Approximate inference for Bayesian smoothing problems on bounded domains
Stochastic partial differential equations and INLA

Daniel Simpson, Håvard Rue, Finn Lindgren
Department of Mathematical Sciences
14 July, 2010
Outline

Smoothing Problems

Using SPDEs to construct a computationally efficient prior

An example
Bayesian surface reconstruction

Suppose we have some observed data \( \{y_i\}_{i=1}^N \) taken at locations \( \{s_i\}_{i=1}^N \in \mathbb{R}^2 \) and

\[
y_i \sim \pi(y_i|x(\cdot), \theta),
\]

where \( x(\cdot) \) is a latent surface we are interested in and \( \theta \) is a vector of parameters.

By Bayes’ rule:

\[
\pi(x(s), \theta|y) \propto \pi(y|x(\cdot), \theta)\pi(x(\cdot)|\theta)\pi(\theta).
\]

How do we chose the prior on \( x(\cdot) \)? What if the region we are interested is bounded?
Deterministic smoothing splines on $\mathbb{R}^2$

Find the function $f(s)$ that:

- Interpolates the points $(s_i, y_i)$: that is $f(s_i) = y_i$
- Isn’t too rough: that is $f(s)$ minimises the bending energy

$$
\int_{\mathbb{R}^2} (\Delta f(x))^2 \, dx,
$$

where $\Delta = \frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}$. 
What does the solution look like?

The smoothing spline has the form

\[ f(x) = \sum_{i=1}^{N} a_i k(\|x - y_i\|), \]

where \( k(r) = r^2 \log(r) \) a the radial basis function.

What if we’re smoothing over something more interesting than \( \mathbb{R}^2 \)?
Smoothing over more interesting domains
Smoothing over more interesting domains

NOTHING CHANGES!
Smoothing over more interesting domains

NOTHING CHANGES!

Except the “radial basis function” $k(x, y)$ (which will no longer be isotropic).
$k(x, y)$—Deterministic interpretation

The kernel function $k(x, y)$ satisfies

$$\Delta^2 k(x, y) = \delta(x - y)$$

where $\Delta^2 = \frac{\partial^4}{\partial s_1^4} + 2\frac{\partial^4}{\partial s_1^2\partial s_2^2} + \frac{\partial^4}{\partial s_2^4}$ on the domain $D$ we want to smooth over.

For almost any domain, $k(x, y)$ cannot be expressed in terms of simple functions!

But they can be computed numerically $\Leftarrow$ Basis for Soap-film smoothing approach of Wood et al..
The stochastic process interpretation

\[ k(x, y) \] is the covariance function of the solution to the stochastic PDE

\[ \Delta u(s) = W(s), \]

where \( W(x) \) is Gaussian white noise on \( D \).

If we use this process as our prior, the MAP estimate will be the deterministic spline.

NB: This prior is completely independent of the location of the data.
Outline

Smoothing Problems

Using SPDEs to construct a computationally efficient prior

An example
Problems with Gaussian random fields

There are two primary problems with the GRF approach:

— It is *impossible* to calculate $k(x, y)$ analytically!

— The covariance function has global support which will mean the covariance matrix is *dense*!

• Solution: Use a computer!

• Better solution: Don’t use it explicitly! (This talk!)


Problems with Gaussian random fields

There are two primary problems with the GRF approach:

— It is *impossible* to calculate $k(x, y)$ analytically!
  
  • Solution: Use a computer!
    
    — *Soap Film Smoothing*, S.N. Wood, M.V. Bravington and S.L. Hedley, JRSSSB, 2008.)

— The covariance function has global support which will mean the covariance matrix is *dense*!
Problems with Gaussian random fields

There are two primary problems with the GRF approach:

— It is *impossible* to calculate $k(x, y)$ analytically!
  
  • Solution: Use a computer!
    
    — *Soap Film Smoothing*, S.N. Wood, M.V. Bravington and S.L. Hedley, JRSSSB, 2008.)
  
    • Better solution: Don’t use it explicitly! (This talk!)

— The covariance function has global support which will mean the covariance matrix is *dense*!
Problems with Gaussian random fields

There are two primary problems with the GRF approach:

— It is impossible to calculate $k(x, y)$ analytically!
  
  • Solution: Use a computer!
    
    — Soap Film Smoothing, S.N. Wood, M.V. Bravington and S.L. Hedley, JRSSSB, 2008.)
    
    • Better solution: Don’t use it explicitly! (This talk!)

— The covariance function has global support which will mean the covariance matrix is dense!
  
  • Solution: Approximate the GRF by a Gaussian Markov random field.
    
How to “work” with the SPDE?

In order to construct a practical GMRF approximation to the Gaussian field, we need to look for the simplest type of field we can—a finite combination of piecewise linear functions:

$$u(s) = \sum_{k=1}^{M} \psi_k(s) w_k$$

for linear basis-functions \(\{\psi_k\}\) and (Gaussian) weights \(\{w_k\}\).
What does the prior look like?

If we take $M$ points (not necessarily related to the data points) in the domain and triangulate it, the prior is an $M$–dimensional GMRF

$$w \sim \text{MVN}(0, \kappa^{-1} Q^{-1}),$$

where the precision matrix $Q$ is given by

$$Q = K^T C^{-1} K,$$

where $K$ is the “Finite Element Representation of the Laplacian” and $C$ is diagonal.

We have an R–package to compute this!
Kriging!

The posterior field will be for the form

\[ x(s), \theta | y \overset{D}{=} \sum_{i=1}^{M} \psi_i(s) w_i^{post}, \]

where

\[ \pi(w^{post}, \theta | y) \propto \pi(y | w, \theta) \pi(w | \theta) \pi(\theta) \]

is the posterior of \( w \).

**THE KRIGING ESTIMATES ARE FREE!**
Outline

Smoothing Problems

Using SPDEs to construct a computationally efficient prior

An example
An interesting region
The problem

Simulated data

Observed data:

\[ y_i | x_i \sim \text{Bin}(n, \logit^{-1}(x_i)), \]

where the spatial effects are modelled by

\[ x | \kappa \sim \text{MVN}(0, \kappa^{-1} Q^{-1}) \]

and the precision prior is \( \kappa \sim \text{IG}(1, 0.001) \).

Take \( n = 100 \) and simulate 300 data points.
The RMS Error in the reconstruction is 0.073.
So where do we go from here?

— This serves a ‘proof of concept’ for fast Bayesian smoothing and Kriging on complex regions.

— It also works on manifolds!
  - The sphere! (David Bolin, Session I9 on Wednesday)
  - We can use this to explicitly incorporate the topography of a region into our models. (Ottavi and Rue, In prep.)

— It is trivial to generate non-stationary priors. But how do we model them?

— It is also trivial to generate non-separable spatiotemporal random fields.