UiO : **Matematisk institutt**

Det matematisk-naturvitenskapelige fakultet

Introduction STK-4051/STK-9051 Computational Statistics Spring 2022

Instructor: Odd Kolbjørnsen, oddkol@math.uio.no

Course originally made by Geir Storvik

Odd Kolbjørnsen

Age: 50
Married

Two sons that work as actors in Oslo

- Data science Advisor
  @Lundin-Energy Lysaker (2014)
- Associate Professor
  Statistics and Data Science @UIO (2017)

- PhD in Mathematical-statistics
  from NTNU (2002)

- Worked 12 years for
  Norsk Regnesentral

Admission to the course

Students admitted at UiO must apply for courses in Stud . Students enrolled in other Master's Degree Programmes can, on app tted to the course if this is cleared by their own study prog

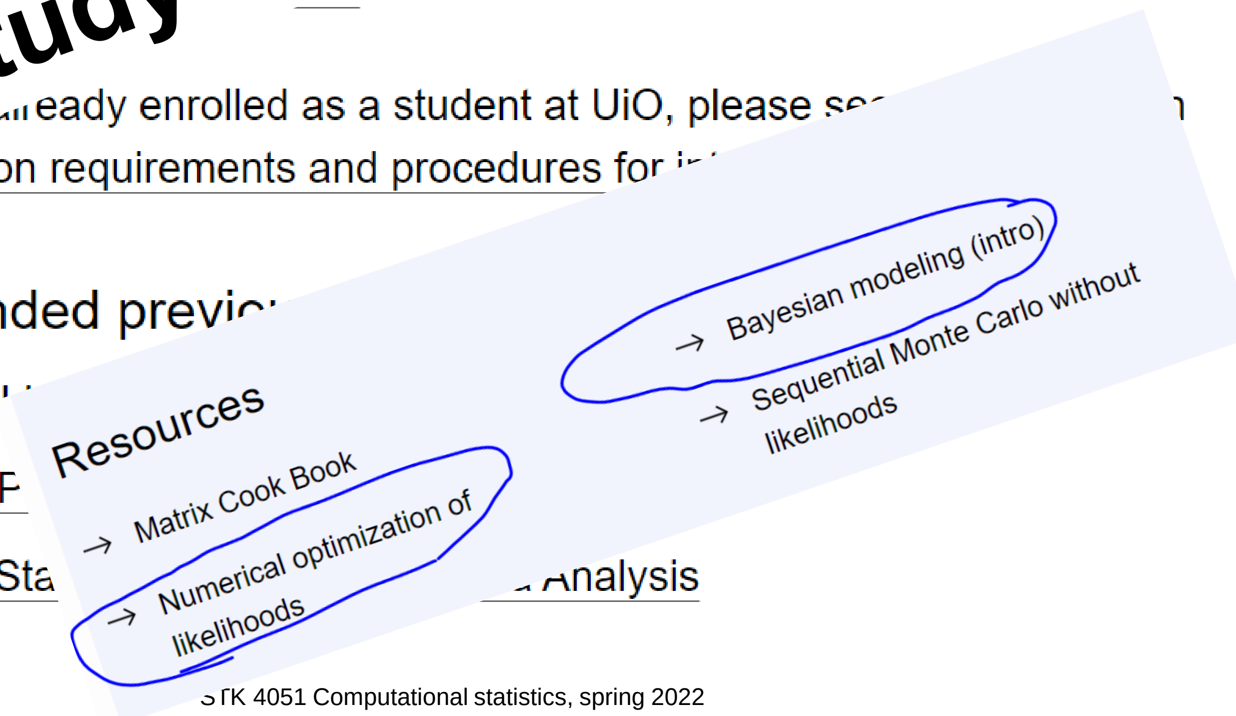
Nordic citizens and he Nordic countries may apply to take this course as

If you are eady enrolled as a student at UiO, please se about admission requirements and procedures for in

Study administration

Recommended previ

- MAT1120 –
- STK1100 – P
- STK1110 – Sta Analysis



Lectures and feedback

- Each week 14.15-17.00
- Ordering: Lecture x2 then exercise
- Physical lectures from next week
- I'll do «forelesninsopptak» and put it on the course webpage.
- I'll put «solution» to exercise online
- We will have a **reference group** of two students to give me feed back. **To be selected next week**.

