Advanced dataanalysis and statistical modelling, Week 13

Hierarchical models summary and outlook

Jan Kloppenborg Møller, Henrik Madsen

May 7, 2018

Outline

Overview

General mixed effect models

Remember that the likelihood of a general mixed effect model can be written as

$$L_M(\theta; \mathbf{y}) = \int_{\mathbb{R}^q} L(\boldsymbol{\theta}; \mathbf{u}, \mathbf{y}) d\mathbf{u} \tag{1}$$

where $\theta = (\boldsymbol{\beta}, \psi)$ and

$$L(\boldsymbol{\theta}; \mathbf{u}, \mathbf{y}) = f_{Y|u}(\mathbf{y}; \mathbf{u}, \boldsymbol{\beta}) f_U(\mathbf{u}; \boldsymbol{\psi}) \tag{2}$$

is the joint likelihood.

General mixed effect models - one level of grouping

If only one level of grouping is present

$$L(\boldsymbol{\theta}; \mathbf{y}) = \prod_{i=1}^{M} \int_{\mathbb{R}^{q_i}} f_{Y|u_i}(\mathbf{y}; \mathbf{u}_i, \boldsymbol{\beta}) f_{U_i}(\mathbf{u}_i; \boldsymbol{\psi}) d\mathbf{u}_i \tag{3}$$

Which simplify the integration greatly, but might still be impossible to solve, even when $q_i = 1$.

Laplace approximation

Approximate the joint log-likelihood by the second order Taylor approximation

$$l_{LA}(\boldsymbol{\theta}; \mathbf{u}, \mathbf{y}) \approx l(\boldsymbol{\theta}; \tilde{\mathbf{u}}, \mathbf{y}) - \frac{1}{2}(\mathbf{u} - \tilde{\mathbf{u}})^T \mathbf{H}(\tilde{\mathbf{u}})(\mathbf{u} - \tilde{\mathbf{u}}) \tag{4}$$

with

$$\tilde{\mathbf{u}} = \arg\max_{\mathbf{u}} l(\mathbf{u}, \boldsymbol{\theta}, \mathbf{y}) \tag{5}$$

and

$$H(\tilde{\mathbf{u}}) = -l''_{uu}(\mathbf{u}, \boldsymbol{\theta}, \mathbf{y})|_{\mathbf{u}=\tilde{\mathbf{u}}} \tag{6}$$

The log likelihood is approximated by

$$l_{M,LA}(\boldsymbol{\theta}, \mathbf{y}) = \log f_{Y|u}(\mathbf{y}; \tilde{\mathbf{u}}, \boldsymbol{\beta}) + \log f_U(\tilde{\mathbf{u}}; \boldsymbol{\psi}) - \frac{1}{2}\log|H(\tilde{\mathbf{u}})| \tag{7}$$

Laplace approximation

As usual the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ is given by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\theta}(l_{M,LA}(\boldsymbol{\theta}, \mathbf{y})) \tag{8}$$

In general we need to solve (8) by numerical methods.

Each step in the numerical procedure require the numerical solution of the second stage model *and* the (an approximation of) the hessian at the estimated random effects.

In general numerical optimisation will speed up significantly if we can supply gradients of the objective function.

Laplace approximation work flow

0. Initialize $\boldsymbol{\theta}$ to some arbitrary value $\boldsymbol{\theta}_0$
1. With current value for $\boldsymbol{\theta}$ optimize joint likelihood w.r.t. $\boldsymbol{u}$ to get $\tilde{\boldsymbol{u}}_{\boldsymbol{\theta}}$ and corresponding Hessian $H(\tilde{\boldsymbol{u}}_{\boldsymbol{\theta}})$.
2. Use $\tilde{\boldsymbol{u}}_{\boldsymbol{\theta}}$ and $H(\tilde{\boldsymbol{u}}_{\boldsymbol{\theta}})$ to approximate $\ell_M(\boldsymbol{\theta})$
3. Compute value and gradient of $\ell_M(\boldsymbol{\theta})$
4. If the gradient is "$> \epsilon$" set $\boldsymbol{\theta}$ to a different value and go to 1.

Notice the huge number of — possibly high dimensional — optimizations that are required.

