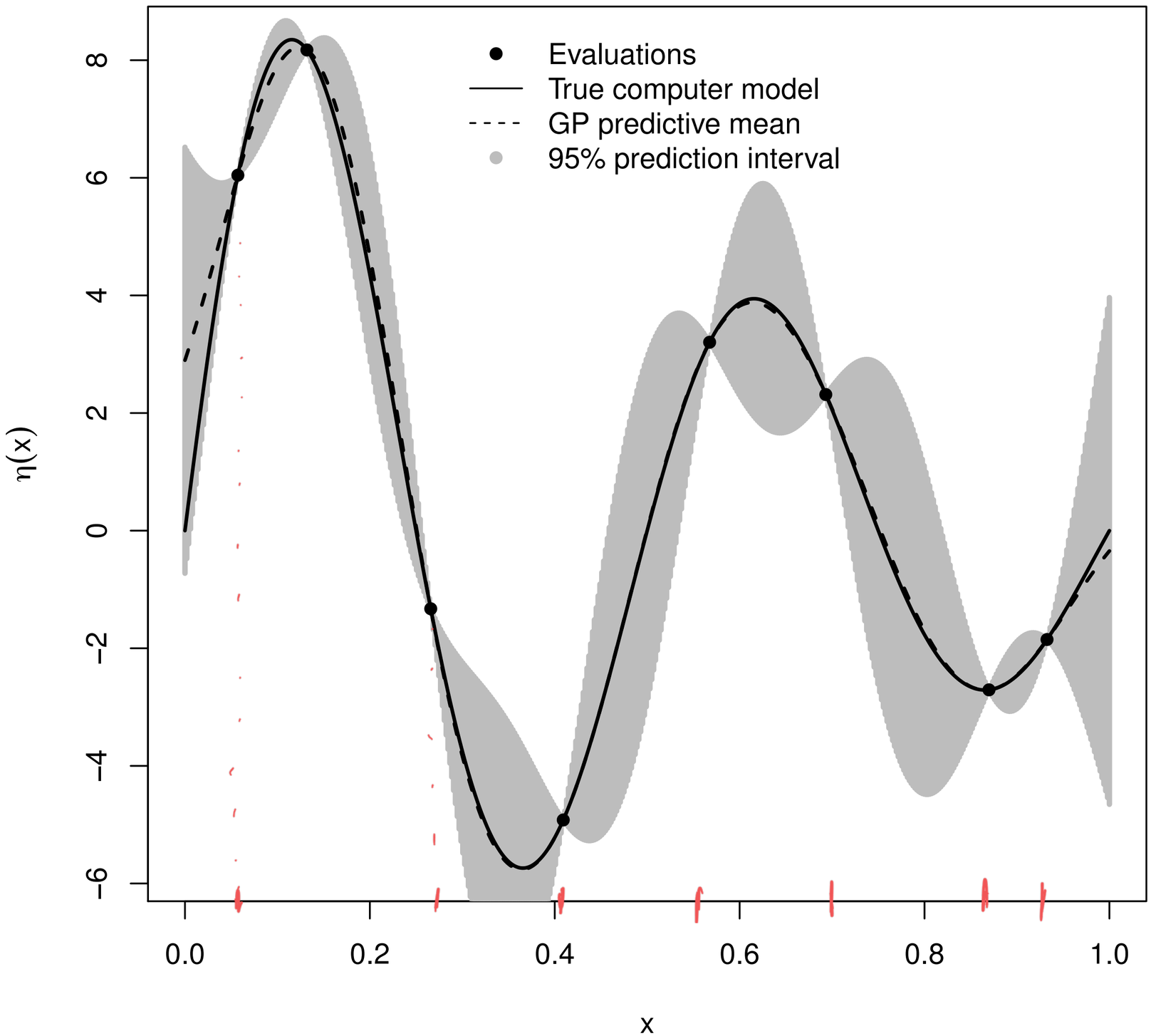
- A commonly-used correlation function is the squared exponential:

$$\kappa(z^i, z^j; \rho) = \exp\left(-\sum_{k=1}^{d} \rho_i \left(z_k^i - z_k^j\right)^2\right).$$

z_k^i is the k^{th} element of z^i

- θ, τ^2 and ρ can be estimated via maximum likelihood or a Bayesian approach taken.

Example



Model discrepancy

Model discrepancy

- Under the basic nonlinear model we are assuming

$$y = \eta(\beta, x) + \epsilon,$$

  i.e. the observed responses are given by the nonlinear model plus some random error.
- However, all models are wrong.
- In Chapter 1, we accounted for this by considering more complex models.
- If $\eta(\beta, x)$ is a mechanistic model, there is not really scope to make it more complex.

