Spatio-temporal statistics (MATH4341) Michaelmas term, 2023

Homework 2: Geostatistics (Kriging and MLE inference)

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

---

Exercise 1. ($\star$) Consider we the geostatistical model $(Z_s)_{s\in\mathcal{S}}$ with

$$Z(s) = \mu(s) + w(s) + \varepsilon(s)$$

where $w(s)$ is a weakly stationary process with mean zero and covariogram $c_w(h; \sigma^2, \phi) = \sigma^2 \exp\left(-\frac{1}{\phi}\|h\|\right)$, $\mu(s;\beta)$ is a deterministic function

$$\mu(s;\beta) = \sum_{j=0}^{p} \psi_j(s)\,\beta_j = (\psi(s))^\top \beta$$

with unknown coefficients $\beta = (\beta_0, ..., \beta_p)^\top$ and known basis functions $\psi(s) = (\psi_0(s), ..., \psi_p(s))^\top$, $\varepsilon(s)$ is a nugget effect process whose covariogram has sill τ^2, and assume that $w(s)$ and $\varepsilon(s)$ are independent Gaussian Processes.

(1) Write down the formula of the covariogram $c\left(h; \left(\sigma^2, \phi, \tau\right)\right)$ of (Z_s).

(2) Consider a re-parametrization $\theta = \left(\sigma^2, \phi, \xi\right)$ where $\xi^2 = \frac{\tau^2}{\sigma^2}$ is called signal to noise ratio. Assume there is available a dataset $\{(s_i, Z_i)\}_{i=1}^n$ where $, Z_i := Z(s_i)$ is a realization of $(Z_s)_{s\in\mathcal{S}}$ at site s_i.

  (a) Let Ψ be a matrix with $[\Psi]_{i,j} = \psi_j(s_i)$. Let D be a matrix such as $[D]_{i,j} = \|s_i - s_j\|$. Consider that you can use convenient notation such as $\exp(D)$ meaning $[\exp(D)]_{i,j} = \exp(D_{i,j})$. Write down the covariance matrix $C(\theta)$ of $Z = (Z_1, ..., Z_n)^\top$ as a function of D and θ.

  (b) Write down the log likelihood function $\log(L(Z;\theta))$ of $Z = (Z_1, ..., Z_n)^\top$ given $\theta = \left(\sigma^2, \phi, \xi\right)$.

(3) Let $r(\cdot)$ (called correlogram) such as $c(\cdot) = \sigma^2 r(\cdot)$. Assume that (ϕ, ξ) as known constants.

  (a) Compute the likelihood equations[1] w.r.t. $\left(\beta, \sigma^2\right)$, and for given (ϕ, ξ).

  (b) Compute the MLE $\hat{\beta}_{(\phi,\xi)}$ of β as a function of (ϕ, ξ)

  (c) Compute the MLE $\hat{\sigma}^2_{(\phi,\xi)}$ of σ^2 as a function of (ϕ, ξ).

  (d) Compute the unbiased estimator of $\tilde{\sigma}^2$ of σ^2.

  Hint: Consider the fitted values $e = (e_1, ..., e_n)^\top$ as $e = [I - H]\,Z$ where $H = \left(\Psi^\top R^{-1} \Psi\right)^{-1} \Psi^\top R^{-1}$, and write $\hat{\sigma}^2_{(\phi,\xi)}$ w.r.t. e.

  Hint: It is given that $\mathrm{E}\left(Z^\top A Z\right) = \mathrm{E}(Z)^\top A \mathrm{E}(Z)^\top + \mathrm{tr}(A\mathrm{Var}(Z))$ when $Z \sim$ Normal

[1] that is, the gradient of the log-likelihood

(4) Compute the so-called log "profiled likelihood" $\log\left(L\left(Z;(\phi,\xi)\right)\right)$ resulting as

$$L\left(Z;(\phi,\xi)\right)=L\left(Z;\beta=\hat{\beta}_{(\phi,\xi)},\sigma^2=\hat{\sigma}^2_{\left(\hat{\beta}_{(\phi,\xi)},\phi,\xi\right)},\phi,\xi\right)$$

by replacing the β with $\hat{\beta}_{(\phi,\xi)}$ and σ^2 with $\hat{\sigma}^2_{\left(\hat{\beta}_{(\phi,\xi)},\phi,\xi\right)}$ in the actual likelihood $L\left(Z;\beta,\theta=\left(\sigma^2,\phi,\xi\right)\right)$. Describe how you would compute suitable values $\left(\hat{\phi},\hat{\xi}\right)$ for the MLE of (ϕ,ξ)

