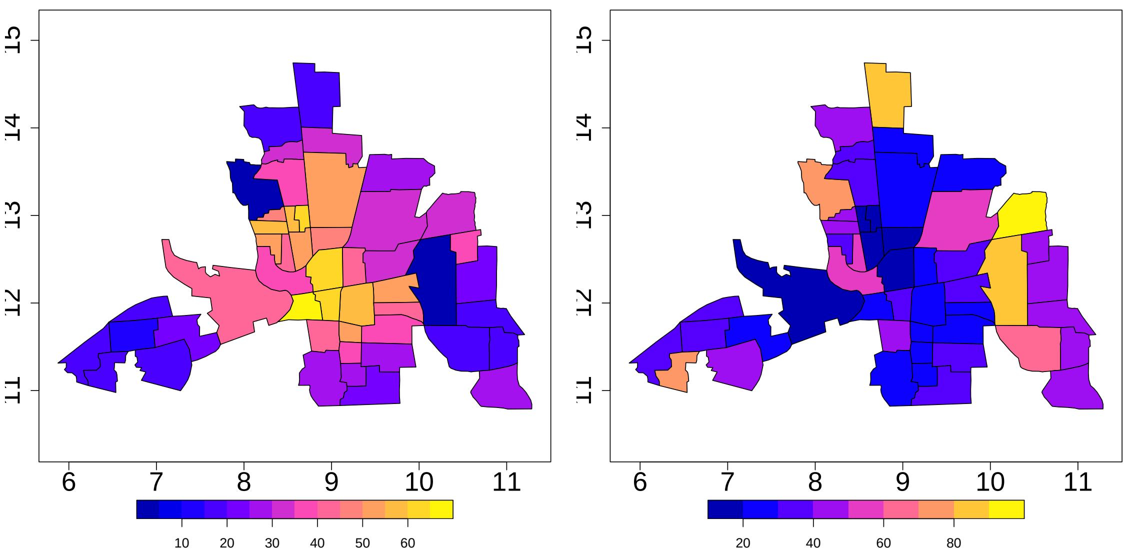
Spatio-temporal statistics (MATH4341) Michaelmas term, 2024

Exercise sheet

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

Part 1. Types of spatial data

($\star$)(Columbus Columbus OH data set) Figure 1a 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 1b 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 1. ($\star$)(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 2 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?

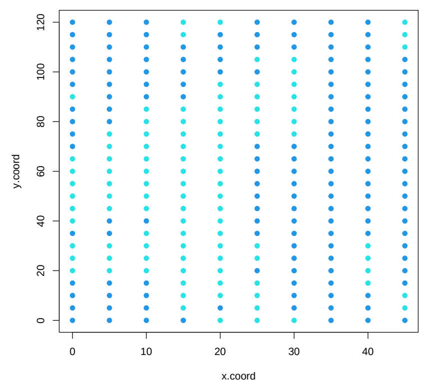


FIGURE 2. Soil chemistry data set

Solution. Point referenced data, or geostatistcal data

Exercise 2. ($\star$)(Wolfcamp-aquifer data) Figure 3 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?

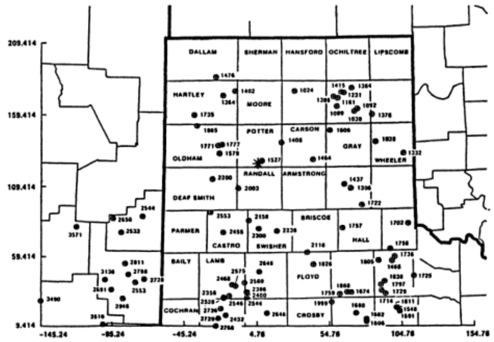


FIGURE 3. Wolfcamp-aquifer data. Piezometric-head levels (feet above sea level) vs coordinates.

Solution. Point referenced data, or geostatistcal data

Exercise 3. ($\star$)(Swiss rainfall data) Figure 4 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?

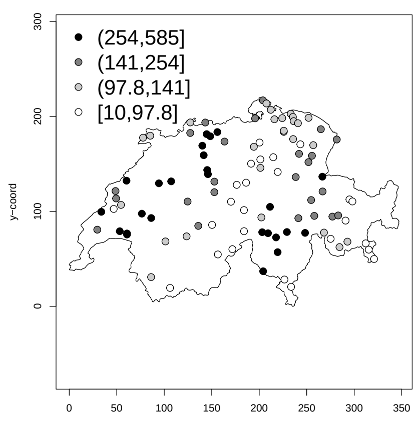


FIGURE 4. Swiss rainfall data

Solution. Point referenced data, or geostatistcal data

Exercise 4. ($\star$)(Air pollution in Piemonte.) Figure 5 presents the average PM10 ($\mu g/m^3$) concentration during October 2005–March 2006 for the 24 monitoring stations in the Piemonte region (Northern Italy). The data (measurements) are at fixed locations at irregular grid points. PM10 is one of the most troublesome pollutants in the area. Environmental agencies need models to predict PM10 at unmonitored sites in order to assess PM10 concentration over an entire region. A geostatistician can build a model which is satisfactory in terms of goodness of fit, interpretability, parsimony, prediction capability and computational costs with purpose to build reliable PM10 concentration maps, equipped with the corresponding uncertainty measure.To which principal spatial statistical are would you associate this problem?

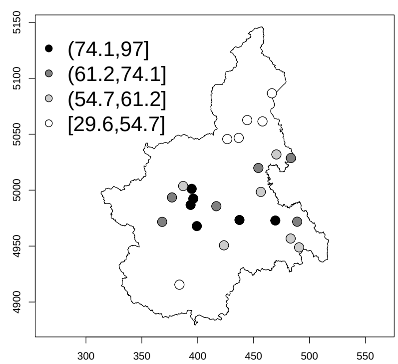


FIGURE 5. (Air pollution data) Average PM10 ($\mu g/m^3$) concentration during October 2005–March 2006 for the 24 monitoring stations in the Piemonte

Solution. Point reference data / geostatistical data

Exercise 5. ($\star$)(Scallop abundance data) The scallop data is based on a 1990 survey cruise in the Atlantic continental shelf off Long Island, New York, U.S.A. They are available from R as `scallop {SemiPar}`. Figure 6 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.

(1) To which principal spatial statistical are would you associate this problem?
(2) Can you suggest a Bayesian hierarchical model for this? Justify your suggestion.

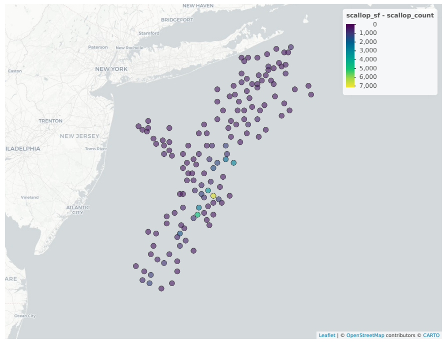


FIGURE 6. Scallop abundance data

Solution.

(1) Point referenced data, or geostatistcal data

(2) It contains $n = 148$ observations $\{(Z_i, s_i)\}_{i=1}^{n}$ where the i-th observation contains the observed scallop catch size Z_i at location specified by coordinate $s_i = (s_{1,i}, s_{2,i})^\top$ that is the degrees latitude north of the Equator $s_{2,i}$, and degrees longitude west of Greenwich $s_{1,i}$.

This is definitely a geostatistics problem. Here, we will present a naive way to model it by reflecting what we discussed earlier.

Data model: One may consider that the observations $\{Z_i\}$ at each location are a realization of the actual mean (hence unknown) scallop catch abundance $Y_i$and as it is a count its distribution can be represented as a Poisson distribution, i.e.

$$Z_i|Y_i, \sigma^2 \overset{\text{ind}}{\sim} \text{Poi}\left(\exp\left(Y_i\right)\right), \quad i = 1, ..., n$$

Spatial process model: One may consider that in log scale the mean scallop catch abundance $Y_i = \log\left(\text{E}\left(Z_i|Y_i, \sigma^2\right)\right)$ is a function where at each finite set of locations $\{s_i\}_{i=1}^{n}$ and that it follows a Normal distribution with a mean μ with $[\mu]_i = \mu\left(s_i\right)$ parametrized as

$$\mu\left(s\right) = \exp\left(\beta_0 + \beta_1 s_1 + \beta_2 s_2 + \beta_{12} s_1 s_2\right), \text{ at a location s=}\left(s_1, s_2\right)^\top$$

with unknown parameter β, and covariance matrix $[C]_{i,j} = c\left(s_i, s_j\right)$ parametrized with covariance function

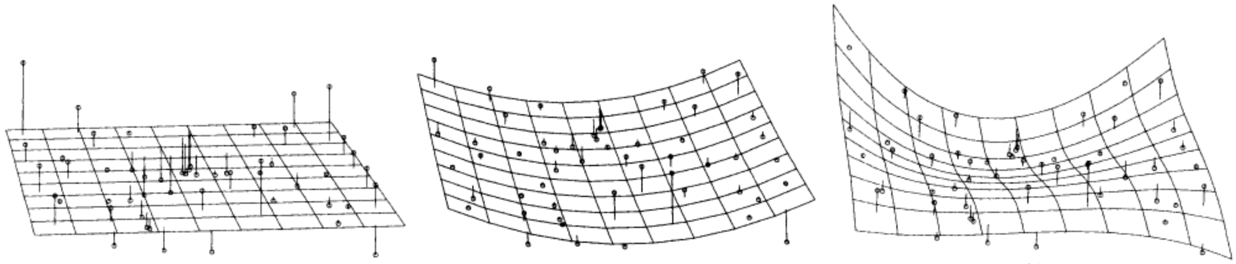
$$c\left(s, s'\right) = \sigma^2 \exp\left(-\phi \left\|s - s'\right\|\right)$$

to impose that nearer locations cause stronger dependences in the model. Here β, ϕ, and σ^2 are unknown parameters.

