INLA for GMRFs (e.g. GLMMs, Spatial Models) with Spatial Examples of Leukemia Cases and Heavy Metal Concentrations"

Frances Lin

June 2022

Background and Introduction

The steps involving the Bayesian inference may appear easy and straightforward:

- ▶ updating prior beliefs about the unknown parameters and
- ▶ obtaining the posterior distribution for the parameters.

However, this is much harder to do in practice since solutions in closed-form may not always be determined.

- ▶ MCMC (Markov chain Monte Carlo) was introduced and represented a breakthrough in Bayesian inference in the early 1990s.
- ▶ Tools such as `WinBugs` (Spiegelhalter et al., 1995), `JAGS` (Plummer, 2016), and `stan` (Stan Development Team, 2015) have also been developed, and
- ▶ Bayesian statistics has gained popularity in many fields.

Background and Introduction

However, MCMC methods

- ▶ not only can be computationally demanding (i.e. requires a large amount of CPU),
- ▶ but also present convergence issues.

INLA (integrated nested Laplace approximation) is a fast alternative to MCMC for Bayesian inference. INLA

- ▶ can be applied to a very flexible class of models named LGMs (latent Gaussian models), which ranges from GLMMs (generalized linear mixed models) to time-series, and spatial and spatio-temporal models.
- ▶ allows for faster and more accurate inference without trading speed for accuracy, and
- ▶ is accessible through the **R** package R-INLA.

Applications

INLA have found spatial or spatio-temporal applications in a wide variety of fields such as environment, ecology, disease mapping, public health, cancer research, energy, economics, risk analysis, etc.

Selected examples include:

- ▶ polio-virus eradication in Pakistan (Mercer et al., 2017);
- ▶ socio-demographic and geographic impact of HPV vaccination (Rutten et al., 2017);
- ▶ topsoil metals and cancer mortality (Lopez-Abente et al., 2017);
- ▶ probabilistic prediction of wind power (Lenzi et al., 2017);
- ▶ applications in spatial econometrics (Bivand et al., 2014; Gomez-Rubio et al., 2015; Gomez-Rubio et al., 2014);
- ▶ predicting extreme rainfall events in space and time (Opitz et al., 2018), etc.

Outline

- ▶ Key Components
  - 0. Bayesian Inference
  - 1. Latent Gaussian Models
  - 2. Additive Models
  - 3. Gaussian Markov Random Fields
  - 4. Additive Models and Gaussian Markov Random Fields
  - 5. Laplace Approximations
- ▶ INLA
- ▶ INLA-SPDE (Stochastic Partial Differential Equations) Approach
- ▶ Discussion
- ▶ Spatial Examples using the package R-INLA
  - 1. Leukemia Incident Cases (Areal data)
  - 2. Heavy Metal Concentrations (Geostatistical data)

0. Bayesian Inference

The posterior distribution is proportional to the likelihood function multiples by the prior distribution

$$f(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta} \propto p(y|\theta)p(\theta),$$

where $p(y|\theta)$ is the likelihood function, $p(\theta)$ is the prior, and $\int p(y|\theta)p(\theta)d\theta$ is the normalizing constant.

▶ Based on the posterior distribution, relevant statistics for the parameters of interest (e.g. marginal distribution, means, variances, and credibility intervals) can be obtained.

▶ However, the integral is generally intractable in closed-form, thus requiring the use of numerical methods such as MCMC.

1. Latent Gaussian Models

The latent Gaussian models (LGMs) is a class of three-stage Bayesian hierarchical models. It involves the following stages:

1. Observations y is assumed to be conditionally independent, given a latent Gaussian random field x and hyperparameter θ_1

$$y|x, \theta_1 \sim \prod_{i \in I} p(y_i|x_i, \theta_1). \qquad \text{\textit{likelihood}}$$

2. The latent field $x|\theta_2$ is assumed to be a GMRF (Gaussian Markov random field) with a sparse precision matrix Q

$$x|\theta_2 \sim p(x|\theta_2) = N(\mu(\theta_2), Q^{-1}(\theta_2)), \qquad \text{\textit{latent field}}$$

where $Q = \Sigma^{-1}$ is the precision matrix and θ_2 is a hyperparameter.

3. The hyperparameters of the latent field that are not necessarily Gaussian are assumed to follow a prior distribution

$$\theta = (\theta_1, \theta_2) \sim p(\theta), \qquad \text{\textit{hyperpriors}}$$

where $p(\cdot)$ is a known distribution.

1. Latent Gaussian Models

Then, the posterior distribution, structured in a hierarchical way, becomes

$$p(x, \theta | y) \propto p(y | x, \theta) p(x, \theta)$$
$$\propto \prod_{i \in I} p(y_i | x_i, \theta) p(x | \theta) p(\theta).$$

For computational reasons and to ensure accurate approximations, the following assumptions hold:

1. Each observation y_i depends only on one component of the latent field x_i, and most components of x will not be observed.

2. The distribution of the latent field x is Gaussian and is close to a Gaussian Markov random field (GMRF) when the *dim* of n is high (10^3 to 10^5).

3. The number of hyperparameters θ is small (~ 2 to 5 but < 20).

2. Additive Models

LGMs (latent Gaussian models) is an umbrella class that generalizes the large number of related variants of additive and/or generalized linear models.

Applications of LGMs include, for example,

- regression models (e.g. GAMs/GLMs/GAMMs/GLMMs/++),
- dynamic models, and
- spatial (e.g. Gaussian and BYM models) and spatio-temporal models.

