Think Continuous: Markovian Gaussian Models in Spatial Statistics

Simpson, Lindgren, and Rue
2012

Serge Aleshin-Guendel

April 11, 2018

Spatial Statistics

- Suppose we have some data that's spatially indexed

$$\{Y_i(s_i), \{X_{ij}\} | s_i \in D_s, i = 1, \cdots, N, j = 1, \cdots, n_i\}$$

where $Y(s)$ are our observed data, $\{X_{ij}\}$ are covariates, and D_s is the region in space we observe data. We'd like to model the observed data while taking into account the spatial nature of the data.

- If we're working in a GLM framework, we could write the desired model as

$$Y_i(s_i)|\eta_i \sim \text{Exponential Family}(\eta_i),$$

where for an appropriate link function g,

$$g(\eta_i) = \sum_{j=1}^{n_i} \beta_j X_{ij} + f(s_i),$$

where f is some spatial random effect.

- How do we specify f?

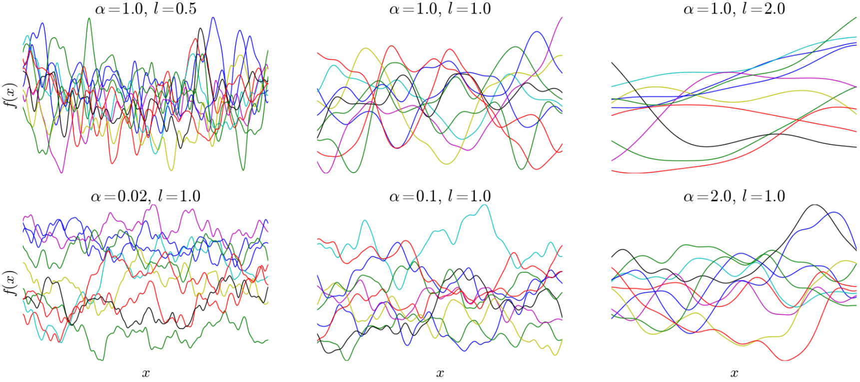
Gaussian Processes

- The canonical way to specify the spatial random effect is through a Gaussian Process (GP), or more generally a Gaussian Field.
- Can think of a GP as a distribution over functions which are indexed by some underlying set, like space (what we care about) or time.
- GPs are completely specified through a mean function μ and a positive semi-definite covariance function (kernel) c. The mean function is usually taken to be known and equal to 0.

Gaussian Processes

Example from Keyon Vafa. Here are some samples from a GP indexed by $\mathbb{R}$, with a rational quadratic covariance function, which is defined as:

$$c(x, x') = \left(1 + \frac{(x - x')^2}{2\alpha l^2}\right)^{-\alpha}.$$



Gaussian Processes

- Key Fact: If $f \sim \mathrm{GP}(0, c)$, then for any finite collection of points in the index set $\{s_1, \cdots, s_n\}$, the process evaluated at these points is jointly Gaussian

  $$\{f(s_1), \cdots, f(s_n)\} \sim \mathrm{MVN}(0, \Sigma),$$

  where $\Sigma_{ij} = c(s_i, s_j)$.
- The covariance matrix Σ is sometimes called the Gram matrix.

Gaussian Processes: Main Problem

- In order to use a GP in practice, you're going to have to either calculate the determinant or find the inverse of the covariance matrix.
- For example you may want to evaluate the density:

$$\pi(f(s)) = \frac{1}{\sqrt{(2\pi)\det(\Sigma)}} \exp\left\{\frac{1}{2} f(s)^T \Sigma^{-1} f(s)\right\}.$$

- Calculating the determinant or finding the inverse of a generic covariance matrix is $\mathcal{O}(n^3)$, which for more than a couple thousand points is intractable.

Gaussian Markov Random Fields

- There are a certain class of multivariate Gaussians that are much more computationally tractable than GPs, called Gaussian Markov Random Fields (GMRFs).
- If $x \sim \text{MVN}(\mu, \Sigma)$ is a GMRF, then for some components x_i and x_j, $x_i \perp\!\!\!\perp x_j | x_{-ij}$.
- The important part is that this leads to a sparse precision matrix, which is $Q = \Sigma^{-1}$.
- In particular, if $x_i \perp\!\!\!\perp x_j | x_{-ij}$, then $Q_{ij} = Q_{ji} = 0$.

