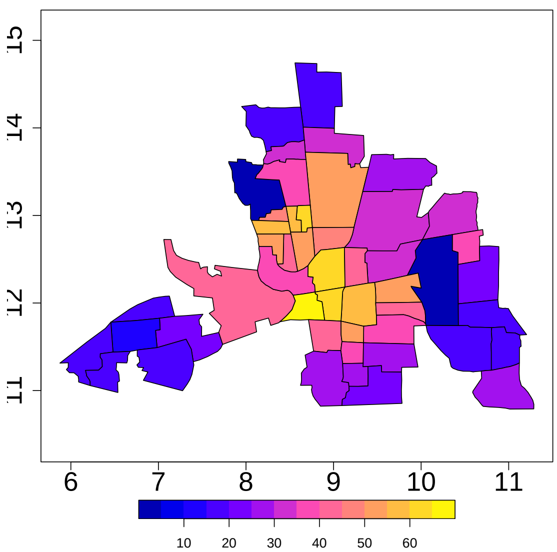
Spatio-temporal statistics (MATH4341) Michaelmas term, 2023

Exercise sheet

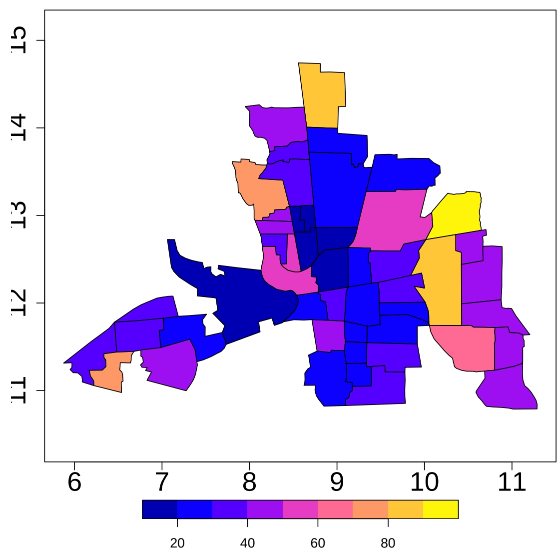
Lecturer: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

Part 1. Types of spatial data

Exercise 1. ($\star$)(Columbus Columbus OH data set) Figure 2a shows the Property crime (number per thousand households) in 49 districts in Columbus in 1980, as well as the average value of the house in USD. Figure 2b presents the corresponding average house value. This is the R dataset `columbus{spdep}`. Interest may lie to find whether high rates of crime are clusterd in a particular areas, and if yes, perhaps what is the association of it with the value of the houses in the area. To which principal spatial statistical are would you associate this problem?



(A) Property crime (number per thousand households)

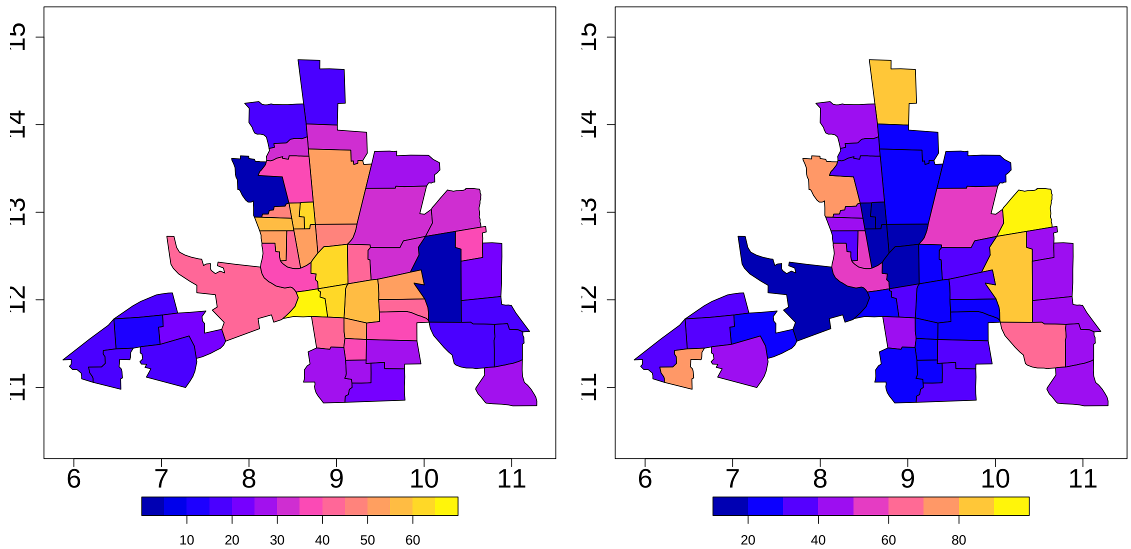


(B) Average value of the house in USD

FIGURE 1. Columbus Columbus OH spatial analysis dataset

Solution. Aerial unit data / spatial data on lattices

Exercise 2. ($\star$)(Columbus Columbus OH data set) Figure 2a shows the Property crime (number per thousand households) in 49 districts in Columbus in 1980, as well as the average value of the house in USD. Figure 2b presents the corresponding average house value. This is the R dataset columbus{spdep}. Interest may lie to find whether high rates of crime are clusterd in a particular areas, and if yes, perhaps what is the association of it with the value of the houses in the area. To which principal spatial statistical are would you associate this problem?