Overview

Hierarchical models - non-Gaussian case

The non-Gaussian case of the hierarchical models, where

$$g(E[\boldsymbol{Y}|\boldsymbol{U}]) = \boldsymbol{X\beta} + \boldsymbol{Zv}$$

and where $g(.)$ is an appropriate link function (Theorem 6.3). And U (with $g(\boldsymbol{v}) = \boldsymbol{u}$) belong to some appropriate distribution(s).
Or

$$\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Zv}$$

with $\boldsymbol{v} = g^{-1}(\boldsymbol{u})$

Hierarchical models - non-Gaussian case

Solution is based on the socalled hierarchical likelihood

$$h = l(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha}; \boldsymbol{y}|\boldsymbol{v}) + l(\boldsymbol{\alpha}; \boldsymbol{v})$$

The marginal for $\boldsymbol{y}$ is given by

$$l_M(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha}; \boldsymbol{y}) = \log\left(\int e^h d\boldsymbol{v}\right)$$

and the normal equation in Remark 6.8 is based on this marginal likelihood (see Lee and Nelder (1996, 2001)).

Overview

Likelihood

- Likelihood function $L(\theta) = P_\theta(Y = y)$
- Log likelihood function $\ell(\theta) = \log(L(\theta))$
- Score function $\ell'(\theta)$
- Maximum likelihood estimate $\widehat{\theta} = \underset{\theta \in \Theta}{\text{argmax}}\ \ell(\theta)$
- Observed information matrix $-\ell''(\widehat{\theta})$
- Invariance
- Assumptotic normality
- Wald statistic and quadratic approximation
- Profile likelihood

The likelihood function

Definition (Likelihood function)

Given the parametric density $f_{\boldsymbol{Y}}(\boldsymbol{y}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta^P$, for the observations $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ the *likelihood function for* $\boldsymbol{\theta}$ is the function

$$L(\boldsymbol{\theta}; \boldsymbol{y}) = c(y_1, y_2, \ldots, y_n) f_Y(y_1, y_2, \ldots, y_n; \boldsymbol{\theta})$$

where $c(y_1, y_2, \ldots, y_n)$ is a constant.

The likelihood function is thus (proportional to) the joint probability density for the actual observations considered as a function of $\boldsymbol{\theta}$.

Statistical methods in likelihood theory

The general results are based on asymptotic theory

- Asymptotic normallity of estimators (Wald statistics), beased on C-R lower bound
- Likelihood ratio test (χ^2-distribution)
- Profile likelihood (confidence intervals)

The general linear model - overview

- The classical GLM leads to a unique way of describing the variations of experiments with a *continuous* variable.
- The classical GLM's include
  - Regression analysis
  - Analysis of variance - ANOVA
  - Analysis of covariance - ANCOVA
- The residuals are assumed to follow a multivariate normal distribution in the classical GLM.
- A general linear model is:

$$\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$$

Estimation of mean value parameters

- Estimator: $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{y}$
- Distribution of estimator $\widehat{\boldsymbol{\beta}} \sim \mathrm{N}_k(\boldsymbol{\beta}, \sigma^2 \, (\boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{X})^{-1})$
- Estimator if variance $\hat{\sigma}^2 = \frac{(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})}{f} \quad \sim \sigma^2 \chi^2_f / f$
- The deviance: $D(\boldsymbol{y}, p(\boldsymbol{y})) = ||\boldsymbol{y} - p(\boldsymbol{y})||^2 = \sum_i (y_i - p(y)_i)^2$

Source	f	Deviance	Test statistic, F
Model versus hypothesis	$m_1 - m_0$	$\lVert p_1(\boldsymbol{y}) - p_0(\boldsymbol{y})\rVert^2$	$\frac{\lVert p_1(\boldsymbol{y}) - p_0(\boldsymbol{y})\rVert^2/(m_1 - m_0)}{\lVert\boldsymbol{y} - p_1(\boldsymbol{y})\rVert^2/(n - m_1)}$
Residual under model	$n - m_1$	$\lVert\boldsymbol{y} - p_1(\boldsymbol{y})\rVert^2$	
Residual under hypothesis	$n - m_0$	$\lVert\boldsymbol{y} - p_0(\boldsymbol{y})\rVert^2$	