STK 4051/9051 in one slide

- Optimization ~ Maximum likelihood
  - Continuous space (Newton-like methods, SGD,++)
  - Discrete/combinatorial
  - Missing/hidden variables (EM)
- Integration ~ Bayesian inference
  - Direct methods low dimensions
  - Variance reduction methods
  - Sequential Monte Carlo
  - Markov chain Monte Carlo
- Numerical methods within statistics

Maximum likelihood Theory

- For independent data:

$$L(\theta) = \prod_{i=1}^{n} f(\mathbf{x}_i|\theta)$$ Likelihood function

- Maximum likelihood estimate: $\hat{\theta}_{ML} = \arg\max_{\theta} L(\theta)$. Typically easier to work with the log-likelihood:

$$\ell(\theta) = \sum_{i=1}^{n} \log(f(\mathbf{x}_i;\theta))$$ Log-Likelihood function

- For smooth likelihoods, necessary requirement:

$\mathbf{s}(\theta) \equiv \ell'(\theta) = \mathbf{0}$, $|\theta|$ equations, called score (a vector)

$\mathbf{J}(\theta) \equiv -\ell''(\theta)$ positive (definite), called observed Fisher information

- Theory:
  - $E[\mathbf{s}(\theta)] = 0$
  - $\mathbf{I}(\theta) \equiv -E[\ell''(\theta)] = E[\mathbf{J}(\theta)] = \text{Var}[\mathbf{s}(\theta)]$, expected Fisher information
  - For large n (and some regularity assumptions)

$$\hat{\theta}_{ML} \approx N(\theta, \mathbf{I}^{-1}(\hat{\theta}_{ML})) \approx N(\theta, \mathbf{J}^{-1}(\hat{\theta}_{ML}))$$

Bayesian statistics

- Frequentistic approach: Data model (likelihood) $f(\mathbf{y}|\boldsymbol{\theta})$.
- Bayesian approach: Include prior information through a density $f(\boldsymbol{\theta})$.
- Prior: Describe our knowledge before data are collected
- Note: Treat θ as a random variable
- Bayes theorem:

$$f(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})}{f(\mathbf{y})}$$

$$f(\mathbf{y}) = \int_{\boldsymbol{\theta}} f(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})d\boldsymbol{\theta}$$

Note: In the book $f(\mathbf{y}|\boldsymbol{\theta}) = L(\boldsymbol{\theta}|\mathbf{y})$.

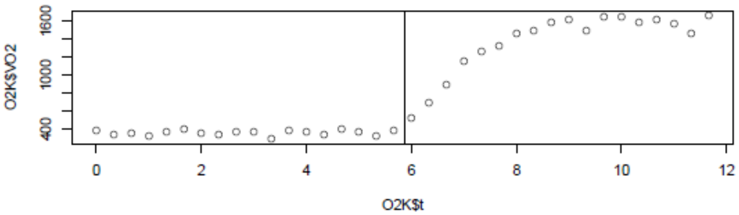
- Bayesian paradigm: All relevant information about $\boldsymbol{\theta}$ is contained in the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$
  - $\hat{\theta}_{post} = E[\boldsymbol{\theta}|\mathbf{y}] = \int_{\boldsymbol{\theta}} \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$
  - Credibility interval (one-dimensional): $\alpha = \Pr(a < \theta < b|\mathbf{y}) = \int_a^b \theta p(\theta|\mathbf{y})d\theta$
- Posterior: updated knowledge based on both prior and data
- Numerical aspect: Bayesian approach change optimization to integration

To day

- Introduction motivating examples
- (Chapter 1) + Chapter 2

Non linear regression

- Change of oxygen uptake ($VO2$) during exercise testing monitored in 2 distinct phases including a resting phase and the 6-minute exercise testing period (Baty et al., 2016)



- The length of the resting phase (λ) is controlled by the experimenter and does not need to be estimated
- 4-parameter regression model for the 2 phases of the $VO2$ kinetics:

$$VO2_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

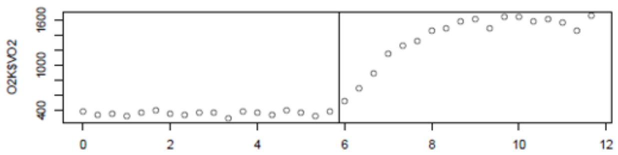
$$\mu_t = \begin{cases} \beta_0 & \text{if } t \leq \lambda \\ \beta_0 + (\beta_1 - \beta_0)(1 - e^{-(t-\lambda)/\tau_1}) & \text{if } \lambda < t \end{cases}$$

$$l(\boldsymbol{\theta}) = -\frac{1}{2}\log(2\pi) - \frac{1}{2\sigma^2}\log(\sigma^2) - \frac{1}{2}\sum_{t=1}^{T}(VO2_t - \mu_t)^2$$

- Aim: Maximum likelihood estimates of $\theta = [\,\beta_0, \beta_1, \tau, \sigma]$.

