Hierarchical Bayesian models

Hierarchical models are an extremely useful tool in Bayesian model building.

Three parts:

- Observation model $\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}$: Encodes information about observed data.
- The latent model $\boldsymbol{x}|\boldsymbol{\theta}$: The unobserved process.
- Hyperpriors for $\boldsymbol{\theta}$: Models for all of the parameters in the observation and latent processes.

Bayesian computing

The posterior distribution is given by

$$\mathrm{p}(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) \propto \mathrm{p}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})\,\mathrm{p}(\boldsymbol{x}|\boldsymbol{\theta})\,\mathrm{p}(\boldsymbol{\theta})d\boldsymbol{\theta}$$

We are interested in the posterior marginal quantities like $\mathrm{p}(x_i|\boldsymbol{y})$ and $\mathrm{p}(\theta_j|\boldsymbol{y})$. This requires the evaluation of integrals of the form

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

The computation of massively high dimensional integrals is at the core of Bayesian computing.

- Hierarchical models might be difficult for MCMC, leading to long computation times.
- Few generic tools, for example OpenBUGS, based on MCMC exist.
- Specialized tools, such as `BayesX`, for specific models are available and mainly based on MCMC.

Approximate inference

Except for the (trivial) cases when everything can be computed exactly (maybe up to very small integration error), we can never do exact inference.

- Integrated nested Laplace approximations (INLA) are a promising alternative to inference via MCMC in latent Gaussian models (Rue et al, 2009, JRSS-B).
- The methodology is particularly attractive if the latent Gaussian model is a Gaussian Markov random field (GMRF) (Rue and Held, 2005).

Latent Gaussian models

- Assume y_i, $i = 1, \ldots, n$ belongs to the exponential family.
- Let $\mathrm{E}(y_i) = \mu_i$ be linked to η_i: $g(\mu_i) = \eta_i$.
- The linear predictor η_i is then additively described by:

$$\eta_i = \alpha + \sum_{l=1}^{L} f^{(l)}(u_{li}) + \sum_{k=1}^{K} \beta_k z_{ki} + \epsilon_i$$

- ϵ_i's are unstructured terms.
- $f^{(l)}$ are unknown functions of covariate l taking value u_{li} for observation i.
- β_k's are linear effects for covariates z_{ki}'s.
- Define $\boldsymbol{x}$ as all η_i, $\{f^{(l)}\}$, β_k, and α, such that

$$\boldsymbol{x}|\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta}))$$

- $\boldsymbol{\theta}$ are hyperparameters to which we can assign priors.

Latent Gaussian models

Latent Gaussian models are a subset of Bayesian additive models.

- Observation model :

$$\pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) \sim \prod_i \pi(y_i|x_i,\boldsymbol{\theta}).$$

- Latent Gaussian model: $\pi(\boldsymbol{x}|\theta)$, usually a GMRF,

  $\pi(\boldsymbol{x}|\theta) \sim \mathcal{N}(\cdot, \boldsymbol{Q}(\boldsymbol{\theta})^{-1})$, with sparse $\boldsymbol{Q}(\boldsymbol{\theta})$.

- The hyperparameters $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

Comment: Methods for sparse matrices are implemented in the R-package `spam` (Furrer and Sain, 2010).

Gaussian Markov Random Field (GMRF)

Theorem

Let $\boldsymbol{x}$ be normal distributed with mean $\boldsymbol{\mu}$ and symmetric positive-definite (SPD) precision matrix $\boldsymbol{Q}$, i.e. $\boldsymbol{Q} > 0$, Then for $i \neq j$,

$$x_i \perp x_j | \boldsymbol{x}_{-ij} \iff Q_{ij} = 0$$

Gaussian Markov Random Field (GMRF) cont.

Definition

A random vector $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$ is called a GMRF with respect to a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q} > 0$, iff its density has the form

$$\pi(\boldsymbol{x}) = (2\pi)^{-n/2} |\boldsymbol{Q}|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

and

$$Q_{ij} \neq 0 \iff \{i,j\} \in \mathcal{E}, \forall i \neq j$$

If $\boldsymbol{Q}$ is completely dense then $\mathcal{G}$ is fully connected. Thus, any normal distribution with a SPD covariance matrix is also a GMRF and vice versa.

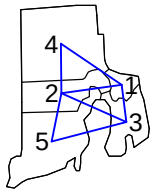
Example: Gaussian Markov random fields

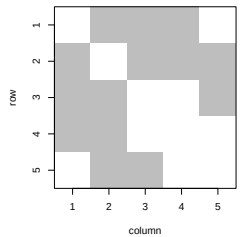
Computational properties: Sparseness of $\boldsymbol{Q}$ allows us to do many computations in an efficient way.

