Progress this week
November 24, 2019

- Added some keywords.
- Turned §1 and §1.1 into complete sentences.
- Added some orientation text to §2.
- Turned §2.1 into complete sentences.
- Added Figure 1 illustrating the triangular mesh representation.
- Added Figure 2 illustrating the barycentric change-of-basis.

Previous word count: —
Previous page count: —

Current word count: 1032
Current page count: 3

REVIEW ARTICLE

The Integrated Nested Laplace Approximation applied to Spatial Point Process Models

Kenneth Flagg and Andrew Hoegh

Montana State University, Bozeman, MT

ARTICLE HISTORY
Compiled November 24, 2019

ABSTRACT
This template is for authors who are preparing a manuscript for a Taylor & Francis journal using the LaTeX document preparation system and the `interact` class file, which is available via selected journals' home pages on the Taylor & Francis website.

KEYWORDS
INLA, spatial prediction, log-Gaussian Cox process, spatial point process

1. Introduction

Spatial prediction is a high-dimensional inference problem. When the goal of statistical modeling is to produce a graphical map of a random variable over space, the model ultimately must be able to predict that random variable at every pixel of the image. A map image will typically be at least several hundred by several hundred pixels, so in total there can easily be hundreds of thousands of pixels requiring predictions. Thus, even when a model has only half a dozen parameters, it may include hundreds of thousands of latent variables.

Spatial point process models further complicate the situation with difficult likelihoods. *(Cite some computational papers — Baddely?)* Both maximum likelihood and Bayesian model fitting require integrating the intensity function over space, but the integral is generally not available in closed form. Many methods have been introduced including quadrature-based approximations *(cite Baddeley)*, pseudodata approaches *(cite Baddeley/Berman/Turner etc)*, and Markov chain Monte Carlo [5].

Development of the integrated nested Laplace approximation (INLA) has made accurate approximate model fitting considerably more feasible for a particular class of log-Gaussian Cox process (LGCP) models. INLA was developed to fit Bayesian hierarchical models with many latent Gaussian variables [7]. A key part of INLA's computational simplicity is that it calculates the posterior distribution of each latent Gaussian variable one at a time; that is, it provides only the posterior marginal distributions rather than the full joint distribution.

When using a LGCP for spatial mapping, two aspects make INLA a suitable approach. First, the LGCP is driven by a spatial Gaussian process (GP), so the latent variables are Gaussian. Second, even though the latent variables are expected to ex-

CONTACT Kenneth Flagg. Email: kenneth.flagg@montana.edu

hibit spatial dependence, their full joint distribution is not needed. In most situations it suffices to map their predicted values, variance, and upper and lower interval bounds pointwise across space.

This article provides a review of recent advances in the fitting of LGCP models via INLA, including dimension reduction by triangulation, a likelihood factorization that avoids gridding, and incorporation of sampling or false negatives.

1.1. Log-Gaussian Cox Process

The LGCP is a Poisson process driven by a latent Gaussian process [6]. A Poisson process is characterized entirely by its intensity function $\lambda(\mathbf{s})$, which gives the mean number of events per unit area, and the process satisfies the following two properties *(find a good citation)*.

(1) The number of events in a region $\mathcal{S}$ follows a Poisson distribution with mean $\int_{\mathcal{S}} \lambda(\mathbf{s})\mathrm{d}\mathbf{s}$.
(2) The numbers of events in disjoint regions are independent.

Commonly, covariates are incorporated via a log-linear model for the intensity,

$$\log \lambda(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta}.$$

The LGCP adds another stochastic layer,

$$\log \lambda(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + Z(\mathbf{s}),$$

where Z is a Gaussian process.

Where the spatial Poisson process has a (stochastically) fixed intensity to be estimated from data, the spatial LGCP induces a hierarchical model where the intensity function is itself a random process to be predicted.

1.2. Integrated Nested Laplace Approximation

INLA fast approximation for marginals useful for mapping [7]
Integrated Nested Laplace Approximation
Bayesian Hierarchical models, many latent Gaussian variables, few parameters
Laplace approximation in general: $\int \exp[h(x)]\mathrm{d}x$, Taylor expansion of $h(x)$
Example from [1]
- $\mathbf{y} = (y_1, \ldots, y_n)'$ independent Gaussian observations
- $y_i \sim \mathrm{N}(\theta, \sigma^2)$
- $\theta \sim \mathrm{N}(\mu_0, \sigma_0^2)$
- $\psi = 1/\sigma^2$, $\psi \sim \mathrm{Gamma}(a, b)$
- The posterior distribution of ψ:

$$p(\psi|\mathbf{y}) \propto \frac{p(\mathbf{y}|\theta, \psi)p(\theta)p(\psi)}{p(\theta|\psi, \mathbf{y})}$$

- Laplace approximation:

$$\tilde{p}(\psi|\mathbf{y}) \propto \frac{p(\mathbf{y}|\theta, \psi)p(\theta)p(\psi)}{\tilde{p}_G(\theta^*|\psi, \mathbf{y})}$$

Repeat for θ, will depend on ψ
Provides marginal posteror for one entry at a time of a vector $\boldsymbol{\theta}$
Available in R through the `R-INLA` package [3].