The parameter τ enters the model such that the problem is non linear

Solving the problem in R



- All parameters are positive, log-transformed first

```
data(O2K,package="nlstools")
tresh = 5.883
```

```
fit.vo2 = function(param,dat,lambda)
{
  tau1 = exp(param[1])
  beta0 = exp(param[2])
  beta1 = exp(param[3])
  fit = (dat$t <= lambda)*beta0 + (dat$t > lambda) *
    (beta0 + (beta1-beta0)*(1-exp(-(dat$t-lambda) / tau1)))
}
```

Mapping from constrained to unconstrained optimization

```
minusloglik = function(param,dat,tresh=5.883)
{
  fit = fit.vo2(param,dat,tresh)
  sig2 = exp(param[4])
  rss = sum((dat$VO2-fit)^2)
  n = ncol(dat)
  loglik = -0.5*n*log(2*pi)-0.5*n*log(sig2)-0.5*rss/sig2
  -loglik
}
```

Function to be optimized (where first input is the parameters to be optimized)

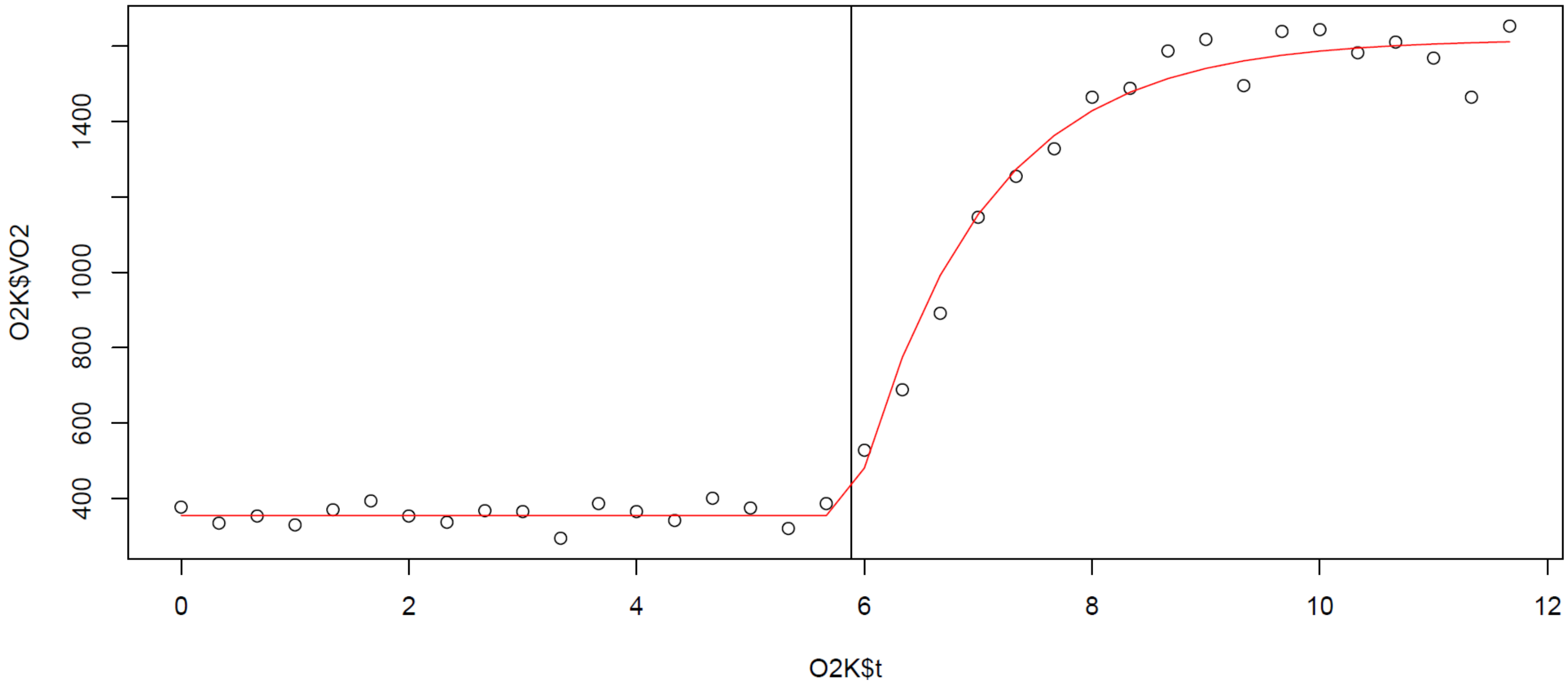
```
res = optim(c(1,log(400),log(1600),0),minusloglik,dat=O2K,tresh=tresh)
```

Initial guess of parameters to be estimated

Control variables (fixed during optimization)

Optimal fit

	Estimate	Initial
tau1	1.116745	2.73
beta0	354.732771	400
beta1	1618.926343	1600
sig2	37689.308635	1.00



Uncertainty (using ML)

- All parameters positive, log-transformed.
- The delta method
  - Assume $\hat{\theta} \approx N(\theta, \sigma_\theta^2)$
  - Then $g(\hat{\theta}) \approx N(g(\theta), \sigma_\theta^2[g'(\hat{\theta})]^2)$
  - Example: $\alpha = \exp(\theta)$,

$$\hat{\alpha} = \exp(\hat{\theta}) \approx N(\exp(\theta), \sigma_\theta^2 \exp(2\hat{\theta}))$$

