1st, choose a covariance model;
2nd, aprroximate the precision matrix $\boldsymbol{Q}$;
3rd, draw approximate inference.

Henrique Laureano
http://leg.ufpr.br/~henrique

December 16, 2019

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

spde2smoothing

Understanding the Stochastic Partial Differential Equation Approach to Smoothing

David L. MILLER, Richard GLENNIE, and Andrew E. SEATON

Correlation and smoothness are terms used to describe a wide variety of random quantities. In time, space, and many other domains, they both imply the same idea: quantities that occur closer together are more similar than those further apart. Two popular statistical models that represent this idea are basis-penalty smoothers (Wood in Texts in statistical science, CRC Press, Boca Raton, 2017) and stochastic partial differential equations (SPDEs) (Lindgren et al. in J R Stat Soc Series B (Stat Methodol) 73(4):423–498, 2011). In this paper, we discuss how the SPDE can be interpreted as a smoothing penalty and can be fitted using the `R` package `mgcv`, allowing practitioners with existing knowledge of smoothing penalties to better understand the implementation and theory behind the SPDE approach.

Supplementary materials accompanying this paper appear online.

Key Words: Smoothing; Stochastic partial differential equations; Generalized additive model; Spatial modelling; Basis-penalty smoothing.

Where? *Journal of Agricultural, Biological, and Environmental Statistics*,

Published online: 19 September 2019

SPDE? An equation to be solved.

$$Df = \epsilon/\tau$$

- f, a stochastic process, called a solution to the SPDE;
- Df is a linear combination of derivatives of f, of different orders;
- ϵ, commonly a white noise process;
- τ, a parameter that controls the variance in the white noise process.
  - changes in f are more variable when τ is reduced and less variable for higher τ

---

f has a covariance structure that is induced by the choice of D.

i.e.,

Find a D that induces the covariance function that you want.

Going a little deeper

$Df = \epsilon$ is a convenient shorthand way to think about the SPDE, but technically, the SPDE only has meaning when stated in an integral form.

$$Df = \epsilon \text{ means that we require } \int Df(x)\phi(x)\ \mathrm{d}x = \int \epsilon(x)\phi(x)\ \mathrm{d}x$$

for every function ϕ with compact support.

The function ϕ is often called the test function.

Integral form makes sense because any stochastic process can be integrated, but not every one can be differentiated.

Ok, but how we solve the SPDE? Finite Element Method (FEM).

$$\text{SPDE solution :} \quad \text{weighted sum, } f(x) = \sum_{j=1}^{M} \beta_j \psi_j(x).$$

Real life $\equiv$ Linear Algebra

The integral form can be written as a matrix equation: $\boldsymbol{P}\beta = \epsilon$ where

- $\boldsymbol{P}$ has $(i,j)^{\text{th}}$ entry $\langle D\psi_i, \psi_j \rangle$;
- ϵ has j^{th} entry $\langle \epsilon, \psi_j \rangle$
  - $\epsilon \sim \text{MVN}(0, \boldsymbol{Q}_e^{-1})$, where $\boldsymbol{Q}_e^{-1}$ has $(i,j)^{\text{th}}$ entry $\langle \psi_i, \psi_j \rangle$
- $\beta \sim \text{MVN}(0, \boldsymbol{Q}^{-1})$, where $\boldsymbol{Q} = \boldsymbol{P}^\top \boldsymbol{Q}_e \boldsymbol{P}$
  - i.e., the SPDE is therefore a way to specify a prior for β.

Summary

Given an SPDE, one can use the FEM to compute $\boldsymbol{Q}$ and therefore simulate $\tilde{\beta}$ from a MVN with precision $\boldsymbol{Q}$. The function $f = \sum_{j=1}^{M} \tilde{\beta}_j \psi_j$ would then be a realization from a stochastic process which is a solution to the SPDE, a stochastic process with the covariance structure implied by D.

Matérn SPDE

$$\kappa^2 f - \Delta f = \epsilon/\tau,$$

i.e. $Df = \epsilon$ with $D = (\kappa^2 - \Delta)^{\alpha/2}\tau$.
D is a linear differential operator only when $\alpha = \nu - d/2 = 2$.

Whittle, P. (1954)[1] shows that the solution of this SPDE has Matérn covariance.

In other words, the $\boldsymbol{Q}$ computed from the FEM is an approx. to the $\boldsymbol{Q}$ one would obtain if you computed $\boldsymbol{\Sigma}$ with the Matérn covariance function and then, at great computational cost, inverted it.

[1] On stationary processes in the plane. *Biometrika 41*(3-4), 434-449.

Basis-penalty smoothing approach

$$\text{penalized likelihood}: \quad l_p(\beta, \lambda) = l(\beta) - J(\beta, \lambda),$$

- For the observations given the form of f, log-likelihood $l(\beta)$;
- To penalize functions that are too wiggly, smoothing penalty $J(\beta, \lambda)$.

