Bayesian animal model using Integrated Nested Laplace Approximations

-A wild house sparrow population case study

Anna M. Holand
Ingelin Steinsland

Animal model workshop with application to ecology at Oulu
November 2014

Outline

- Lecture
  - Motivating cases
  - Animal models
  - Bayesian inference
  - INLA
  - Animal model and INLA
- Computer exercise 1: `AnimalINLA` R package
- Computer exercise 2: `R` – `INLA` package

Some history - $\mathtt{R-INLA}$

- In the beginning there was

  GMRFLib

  a C library for fast computations for GMRFs.
- GMRFLib became

  INLA

  a C library for fast approximate inference, accessed through .ini files
- After much wailing and gnashing of teeth there came

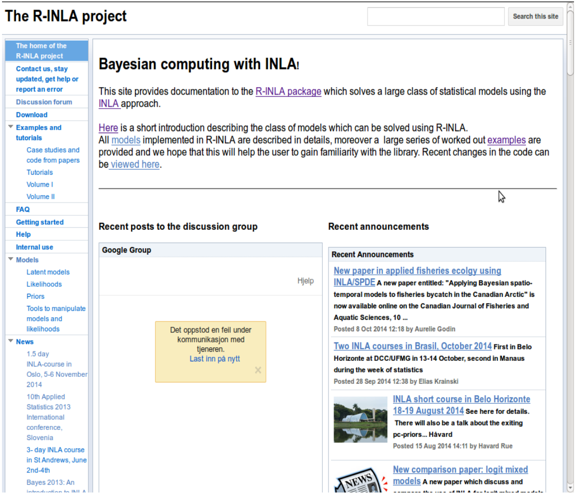
  $\mathtt{R-INLA}$

  which takes `R` code and writes an appropriate .ini file for the INLA C-program to read. (This is why $\mathtt{R-INLA}$ is not on CRAN).

R − INLA

- Can be downloaded from http : //www.math.ntnu.no/ hrue/GMRFLib/R − INLA/ or www.r − inla.org
- Available from Unix, Windows and Mac

R — INLA



The R-INLA project

Search this site

The home of the R-INLA project
Contact us, stay updated, get help or report an error
Discussion forum
Download
Examples and tutorials
Case studies and code from papers
Tutorials
Volume I
Volume II
FAQ
Getting started
Help
Internal use
Models
Latent models
Likelihoods
Priors
Tools to manipulate models and likelihoods
News
1.5 day INLA-course in Oslo, 5-6 November 2014
10th Applied Statistics 2013 International conference, Slovenia
3- day INLA course in St Andrews, June 2nd-4th
Bayes 2013: An

Bayesian computing with INLA!

This site provides documentation to the R-INLA package which solves a large class of statistical models using the INLA approach.

Here is a short introduction describing the class of models which can be solved using R-INLA.
All models implemented in R-INLA are described in details, moreover a large series of worked out examples are provided and we hope that this will help the user to gain familiarity with the library. Recent changes in the code can be viewed here.

Recent posts to the discussion group

Google Group

Hjelp

Det oppstod en feil under kommunikasjon med tjeneren.
Last inn på nytt

Recent announcements

Recent Announcements

New paper in applied fisheries ecolgy using INLA/SPDE A new paper entitled: "Applying Bayesian spatio-temporal models to fisheries bycatch in the Canadian Arctic" is now available online on the Canadian Journal of Fisheries and Aquatic Sciences, 10 ...
Posted 8 Oct 2014 12:18 by Aurelie Godin

Two INLA courses in Brasil, October 2014 First in Belo Horizonte at DCC/UFMG in 13-14 October, second in Manaus during the week of statistics
Posted 28 Sep 2014 12:38 by Elias Krainski

INLA short course in Belo Horizonte 18-19 August 2014 See here for details. There will also be a talk about the exiting pc-priors... Håvard
Posted 15 Aug 2014 14:11 by Havard Rue

New comparison paper: logit mixed models A new paper which discuss and

All functions are provided with a help file which can be viewed using >`?inla`

Else, maybe a bit poor/outdated manuals...

The structure of an `R` program using INLA

There are essentially three parts to an INLA program:

1. The data organization (important!)
2. The *formula* -notation inherited from `R`'s native `glm` function
3. The call to the INLA program

Running INLA on a remote machine

It is also possible to run INLA on a remote (Linux) machine.

- It takes some effort to set up (see the FAQ at $\texttt{r} - \texttt{inla.org}$)
- But INLA is programmed using OpenMP and can make use of multiple cores
- The option in the `inla` call is $\texttt{inla.call} = "\texttt{remote}"$

We have a 12 core machine with lots of memory that we use for big problems. It works well!