Comment: An intrinsic GMRF is is a GMRF where the precision matrix $\boldsymbol{Q}$ is singular (i.e. it has at least one zero eigenvalue).

Five counties of the US state Rhode Island

Conditional dependence defined over neighborhood structure.





Adjacency matrix

3	-1	-1	- 1	0
-1	4	-1	-1	-1
-1	-1	3	0	-1
-1	-1	0	2	0
0	-1	-1	0	2

Structure matrix

Main interest

The posterior distribution is given by

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta}) \prod_i \pi(y_i|x_i, \boldsymbol{\theta})$$

We are mainly interested in the posterior marginals

$$\pi(x_i|\boldsymbol{y}) = \int_{\boldsymbol{\theta}} \underbrace{\int_{\boldsymbol{x}_{-i}} \pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{x}_{-i}}_{\pi(x_i, \boldsymbol{\theta}|\boldsymbol{y}) = \pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})} \; d\boldsymbol{\theta} \tag{4}$$

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

INLA approximates (4) by

$$\widetilde{\pi}(x_i|\boldsymbol{y}) = \sum_k \widetilde{\pi}(x_i|\boldsymbol{\theta}_k, \boldsymbol{y}) \, \widetilde{\pi}(\boldsymbol{\theta}_k|\boldsymbol{y}) \, \Delta_k.$$

That means, compute approximations for $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ and $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$, and use numerical integration (a finite sum) to integrate out $\boldsymbol{\theta}$.

Approximating $\pi(\boldsymbol{\theta}|\boldsymbol{y})$

- From $\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y}) = \pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y}) \times \pi(\boldsymbol{\theta}|\boldsymbol{y}) \times \pi(\boldsymbol{y})$ it follows that

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})} \text{ for all } \boldsymbol{x}.$$

- INLA approximates $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ using

$$\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}.$$

  which is also known as Laplace approximation.

- Here $\widetilde{\pi}_G$ is the Gaussian (GMRF) approximation to $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ and $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the mode of $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$.

The GMRF approximation

Let $\boldsymbol{x}$ denote a GMRF with precision matrix $\boldsymbol{Q}$ and mean $\boldsymbol{\mu}$. Approximate

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y}) \propto \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \sum_{i=1}^n \log \pi(y_i|x_i)\right)$$

by using a second-order Taylor expansion of $\log \pi(y_i|x_i)$ around $\boldsymbol{\mu}_0$, say.

Recall

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 = a + bx - \frac{1}{2}cx^2$$

with $b = f'(x_0) - f''(x_0)x_0$ and $c = -f''(x_0)$.

The GMRF approximation (II)

Thus,

$$\tilde{\pi}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y}) \propto \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \sum_{i=1}^n (a_i + b_i x_i - 0.5 c_i x_i^2)\right)$$
$$\propto \exp\left(-\frac{1}{2}\boldsymbol{x}^T(\boldsymbol{Q} + \text{diag}(\boldsymbol{c}))\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x}\right)$$

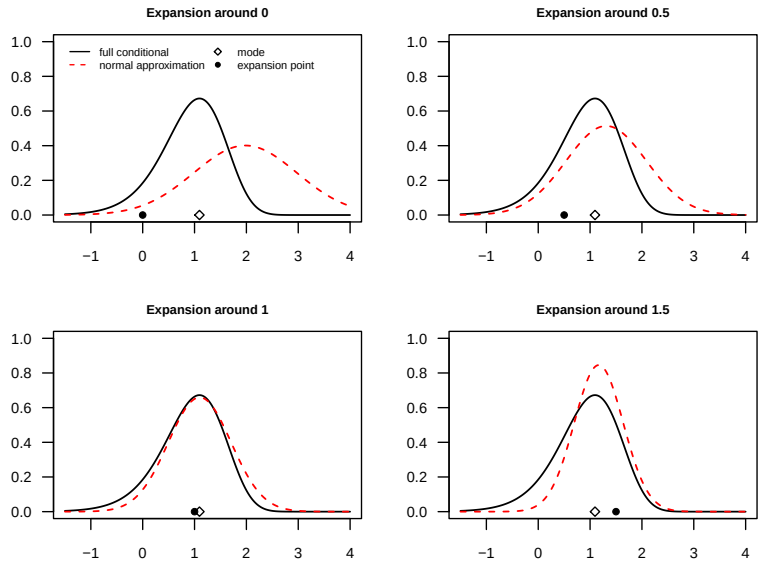
to get a Gaussian approximation with precision matrix $\boldsymbol{Q} + \text{diag}(\boldsymbol{c})$ and mean given by the solution of $(\boldsymbol{Q} + \text{diag}(\boldsymbol{c}))\boldsymbol{\mu} = \boldsymbol{b}$. The canonical parameterization is

$$\mathcal{N}_C(\mathbf{b}, \boldsymbol{Q} + \text{diag}(\boldsymbol{c}))$$

which corresponds to

$$\mathcal{N}((\boldsymbol{Q} + \text{diag}(\boldsymbol{c}))^{-1}\mathbf{b}, (\boldsymbol{Q} + \text{diag}(\boldsymbol{c}))^{-1}).$$