(A) Property crime (number per thousand households)

(B) Average value of the house in USD

FIGURE 2. Columbus Columbus OH spatial analysis dataset

Solution. Aerial unit data / spatial data on lattices

Exercise 3. (⋆)(Soil chemistry properties data set.) It contains measurements of various chemical properties of soil samples collected at different locations in a field. These properties include: the acidity or alkalinity of the soil (PH), the salt concentration in the soil (Salinity), and others. It is the R dataset soil250{geoR}. Figure 3 presents the locations these measurements are taken. The data (measurements) are in fixed locations at a regular grid of points. The domain scientist would be interested in the nutrient levels and pH to assess soil fertility and make recommendations for agricultural practices. The statistician could (i.) estimate/predict values of soil properties at unsampled locations based on measurements at sampled locations; and (ii.) assess the spatial variability of soil properties (nutrient levels and pH) to identify regions with high or low variability. To which principal spatial statistical are would you associate this problem?

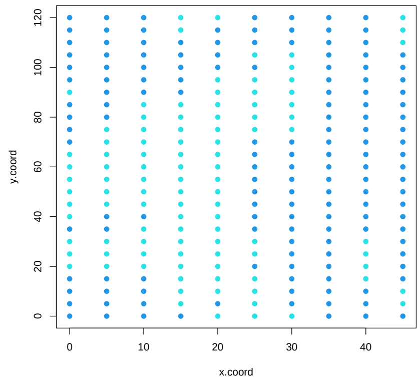


FIGURE 3. Soil chemistry data set

Solution. Point referenced data, or geostatistcal data

Exercise 4. ($\star$)(Scallop abundance data) Figure 4 presents 148 locations (degrees of longitude & latitude) in the Atlantic waters off the coasts of New Jersey and Long Island New York as coordinates and the size of scallop catch at the corresponding location as the dot size. The sites are at fixed locations within an irregular grid of points. Sustainable scallop abundance is critical for the long-term economic viability of the fishing industry. A healthy and stable scallop population supports a consistent source of income for fishermen and related businesses. To which principal spatial statistical are would you associate this problem?

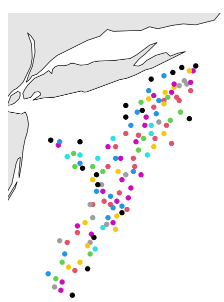


FIGURE 4. Scallop abundance data

Solution. Point referenced data, or geostatistcal data

Exercise 5. ($\star$)(Wolfcamp-aquifer data) Figure 5 presents locations and levels (in feet above sea level) of piezometric head for the aquifer; they are obtained by drilling a narrow pipe into the aquifer and letting the water find its own level in the pipe. After rigorous screening of unsuitable wells, 85 remained. There is interest to find where the radionuclide contamination would flow from the site in Deaf Smith County, Texas. Beneath Deaf Smith County is a deep brine aquifer known as the Wolfcamp aquifer, a potential pathway for any radionuclides leaking from the repository. The predicted direction of flow can be used to determine locations of downgradient and upgradient wells for a groundwater monitoring system. A first direction in analyzing this spatial data set is to draw a map of a predicted surface based on the (irregularly located) 85 data. To which principal spatial statistical are would you associate this problem?

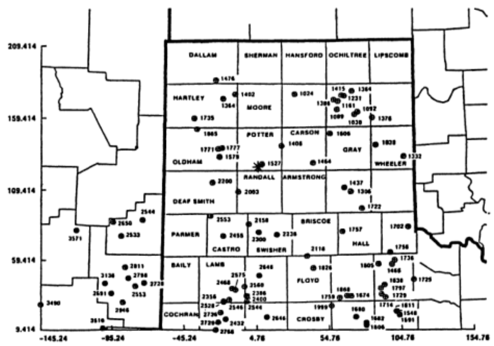


FIGURE 5. Wolfcamp-aquifer data. Piezometric-head levels (feet above sea level) vs coordinates.

Solution. Point referenced data, or geostatistcal data

Exercise 6. (⋆)(Swiss rainfall data) Figure 6 presents the locations of the 100 locations in Switzerland as dots whose size and color indicates the amount of the corresponding rainfall measurements (in 10th of mm) taken on May 8, 1986. This is the R data set `SIC{geoR}`. Observation sites are irregularly spaced, and fixed. A scientific objective may be to analyzing rainfall patterns with purpose to optimize crop planting and irrigation schedules. A statistician is able to estimate rainfall values at unsampled locations based on available measurements, create maps that represent the spatial distribution of rainfall, or quantify the uncertainty associated with rainfall estimates and predictions, which are important for risk assessment and decision-making. To which principal spatial statistical are would you associate this problem?