Hierarchical model: To sum up, we have build the hierarchical model

(1) $$\begin{cases} Z_i|Y_i, \sigma^2 \overset{\text{ind}}{\sim} & \text{Poi}\left(\exp\left(Y_i\right)\right), \text{ data model} \\ Y|\sigma^2, \beta, \phi \sim & \text{N}_n\left(S\beta, C\right), \quad \text{spatial process model} \end{cases}$$

Figure 7 shows the hierarchical spatial model (1) for different values of $\theta = \left(\sigma^2, \beta, \phi\right)$; the surface corresponds to the spatial process $\left\{Y\left(s\right); s \in \mathbb{R}^2\right\}$ and is presented at



(A) (B) (C)

FIGURE 7. Examples representing the hierarchical spatial model for different values of $\theta = \left(\sigma^2, \beta, \phi\right)$

three different instances each of them with different values for (β, ϕ), while the dots correspond to the observations $\{(Z(s_i), s_i)\}_{i=1}^{n}$ and their deviation from the spatial process is controlled by σ^2 .

Bayesian hierarchical model: If we work on the fully Bayesian framework, we can complete the model with priors on $\theta = \left(\sigma^2, \beta, \phi\right)$ for instance $\sigma^2 \sim \mathrm{IG}(\kappa_\sigma, \lambda_\sigma)$, $\phi \sim \mathrm{IG}(\kappa_\phi, \lambda_\phi)$, and $\beta \sim \mathrm{N}_4(b, Iv)$, with some known hyper-parameters $\kappa_\sigma, \lambda_\sigma, \kappa_\phi, \lambda_\phi, b, v$. To sum up, we have build the Bayesian model

$$\begin{cases} Z_i|Y_i, \sigma^2 \overset{\text{ind}}{\sim} & \mathrm{Poi}\left(\exp\left(Y_i\right)\right), \text{ data model} \\ Y|\sigma^2, \beta, \phi \sim & \mathrm{N}_n\left(S\beta, C\right), \text{ spatial process model} \\ \beta \sim & \mathrm{N}_4\left(b, Iv\right), \text{ hyper-parameter prior model} \\ \sigma^2 \sim & \mathrm{IG}\left(\kappa_\sigma, \lambda_\sigma\right), \text{ hyper-parameter prior model} \\ \phi \sim & \mathrm{IG}\left(\kappa_\phi, \lambda_\phi\right), \text{ hyper-parameter prior model} \end{cases}$$

Exercise 6. ($\star$)An example of spatio-temporal data where aerial spatial data are time referenced is given in Figure 8 which shows a spatio-temporal dataset representing the population of the counties of Ohio, USA, from 1968 to 1988. The dataset is available from the SpatialEpiApp R package. Interest lies in not only how the population is only distributed over the spatial domain but also how it evolves during the time. To which principal spatial statistical are would you associate this problem?

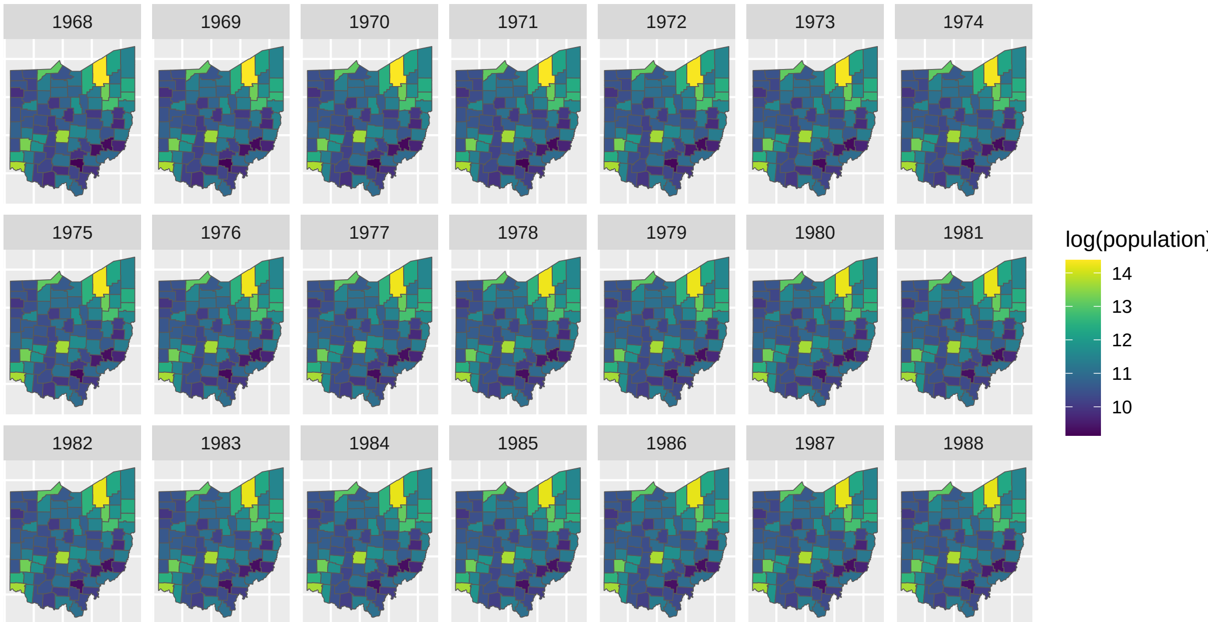


FIGURE 8. Population of the counties of Ohio, USA, from 1968 to 1988.

Solution. Aerial spatio-temporal data

Part 2. **Point referenced data / Geostatistics**

Exercise 7. ($\star$)If $c:\mathbb{R}^d\rightarrow\mathbb{R}$ 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 $\{s_1,...,s_n\}\subseteq S$

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

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

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

Well it is

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

Exercise 8. ($\star$) Show that if $c_1(\cdot,\cdot)$ and $c_2(\cdot,\cdot)$ are are covariance functions (are non-negative definite) then so are $c_3(\cdot,\cdot)=bc_1(\cdot,\cdot)+dc_2(\cdot,\cdot)$ with $b,d\geq 0$ and $c_4(\cdot,\cdot)=c_1(\cdot,\cdot)c_2(\cdot,\cdot)$.

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

$$\begin{aligned}
\sum_{i=1}^{n}a_i\sum_{j=1}^{n}a_jc_3(s_i,s_j) &= \sum_{i=1}^{n}a_i\sum_{j=1}^{n}a_j(bc_1(s_i,s_j)+dc_2(s_i,s_j))\\
&=\underbrace{\sum_{i=1}^{n}a_i\sum_{j=1}^{n}a_jbc_1(s_i,s_j)}_{\geq 0}+\underbrace{\sum_{i=1}^{n}a_i\sum_{j=1}^{n}a_jdc_2(s_i,s_j)}_{\geq 0}\\
&\geq 0
\end{aligned}$$

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

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

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

Exercise 9. ($\star$) 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

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

Exercise 10. ($\star$) 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

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

where

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

hence

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

(Given as Formative assessment 1)

Exercise 11. ($\star$) Let $Z = \left(Z\left(s\right) : s \in \mathbb{R}^d\right)$ be an intrinsic random field with $\mathrm{E}\left(Z\left(s\right) - Z\left(t\right)\right) = 0$ and let $\gamma : \mathbb{R}^d \to \mathbb{R}$ be its semivariogram.

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

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

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

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

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

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

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

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

Let $c_I\left(s, t\right) = \mathrm{E}\left(\left(Z\left(s_i\right) - Z\left(s_0\right)\right)\left(Z\left(s_j\right) - Z\left(s_0\right)\right)\right)$.

(2) It is

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

It is

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

hence

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

(Given as Formative assessment 1)

Exercise 12. ($\star$) Consider the zero-mean random field $Z=\left(Z(s): s\in\mathbb{R}^d\right)$ with covariogram 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 semivariogram for the random field $\left(Z(s): s\in\mathbb{R}^d\right)$
(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.

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

Solution.

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

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

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

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

  then

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

- The nugget effect is the limiting value of the semicovariogram as $h\to0$ from above, hence it is $\gamma(h)\to\nu^2$ as $h\to0^+$.
- 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.

