STAT 8025

Lecture 3: Geostatistical Models

Dr. Emily Lei Kang

Division of Statistics & Data Science
Department of Mathematical Sciences
University of Cincinnati

- Gaussian Process
  - Multivariate normal distribution
- Specifying the mean function
- Covariance function
  - Assumptions
  - Common covariance functions

Geostatistical/Point-Referenced Data

- Let $\{Y(s) : s \in \mathcal{D}\}$ denote the spatial process in a spatial domain $\mathcal{D}$
  - With $\mathcal{D} \subset \mathcal{R}$, $Y(s)$ can be visualized with a curve
  - With $\mathcal{D} \subset \mathcal{R}^2$, $Y(s)$ can be visualized with a surface
- We observe this process $\{Y(\cdot)\}$ at a set of finite locations, say, $s_1, \ldots, s_n$
- Goals when analyzing point-referenced data:
  - Mapping: smoothing and making prediction at unobserved locations
    - Generating $\hat{Y}(s_0)$, at any prediction location (with associated uncertainty e.g., confidence interval, standard error)
  - Accounting for spatial dependence in regression models testing null hypotheses of no spatial structure
    - Better investigate β, e.g., hypothesis testing, confidence interval

Gaussian Process

- Gaussian process (GP) is popular in spatial statistics, computer experiments, and machine learning.
- It is a model for a random process at all locations in $\mathcal{D}$
- It is related to univariate and multivariate normal distributions and possess nice properties related to predictions.
- For any n and any $\mathsf{s}_1, \ldots, \mathsf{s}_n \in \mathcal{D}$, the joint distribution of $(Y(\mathsf{s}_1), \ldots, Y(\mathsf{s}_n))$ is multivariate normal
- Two components to specify this process:
  - Mean function: $\mu(\cdot)$
  - Covariance function: $C(\cdot, \cdot)$

Gaussian Process

For any n and any $s_1, \ldots, s_n \in \mathcal{D}$

$$\begin{pmatrix} Y(s_1) \\ Y(s_2) \\ \vdots \\ Y(s_n) \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

where
$\boldsymbol{\mu} = (\mu(s_1), \ldots, \mu(s_n))'$ and

$$\Sigma = \begin{pmatrix} C(s_1, s_1) & C(s_1, s_2) & \cdots & C(s_1, s_n) \\ C(s_2, s_1) & C(s_2, s_2) & \cdots & C(s_2, s_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(s_n, s_1) & C(s_n, s_2) & \cdots & C(s_n, s_n) \end{pmatrix}$$

- Approapriate modeling the functions $\mu(\cdot)$ and $C(\cdot, \cdot)$ is key to estimation, infernce and prediction.

The Multivariate Normal Distribution

- Univariate normal: $Y \sim \mathcal{N}(\mu, \sigma^2)$, fully specified by its mean and variance
- Multivariate normal: p-dimensional random vector $\mathrm{Y} = (Y_1, \ldots, Y_p)' \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ and $\Sigma = Cov(\mathrm{Y})$

$$E(\mathrm{Y}) = (E(Y_1), \ldots, E(Y_p))' = \boldsymbol{\mu}$$

$$\Sigma = Cov(\mathrm{Y}) = \begin{pmatrix} Var(Y_1) & Cov(Y_1, Y_2) & \cdots & Cov(Y_1, Y_p) \\ Cov(Y_2, Y_1) & Var(Y_2) & \cdots & Cov(Y_2, Y_p) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_p, Y_1) & Cov(Y_p, Y_2) & \cdots & Var(Y_p) \end{pmatrix}$$

For $Y \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$, its probability density function (pdf) is

$$f(y) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(y - \boldsymbol{\mu})' \Sigma^{-1} (y - \boldsymbol{\mu}) \right\}$$

- $\Omega = \Sigma^{-1}$ is called the precision matrix

GP is fully specified once we specify its mean function $\mu(\cdot)$ and covariance function $C(\cdot, \cdot)$