Solution. It is

(1) It is

$$\begin{aligned} c\left(h;\left(\sigma^2,\phi,\tau\right)\right) =& c_\delta\left(h;\sigma^2,\phi\right)+c_\varepsilon\left(h;\tau\right) \\ =& \sigma^2\exp\left(-\frac{1}{\phi}\|h\|\right)+\tau 1_{\{0\}}\left(h\right) \end{aligned}$$

(2) It is

(a)

$$C\left(\sigma^2,\phi,\xi\right)=\sigma^2\exp\left(-\frac{1}{\phi}D\right)+\sigma^2\xi^2 I$$

(b)

$$\begin{aligned} 2\log\left(L\left(Z;\theta\right)\right) =& 2\log\left(\mathrm{N}\left(Z|\Psi\beta,C\left(\theta\right)\right)\right) \\ =& -n\log\left(\sigma^2\right)-\log\left(\left|\exp\left(-\frac{1}{\phi}D\right)+\xi^2 I\right|\right) \\ & -\frac{1}{\sigma^2}\left(Z-\Psi\beta\right)^\top\left(\exp\left(-\frac{1}{\phi}D\right)+\xi^2 I\right)^{-1}\left(Z-\Psi\beta\right) \end{aligned}$$

(3) It is

$$\begin{aligned} 2\log\left(L\left(Z;\theta\right)\right) =& -n\log\left(\sigma^2\right)-\log\left(\left|\exp\left(-\frac{1}{\phi}D\right)+\xi^2 I\right|\right) \\ =& \quad -\frac{1}{\sigma^2}\left(Z-\Psi\beta\right)^\top\left(\exp\left(-\frac{1}{\phi}D\right)+\xi^2 I\right)^{-1}\left(Z-\Psi\beta\right) \end{aligned}$$

Let $R_{(\phi,\xi)}$ matrix with $\left[R_{(\phi,\xi)}\right]_{i,j}=r\left(s_i-s_j|\phi,\xi\right)$

(a) So the likelihood equations are $0=\nabla_{(\beta,\sigma^2)}\log\left(L\left(Z;\theta\right)\right)$

$$\begin{cases} 0=\Psi^\top\left(R_{(\phi,\xi)}\right)^{-1}Z-\Psi^\top\left(R_{(\phi,\xi)}\right)^{-1}\Psi\beta \\ 0=\frac{n}{2}\frac{1}{\sigma^2}+\frac{1}{2(\sigma^2)^2}\left(Z-\Psi\beta\right)^\top\left(R_{(\phi,\xi)}\right)^{-1}\left(Z-\Psi\beta\right) \end{cases}$$

(b) It is

$$\hat{\beta}_{(\phi,\xi)}=\left(\Psi^\top\left(R_{(\phi,\xi)}\right)^{-1}\Psi\right)^{-1}\Psi^\top\left(R_{(\phi,\xi)}\right)^{-1}Z$$

(c) It is

$$\hat{\sigma}_{(\beta,\phi,\xi)}=\frac{1}{n}\left(Z-\Psi\beta\right)^\top\left(R_{(\phi,\xi)}\right)^{-1}\left(Z-\Psi\beta\right)$$

and by subtituting I get

$$
\begin{aligned}
\hat{\sigma}_{(\phi,\xi)} =& \hat{\sigma}_{\left(\hat{\beta}_{(\phi,\xi)},\phi,\xi\right)} = \frac{1}{n}\left(Z - \Psi\hat{\beta}_{(\phi,\xi)}\right)^\top \left(R_{(\phi,\xi)}\right)^{-1}\left(Z - \Psi\hat{\beta}_{(\phi,\xi)}\right) \\
=& \frac{1}{n}\left(Z - \Psi\hat{\beta}_{(\phi,\xi)}\right)^\top \left(R_{(\phi,\xi)}\right)^{-1}\left(Z - \Psi\hat{\beta}_{(\phi,\xi)}\right)
\end{aligned}
$$

(d) It is

$$
e = Z - \Psi\hat{\beta}_{(\phi,\xi)} = (I - H)\, Z
$$

So

$$
\begin{aligned}
n\hat{\sigma}_{(\phi,\xi)} =& Z^\top (I - H)\left(R_{(\phi,\xi)}\right)^{-1}(I - H)\, Z \\
=& \left[(I - H)\, Z\right]^\top \left(R_{(\phi,\xi)}\right)^{-1}\left[(I - H)\, Z\right] \\
=& e^\top R_{(\phi,\xi)} e
\end{aligned}
$$