Model discrepancy

- Let $\mu(x)$ be true system depending on x.
- We observe

$$y = \mu(x) + \epsilon.$$

- $\eta(\beta, x)$ is our model and is our best guess at $\mu(x)$ where β are the true value of the parameters.
- Assume

$$\mu(x) = \eta(\beta, x) + \delta(x),$$

where $\delta(x)$ is the difference between reality and our model, i.e. the model discrepancy.

- Therefore

$$y = \eta(\beta, x) + \delta(x) + \epsilon.$$

- The model discrepancy is an unknown function.
- Taking a Bayesian approach, a prior is placed on this function. In particular, the Kennedy & O'Hagan (2001) framework places a Gaussian process prior on this function.

Example - Illustrative

- This example is adapted from Brynjarsdottir & O'Hagan (2014).
- Suppose reality is such that

$$\mu(x) = \frac{\beta x}{1+x/20},$$

where $\beta = 0.65$ is the true value of nonlinear parameter.

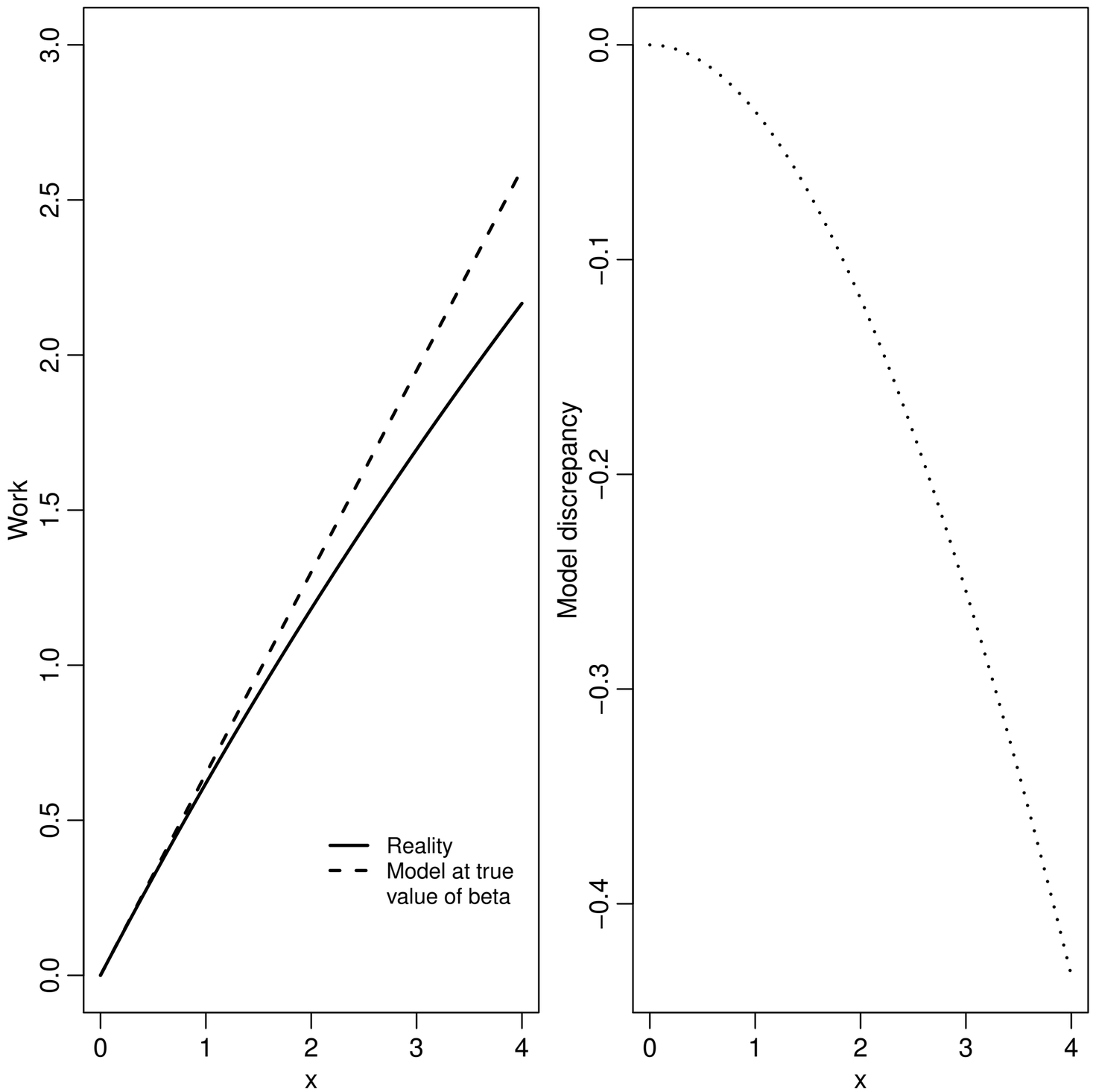
- Our model is such that

$$\eta(\beta, x) = \beta x.$$

- The model discrepancy is then

$$\delta(x) = \frac{-\beta x^2}{20+x}.$$

Example - Illustrative



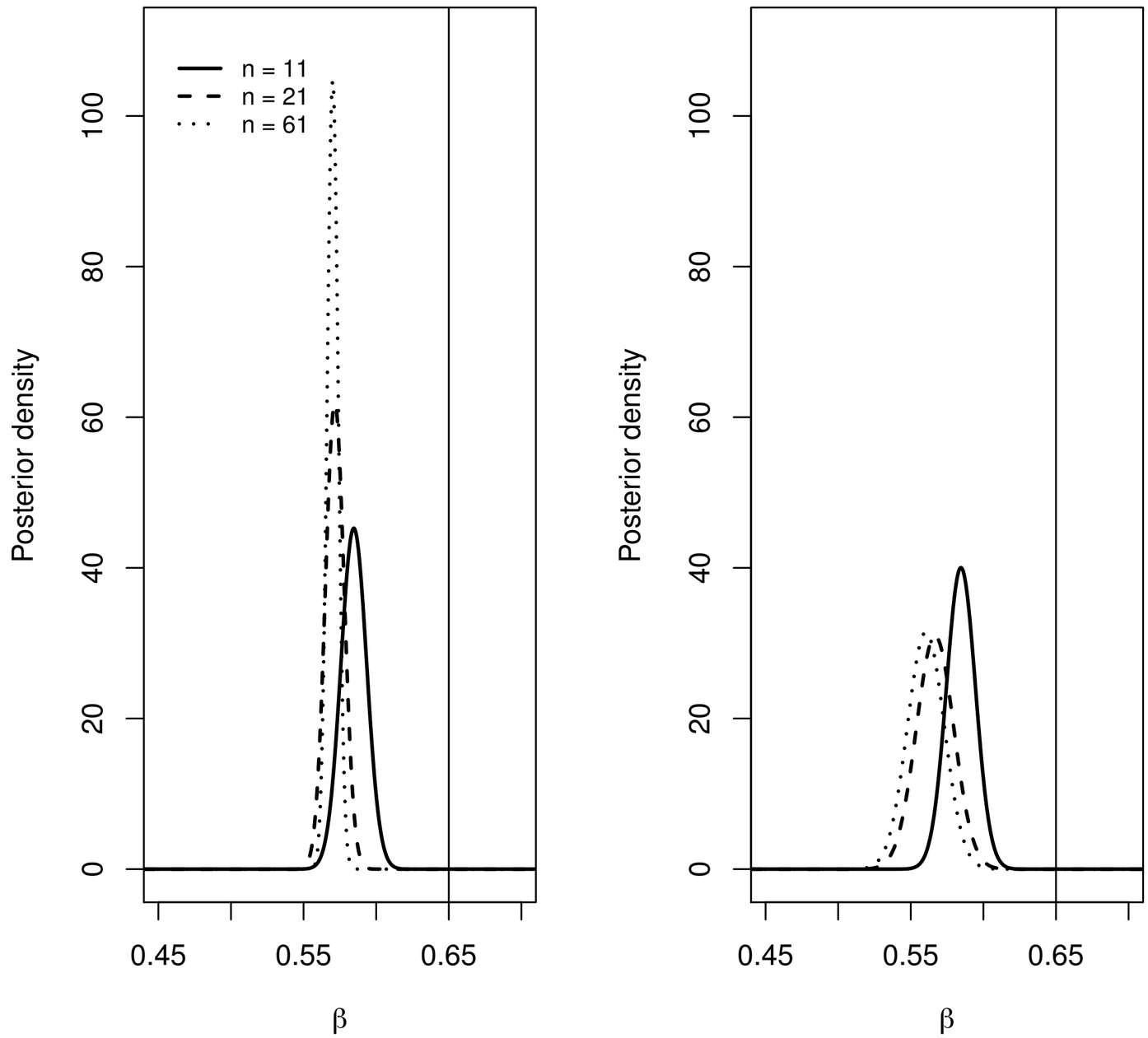
Example - Illustrative

- Consider an experiment to achieve the three aims of
  1. Interpolation prediction, i.e. predict value of y for $x = 2$;
  2. Extrapolation prediction, i.e. predict value of y for $x = 6$;
  3. Estimate value of β.