Consider the Bayesian structured additive model setup, for example,

$$y \sim \prod_{i}^{N} p(y_i|x_i, \theta),$$

2. Additive Models

then the mean μ_i (for observation y_i) can be linked to the linear predictor η_i through a link function g

$$\eta_i = g(\mu_i) = \alpha + \sum_j \beta_j z_{ji} + \sum_k f_k(w_{ki}) + \varepsilon_i,$$

where

- ▶ α is the overall intercept,
- ▶ β are linear effects of fixed covariates z,
- ▶ $\{f_k\}$, which are used to represent specific Gaussian processes, are nonlinear/smooth effects of some covariates w, and
- ▶ ε are iid random effects.

GLMs (generalized linear models) is a special case with the expression $\alpha + \sum_j \beta_j z_j$ (i.e. $f(\cdot) = 0$). GAMs (generalized additive models) is another special case with the expression $\alpha + \sum_k f_k(w_k)$.

2. Additive Models

The model is a LGM *iff* the joint distribution of the latent field

$$x = (\eta, \alpha, \beta, f(\cdot))$$

is Gaussian. I.e.

$$x|\theta = (\eta, \alpha, \beta, f(\cdot))|\theta \sim N(\mu(\theta), Q^{-1}(\theta)).$$

This can be achieved by assigning Gaussian priors to all terms (the intercept and the parameter of the fixed effects) in x.

If we further assume conditional independence of x, then this latent field x is a Gaussian Markov random field.

3. Gaussian Markov Random Fields

A GMRF (Gaussian Markov Random Field) is a random vector that follows a multivariate normal distribution with additional conditional independence properties:

- for $i \neq j$, x_i and x_j are conditionally independent, given the remaining elements x_{-ij}.

Undirected graphs G are typically used to represent the conditional independence properties of the GMRF. An undirected graph G consists of a set of nodes V and edges E

$$G = (V, E),$$

where V is a set of nodes $\{1, ..., n\}$ and E is a set of edges $\{i, j\}$, where $i \neq j \in V$.

3. Gaussian Markov Random Fields

Let x be a GMRF with respect to a graph $G = (V, E)$, then it is equivalent to say that x_i and x_j are conditionally independent, given the remaining elements x_{-ij}

$$x_i \perp x_j | x_{-ij} \quad if \quad i, j \in E, i \neq j,$$

where $-ij$ refers to all elements other than i and j. This is referred to as the pairwise Markov property.

Equivalent properties include the local Markov property and global Markov property.

3. Gaussian Markov Random Fields

The Markov assumption in the GMRFs results in a sparse precision matrix. When a matrix is sparse (with lots of elements $= 0$), the computational cost tends also to be lower, allowing for much faster computation.
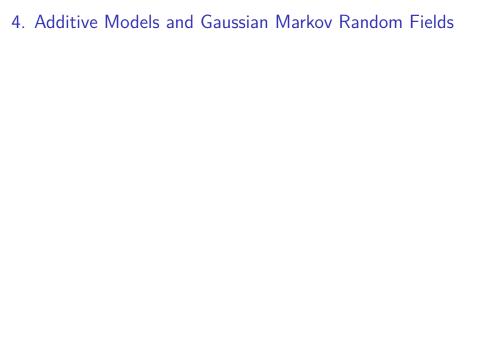
Recall that $x \sim N(0, Q = \Sigma^{-1})$ and

$$x_i \perp x_j \iff \Sigma_{ij} = 0,$$
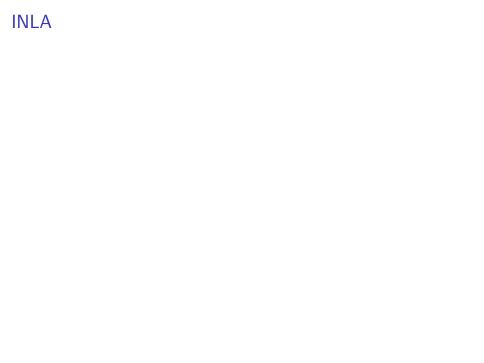
where Σ is the covariance matrix. For Σ to be sparse requires the marginal independence assumption, which can be unreasonable. On the other hand, it can be shown that

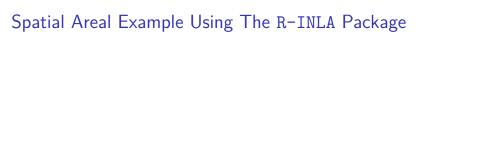$$x_i \perp x_j | x_{-ij} \iff Q_{ij} = 0,$$

where Q is the precision matrix ($Q = \Sigma^{-1}$), and conditional independence is a more reasonable assumption and their properties are encoded in the precision matrix.
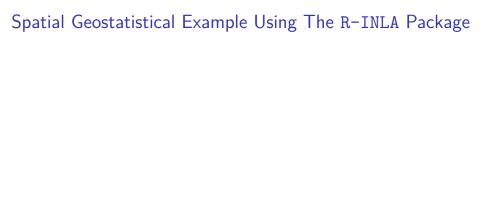
4. Additive Models and Gaussian Markov Random Fields

5. Laplace Approximations

INLA

INLA-SPDE (Stochastic Partial Differential Equations) Approach

Spatial Areal Example Using The R-INLA Package

Spatial Geostatistical Example Using The R-INLA Package

Thank you!

INLA for GMRFs (e.g. GLMMs, Spatial Models) with Spatial Examples of Leukemia Cases and Heavy Metal Concentrations

Frances Lin

PhD student, Dept. of Statistics, Oregon State University

Full report, code, etc. are available at

GitHub: franceslinyc, INLA-with-Spatial-data-2022