Lecture 2: Geostatistical Processes and the Gaussian Process

Geostatistical Processes

- stochastic processes defined continuously over space
- general applied review can be found
  - Diggle et. al. (1998)
  - Cressie (2015)
- for a measure theoretic overview
  - Adler and Taylor (2007) Random Fields & Geometry

- example: temperature fields across regions
  R code: temperature_measurments_Ukraine.R

Definition: A geostatistical process is a stochastic process

$$\{X(t) : t \in D \subset \mathbb{R}^p\}$$

- the stochastic process is defined on a probability space (see Adler & Taylor, 2007)
- the location t varies continuously in D
- can consider this spatial process to be an uncountable collection of random variables indexed by their location t
- natural to assume that $X(t)$ are dependent (goal: model the dependence)
- R code: gaussian-process-realizations.R
  4 realizations (samples) from a geostatistical model

Common questions of interest:

- What is the average value of $X(t)$ across spatial domain $D \in \mathbb{R}^p$
- What is the value of $X(t_0)$ at an unobserved location $t_0 \in D$
- What is the covariance between the values at two locations

$t_0 \in D$

- What is the covariance between the values at two locations i.e. $Cov(X(t), X(s))$ for $t, s \in D$

<u>Definition</u>: The mean function of $\{X(t) : t \in D\}$ is the expected value of the geostatistical process as a function of spatial location

$$m(t) = E(X(t)), \quad t \in D$$

<u>Definition</u>: The covariance function of $\{X(t) : t \in D\}$ is defined as

$$\begin{aligned} C(t,s) &= Cov(X(t), X(s)) \\ &= E\{[X(t) - m(t)][X(s) - m(s)] \quad t, s \in D \end{aligned}$$

Covariance measures strength of dependence between two random variables $X(t)$ and $X(s)$. Properties:

1. $C(t,s) = C(s,t)$ for all $t, s \in D$
2. When $t = s$ we have
$$C(t,t) = Cov(X(t), X(t)) = Var(X(t))$$
3. $C(t,s)$ is non-negative definite

<u>Definition</u>: A geostatistical process $\{X(t) : t \in D\}$ is called stationary if:

1. $m(X(t)) = m$ for all $t \in D$
2. $Cov(X(t), X(t+h)) = C(t, t+h) = C(h)$ for all $t \in D$ and spatial lag h

<u>Definition</u>: A stationary geostatistical process $\{X(t) : t \in D\}$ is isotropic when

$$C(t,s) = C(\|t - s\|)$$

If the covariance function of the stationary process depends both on the direction and the distance between the locations, then the process is called anisotropic.

Smoothness of stochastic process

$D \in \mathbb{R}$

Definition: A geostatistical process $\{X(t) : t \in D\}$ is mean-square (MS) continuous if

$$\lim_{h \to 0} E\left[|X(t+h) - X(t)|^2 \right] = 0$$

Definition: A geostatistical process $\{X(t) : t \in D\}$ is mean-square differentiable if there exists a process $\frac{d}{dt}X(t)$ such that

$$\lim_{h \to 0} E\left[\left| \frac{X(t+h) - X(t)}{h} - \frac{d}{dt}X(t) \right|^2 \right] = 0$$

Example: Brownian motion is MS-continuous everywhere but MS-differentiable nowhere

Gaussian Processes (GP)

A type of geostatistical process that can be interpreted as a generalization of the multivariate normal model from random vectors to random functions

Definition: The process $\{X(t) : t \in D\}$ is a Gaussian process if the joint distribution of any finite subcollection of the random variables follows a MVN distribution

A GP is completely specified by

1. $m(t) = E(X(t)), \quad t \in D$
2. $C(t, s) = E\{[X(t) - m(t)][X(s) - m(s)]\}$

$t, s \in D$

2. $C(t,s) = E\{[X(t)-m(t)][X(s)-m(s)]\}$
$t, s \in D$

We write the GP model as $X \sim GP(m(t), C(t,s))$

The probability density function (pdf) of $X = (X(t_1), \dots, X(t_n))^T$ is

$$f(x) = (2\pi)^{-n/2} |C|^{-1/2} \exp\left[-\frac{1}{2}(x-m)^T C^{-1}(x-m)\right],$$

$$x \in \mathbb{R}^n$$

where $m = (m(t_1), \dots, m(t_n))$ and (j,k) element of the $n \times n$ covariance matrix C is $C(t_j, t_k)$

Gaussian process covariance models

Matérn class of covariance function

$$C_{\ell,\nu}(t,s) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{2\nu}{\ell} \|t-s\|\right)^{\nu} K_\nu\left(\frac{2\nu}{\ell}\|t-s\|\right)$$

- K_ν is the modified Bessel function of the 2^{nd} kind
- Γ is the gamma function
- $\|\cdot\|$ is the Euclidian norm
- $\ell \in \mathbb{R}^+$: length-scale parameter controls rate at which correlation decreases with distance in input space
- $\nu \in \mathbb{R}^+$: smoothess parameter controls the MS differentiability of the resulting GP

This covariance family contains a popular covariance called square exponential as a limiting case as $\nu \to \infty$

$$C(t,s) = \exp\left\{-\frac{\|t-s\|_2^2}{2\ell^2}\right\}$$