LECTURE 13: OTHER COVARIANCE FUNCTIONS

CLASS INTRO

INTRO QUESTIONS

- Compare the strengths and weaknesses of the model fitting tools we saw last class.
- For Today:
  - Building Covariance Functions

GEOSTATISTICAL THEORY

STOCHASTIC PROCESS SPECIFICATION

- Given the assumption that a Gaussian process is reasonable for the spatial process, a valid covariance function needs to be specified.
- Up to this point, we have largely worked with isotropic covariance functions. In particular, the spherical and exponential covariance functions have primarily been used.
- However, a Gaussian process is flexible and can use any valid covariance function.

COVARIANCE FUNCTIONS

- A valid covariance function $C(\boldsymbol{h})$, defined as $Cov(Y(\boldsymbol{s}), Y(\boldsymbol{s}+\boldsymbol{h}))$, for any finite set of sites $\boldsymbol{s_1}, \ldots, \boldsymbol{s_n}$ and $a_1, \ldots, a_n$ should satisfy
$$Var\left[\sum_i a_i Y(\boldsymbol{s_i})\right] = \sum_{i,j} a_i a_j Cov(Y(\boldsymbol{s_i}), Y(\boldsymbol{s_j})) = \sum_{i,j} a_i a_j C(\boldsymbol{s_i} - \boldsymbol{s_j}) \geq 0$$
with strict inequality if all the a_i are not zero.
- In other words, $C(\boldsymbol{h})$ needs to be a positive definite function, which includes the following properties

1. $C(\mathbf{0}) \geq 0$
2. $|C(\boldsymbol{h})| \leq C(\mathbf{0})$

BOCHNER'S THEOREM

- Bochner's Theorem provides necessary and sufficient conditions for positive definiteness, stating $C(\boldsymbol{h})$ is positive definite, if and only if,

  $$C(\boldsymbol{h}) = \int \cos(\boldsymbol{w}^T \boldsymbol{h}) G(d\boldsymbol{w})$$

  where $G()$ is bounded, positive, and symmetric around 0 in $\mathcal{R}^r$.
- This equation is also related to the Fourier transformation, of $C(\boldsymbol{h})$, and the associated spectral distribution.

CONSTRUCTING COVARIANCE FUNCTIONS

- There are three approaches for building correlation functions. For all cases let $C_1, \dots, C_m$ be valid correlation functions

1. *Mixing:* $C(\boldsymbol{h}) = \sum_i p_i C_i$ is also valid if $\sum_i p_i = 1$.
2. *Products:* $C(\boldsymbol{h}) = \prod_i C_i$
3. *Convolution:* $C_{12}(\boldsymbol{h}) = \int C_1(\boldsymbol{h} - \boldsymbol{t}) C_2(\boldsymbol{t}) d\boldsymbol{t}$ again this is based on a Fourier transform.

CHOOSING COVARIANCE STRUCTURE

Suppose a collaborator suggests two possible valid covariance functions and asks you which one is better. The collaborator also wants to select important covariates related to their study. How would you proceed.

1. Pick one or two questions to ask your collaborator.
2. Suggest two strategies for selecting the best model.

SMOOTHNESS

- Many one-parameter isotropic covariance functions will be quite similar.
- Another consideration for choosing the correlation function is the theoretical smoothness property of the correlation.
- Again the spatial process is a continuous surface that is observed at a finite number of locations, or when latent random effects are implied at observed locations.
- The Matern class of covariance functions contains a parameter, ν, to control smoothness. With $\nu = \infty$ this is a Gaussian correlation function and with $\nu = 1/2$ this results in an exponential correlation function.

MATERN COVARIANCE MOTIVATION

- “Expressed in a different way, use of the Matern covariance function as a model enables the data to inform about ν; we can learn about process smoothness despite observing the process at only a finite number of locations.”
- The Matern covariance function was originally derived as the characteristic function from a Cauchy spectral density.