Residual plots

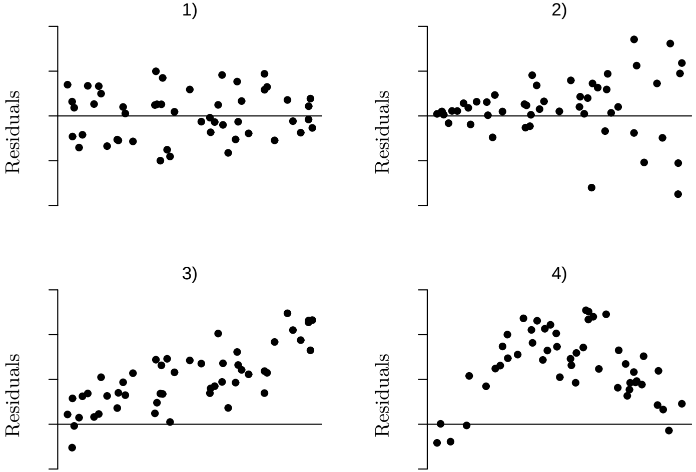


Figure: Residual plots

Confidence intervals and confidence regions

Confidence intervals for individual parameters

$100(1-\alpha)\%$ confidence interval for β_j is found as

$$\widehat{\beta}_j \pm t_{1-\alpha/2}(n-m_0)\widehat{\sigma}\sqrt{(\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}_{jj}}$$

Predictions

Most often the parameters are unknown but assume that there exist some estimates of $\boldsymbol{\theta}$. Assume also that the estimates are found by the estimator

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

then the variance of the prediction error can be stated.

G(eneral)LM vs G(eneralized)LM

General linear models	**Generalized linear models**
Normal distribution	Exponential dispersion family
Mean value linear	Function of mean value linear
Independent observations	Independent observations
Same variance	Variance function of mean
Easy to apply	Almost as easy to apply

Exponential families of distributions

Definition (An exponential dispersion family)

A family of probability densities which can be written on the form

$$f_Y(y;\theta) = c(y,\lambda)\exp(\lambda\{\theta y - \kappa(\theta)\})$$

is called an *exponential dispersion family* of distributions. The parameter $\lambda > 0$ is called the *precision parameter*.

- Basic idea: separate the mean value related distributional properties described by the *cumulant generator* $\kappa(\theta)$ from features as sample size, common variance, or common over-dispersion.
- In some cases the precision parameter represents a known number of observations as for the binomial distribution, or a known shape parameter as for the gamma (or χ^2-) distribution.
- In other cases the precision parameter represents an unknown dispersion like for the normal distribution, or an over-dispersion that is not related to the mean.

Exponential family densities as a statistical model

- Then the joint density, using the canonical parameter, is

$$f(\boldsymbol{y};\boldsymbol{\theta}) = \exp\left[\sum_{i=1}^{n} w_i(\theta_i y_i - \kappa(\theta_i))\right]\prod_{i=1}^{n} c(y_i, w_i)$$

- A *generalized linear model* for $Y_1, Y_2, \ldots, Y_n$ describes an affine hypothesis for $\eta_1, \eta_2, \ldots, \eta_n$, where

$$\eta_i = g(\mu_i)$$

  is a transformation of the mean values $\mu_1, \mu_2, \ldots, \mu_n$.
- With a linear hypotyhesis

$$\boldsymbol{\eta} - \boldsymbol{\eta}_0 = \boldsymbol{X}\boldsymbol{\beta} \text{ with } \boldsymbol{\beta} \in \mathbb{R}^k,$$

- The estimates must be found by an iterative procedure.

Specification of a generalized linear model

a) Distribution / Variance function:
Specification of the distribution – or the *variance function* $V(\mu)$.

b) Link function:
Specification of the *link function* $g(\cdot)$, which describes a function of the mean value which can be described linearly by the explanatory variables.

c) Linear predictor:
Specification of the linear dependency

$$g(\mu_i) = \eta_i = (\boldsymbol{x}_i)^T \boldsymbol{\beta}.$$

d) Precision (optional):
If needed the precision is formulated as *known individual weights*, $\lambda_i = w_i$, or as a *common dispersion parameter*, $\lambda = 1/\sigma^2$, or a *combination* $\lambda_i = w_i/\sigma^2$.

