Bayesian Hierarchical Models

Gavin Shaddick,  
Millie Green, Matthew Thomas  
University of Bath

6th - 9th December 2016
IMPLEMENTING BAYESIAN MODELS USING R-INLA
OUTLINE

R and RStudio

Statistical Analyses

Packages

R-INLA
R and RStudio
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
Statistical Analyses
**Statistical Analysis in R**

- R comes with many statistical tools already installed
  - descriptive statistics
  - visualisation
  - statistical tests
  - model fitting.
CAN R DO MORE?

- The default installation of R has a comprehensive set of tools for statistical analyses.
- To meet the specific needs of data scientists, many other statistical tools are readily available in the form of packages.
- Packages are collections of functions and data.
- "During the last decade, the momentum coming from both academia and industry has lifted R to become the single most important tool for computational statistics, visualisation and data science."
R Packages: Examples used in this course

- ggplot2
- raster
- Rmisc
- mgcv
- maptools
- ... many many more!!
A list of R Packages can be seen and downloaded from https://cran.r-project.org

Available CRAN Packages By Name

A3
abbyyR
abc
ABCAnalysis
abc.data
abcdeFBA
ABCoptim
ABCp2
ABC.RAP
abcrf
abctools
abdd
abf2
ABHgenotypeR
abind
abdn
abodOutlier
AbsFilterGSEA
abundant

Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
Access to Abbyy Optical Character Recognition (OCR) API
Tools for Approximate Bayesian Computation (ABC)
Computed ABC Analysis
Data Only: Tools for Approximate Bayesian Computation (ABC)
ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package
Implementation of Artificial Bee Colony (ABC) Optimization
Approximate Bayesian Computational Model for Estimating P2
Array Based CpG Region Analysis Pipeline
Approximate Bayesian Computation via Random Forests
Tools for ABC Analyses
The Analysis of Biological Data
Load Gap-Free Axon ABF2 Files
Easy Visualization of ABH Genotypes
Combine Multidimensional Arrays
Modelling Multivariate Data with Additive Bayesian Networks
Angle-Based Outlier Detection
Improved False Positive Control of Gene-Permuting GSEA with Absolute Filtering
Abundant regression and high-dimensional principal fitted components
R-INLA
R-INLA

The R–INLA package provides a practical implementation of Integrated Nested Laplace Approximations (INLA).

The class of models that can be expressed in this form and thus can be used with R–INLA is very large and includes, amongst others, the following:

- Dynamic linear models.
- Stochastic volatility models.
- Generalised linear (mixed) models.
- Generalised additive (mixed) models.
- Spline smoothing.
- Semi-parametric regression.
- Disease mapping.
- Log-Gaussian Cox-processes.
- Model-based geostatistics.
- Spatio-temporal models.
- Survival analysis.
The syntax of R–INLA

- There are three main parts to fitting a model using R–INLA:
  1. The data.
  2. Defining the model formula.
  3. The call to the INLA program.

- The basic syntax of running models in R–INLA is very similar in appearance to that of `glm` in R and takes the general form `formula, data, family` but with the addition of the specification of the nature of the random effects, `f()`.

- For the latter component, common examples include
  - `f(i, model="iid")` (independent)
  - `f(i, model="rw")` (random walk of order one)
  - `f(i, model="ar")` (autoregressive of order p).
Fitting a Poisson regression model in R–INLA

- An extension of the standard Poisson model to include log–normal random effects in the linear predictor

\[ \log \mu_l = \beta_0 + \beta_{0i} + \beta_1 X_l + \beta_d X_l + \epsilon_l \quad (1) \]

where \( \beta_l \) represents the effect of exposure, \( \beta_d \) is the effect of an area-level covariate and \( \beta_{0i} \) denotes the random effect for area \( i \).

- The syntax of the R–INLA code to fit this model is very similar to that of a standard \texttt{glm} in R.
Fitting a Poisson regression model in R–INLA

```r
> formula = Y ~ X1+X2 + f(i, model="iid")
> model = inla(formula, family="poisson", data=data)

Call:
"inla(formula = formula, family = "poisson", data = data)"

Time used:
Pre-processing          Running inla Post-processing        Total
          0.278389          0.286911            0.125699          0.690999

Integration Strategy: Central Composite Design

Model contains 1 hyperparameters
The model contains 3 fixed effect (including a possible intercept)

Likelihood model: poisson

The model has 1 random effects:
1. 'i' is a IID model
Fitting a Poisson regression model in R–INLA

```r
> summary(model)

Call:
"inla(formula = formula, family = "poisson", data = data)"

Time used:
Pre-processing Running inla Post-processing Total
0.2784 0.2869 0.1257 0.6910

Fixed effects:
    mean   sd 0.025quant 0.5quant 0.975quant
```
R and RStudio Statistical Analyses Packages

**Fitting a Poisson regression model in R-INLA**

\[
\begin{align*}
\text{(Intercept)} & \quad 2.4960 & 0.0713 & 2.3553 & 2.4962 & 2.6355 \\
X1 & \quad 0.1187 & 0.0310 & 0.0578 & 0.1186 & 0.1796 \\
X2 & \quad 0.0578 & 0.0074 & 0.0433 & 0.0578 & 0.0722
\end{align*}
\]

**Random effects:**

<table>
<thead>
<tr>
<th>Name</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>IID model</td>
</tr>
</tbody>
</table>

**Model hyperparameters:**

<table>
<thead>
<tr>
<th>Precision for i</th>
<th>mean</th>
<th>sd</th>
<th>0.025quant</th>
<th>0.5quant</th>
<th>0.975quant</th>
<th>4.525</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.784</td>
<td>0.3548</td>
<td>3.131</td>
<td>3.769</td>
<td>4.525</td>
<td></td>
</tr>
</tbody>
</table>

**Expected number of effective parameters (std dev):**

321.42(3.926)

**Number of equivalent replicates:** 1.223

**Marginal Likelihood:** -1513.92
Fitting models in R–INLA

Future details can be found on the R–INLA webpage: