

## Homework 2: Geostatistics (Kriging and MLE inference)

Lecturer: Georgios P. Karagiannis

georgios.karagiannis@durham.ac.uk

**Exercise 1.** (★) 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  $\beta = (\beta_0, \dots, \beta_p)^\top$  and known basis functions  $\psi(s) = (\psi_0(s), \dots, \psi_p(s))^\top$ ,  $\varepsilon(s)$  is a nugget effect process whose covariogram has sill  $\tau^2$ , and assume that  $w(s)$  and  $\varepsilon(s)$  are independent Gaussian Processes.

- (1) Write down the formula of the covariogram  $c(h; (\sigma^2, \phi, \tau))$  of  $(Z_s)$ .
- (2) Consider a re-parametrization  $\theta = (\sigma^2, \phi, \xi)$  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  $\Psi$  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  $Z = (Z_1, \dots, Z_n)^\top$  as a function of  $D$  and  $\theta$ .
  - (b) Write down the log likelihood function  $\log(L(Z; \theta))$  of  $Z = (Z_1, \dots, Z_n)^\top$  given  $\theta = (\sigma^2, \phi, \xi)$ .
- (3) Let  $r(\cdot)$  (called correlogram) such as  $c(\cdot) = \sigma^2 r(\cdot)$ . Assume that  $(\phi, \xi)$  as known constants.
  - (a) Compute the likelihood equations<sup>1</sup> w.r.t.  $(\beta, \sigma^2)$ , and for given  $(\phi, \xi)$ .
  - (b) Compute the MLE  $\hat{\beta}_{(\phi, \xi)}$  of  $\beta$  as a function of  $(\phi, \xi)$
  - (c) Compute the MLE  $\hat{\sigma}_{(\phi, \xi)}^2$  of  $\sigma^2$  as a function of  $(\phi, \xi)$ .
  - (d) Compute the unbiased estimator of  $\tilde{\sigma}^2$  of  $\sigma^2$ .

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

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

<sup>1</sup>that is, the gradient of the log-likelihood

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

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

by replacing the  $\beta$  with  $\hat{\beta}_{(\phi, \xi)}$  and  $\sigma^2$  with  $\hat{\sigma}_{(\hat{\beta}_{(\phi, \xi)}, \phi, \xi)}^2$  in the actual likelihood  $L (Z; \beta, \theta = (\sigma^2, \phi, \xi))$ .

Describe how you would compute suitable values  $(\hat{\phi}, \hat{\xi})$  for the MLE of  $(\phi, \xi)$

**Solution.** It is

(1) It is

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

(2) It is

(a)

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

(b)

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

(3) It is

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

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

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

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

(b) It is

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

(c) It is

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

and by substituting I get

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

(d) It is

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

So

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

where

$$E[e] = 0$$

then

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

So it is

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

because

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

(e) It is

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

so it is Normal as a linear combination of normal random variables, with mean

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

and Variance

$$\begin{aligned}\text{Var}(\hat{\beta}_{(\phi,\xi)}) &= \left( \Psi^\top (R_{(\phi,\xi)})^{-1} \Psi \right)^{-1} \Psi^\top (R_{(\phi,\xi)})^{-1} \underbrace{\text{Var}(Z)}_{= \sigma^2 R_{(\phi,\xi)}} \\ &= \left( R_{(\phi,\xi)} \right)^{-1} \Psi \left( \Psi^\top (R_{(\phi,\xi)})^{-1} \Psi \right)^{-1} = \left( \Psi^\top (R_{(\phi,\xi)})^{-1} \Psi \right)^{-1}\end{aligned}$$

(4) It is

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

where obviously

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

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

**Exercise 2.** (★) Let  $(Z_s)_{s \in \mathcal{S}}$  be a specified statistical model. Assume that process  $(Z_s)_{s \in \mathcal{S}}$  has known mean  $\mu(s) = E(Z(s))$  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  $Z_{\text{SK}}(s_0)$  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  $Z = (Z_1, \dots, Z_n)^\top$ . Let  $w = (w_1, \dots, w_n)^\top$ .

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

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

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  $\sigma_{\text{SK}} = \sqrt{E(Z_{\text{SK}}(s_0) - Z(s_0))^2}$ .

**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

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

which is satisfied given the assumption

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

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

(2) It is

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

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

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

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

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

$$\begin{aligned} 0 &= \nabla_w E(Z_{\text{SK}}(s_0) - Z(s_0))^2 = \nabla_w \text{Var}(Z_{\text{SK}}(s_0) - Z(s_0)) \\ &= 2Cw - 2C_0 \end{aligned}$$

So I get

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

So

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

(4) It is

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