Properties of the ML estimator

- Under the hypothesis $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$ we have asymptotically

$$\frac{\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}}{\sqrt{\sigma^2}} \in N_k(\mathbf{0}, [\boldsymbol{X}^T \boldsymbol{W}(\boldsymbol{\beta})\boldsymbol{X}]^{-1}),$$

  with $\boldsymbol{W}(\boldsymbol{\beta})$ depending on the link "g" and the variance function.
- $\mathrm{D}[\widehat{\boldsymbol{\eta}}] \approx \widehat{\sigma}^2 \boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{X}^T$

Residuals

Residuals represents the difference between the data and the model. In the classical GLM the residuals are $r_i = y_i - \widehat{\mu}_i$. These are called response residuals for GLM's. Since the variance of the response is not constant for most GLM's we need some modification. We will look at:

- Deviance residuals
- Pearson residuals

Likelihood ratio tests

- The approximative normal distribution of the ML-estimator implies that many distributional results from the classical GLM-theory are carried over to generalized linear models as approximative (asymptotic) results.
- An example of this is the likelihood ratio test.
- For generalized linear models, this is not possible, and hence we shall use the asymptotic results for the logarithm of the likelihood ratio.

Source	f	Deviance	Mean deviance	Goodness of fit interpretation
Model $\mathcal{H}_{null}$	$k-1$	$\mathrm{D}\left(\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}); \widehat{\boldsymbol{\mu}}_{null}\right)$	$\dfrac{\mathrm{D}\left(\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}); \widehat{\boldsymbol{\mu}}_{null}\right)}{k-1}$	$G^2(\mathcal{H}_{null} \vert \mathcal{H}_1)$
Residual (Error)	$n-k$	$\mathrm{D}\left(\boldsymbol{y}; \boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})\right)$	$\dfrac{\mathrm{D}\left(\boldsymbol{y}; \boldsymbol{\mu}(\widehat{\boldsymbol{\beta}})\right)}{n-k}$	$G^2(\mathcal{H}_1)$
Corrected total	$n-1$	$\mathrm{D}\left(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}_{null}\right)$		$G^2(\mathcal{H}_{null})$

Formulation of the random model

Definition (One-way model with random effects)

Consider the random variables $Y_{i,j}, i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n_i$

$$Y_{ij} = \mu + U_i + \epsilon_{ij} \ ,$$

with $\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$ and $U_i \sim \mathrm{N}(0, \sigma_b^2)$, and independence.

The one-way model as a hierarchical model

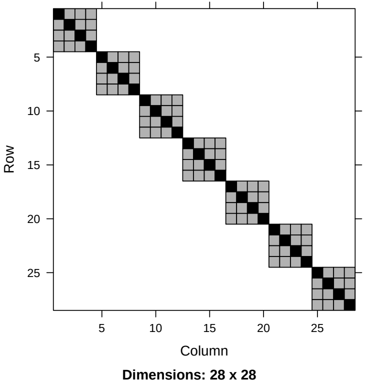
Putting $\mu_i = \mu + U_i$ we may formulated as a *hierarchical model*

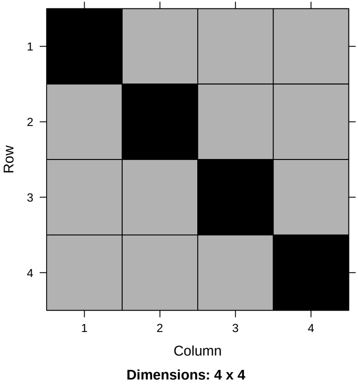
$$Y_{ij}|\mu_i \sim \mathrm{N}(\mu_i, \sigma^2) \ ;$$

μ_i is modeled as a realization of a random variable,

$$\mu_i \sim \mathrm{N}(\mu, \sigma_b^2),$$

Covariance structure for the whole set of observations





A general linear mixed effects model

A general linear mixed model can be presented in matrix notation by:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\varepsilon}, \quad \text{where } \mathbf{U} \sim N(0, \mathbf{G}) \text{ and } \varepsilon \sim N(0, \mathbf{R}).$$

- $\mathbf{Y}$ is the observation vector
- $\mathbf{X}$ is the design matrix for the fixed effects
- $\boldsymbol{\beta}$ is the vector containing the fixed effect parameters
- $\mathbf{Z}$ is the design matrix for the random effects
- $\mathbf{U}$ is the vector of random effects
  - It is assumed that $\mathbf{U} \sim N(\mathbf{0}, \mathbf{G})$
  - cov$(U_i, U_j) = G_{i,j}$ (typically $\mathbf{G}$ has a very simple structure (for instance diagonal))
- $\boldsymbol{\varepsilon}$ is the vector of residual errors
  - It is assumed that $\varepsilon \sim N(\mathbf{0}, \mathbf{R})$
  - cov$(\varepsilon_i, \varepsilon_j) = R_{i,j}$ (typically $\mathbf{R}$ is diagonal, but we shall later see some useful exceptions for repeated measurements)

