Spatio-temporal statistics (MATH4341) Michaelmas term, 2023

Homework 1: Geostatistics (building concepts)

Lecturer: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

As formative assessment 1, submit the solutions to all the Exercises

---

Exercise 1. ($\star$) Let $Z = (Z_s)_{s\in\mathbb{R}^d}$ be an intrinsically stationary stochastic process, and let $\gamma : \mathbb{R}^d \to \mathbb{R}$ be its semivariogram. Assume $a \in \mathbb{R}^n$s.t. $\sum_{i=1}^{n} a_i = 0$.

(1) Let $a \in \mathbb{R}^n$ be a vector of constants. Show that

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

where $c_Y(s,t) = \operatorname{E}(Y(s)Y(t))$, and $Y_s = Z_s - Z_0$.

(2) Show that

$$c_Y(s,t) = \gamma(s) + \gamma(t) - \gamma(s-t)$$

(3) Show that for all $\forall \{s_1, ..., s_n\} \subseteq S$ it is

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

Solution. Assume $0 \in S$, and a random variable $Z(0)$. Let $Y_s = Z_s - Z_0$.

(1) It is

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^{n} a_i Z(s_i)\right) =& \operatorname{Var}\left(\sum_{i=1}^{n} a_i Z(s_i) - \overbrace{\sum_{i=1}^{n} a_i Z(0)}^{0=}\right) = \operatorname{Var}\left(\sum_{i=1}^{n} a_i Y(s_i)\right)\\ =& \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \operatorname{E}(Y(s_i)Y(s_j)) = c_Y(s,t)\end{aligned}$$

(2) For $\operatorname{E}(Y(s_i)) = 0$ it is

$$\begin{aligned}\gamma(s-t) &= \frac{1}{2}\operatorname{E}(Z(s) - Z(0) + Z(t) - Z(0))^2\\ &= \frac{1}{2}(2\gamma(s) + 2\gamma(t) - 2c_Y(s,t))\\ \implies c_Y(s,t) &= \gamma(s) + \gamma(t) - \gamma(s-t)\end{aligned}$$

(3) It is

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

hence

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

Exercise 2. ($\star$) Consider the zero-mean geostatistical process $Z=\left(Z_s\right)_{s \in \mathbb{R}^d}$ with a weakly stationary and isotropic covariance 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 semi-variogram for the geostatistical process $\left(Z_s\right)$

(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 covariance function at distance 0, that is $c(0) = \nu^2 + \xi^2$. Or since analogously, it is $\lim_{\|h\|\to\infty} \gamma(h)$. So,

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

  then

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

- The nugget effect is the limiting value of the semi-variogram as $h \to 0$ from above, hence it is $\gamma(h) \to \nu^2$ as $h \to 0^+$.
- 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.