INLA for GMRFs (e.g. GLMMs, Spatial Models) with An Example of Leukemia Cases

Frances Lin

June 2022

1. Introduction

1.1 Background and Introduction

The steps involving the Bayesian inference may appear easy and straightforward: updating prior beliefs about the unknown parameters with observed data 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.

The simulation-based inference through the idea of MCMC (Markov chain Monte Carlo) was introduced and represented a breakthrough in Bayesian inference (Robert & Casella, 1999) in the early 1990s. MCMC tools such as `WinBugs` (Spiegelhalter et al., 1995), `JAGS` (Plummer, 2016), and `stan` (Stan Development Team, 2015) have also been developed. Bayesian statistics has gained popularity in many fields. However, these MCMC methods, based on sampling, not only are 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 that does not require sampling. INLA can be applied to a very wide and flexible class of models named LGMs (latent Gaussian models), which ranges from GLMMs (generalized linear mixed models) to time-series, spatial and spatio-temporal models. INLA also allows for faster and more accurate inference without trading speed for accuracy, and it is accessible through the **R** package `R-INLA`.

1.2 Applications

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

Some selected examples include: environmental risk factors to liver fluke in cattle (Innocent et al., 2017); modelling recovering fish populations (Boudreau et al., 2017); polio-virus eradication in Pakistan (Mercer et al., 2017); cortical surface fMRI data (Mejia 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) with spatially misaligned data; ethanol and gasoline pricing (Laurini, 2017); applications in spatial econometrics (Bivand et al., 2014; Gomez-Rubio et al., 2015; Gomez-Rubio et al., 2014); probabilistic prediction of wind power (Lenzi et al., 2017); modeling landslides as point processes (Lombardo et al., 2018); predicting extreme rainfall events in space and time (Opitz et al., 2018), etc.

2. Key Components

2.0. 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 such as marginal distribution, means, variances, quantiles, credibility intervals, etc. can be obtained.

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

2.1. Latent Gaussian Models

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

In the first stage, observations (or data) y is assumed to be conditionally independent, given a latent Gaussian random field x (joint distribution of all parameters in the linear predictor) and hyperparameter θ_1

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

In the second stage, the latent field $x|\theta_2$ is assumed to be a GMRF (Gaussian Markov random field) with a sparse precision matrix Q

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

where $Q = \Sigma^{-1}$ is the precision matrix and θ_2 is a hyperparameter. The versatility of the model class can be specified through the unobserved latent field. The latent field includes all random terms and captures the underlying dependence structure of the data. Latent field can be, for example, covariates, unstructured random effects (e.g. white noise), structure random effects (e.g. temporal dependency, spatial dependency, smoothness terms) (Bolin, 2015).

In the last stage, the hyperparameters of the latent field that are not necessarily Gaussian are assumed to follow a prior distribution

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

where $p(\cdot)$ is a known distribution. The hyperparameters control the likelihood for the data and/or the latent Gaussian field and can be used to account for variability and strength of dependence. Hyperparameters can be, for example, variance of observation noise, variance of the unstructured random field, range of a structured random effect (Bolin, 2015).

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 ($\sim$ 2 to 5 but < 20).

2.2. Additive Models

LGMs (latent Gaussian models) is an umbrella class that generalizes the large number of related additive and/or generalized linear models. Consider the generalized linear mixed model setup, for example,

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

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. The model components f_k are what make LGMs flexible. Examples of f_k include spatially or temporally correlated effects, smoothing and stochastic spline, measure errors, and random effects with different types of correlations.

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 $\alpha + \sum_k f_k(w_k)$.

The model is a LGM iff the joint distribution of

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

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.

2.3. 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}.

More specifically, 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 $i \neq j \in V$ (Bolin, 2015).

Formally, a random vector $x = (x_1, ..., x_n)^T \in \mathbb{R}$ is called a GMRF with respect to a labelled graph $G = (V, E)$ with mean μ and positive definite matrix Q iff its density has the form

$$p(x) = (2\pi)^{-n/2}|Q|^{1/2} exp(-\frac{1}{2}(x-\mu)^T Q(x-\mu))$$

and

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

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}

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

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.