The restricted/residual maximum likelihood method

- The maximum likelihood method tends to give (slightly) too low estimates of the random effects parameters. We say it is *biased downwards*
- The *restricted/residual maximum likelihood (REML)* method modifies the maximum likelihood method by maximizing:

$$\ell_{re}(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\psi}) \propto -\frac{1}{2}\left\{\log|\mathbf{V}(\boldsymbol{\psi})| + (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{V}(\boldsymbol{\psi}))^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \log|\mathbf{X}^T(\mathbf{V}(\boldsymbol{\psi}))^{-1}\mathbf{X}|\right\}$$

  which gives unbiased estimates (at least in balanced cases)
- The REML method is generally preferred in mixed models

Estimation of random effects

- Formally, the random effects, $\boldsymbol{U}$ are not parameters in the model, and the usual likelihood approach does not make much sense for "estimating" these random quantities.
- It is, however, often of interest to assess these "latent", or "state" variables.
- We formulate a so-called *hierarchical likelihood* by writing the joint density for observable as well as unobservable random quantities.

$$
\begin{aligned}
f(\mathbf{y},\mathbf{u};\beta,\psi) =& f_{Y|u}(\mathbf{y};\beta) f_U(\mathbf{u};\psi) \\
=& \frac{1}{(\sqrt{2})^N\sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{u})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{u})}\times \\
& \frac{1}{(\sqrt{2})^q\sqrt{|\boldsymbol{\Psi}|}} e^{-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Psi}^{-1}\mathbf{u}}
\end{aligned}
$$

General mixed effects models

The general mixed effects model can be represented by its likelihood function:

$$L_M(\boldsymbol{\theta}; \boldsymbol{y}) = \int_{\mathbb{R}^q} L(\boldsymbol{\theta}; \boldsymbol{u}, \boldsymbol{y}) d\boldsymbol{u}$$

- $\boldsymbol{y}$ is the observed random variables
- $\boldsymbol{u}$ is the q unobserved random variables
- $\boldsymbol{\theta}$ is the model parameters to be estimated

The likelihood function L is the joint likelihood of both the observed and the unobserved random variables.

The likelihood function for estimating $\boldsymbol{\theta}$ is the marginal likelihood L_M obtained by integrating out the unobserved random variables.

The Laplace approximation

- We need to calculate the difficult integral

$$L_M(\boldsymbol{\theta}, \boldsymbol{y}) = \int_{\mathbb{R}^q} L(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y}) d\boldsymbol{u}$$

- So we set up an approximation of $\ell(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y}) = \log L(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y})$

$$\ell(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y}) \approx \ell(\boldsymbol{\theta}, \hat{\boldsymbol{u}}_{\boldsymbol{\theta}}, \boldsymbol{y}) - \frac{1}{2}(\boldsymbol{u} - \hat{\boldsymbol{u}}_{\boldsymbol{\theta}})^T \left(-\ell''_{uu}(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y})|_{\boldsymbol{u}=\hat{\boldsymbol{u}}_{\boldsymbol{\theta}}}\right)(\boldsymbol{u} - \hat{\boldsymbol{u}}_{\boldsymbol{\theta}})$$

- Which (for given $\boldsymbol{\theta}$) is the 2. order Taylor approximation around:

$$\hat{\boldsymbol{u}}_{\boldsymbol{\theta}} = \underset{\boldsymbol{u}}{\text{argmax}}\ L(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y})$$

Heirarchical Generalized Models

- A characteristic property of the generalized linear models is that **the variance,** $\mathrm{Var}[Y]$ **is a known function**, $V(\mu)$, that only depends on μ
$$\mathrm{Var}[Y_i] = \lambda_i V(\mu) = \frac{\sigma^2}{w_i} V(\mu)$$
where w_i denotes a known **weight**, associated with the i'th observation, and where σ^2 denotes a common **dispersion parameter**
- The dispersion parameter σ^2 serve to **express overdispersion** in situations where the residual deviance is too large.
- An alternative method for modeling overdispersion, is by **hierarchical models**, analogous to the mixed effects models for the normally distributed observations.

Example: Hierarchical Binomial-Beta distribution model

The natural conjugate distribution to the binomial is a Beta-distribution.