(Given as Formative assessment 2)

Exercise 13. ($\star$) 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 $c_w\left(h;\sigma^2,\phi\right)=\sigma^2\exp\left(-\frac{1}{\phi}\left\|h\right\|\right)$, $\mu(s;\beta)$ 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 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 $\theta = (\sigma^2, \phi, \xi)$.

(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. (β, σ^2), 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 $\mathrm{E}\left(Z^\top AZ\right) = \mathrm{E}(Z)^\top A\mathrm{E}(Z)^\top + \mathrm{tr}(A\mathrm{Var}(Z))$ 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 $\left(\hat{\phi}, \hat{\xi}\right)$ for the MLE of (ϕ, ξ)

Solution. It is

(1) It is

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

(2) It is

(a)

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

(b)

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

(3) It is

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

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

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

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

So

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

where

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

then

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

So it is

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

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

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

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 14. ($\star$) Let $(Z(s):s\in\mathcal{S})$ be a specified statistical model. Assume that $(Z(s):s\in\mathcal{S})$ is weakly stationary with unknown constant mean $\mu=\mathrm{E}(Z(s))$ and known covariogram $c(\cdot)$. Assume there is available a dataset $\{(s_i,Z_i:=Z(s_i))\}_{i=1}^n$ and assume they are realizations of $(Z_s)_{s\in\mathcal{S}}$. Assume that the matrix C such as $[C]_{i,j}=c(\|s_i-s_j\|)$ 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 $\sigma_{\mathrm{KM}}=\sqrt{\mathrm{E}\left(\mu_{\mathrm{KM}}-\mu\right)^2}$ as a function of C

Solution. 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

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

which is satisfied given the assumption

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

(2) It is

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

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

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

The Lagrange function is

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

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

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

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

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

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

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

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

(Given as Formative assessment 2)

Exercise 15. ($\star$) 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

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

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

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

(1) It is

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

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

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

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

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

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

where $C_0=\mathrm{Cov}\left(Z, Z\left(s_0\right)\right)$, 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

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

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

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

Exercise 16. ($\star$) Assume a spatial model

$$Z(s)=\mu+\delta(s), \; s \in \mathcal{S} \tag{2}$$

with unknown mean $\mu \in \mathbb{R}$. Assume a set of n observed realizations $Z_i:=Z\left(s_i\right)$ of (2) at sites s_i for $i=1, \ldots, n$. Assume that $Z(s)$ is a weak stationary stochastic process with known covariogram $c(\cdot)$. 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 $c(h)$ and not the semi-variogram.

Solution.

Exercise 17. ($\star$) Let $(Z(s); s \in \mathcal{S})$ be a specified statistical model. Assume that $(Z(s); s \in \mathcal{S})$ is an intrinsic stationary process with unknown constant mean $\mu = \mathrm{E}(Z(s))$ and known semi-variogram $\gamma(\cdot)$. Assume there is available a dataset $\{(s_i, Z_i := Z(s_i))\}_{i=1}^{n}$. Consider the "Kriging" estimator $Z_{\mathrm{OK}}(s_0)$ of $Z(s_0)$ at any unseen spatial location s_0 as the BLUE (Best Linear Unbiased Estimator)

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

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 $w = (w_1, ..., w_n)$ so that the Kriging estimator $Z_{\mathrm{OK}}(s_0)$ to be unbiased.

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

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

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

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

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

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

Exercise 18. ($\star$)

Inventory of useful formulas.

[Normal distr. conditioning] Let $x_1 \in \mathbb{R}^{d_1}$, and $x_2 \in \mathbb{R}^{d_2}$. If

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathrm{N}_{d_1+d_2}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_1 & \Sigma_{21}^\top \\ \Sigma_{21} & \Sigma_2 \end{bmatrix}\right)$$

then it is

$$x_2 | x_1 \sim \mathrm{N}_{d_2}\left(\mu_{2|1}, \Sigma_{2|1}\right)$$

where

$$\mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_1^{-1}(x_1 - \mu_1) \ \text{ and } \ \Sigma_{2|1} = \Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{21}^\top$$

Consider the Bayesian Kriging from your lecture notes:

$$Z(s) = Y(s) + \varepsilon(s), \ \ s \in \mathcal{S}$$

where

$$\varepsilon(\cdot) \sim \mathrm{GP}(0, c_\varepsilon(\cdot, \cdot|\tau))$$

with $c_\varepsilon(s, s'|\tau) = \tau^2 1_{\{0\}}(\|s - s'\|)$ and

$$Y(\cdot)|\beta, \theta \sim \mathrm{GP}\left(\mu(\cdot|\beta), c_Y\left(\cdot, \cdot|\sigma^2, \phi\right)\right)$$

with mean function $\mu(\cdot|\beta)$ (to be specified later) labeled by unknown parameter β and covariance function $c_Y\left(\cdot, \cdot|\sigma^2, \phi\right)$.

Assume there is available a dataset $\{(s_i, Z_i)\}_{i=1}^n$ where $Z_i = Z(s_i)$ is a realization of a stochastic process (Z_s).

(1) Write the hierarchical spatial model $Z(\cdot)|Y(\cdot), \beta, \varphi$ and $Y(\cdot)|\beta, \varphi$ where $\varphi = \left(\sigma^2, \phi, \tau\right)^\top$.

(2) Write the marginal process $Z(\cdot)|\beta, \varphi$ where $\varphi = \left(\sigma^2, \phi, \tau\right)^\top$, its mean function denoted as $\mu(\cdot|\cdot)$, and its covariance function denoted as $c(\cdot|\cdot)$.

Hint:: Let Y and X be independent random variables with $X \sim \mathrm{N}(\mu_X, \Sigma_X)$, $Y \sim \mathrm{N}(\mu_Y, \Sigma_Y)$. Let A and B be fixed matrices. Let c be a fixed vector. Then

$$AX + BY + c \sim \mathrm{N}\left(A\mu_X + B\mu_Y + c, A\Sigma_X A^\top + B\Sigma_Y B^\top\right)$$

(3) Compute the predictive process $Z(\cdot)|Z, \beta, \varphi$ as

$$Z(\cdot)|Z, \beta, \varphi \sim \mathrm{GP}\left(\mu_1(\cdot|\beta, \varphi), c_1(\cdot, \cdot|\varphi)\right)$$

with

$$\begin{aligned} c_1\left(s, s'|\varphi\right) =& c(s, s|\varphi) + (C(S, s|\varphi))^\top (C(S, S|\varphi))^{-1} C\left(S, s'|\varphi\right) \\ \mu_1(s|\beta, \varphi) =& \mu(s|\beta) - (C(S, s|\varphi))^\top (C(S, S|\varphi))^{-1} (\mu(S|\beta) - Z) \end{aligned}$$

Hint: See the Conditional Normal formula above.

(4) Assume $\mu(s|\beta) = \psi(s)^\top \beta$. Consider a conjugate prior $\beta \sim \mathrm{N}(b, B)$ on β where $B > 0$.

(a) Write down the Bayesian statistical model involving layers $[Z|\beta, \varphi]$, and $[\beta|\varphi]$.

(b) Compute the posterior distribution as

$$\beta|Z, \varphi \sim \mathrm{N}(b_n(\varphi), B_n(\varphi))$$

with

$$\begin{aligned} B_n(\varphi) =& \left(B^{-1} + \Psi^\top (C(S, S|\varphi))^{-1} \Psi\right)^{-1} \\ b_n(\varphi) =& B_n(\varphi)\ \left(B^{-1} b + \Psi^\top (C(S, S|\varphi))^{-1} Z\right) \end{aligned}$$

where $C(S, S|\varphi)$ is a matrix with $[C(S, S|\varphi)]_{i,j} = c(s_i, s_j|\varphi)$.

Hint: Use the following identity

$$(y - \Phi\beta)^\top \Sigma^{-1} (y - \Phi\beta) + (\beta - \mu)^\top V^{-1} (\beta - \mu) = (\beta - \mu^*)^\top (V^*)^{-1} (\beta - \mu^*) + S^*;$$

$$V^* = \left(V^{-1} + \Phi^\top \Sigma^{-1} \Phi\right)^{-1}; \qquad \mu^* = V^*\left(V^{-1}\mu + \Phi^\top \Sigma^{-1} y\right)$$

$$S^* = \mu^\top V^{-1} \mu - (\mu^*)^\top (V^*)^{-1} (\mu^*) + y^\top \Sigma^{-1} y;$$

(c) Compute the (posterior) predictive process $Z(\cdot)|Z,\varphi$ given the data Z and given the parameters φ as

$$Z(\cdot)|Z,\varphi \sim \mathrm{GP}\left(\mu_2(\cdot|\varphi), c_2(\cdot,\cdot|\varphi)\right)$$

with

$$\begin{aligned}\mu_2(s|\varphi) =& \left(\psi(s) - \Psi^\top C^{-1} C(s)\right)^\top \left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1} B^{-1} b \\ &+ \left[(C(s))^\top + \left(\psi(s) - \Psi^\top C^{-1} C(s)\right)^\top \left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\Psi\right] C^{-1} Z \\ c_2(s,s'|\varphi) =& c(s,s'|\varphi) - (C(s))^\top C^{-1} C(s') \\ &+ \left(\psi(s) - \Psi^\top C^{-1} C(s)\right)^\top \left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1} \left(\psi(s') - \Psi^\top C^{-1} C(s')\right)\end{aligned}$$

with column vector $C(s) := (c(s,s_1|\varphi), ..., c(s,s_n|\varphi))^\top$, and matrix $C := C(S,S|\varphi)$.