Compare to nls()

```
------
Formula: VO2 ~ (t <= 5.883) * VO2rest + (t > 5.883) * (VO2rest + (VO2peak -
    VO2rest) * (1 - exp(-(t - 5.883)/mu)))

Parameters:
        Estimate Std. Error t value Pr(>|t|)
VO2rest 3.568e+02  1.141e+01   31.26   <2e-16 ***
VO2peak 1.631e+03  2.149e+01   75.88   <2e-16 ***
mu      1.186e+00  7.661e-02   15.48   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 49.59 on 33 degrees of freedom
```

```
res = optim(c(1,log(400),log(1600),0),minusloglik,dat=O2K,tresh=tresh,
            hessian=TRUE)
hat.param = exp(res$par)
var.theta = diag(solve(res$hessian))
var.param = var.theta*exp(2*res$par)
M = cbind(hat.param,sqrt(var.param))
rownames(M) = c("tau1","beta0","beta1","sig2")
colnames(M) = c("Estimate","Std.Err")
show(M)
          Estimate       Std.Err
tau1      1.116745  2.771562e-01
beta0   354.732771  4.457839e+01
beta1  1618.926343  7.938112e+01
sig2  37689.308635  3.690964e+04
```

Lasso

- Ordinary linear regression

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), i = 1, \ldots, n$$

  Least squares/ML problematic when p large compared to n.

- Lasso: Minimize wrt $\boldsymbol{\beta}$:

$$g(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

- Optimization problem in $p + 1$ dimensions.

Lasso on Diabetes data

- Blood and other measurements in diabetics
- $n = 442, p = 10$
- Direct use of `optim` in R (100 000 iterations):
  - $\boldsymbol{\beta}^{(0)} = \mathbf{0}$: $g(\hat{\boldsymbol{\beta}}_{\text{lasso}}) = 1\,771\,743$
  - $\boldsymbol{\beta}^{(0)} = \hat{\boldsymbol{\beta}}_{LS}$: $g(\hat{\boldsymbol{\beta}}_{\text{lasso}}) = 1\,278\,314$
  - Dangerous to use gradient-based methods since derivative is non-continuous!
- Including interactions: $p = 64$
- Direct use of `optim` in R (100 000 iterations):
  - $\boldsymbol{\beta}^{(0)} = \mathbf{0}$: $g(\hat{\boldsymbol{\beta}}_{\text{lasso}}) = 2\,943\,334$
  - $\boldsymbol{\beta}^{(0)} = \hat{\boldsymbol{\beta}}_{LS}$: $g(\hat{\boldsymbol{\beta}}_{\text{lasso}}) = 1\,119\,003$
- Selecting right method for the problem is an art
- Better methods exists!

ADMM will be discussed for solving Lasso

Model selection

- Genetic association studies: Which genes influence a certain phenotype (presence of cancer, size, etc)
- Linear model including all possible variables:

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i$$

- Reasonable to assume that some x_{ij}'s do not influence the response, modification:

$$Y_i = \beta_0 + \sum_{j=1}^{p} \gamma_j \beta_j x_{ij} + \varepsilon_i$$

  where $\gamma_j \in \{0, 1\}$.
- 2^p possible models, how to choose the best one?
  - $p = 20, 2^p = 1\,048\,576$, $p = 100, 2^p = 1.267651 * 10^{30}$
- Combinatorial problem, discrete optimization

Censored Data

- Motorette data of Schmee and Hahn (1979)
- Time to failure at different temperatures

150°	170°	190°	220°
8064*	1764	408	408
8064*	2772	408	408
8064*	3444	1344	504
8064*	3542	1344	504
8064*	3780	1440	504
8064*	4860	1680*	528*
8064*	5196	1680*	528*
8064*	5448*	1680*	528*
8064*	5448*	1680*	528*
8064*	5448*	1680*	528*

- Observations with * are censored, no failure when monitoring stopped.
- Possible model:

$$t_i = \beta_0 + \beta_1 v_i + \varepsilon_i, \quad \varepsilon_i \sim (0, \sigma^2)$$

  with $v_i = 1000/(\text{temperature} + 273.2)$ and $t_i = \log_{10}(i\text{th failure time})$
- Problem: Density for censored data not directly available:

$$f(t_i^*; \boldsymbol{\theta}) = \Pr(T \geq t^*; \boldsymbol{\theta}) = \int_{t^*}^{T} \phi(t; \beta_0 + \beta_1 v_i, \sigma) dt$$

- Need combination of optimization and integration

Hierarchical model

- Salmonella experiment (Breslow, 1984)
- Three plates have been processed at each dose i of quinoline
- The number of revertant colonies of TA98 Salmonella measured:

Dose of quinoline (μg per plate)					
0	10	33	100	333	1000
15	16	16	27	33	20
21	18	26	41	38	27
29	21	33	60	41	42

- Possible model:

$$\begin{aligned} y_{ij} \sim& \text{Poisson}(\mu_{ij}) \\ \log(\mu_{ij}) =& \alpha + \beta \log(x_i + 10) + \gamma x_i + \lambda_{ij} \\ \lambda_{ij} \sim& \text{Normal}(0, \sigma^2) \end{aligned}$$

- With λ_{ij} fixed this is an ordinary Poisson regression model (within the family of generalized linear models)
- With λ_{ij} random the model becomes a bit more complicated.
- Likelihood for data given $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \sigma^2)$ not analytically available:

$$f(y_{ij}; \boldsymbol{\theta}) = \int_0^\infty f(y_{ij}|\lambda_{ij}; \boldsymbol{\theta}) f(\lambda_{ij}; \boldsymbol{\theta}) d\lambda_{ij}$$

- Maximum likelihood requires combination of integration and optimization

Hierarchical Bayesian model

- Salmonella experiment much used to illustrate Bayesian approach
- Extension of model:

$$\begin{aligned}
y_{ij} \sim & \text{Poisson}(\mu_{ij}) \\
\log(\mu_{ij}) = & \alpha + \beta\log(x_i + 10) + \gamma x_i + \lambda_{ij} \\
\lambda_{ij} \sim & \text{Normal}(0, \sigma^2) \\
\alpha, \beta, \gamma \overset{ind}{\sim} & N(0, 1000) \\
\sigma^2 \sim & \text{InverseGamma}(0.001, 0.001)
\end{aligned}$$

- Parameters $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \sigma^2)$ are now given a stochastic model
- Such models can be used to include prior information about parameters.
- Aim now:

$$f(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})}{f(\mathbf{y})}$$

$$f(\mathbf{y}|\boldsymbol{\theta}) = \int_{\boldsymbol{\lambda}} f(\mathbf{y}|\boldsymbol{\lambda}, \boldsymbol{\theta}) f(\boldsymbol{\lambda}|\boldsymbol{\theta}) d\boldsymbol{\lambda} \qquad \text{18 dimensional integral}$$

$$f(\mathbf{y}) = \int_{\boldsymbol{\theta}} f(\mathbf{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

Bayesian statistics

- Frequentistic approach: Data model (likelihood) $f(\mathbf{y}|\theta)$.
- Bayesian approach: Include prior information through a density $f(\theta)$.
- Prior: Describe our knowledge before data are collected
- Note: Treat θ as a random variable
- Bayes theorem:

$$f(\theta|\mathbf{y}) = \frac{f(\theta)f(\mathbf{y}|\theta)}{f(\mathbf{y})}$$

$$f(\mathbf{y}) = \int_{\theta} f(\theta)f(\mathbf{y}|\theta)d\theta$$

  Note: In the book $f(\mathbf{y}|\theta) = L(\theta|\mathbf{y})$.
- Bayesian paradigm: All relevant information about θ is contained in the posterior distribution $p(\theta|\mathbf{y})$
  - $\hat{\theta}_{post} = E[\theta|\mathbf{y}] = \int_{\theta} \theta p(\theta|\mathbf{y})d\theta$
  - Credibility interval (one-dimensional): $\alpha = \Pr(a < \theta < b|\mathbf{y}) = \int_a^b \theta p(\theta|\mathbf{y})d\theta$
- Posterior: updated knowledge based on both prior and data
- Numerical aspect: Bayesian approach change optimization to integration

Tracking possition using GPS measurements

- Simplified model

$$y_{t,i} = \sqrt{(v_t^x - s_{t,i}^x)^2 + (v_t^y - s_{t,i}^y)^2 + (v_t^z - s_{t,i}^z)^2} + \varepsilon_{t,i}, \quad i = 1, 2, \ldots, n_t$$

  with $\{\varepsilon_{t,i}\}$ independent noise terms.
- Assume available model for movement: $p(\mathbf{v}_t|\mathbf{v}_{t-1})$.
- Aim:

$$p(\mathbf{v}_t|\mathbf{y}_{1:t}) = \int_{\mathbf{v}_{1:t-1}} p(\mathbf{v}_{1:t}|\mathbf{y}_{1:t})d\mathbf{v}_{1:t-1} = \int_{\mathbf{v}_{1:t-1}} \frac{p(\mathbf{v}_{1:t})p(\mathbf{y}_{1:t}|\mathbf{v}_{1:t})}{p(\mathbf{y}_{1:t})}d\mathbf{v}_{1:t-1}$$

$$p(\mathbf{v}_{1:t}) = p(\mathbf{v}_1)\prod_{s=2}^{t} p(\mathbf{v}_s|\mathbf{v}_{s-1})$$

$$p(\mathbf{y}_{1:t}) = \int_{\mathbf{v}_{1:t}} p(\mathbf{v}_{1:t})p(\mathbf{y}_{1:t}|\mathbf{v}_{1:t})d\mathbf{v}_{1:t}$$