- Mean function $\mu(\cdot)$: a function from $\mathcal{D} \to \mathcal{R}$

$$E(Y(\mathrm{s})) = \mu(\mathrm{s})$$

- Covariance function $C(\cdot, \cdot)$: a function from $\mathcal{D} \times \mathcal{D} \to \mathcal{R}$

$$Cov(Y(\mathrm{s}), Y(\mathrm{u})) = C(\mathrm{s}, \mathrm{u})$$

- For any n and any $\mathrm{s}_1, \ldots, \mathrm{s}_n \in \mathcal{D}$,

$$(Y(\mathrm{s}_1), \ldots, Y(\mathrm{s}_n))' \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$$

  - $\boldsymbol{\mu} = (\mu(\mathrm{s}_1), \ldots, \mu(\mathrm{s}_n))'$
  - The (ij)-th element of Σ is $C(\mathrm{s}_i, \mathrm{s}_j)$

Mean Function

- Some common choices:
  - Constant $\mu(\mathrm{s}) = \beta$
  - Regression type: $\mu(\mathrm{s}) = X(\mathrm{s})'\boldsymbol{\beta} = \sum_{i=1}^{p} X_i(\mathrm{s})\beta_i$
    - We can include spatial coordinates as covariates, e.g., quadratic terms
- "Mean vs Covariance"
  - One's mean can be the other's second-order variation to be modeled in covariance term
  - A simple mean function will leave more variation to the second-order variation and strong covariance
  - A very sophisticated mean structure may result in residuals approximately independence

- We usually assume the correlation decays with distance, but it may not always hold.
- Unlike the mean function for which any function can be valid, covariance functions need to satisfy a few conditions.
- We may also consider some stronger assumptions.

The covariance function $C(\cdot, \cdot)$ satisfies:

1. Symmetry, i.e.

$$C(\mathrm{s}, \mathrm{u}) = C(\mathrm{u}, \mathrm{s})$$

   for all $\mathrm{s}, \mathrm{u} \in \mathcal{D}$.

2. Nonnegative definiteness,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j C(\mathrm{s}_i, \mathrm{s}_j) \geq 0$$

   for all n, all sequences $\{a_i : i = 1, \ldots, n\}$ and all sequences of spatial locations $\{\mathrm{s}_i : i = 1, \ldots, n\}$. (Nonnegative definiteness).

   - Why?
   - This may not be easy to prove for a given function.

- For a random process $\{Y(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$, we only observe $\mathbf{Y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))'$. We usually need to make some assumptions about its covariance function $C(\cdot, \cdot)$.
- We cannot use sample covariance matrix:
  - We only observe data at n locations. No replicates
  - This is also why we usually need to make assumptions about $C(\cdot, \cdot)$
  - Different from typical examples in Functional Data Analysis, e.g., Yao et al. (2005) Functional Data Analysis for Sparse Longitudinal Data, JASA, 100, 577-590.
- A basic assumption is **stationarity**!

Covariance Assumptions

- Stationarity: Invariance to shifts
  - Weak (second-order) stationarity
  - Intrinsic stationarity
  - Strict (strong) stationarity

Weak stationarity (second-order stationarity)

- Mean stationarity: The mean is constant

$$E(Y(\mathrm{s})) = \mu(\mathrm{s}) = \mu, \text{ for all } \mathrm{s} \in \mathcal{D}$$

- Covariance stationarity: the covariance at two sites depends on only the sites' **relative positions**.

$$C(\mathrm{s}, \mathrm{u}) = C(\mathrm{s} - \mathrm{u}) = C(\mathrm{d})$$

with $\mathrm{d} = \mathrm{s} - \mathrm{u}$.

Intrinsic stationarity

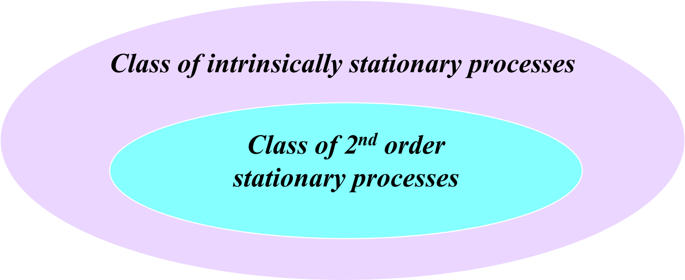
Intrinsic stationarity is more general compared to weak stationarity