(d) Compute the marginal likelihood $\Pr(Z|\varphi)$ in the form

$$\Pr\left(Z|\sigma^2,\varphi\right) = \mathrm{N}\left(Z|\Psi b, \left(C^{-1} - C^{-1}\Psi\left(B^{-1} + \Psi^\top B^{-1}\Psi\right)^{-1}\Psi^\top C^{-1}\right)^{-1}\right)$$

where Ψ is a matrix with $[\Psi]_{i,j} = \psi_j(s_i)$, and R is a matrix with $[C]_{i,j} = c(s_i,s_j|\varphi)$.

Hint-2:: It is

$$\int \mathrm{N}(Z|\Psi\beta, C)\,\mathrm{N}(\beta|b,B)\,\mathrm{d}\beta = \mathrm{N}\left(Z|\Psi b, C + \Psi B \Psi^\top\right)$$

Hint 3:: [Woodbury matrix identity]

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1} + VA^{-1}U\right)^{-1} VA^{-1}$$

(5) Consider non-informative prior $\Pr(\beta) \propto 1$ for β by specifying $b \to 0$ and letting $B^{-1} \to 0$. Argue whether such a prior can be used. Recompute the (asymptotic) quantities $\Pr(Z|\varphi)$, $[Z(\cdot)|Z,\varphi]$ under this new prior in the limit.

Solution.

(1) The hierarchical model is

$$\begin{aligned}Z(\cdot)|Y(\cdot),\tau &\sim \mathrm{GP}\left(Y(\cdot), c_\varepsilon(\cdot,\cdot|\tau)\right) \\ Y(\cdot)|\beta, &\sim \mathrm{GP}\left(\mu(\cdot|\beta), c_Y\left(\cdot,\cdot|\sigma^2,\phi\right)\right)\end{aligned}$$

(2) We use the additive property of the Gaussian distribution (in Hint) it is

$$Z(\cdot)|\beta,\varphi \sim \mathrm{GP}\left(\mu(\cdot|\beta), c(\cdot,\cdot|\varphi)\right)$$

where

$$c(s,s'|\varphi) = c_Y\left(s,s'|\sigma^2,\phi\right) + c_\varepsilon\left(s,s'|\tau\right)$$

(3) Assume a vector of "unseen" sites $S_* = (s_{*,1}, ..., s_{*,q})^\top$ for any $q \in \mathbb{N}_0$. Let convenient notation $Z := Z(S)$, and $Z_* := Z(S_*)$. The join marginal distribution of $(Z_*, Z)^\top$ given β,

$\varphi = \left(\sigma^2, \phi, \tau\right)^\top$ is

$$\begin{pmatrix} Z_* \\ Z \end{pmatrix} |\beta, \varphi \sim \mathrm{N}\left(\begin{pmatrix} \mu\left(S_*;\beta\right) \\ \mu\left(S;\beta\right) \end{pmatrix}, \begin{pmatrix} C\left(S_*, S_*|\varphi\right) & \left(C\left(S_*, S|\varphi\right)\right)^\top \\ C\left(S_*, S|\varphi\right) & C\left(S, S|\varphi\right) \end{pmatrix}\right)$$

by using convenient notation $\left[C\left(S_*, S|\varphi\right)\right]_{i,j} = s\left(s_{*,i}, s_j|\varphi\right)$ and $\left[\mu\left(S;\beta\right)\right]_i = \mu\left(s_i;\beta\right)$. By conditioning the Normal distribution (see Hint), I get

$$Z_*|Z, \beta, \varphi \sim \mathrm{N}\left(\mu_*\left(S_*|\beta, \varphi\right), C_*\left(S_*, S_*|\varphi\right)\right)$$

where

$$\begin{aligned} C_1\left(S_*, S_*|\varphi\right) =& C\left(S_*, S_*|\varphi\right) + \left(C\left(S, S_*|\varphi\right)\right)^\top \left(C\left(S, S|\varphi\right)\right)^{-1} C\left(S, S_*|\varphi\right) \\ \mu_1\left(S_*|\beta, \varphi\right) =& \mu\left(S_*|\beta\right) - \left(C\left(S, S_*|\varphi\right)\right)^\top \left(C\left(S, S|\varphi\right)\right)^{-1} \left(\mu\left(S|\beta\right) - Z\right) \end{aligned}$$

As it is for any length of of any vector S_*, then it is a Gaussian process

$$Z\left(\cdot\right)|Z, \beta, \varphi \sim \mathrm{GP}\left(\mu_1\left(\cdot|\beta, \varphi\right), c_1\left(\cdot, \cdot|\varphi\right)\right)$$

with

$$\begin{aligned} c_1\left(s, s'|\varphi\right) =& c\left(s, s'|\varphi\right) + \left(C\left(S, s|\varphi\right)\right)^\top \left(C\left(S, S|\varphi\right)\right)^{-1} C\left(S, s'|\varphi\right) \\ \mu_1\left(s|\beta, \varphi\right) =& \mu\left(s|\beta\right) - \left(C\left(S, s|\varphi\right)\right)^\top \left(C\left(S, S|\varphi\right)\right)^{-1} \left(\mu\left(S|\beta\right) - Z\right) \end{aligned}$$

(4)

(a) The Bayesian model is

$$\begin{cases} Z|\beta, \varphi \sim \mathrm{N}\left(\Psi\beta, C\left(S, S|\varphi\right)\right) \\ \beta \sim \mathrm{N}\left(b, B\right) \end{cases} \tag{3}$$

(b) Let $C := C\left(S, S|\varphi\right)$. The posterior distribution (by using Bayes theorem) is

$$\begin{aligned} \Pr\left(\beta|Z, \varphi\right) \propto& \Pr\left(Z|\beta, \varphi\right)\Pr\left(\beta|\varphi\right) \\ =& \mathrm{N}\left(Z|\Psi\beta, C\right)\mathrm{N}\left(\beta|b, B\right) \\ \propto& \exp\left(-\frac{1}{2}\left(Z - \Psi\beta\right)^\top C^{-1}\left(Z - \Psi\beta\right)\right)\exp\left(-\frac{1}{2}\left(\beta - b\right)^\top B^{-1}\left(\beta - b\right)\right) \\ =& \exp\left(-\frac{1}{2}\left[\left(Z - \Psi\beta\right)^\top C^{-1}\left(Z - \Psi\beta\right) + \left(\beta - b\right)^\top B^{-1}\left(\beta - b\right)\right]\right) \end{aligned}$$

By using the Hint I have

$$\left(Z - \Psi\beta\right)^\top C^{-1}\left(Z - \Psi\beta\right) + \left(\beta - b\right)^\top B^{-1}\left(\beta - b\right) = \left(\beta - b_n\right)^\top \left(B_n\right)^{-1}\left(\beta - b_n\right) + R_n$$

where by denoting $B_n := B_n\left(\varphi\right)$, and $b_n := b_n\left(\varphi\right)$ I get

$$\begin{aligned} B_n =& \left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1} \\ b_n =& B_n\left(B^{-1}b + \Psi^\top C^{-1}Z\right) \end{aligned}$$

and R_n is a "constant" quantity that does not contain any β. Hence

$$\begin{aligned}\Pr(\beta|Z,\varphi) &\propto \exp\left(-\frac{1}{2}(\beta-b_n)^\top (B_n)^{-1}(\beta-b_n)-\frac{1}{2}R_n\right)\\ &\propto \exp\left(-\frac{1}{2}(\beta-b_n)^\top (B_n)^{-1}(\beta-b_n)\right)\end{aligned}$$

Well, from the above, I recognize the kernel of the Multivariate Normal distribution, as

$$\beta|Z,\varphi \sim \mathrm{N}\left(b_n(\varphi), B_n(\varphi)\right)$$

(c) Assume a vector of "unseen" sites $S_* = (s_{*,1},...,s_{*,q})^\top$ for any $q \in \mathbb{N}-\{0\}$. Let convenient notation $Z := Z(S)$, and $Z_* := Z(S_*)$. I have already computed

$$\Pr(Z_*|Z,\beta,\varphi) = \mathrm{N}\left(Z_*|\mu_1(S_*|\beta,\varphi), C_1(S_*,S_*|\varphi)\right)$$

from the previous part. It is

$$\begin{aligned}\Pr(Z_*|Z,\varphi) &= \int \Pr(Z_*|Z,\beta,\varphi)\Pr(\beta|Z,\varphi)\,\mathrm{d}\beta\\ &= \int \mathrm{N}\left(Z_*|\mu_1(S_*|\beta,\varphi), C_1(S_*,S_*|\varphi)\right)\mathrm{N}(\beta|b_n,B_n)\,\mathrm{d}\beta\end{aligned}$$