Model dynamics

- Linear dynamics

$$\begin{aligned}\mathbf{v}_t =&(v_t^x, v_t^y, v_t^z, \dot{v}_t^x, \dot{v}_t^y, \dot{v}_t^z)^T\\ =&\mathbf{\Phi}\mathbf{v}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \sigma_Q^2\mathbf{Q})\end{aligned}$$

where

$$\mathbf{\Phi} = \begin{pmatrix}\mathbf{I}_3 & \mathbf{I}_3\\ \mathbf{0} & \mathbf{I}_3\end{pmatrix}, \qquad \mathbf{Q} = \begin{pmatrix}\frac{q_v^2 D_t^3}{3}\mathbf{I}_3 & \frac{q_{cd}^2 D_t}{2}\mathbf{I}_3\\ \frac{q_{cd}^2 D_t}{2}\mathbf{I}_3 & q_{cb}^2 D_t\mathbf{I}_3\end{pmatrix}$$

- Combined model

$$\begin{aligned}\mathbf{v}_t =&\mathbf{\Phi}\mathbf{v}_{t-1} + \boldsymbol{\eta}_t,\\ y_{t,i} =&\sqrt{(v_t^x - s_{t,i}^x)^2 + (v_t^y - s_{t,i}^y)^2 + (v_t^z - s_{t,i}^z)^2} + \varepsilon_{t,i}, \quad i = 1, 2, \dots, n_t\end{aligned}$$

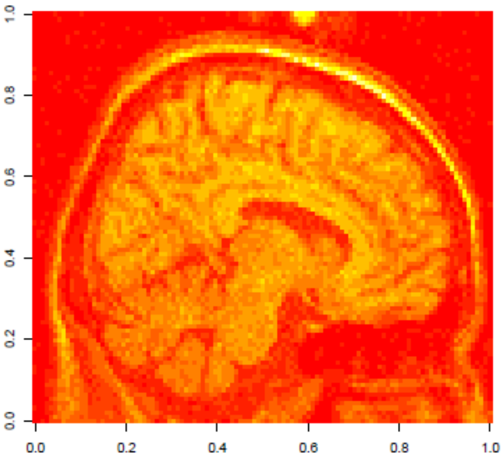
example of a state space model

- Challenge: Compute $p(\mathbf{v}_t|\mathbf{y}_{1:t})$ for each t in real time
- Need to utilize computation performed on previous time step

Image analysis

- MRI tissue classification problem
- Three major tissue classes (cerebrospinal fluid (CSF), gray matter (GM), white matter (WM))
- Intensities assumed normally distributed with class-dependent means and variances:

$$y_{ij}|C_{ij} = k \sim N(\mu_k, \sigma_k^2)$$



-
- Bayes formula ($\pi_k = \Pr(C_{ij} = k)$):

$$\Pr(C_{ij} = k|y_{ij}) = \frac{\pi_k f(y_{ij}|C_{ij} = k)}{\sum_{l=1}^{3} \pi_l f(y_{ij}|C_{ij} = l)}$$

- Easy to calculate individually for each square (pixel)

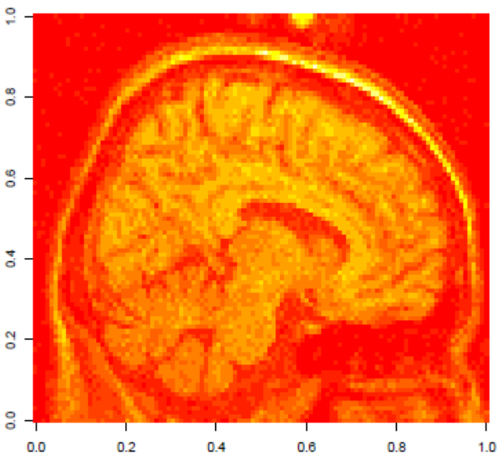
Image analysis spatial structure

- Expect some smoothness in class-structure
- Markov Random field/Potts model:

$$\begin{aligned}\Pr(\mathbf{C}) =& \Pr(C_{11}, \ldots., C_{n_1 n_2})\\ =& \frac{1}{Z} e^{-\beta \sum_{||(i,j)-(i'j')||=1} I(C_{ij} \neq C_{i'j'})}\end{aligned}$$

- Now interested in

$$\Pr(\mathbf{C}|\mathbf{y}) = \frac{\Pr(\mathbf{C}) \prod_{ij} f(y_{ij}|C_{ij})}{\sum_{\mathbf{C}'} \Pr(\mathbf{C}') \prod_{ij} f(y_{ij}|C'_{ij})}$$



