NTNU
Norwegian University of
Science and Technology

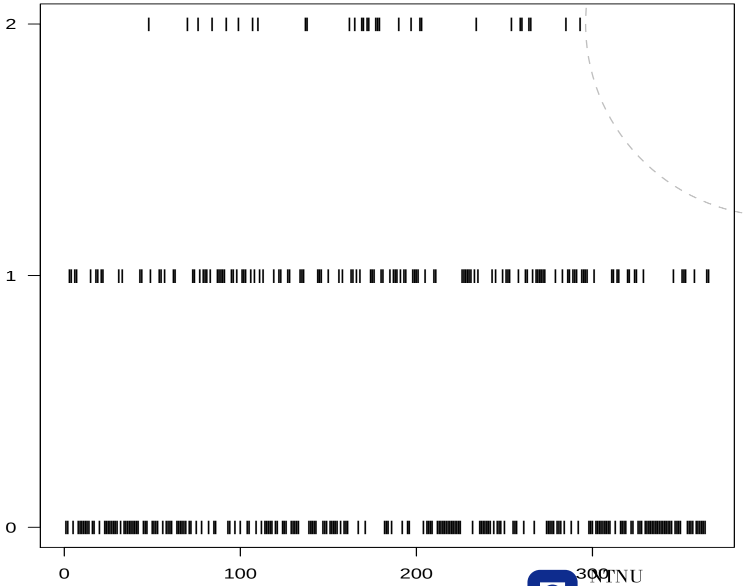
INLA - Introduction

Elias T. Krainski

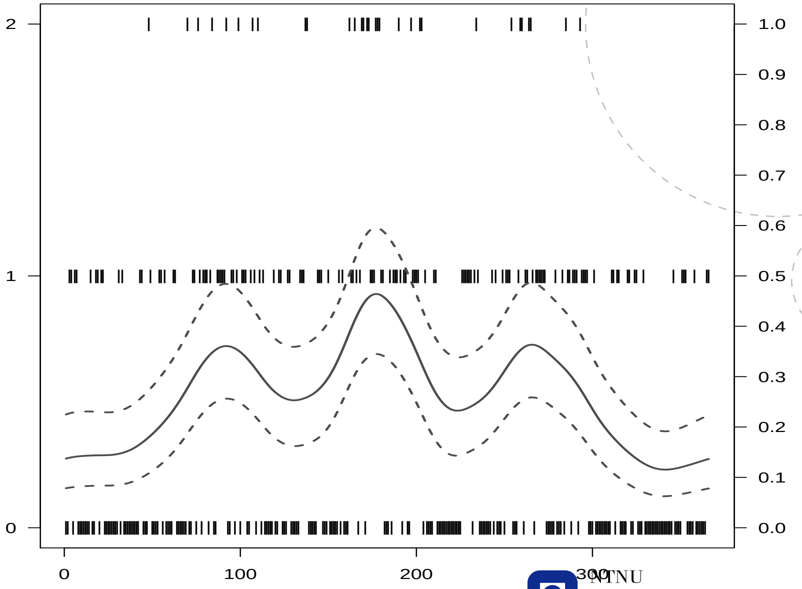
Outline

NTNU
Norwegian University of Science and Technology

Number of raining days in Tokyo, for each yearly day in two years



Number of raining days in Tokyo, for each yearly day in two years



NTNU
Norwegian University of Science and Technology

A model for Tokyo data

Observation model

$$y_i \sim \text{Binomial}(n_i, p_i),$$

for $i = 1, 2, ..., 366$

$$n_i = \begin{cases} 1, & \text{for 29 February} \\ 2, & \text{other days} \end{cases} \tag{1}$$

$$y_i \in \begin{cases} \{0, 1\}, & \text{for 29 February} \\ \{0, 1, 2\}, & \text{other days} \end{cases} \tag{2}$$

$$p_i = \frac{1}{1 + exp(-x_i)}$$

probability on day i depends on x_i

NTNU
Norwegian University of
Science and Technology

Smoothing $\boldsymbol{x}$

— Let

$$x_i|\boldsymbol{x}_{-i} \sim N(\overline{x}_i, \frac{\sigma^2}{2})$$

where

$$\overline{x}_i = \begin{cases} \frac{x_2+x_{366}}{2} & \text{if} \quad i = 1 \\ \frac{x_{i-1}+x_{i+1}}{2} & \text{if} \quad 1 < i < 366 \\ \frac{x_{365}+x_1}{2} & \text{if} \quad i = 366 \end{cases} . \quad (3)$$

and $\theta = 1/\sigma^2$

— θ is controls the variation of $\boldsymbol{x}$

- so, related to variantion of p_i

— as $\theta > 0$: people usually use $\pi(\theta) \sim$ Gamma(a, b)