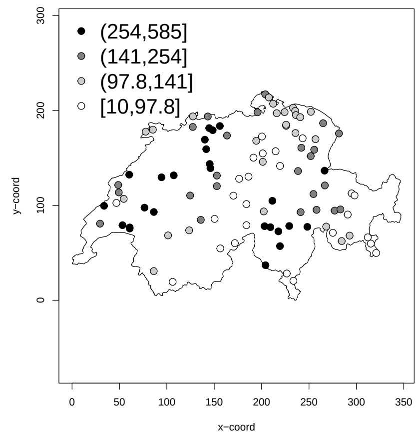


FIGURE 6. Swiss rainfall data

Solution. Point referenced data, or geostatistcal data

Part 2. INLA

Exercise 7. ($\star$)Consider the model

$$\begin{cases} z_i|\eta_i \sim \text{Poisson}\left(\exp\left(\eta_i\right)\right) \quad i = 1, ..., n \\ \eta_i = \beta_0 + \beta_1 w_i + u_{j(i)} \\ u \sim \text{N}_m\left(0, I\tau^{-1}\right) \end{cases}$$

where $\{w_i\}$ are covariates, $j(i)$ is a known mapping from $1:n$ to $1:m$ (given below in the dataset as idx).

For training use the following data set $\{(z_i, w_i)\}_{i=1}^{n}$ by running

```
rm(list=ls())
# generate the dataset
set.seed(123456L)
n = 50;
m = 10
w = rnorm(n, sd = 1/3)
u = rnorm(m, sd = 1/4)
intercept = 0;
beta = 1
idx = sample(1:m, n, replace = TRUE)
z = rpois(n, lambda = exp(intercept + beta * w + u[idx]))
table(z, dnn=NULL)
```

Do the following, by using R-INLA

(1) Run inla{INLA} in order to train the above model, and generate an inla object (that you will call it out.inla). For the function inla{INLA} specify the formula, data, and family

arguments. To approximate the conditional pdf of latent variables of the GMRF use the Gaussian approximation. For the rest parameters just use the default R-INLA options.

(2) Print a summary of the marginal posteriors

(3) Produce and print the marginal posterior pdf of $\Pr(\beta_1|z)$.

Solution.

(1)
```
my.data = data.frame(z, w, idx)
formula = z ~ 1 + w + f(idx, model="iid")
out.inla = inla(formula, data = my.data,
          family = "poisson",
          control.inla = list(strategy = "gaussian")
          )
```

(2)
```
summary(out.inla)
```

```
> res.predict$summary.linear.predictor[7,]
                 mean        sd 0.025quant 0.5quant 0.975quant     mode          kld
Predictor.07 3.021652 0.1847223   2.639797 3.029161   3.362469 3.045581 1.224947e-07
```

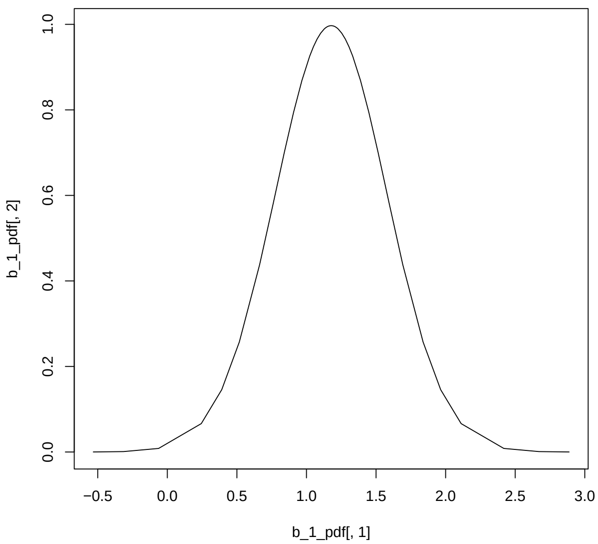
```
Call:
   c("inla.core(formula = formula, family = family, contrasts = contrasts, ", " data =
   data, quantiles = quantiles, E = E, offset = offset, ", " scale = scale, weights =
   weights, Ntrials = Ntrials, strata = strata, ", " lp.scale = lp.scale, link.covariates
   = link.covariates, verbose = verbose, ", " lincomb = lincomb, selection = selection,
   control.compute = control.compute, ", " control.predictor = control.predictor,
   control.family = control.family, ", " control.inla = control.inla, control.fixed =
   control.fixed, ", " control.mode = control.mode, control.expert = control.expert, ", "
   control.hazard = control.hazard, control.lincomb = control.lincomb, ", " control.update
   = control.update, control.lp.scale = control.lp.scale, ", " control.pardiso =
   control.pardiso, only.hyperparam = only.hyperparam, ", " inla.call = inla.call,
   inla.arg = inla.arg, num.threads = num.threads, ", " blas.num.threads =
   blas.num.threads, keep = keep, working.directory = working.directory, ", " silent =
   silent, inla.mode = inla.mode, safe = FALSE, debug = debug, ", " .parent.frame =
   .parent.frame)")
Time used:
    Pre = 0.763, Running = 0.224, Post = 0.0168, Total = 1
Fixed effects:
              mean    sd 0.025quant 0.5quant 0.975quant   mode kld
(Intercept) -0.069 0.153     -0.370   -0.069      0.232 -0.069   0
w            1.178 0.401      0.391    1.178      1.964  1.178   0

Random effects:
  Name    Model
    idx IID model

Model hyperparameters:
                      mean       sd 0.025quant 0.5quant 0.975quant   mode
Precision for idx 19980.67 19912.46     599.82 13939.51   74289.61 214.74

Marginal log-Likelihood:  -69.62
 is computed
Posterior summaries for the linear predictor and the fitted values are computed
(Posterior marginals needs also 'control.compute=list(return.marginals.predictor=TRUE)')
```