Gaussian Markov Random Fields

- When you specify a Gaussian in terms of a sparse precision matrix, the cost to compute determinants and find the covariance matrix goes down to $\mathcal{O}(n^{3/2})$, which is much more tractable than the generic case.
- However, nearly all GMRFs in the spatial literature are specified discretely, rather than continuously, through graphs or lattices.

What We're Working Towards

- We'd like to be able to specify, and actually use, a continuous spatial model that's computationally tractable and accurate.
- We know how to specify continuous spatial models using Gaussian Processes, and we know that GMRFs are computationally tractable.
- So how do we represent a GP as a GMRF?

Markovian Gaussian Processes

- When the index space of a GP is one dimensional, like with time, Markov properties are relatively straightforward to think about/ encode in a covariance function.
- When the index of a GP is space, like we want, it's harder to nail down what being Markov means.

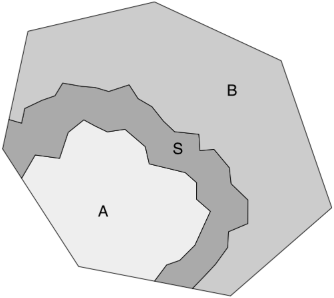


Fig. 1. An illustration of the spatial Markov property. If, for any appropriate set S, $\{x(s) : s \in A\}$ is independent of $\{x(s) : s \in B\}$ given $\{x(s) : s \in S\}$, then the field $x(s)$ has the spatial Markov property.

Some Theory

- The Fourier transform of the covariance function of a stationary GP on $\mathbb{R}^2$, called the power spectrum, is given by

$$R(k) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \exp(-ik^T h) c(h) dh.$$

- Note that when we evaluate the kernel at a point, $c(h)$, we're using the stationarity of the kernel (really evaluating it at two points that differ by h).
- Apparently some Russian in the 70s showed that a stationary GP is Markov iff $R(k) = \frac{1}{p(k)}$ for some positive symmetric polynomial p.
- This doesn't seem that helpful!

Some Theory

- Can define the covariance operator of a Markovian GP as

$$C[f](h) = \int_{R^2} c(h' - h) f(h') dh' = \int_{R^2} \exp(-ik^T h) \frac{\hat{g}(k)}{p(k)} dk,$$

  where g is a smooth function that goes to zero rapidly at infinity, $\hat{g}$ is it's Fourier transform, and p the inverse of the power spectrum.
- Can think of this as the functional analog of the covariance matrix (can talk about eigenvalues and eigenfunctions of the operator).

Some Theory

- The covariance operator has inverse, which is the (surprise) precision operator, defined for a Markovian GP as

$$Q[f](h) = \int_{R^2} \exp(-ik^T h)\hat{g}(k)p(k)dk = \sum_{|i|\leq \ell} a_i D^i f(h),$$

  where a_i are the coefficients of p, ℓ is the degree of p, and D^i are appropriate differential operators.
- Ok so what's the point?
- It can be shown that a GP is Markovian iff it's precision operator is of this differential form.
- Note that the differential operator is **local**, whereas the integral form of the covariance is **global**!
- So we have some guidance on how to find a Markov GP!

The Matern Kernel

- One of the most common covariance functions used in practice is the Matern covariance, given by:

$$c(s, s') = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa||s - s'||)^{\nu} K_{\nu}(\kappa||s - s'||),$$

  where σ is a variance parameter, ν is a smoothness parameter, κ is a range parameter, and K_ν is the modified Bessel function of the second kind with parameter ν.
- Consider in $\mathbb{R}^2$, the following stochastic partial differential equation (SPDE)

$$(\kappa - \Delta)^{\alpha/2} f(s) = W(s),$$

  where $\alpha = \nu + 1 > 0$, $\Delta = \frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}$ is the 2 dimensional Laplacian (the trace of the Hessian), and $W(s)$ is Gaussian white noise
- It turns out that the solution $f(s)$ to this SPDE is a GP with a Matern covariance with the respective parameters!

The Markov Connection

- It can be shown that when α is an integer, the precision operator of $f(s)$ as defined by the SPDE is

$$Q = (\kappa - \Delta)^{\alpha}.$$

- From our bit of theory, we see that GPs with Matern kernels and integer ν are actually Markov!
- Now how do we take advantage of this fact in order to find a computationally tractable representation of GPs with Matern kernels?