Denote $\Psi_* = \Psi(S_*)$, $C_* = C(S_*,S|\varphi)$, and $C_{**} = C(S_*,S_*|\varphi)$. Notice that

$$\begin{aligned}\mu_1(S_*) =& \Psi_*\beta - C_*C^{-1}(\Psi\beta - Z)\\ =& \left[\Psi_* - C_*C^{-1}\Psi\right]\beta + C_*C^{-1}Z\end{aligned}$$

Hence, for given/fixed Z,φ, it is

$$Z_* = C_*C^{-1}Z + \left[\Psi_* - C_*C^{-1}\Psi\right]\beta + \zeta, \quad \zeta \sim \mathrm{N}\left(0, C_1(S_*,S_*)\right)$$

Hence, because $\beta \sim \mathrm{N}(b_n,B_n)$, and because $Z_*|Z,\varphi$ is a linear combination of the Normally distributed random vector $\beta \sim \mathrm{N}(b_n,B_n)$, $Z_*|Z,\varphi$ follows a Normal distribution, with mean

$$\begin{aligned}\mu_2(S_*) =& \mathrm{E}_{\beta\sim\mathrm{N}(b_n,B_n)}\left(Z_*|\mu_1(S_*), C_1(S_*,S_*)\right)\\ =& \left(\Psi_* - C_*C^{-1}\Psi\right)\mathrm{E}_{\beta\sim\mathrm{N}(b_n,B_n)}(\beta) + C_*C^{-1}Z\\ =& \left(\Psi_* - C_*C^{-1}\Psi\right)b_n + C_*C^{-1}Z\\ =& \left(\Psi_* - C_*C^{-1}\Psi\right)\left(B^{-1}+\Psi^\top C^{-1}\Psi\right)^{-1}\left(B^{-1}b + \Psi^\top C^{-1}Z\right) + C_*C^{-1}Z\\ =& \left(\Psi_* - C_*C^{-1}\Psi\right)\left(B^{-1}+\Psi^\top C^{-1}\Psi\right)^{-1}B^{-1}b\\ &+ \left[\left(\Psi_* - C_*C^{-1}\Psi\right)\left(B^{-1}+\Psi^\top C^{-1}\Psi\right)^{-1}\Psi^\top + C_*\right]C^{-1}Z\end{aligned}$$

and with covariance matrix

$$\begin{aligned}
C_2(S_*, S_*) =& \mathrm{Var}_{\beta\sim \mathrm{N}(b_n,B_n)}\left(Z_*|\mu_1(S_*), C_1(S_*,S_*)\right) \\
=& \mathrm{Var}_{\beta\sim \mathrm{N}(b_n,B_n)}\left(\left[\Psi_* - C_* C^{-1}\Psi\right]\beta\right) + \mathrm{Var}_{\zeta\sim \mathrm{N}(0,C_1(S_*,S_*))}(\zeta) \\
=& \left[\Psi_* - C_* C^{-1}\Psi\right] B_n \left[\Psi_* - C_* C^{-1}\right]^\top + C_1(S_*,S_*) \\
=& \left[\Psi_* - C_* C^{-1}\Psi\right]\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\left[\Psi_* - C_* C^{-1}\Psi\right]^\top \\
& + C_{**} + C_* C^{-1}(C_*)^\top
\end{aligned}$$

Recall that $C(s) = (c(s_1, s|\varphi), ..., c(s_n, s|\varphi))^\top$ is a column vector.
Since this is for any vector S_* of any length, then

$$Z(\cdot)|Z,\varphi \sim \mathrm{GP}(\mu_2(\cdot|\varphi), c_2(\cdot,\cdot|\varphi))$$

with mean function at s

$$\begin{aligned}
\mu_2(s|\varphi) =& \left(\psi(s) - (C(s))^\top C^{-1}\Psi\right)\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1} B^{-1} b \\
& + \left[\left(\psi(s) - (C(s))^\top C^{-1}\Psi\right)\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\Psi^\top + (C(s))^\top\right] C^{-1} Z
\end{aligned}$$

and covariance function as s, s'

$$\begin{aligned}
c_2(s,s'|\varphi) =& \left[\psi(s) - (C(s))^\top C^{-1}\Psi\right]\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\left[\psi(s') - \left(C(s')\right)^\top C^{-1}\Psi\right]^\top \\
& + c\left(s,s'|\varphi\right) + (C(s))^\top C^{-1} C\left(s'\right)
\end{aligned}$$

Recall that $C(s) = (c(s_1, s|\varphi), ..., c(s_n, s|\varphi))^\top$ is a column vector.

(d) It is, from Hint-2,

$$\begin{aligned}
\Pr(Z|\varphi) =& \int \Pr(Z|\beta,\varphi)\Pr(\beta)\,\mathrm{d}\beta \\
=& \int \mathrm{N}(Z|\Psi\beta, C(S,S|\varphi))\,\mathrm{N}(\beta|b,B)\,\mathrm{d}\beta \\
=& \int \mathrm{N}(Z|\Psi\beta, C(S,S|\varphi))\,\mathrm{N}\left(\Psi\beta|\Psi b, \Psi B\Psi^\top\right)\mathrm{d}\beta \\
=& \mathrm{N}\left(Z|\Psi b, C(S,S|\varphi) + \Psi B\Psi^\top\right)
\end{aligned}$$

By letting $C := C(S,S|\varphi)$ and using the Hint I get

$$\left(\left(C + \Psi B\Psi^\top\right)^{-1}\right)^{-1} = \left(C^{-1} - C^{-1}\Psi\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\Psi^\top C^{-1}\right)^{-1}$$

(5) Denote $C = C(S, S|\varphi)$. It is

$$\lim_{\substack{B^{-1}\to 0\\ b\to 0}} \Pr(Z|\varphi) = \lim_{\substack{B^{-1}\to 0\\ b\to 0}} \mathrm{N}\left(Z|\Psi b, C + \Psi B \Psi^{\top}\right)$$
$$\propto \mathrm{N}\left(Z|0, \left(C^{-1} - C^{-1}\Psi\left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\Psi^{\top}C^{-1}\right)^{-1}\right)$$
$$<\infty$$

namely the bottom part of the fraction of the posterior of $\beta|Z,\varphi$ is bounded (finite); this implies that the posterior is proper. The posterior of $\beta|Z,\varphi$ has density such as

$$\Pr(\beta|Z,\varphi) \propto \exp\left(-\frac{1}{2}(\beta - b_n)^{\top} B_n^{-1}(\beta - b_n)\right)$$

then by computing the limit

$$\lim_{\substack{B^{-1}\to 0\\ b\to 0}} \exp\left(-\frac{1}{2}(\beta - b_n)^{\top} B_n^{-1}(\beta - b_n)\right) =$$
$$\exp\left(-\frac{1}{2}\left(\beta - \left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\Psi^{\top}C^{-1}Z\right)^{\top}\left(\Psi^{\top}C^{-1}\Psi\right)\left(\beta - \left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\Psi^{\top}C^{-1}Z\right)\right)$$

Hence the limiting case is

$$\beta|Z,\varphi \overset{\text{approx}}{\sim} \mathrm{N}\left(\left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\Psi^{\top}C^{-1}Z,\ \left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\right)$$

Hence the predictive process becomes

$$Z(\cdot)|Z,\varphi \overset{\text{approx}}{\sim} \mathrm{GP}\left(\mu_3(\cdot|\varphi), c_3(\cdot,\cdot|\varphi)\right)$$

$$\mu_3(s|\varphi) = \left[\left(\psi(s) - (C(s))^{\top}C^{-1}\Psi\right)\left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\Psi^{\top} + (C(s))^{\top}\right]C^{-1}Z$$
$$c_3(s, s'|\varphi) = \left(\psi(s) - (C(s))^{\top}C^{-1}\Psi\right)\left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\left(\psi(s') - (C(s'))^{\top}C^{-1}\Psi\right)^{\top}$$
$$+ c(s, s'|\varphi) + (C(s))^{\top}C^{-1}C(s')$$

Exercise 19. ($\star$) Consider the Bayesian Kriging in the Lecture notes, in particular the "Gaussian process regression".

The Bayesian hierarchical model summaries to

$$\begin{cases} Z|Y,\sigma^2 \sim \mathrm{N}\left(Y, I\sigma^2\right) & \text{data model} \\ Y|\beta,\theta \sim \mathrm{N}\left(\mu(S), c(S,S|\theta)\right) & \text{spatial process model} \\ \beta \sim \mathrm{N}(b, B) & \text{hyper-prior model} \end{cases} \quad (4)$$

The estimates $\hat{\theta}$ and $\hat{\sigma}^2$ of the unknown fixed hyper-parameters θ and σ^2 are said to be computed as

$$\left(\hat{\theta}, \hat{\sigma}^2\right) = \underset{\theta, \sigma^2}{\arg\min} \left(-2 \log \left(\mathrm{N}\left(Z|\Psi b, C(\theta) + I\sigma^2 + \Psi B \Psi^\top\right)\right)\right)$$

Show that if I consider noiseless observations (which can be achieved by $\sigma^2 \to 0$) and uniform hyper-prior for β i.e. $\Pr(\beta) \propto 1$ (which can be achieved a $b \to 0$ and $B^{-1} \to 0$) then I get the following results.