Outline

NTNU
Norwegian University of
Science and Technology

Bayesian hierarchical model

- $\boldsymbol{y}$: observed response data
- θ_1: likelihood parameter(s)

$$\boldsymbol{y}|\boldsymbol{x},\theta_1 \sim \pi(\boldsymbol{y}|\boldsymbol{x},\theta_1) = \prod_{i=1}^{n} \pi(y_i|x,\boldsymbol{\theta_1}) \qquad \text{(ind. cond.)}$$

NTNU
Norwegian University of
Science and Technology

Bayesian hierarchical model

- $\boldsymbol{y}$: observed response data
- θ_1: likelihood parameter(s)

$$\boldsymbol{y}|\boldsymbol{x},\theta_1 \sim \pi(\boldsymbol{y}|\boldsymbol{x},\theta_1) = \prod_{i=1}^{n} \pi(y_i|x,\boldsymbol{\theta_1}) \qquad \text{(ind. cond.)}$$

- $\boldsymbol{x}$: latent/unobserved field
  - Gaussian $\rightarrow$ to use INLA
- θ_2: latent field parameter(s)

$$\boldsymbol{x}|\theta_2 \sim \pi(\boldsymbol{x}|\theta_2) = N(\boldsymbol{0}, \boldsymbol{Q}(\theta_2)^{-1})$$

NTNU
Norwegian University of
Science and Technology

Bayesian hierarchical model

— $\boldsymbol{y}$: observed response data

— θ_1: likelihood parameter(s)

$$\boldsymbol{y}|\boldsymbol{x},\theta_1 \sim \pi(\boldsymbol{y}|\boldsymbol{x},\theta_1) = \prod_{i=1}^{n} \pi(y_i|x,\boldsymbol{\theta_1}) \qquad \text{(ind. cond.)}$$

— $\boldsymbol{x}$: latent/unobserved field

- Gaussian $\rightarrow$ to use INLA

— θ_2: latent field parameter(s)

$$\boldsymbol{x}|\theta_2 \sim \pi(\boldsymbol{x}|\theta_2) = N(\boldsymbol{0}, \boldsymbol{Q}(\theta_2)^{-1})$$

— in short: $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$ (hyperparameter)

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}) \rightarrow \text{to be Bayesian}$$

NTNU
Norwegian University of Science and Technology

$\pi(\boldsymbol{y}|\boldsymbol{x},\theta)$: likelihood

Depends on

— which kind of data values we have

- binary (yes/no response, binary image)
- counts (people infected with a disease in each area)
- continuous - or + (stock return, temperature)
- continuous + (rainfall amount, fish weight)
- survival (recovery time, time to death)

NTNU
Norwegian University of
Science and Technology

$\pi(\boldsymbol{y}|\boldsymbol{x},\theta)$: likelihood

Depends on

— which kind of data values we have

- binary (yes/no response, binary image)
- counts (people infected with a disease in each area)
- continuous - or + (stock return, temperature)
- continuous + (rainfall amount, fish weight)
- survival (recovery time, time to death)

— the way it is collected

- **usual**: each observational unit gives one value
- each observational unit gives more than one value
- point process: locations (time, spatial) of events

$\pi(\boldsymbol{x}|\boldsymbol{Q}(\theta))$: The latent field prior

— It is

- unobserved
- called Gaussian latent random field
- the most important ingredient in INLA

$\pi(\boldsymbol{x}|\boldsymbol{Q}(\theta))$: The latent field prior

— It is

- unobserved
- called Gaussian latent random field
- the most important ingredient in INLA

— it represents

- covariate coefficients
- unobserved effects

$\pi(\boldsymbol{x}|\boldsymbol{Q}(\theta))$: The latent field prior

- It is
  - unobserved
  - called Gaussian latent random field
  - the most important ingredient in INLA
- it represents
  - covariate coefficients
  - unobserved effects
- it can be
  - unstructured (Tokyo: p_i doesn't depends p_j)
  - structured (Tokyo: p_i depends on neighbour days)
  - more than one (structured(s) + unstructured(s) + covariate(s))

NTNU
Norwegian University of
Science and Technology

$\pi(\theta)$: The θ prior

- parameters from the likelihood and $\boldsymbol{x}$ distribution
- examples (likelihood):
  - precision parameter of the Gaussian or Gamma
  - dispersion parameter in Beta, negative binomial
  - zero-inflation probability