To estimate the optimal smoothing parameter λ and the coefficients β: REstricted Maximum Likelihood (REML).

Similar to the SPDE approach:

- The function f is a sum of basis functions multiplied by coefficients.

Difference:

- Rather than specify an SPDE and deduce a covariance structure, a smoothing penalty is used to induce correlation.

Going a little deeper in the smoothing penalty

Smoothing penalty leads to an optimal curve, the smoothing spline[2]. The penalty for smoothing splines takes the form
$J(\beta, \lambda) = \lambda \int (Df)^2 = \lambda \langle Df, Df \rangle$.

$$\text{When } f(x) = \sum_{j=1}^{M} \beta_j \psi_j(x), \text{ we have } J(\beta, \lambda) = \lambda \beta^\top \boldsymbol{S} \beta$$

where $\boldsymbol{S}$ is a $M \times M$ matrix with $(i, j)^{\text{th}}$ entry $\langle D\psi_i, D\psi_j \rangle$.

Rewriting the penalized log-likelihood as a likelihood,

$$\exp\{l_p(\beta, \lambda)\} = \exp\{l(\beta)\} \times \exp(-\lambda \beta^\top \boldsymbol{S} \beta),$$

$\exp(-\lambda \beta^\top \boldsymbol{S} \beta)$ is $\propto$ to a $\text{MVN}(0, \boldsymbol{S}_\lambda^{-1} = (\lambda \boldsymbol{S})^{-1})$.
The penalized likelihood is equivalent to assigning the prior $\beta \sim \text{MVN}(0, \boldsymbol{S}_\lambda^{-1})$.

[2] Wahba, G. (1990). *Spline methods for observational data*. SIAM, USA.

Connection: SPDE model as a basis-penalty smoother

- For a given differential operator D, the approx. $\boldsymbol{Q}$ for the SPDE is the same as the precision matrix $\boldsymbol{S}_\lambda$ computed using the smoothing penalty $\langle Df, Df \rangle$;
- Differences between the basis-penalty approach and the SPDE finite element approx., when using the same basis and differential operator, are differences in implementation only.

Lindgren, F., Rue, H. and Lindström, J. (2011)[a]

[a] An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach (with discussion). *Journal of the Royal Statistical Society: Series B 73*(4), 423-498

An approx. solution to the SPDE is given by representing f as a sum of linear (specifically, B-spline) basis functions multiplied by coefficients; the coefs of these basis form a GMRF.

Matérn penalty

$$D = \tau(\kappa^2 - \Delta) \quad \Rightarrow \quad \text{smoothing penalty} : \tau \int (\kappa^2 f - \Delta f)^2 \, \mathrm{d}x.$$

- inverse correlation range κ: higher values lead to less smooth functions;
- smoothing parameter τ controls the overall smoothness of f.

In matrix form, this leads to the smoothing matrix

$$\boldsymbol{S} = \tau(\kappa^4 \boldsymbol{C} + 2\kappa^2 \boldsymbol{G}_1 + \boldsymbol{G}_2) \quad \text{where}$$

$\boldsymbol{C}, \boldsymbol{G}_1, \boldsymbol{G}_2$ are all $M \times M$ sparse matrices with $(i,j)^{\text{th}}$ entries $\langle \psi_i, \psi_j \rangle$, $\langle \psi_i, \nabla \psi_j \rangle$, and $\langle \nabla \psi_i, \nabla \psi_j \rangle$.

The matrix $\boldsymbol{S}$ is equal to the matrix $\boldsymbol{Q} = \boldsymbol{P}^\top \boldsymbol{Q}_e \boldsymbol{P}$ computed using the FEM.

Fitting the Matérn SPDE in `mgcv`

`mgcv` allows the specification of new basis-penalty smoothers.

step-by-step

- `INLA::inla.mesh.(1d or 2d)` to create a mesh;
- `INLA::inla.mesh.fem` to calculate $\boldsymbol{C}$, $\boldsymbol{G}_1$, and $\boldsymbol{G}_2$;
- Connect the basis representation of f to the observation locations,
  - The full design matrix is given by combining the fixed effects design matrix $\boldsymbol{X}_c$ and the contribution for f, $\boldsymbol{A}$ - the projection matrix found using `INLA::inla.spde.mesh.A`;
- Use REML to findo optimal κ, τ and β.

Some final remarks,

- As REML is an empirical Bayes procedure, we expect point estimates for $\hat{\beta}$ to coincide with `R-INLA`;
- A uniform prior is implied for the smoothing parameters τ and κ;
- `R-INLA` allows for similar estimation by just using the modes of the hyperparameters κ and τ (`int.strategy="eb"`).

To finish, let's check some [code].