(3) b_1_pdf = out.inla$marginals.fixed$w
plot(b_1_pdf[,1], b_1_pdf[,2], type ="l")



Part 3. **Point referenced data / Geostatistics**

Exercise 8. ($\star$)If $c:\mathbb{R}^d \to \mathbb{R}$ is the covariogram of a weakly stationary random field $Z=(Z_s)_{s\in\mathbb{R}^d}$ then $c(\cdot)$ is semi-positive definite; i.e. for all $n\in\mathbb{N}$, $a\in\mathbb{R}^n$, and $\{s_1,...,s_n\}\subseteq S$

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

Solution. To show that $c(\cdot)$ is semi-positive definite, I need to show that $\forall a\in\mathbb{R}^n - \{0\}$ it is

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

Well it is

$$\begin{aligned} 0 \leq \mathrm{Var}\left(\sum_{i=1}^{n} a_i Z(s_i)\right) &= \mathrm{Cov}\left(\sum_{i=1}^{n} a_i Z(s_i), \sum_{j=1}^{n} a_j Z(s_j)\right) \\ &= \sum_{i=1}^{n} a_i \sum_{j=1}^{n} a_j \mathrm{Cov}(Z(s_i), Z(s_j)) \\ &= \sum_{i=1}^{n} a_i \sum_{j=1}^{n} a_j a_j c(s_i, s_j) = \sum_{i=1}^{n} a_i \sum_{j=1}^{n} a_j c(s_i - s_j) \end{aligned}$$

Exercise 9. ($\star$) Show that if $c_1(\cdot,\cdot)$ and $c_2(\cdot,\cdot)$ are are covariance functions (are non-negative definite) then so are $c_3(\cdot,\cdot) = bc_1(\cdot,\cdot) + dc_2(\cdot,\cdot)$ and $c_4(\cdot,\cdot) = c_1(\cdot,\cdot)\,c_2(\cdot,\cdot)$.

Solution. For all $n \in \mathbb{N}$ and $a_1, ...a_n$

$$\begin{aligned}
\sum_{i=1}^{n} a_i \sum_{j=1}^{n} a_j c_3 (s_i, s_j) &= \sum_{i=1}^{n} a_i \sum_{j=1}^{n} a_j (b c_1 (s_i, s_j) + d c_2 (s_i, s_j)) \\
&= \underbrace{\sum_{i=1}^{n} a_i \sum_{j=1}^{n} a_j b c_1 (s_i, s_j)}_{\geq 0} + \underbrace{\sum_{i=1}^{n} a_i \sum_{j=1}^{n} a_j d c_2 (s_i, s_j)}_{\geq 0} \\
&\geq 0
\end{aligned}$$

Regarding c_4, assume independent stochastic processes $(Y_s)_{s \in S}$ and $(X_s)_{s \in S}$ with mean zero and covariance functions $c_1 (\cdot, \cdot)$ and $c_2 (\cdot, \cdot)$ correspondingly. Let stochastic processes $(Z_s)_{s \in S}$ with $Z_s = Y_s X_s$. Then

$$\begin{aligned}
\operatorname{Cov}(Z_s, Z_t) &= \operatorname{Cov}(Y_s X_s, Y_t X_t) \\
&= \operatorname{E}(Y_s X_s Y_t X_t) \\
&= \operatorname{E}(Y_s Y_t X_s X_t), \ \text{ but } \ Y_s \perp X_s \\
&= \operatorname{E}(X_s X_t) \operatorname{E}(X_s X_t) \\
&= \operatorname{Cov}(X_s, X_t) \operatorname{Cov}(Y_s, Y_t) \\
&= c_1 (s, t) c_2 (s, t) = c_4 (s, t)
\end{aligned}$$

that is $c_4 (\cdot, \cdot)$ is a covariance function of a stochastic processes.

Exercise 10. ($\star$) Consider the Gaussian c.f. $c(h) = \sigma^2 \exp\left(-\beta \|h\|_2^2\right)$ for $\sigma^2, \beta > 0$ and $h \in \mathbb{R}^d$. Compute the spectral density from Bochner's theorem

Solution. It is

$$\begin{aligned}
f(\omega) &= \left(\frac{1}{2\pi}\right)^d \int_{\mathbb{R}^d} \exp\left(-i\omega^\top h\right) \sigma^2 \exp\left(-\beta \|h\|_2^2\right) \mathrm{d}h \\
&= \sigma^2 \left(\frac{1}{2\pi}\right)^d \prod_{j=1}^{d} \int_{\mathbb{R}} \exp\left(-i\omega_j h_j - \beta h_j^2\right) \mathrm{d}h \\
&= \sigma^2 \left(\frac{1}{2\pi}\right)^d \prod_{j=1}^{d} \int_{\mathbb{R}} \exp\left(-\beta \left(h_j - (-i\omega / (2\beta))\right)^2\right) \mathrm{d}h_j \\
&= \sigma^2 \left(\frac{1}{4\pi\beta}\right)^{d/2} \exp\left(-\|\omega\|_2^2 / (4\beta)\right)
\end{aligned}$$