$\pi(\theta)$: The θ prior

- parameters from the likelihood and $\boldsymbol{x}$ distribution
- examples (likelihood):
  - precision parameter of the Gaussian or Gamma
  - dispersion parameter in Beta, negative binomial
  - zero-inflation probability
- examples (latent field)
  - precision parameter in, usually, all of those
  - correlation parameter (in some for time series modelling)
  - range parameter (in some for spatial data modelling)

NTNU
Norwegian University of
Science and Technology

Outline

Tokyo example

Hierarchical model

On the Tokyo model

Heart model example

Bayesian inference

INLA overview

NTNU
Norwegian University of
Science and Technology

A model for Tokyo data

Observation model

$$y_i \sim \text{Binomial}(n_i, p_i)$$

$$p_i = \frac{1}{1 + \exp(-x_i)}$$

the likelihood has no θ

$$\pi(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{366} \pi(y_i|x_i)$$

NTNU
Norwegian University of
Science and Technology

Latent model

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}) \propto \exp\left\{-\frac{\theta}{2}\left[(x_1 - x_{366})^2 + \sum_{i=2}^{366}(x_i - x_{i-1})^2\right]\right\} \quad (4)$$

$$= \exp\left\{-\frac{\theta}{2}\boldsymbol{x}^T\boldsymbol{R}\boldsymbol{x}\right\} \quad (5)$$

Latent model

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}) \propto \exp\left\{-\frac{\theta}{2}\left[(x_1 - x_{366})^2 + \sum_{i=2}^{366}(x_i - x_{i-1})^2\right]\right\} \quad (4)$$

$$= \exp\left\{-\frac{\theta}{2}\boldsymbol{x}^T\boldsymbol{R}\boldsymbol{x}\right\} \quad (5)$$

where $\boldsymbol{R} = \begin{pmatrix} 2 & -1 & & & & & -1 \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & & \ddots & & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ -1 & & & & & -1 & 2 \end{pmatrix}$

$\boldsymbol{Q}(\theta) = \theta\boldsymbol{R}$

NTNU
Norwegian University of
Science and Technology

Latent model **warning**

$$\exp\left\{-\frac{\theta}{2}\left[(x_1 - x_{366})^2 + \sum_{i=2}^{366}(x_i - x_{i-1})^2\right]\right\} \quad (6)$$

$$(7)$$

intrinsic/improper

$$\begin{aligned} x_i = 20, \quad x_{i-1} = 10 \quad &\rightarrow \quad x_i - x_{i-1} = 10 \\ x_i = 10020, \quad x_{i-1} = 10010 \quad &\rightarrow \quad x_i - x_{i-1} = 10 \end{aligned}$$

constraint or take the intercept out

NTNU
Norwegian University of
Science and Technology

$\pi(\theta)$ problem

- Tokyo example: $Q(\boldsymbol{\theta}) = \theta \boldsymbol{R}$
  - bigger θ less variation of $\boldsymbol{x}$
    - related to the variation of p_i
- $\theta > 0$: people usually use $\theta \sim \text{Gamma}(a, b)$
- improper distribution: θ values depends on $\boldsymbol{R}$
  - hard to interpret θ (a=?????, b=?????)

NTNU
Norwegian University of
Science and Technology

$\pi(\boldsymbol{x}|\theta = 1)$ and n

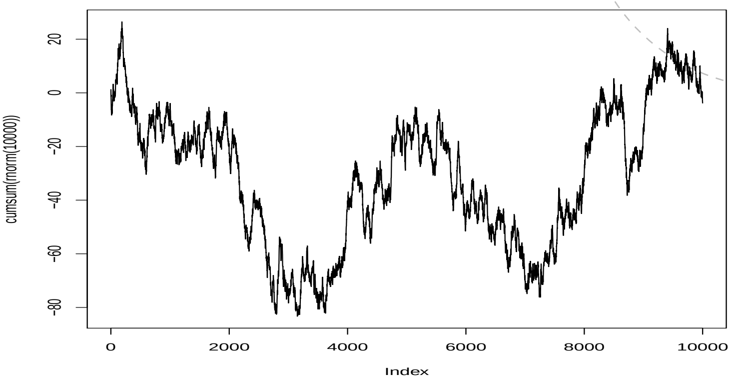
The marginal variance and n relation

```
rw.var <- function(n, order) {
    R <- as.matrix(INLA::inla.rw(n, order=order))
    mean(diag(INLA::inla.ginv(R, rankdef=order)))
}
n <- c(10, 100, 366, 1000); names(n) <- n
rbind(rw1=sapply(n, rw.var, order=1),
      rw2=sapply(n, rw.var, order=2))

##       10      100         366         1000
## rw1 1.65   16.665    60.99954     166.6665
## rw2 2.40 2381.190 116733.95702 2380955.1304
```

NTNU
Norwegian University of
Science and Technology

$\pi(\boldsymbol{x}|\theta = 1)$: one realization



We need to control the marginal variance!