2. Methodology

INLA provides an efficient computational framework for fitting Bayesian models with latent Gaussian variables. On top of this framework are built several tools to further simplify the fitting of spatial LGCP models. The stochastic partial differential equation (SPDE) approach provides dimension reduction for the spatial GP [4]. The SPDE approach employs a numerical integration scheme which can also be used to approximate the LGCP likelihood and negate the need to grid the events into Poisson counts [8]. With these computational improvements, researchers are now able to efficiently fit LGCP models to incompletely-observed point patterns [9].

2.1. The SPDE Approach

Because the LGCP includes a Gaussian process (GP), efficient computation for Gaussian processes is critical when working with LGCP models.

The GP imposes a dense covariance matrix on the latent variables [1]. For GPs with a Matérn covariance function, a Gaussian Markov random field (GMRF) approximation can simplify computation, requiring only a sparse covariance structure.

The GMRF approximation is motivated by the fact that Gaussian fields with Matérn covariances are solutions to the below stochastic partial differential equation (SPDE) [4].

$$(\kappa^2 - \Delta)^{\alpha/2} Z(\mathbf{s}) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d, \kappa > 0, \alpha = \nu + d/2, \nu > 0$$

Here, $\mathcal{W}$ is a Gaussian white noise process with variance 1, and Δ is the Laplacian operator. The stationary solution Z is a Gaussian field with a Matérn covariance function with scaling parameter κ (approximately inversly proportional to the range) and smoothness parameter ν. The variance is a function of κ, ν, and d.

Lindgren, Rue, and Lindström investigate the limit as $\nu \to 0$ for $d = 2$, finding that the solution is a GMRF on a unit lattice [4]. They then construct approximations for positive integer values of ν by ν-fold convolution of Z with itself. Finally, they use a finite element method to genrealize the approximation to arbitrary triangulations of the support. This approximation has considerable computational benefits because the GMRF has a sparse covariance structure; the only nodes with nonzero covariances are those directly connected by edges in the triangulation.

In practice, the SPDE approach goes as follows. Choose nodes $\mathbf{s}_i$ at which to model $Z(\mathbf{s}_i)$, then build a triangular mesh using these nodes. Typically the nodes will include locations where data or covariates are available, then the rest will be filled in with a Delaunay triangulation under some edge length constraints. The $Z(\mathbf{s}_i)$ are be modeled as a GMRF where the distribution of each $Z(\mathbf{s}_i)$ depends only on the $Z(\mathbf{s}_j)$ where $\mathbf{s}_i$ and $\mathbf{s}_j$ are connected by an edge. The GMRF representation is assumed to be a piecwise linear approximation of the continuous Gaussian field; values of $Z(\mathbf{s})$ for $\mathbf{s}$ not in the set of nodes are predicted using linear interpolation using the barycentric coordinates of $\mathbf{s}$.

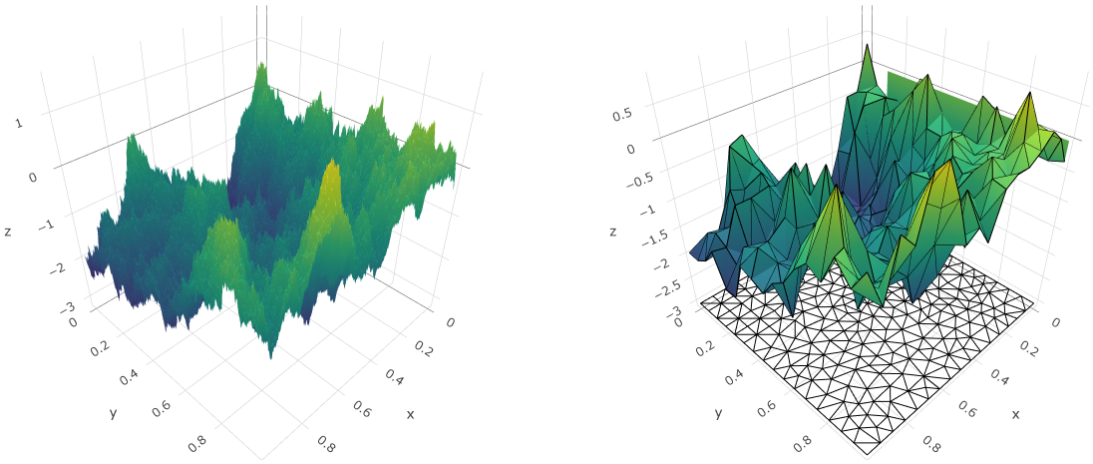

Figure 1. A realization of a spatial Gaussian process (left) and an approximation of that realization over a triangular mesh (right).