Exercise 11. ($\star$) Consider the Exponential c.f. $c(h) = \sigma^2 \exp\left(-\beta \|h\|_1^1\right)$ for $\sigma^2, \beta > 0$ and $h \in \mathbb{R}^d$. Compute the spectral density from Bochner's theorem

Solution. It is

$$f(\omega) = \left(\frac{1}{2\pi}\right)^d \int_{\mathbb{R}^d} \exp\left(-i\omega^\top h\right) \sigma^2 \exp\left(-\beta \|h\|_1^1\right) \mathrm{d}h$$
$$= \sigma^2 \left(\frac{1}{2\pi}\right)^d \prod_{j=1}^d \int_{\mathbb{R}} \exp\left(-i\omega_j h_j - \beta |h_j|\right) \mathrm{d}h_j$$

where

$$\begin{aligned}
\int_{\mathbb{R}} \exp\left(-i\omega_j h_j - \beta |h_j|\right) &= \int_{-\infty}^0 \exp\left(-i\omega_j h_j - \beta |h_j|\right) \mathrm{d}h_j + \int_0^\infty \exp\left(-i\omega_j h_j - \beta |h_j|\right) \mathrm{d}h_j \\
&= \int_{-\infty}^0 \exp\left(-i\omega_j h_j + \beta h_j\right) \mathrm{d}h_j + \int_0^\infty \exp\left(-i\omega_j h_j - \beta h_j\right) \mathrm{d}h_j \\
&= \int_{-\infty}^0 \exp\left(-\left(i\omega_j - \beta\right) h_j\right) \mathrm{d}h_j + \int_0^\infty \exp\left(-\left(i\omega_j + \beta\right) h_j\right) \mathrm{d}h_j \\
&= \int_0^\infty \exp\left(-\left(\beta - i\omega_j\right) h_j\right) \mathrm{d}h_j + \int_0^\infty \exp\left(-\left(i\omega_j + \beta\right) h_j\right) \mathrm{d}h_j \\
&= \frac{1}{(\beta - i\omega_j)} + \frac{1}{(\beta + i\omega_j)} = \frac{2\beta}{\beta^2 + \omega_j^2}
\end{aligned}$$

hence

$$f(\omega) = \sigma^2 \left(\frac{\beta}{\pi}\right)^d \prod_{j=1}^d \frac{1}{\beta^2 + \omega_j^2}$$

(Given as Formative assessment 1)

Exercise 12. ($\star$) 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

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

where $c_Y(s,t) = \mathrm{E}\left(Y(s) Y(t)\right)$, 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

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

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

(1) It is

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

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

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

(3) It is

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

hence

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

(Given as Formative assessment 1)

Exercise 13. ($\star$) Consider the zero-mean geostatistical process $Z = (Z_s)_{s\in\mathbb{R}^d}$ with a weakly stationary and isotropic covariance function given by

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

(1) Compute the semi-variogram for the geostatistical process (Z_s)
(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.

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

Solution.

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

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

then

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

(2)

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

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

  then

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

- 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.

Exercise 14. ($\star$) Consider we the geostatistical model $(Z_s)_{s\in\mathcal{S}}$ with

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

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

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

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.

(1) Write down 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 $[D]_{i,j} = \|s_i - s_j\|$. Write down the covariance matrix $C(\theta)$ of $Z = (Z_1, ..., Z_n)^\top$ as a function of D and θ.

Hint:: You can use convenient notation such as $\exp(D)$ meaning $[\exp(D)]_{i,j} = \exp(D_{i,j})$.

(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 $c(\cdot) = \sigma^2 r(\cdot)$. For the following, 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

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

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

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 for the MLE $\left(\hat{\phi}, \hat{\xi}\right)$ of (ϕ, ξ)

Solution. It is

(1) It is

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

(2) It is

(a)

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

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

(b)

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

(3) It is

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

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

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

(b) It is

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

(c) It is

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

and by subtituting I get

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

(d) It is

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

So

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

where

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

then

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

So it is

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

because

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

(e) It is

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

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

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

and Variance

$$
\begin{aligned}
\mathrm{Var}\left(\hat{\beta}\left(\phi,\xi\right)\right)=&\left(\Psi^{\top}\left(R\left(\theta\right)\right)^{-1}\Psi\right)^{-1}\Psi^{\top}\left(R\left(\theta\right)\right)^{-1}\cancelto{=\sigma^{2}R\left(\theta\right)}{\mathrm{Var}\left(Z\right)}\\
&\left(R\left(\theta\right)\right)^{-1}\Psi\left(\Psi^{\top}\left(R\left(\theta\right)\right)^{-1}\Psi\right)^{-1}=\left(\Psi^{\top}\left(R\left(\theta\right)\right)^{-1}\Psi\right)^{-1}
\end{aligned}
$$