(1) The asymptotic posterior of β under the assumptions of noiseless data and uniform prior of β is proper.

Hint 1: The posterior of β under the Bayesian model (4) is

$$\beta|Z, \theta, \sigma^2 \sim \mathrm{N}\left(b_n\left(\theta, \sigma^2\right), B_n\left(\theta, \sigma^2\right)\right)$$

with

$$\begin{aligned} B_n\left(\theta, \sigma^2\right) &= \left(B^{-1} + \Psi^\top \left(C(\theta) + I\sigma^2\right)^{-1} \Psi\right)^{-1} \\ b_n\left(\theta, \sigma^2\right) &= B_n\left(\theta, \sigma^2\right) \left(B^{-1} b + \Psi^\top \left(C(\theta) + I\sigma^2\right)^{-1} Z\right) \end{aligned}$$

Hint 2: (Woodbury matrix identity)

$$(A + BCD)^{-1} = A^{-1} - A^{-1} B \left(C^{-1} + D A^{-1} B\right)^{-1} D A^{-1}$$

(2) The asymptotic predictive random field $Y(\cdot)|Z, \theta$ is

$$Y(\cdot)|Z, \theta \sim \mathrm{GP}\left(\tilde{\mu}_2(\cdot|\theta), \tilde{c}_2(\cdot, \cdot|\theta)\right)$$

$$\begin{aligned} \tilde{\mu}_2(s|\theta) &= \left[\left(\psi(s) - (C(s|\theta))^\top C^{-1} \Psi\right) \left(\Psi^\top C(\theta)^{-1} \Psi\right)^{-1} \Psi^\top + (C(s|\theta))^\top\right] C^{-1} Z \\ \tilde{c}_2\left(s, s'|\theta\right) &= \left[\psi(s) - (C(s|\theta))^\top C(\theta)^{-1} \Psi\right] \left(\Psi^\top C(\theta)^{-1} \Psi\right)^{-1} \left[\psi\left(s'\right) - \left(C\left(s'|\theta\right)\right)^\top C(\theta)^{-1} \Psi\right]^\top \\ &\quad + c\left(s, s'|\theta\right) + (C(s|\theta))^\top C(\theta)^{-1} C\left(s'|\theta\right) \end{aligned}$$

aka the same as those of Universal Kriging. Here $C(\theta) := C(S, S|\theta)$ and $C(s|\theta) = (c(s, s_1|\theta), ..., c(s, s_n|\theta))^\top$.

Hint: The predictive random field $Y(\cdot)|Z, \theta, \sigma^2$ under the Bayesian model (4) is

$$Y(\cdot)|Z, \theta, \sigma^2 \sim \mathrm{GP}\left(\mu_2\left(\cdot|\theta, \sigma^2\right), c_2\left(\cdot, \cdot|\theta, \sigma^2\right)\right)$$

$$\begin{aligned} \mu_2\left(s|\theta, \sigma^2\right) &= \psi(s) b_n\left(\theta, \sigma^2\right) - (C(s|\theta))^\top \left(C(\theta) + I\sigma^2\right)^{-1} \left(\Psi b_n\left(\theta, \sigma^2\right) - Z\right) \\ c_2\left(s, s'|\theta, \sigma^2\right) &= c_1\left(s, s'|\theta, \sigma^2\right) \\ &\quad + \left[\psi(s) - (C(s|\theta))^\top \left(C(\theta) + I\sigma^2\right)^{-1} \Psi\right] B_n\left(\theta, \sigma^2\right) \left[\psi(s) - (C(s|\theta))^\top \left(C(\theta) + I\sigma^2\right)^{-1} \Psi\right] \\ c_1\left(s, s'|\theta, \sigma^2\right) &= c\left(s, s'|\theta\right) + (C(S, s|\theta))^\top \left(C(S, S|\theta) + I\sigma^2\right)^{-1} C\left(S, s'|\theta\right) \end{aligned}$$

(3) the estimate $\hat{\theta}$ of the unknown fixed hyper-parameter θ is computed as

$$\hat{\theta} = \arg\min_{\theta,\sigma^2} \left(-\frac{n}{4} \log \left(\left| C(\theta)^{-1} - C^{-1}\Psi \left(\Psi^\top C(\theta)^{-1} \Psi \right)^{-1} \Psi^\top C(\theta)^{-1} \right| \right) \right.$$
$$\left. + Z^\top \left[C(\theta)^{-1} - C(\theta)^{-1} \Psi \left(\Psi^\top C(\theta)^{-1} \Psi \right)^{-1} \Psi^\top C(\theta)^{-1} \right] Z \right)$$

where $C(\theta) := C(S, S|\theta)$.

Hint: *The* estimates $\hat{\theta}$ and $\hat{\sigma}^2$ of the unknown fixed hyper-parameters θ and σ^2 under the Bayesian model (4) is are

$$\left(\hat{\theta}, \hat{\sigma}^2 \right) = \arg\min_{\theta,\sigma^2} \left(-2 \log \left(\mathrm{N} \left(Z | \Psi b, C(\theta) + I\sigma^2 + \Psi B \Psi^\top \right) \right) \right)$$

Solution. The noiseless data and the non-informative prior can be taken as $\sigma^2 \to 0$, $b \to 0$ and $B^{-1} \to 0$. So I will pass the limit in the distributions.

(1) As $\sigma^2 \to 0$, $b \to 0$ and $B^{-1} \to 0$, it is

$$B_n\left(\theta, \sigma^2\right) = \left(B^{-1} + \Psi^\top \left(C(\theta) + I\sigma^2 \right)^{-1} \Psi \right)^{-1}$$
$$\to \left(\Psi^\top \left(C(\theta) \right)^{-1} \Psi \right)^{-1} = \tilde{B}_n(\theta)$$

and

$$b_n\left(\theta, \sigma^2\right) = B_n\left(\theta, \sigma^2\right) \left(B^{-1} b + \Psi^\top \left(C(\theta) + I\sigma^2 \right)^{-1} Z \right)$$
$$\to \left(\Psi^\top \left(C(\theta) \right)^{-1} \Psi \right)^{-1} \Psi^\top \left(C(\theta) \right)^{-1} Z = \tilde{b}_n(\theta)$$

Hence

$$\beta | Z, \theta \sim \mathrm{N}\left(\tilde{b}_n(\theta), \tilde{B}_n(\theta) \right)$$

which is a proper posterior distribution and hence can be used for inference.

(2) As $\sigma^2 \to 0$, $b \to 0$ and $B^{-1} \to 0$, it is

$$\begin{aligned}
c_2\left(s, s'|\theta, \sigma^2\right) =& c_1\left(s, s'|\theta, \sigma^2\right) \\
&+ \left[\psi(s) - \left(C(s|\theta) \right)^\top \left(C(\theta) + I\sigma^2 \right)^{-1} \Psi \right] B_n\left(\theta, \sigma^2\right) \left[\psi(s) - \left(C(s|\theta) \right)^\top \left(C(\theta) + I\sigma^2 \right)^{-1} \Psi \right] \\
\to& \left[\psi(s) - \left(C(s|\theta) \right)^\top C^{-1} \Psi \right] \left(\Psi^\top C(\theta)^{-1} \Psi \right)^{-1} \left[\psi(s') - \left(C(s'|\theta) \right)^\top C(\theta)^{-1} \Psi \right]^\top \\
&+ c\left(s, s'|\theta\right) + \left(C(s|\theta) \right)^\top C(\theta)^{-1} C\left(s'|\theta\right) \\
=& \tilde{c}_2\left(s, s'|\theta\right)
\end{aligned}$$

and

$$\mu_2\left(s|\theta,\sigma^2\right)=\psi\left(s\right)b_n\left(\theta,\sigma^2\right)-\left(C\left(s|\theta\right)\right)^\top\left(C\left(\theta\right)+I\sigma^2\right)^{-1}\left(\Psi b_n\left(\theta,\sigma^2\right)-Z\right)$$

$$\to\psi\left(s\right)\left(\Psi^\top C\left(\theta\right)^{-1}\Psi\right)^{-1}\Psi^\top C\left(\theta\right)^{-1}Z-\left(C\left(s|\theta\right)\right)^\top C^{-1}\left(\Psi\left(\Psi^\top C\left(\theta\right)^{-1}\Psi\right)^{-1}\Psi^\top C\left(\theta\right)^{-1}Z-Z\right)$$

$$=\left[\left(\psi\left(s\right)-\left(C\left(s|\theta\right)\right)^\top C\left(\theta\right)^{-1}\Psi\right)\left(\Psi^\top C\left(\theta\right)^{-1}\Psi\right)^{-1}\Psi^\top+\left(C\left(s|\theta\right)\right)^\top\right]C\left(\theta\right)^{-1}Z$$