where

$$
\mathrm{E}\left[e\right] = 0
$$

then

$$
\begin{aligned}
\mathrm{E}\left(n\hat{\sigma}\left(\phi,\xi\right)\right) =& \mathrm{E}\left(Z^\top (I - H)\left(R_{(\phi,\xi)}\right)^{-1}(I - H)\, Z\right) \\
=& \overset{=0}{\cancel{\left(\mathrm{E}\left[e\right]\right)^\top}}\left(R_{(\phi,\xi)}\right)^{-\frac{1}{2}}\overset{=0}{\cancel{\mathrm{E}\left[e\right]}} + \mathrm{tr}\left(\left(R_{(\phi,\xi)}\right)^{-1}\mathrm{Var}\left(e\right)\right) \\
=& \mathrm{tr}\left(\left(R_{(\phi,\xi)}\right)^{-1}\mathrm{Var}\left((I - H)\, Z\right)\right) \\
=& \mathrm{tr}\left(\left(R_{(\phi,\xi)}\right)^{-1}(I - H)\,\sigma^2 R_{(\phi,\xi)}\,(I - H)\right) = \sigma^2\mathrm{tr}\left(\left(R_{(\phi,\xi)}\right)^{-1}(I - H)\, R_{(\phi,\xi)}\,(I - H)\right) \\
=& \mathrm{tr}\left((I - H)\right) = \sigma^2\left(n - p\right)
\end{aligned}
$$

So it is

$$
\tilde{\sigma}\left(\beta,\phi,\xi\right) = \frac{1}{n - p}\left(Z - \Psi\beta\right)^\top\left(R_{(\phi,\xi)}\right)^{-1}\left(Z - \Psi\beta\right)
$$

because

$$
\mathrm{E}\left(\tilde{\sigma}\left(\beta,\phi,\xi\right)\right) = \sigma^2
$$

(4) It is

$$
\begin{aligned}
\log\left(L\left(Z;(\phi,\xi)\right)\right) = L\left(Z;\beta = \hat{\beta}_{(\phi,\xi)},\sigma^2 = \hat{\sigma}^2_{\left(\hat{\beta}_{(\phi,\xi)},\phi,\xi\right)},\phi,\xi\right) \\
-\frac{n}{2}\log\left(\hat{\sigma}^2_{\left(\hat{\beta}_{(\phi,\xi)},\phi,\xi\right)}\right) - \frac{1}{2}\log\left(\left|R_{(\phi,\xi)}\right|\right)
\end{aligned}
$$

where obviously

$$
0 = \nabla_{(\phi,\xi)}\log\left(L\left(Z;(\phi,\xi)\right)\right)\Big|_{(\phi,\xi)=\left(\hat{\phi},\hat{\xi}\right)}
$$

cannot be solved numerically. The Newton method or the gradient descent method can be used to maximize $\log\left(L\left(Z;(\phi,\xi)\right)\right)$.

Exercise 2. ($\star$) Let $(Z_s)_{s\in\mathcal{S}}$ be a specified statistical model. Assume that process $(Z_s)_{s\in\mathcal{S}}$ has known mean $\mu(s) = \mathrm{E}(Z(s))$ and known covariance function $c(\cdot,\cdot)$. Assume there is available a dataset $\{(s_i, Z_i := Z(s_i))\}_{i=1}^{n}$. Assume that the matrix C such as $[C]_{i,j} = c(s_i, s_j)$ has an inverse. Consider the "Kriging" estimator μ_{SK} Consider the "Kriging" estimator $Z_{\text{SK}}(s_0)$ of $Z(s_0)$ at an unseen spatial location s_0 as the BLUE (Best Linear Unbiased Estimator)

$$Z_{\text{SK}}(s_0) = w_{n+1} + \sum_{i=1}^{n} w_i Z(s_i) = w_{n+1} + w^\top Z,$$

for some unknown $\{w_i\}$ that we need to learn, and $Z = (Z_1, ..., Z_n)^\top$. Let $w = (w_1, ..., w_n)^\top$.

(1) Find sufficient conditions on $w = (w_1, ..., w_n)^\top$ so that the Kriging estimator $Z_{\text{SK}}(s_0)$ to be unbiased.

(2) Derive the MSE of $Z_{\text{SK}}(s_0)$ as

$$\mathrm{E}(Z_{\text{SK}}(s_0) - Z(s_0))^2 = w^\top C w + c(s_0, s_0) - 2w^\top C_0$$

where C_0 is a vector such as $[C_0]_i = c(s_0, s_i)$.

(3) Derive the Kriging estimator of $Z(s_0)$ as

$$Z_{\text{SK}}(s_0) = \mu(s_0) + C_0^\top C^{-1}[Z - \mu(s_{1:n})]$$

where $\mu(s_{1:n})$ is a vector such as $[\mu(s_{1:n})]_i = \mu(s_i)$.