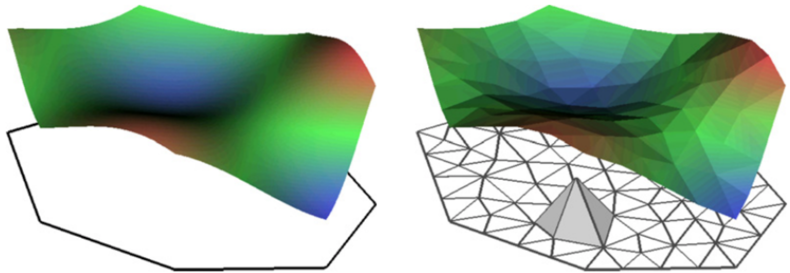
Approximating an SPDE: Finite Element Method

- We'll consider approximations to $f(s)$ of the form

$$f(s) \approx f_a(s) = \sum_{i=1}^{n} w_i \phi_i(s),$$

  where the weights w_i are jointly Gaussian and ϕ_i are appropriately chosen basis functions.
- So how do we choose the basis functions so that the approximation has a Markov form?
- The first step is to triangulate our spatial region of interest.
- Basis functions then correspond to piecewise linear functions defined on the verticies of the triangulation.

Triangulation



(a) A continuous function.

(b) A piecewise linear approximation.

Fig. 2. Piecewise linear approximation of a function over a triangulated mesh.

Weak Solutions and Weighted Basis Functions

- Going to just consider $\alpha = 2$, or equivalently $\nu = 1$.
- Letting $D_s \in \mathbb{R}^2$, we have that any solution to the SPDE defined before also satisfies, for any suitable $\psi(s)$,

$$\int_{D_s} \psi(s)(\kappa - \Delta) f(s) = \int_{D_s} \psi(s) W(ds).$$

- So how do we choose ψ? Can't test all possible functions, so how about just $\{\phi_j\}_{j=1}^{n}$?

Weak Solutions and Weighted Basis Functions

- Using the basis functions as our test functions, we arrive then at the following system of linear equations after using Green's formula (something you might have seen in multivariable calculus):

$$\sum_{i=1}^{n}\left(\kappa^2\int_{D_s}\phi_i(s)\phi_j(s)ds+\int_{D_s}\nabla\phi_i(s)\nabla\phi_j(s)ds\right)w_i=$$

$$\int_{D_s}\phi_j(s)W(ds),$$

  for $j=1,\cdots,n$.
- Due to the piecewise linear nature of the basis functions, all of the integrals can be easily computed! In particular, the white noise integral is Gaussian with zero mean and covariance given by $\tilde{C}_{ij}=\int_{D_s}\phi_i(s)\phi_j(s)ds$.

The Final Approximation

- Letting

$$G_{ij} = \int_{D_s} \nabla\phi_i(s)\nabla\phi_j(s)ds,$$

  we arrive at

$$(\kappa\tilde{C} + G)w \sim \mathrm{MVN}(0, \tilde{C}).$$

- However, $\tilde{C}^{-1}$ is dense, so we approximate it with

$$C = \mathrm{diag}\left(\int_{D_s} \phi_i(s)ds, i = 1, \cdots, n\right).$$

- The solution w is then a GMRF with sparse precision matrix

$$Q = (\kappa\tilde{C} + G)^T C(\kappa\tilde{C} + G).$$

Generalizations

- Can generalize this approach to non-stationary, anisotropic, GPs on general manifolds!
- All you have to do is appropriately alter the SPDE.

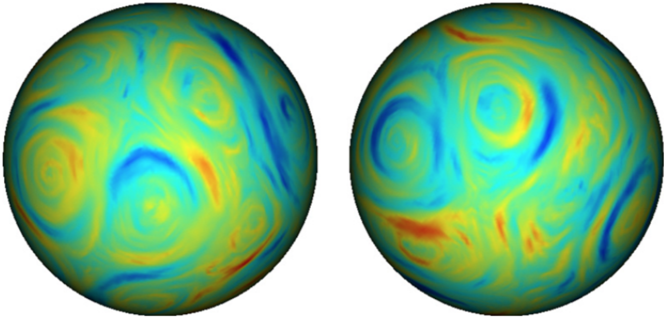


Fig. 5. A sample from a non-stationary anisotropic random field on the sphere.