NTNU
Norwegian University of
Science and Technology

$\pi(\theta)$ solution

1. scale the model $\rightarrow$ easy to interpret θ
   - Tutorial on `scale.option` at `www.r-inla.org/`

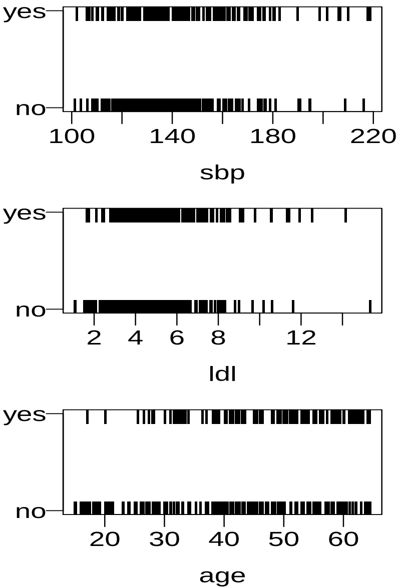
$\pi(\theta)$ solution

1. scale the model $\rightarrow$ easy to interpret θ
   - Tutorial on `scale.option` at `www.r-inla.org/`
2. AND (new idea) Penalized complexity prior
   - P0: basic model: $p_i = p_0$
   - P1: complex model: p_i varies
   - Kullback-Leibler divergence (KLD)
     - a distance from P1 model to P0, KLD(P0/P0) = 0
   - allow variation on p_i
   - AND supports the basic model
     - Gamma(a, b) always overfits

NTNU
Norwegian University of
Science and Technology

Outline

NTNU
Norwegian University of
Science and Technology



Heart data, from `catdata` package

NTNU
Norwegian University of
Science and Technology

Model?

- Generalized Linear Model?
  - linear predictor
  - nonlinear link (logit, probit, and others)
- Nonlinear effect from covariate?
  - parametric nonlinear function?
  - non parametric nonlinear function?
- Bayesian?

```
l <- glm(y ~ age, family=binomial, data=heart)
b <- inla(y ~ age, family='binomial', data=heart)
round(cbind(l=coef(summary(l))[,1:2], b=b$summary.fix[,1:2])

##             l.Estimate l.Std. Error  b.mean   b.sd
## (Intercept)    -3.5217       0.4160 -3.5257 0.4160
## age             0.0641       0.0085  0.0643 0.0085
```

GxxMs: Different names for the same thing

GLMM/GAM/GAMM/+++

- Perhaps the most important class of statistical models
- Many “different” models belong to this class
- No good (enough) MCMC solution around
- Even frequentist approaches does not scale well computationally

Back to linear models

Consider the linear model

$$y_i = \beta_0 \boldsymbol{F}_{i1} + \beta_1 \boldsymbol{F}_{i2} + \beta_2 \boldsymbol{F}_{i3} + u_i + \epsilon_i$$

where $\boldsymbol{F}$ is the design matrix (with ones at first column)

- y_i is an observation
- μ is the intercept
- β_0, β_1 and β_2 are the regression coefficients
- $\boldsymbol{u}$ is a random effect
- ϵ_i is i.i.d. normal observation noise.

How it works in a Bayesian framework?

NTNU
Norwegian University of
Science and Technology

Bayesian linear models

Linear model:

$$y_i = \beta_0 \boldsymbol{F}_{i1} + \beta_1 \boldsymbol{F} i2 + \beta_2 \boldsymbol{F}_{i1} + u_i + \epsilon_i$$

Bayesian model: chose priors. Usual choices:

- $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T \sim N(\mathbf{0}, \tau_{\text{fix}}^{-1} \boldsymbol{I})$, where τ_{fix} is a small number
- $\boldsymbol{u} \sim N(\mathbf{0}, \boldsymbol{Q}_{\boldsymbol{u}}^{-1})$ where the *precision matrix* $\boldsymbol{Q}$ is known
- $\epsilon \sim N(\mathbf{0}, \tau_{\text{n}}^{-1} \boldsymbol{I})$

NTNU
Norwegian University of
Science and Technology

What does this look like? (Horror slide!)