Model specification the R-INLA package

The INLA framework supports latent GMRF models of the following type

$$y_i \mid \eta_i, \boldsymbol{\theta_1} \sim \pi(y_i \mid \eta_i, \boldsymbol{\theta_1}) \quad (1)$$

$$\eta_i \mid \boldsymbol{\theta_2} = \alpha + \sum_k \beta_k(z_{ki})(\boldsymbol{\theta_2}) + \sum_j f^j(u_{ij})(\boldsymbol{\theta_2}) + \epsilon_i(\boldsymbol{\theta_2}) \quad (2)$$

- $\boldsymbol{x} = (\alpha, f^j(), \beta_k, \epsilon_i,)$ ($\boldsymbol{\eta}$ a linear combination of $\boldsymbol{x}$)
- $\{f^j()\} \sim \mathcal{N}(0, Q_f^{-1}(\tau_2))$:
  non-linear effects of the covariates $\boldsymbol{u}$ with unknown functions (structured random effects)
- $\beta_k \sim \mathcal{N}(0, \tau_1^{-1})$:
  linear effect of covariates $\boldsymbol{z}$ (fixed effects)
- ϵ_i:
  unstructured random effect of observation i.
- $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$:
  hyperparameters (likelihood and latent field)

Model specification the R-INLA package

The INLA framework supports latent GMRF models of the following type

$$y_i \mid \eta_i, \boldsymbol{\theta_1} \sim \pi(y_i \mid \eta_i, \boldsymbol{\theta_1}) \quad (1)$$

$$\eta_i \mid \boldsymbol{\theta_2} = \alpha + \sum_k \beta_k(z_{ki})(\boldsymbol{\theta_2}) + \sum_j f^j(u_{ij})(\boldsymbol{\theta_2}) + \epsilon_i(\boldsymbol{\theta_2}) \quad (2)$$

- $\boldsymbol{x} = (\alpha, f^j(), \beta_k, \epsilon_i,)$ ($\boldsymbol{\eta}$ a linear combination of $\boldsymbol{x}$)
- $\{f^j()\} \sim \mathcal{N}(0, Q_f^{-1}(\tau_2))$:
  non-linear effects of the covariates $\boldsymbol{u}$ with unknown functions (structured random effects)
- $\beta_k \sim \mathcal{N}(0, \tau_1^{-1})$:
  linear effect of covariates $\boldsymbol{z}$ (fixed effects)
- ϵ_i:
  unstructured random effect of observation i.
- $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$:
  hyperparameters (likelihood and latent field)

Model specification the R-INLA package

The INLA framework supports latent GMRF models of the following type

$$y_i \mid \eta_i, \boldsymbol{\theta_1} \sim \pi(y_i \mid \eta_i, \boldsymbol{\theta_1}) \quad (1)$$

$$\eta_i \mid \boldsymbol{\theta_2} = \alpha + \sum_k \beta_k(z_{ki})(\boldsymbol{\theta_2}) + \sum_j f^j(u_{ij})(\boldsymbol{\theta_2}) + \epsilon_i(\boldsymbol{\theta_2}) \quad (2)$$

- $\boldsymbol{x} = (\alpha, f^j(), \beta_k, \epsilon_i,)$ ($\boldsymbol{\eta}$ a linear combination of $\boldsymbol{x}$)
- $\{f^j()\} \sim \mathcal{N}(0, Q_f^{-1}(\tau_2))$:
  non-linear effects of the covariates $\boldsymbol{u}$ with unknown functions (structured random effects)
- $\beta_k \sim \mathcal{N}(0, \tau_1^{-1})$:
  linear effect of covariates $\boldsymbol{z}$ (fixed effects)
- ϵ_i:
  unstructured random effect of observation i.
- $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$:
  hyperparameters (likelihood and latent field)

Model specification the R-INLA package

The INLA framework supports latent GMRF models of the following type

$$y_i \mid \eta_i, \boldsymbol{\theta_1} \sim \pi(y_i \mid \eta_i, \boldsymbol{\theta_1}) \tag{1}$$

$$\eta_i \mid \boldsymbol{\theta_2} = \alpha + \sum_k \beta_k(z_{ki})(\boldsymbol{\theta_2}) + \sum_j f^j(u_{ij})(\boldsymbol{\theta_2}) + \epsilon_i(\boldsymbol{\theta_2}) \tag{2}$$