- Mean stationarity: $E(Y(\mathrm{s})) = \mu(\mathrm{s}) = \mu,$ for all $\mathrm{s} \in \mathcal{D}$
- $\mathrm{E}[(Y(\mathrm{s}) - Y(\mathrm{u}))^2]$ depends on only the sites' **relative positions** $\mathrm{s} - \mathrm{u}$, i.e. there is a function γ such that:

$$\mathrm{E}[(Y(\mathrm{s}) - Y(\mathrm{u}))^2] = 2\gamma(\mathrm{d})$$

  with $\mathrm{d} = \mathrm{s} - \mathrm{t}$.
- $\gamma(\cdot)$ is called the semivariogram, and $2\gamma(\cdot)$ is the variogram.

Weak stationarity $\Longrightarrow$ intrinsic stationarity, but not vice versa.



An example is given in HW 1.

Strong (Strict) Stationarity

The joint probability distribution of the data depends only on the relative positions of the sites at which the data were taken, i.e. the joint distribution of

$$(Y(\mathrm{s}_1), Y(\mathrm{s}_2), \ldots, Y(\mathrm{s}_n))$$

is the same as

$$(Y(\mathrm{s}_1 + \mathrm{d}), Y(\mathrm{s}_2 + \mathrm{d}), \ldots, Y(\mathrm{s}_n + \mathrm{d}))$$

for any n spatial locations $\mathrm{s}_1, \ldots, \mathrm{s}_n$ and any d.

- For GP, a weakly stationary GP is also strongly stationary

Isotropy

- Isotropy: Invariance to shift and rotation
- Isotropic covariance function:

$$C(\mathrm{s}, \mathrm{u}) = C(h), \text{ where } h = \|\mathrm{s} - \mathrm{u}\|$$

- A covariance is not isotropic is called anisotropic. Examples?

Nugget Effect

- We often decompose the errors into two parts: spatial + non-spatial
- The non-spatial component is usually modeled as iid measurement errors, called the nugget effect in spatial statistics.

$$\epsilon(s) = \epsilon_s(s) + \epsilon_{ne}(s)$$

where

- $\epsilon_s(s)$ are spatially correlated with $var(\epsilon_s(s)) = \sigma^2$ and correlation function $\rho(d)$
- $\epsilon_{ne}(s)$ are iid measurement errors, with $var(\epsilon_{ne}(s)) = \tau^2$ called the nugget

The covariance for ϵ is then represented as

$$C(\mathrm{d};\theta) = \begin{cases} \tau^2 + \sigma^2\rho(0) = \tau^2 + \sigma^2 & \text{if } \mathrm{d} = 0 \\ \sigma^2\rho(\mathrm{d}) & \text{if } \|\mathrm{d}\| > 0. \end{cases}$$

- τ^2 is called the *nugget*.
- σ^2 is called the *partial sill*
- The total variance is called the *sill*, $var(Y(\mathrm{s})) = \sigma^2 + \tau^2$

Some Isotropic Covariance Functions

Model	$C(h)$
Spherical	$C(h) = \begin{cases} 0 & \text{if } h \geq \phi \\ \sigma^2[1 - \frac{3}{2}\frac{h}{\phi} + \frac{1}{2}(\frac{h}{\phi})^3] & \text{if } 0 < h < \phi \\ \tau^2 + \sigma^2 & \text{if } h = 0 \end{cases}$
Exponential	$C(h) = \begin{cases} \sigma^2 \exp(-\frac{h}{\phi}) & \text{if } h > 0 \\ \tau^2 + \sigma^2 & \text{if } h = 0 \end{cases}$
Powered exponential	$C(h) = \begin{cases} \sigma^2 \exp(-(\frac{h}{\phi})^p) & \text{if } h > 0 \\ \tau^2 + \sigma^2 & \text{if } h = 0 \end{cases}$
Matérn at $\nu = 3/2$	$C(h) = \begin{cases} \sigma^2(1 - \frac{\sqrt{3}h}{\phi}) \exp(-\frac{\sqrt{3}h}{\phi}) & \text{if } h > 0 \\ \tau^2 + \sigma^2 & \text{if } h = 0 \end{cases}$