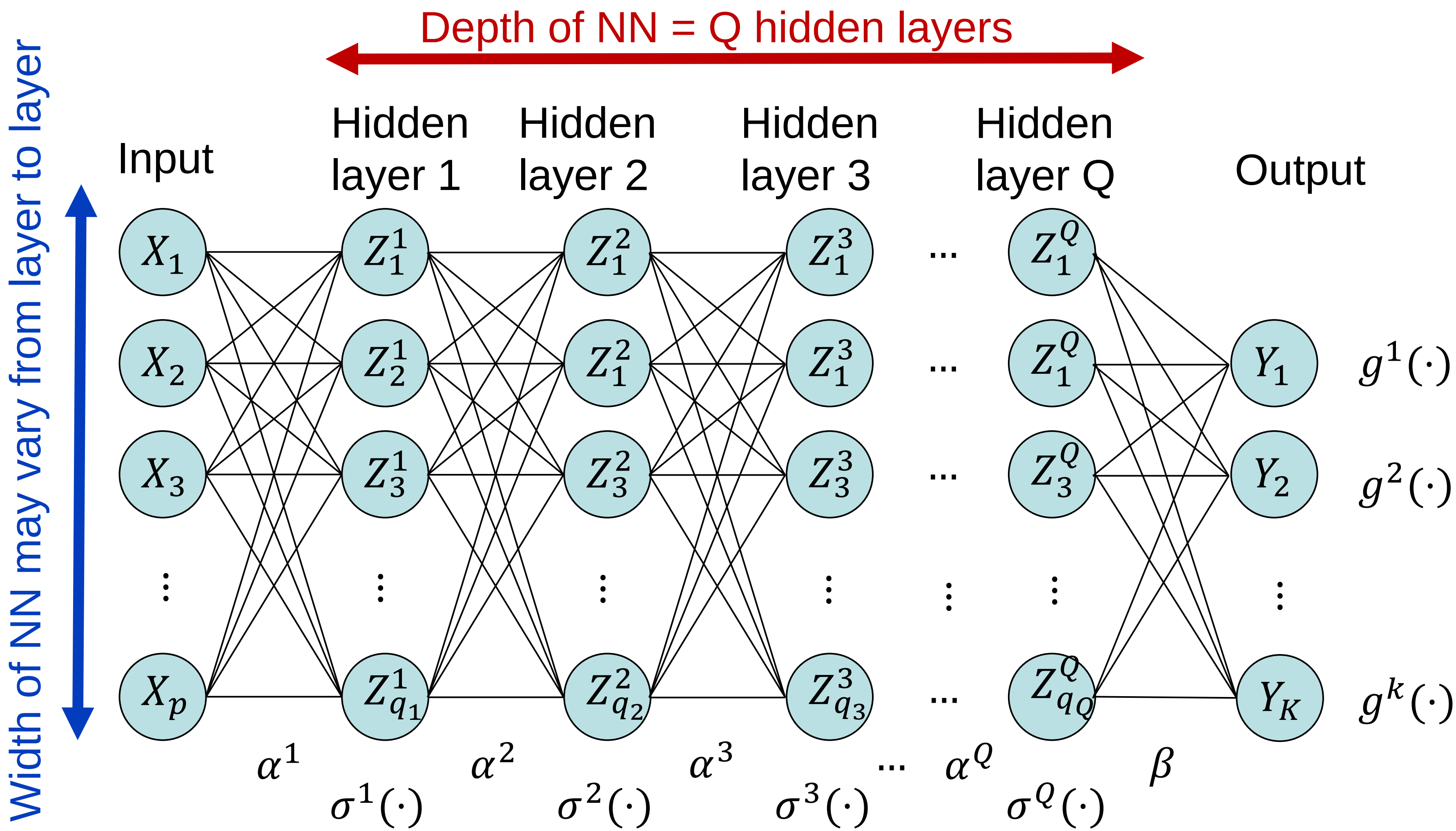
-
- The sum in the denominator contains K^n terms,
  - $K =$ number of class
  - $n =$ number of pixels.
- Discrete type of "integration"

Machine Learning

- Large data sets
- Complex models
- Algorithms
- Sparse coding
- Deep neural nets

- Stochastic gradient

Deep neural network



Numerical challenges in statistics

- Optimization
  - Maximum likelihood
  - Penalized likelihood (Lasso, Ridge, Regularization)
  - Model selection (Discrete)
  - Classification (Discrete)
- Integration
  - Missing / censored data
  - Bayesian approaches

Optimization and Integration

- Aim: Maximize $g(\theta)$
- Newton's method

$$\theta^{(t+1)} = \theta^t - \frac{g'(\theta^t)}{g''(\theta^t)}$$

- Gradient descent:

$$\theta^{(t+1)} = \theta^t + \alpha g'(\theta^t)$$

- Stochastic gradient descent:

$$\theta^{(t+1)} = \theta^t + \alpha_t \widehat{g'(\theta^t)}$$

$\widehat{g'(\theta^t)}$ estimate of $g'(\theta^t)$.