$(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{\beta})$ are jointly Gaussian!

$$\pi(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\beta}) \propto \exp\left(-\frac{\tau_\text{n}}{2}(\boldsymbol{y}-\boldsymbol{u}-\boldsymbol{F}^T\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{u}-\boldsymbol{F}^T\boldsymbol{\beta})\right)$$
$$= \exp\left(-\frac{\tau_\text{n}}{2}\begin{pmatrix}\boldsymbol{y}^T & \boldsymbol{u}^T & \boldsymbol{\beta}^T\end{pmatrix}\begin{pmatrix}\boldsymbol{I} & -\boldsymbol{I} & -\boldsymbol{F}^T \\ -\boldsymbol{I} & \boldsymbol{I} & \boldsymbol{F}^T \\ -\boldsymbol{F} & \boldsymbol{F} & \boldsymbol{F}^T\boldsymbol{F}\end{pmatrix}\begin{pmatrix}\boldsymbol{y} \\ \boldsymbol{u} \\ \boldsymbol{\beta}\end{pmatrix}\right)$$

It follows that

$$\pi(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{\beta}) = \pi(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\beta})\pi(\boldsymbol{u})\pi(\boldsymbol{\beta})$$
$$\propto \exp\left(-\frac{\tau_\text{n}}{2}\begin{pmatrix}\boldsymbol{y}^T & \boldsymbol{u}^T & \boldsymbol{\beta}^T\end{pmatrix}\begin{pmatrix}\boldsymbol{I} & -\boldsymbol{I} & -\boldsymbol{F}^T \\ -\boldsymbol{I} & \boldsymbol{I}+\tau_\text{n}^{-1}\boldsymbol{Q}_{\boldsymbol{u}} & \boldsymbol{F}^T \\ -\boldsymbol{F} & \boldsymbol{F} & \boldsymbol{F}^T\boldsymbol{F}+\frac{\tau_\text{fix}}{\tau_\text{n}}\boldsymbol{I}\end{pmatrix}\begin{pmatrix}\boldsymbol{y} \\ \boldsymbol{u} \\ \boldsymbol{\beta}\end{pmatrix}\right)$$

How can we use this?

From multivariate Gaussian distributions:
If

$$\boldsymbol{x} \equiv \begin{pmatrix} \boldsymbol{x}_A \\ \boldsymbol{x}_B \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \begin{pmatrix} \boldsymbol{Q}_{AA} & \boldsymbol{Q}_{AB} \\ \boldsymbol{Q}_{BA} & \boldsymbol{Q}_{BB} \end{pmatrix}^{-1} \right),$$

then the conditional distribution is given by

$$\boldsymbol{x}_A | \boldsymbol{x}_B \sim N\left(\boldsymbol{\mu}_A - \boldsymbol{Q}_{AA}^{-1} \boldsymbol{Q}_{AB} (\boldsymbol{x}_B - \boldsymbol{\mu}_B), \boldsymbol{Q}_{AA}^{-1} \right).$$

We can easily compute the marginal distributions for $u_i | \boldsymbol{y}$ and $\beta_i | \boldsymbol{y}$.

NTNU
Norwegian University of
Science and Technology

Non Gaussian likelihood?

General framework:

— include the linear predictor in $\boldsymbol{x}$

- with a small fixed variance

—

$$\pi(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{\#\text{data}} \pi(y_i|x_i)$$

— Gaussian approximantion does the rest

NTNU
Norwegian University of
Science and Technology

Further Examples

— Dynamic linear models
— Stochastic volatility models (famously difficult with MCMC)
— Generalised linear (mixed) models
— Generalised additive (mixed) models
— Spline smoothing
— Semiparametric regression
— Space-varying (semiparametric) regression models
— Disease mapping
— Log-Gaussian Cox-processes
— Model-based geostatistics (*)
— Spatio-temporal models
— Survival analysis
— +++

NTNU
Norwegian University of
Science and Technology

Outline

NTNU
Norwegian University of Science and Technology

On our Bayesian hierarchical model

— Inference on (what we know about) θ and $\boldsymbol{x}$ given $\boldsymbol{y}$
- in maths: $\pi(\boldsymbol{x}|\boldsymbol{y})$ and $\pi(\theta|\boldsymbol{y})$