- $\boldsymbol{x} = (\alpha, f^j(), \beta_k, \epsilon_i,)$ ($\boldsymbol{\eta}$ a linear combination of $\boldsymbol{x}$)
- $\{f^j()\} \sim \mathcal{N}(0, Q_f^{-1}(\tau_2))$:
  non-linear effects of the covariates $\boldsymbol{u}$ with unknown functions (structured random effects)
- $\beta_k \sim \mathcal{N}(0, \tau_1^{-1})$:
  linear effect of covariates $\boldsymbol{z}$ (fixed effects)
- ϵ_i:
  unstructured random effect of observation i.
- $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$:
  hyperparameters (likelihood and latent field)

Model specification the R-INLA package

The INLA framework supports latent GMRF models of the following type

$$y_i \mid \eta_i, \boldsymbol{\theta_1} \sim \pi(y_i \mid \eta_i, \boldsymbol{\theta_1}) \quad (1)$$

$$\eta_i \mid \boldsymbol{\theta_2} = \alpha + \sum_k \beta_k(z_{ki})(\boldsymbol{\theta_2}) + \sum_j f^j(u_{ij})(\boldsymbol{\theta_2}) + \epsilon_i(\boldsymbol{\theta_2}) \quad (2)$$

- $\boldsymbol{x} = (\alpha, f^j(), \beta_k, \epsilon_i,)$ ($\boldsymbol{\eta}$ a linear combination of $\boldsymbol{x}$)
- $\{f^j()\} \sim \mathcal{N}(0, Q_f^{-1}(\tau_2))$:
  non-linear effects of the covariates $\boldsymbol{u}$ with unknown functions (structured random effects)
- $\beta_k \sim \mathcal{N}(0, \tau_1^{-1})$:
  linear effect of covariates $\boldsymbol{z}$ (fixed effects)
- ϵ_i:
  unstructured random effect of observation i.
- $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$:
  hyperparameters (likelihood and latent field)

Model specification the R-INLA package

The INLA framework supports latent GMRF models of the following type

$$y_i \mid \eta_i, \boldsymbol{\theta_1} \sim \pi(y_i \mid \eta_i, \boldsymbol{\theta_1}) \tag{1}$$

$$\eta_i \mid \boldsymbol{\theta_2} = \alpha + \sum_k \beta_k(z_{ki})(\boldsymbol{\theta_2}) + \sum_j f^j(u_{ij})(\boldsymbol{\theta_2}) + \epsilon_i(\boldsymbol{\theta_2}) \tag{2}$$

- $\boldsymbol{x} = (\alpha, f^j(), \beta_k, \epsilon_i,)$ ($\boldsymbol{\eta}$ a linear combination of $\boldsymbol{x}$)
- $\{f^j()\} \sim \mathcal{N}(0, Q_f^{-1}(\tau_2))$:
  non-linear effects of the covariates $\boldsymbol{u}$ with unknown functions (structured random effects)
- $\beta_k \sim \mathcal{N}(0, \tau_1^{-1})$:
  linear effect of covariates $\boldsymbol{z}$ (fixed effects)
- ϵ_i:
  unstructured random effect of observation i.
- $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$:
  hyperparameters (likelihood and latent field)

The inla function

inla(): Performs Bayesian analysis of additive models

```
> result<- inla(
          formula, # This describes your latent field
          family ="gaussian", #The likelihood distribution
          data = dat # A list or dataframe
          #This is all that is needed for a basic call

          verbose = TRUE, #shows the steps of the process
          keep = FALSE, #keeps the output)

          #Then there are some "control statements"
          #that allow you to customise some things
          control.pedictor = list(A = ObservationMatrix)
          )...
```

Likelihood functions

- "gaussian" item "gev"
- "binomial"
- "coxph"
- "Exponential"
- "laplace"
- "sn" (Skew Normal)
- "stocvol", "stochvol.nig","stochvol.t"
- "T"
- "weibull"
- Many others: go to $\texttt{http}://\texttt{r}-\texttt{inla.org}$

Control statements

The `control.xxx` statements control various parts of the INLA program

- `control.compute`
  - `dic`, `mlik`, `cpo` – Compute measures of fit
- `control.inla`
  - `strategy` and `int.strategy` contain useful advanced features, dz=1, diff.logdens=2.5, h=0.01
- `control.family`
  - `hyper = list(prior = "", param = c(, ), fixed = FALSE)` control variance of the likelihood
- `control.mode`
  - `x`, θ, `result`– Gives modes to INLA
  - `restart = TRUE`– Tells INLA to try to improve on the supplied mode

Various other – see help!