- Assume we are interested in $I = \int h(\mathbf{x})d\mathbf{x}$.
- Select a density $f(\mathbf{x})$ and write

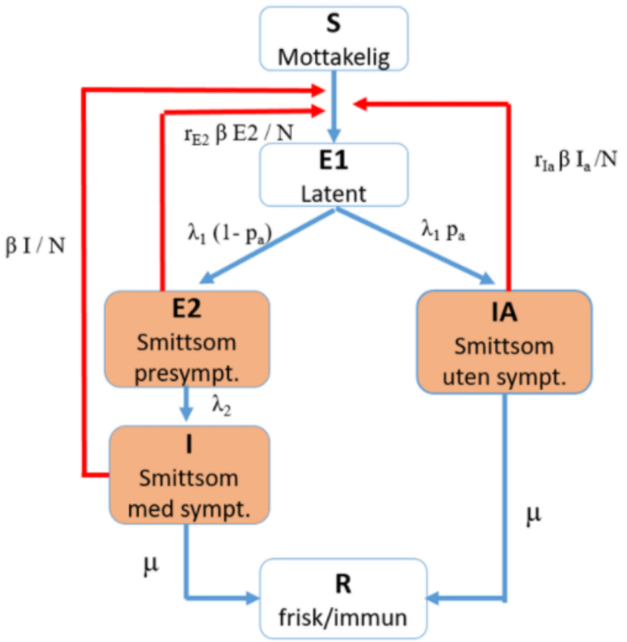
$$\begin{aligned} I =& \int \frac{h(\mathbf{x})}{f(\mathbf{x})} f(\mathbf{x}) d\mathbf{x} \\ =& E\left[\frac{h(\mathbf{X})}{f(\mathbf{X})}\right], \quad \text{where } \mathbf{X} \sim f(\mathbf{x}) \end{aligned}$$

- Assume $\mathbf{x}_1, ..., \mathbf{x}_n \sim f(\mathbf{x})$:

$$\widehat{I} = \frac{1}{n}\sum_{i=1}^n \frac{h(\mathbf{x}_i)}{f(\mathbf{x}_i)}$$

- This is Monte Carlo integration
- Can in principle work in any dimension ($\text{Var}(\widehat{I}) \propto n^{-1}$)
- In the course:
  - Simple Monte Carlo in low dimensions (chapter 6)
  - Markov chain Monte Carlo in high dimensions (chapters 7 and 8)
  - Sequential Monte Carlo for state-space models (sec 6.3)

Covid-19 outbreake



As we speak people around the world are making statistical models and designing algorithms to predict effect of government policies on the Covid-19 outbreak

Solveig Engebretsen
Reference paper for FHI
Work at NR
Guest lecture last year

https://www.fhi.no/sv/smittsomme-sykdommer/corona/koronavirus-modellering/

Syllabus -requirements

- Main textbook: Givens and Hoeting (2012)
  - Chapter 1 - Background Will only be referred to when needed
  - Chapter 2 - Optimization General methods, will briefly be discussed
  - Chapter 3 - Combinatorial optimization
  - Chapter 4 - The EM algorithm
  - Chapter 5 - Numerical integration General methods, will briefly be discussed
  - Chapter 6 - Monte Carlo methods (additional note on SMC)
  - Chapter 7 - Markov Chain Monte Carlo
  - Chapter 8 - Advanced topics in MCMC Selected sections
  - Chapter 9 - Bootstrap - Section 9.1 & Section 9.2
- Some additional material
  - Alternating directions methods of moments (ADMM)
  - Stochastic gradient method
  - Variational inference

Course structure

- Focus on methods
- Focus on implementing algorithms
  - Will mainly use R, not that efficient
  - For most methods, there exist efficient software
  - Focus on learning through implementation
- Some theory on why and how methods work
- Examination:
  - Project and written exam (or exam at home)
    - Dates will be published on the course page
  - Project divided into two parts
  - No grades given separately on projects, but feedback will be given

How to work (my reccomendation)

- Before lecture
  - Read book / note
- Attend lecture
- After lecture
  - Read book / note [if you did not do it before]
  - go through R-code example
  - do exercises
- Attend exercise
- After exercise
  - do exercises [if you did not d
    wrt code, go through R-code
    make sure that you understand
- Always possible
  - Send mail with questions to me
  - Talk to me (when we meet in re

Big payoff when doing the
compulsory exercise
(and for life)

You get reply
by mail or in class (most often)