— considering $\pi(\boldsymbol{y}|\boldsymbol{x},\theta)$, $\pi(\boldsymbol{x}|\theta)$ and $\pi(\theta)$

NTNU
Norwegian University of
Science and Technology

On our Bayesian hierarchical model

- Inference on (what we know about) $\boldsymbol{\theta}$ and $\boldsymbol{x}$ given $\boldsymbol{y}$
  - in maths: $\pi(\boldsymbol{x}|\boldsymbol{y})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$
- considering $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$, $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$
- using the Bayes theorem,

$$\pi(\boldsymbol{x}|\boldsymbol{y}) = \int \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \int \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{x}$$

NTNU
Norwegian University of
Science and Technology

On our Bayesian hierarchical model

- Inference on (what we know about) $\boldsymbol{\theta}$ and $\boldsymbol{x}$ given $\boldsymbol{y}$
  - in maths: $\pi(\boldsymbol{x}|\boldsymbol{y})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$
- considering $\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})$, $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$
- using the Bayes theorem,

$$\pi(\boldsymbol{x}|\boldsymbol{y}) = \int \pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \int \pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{x}$$

- even more...
  - $\pi(\theta_j|\boldsymbol{y})$, $j = 1, ..., \dim(\boldsymbol{\theta})$
  - $\pi(x_i|\boldsymbol{y})$, $i = 1, ..., \dim(\boldsymbol{x})$

NTNU
Norwegian University of
Science and Technology

The inference problem

— we have to compute

$$\pi(x_i|\boldsymbol{y}) \propto \int_{x_{\{-i\}}} \int_{\boldsymbol{\theta}} \pi(y|\boldsymbol{x}, \theta)\pi(\boldsymbol{x}|\theta)\pi(\theta)\, d\theta\, d\boldsymbol{x}_{\{-i\}}$$

NTNU
Norwegian University of
Science and Technology

The inference problem

— we have to compute

$$\pi(x_i|\boldsymbol{y}) \propto \int_{x_{\{-i\}}} \int_{\boldsymbol{\theta}} \pi(y|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}\, d\boldsymbol{x}_{\{-i\}}$$

and

$$\pi(\theta_j|\boldsymbol{y}) \propto \int_{x} \int_{\boldsymbol{\theta}_{\{-j\}}} \pi(y|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}_{\{-j\}}\, d\boldsymbol{x}$$

NTNU
Norwegian University of
Science and Technology

The inference problem

— we have to compute

$$\pi(x_i|\boldsymbol{y}) \propto \int_{x_{\{-i\}}} \int_{\boldsymbol{\theta}} \pi(y|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}\, d\boldsymbol{x}_{\{-i\}}$$

and

$$\pi(\theta_j|\boldsymbol{y}) \propto \int_{x} \int_{\boldsymbol{\theta}_{\{-j\}}} \pi(y|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}_{\{-j\}}\, d\boldsymbol{x}$$

— remember

- dim($\boldsymbol{\theta}$) is small
- dim($\boldsymbol{x}$) is not small
- we have to compute very high dimensional integrals

NTNU
Norwegian University of Science and Technology

The inference problem

- we have to compute

$$\pi(x_i|\boldsymbol{y}) \propto \int_{x_{\{-i\}}} \int_{\boldsymbol{\theta}} \pi(y|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}\, d\boldsymbol{x}_{\{-i\}}$$

  and

$$\pi(\theta_j|\boldsymbol{y}) \propto \int_{x} \int_{\boldsymbol{\theta}_{\{-j\}}} \pi(y|\boldsymbol{x},\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}_{\{-j\}}\, d\boldsymbol{x}$$

- remember
  - dim($\boldsymbol{\theta}$) is small
  - dim($\boldsymbol{x}$) is not small
  - we have to compute very high dimensional integrals
- typically they are not analytically tractable
  - $\rightarrow$ we have to approach

NTNU
Norwegian University of
Science and Technology

using MCMC

— single-site: compute (the expressions) for

- $p(\theta_j|\boldsymbol{\theta}_{-j}, \boldsymbol{x}, \boldsymbol{y})$
- $p(x_i|\boldsymbol{x}_{-i}, \boldsymbol{\theta}, \boldsymbol{y})$

NTNU
Norwegian University of
Science and Technology

using MCMC

— single-site: compute (the expressions) for
  - $p(\theta_j|\boldsymbol{\theta}_{-j}, \boldsymbol{x}, \boldsymbol{y})$
  - $p(x_i|\boldsymbol{x}_{-i}, \boldsymbol{\theta}, \boldsymbol{y})$