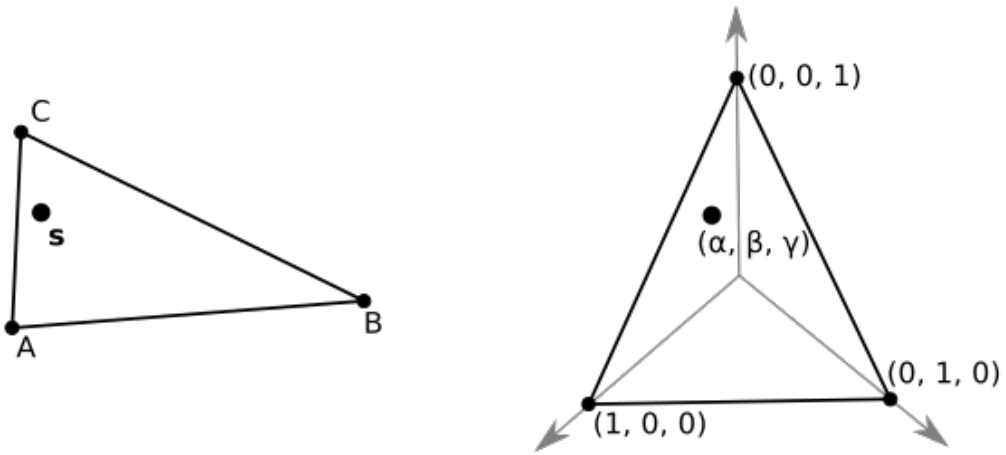

Figure 2. An illustration of the transformation from a mesh triangle to the simplex. (α, β, γ) are the barycentric coordinates of $\mathbf{s}$

The SPDE approach is implemented in the `R-INLA` package along with tools for constructing triangulation meshes.

2.2. Going Off the Grid

Log-likelihood:

$$\ell(Z) = C - \int \lambda(\mathbf{s}) d\mathbf{s} + \sum \log\left(\lambda(\mathbf{s_i})\right)$$

[8]:

$$\ell(Z) \approx C - \sum_i \tilde{\alpha}_i \exp\left[\sum_j z_j \phi_j(\tilde{\mathbf{s}}_i)\right] + \sum_i \sum_j z_j \phi_j(\mathbf{s_i})$$

(Poisson distribution)

2.3. Variable Sampling Effort

Observed a thinned process

Thinning process can be known or unknown

Idea has been around a while but most applications have used gridding [2].

Scale SPDE node integration weights by thinning probabilities when known

Incorporate log-linear model for thinning probability when unknown [9]

3. Applications

3.1. Simulation Study

3.2. Data Application

Examples with data, maybe `bei` dataset or Victorville

4. Conclusion and Discussion

References

[1] M. Blangiardo and M. Cameletti, *Spatial and Spatio-temporal Bayesian Models with R-INLA*, Wiley, 2015.

[2] A. Chakraborty, A.E. Gelfand, A.M. Wilson, A.M. Latimer, and J.A. Silander, *Point pattern modelling for degraded presence-only data over large regions*, Journal of the Royal Statistical Society: Series C (Applied Statistics) 60 (2011), pp. 757–776.

[3] F. Lindgren and H. Rue, *Bayesian spatial modelling with R-INLA*, Journal of Statistical Software 63 (2015), pp. 1–25.

[4] F. Lindgren, H. Rue, and J. Lindström, *An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73 (2011), pp. 423–498.

[5] J. Møller and R. Waagepetersen, *Modern spatial point process modelling and inference*, Scandinavian Journal of Statistics 34 (2007), pp. 643–711.
[6] J. Møller, A.R. Syversveen, and R.P. Waagepetersen, *Log gaussian cox processes*, Scandinavian journal of statistics 25 (1998), pp. 451–482.
[7] H. Rue, S. Martino, and N. Chopin, *Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations*, Journal of the royal statistical society: Series b (statistical methodology) 71 (2009), pp. 319–392.
[8] D. Simpson, J.B. Illian, F. Lindgren, S.H. Sørbye, and H. Rue, *Going off grid: Computationally efficient inference for log-gaussian cox processes*, Biometrika 103 (2016), pp. 49–70.
[9] Y. Yuan, F.E. Bachl, F. Lindgren, D.L. Borchers, J.B. Illian, S.T. Buckland, H. Rue, T. Gerrodette, *et al.*, *Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales*, The Annals of Applied Statistics 11 (2017), pp. 2270–2297.