$$=\tilde{\mu}_2\left(s|\theta\right)$$

(3) Similarly

$$-2\log\left(\mathrm{N}\left(Z|\Psi b,C\left(\theta\right)+I\sigma^2+\Psi B\Psi^\top\right)\right)$$

$$=-\frac{1}{4}\log\left(\left|\left(C\left(\theta\right)+I\sigma^2+\Psi B\Psi^\top\right)^{-1}\right|\right)+\left(Z-\Psi b\right)^\top\left(C\left(\theta\right)+I\sigma^2+\Psi B\Psi^\top\right)^{-1}\left(Z-\Psi b\right)$$

it is

$$\left(C\left(\theta\right)+I\sigma^2+\Psi B\Psi^\top\right)^{-1}=$$

$$\left(C\left(\theta\right)+I\sigma^2\right)^{-1}+\left(C\left(\theta\right)+I\sigma^2\right)^{-1}\Psi\left(B^{-1}+\Psi^\top\left(C\left(\theta\right)+I\sigma^2\right)^{-1}\Psi\right)\Psi^\top\left(C\left(\theta\right)+I\sigma^2\right)^{-1}\to$$

$$C\left(\theta\right)^{-1}+C\left(\theta\right)^{-1}\Psi\left(\Psi^\top C\left(\theta\right)^{-1}\Psi\right)\Psi^\top C\left(\theta\right)^{-1}$$

$$Z-\Psi b\to Z$$

hence the result

$$\hat{\theta}=\arg\min_{\theta}\left(-\frac{n}{4}\log\left(\left|C\left(\theta\right)^{-1}-C^{-1}\Psi\left(\Psi^\top C\left(\theta\right)^{-1}\Psi\right)^{-1}\Psi^\top C\left(\theta\right)^{-1}\right|\right)\right.$$
$$\left.+Z^\top\left[C\left(\theta\right)^{-1}-C\left(\theta\right)^{-1}\Psi\left(\Psi^\top C\left(\theta\right)^{-1}\Psi\right)^{-1}\Psi^\top C\left(\theta\right)^{-1}\right]Z\right)$$

Exercise 20. ($\star$) Assume that $\left(Z\left(s\right):s\in\mathcal{S}\right)$ is a a random field with semivariogram $\gamma\left(\cdot\right)$.

(1) Then

$$\mathrm{Var}\left(\int Z\left(s+t\right)g\left(t\right)\right)=-\frac{1}{2}\int g\left(s\right)g\left(t\right)\gamma\left(s-t\right)\mathrm{d}s\mathrm{d}t$$

for any $g\left(t\right)$ be an integrable function of t such as

$$\int g\left(t\right)\mathrm{d}t=0$$

(2) Semivariogram $\gamma\left(\cdot\right)$ is conditionally negative definite (c.n.d.): if for all $n\in\mathbb{N}$, for all function g over $\mathcal{S}$ such that $\int g\left(t\right)\mathrm{d}t=0$

$$\int g\left(s\right)g\left(t\right)\gamma\left(s-t\right)\mathrm{d}s\mathrm{d}t\leq 0,$$

Solution.

(1) It is

$$
\begin{aligned}
\mathrm{Var}\left(\int Z(s+t)\,g(t)\,\mathrm{d}t\right) =& \mathrm{Var}\left(\int Z(s+t)\,g(t)\,\mathrm{d}t - Z(0)\cancelto{0}{\int g(t)\,\mathrm{d}t}\right) \\
=& \mathrm{Var}\left(\int \left[Z(s+t) - Z(s_0)\right] g(t)\,\mathrm{d}t\right) \\
=& \mathrm{E}\left(\left[\int \left[Z(s+t) - Z(s_0)\right] g(t)\,\mathrm{d}t\right]\left[\int \left[Z(s+t') - Z(s_0)\right] g(t')\,\mathrm{d}t'\right]\right) \\
=& \int\int \mathrm{E}\left(\left[Z(s+t) - Z(s_0)\right]\left[Z(s+t') - Z(s_0)\right]\right) g(t)\,g(t')\,\mathrm{d}t\mathrm{d}t' \\
=& \frac{1}{2}\int\int \left[\gamma(s+t-s_0) + \gamma(s+t-s_0) - \gamma(t-t')\right] g(t)\,g(t')\,\mathrm{d}t\mathrm{d}t' \\
=& 0 + 0 - \frac{1}{2}\int\int \gamma(t-t')\,g(t)\,g(t')\,\mathrm{d}t\mathrm{d}t'
\end{aligned}
$$

(2) It is

$$
\int g(s)\,g(t)\,\gamma(s-t)\,\mathrm{d}s\mathrm{d}t = -\,2\mathrm{Var}\left(\int Z(s+t)\,g(t)\right) \leq 0
$$

Exercise 21. ($\star$) Let $(Z(s) : s \in \mathcal{S})$ be a random field with mean $\mu(s)$ at $s \in \mathcal{S}$, and covariance function $c(s, s')$ at $s, s' \in \mathcal{S}$.. Let $f(t)$ be an integrable function of t such as $\int |f(t)|\,\mathrm{d}t < \infty$. Let regularized random field $(Z_f(s) : s \in \mathcal{S})$ defined as

$$
Z_f(s) = \int Z(s+t)\,f(t)\,\mathrm{d}t
$$

(1) Show that the mean of $Z_f(s)$ is

$$
\begin{aligned}
\mu_f(s) =& \mathrm{E}\left(Z_f(s)\right) \\
=& \int \mu(s+t)\,f(t)\,\mathrm{d}t
\end{aligned}
$$

(2) Show that the covariance function of of $Z_f(s)$ is

$$
\begin{aligned}
c_f\left(s, s'\right) =& \mathrm{Cov}\left(Z_f(s), Z_f\left(s'\right)\right) \\
=& \int\int c\left(s+t, s'+t'\right) f(t)\,f\left(t'\right)\mathrm{d}t\mathrm{d}t
\end{aligned}
$$

(3) Assume that $(Z(s) : s \in \mathcal{S})$ is a a random field with semivariogram $\gamma(h)$. Show that for any $g(t)$ be an integrable function of t such as

$$
\int g(t)\,\mathrm{d}t = 0
$$

it is

$$
\mathrm{Var}\left(\int Z(s+t)\,g(t)\right) = -\frac{1}{2}\int g(s)\,g(t)\,\gamma(s-t)\,\mathrm{d}s\mathrm{d}t
$$

(4) Assume that $(Z(s) : s \in \mathcal{S})$ is a a random field semivariogram with semivariogram $\gamma(h)$. Show that

$$\begin{aligned}
\gamma_f(h) =& \frac{1}{2}\mathrm{Var}\left(Z_f(h) - Z_f(0)\right) \\
=& \int\int f(s) f(t-h) \gamma(s-t) \,\mathrm{d}s\mathrm{d}t - \int\int f(s) f(t) \gamma(s-t) \,\mathrm{d}s\mathrm{d}t
\end{aligned}$$

Hint:: Assume that $(Z(s) : s \in \mathcal{S})$ is a a random field with semivariogram $\gamma(h)$. Then

$$\mathrm{Var}\left(\int Z(s+t) g(t)\right) = -\frac{1}{2}\int g(s) g(t) \gamma(s-t) \,\mathrm{d}s\mathrm{d}t$$

for any $g(t)$ be an integrable function of t such as

$$\int g(t) \,\mathrm{d}t = 0$$

Solution.

(1) Straightforward by exchanging the integrals
(2) Straightforward by exchanging the integrals
(3) It is

$$\begin{aligned}
\gamma_f(h) =& \frac{1}{2}\mathrm{Var}\left(Z_f(h) - Z_f(0)\right) = \frac{1}{2}\mathrm{Var}\left(\int Z(h+t) f(t) \,\mathrm{d}t - \int Z(0+t) f(t) \,\mathrm{d}t\right) \\
=& \frac{1}{2}\mathrm{Var}\left(\int Z(t) \left[f(t-h) - f(t)\right] \mathrm{d}t\right) = \frac{1}{2}\mathrm{Var}\left(\int Z(t) \left[\overbrace{f(t-h) - f(t)}^{g(t)=}\right] \mathrm{d}t\right) \\
=& -\frac{1}{2}\int\int g(t) g(t) \gamma(s-t) \,\mathrm{d}s\mathrm{d}t \\
=& -\frac{1}{2}\int\int \left[f(t-h) - f(t)\right]\left[f(s-h) - f(s)\right] \gamma(s-t) \,\mathrm{d}s\mathrm{d}t \\
=& \int\int f(s) f(t-h) \gamma(s-t) \,\mathrm{d}s\mathrm{d}t - \int\int f(s) f(t) \gamma(s-t) \,\mathrm{d}s\mathrm{d}t
\end{aligned}$$

as $\int g(t) \,\mathrm{d}t = 0$.