MATERN COVARIANCE SPECIFICATION

- The Matern covariance function is written as
  $$\frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\phi d)^{\nu} K_{\nu}(\phi d),$$
  where $\Gamma()$ is a gamma function and K_{ν} is the modified Bessel function of order ν.

NONSTATIONARY MODEL OVERVIEW

- Let $w(\mathbf{s})$ be a mean 0 and variance 1 stationary spatial process with correlation function ρ. Then consider $v(\mathbf{s}) = \sigma(\mathbf{s})w(\mathbf{s})$.
- What is the variance of $v(\mathbf{s})$? What about $cov(v(\mathbf{s}), v(\mathbf{s}'))$?
- The variance is $\sigma^2(\mathbf{s})$ and the covariance is $\sigma(\mathbf{s})\sigma(\mathbf{s}')\rho(\mathbf{s} - \mathbf{s}')$
- This formulation results in nonhomogenous variance; spatial correlations are still stationary.
- $\sigma(\mathbf{s})$ could be modeled as a function of covariates.

MORE NONSTATIONARY IDEAS

- Nonstationarity can also me included through deformation and kernel mixing.
- Deformation transforms a geographic area into a new transformed area that is stationary and isotropic.
- Kernel mixing is another approach that builds a non-stationary function that is similar to a set of basis functions.

ANISOTROPY

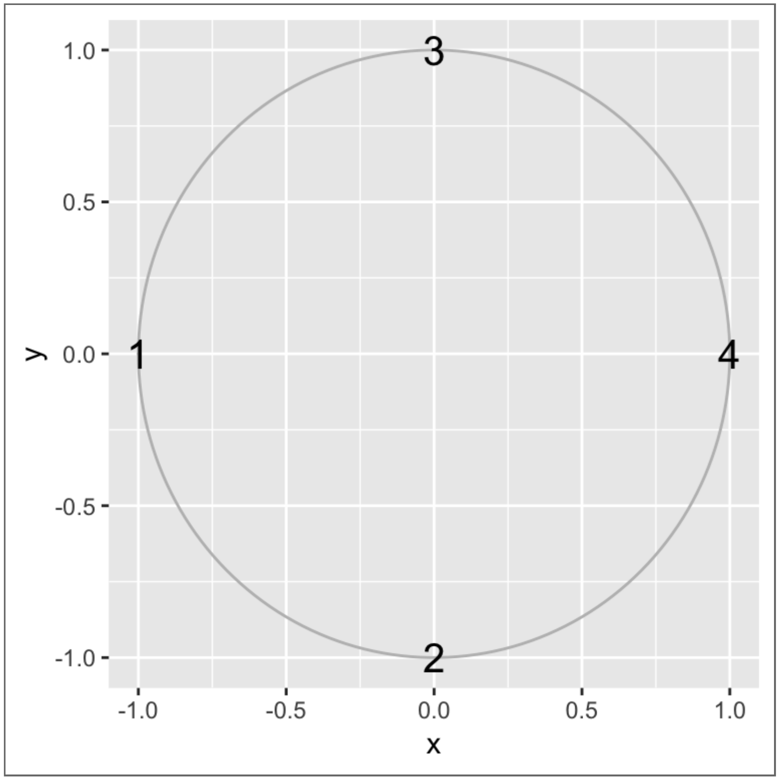
- Geometric anisotropy refers to the case where the coordinate space is anisotropic, but can be transformed to an isotropic space.
- The transformation can result in a rotation or stretching of coordinate axes.
- In general consider the correlation function,
  $\rho(\boldsymbol{h};\phi) = \phi_0(||L\boldsymbol{h}||;\phi)$
  where L is a $d \times d$ matrix that controls the transformation.