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, but this can be an unreasonable assumption. 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 (the inverse of the covariance matrix), and conditional independence is a more reasonable assumption and their properties are encoded in the precision matrix (Rue & Held, 2005).

2.4. Additive Models and Gaussian Markov Random Fields

2.5. Laplace Approximations

3. INLA

3.1. INLA

3.2. INLA-SPDE (Stochastic Partial Differential Equations) Approach

4. Discussion

5. Spatial Examples

Two data sets are considered. The `NY8` data set is areal data.

. . . . . . is point-referenced (or geostatistical) data.

Since the focus of this part of the project is on the use of `R-INLA` package, the details of the models used for each data set are briefly introduced here.

5.1. A Spatial (Areal) Example of Leukemia Cases

The `NY8` data set contains the number of incident leukemia cases per census tract in an eight-country region of upstate New York from 1978-1982 (Waller & Gotway, 2004; Bivand et al., 2008). The `NY8` data set can be accessed from the **R** package `DClusterm`, and it is a `SpatialPolygonsDataFrame` object.

For this data set, a total of 5 models (fixed effects, random effects (iid), ICAR, BYM and Leroux et al.) are fitted, and results include criteria for model selection (marginal log-likelihood, DIC and WAIC) and plots of predicted values.

Since the number of incident leukemia cases is count, Poisson GLMs with fixed effects and random effects are fitted. Since there is spatial dependence in the data, spatial models (GLMs with spatial random effects) such as ICAR (Intrinsic Conditional autoregressive), BYM and Leroux et al. model are also considered.

Without going into details, recall that the GLMs have the following form

$$Y = X\beta + Z\alpha + \varepsilon,$$

where β is a vector of fixed effects with design matrix X, α is a vector of random effects with design matrix Z, and ε is an error term, where it is assumed that $\varepsilon_i \sim N(0, \sigma^2), i = 1, ..., n$. The vector of random effects α is modeled as MVN (it is assumed that)

$$\alpha \sim N(0, \sigma_\alpha^2 \Sigma),$$

where the covariance matrix Σ is defined such that it induces higher correlation with adjacent areas.

There are a few ways to include spatial dependence in Σ, and in spatial areal model especially, it is more common to model the precision matrix Q directly, where $Q = \Sigma^{-1}$. In ICAR (Intrinsic CAR), $\Sigma^{-1} = diag(n_i) - W$, where n_i is the number of neighbors of area i. In Leroux et al.'s model (mixture of matrices), $\Sigma^{-1} = ((1-\lambda)I_n + \lambda M), \lambda \in (0, 1)$, where M is precision of intrinsic CAR specification. The BYM (Besag, York and Mollié) model includes two latent random effects: an ICAR latent effect and a Gaussian iid latent effect.

Results show that for spatially dependent data, spatial models generally perform better than GLM with fixed or random (iid) effects. On the other hand, it is no surprise that the baseline model (fixed effects model) appears to be the poorest fit of all.

	Marg_logLik	DIC	WAIC
fixed	-514.4	1016	1017
iid	-512.1	979.2	983.6
ICAR	-718.9	968.3	972.2
BYM	-458.9	967.4	971.9
Leroux	-508.3	967.3	971.3

5.2. A Spatial (Geostatistical) Example of Heavy Metal Concentrations

Reference

Bakka, H., Rue, H., Fuglstad, G. A., Riebler, A., Bolin, D., Illian, J., Krainski, E., & Lindgren, F. (2018). Spatial modeling with R-INLA: A review. Wiley Interdisciplinary Reviews: Computational Statistics, 10(6), e1443.

Bolin, D. (2015). *Lecture 1: Introduction Gaussian Markov random fields.* Personal Collection of D. Bolin, Chalmers University of Technology, Gothenburg Sweden.

Bolin, D. (2015). *Lecture 2: Definitions and properties Gaussian Markov random fields.* Personal Collection of D. Bolin, Chalmers University of Technology, Gothenburg Sweden.

Morrison, K. (2017). A gentle INLA tutorial. Precision Analytics. https://www.precision-analytics.ca/articles/a-gentle-inla-tutorial/.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with INLA: a review. Annual Review of Statistics and Its Application, 4, 395-421.

Rue, H., & Held, L. (2005). Gaussian Markov random fields: theory and applications. Chapman and Hall/CRC.