The GMRF approximation

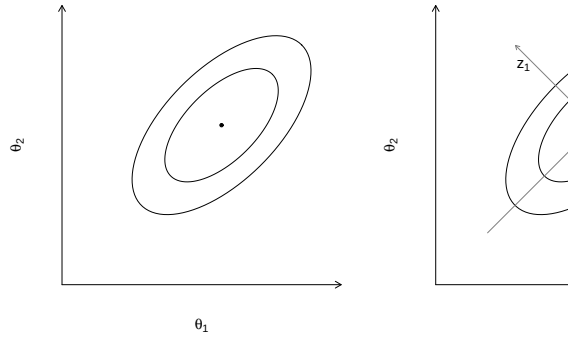


Exploring $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$

$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is "numerically explored" to find suitable support points $\boldsymbol{\theta}_k$.

Main use: Select good evaluation points $\boldsymbol{\theta}_k$ for the numerical integration when approximating $\tilde{\pi}(x_i|\boldsymbol{y})$

- Locate the mode
- Compute the Hessian to construct principal components
- Grid-search to locate bulk of the probability mass



All points found have equal area weight Δ_k.

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

For approximating the first component $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ we can use

- a Gaussian approximation, easily extractable from $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$. However, errors in location and/or lack of skewness possible.
- a Laplace approximation

$$\tilde{\pi}_{\text{LA}}(x_i|\boldsymbol{\theta}, \boldsymbol{y}) \propto \left.\frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\tilde{\pi}_{\text{GG}}(\boldsymbol{x}_{-i}|x_i, \boldsymbol{\theta}, \boldsymbol{y})}\right|_{\boldsymbol{x}_{-i}=\boldsymbol{x}^*_{-i}(x_i,\boldsymbol{\theta})}.$$

  The approximation is very accurate but very expensive.
- a simplified Laplace approximation based on fitting a skew-normal distribution to a series expansion of $\tilde{\pi}_{LA}$.

INLA: Overview

Step I Approximate $\pi(\boldsymbol{\theta}|y)$ using the Laplace approximation and select good evaluation points $\boldsymbol{\theta}_k$.

Step II For each $\boldsymbol{\theta}_k$ and i approximate $\pi(x_i|\boldsymbol{\theta}_k, \boldsymbol{y})$ using the Laplace or simplified Laplace approximation for selected values of x_i .

Step III For each i, sum out $\boldsymbol{\theta}_k$

$$\tilde{\pi}(x_i|\boldsymbol{y}) = \sum_k \tilde{\pi}(x_i|\boldsymbol{\theta}_k, \boldsymbol{y}) \times \tilde{\pi}(\boldsymbol{\theta}_k|\boldsymbol{y}) \times \Delta_k.$$

Build a log spline corrected Gaussian to represent $\tilde{\pi}(x_i|\boldsymbol{y})$.

INLA features

INLA fully incorporates posterior uncertainty with respect to hyperparameters ⇒ tool for full Bayesian inference

- Marginal posterior densities of all (hyper-)parameters
- Posterior mean, median, quantiles, std. deviation, etc.

The approach can be used for predictions, model assessment, ...

Model description

Hierarchical GMRF models supported by `inla`

- Time series models
- Generalized linear (mixed) models
- Generalized additive (mixed) models
- Disease mapping, see exercise
- Spatio-temporal models
- Survival models
- Univariate volatility models
- ...

The INLA-program

- The inla program is written in C and built on the open source GMRFLib library (Rue and Held, 2005)
- There is no need for C programming when applying inla.
- An R-package called INLA, which works as an interface, is available.

  `http:www.r-inla.org`

  and can be installed by typing

  source("http://www.math.ntnu.no/inla/givemeINLA-testing.R")

  in the R-terminal, see `http://www.r-inla.org/download`
- Its usage is similar to the `glm(.)` routine in R.