— draw samples from such conditionals
  - WinBUGS, OpenBUGS, JAGS, and others

— use these samples to summarize $p(\boldsymbol{x})$ and $p(\boldsymbol{\theta})$

NTNU
Norwegian University of
Science and Technology

using MCMC

- single-site: compute (the expressions) for
  - $p(\theta_j | \boldsymbol{\theta}_{-j}, \boldsymbol{x}, \boldsymbol{y})$
  - $p(x_i | \boldsymbol{x}_{-i}, \boldsymbol{\theta}, \boldsymbol{y})$
- draw samples from such conditionals
  - WinBUGS, OpenBUGS, JAGS, and others
- use these samples to summarize $p(\boldsymbol{x})$ and $p(\boldsymbol{\theta})$
- warning
  - sampling from $x_i | \boldsymbol{x}_{-i}, \boldsymbol{\theta}, y$
    - slow convergence when strong dependence
    - does not works for our example...
  - better: draw joint sample from $\boldsymbol{x} | \boldsymbol{\theta}, y$
  - best: use INLA

NTNU
Norwegian University of Science and Technology

Outline

NTNU
Norwegian University of Science and Technology

What INLA does

- INLA does:
  - compute marginals of $\pi(x_i|\boldsymbol{y})$ and $\pi(\theta_j|\boldsymbol{y})$
- how?
  - approach $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ to approach $\pi(\boldsymbol{\theta}|\boldsymbol{y})$
  - explore $\pi(\boldsymbol{\theta}|\boldsymbol{y})$
    - approach $\pi(\theta_j|\boldsymbol{y})$
  - approach $\pi(x_i|\boldsymbol{x}_{-i})$

NTNU
Norwegian University of
Science and Technology

Important ingredient

The GMRF-approximation

$$\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y}) \quad \propto \quad \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \sum_i \log \pi(y_i|x_i)\right)$$

NTNU
Norwegian University of
Science and Technology

Important ingredient

The GMRF-approximation

$$\begin{aligned}\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y}) &\propto \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \sum_i \log\pi(y_i|x_i)\right)\\ &\approx \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{Q}+\text{diag}(\boldsymbol{c}))(\boldsymbol{x}-\boldsymbol{\mu})\right)\\ &= \pi_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})\end{aligned}$$

$c_i = -\frac{dl_i^2}{dx_i^2}$ where $l_i = \log(\pi(y_i|x_i))$, $i = 1, ..., \#$ data

NTNU
Norwegian University of Science and Technology

Important ingredient

The GMRF-approximation

$$
\begin{aligned}
\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y}) &\propto \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \sum_i \log \pi(y_i|x_i)\right) \\
&\approx \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{Q} + \text{diag}(\boldsymbol{c}))(\boldsymbol{x}-\boldsymbol{\mu})\right) \\
&= \pi_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})
\end{aligned}
$$

$c_i = -\frac{dl_i^2}{dx_i^2}$ where $l_i = \log(\pi(y_i|x_i))$, $i = 1, ..., \#$ data

— Markov and computational properties (on $\boldsymbol{Q}$) are preserved

Important ingredient

The GMRF-approximation

$$\begin{aligned}\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y}) &\propto \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \sum_i \log \pi(y_i|x_i)\right) \\ &\approx \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{Q}+\text{diag}(\boldsymbol{c}))(\boldsymbol{x}-\boldsymbol{\mu})\right) \\ &= \pi_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})\end{aligned}$$

$c_i = -\frac{dl_i^2}{dx_i^2}$ where $l_i = \log(\pi(y_i|x_i))$, $i = 1, ..., \#$ data

— Markov and computational properties (on $\boldsymbol{Q}$) are preserved

— $\widetilde{\pi}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ costs

- temporal: $O(n)$
- spatial: $O(n\log(n))$

If $\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}$ is *Gaussian*, the “approximation” is exact.