(4) It is

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

where obviously

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

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

Exercise 15. $(\star)$ Let $\left(Z_s\right)_{s\in\mathcal{S}}$ be a specified statistical model. Assume that $\left(Z_s\right)_{s\in\mathcal{S}}$ is weakly stationary with unknown constant mean $\mu=\mathrm{E}\left(Z\left(s\right)\right)$ and known covariogram $c\left(\cdot\right)$. Assume there is available a dataset $\left\{\left(s_i,Z_i:=Z\left(s_i\right)\right)\right\}_{i=1}^{n}$ and assume they are realizations of $\left(Z_s\right)_{s\in\mathcal{S}}$. Assume that the matrix C such as $\left[C\right]_{i,j}=c\left(\left\|s_i-s_j\right\|\right)$ has an inverse. Consider the "Kriging" estimator

μ_{KM} of μ as the BLUE (Best Linear Unbiased Estimator)

$$\mu_{\mathrm{KM}} = \sum_{i=1}^{n} w_i Z(s_i) = w^\top Z,$$

for some unknown $\{w_i\}$ that we need to learn.

(1) Find sufficient conditions on $w = (w_1, ..., w_n)$ so that the Kriging estimator μ_{KM} to be unbiased.

(2) Assume C is invertable. Compute the MSE of μ_{KM} as a function of $w = (w_1, ..., w_n)$ and C

(3) Derive the Kriging estimator μ_{KM} of μ as a function of C

(4) Derive the Kriging standard error as $\sigma_{\mathrm{KM}} = \sqrt{\mathrm{E}(\mu_{\mathrm{KM}} - \mu)^2}$ as a function of C

Solution 16. The method is called Kriging the Mean, and hence we denote it as KM.

(1) It is

$$\mu_{\mathrm{KM}} = \sum_{i=1}^{n} w_i Z(s_i) = w^\top Z,$$

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

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

$$\mathrm{E}(\mu_{\mathrm{KM}} - \mu) = \mathrm{E}\left(\sum_{i=1}^{n} w_i Z(s_i) - \mu\right) = \cancelto{=1}{\sum_{i=1}^{n} w_i}\cancelto{=\mu}{\mathrm{E}(Z(s_i))} - \mu$$

which is satisfied given the assumption

$$\sum_{i=1}^{n} w_i = 1 \Longleftrightarrow 1^\top w = 1 \quad \text{(ASSUMPTION)}$$

(2) It is

$$\begin{aligned}
\mathrm{E}(\mu_{\mathrm{KM}} - \mu)^2 &= \mathrm{E}\left(\mu_{\mathrm{KM}}^2 + \mu^2 - 2\mu_{\mathrm{KM}}\mu\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \mathrm{E}(Z(s_i) Z(s_j)) - \cancelto{=1}{\sum_{i=1}^{n} w_i}\cancelto{=1}{\sum_{j=1}^{n} w_j}\mu \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j (c(s_i - s_j) - \mu) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j c(s_i - s_j) = w^\top C w
\end{aligned}$$

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

$$w^{\mathrm{KM}} = \arg\min_{w} \mathrm{E}(\mu_{\mathrm{KM}} - \mu)^2, \text{ subject to } \sum_{i=1}^{n} w_i = 1$$

The Lagrange function is

$$\begin{aligned}
\mathfrak{L}(w, \lambda) &= \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j c(s_i - s_j) - 2\lambda\left(\sum_{i=1}^{n} w_i - 1\right) \\
&= w^\top C w - 2\lambda\left(1^\top w - 1\right)
\end{aligned}$$

The Kriging to mean equations are $0 = \nabla_{w,\lambda}\mathfrak{L}(w,\lambda)$ producing

$$\begin{cases} 0 = 2\sum_{j=1}^{n} w_j^{\mathrm{KM}} c(s_i - s_j) - 2\lambda \quad \forall i = 1, ..., n \\ 1 = \sum_{i=1}^{n} w_i^{\mathrm{KM}} \end{cases}$$

$$\begin{cases} 2Cw^{\mathrm{KM}} - 2\lambda 1 = 0 \\ 1^\top w^{\mathrm{KM}} = 1 \end{cases}$$

Given that C^{-1} exists, I multiply by $1^\top C^{-1}$ and I get

$$21^\top \overset{=I}{\cancel{C^{-1}}C} w - 21^\top C^{-1}\lambda 1 = 0$$

so

$$\lambda = \frac{1}{1^\top C^{-1} 1}$$

I substitute and I get

$$w^{\mathrm{KM}} = \frac{C^{-1}1}{1^\top C^{-1} 1}$$

So

$$\mu_{\mathrm{KM}} = \left(\frac{C^{-1}1}{1^\top C^{-1} 1}\right)^\top Z$$

(4) It is

$$\sigma_{\mathrm{KM}} = \sqrt{\mathrm{E}(\mu_{\mathrm{KM}} - \mu)^2} = \sqrt{\left(\frac{C^{-1}1}{1^\top C^{-1} 1}\right)^\top C \frac{C^{-1}1}{1^\top C^{-1} 1}} = \frac{1}{\sqrt{1^\top C^{-1} 1}}$$

Exercise 17. ($\star$) 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 $Z_{\mathrm{SK}}(s_0)$ of $Z(s_0)$ at an unseen spatial location s_0 as the BLUE (Best Linear Unbiased Estimator)

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

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 $Z_{\mathrm{SK}}(s_0)$ to be unbiased.