- All decay as distance increases
- ϕ is called the range parameter
- Powered exponential with $k = 2$ is called squared-exponential (Gaussian) covariance function, commonly used in computer experiments

Matérn Class of Covariance Function

$$C(h) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\sqrt{2\nu}\frac{h}{\phi})^{\nu} K_{\nu}(\sqrt{2\nu}\frac{h}{\phi}) & \text{if } h > 0 \\ \tau^2 + \sigma^2 & \text{if } h = 0 \end{cases}$$

where K_ν is the modified Bessel function of order ν (computationally tractable in geoR) and $\Gamma(\cdot)$ is the gamma function

- ϕ is the range parameter
- ν is the smoothness parameter:
  - $\nu = 1/2 \implies$ exponential
  - $\nu = 3/2 \implies$ convenient closed form for $C(h)$ and $\gamma(h)$
  - $\nu \to \infty \implies$ squared exponential
- $\frac{\partial^k \epsilon_s(s)}{\partial s}$ exists if $k < \nu$

- Valid covariances on a sphere are also available
- Reading assignment: Jeong and Jun (2015) Covariance models on the surface of a sphere: when does it matter? STAT, 4, 167-182.

Anisotropic Covariance Functions

Example:

$$C(\mathrm{s}, \mathrm{u}) = \sigma^2 \exp\left(-\sqrt{(\mathrm{s}-\mathrm{u})' \mathrm{A} (\mathrm{s}-\mathrm{u})}\right)$$

for $\mathrm{s}, \mathrm{u} \in \mathcal{D} \subset \mathcal{R}^2$; A is a 2×2 matrix.

- $\mathrm{A} = \frac{1}{\phi} \mathrm{I} \rightarrow$ we get the isotropic exponential covariance function
- $\mathrm{A} = diag(\frac{1}{\phi_1}, \frac{1}{\phi_2}) \rightarrow$ we have different range parameters for the N-S and W-E directions
- When A is a generic matrix, we are allowing different correlations in elliptical directions

From basic covariance models, we can construct more complicated models using the following rules:

- If C_1 and C_2 are valid covariances, then so is $C(\cdot) \equiv C_1(\cdot) + C_2(\cdot)$.
- If C_0 is a valid covariance and $b > 0$, then $C(\cdot) \equiv b \cdot C_0(\cdot)$ is a valid covariance.
- If C_1 and C_2 are valid covariances, then so is $C(\cdot) \equiv C_1(\cdot) \cdot C_2(\cdot)$.
    - **Separability**: for example, a (2-D) covariance function C is said to be separable if

      $$C(\mathrm{h}) = C(h_1, h_2) = C_1(h_1)C_2(h_2)$$

      for two valid (in $\mathbb{R}$) covariance functions C_1 and C_2.
- A valid isotropic covariance function in $\mathbb{R}^{d_1}$ may not be a valid isotropic covariance function in $\mathbb{R}^{d_2}$ where $d_2 > d_1$. The converse is true.

Intrinsic Stationarity and the Semivariogram

- Intrinsic stationarity:
  - the mean is constant, $E(Y(\mathrm{s})) = \mu$
  - there is a function γ such that:

$$\frac{1}{2}\mathrm{E}[(Y(\mathrm{s}) - Y(\mathrm{u}))^2] = \gamma(\mathrm{s} - \mathrm{u})$$

    γ is called semivariogram.

- Some facts:
  - A weakly (second-order) stationary random process with covariance $C(\cdot, \cdot)$ is intrinsically stationary, with semivariogram

$$\gamma(\mathrm{h}) = C(0) - C(\mathrm{h})$$

    the converse is not true in general.
  - Note that separability of the covariance function does not imply that

$$\gamma(\mathrm{h}) = \gamma_1(h_1)\gamma_2(h_2)$$

Issues choosing a model for the semivariogram:

1. Vanishes at 0, i.e., $\gamma(0) = 0$ by definition.
2. Evenness, i.e. $\gamma(-h) = \gamma(h)$.
3. Conditional non-positive definiteness, i.e.