GEOMETRIC ANISOTROPY MODEL

- Let $\boldsymbol{Y}(\boldsymbol{s}) = \mu(\boldsymbol{s}) + w(\boldsymbol{s}) + \epsilon(\boldsymbol{s})$, thus $\boldsymbol{Y}(\boldsymbol{s}) \sim N(\mu(\boldsymbol{s}), \Sigma(\tau^2, \sigma^2, \phi, B))$, where $B = L^T L$.
- The covariance matrix is defined as $\Sigma(\tau^2, \sigma^2, \phi, B)) = \tau^2 I + \sigma^2 H((\boldsymbol{h}^T B \boldsymbol{h}^T)^{\frac{1}{2}})$, where $H((\boldsymbol{h}^T B \boldsymbol{h}^T)^{\frac{1}{2}})$ has entries of $\rho((\boldsymbol{h}_{\boldsymbol{ij}}^T B \boldsymbol{h}_{\boldsymbol{ij}}^T)^{\frac{1}{2}}))$ with $\rho()$ being a valid covariance function, typically including ϕ and $\boldsymbol{h}_{\boldsymbol{ij}} = \boldsymbol{s}_i - \boldsymbol{s}_j$.

GEOMETRIC ANISOTROPY VISUAL

- Consider four points positioned on a unit circle.



- How far apart are each set of points?

GEOMETRIC ANISOTROPY EXERCISE 1

Now consider a set of correlation functions. For each, calculate the correlation matrix and discuss the impact of B on the correlation. Furthermore, how does B change the geometry of the correlation?

1. $\rho() = \exp(-\boldsymbol{h}_{ij}^T B \boldsymbol{h}_{ij}^T)^{\frac{1}{2}})$, where $B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
2. $\rho() = \exp(-\boldsymbol{h}_{ij}^T B \boldsymbol{h}_{ij}^T)^{\frac{1}{2}})$, where $B = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$
3. $\rho() = \exp(-\boldsymbol{h}_{ij}^T B \boldsymbol{h}_{ij}^T)^{\frac{1}{2}})$, where $B = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}$

GEOMETRIC ANISOTROPY: SOLUTION 1

1. $\rho() = \exp(-\boldsymbol{h}_{\boldsymbol{ij}}^T I \boldsymbol{h}_{\boldsymbol{ij}}^T)^{\frac{1}{2}}))$

1.000	0.243	0.243	0.135
0.243	1.000	0.135	0.243
0.243	0.135	1.000	0.243
0.135	0.243	0.243	1.000

GEOMETRIC ANISOTROPY: SOLUTION 2

2. $\rho() = \exp(-\boldsymbol{h}_{\boldsymbol{ij}}^T B \boldsymbol{h}_{\boldsymbol{ij}}^T)^{\frac{1}{2}}))$, where $B = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$

1.000	0.177	0.177	0.059
0.177	1.000	0.135	0.177
0.177	0.135	1.000	0.177
0.059	0.177	0.177	1.000

GEOMETRIC ANISOTROP: SOLUTION 3

3. $\rho() = \exp(-\boldsymbol{h}_{\boldsymbol{ij}}^T B \boldsymbol{h}_{\boldsymbol{ij}}^T)^{\frac{1}{2}}))$, where $B = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}$

1.000	0.243	0.086	0.031
0.243	1.000	0.135	0.086
0.086	0.135	1.000	0.243
0.031	0.086	0.243	1.000

MORE GEOMETRIC ANISOTROPY

- The matrix B relates to the orientation of a transformed ellipse.
- The (effective) range for any angle η is determined by the equation
  $\rho(r_\eta(\tilde{\boldsymbol{h}}_\eta^T B \tilde{\boldsymbol{h}}_\eta^T)^{\frac{1}{2}}) = .05,$
  where $\tilde{\boldsymbol{h}}_\eta$ is a unit vector in the direction η.

FITTING GEOMETRIC ANISOTROPY MODELS

- Okay, so if we suspect that geometric anisotrophy is present, how do we fit the model? That is, what is necessary in estimating this model?
- In addition to σ^2 and τ^2 we need to fit B.
- What about ϕ? What is ϕ when $B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$?

PRIORS FOR B

- While B is a matrix, it is just another unknown parameter.
- Hence, to fit a Bayesian model we need a prior distribution for B.
- One option for the positive definite matrix is the Wishart distribution, which is a bit like a matrix-variate gamma distribution.

UP NEXT

- Spatial GLMs
- Model Selection
- Areal Data