Reparameterized SGLMM (Spatial Generalized Linear Mixed Model): A Review

Frances Lin

June 2022

Background and Introduction

The SGLMM (spatial generalized linear mixed model)

- is a hierarchical model that introduces spatial dependence through a GMRF (Gaussian Markov random field) (Besag et al., 1991).
- was initially proposed for prediction for count data, but
- has later been applied to estimation and prediction for other types of data (e.g. binary) and
- found applications in many fields (e.g. ecology, geology, epidemiology, image analysis, and forestry).

Background and Introduction

The SGLMM has been the dominant model for areal data because of

- its flexible hierarchical specification,
- the availability of the software `WinBUGS` for (Bayesian) data analysis (Lunn et al., 2000) and
- various theoretical and computational advantages over the competitive model named automodel.

However, SGLMMs suffer from two major shortcomings:

i) variance inflation due to spatial confounding and
ii) computational challenges posed by high dimensional latent variables (random effects).

Background and Introduction

On the other hand, while another model named RHS model

- seeks to alleviate confounding (Reich et al., 2006),
- can result in random-effects structure with negative spatial dependence (i.e. repulsion) that is not typically applied in practice.

A reparameterized model is proposed, and it is able to

- alleviate confounding, and, at the same time,
- include patterns of positive spatial dependence (i.e. attraction) while excluding patterns of repulsion.

The proposed model is also one of the first dimension reduction techniques for spatial areal models.

Outline

- Traditional SGLMM (Spatial Generalized Linear Mixed Model)
- Spatial confounding
- Sparse reparameterization of the areal SGLMM
-

Traditional SGLMM

Let

$$G = (V, E)$$

be an undirected, labelled graph, where $V = \{1, 2, ..., n\}$ is a set of vertices (nodes) and $E = \{i, j\}$ is a set of edges, where $i, j \in V$, $i \neq j$.

- Each vertex represents an area of interest and each edge represents the proximity of areas i and j.
- G is represented using an adjacency matrix A, which is a nxn matrix with $diag(A) = 0$ and entries $A_{ij} = 1\{(i, j) \in E, i \neq j\}$, where $1(\cdot)$ is an indicator function.

Traditional SGLMM

Further let $Z = (Z_1, ...Z_n)^T$ be the random field of interest, where Z_i is the random variable associated with vertex i. Then, the first stage of the model is given by

$$g(E(Z_i|\beta, W_i)) = X_i\beta + Wi, \tag{1}$$

where

- g is a link function,
- X_i is the ith row of the design matrix X,
- β is a p-vector of regression parameters and
- W_i is a spatial random effect associated with vertex i.

Different types of data require different canonical choices of the link function g (e.g. the logit function for spatial binary data and the logarithm function for spatial count data).

GMRF prior for the random effects

The field of random effects $W = (W_1, ..., W_n)^T$, through which spatial dependence is incorporated, is assumed to follow the intrinsic conditionally autoregressive or GMRF prior

$$p(W|\tau) \propto \tau^{rank(Q)/2} exp(-\frac{\tau}{2} W^T QW), \qquad (2)$$

where τ is a smoothing parameter and $Q = diag(A1) - A$ is a precision matrix. The precision matrix Q incorporates both dependence and prior uncertainty.

Note. A SGLMM

1. can be reformatted by replacing GMRF with GP (Gaussian process) for point-referenced data or geostatistical data.
2. is restricted to a Bayesian or restricted maximum likelihood (REML) analysis since the prior (2) is improper.

Spatial confounding