$$\sum_i \sum_j \lambda_i \lambda_j \gamma(s_i - s_j) \leq 0$$

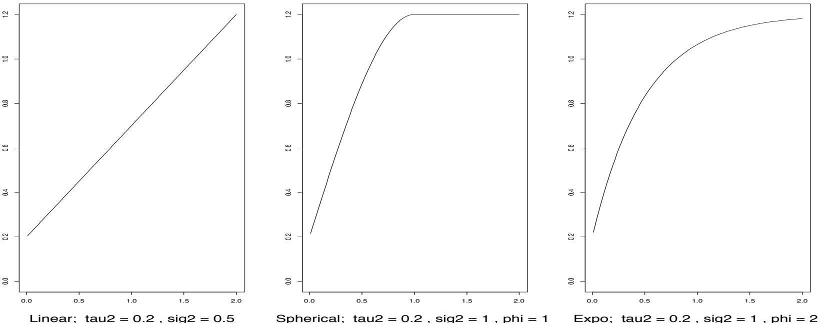
for all $s_1, s_2, \ldots$ and all $\lambda_1, \lambda_2, \ldots$ such that $\sum_i \lambda_i = 0$.
(Related question in HW 2)

- While $\gamma(0) = 0$ by definition, $\gamma(0+) = \lim_{h \to 0+} \gamma(h)$ is the nugget, denoted by τ^2
- $\lim_{h \to \infty} \gamma(h)$ is the sill, and it is equivalent to $C(0)$ (The *sill minus the nugget* is the *partial sill*)
- The value h when $\gamma(\cdot)$ first reaches and stay at the sill is called the range, R (same as h at which $C(\cdot)$ reaches and stays at 0)
- Example: Spherical semivariogram

$$\gamma(h) = \begin{cases} \tau^2 + \sigma^2 & \text{if } h \geq \phi \\ \tau^2 + \sigma^2[\frac{3}{2}\frac{h}{\phi} - \frac{1}{2}(\frac{h}{\phi})^3] & \text{if } 0 < h < \phi \\ 0 & o.w.(\text{h=0}) \end{cases}$$

  - Nugget: τ^2
  - Sill: $\tau^2 + \sigma^2$, and σ^2 is the partial sill
  - Range: $R = \phi$, and $\frac{1}{\phi}$ is called the decay parameter

Examples: Linear (left), Spherical (middle), and Exponential (right)



- The linear model: no corresponding 2nd-order stationary process exists
- Both sill and range can be infinite

University of Cincinnati

Some Isotropic Semivariograms

Model	Semivariogram $\gamma(h)$
Linear	$\gamma(h) = \begin{cases} \tau^2 + \sigma^2 h & \text{if } h > 0 \\ 0 & o.w. \end{cases}$
Spherical	$\gamma(h) = \begin{cases} \tau^2 + \sigma^2 & \text{if } h \geq \phi \\ \tau^2 + \sigma^2[\frac{3}{2}\frac{h}{\phi} - \frac{1}{2}(\frac{h}{\phi})^3] & \text{if } 0 < h < \phi \\ 0 & o.w. \end{cases}$
Exponential	$\gamma(h) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\frac{h}{\phi})) & \text{if } h > 0 \\ 0 & o.w. \end{cases}$
Powered exponential	$\gamma(h) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\lvert\frac{h}{\phi}\rvert^p)) & \text{if } h > 0 \\ 0 & o.w. \end{cases}$

Properties of $C(\mathrm{h})$ and $\gamma(\mathrm{h})$, $\mathrm{h} \in \mathbb{R}^d$

- When $\mathrm{h} = 0$, $C(0) = \sigma^2 + \tau^2$; $\gamma(0) = 0$
- $C(\mathrm{h}) = C(-\mathrm{h})$, $\gamma(\mathrm{h}) = \gamma(-\mathrm{h})$
- as $\mathrm{h} \to 0$, $C(\mathrm{h}) \to$ partial sill σ^2; $\gamma(\mathrm{h}) \to \tau^2$ nugget
- $C(\cdot)$ nonnegative-definite; $\gamma(\cdot)$ conditionally negative-definite

Summary

- Gaussian process
- Mean function
- Covariance function and Semivariogram

Preview:

- Estimation and Prediction