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

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

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

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

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

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

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

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

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

$$\begin{aligned}
\mathrm{E}\left(Z_{\text{SK}}(s_0) - Z(s_0)\right)^2 =& \mathrm{Var}\left(Z_{\text{SK}}(s_0) - Z(s_0)\right) = \mathrm{Var}\left(w_{n+1} + w^\top Z - Z(s_0)\right) \\
=& \mathrm{Var}\left(w_{n+1} + w^\top Z\right) + \mathrm{Var}\left(Z(s_0)\right) - 2\mathrm{Cov}\left(w_{n+1} + w^\top Z, Z(s_0)\right) && (1) \\
=& w^\top C w + c(s_0, s_0) - 2 w^\top \mathrm{Cov}\left(Z, Z(s_0)\right) && (2) \\
=& w^\top C w + c(s_0, s_0) - 2 w^\top C_0 && (3)
\end{aligned}$$

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

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

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

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

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

So I get

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

So

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

(3) It is

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

Exercise 18. ($\star$) Assume a spatial model

$$Z\left(s\right) = \mu + \delta\left(s\right),\ \ s \in \mathcal{S} \tag{4}$$

with unknown mean $\mu \in \mathbb{R}$. Assume a set of n observed realizations $Z_i := Z\left(s_i\right)$ of (4) at sites s_i for $i = 1, ..., n$. Assume that $Z\left(s\right)$ is a weak stationary stochastic process with known covariogram $c\left(\cdot\right)$. Derive the formula for the Ordinary Kriging predictor $Z_0 := Z\left(s_0\right)$ at spatial location s_0 and its kriging variance as function of the covariogram $c\left(h\right)$ and not the semi-variogram.

Exercise 19. ($\star$) Let $\left(Z_s\right)_{s \in \mathcal{S}}$ be a specified statistical model. Assume that $\left(Z_s\right)_{s \in \mathcal{S}}$ is an intrinsic stationary process with unknown constant mean $\mu\left(s\right) = \mathrm{E}\left(Z\left(s\right)\right)$ and known semi-variogram $\gamma\left(\cdot\right)$. Assume there is available a dataset $\{\left(s_i, Z_i := Z\left(s_i\right)\right)\}_{i=1}^n$. Consider the "Kriging" estimator $Z_{\mathrm{OK}}\left(s_0\right)$ of $Z\left(s_0\right)$ at any unseen spatial location s_0 as the BLUE (Best Linear Unbiased Estimator)

$$Z_{\mathrm{OK}}\left(s_0\right) = w_{n+1} + \sum_{i=1}^n w_i Z\left(s_i\right) = w_{n+1} + w^\top Z$$

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

(1) Find sufficient conditions on $w = \left(w_1, ..., w_n\right)$ so that the Kriging estimator $Z_{\mathrm{OK}}\left(s_0\right)$ to be unbiased.

(2) Derive the MSE of $Z_{\mathrm{OK}}\left(s_0\right)$ as

$$\mathrm{E}\left(Z_{\mathrm{OK}}\left(s_0\right) - Z\left(s_0\right)\right)^2 = -w^\top \mathbf{\Gamma} w + 2 w^\top \boldsymbol{\gamma}_0$$

where $\boldsymbol{\gamma}_0 = \left(\gamma\left(s_0 - s_i\right), ..., \gamma\left(s_0 - s_n\right)\right)^\top$ and Γ with $\left[\mathbf{\Gamma}\right]_{i,j} = \gamma\left(s_i - s_j\right)$

(3) Assume Γ is invertable matrix. Derive the Kriging estimator of $Z\left(s_0\right)$ as

$$Z_{\mathrm{OK}}\left(s_0\right) = \mathbf{\Gamma}^{-1}\left(\boldsymbol{\gamma}_0 + \frac{1 - 1^\top \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0}{1^\top \mathbf{\Gamma}^{-1} 1} 1\right) Z$$

(4) Derive the Kriging standard error of $Z_{\mathrm{OK}}\left(s_0\right)$ as

$$\sigma_{\mathrm{SK}} = \sqrt{\boldsymbol{\gamma}_0 \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 - \frac{\left(1 - 1^\top \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0\right)^2}{1^\top \mathbf{\Gamma}^{-1} 1}}$$

.

Solution. The method is called Ordinary Kriging, and hence we denote it as OK.