Using different link functions

INLA typically implements the canonical link functions, which, in this case is the logit link. Sometimes, you want other things

- control.data = list(link = "logit")
- control.data = list(link = "probit")
- control.data = list(link = "cloglog")

`formula`: specifying the latent field

The latent field is specified using the "standard" `R` method
$\texttt{formula} = \texttt{y} \sim 1 + \texttt{covariate}(\beta) + \texttt{f}(...)$.

- `y` is the name of your data in the data frame
- An intercept is fitted *automatically!* Use -1 in your formula to avoid it
- The fixed effects (covariates) are taken as i.i.d. normal with a common prior (this can be changed)
- The `f()` function contains the (non-linear) random effect specifications

`formula`: specifying the latent field

The latent field is specified using the "standard" `R` method
$\texttt{formula} = \texttt{y} \sim 1 + \texttt{covariate}(\beta) + \texttt{f}(\ldots)$.

- `y` is the name of your data in the data frame
- An intercept is fitted *automatically!* Use -1 in your formula to avoid it
- The fixed effects (covariates) are taken as i.i.d. normal with a common prior (this can be changed)
- The `f()` function contains the (non-linear) random effect specifications

`formula`: specifying the latent field

The latent field is specified using the "standard" `R` method $\texttt{formula} = \texttt{y} \sim 1 + \texttt{covariate}(\beta) + \texttt{f}(...)$.

- `y` is the name of your data in the data frame
- An intercept is fitted *automatically!* Use -1 in your formula to avoid it
- The fixed effects (covariates) are taken as i.i.d. normal with a common prior (this can be changed)
- The `f()` function contains the (non-linear) random effect specifications

`formula`: specifying the latent field

The latent field is specified using the "standard" `R` method
$\texttt{formula} = \texttt{y} \sim 1 + \texttt{covariate}(\beta) + \texttt{f}(...)$.

- `y` is the name of your data in the data frame
- An intercept is fitted *automatically!* Use -1 in your formula to avoid it
- The fixed effects (covariates) are taken as i.i.d. normal with a common prior (this can be changed)
- The `f()` function contains the (non-linear) random effect specifications

Specifying random effects

Random effects are added to the formula through the function
`f(name, model = "...", hyper = ..., replicate = ..., constr = FALSE, cyclic = FALSE)`

- `name`– the name of the random effect. Also refers to the values on the data which are used for various things
- `model`– the latent model. E.g. "iid", "generic0", "rw2, "ar1", etc
- `hyper`– specify the prior on the hyperparametrs
- `replicate`– for replicates
- `constr`– Sum to zero contraints?
- `cyclic`– Are you cyclic (RW1, RW2 and AR1)

`formula`: specifying the latent field - `f()`

The implemented models are

- `iid`: random effects
- `rw1`, `rw2`, `ar1`: smooth effect of covariates or time effect
- `seasonal`: seasonal effect
- `besag`: spatial effect (CAR model)
- `generic`: user defined precision matrix

Additional functions of the INLA package

- `summary()`: produces a summary of the main results from a fitted model
- `plot()`: produces some plots from the fitted model
- \$ *tab*: all the results/output

Internal scale-Input/Output

Internal scale:
the precision parameter τ is represented as

$$\theta = \log \tau \tag{3}$$

and the prior is defined on θ.

Hyperparameter spesication and default values

hyper

theta

name log precision

short.name prec

prior loggamma

param 1 5e-05

initial 4

fixed FALSE

to.theta function(x) log(x)

from.theta function(x) exp(x)

Internal scale-Input/Output

The variances (σ_a^2) is the inverse of the precision (τ)

```
##transform the marginal
marg=inla.tmarginal(function(x) 1/x, model$marginals.hyperp
"Precision for ...")

##Find he expectation and quantiles from transformed margi
inla.emarginal(function(x) x,marg)
inla.qmarginal(c(0.025,0.5,0.975), marg)

##Find DIC
DIC=model$dic$dic
```

Examples from house sparrow dataset

- Repeated measures-individual random effect
- Non-linear effects
- Maternal effects, covariance
- (Multivariate)

Computer exercises 2

Exercises:

- Look at the output from `animal.inla()` - run inla call and formula in `R-INLA`
- Run with multiple measurements data, find individual random effect, age as RW1 model
- Run different models for maternal effect
  - Look at the summary (plot)
  - Transform the marginals of hyper-parameter to get the variances
  - Take out variables and compare DIC values
  - Find individual breeding values and summary of breeding values (Hint: `model$marginals.random$name[1]`, `model$summary.random$names[1]`)
  - +++