(4) Compute the Kriging standard error $\sigma_{\text{SK}} = \sqrt{\mathrm{E}(Z_{\text{SK}}(s_0) - Z(s_0))^2}$.

Solution. The method is called Simple Kriging, and hence we denote it as SK.

(1) It is

$$Z_{\text{SK}}(s_0) = w_{n+1} + \sum_{i=1}^{n} w_i Z(s_i) = w_{n+1} + w^\top Z,$$

where $\{w_i\}$ is a set of unknown weights to be learned.

We assume that assume zero systematic error (unbiasness), hence

$$\mathrm{E}(Z_{\text{SK}}(s_0) - Z(s_0)) = \mathrm{E}\left(w_{n+1} + \sum_{i=1}^{n} w_i Z(s_i) - Z(s_0)\right) = w_{n+1} + \sum_{i=1}^{n} w_i \mu(s_i) - \mu(s_0)$$

which is satisfied given the assumption

$$w_{n+1} = \mu(s_0) - \sum_{i=1}^{n} w_i \mu(s_i) \Longleftrightarrow w_{n+1} = \mu(s_0) - w^\top \mu(s_{1:n})$$

where $w = (w_1, ..., w_n)^\top$.

(2) It is

$$
\begin{aligned}
\mathrm{E}\left(Z_{\mathrm{SK}}\left(s_0\right)-Z\left(s_0\right)\right)^2 = & \operatorname{Var}\left(Z_{\mathrm{SK}}\left(s_0\right)-Z\left(s_0\right)\right)=\operatorname{Var}\left(w_{n+1}+w^{\top} Z-Z\left(s_0\right)\right) \\
= & \operatorname{Var}\left(w_{n+1}+w^{\top} Z\right)+\operatorname{Var}\left(Z\left(s_0\right)\right)-2 \operatorname{Cov}\left(w_{n+1}+w^{\top} Z, Z\left(s_0\right)\right) \\
= & w^{\top} C w+c(s_0, s_0)-2 w^{\top} \operatorname{Cov}\left(Z, Z\left(s_0\right)\right) \\
= & w^{\top} C w+c(s_0, s_0)-2 w^{\top} C_0
\end{aligned}
$$

where $C_0=\operatorname{Cov}\left(Z, Z\left(s_0\right)\right)$, i.e. $\left[C_0\right]_j=c\left(s_j, s_0\right)$.

(3) To learn the unknown weights $\{w_i\}$ we need to solve

$$
w^{\mathrm{SK}}=\arg \min_w \mathrm{E}\left(Z_{\mathrm{SK}}\left(s_0\right)-Z\left(s_0\right)\right)^2, \text{ subject to } w_{n+1}=\mu\left(s_0\right)-w^{\top} \mu\left(s_{1: n}\right)
$$

As $\mathrm{E}\left(\mu_{\mathrm{SK}}-Z\left(s_0\right)\right)^2$ does not depend on w_{n+1} we minimize

$$
\begin{aligned}
0 & =\nabla_w \mathrm{E}\left(Z_{\mathrm{SK}}\left(s_0\right)-Z\left(s_0\right)\right)^2=\nabla_w \operatorname{Var}\left(Z_{\mathrm{SK}}\left(s_0\right)-Z\left(s_0\right)\right) \\
& =2 C w-2 C_0
\end{aligned}
$$

So I get

$$
w_{\mathrm{SK}}=C^{-1} C_0
$$

So

$$
\begin{aligned}
Z_{\mathrm{SK}}\left(s_0\right) & =w_{n+1}+C^{-1} C_0 Z \\
& =\mu\left(s_0\right)-\left(C^{-1} C_0\right)^{\top} \mu\left(s_{1: n}\right)+\left(C^{-1} C_0\right)^{\top} Z \\
& =\mu\left(s_0\right)+C_0^{\top} C^{-1}\left[Z-\mu\left(s_{1: n}\right)\right]
\end{aligned}
$$

(4) It is

$$
\begin{aligned}
\sigma_{\mathrm{SK}} & =\sqrt{\mathrm{E}\left(Z_{\mathrm{SK}}\left(s_0\right)-Z\left(s_0\right)\right)^2} \\
& =\sqrt{w_{\mathrm{SK}}^{\top} C w_{\mathrm{SK}}+c(s_0, s_0)-2 w_{\mathrm{SK}}^{\top} C_0} \\
& =\sqrt{c(s_0, s_0)-C_0^{\top} C^{-1} C_0}
\end{aligned}
$$