Theorem

Consider the generalized one-way random effects model for $Z_1, Z_2, \dots, Z_k$ *given by*

$$\begin{aligned} Z_i|p_i &\sim B(n, p_i) \\ p_i &\sim \mathrm{Beta}(\alpha, \beta) \end{aligned}$$

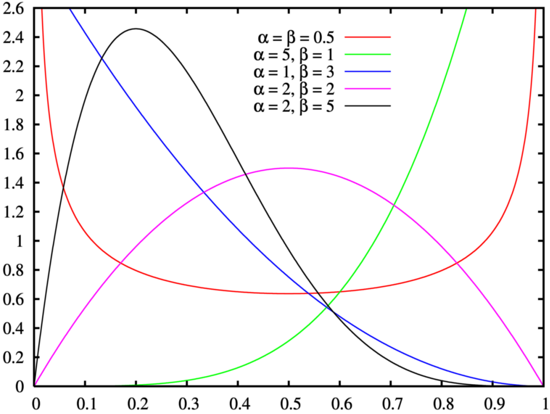
i.e. the conditional distribution of Z_i *given* p_i *is a Binomial distribution, and the distribution of the mean value* p_i *is a Beta distribution. Then the marginal distribution of* Z_i *is a Polya distribution*
with probability function

$$P[Z = z] = g_Z(z) = \binom{n}{z} \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)} \frac{\Gamma(\beta + n - z)}{\Gamma(\beta)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + n)}$$

for $z = 0, 1, 2, \dots, n$.

Why use a Beta to describe variation between samples?

- It has the desired support
- It is a very flexible distribution



- Last but not least the integral can be directly calculated.

Solutions

- Conjugate priors
- Laplace approximation
- Automatic differentiation
  - Speed calculations
  - More accrurate evaluation of derivatives

Overview

Section for dynamical systems and my profile

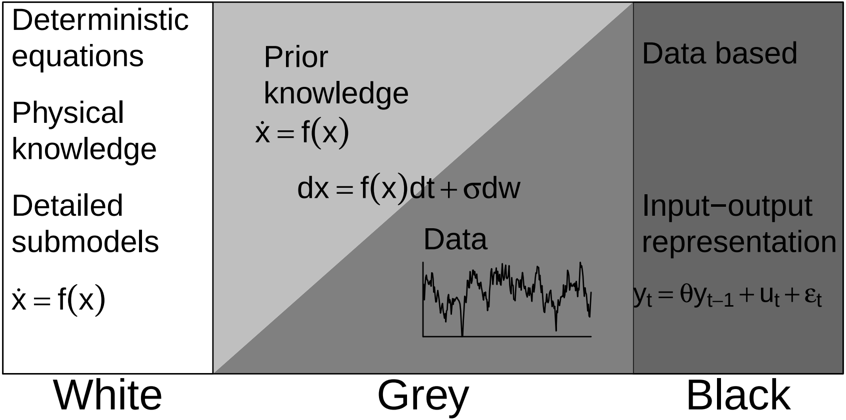
- About 30 PhD-students/post docs and permanent staff
- Work on dynamical systems (stochastic and deterministic), e.g.
  - ODE's, and PDE (deterministic)
  - Timeseries, and stochastic differential equations (stochastic)

Apart from teaching I work with time series and stochastic differential equations models, in different applications.
Some advanced level courses (on stochastics and statistics) thought by the section is:

- Time series analysis (02417), Advanced time series analasis (02427), Diffisions and stochastic differential equations (02425),
- Decision-Making Under Uncertainty (02435)
- Stochastic adaptive control (02421), Advanced system identification (02904)

The grey box modelling concept

- Combines prior physical knowledge with information in data.
- The model is not completely described by physical equations, but equations and the parameters are physically interpretable.



Grey box model for nutrient flow in an estuary

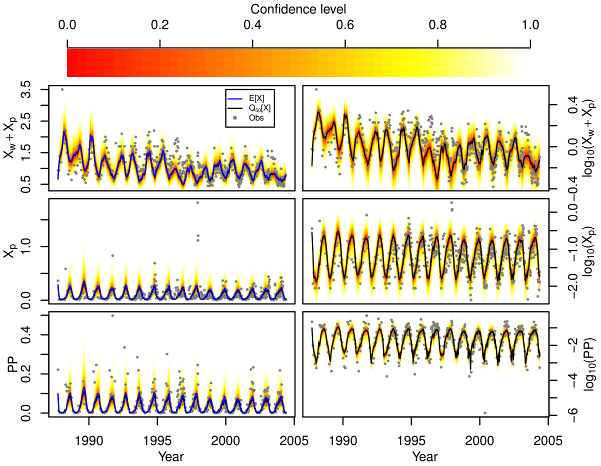
Continuous time stochastic state space formulation