$n = 11, 21, 61$

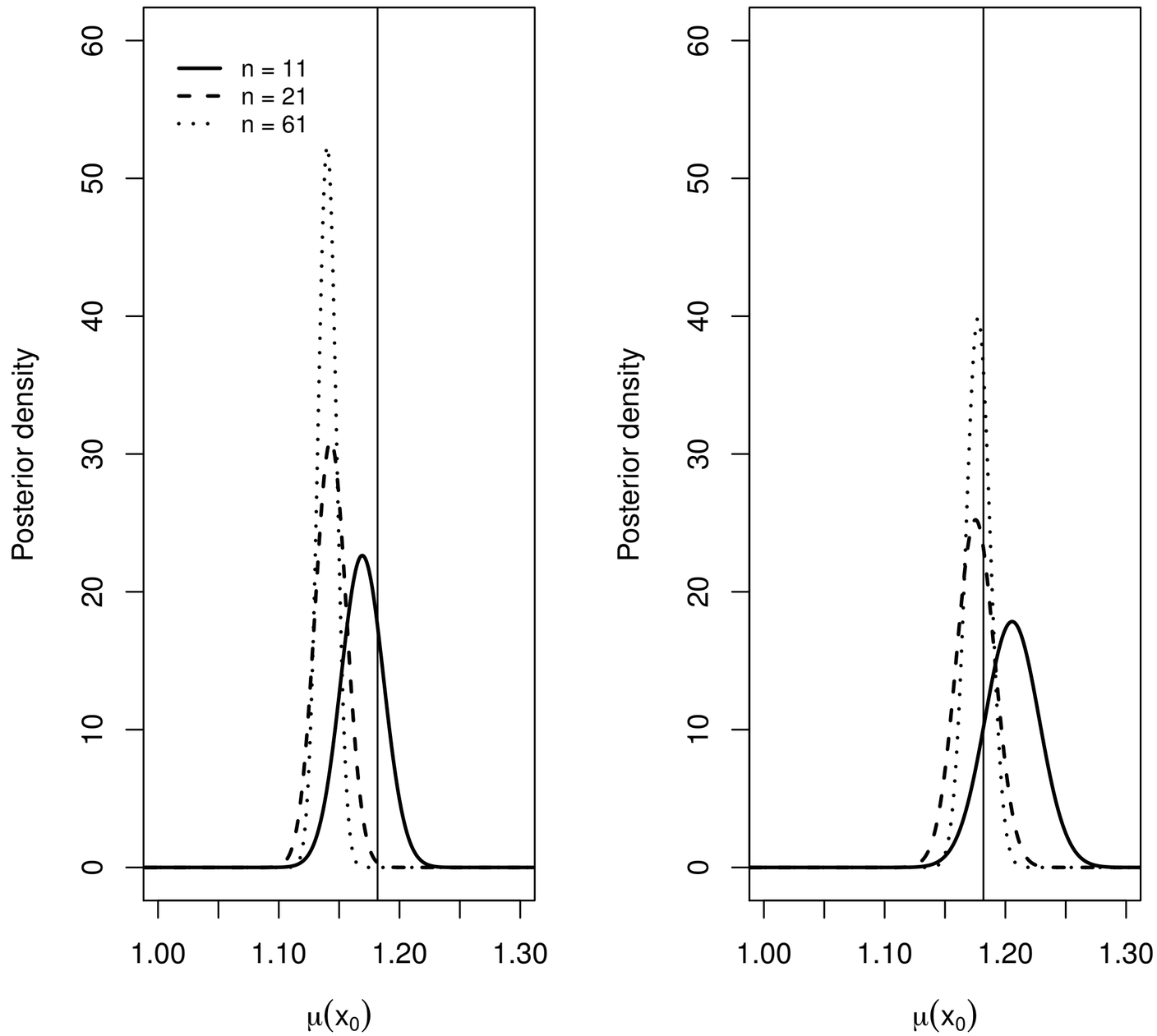
- We observe the response y at n values of $x \in [0, 4]$.
- We take two different approaches:
  1. Ignore mode discrepancy - just fit basic nonlinear model;
  2. Use Kennedy & O'Hagan framework with nonlinear model ~~with~~ and model discrepancy.

Posterior density of β under basic nonlinear model (left) and nonlinear model with model discrepancy (right).