The R-package `r-inla`

The R-package `r-inla` allows the user to combine different types of likelihood models with different regression models.

```
> formula = y ~ a + b + a:b
>     + f(idx, model="iid", hyper=list(...), ...) + ...
```

Once the linear predictor is specified, a basic call to fit the model with R-INLA takes the following form:

```
> result = inla(formula, data=data.frame(y,a,b,idx), family="binomial", ...)
```

After the computations the variable result will hold an S3 object of class "inla", from which summaries, plots, and posterior marginals can be obtained.

The Bayesian age-period-cohort model

- Data $y_{ij} \sim Po(E_{ij} \exp(\eta_{ij}))$
- Linear predictor: $\eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_k + z_{ij}$

  with age effect α_i, period effect β_j, cohort effect γ_k and additional random effect $z_{ij} \sim \mathcal{N}(0, \kappa_z^{-1})$.
- $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ are modelled as a first order random walk

$$\alpha_i \sim \mathcal{N}(\alpha_{i-1}, \kappa^{-1})$$

  and $\boldsymbol{z}$ as an i.i.d. Gaussian effect.
- Sum-to-zero constraints are imposed for $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$
- Precisions $\kappa_{\text{age}}, \kappa_{\text{period}}, \kappa_{\text{cohort}}, \kappa_z$

Preparation of the data for the R-interface of INLA

Illustration:

$I = 2,\ J = 3,\ K = 4$

i				
2 ←	35-40	$\frac{20}{260}$	$\frac{12}{270}$	$\frac{10}{290}$
1 ←	30-35	$\frac{3}{250}$	$\frac{9}{230}$	$\frac{7}{260}$
		1980	1985	1990
		↓	↓	↓
	j	1	2	3

Input data file for R

y	E	i	j	k	z
3	250	1	1	2	1
20	260	2	1	1	2
9	230	1	2	3	3
12	270	2	2	2	4
7	260	1	3	4	5
10	290	2	3	3	6

Call INLA using the R-interface

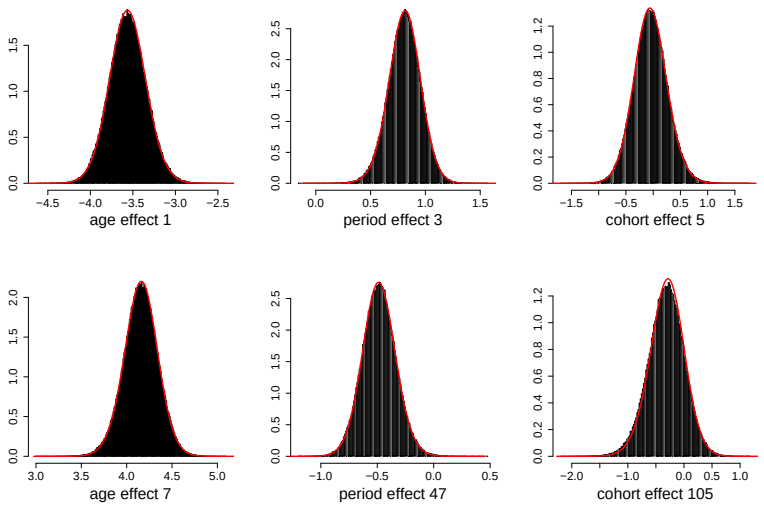
Specify the formula object

```
>     library(INLA)
>     # parameters of gamma prior for precision
>     my.hyper <- list(prec=list(prior="loggamma", param=c(1.0, 0.005)))
>     formula <- y ~ f(i, model="rw1", hyper=my.hyper, constr=T)
>         + f(j, model="rw1", hyper=my.hyper, constr=T)
>         + f(k, model="rw1", hyper=my.hyper, constr=T)
>         + f(z, model="iid", hyper=my.hyper)
```

Call INLA:

```
>     result <- inla(formula, family="poisson", data=data, E=E, verbose=T)
```

Comparison of posterior marginals to MCMC



INLA object

INLA returns an object of class "inla". This is a list object contains (besides others) the following arguments:

- `summary.fixed`: Matrix containing the mean and standard deviation (plus, possibly quantiles and cdf) of the the fixed effects of the model.
- `marginals.fixed`: A list containing the posterior marginal densities of the fixed effects of the model.
- `summary.random`: Analogous to summary.fixed for the random effects
- `marginals.random`: Analogous to summary.marginals for the random effects
- `summary.hyperpar`: Analogous to summary.fixed for the hyperparameters
- `marginals.hyperpar`: Analogous to summary.marginals for the hyperparameters
- ...

Let's call the returned inla-object "result". You can use

- `summary(result)`
- `plot(result)`

to summarize and plot the output information.