$$d\begin{bmatrix} X_{w,t} \\ X_{p,t} \end{bmatrix} = \begin{bmatrix} N_{ex,t}Q_t \\ 0 \end{bmatrix} dt + \begin{bmatrix} -Q_t - a_{wp} - a_{wl} & a_{pw} \\ a_{wp} & -a_{wp} - Q_t \end{bmatrix} \begin{bmatrix} X_{w,t} \\ X_{p,t} \end{bmatrix} dt$$
$$+ \begin{bmatrix} \sigma_w X_{w,t} & 0 \\ 0 & \sigma_p X_{p,t} \end{bmatrix} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} d\boldsymbol{w}_t,$$

discrete time observation

$$\begin{bmatrix} \log(Y_{TN,k}) \\ \log(Y_{p,k}) \\ \log(Y_{pp,k}) \end{bmatrix} = \begin{bmatrix} \log(X_{w,t_k} + X_{p,t_k}) \\ \log(X_{p,t_k}) \\ \log(a_{wp} X_{w,t_k}) \end{bmatrix} + \begin{bmatrix} e_{TN,k} \\ e_{p,k} \\ e_{pp,k} \end{bmatrix},$$

Results



Confidence level
0.0
0.2
0.4
0.6
0.8
1.0
E[X]
Q_{50}[X]
Obs
$X_w + X_p$
X_p
PP
$\log_{10}(X_w + X_p)$
$\log_{10}(X_p)$
$\log_{10}$(PP)
Year
Year
1990
1995
2000
2005

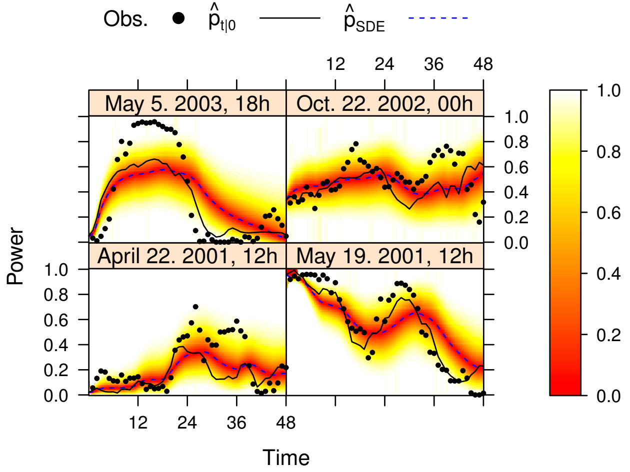
Wind power forecast

An increasing part of electricity supply is generated by wind

- Wind power cover about 27% of total system load
- Renewables should cover 50% in 2020 and 100% of total system load in 2035

With the large penetration of wind accurate forecasts (including uncertainties) are needed on all timescales, some metohods are

- Adaptive time series model, using MET-forecast (e.g. WPPT)
- Combining several MET-forecast
- Time-adaptive quantile regression (uncertainty)
- Stochastic differential equations
- and many more..



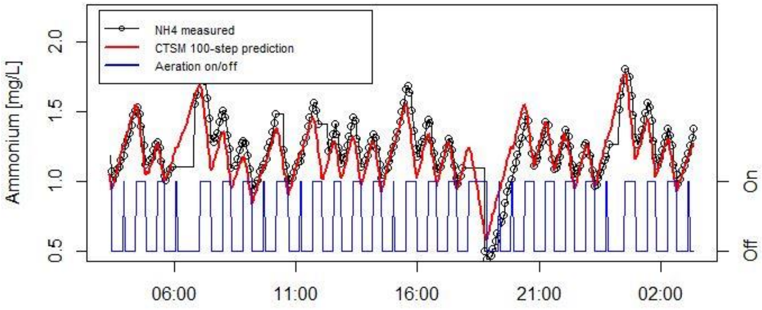
$$dx_{t,i} = -\theta \cdot (x_{t,i} - \hat{p}_{t|0,i} - c\hat{p}_{t|0,i}(1 - \hat{p}_{t|0,i})(1 - 2x_{t,i}))dt +$$
$$2\sqrt{\theta\alpha\hat{p}_{t|0,i}(1 - \hat{p}_{t|0,i})x_{t,i} \cdot (1 - x_{t,i})}dw_{t,i}$$