INLA, $\pi(\theta_j|\boldsymbol{y})$

— Considering

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}$$

NTNU
Norwegian University of
Science and Technology

INLA, $\pi(\theta_j|\boldsymbol{y})$

— Considering

$$
\begin{aligned}
\pi(\boldsymbol{\theta}|\boldsymbol{y}) &= \frac{\pi(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})} \\
&\propto \frac{\pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})}{\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}
\end{aligned}
$$

INLA, $\pi(\theta_j|\boldsymbol{y})$

— Considering

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{y}) &= \frac{\pi(\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})} \\ &\propto \frac{\pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})}{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}\end{aligned}$$

— Gaussian approximation to denominator

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \approx \frac{\pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})}{\pi_{\mathrm{G}}(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}\Big|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}$$

INLA, $\pi(\theta_j|\boldsymbol{y})$

— Considering

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{y}) &= \frac{\pi(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})} \\ &\propto \frac{\pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})}{\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}\end{aligned}$$

— Gaussian approximation to denominator

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \approx \frac{\pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})}{\pi_{\mathrm{G}}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})} \Big|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}$$

— mode of $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ (optimization)
- explore $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$
  — approach $\pi(\theta_j|\boldsymbol{y})$ (numerical integration)

NTNU
Norwegian University of
Science and Technology

INLA, $\pi(x_i|\boldsymbol{y}, \boldsymbol{\theta})$

Approaching $\pi(x_i|\boldsymbol{y}, \boldsymbol{\theta})$

- Problem
  - dim($\boldsymbol{x}$)=n is not small
  - n marginals to compute
- Laplace approximation

$$\widetilde{\pi}(x_i \mid \boldsymbol{y}, \boldsymbol{\theta}) \approx \left.\frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})}{\widetilde{\pi}_{GG}(\boldsymbol{x}_{-i}|x_i, \boldsymbol{y}, \boldsymbol{\theta})}\right|_{\boldsymbol{x}_{-i}=\boldsymbol{x}^*_{-i}(x_i,\boldsymbol{\theta})}$$

NTNU
Norwegian University of Science and Technology

INLA, $\pi(x_i|\boldsymbol{y},\boldsymbol{\theta})$

Approaching $\pi(x_i|\boldsymbol{y},\boldsymbol{\theta})$

- Problem
  - dim($\boldsymbol{x}$)=n is not small
  - n marginals to compute
- Laplace approximation

$$\widetilde{\pi}(x_i \mid \boldsymbol{y},\boldsymbol{\theta}) \approx \left.\frac{\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})}{\widetilde{\pi}_{GG}(\boldsymbol{x}_{-i}|x_i,\boldsymbol{y},\boldsymbol{\theta})}\right|_{\boldsymbol{x}_{-i}=\boldsymbol{x}^*_{-i}(x_i,\boldsymbol{\theta})}$$

- simpler/cruder (fast) approximation (from $\pi_G(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$)

$$\hat{\pi}(x_i|\boldsymbol{y},\boldsymbol{\theta}) = N(x_i;\mu_i(\boldsymbol{\theta}),\sigma_i^2(\boldsymbol{\theta}))$$

NTUU
Norwegian University of Science and Technology

INLA, $\pi(x_i|\boldsymbol{y})$

Approaching $\pi(x_i|\boldsymbol{y},\boldsymbol{\theta})$

- integrate $\boldsymbol{\theta}$ out from $\widetilde{\pi}(x_i \mid \boldsymbol{y},\boldsymbol{\theta})$
- select values for $\boldsymbol{\theta}$
- use weighted sum

$$\widetilde{\pi}(x_i \mid \boldsymbol{y}) \propto \sum_j \widetilde{\pi}(x_i|\boldsymbol{y},\boldsymbol{\theta}_j) \times \widetilde{\pi}(\boldsymbol{\theta}_j|\boldsymbol{y})$$

Remarks

1. Expect $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ to be accurate, since
   - $\boldsymbol{x}|\boldsymbol{\theta}$ is *a priori* Gaussian
   - Likelihood models are 'well-behaved' so

     $$\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$$

     is *almost* Gaussian.
2. There are no distributional assumptions on $\boldsymbol{\theta}|\boldsymbol{y}$
3. Similar remarks are valid to

$$\widetilde{\pi}(x_i \mid \boldsymbol{\theta}, \boldsymbol{y})$$

NTNU
Norwegian University of
Science and Technology

How can we assess the error in the approximations?

Tool 1: Compare a sequence of improved approximations

1. Gaussian approximation
2. Simplified Laplace
3. Laplace

No big differences $\rightarrow$ good approximation

NTNU
Norwegian University of Science and Technology

How can we assess the error in the approximations?

Tool 2: Estimate the "effective" number of parameters as defined in the Deviance Information Criteria:

$$p_{\mathrm{D}}(\theta) = \overline{D}(\boldsymbol{x};\theta) - D(\overline{\boldsymbol{x}};\theta)$$

and compare this with the number of observations
Low ratio is good.
This criteria has theoretical justification.

NTNU
Norwegian University of Science and Technology