

Homework 1: Point referenced data (building concepts)

Lecturer: Georgios P. Karagiannis

georgios.karagiannis@durham.ac.uk

Exercise 1. (★) Let $Z = (Z(s) : s \in \mathbb{R}^d)$ be an intrinsic random field with $E(Z(s) - Z(t)) = 0$ and let $\gamma : \mathbb{R}^d \rightarrow \mathbb{R}$ be its semivariogram.

(1) Let $a \in \mathbb{R}^n$ be a vector of constants. Consider sites $\{s_1, \dots, s_n \subseteq \mathbb{R}^d\}$ Show that

$$\text{Var} \left(\sum_{i=1}^n a_i Z(s_i) \right) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j c_I(s_i, s_j)$$

where $c_I(s, t) = \gamma(s - s_0) + \gamma(t - s_0) - \gamma(s - t)$ at some additional $s_0 \in \mathbb{R}^d$.

(2) Show that for all $n \in \mathbb{N}$, $(a_1, \dots, a_n) \subseteq \mathbb{R}^n$ s.t. $\sum_{i=1}^n a_i = 0$, and for all $(s_1, \dots, s_n) \subseteq S^n$, it is

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(s_i - s_j) \leq 0$$

Solution. Assume origin $s_0 \in \mathbb{R}^d$ with random $Z(s_0)$.

(1) I use $Z(s_0)$ at some location let's say s_0 . It is

$$\begin{aligned} \text{Var} \left(\sum_{i=1}^n a_i Z(s_i) \right) &= \text{Var} \left(\sum_{i=1}^n a_i Z(s_i) - \overbrace{\sum_{i=1}^n a_i Z(s_0)}^{0=} \right) = \text{Var} \left(\sum_{i=1}^n a_i (Z(s_i) - Z(s_0)) \right) \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j E((Z(s_i) - Z(s_0))(Z(s_j) - Z(s_0))) \end{aligned}$$

Let $c_I(s, t) = E((Z(s_i) - Z(s_0))(Z(s_j) - Z(s_0)))$.

(2) It is

$$\begin{aligned} \gamma(s - t) &= \frac{1}{2} E(Z(s) - Z(s_0) + Z(t) - Z(s_0))^2 \\ &= \frac{1}{2} (2\gamma(s - s_0) + 2\gamma(t - s_0) - 2c_I(s, t)) \\ \implies c_I(s, t) &= \gamma(s - s_0) + \gamma(t - s_0) - \gamma(s - t) \end{aligned}$$

It is

$$\begin{aligned}
0 \leq \text{Var} \left(\sum_{i=1}^n a_i Z(s_i) \right) &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j c_I(s_i, s_j) \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j (\gamma(s_i) + \gamma(s_j) - \gamma(s_i - s_j)) \\
&= \sum_{i=1}^n a_i \gamma(s_i) \sum_{j=1}^n \cancel{a_j} + \sum_{j=1}^n a_j \gamma(s_j) \sum_{i=1}^n \cancel{a_i} - \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(s_i - s_j)
\end{aligned}$$

hence

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(s_i - s_j) \leq 0$$

Exercise 2. (★) Consider the zero-mean random field $Z = (Z(s) : s \in \mathbb{R}^d)$ with covariogram function given by

$$c(h) = \begin{cases} \xi^2 (1 + \rho \|h\|) \exp(-\rho \|h\|), & h > 0 \\ \nu^2 + \xi^2, & h = 0 \end{cases}$$

- (1) Compute the semivariogram for the random field $(Z(s) : s \in \mathbb{R}^d)$
- (2) What are the nugget, sill and partial sill for this covariance model? Justify your answer.
- (3) Would the slightly altered covariance function defined below be a good model for spatial data for $\phi > 0$? Justify your answer.

$$c(h) = \begin{cases} \xi^2 (1 + \rho \|h\|) \exp(-\rho \|h\|) + \phi, & h > 0 \\ \nu^2 + \xi^2 + \phi, & h = 0 \end{cases}$$

Solution.

- (1) For all $h \neq 0$, it is

$$\begin{aligned}
\gamma(h) &= c(0) - c(h), \\
&= \nu^2 + \xi^2 - \xi^2 (1 + \rho \|h\|) \exp(-\rho \|h\|) \\
&= \nu^2 + \xi^2 (1 - (1 + \rho \|h\|) \exp(-\rho \|h\|))
\end{aligned}$$

then

$$\gamma(h) = \begin{cases} \nu^2 + \xi^2 (1 - (1 + \rho \|h\|) \exp(-\rho \|h\|)) & h > 0 \\ 0 & h = 0 \end{cases}$$

- (2)

- The sill is the covariogram function at distance 0, that is $c(0) = \nu^2 + \xi^2$. Or since analogously, it is $\lim_{\|h\| \rightarrow \infty} \gamma(h)$. So,

$$\begin{aligned} \lim_{\|h\| \rightarrow \infty} (\|h\| \exp(-\rho \|h\|)) &= \lim_{\|h\| \rightarrow \infty} (\|h\| / \exp(\rho \|h\|)) \\ &= \lim_{\|h\| \rightarrow \infty} (\|h\| / \exp(\rho \|h\|)) = \lim_{\|h\| \rightarrow \infty} (\exp(-\rho \|h\|)) = 0 \end{aligned}$$

then

$$\lim_{\|h\| \rightarrow \infty} \gamma(h) = \nu^2 + \xi^2$$

- The nugget effect is the limiting value of the semicovariogram as $h \rightarrow 0$ from above, hence it is $\gamma(h) \rightarrow \nu^2$ as $h \rightarrow 0^+$.
 - The partial sill is the sill minus the nugget and is hence ξ^2 .
- (3) No, it would be unrealistic because if $\phi > 0$ then the covariance is always positive for infinitely large distances h . In practical terms this means that two points will always be correlated however far apart they are, it would be unrealistic.