Waste water treatment - Peter A. Stentoft

- Waste water treatment account for 1-2% of total societal electricity comsumption
- Better control of aeration can substantially limit this comsumption

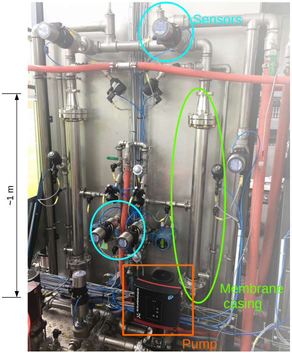
Waste water treatment

- Collaboration between DTU-COMPUTE, DTU- ENVIRONMENT, and Veolia Krüger
- Model Predictive Control for Nitrate using SDE's and optimization
- Devellopment of Phosphor models and data quality evaluations is still needed.



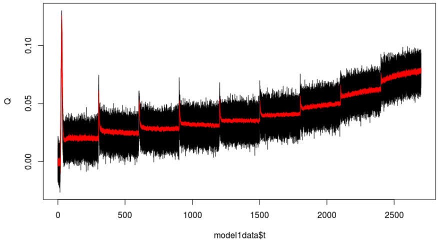
Membrane modelling - Goran Goranovic

Pilot plant AAU - E



$$Q_t = \frac{P_t}{R_0 + R_t} + \text{“}error\text{”}$$

$$dR_t = f(P_t, t, \tau_t)dt + \sigma R_t dw$$



Occupant bahaivor - Jon AR. Liisberg

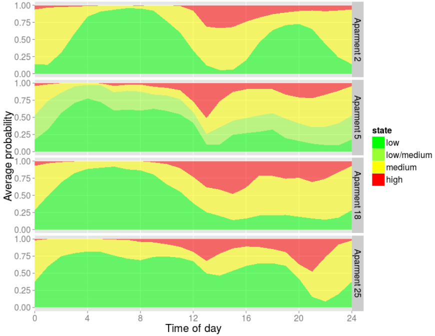
Metering and weather data from and nearby an apartment building in Catalonia, with 44 apartments. Also an occupant survey was available. Hourly observations from July 2012 to December 2013 consists of:

Variable	**description**
x_e	Electricity consumption in kWh
x_{sh}	Space heating in kWh
x_{hw}	Hot water consumption in kWh
x_w	Water consumption in litters
T_a	Ambient temperature in $^\circ C$
G	Solar radiation in W/m^2
W_s	Average wind speed in m/s
W_d	Average wind direction in $^\circ$
P	Precipitation in mm

Homogeneous HMM, distinct patterns

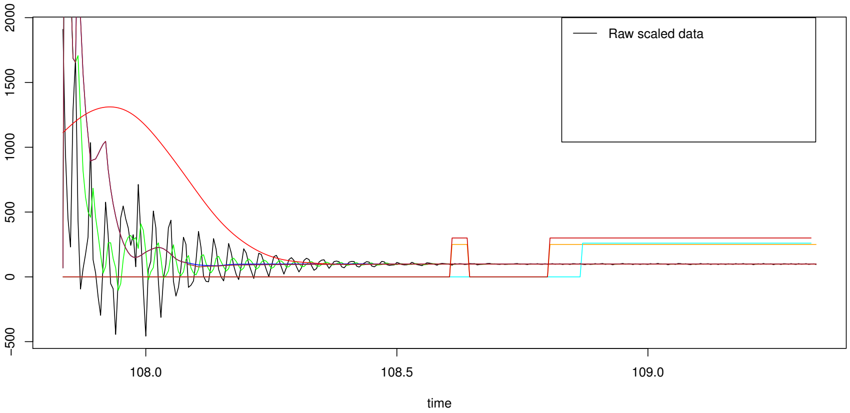
Comparing different patterns with occupant survey, indication of common factors were observed.

- no. residents
- Income (work, pension or subsidies)



Fast identification

Task: Based on data identify steady state before the the system settles



Other possible projects

Methodological work

- Fast evaluation of profile likelihood
- Bifurcation in stochastic differential equations
- Properties of quantile regression models

Overview

Evaluation

Any comments to the evaluation?

Overview