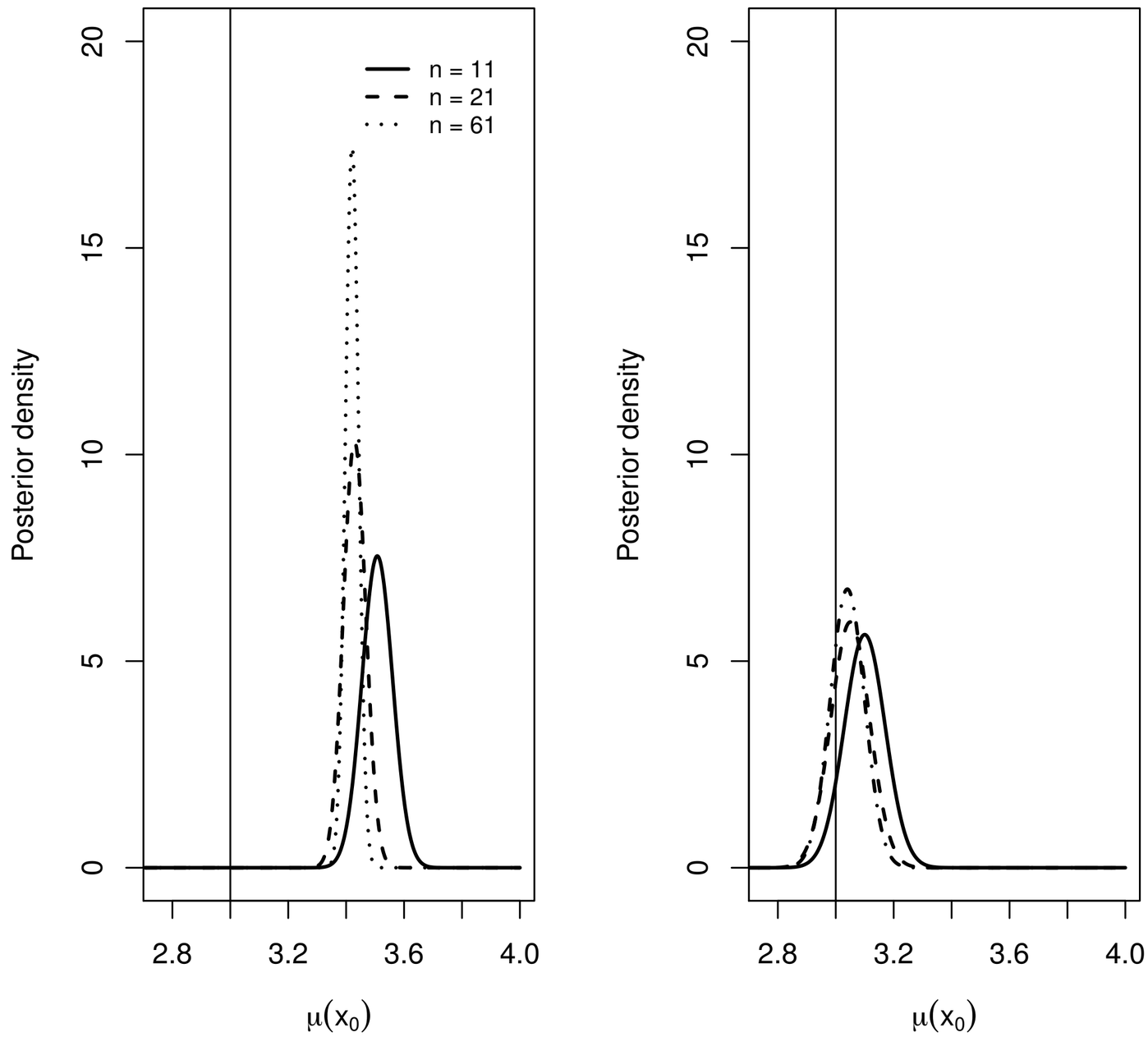
Example - Interpolation

Posterior density of $\mu(x_0)$ (with $x_0 = 2$) under basic nonlinear model (left) and nonlinear model with model discrepancy (right).



Example - Extrapolation

Posterior density of $\mu(x_0)$ (with $x_0 = 6$) under basic nonlinear model (left) and nonlinear model with model discrepancy (right).



Model discrepancy - discussion

- Modelling the model discrepancy with a Gaussian process alleviated the problem with interpolation prediction but not for extrapolation prediction or parameter estimation.
- Brynjarsdottir & O'Hagan (2014) considered using a constrained Gaussian process to incorporate prior information on the model discrepancy (e.g. value at $x = 0$ and monotoncity) and this eased the problem for parameter estimation but not for extrapolation prediction.
- How to account for model discrepancy remains an open research problem.