(1) It is

$$Z_{\mathrm{OK}}\left(s_0\right)=w_{n+1}+\sum_{i=1}^{n} w_i Z\left(s_i\right)=w_{n+1}+w^{\top} Z,$$

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

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

Unbiasness is satisfied given the assumption $w_{n+1}=0$, and

$$\sum_{i=1}^{n} w_i=1 \Longleftrightarrow 1^{\top} w=1 \quad \text{(ASSUMPTION)}$$

(2) The MSE of $Z_{\mathrm{OK}}\left(s_0\right)$ is

$$\begin{aligned}
\operatorname{MSE}\left(Z_{\mathrm{OK}}\left(s_0\right)\right) = & \mathrm{E}\left(Z_{\mathrm{OK}}\left(s_0\right)-Z\left(s_0\right)\right)^2=\mathrm{E}\left(\sum_{i=1}^{n} w_i Z\left(s_i\right)-\underbrace{\sum_{i=1}^{n} w_i}_{=1} Z\left(s_0\right)\right)^2 \\
= & \mathrm{E}\left(\sum_{i=1}^{n} w_i\left(Z\left(s_i\right)-Z\left(s_0\right)\right)\right)^2 \\
= & -\mathrm{E}\left(\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j\left(Z\left(s_i\right)-Z\left(s_j\right)\right)^2-2 \sum_{i=1}^{n} w_i\left(Z\left(s_i\right)-Z\left(s_0\right)\right)^2\right) \\
= & -\sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j \frac{1}{2} \mathrm{E}\left(Z\left(s_i\right)-Z\left(s_j\right)^2\right)+2 \sum_{i=1}^{n} w_i \frac{1}{2} \mathrm{E}\left(Z\left(s_i\right)-Z\left(s_0\right)^2\right) \\
= & -\sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j \gamma\left(s_i-s_j\right)+2 \sum_{i=1}^{n} w_i \gamma\left(s_i-s_0\right) \\
= & -w^{\top} \boldsymbol{\Gamma} w+2 w^{\top} \boldsymbol{\gamma}_0
\end{aligned}$$

where $w=\left(w_1, \ldots, w_n\right)^{\top}$, $\boldsymbol{\gamma}_0=\left(\gamma\left(s_0-s_i\right), \ldots, \gamma\left(s_0-s_n\right)\right)^{\top}$, and $[\boldsymbol{\Gamma}]_{i, j}=\gamma\left(s_i-s_j\right)$.

(3) The Lagrange multiplier function to minimize the MSE under the assumption is

$$\begin{aligned}
\mathfrak{L}(w, \lambda) & =-\sum_{i=1}^{n} w_i w_j \gamma\left(s_i-s_j\right)+2 \sum_{i=1}^{n} w_i \gamma\left(s_0-s_i\right)-\lambda\left(\sum_{i=1}^{n} w_i-1\right) \\
& =-w^{\top} \boldsymbol{\Gamma} w+2 w^{\top} \boldsymbol{\gamma}_0-\lambda\left(1^{\top} w-1\right)
\end{aligned}$$

The OK system of equations is $0 = \nabla_{(\{w_i\},\lambda)} L(w,\lambda)\big|_{(w,\lambda)}$ producing

$$\begin{cases} 0 = -2\sum_{j=1}^{n} w_j^{\mathrm{OK}} \gamma(s_i - s_j) + 2\gamma(s_0 - s_i) - \lambda, \quad i = 1, ..., n \\ 1 = \sum_{i=1}^{n} w_i^{\mathrm{OK}} \end{cases} \iff$$

$$\begin{cases} 0 = -2\boldsymbol{\Gamma} w_{\mathrm{OK}} + 2\boldsymbol{\gamma}_0 - \lambda_{\mathrm{OK}} 1 \\ 1 = 1^{\top} w_{\mathrm{OK}} \end{cases}$$

Assuming $\boldsymbol{\Gamma}$ is invertable and multiplying by $1^{\top}\boldsymbol{\Gamma}^{-1}$ it is

$$0 = -2\boldsymbol{\Gamma} w_{\mathrm{OK}} + 2\boldsymbol{\gamma}_0 - \lambda_{\mathrm{OK}} 1 \iff$$

$$0 = -2\cancelto{=1}{1^{\top}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma} w_{\mathrm{OK}}} + 2 1^{\top}\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0 - 1^{\top}\boldsymbol{\Gamma}^{-1}\lambda 1 \iff$$

$$\lambda_{\mathrm{OK}} = 2\frac{1^{\top}\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0 - 1}{1^{\top}\boldsymbol{\Gamma}^{-1}1}$$

By substitution I get

$$w_{\mathrm{OK}} = \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{\gamma}_0 + \frac{1 - 1^{\top}\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0}{1^{\top}\boldsymbol{\Gamma}^{-1}1} 1\right)$$

Hence

$$Z_{\mathrm{OK}}(s_0) = w_{\mathrm{OK}} Z = \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{\gamma}_0 + \frac{1 - 1^{\top}\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0}{1^{\top}\boldsymbol{\Gamma}^{-1}1} 1\right) Z$$

(4) It is

$$\begin{aligned} \sigma_{\mathrm{OK}}(s_0) &= \sqrt{\mathrm{MSE}(Z_{\mathrm{OK}}(s_0))} \\ &= \sqrt{-w^{\top}\boldsymbol{\Gamma} w + w^{\top}\boldsymbol{\gamma}_0} \\ &= \sqrt{\boldsymbol{\gamma}_0 \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0 - \frac{\left(1 - 1^{\top}\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0\right)^2}{1^{\top}\boldsymbol{\Gamma}^{-1}1}} \end{aligned}$$