Exercise 22. $(\star)$ Show that the extension variance $\sigma_E^2(v, V)$ of a small volume v to a larger volume V is obtained by

$$\sigma_E^2(v, V) = 2\bar{\gamma}(v, V) - \bar{\gamma}(v, v) - \bar{\gamma}(V, V)$$

where

$$\bar{\gamma}(v, V) = \frac{1}{|v||V|}\int_{s\in v}\int_{s\in V} \gamma\left(s - s'\right) \mathrm{d}s\mathrm{d}s'$$

Solution. Essentially I need to show that that

$$\begin{aligned}\mathrm{Var}\left(Z\left(A\right)-Z\left(B\right)\right)=&\frac{1}{\left|A\right|\left|B\right|}\int_{x\in A}\int_{y\in B}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y\\&-\frac{1}{\left|A\right|\left|A\right|}\int_{x\in A}\int_{y\in A}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y\\&-\frac{1}{\left|B\right|\left|B\right|}\int_{x\in B}\int_{y\in B}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y\end{aligned}$$

where I use A, B instead of v, V and x, y instead of s, s' for clarity on notation.

It is

$$\begin{aligned}\mathrm{Var}\left(Z\left(A\right)-Z\left(B\right)\right)=&\mathrm{Cov}\left(Z\left(A\right)-Z\left(B\right),Z\left(A\right)-Z\left(B\right)\right)\\=&\mathrm{Cov}\left(Z\left(A\right),Z\left(A\right)\right)+\mathrm{Cov}\left(Z\left(B\right),Z\left(B\right)\right)-2\mathrm{Cov}\left(Z\left(A\right),Z\left(B\right)\right)\\=&\frac{1}{\left|A\right|\left|A\right|}\int_{x\in A}\int_{y\in A}\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y\\&+\frac{1}{\left|B\right|\left|B\right|}\int_{x\in B}\int_{y\in B}\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y\\&-2\frac{1}{\left|A\right|\left|B\right|}\int_{x\in A}\int_{y\in B}\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y\end{aligned}$$

OK, now I need to write all these Cov as γ ; I know that

$$\begin{aligned}\gamma\left(x-y\right)&=\frac{1}{2}\mathrm{Var}\left(Z\left(x\right)-Z\left(y\right)\right)\\&=\frac{1}{2}\mathrm{Var}\left(Z\left(x\right)\right)+\frac{1}{2}\mathrm{Var}\left(Z\left(y\right)\right)-\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\end{aligned}$$

that is

$$\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)=\frac{1}{2}\mathrm{Var}\left(Z\left(x\right)\right)+\frac{1}{2}\mathrm{Var}\left(Z\left(y\right)\right)-\gamma\left(x-y\right)$$

Now I 'll gonna put all these in the quantity of interest, one by one

$$\begin{aligned}\frac{1}{\left|A\right|\left|A\right|}\int_{x\in A}\int_{y\in A}\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y=&\frac{1}{\left|A\right|\left|A\right|}\int_{x\in A}\int_{y\in A}\frac{1}{2}\mathrm{Var}\left(Z\left(x\right)\right)\mathrm{d}x\mathrm{d}y\\&+\frac{1}{\left|A\right|\left|A\right|}\int_{x\in A}\int_{y\in A}\frac{1}{2}\mathrm{Var}\left(Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y\\&-\frac{1}{\left|A\right|\left|A\right|}\int_{x\in A}\int_{y\in A}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y\\=&\frac{1}{\left|A\right|}\int_{x\in A}\mathrm{Var}\left(Z\left(x\right)\right)\mathrm{d}x\\&-\frac{1}{\left|A\right|^2}\int_{x\in A}\int_{y\in A}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y\end{aligned}$$

and by symmetry

$$\frac{1}{|B|\,|B|}\int_{x\in B}\int_{y\in B}\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y = \frac{1}{|B|}\int_{x\in B}\mathrm{Var}\left(Z\left(x\right)\right)\mathrm{d}x - \frac{1}{|B|^2}\int_{x\in B}\int_{y\in B}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y$$

and finally,

$$\begin{aligned}\frac{1}{|A|\,|B|}\int_{x\in A}\int_{y\in B}\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y =& \frac{1}{|A|\,|B|}\int_{x\in A}\int_{y\in B}\frac{1}{2}\mathrm{Var}\left(Z\left(x\right)\right)\mathrm{d}x\mathrm{d}y\\ &+ \frac{1}{|A|\,|B|}\int_{x\in A}\int_{y\in B}\frac{1}{2}\mathrm{Var}\left(Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y\\ &- \frac{1}{|A|\,|B|}\int_{x\in A}\int_{y\in B}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y\\ =& \frac{1}{2}\frac{1}{|A|}\int_{x\in A}\mathrm{Var}\left(Z\left(x\right)\right)\mathrm{d}x\\ &+ \frac{1}{2}\frac{1}{|B|}\int_{x\in B}\mathrm{Var}\left(Z\left(x\right)\right)\mathrm{d}x\\ &- \frac{1}{|A|\,|B|}\int_{x\in A}\int_{y\in B}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y\end{aligned}$$

Putting all these together, we get the result.

Exercise 23. ($\star$) Suppose a large volume V is partitioned into n smaller units v of equal size. Show that the dispersion variance $\sigma^2\left(v|V\right) = \frac{1}{n}\sum_{j=1}^{n}\sigma_E^2\left(v_j,V\right)$ can be written in term of variogram integrals

$$\bar{\gamma}\left(v,V\right) = \frac{1}{|v|\,|V|}\int_{s\in v}\int_{s\in V}\gamma\left(s-s'\right)\mathrm{d}s\mathrm{d}s'$$

as

$$\sigma^2\left(v|V\right) = \bar{\gamma}\left(V,V\right) - \bar{\gamma}\left(v,v\right)$$

Solution.

$$
\begin{aligned}
\sigma^2(v|V) =& \frac{1}{n}\sum_{j=1}^{n}\sigma_E^2(v_j,V)\\
=& \frac{1}{n}\sum_{j=1}^{n}\left[2\bar{\gamma}(v_j,V)-\bar{\gamma}(v_j,v_j)-\bar{\gamma}(V,V)\right]\\
=& \frac{2}{n}\sum_{j=1}^{n}\bar{\gamma}(v_j,V)-\frac{1}{n}\sum_{j=1}^{n}\bar{\gamma}(v_j,v_j)-\frac{1}{n}\sum_{j=1}^{n}\bar{\gamma}(V,V)\\
=& \frac{2}{n}\sum_{j=1}^{n}\frac{1}{|v_j|\,|V|}\int_{s\in v_j}\int_{s\in V}\gamma\left(s-s'\right)\mathrm{d}s\mathrm{d}s'\\
&-\frac{1}{n}\sum_{j=1}^{n}\bar{\gamma}(v_j,v_j)-\frac{1}{n}\sum_{j=1}^{n}\bar{\gamma}(V,V)\ \text{(but all } v_j \text{ are of the same size as } v)\\
=& 2\frac{1}{n\,|v|\,|V|}\sum_{j=1}^{n}\int_{s\in v_j}\int_{s\in V}\gamma\left(s-s'\right)\mathrm{d}s\mathrm{d}s'-\bar{\gamma}(v,v)-\bar{\gamma}(V,V)\\
=& 2\frac{1}{\underbrace{n\,|v|}_{=|V|}\,|V|}\underbrace{\sum_{j=1}^{n}\int_{s\in v_j}}_{\int_{s\in V}}\int_{s\in V}\gamma\left(s-s'\right)\mathrm{d}s\mathrm{d}s'-\bar{\gamma}(v,v)-\bar{\gamma}(V,V)\\
=& 2\frac{1}{n}\sum_{j=1}^{n}\bar{\gamma}(V,V)-\bar{\gamma}(v,v)-\bar{\gamma}(V,V)=\quad\bar{\gamma}(V,V)-\bar{\gamma}(v,v)
\end{aligned}
$$

Exercise 24. ($\star$) Consider a statistical model which is a stochastic process $(Z_s)_{s\in\mathbb{R}}$ (so s has dimension 1), where $Z(\cdot)\sim\mathrm{GP}(\mu(\cdot),c(\cdot,\cdot))$ with mean function $\mu(s)=1$ and covariance function $c(s,t)=\exp\left(-(s-t)^2\right)$ for any $s\in\mathbb{R}$ and $t\in\mathbb{R}$. Assume there is available a dataset $\{(Z_i,s_i)\}_{i=1}^{n}$ where $Z_i=Z(s_i)$ and $s_i\in\mathbb{R}$ are point sites.

(1) Compute the length $|v|$ of the block $v=[a,b]\subset\mathbb{R}$.
(2) Compute the block mean $\mu(v)$ for some block $v=[a,b]\subset\mathbb{R}$ and point $s\in\mathbb{R}$.
(3) Compute the block covariance function $c(v,s)$ for some block $v=[a,b]\subset\mathbb{R}$ and point $s\in\mathbb{R}$.
(4) Compute the block covariance function $c(v,v')$ for some blocks $v=[a,b]\subset\mathbb{R}$ and $v'=[a',b']\subset\mathbb{R}$.
(5) Denote $Z=(Z_1,...,Z_n)^\top$, and $S=\{s_1,...,s_n\}$. Let $v=[a,b]\subset\mathbb{R}$ and $v'=[a',b']\subset\mathbb{R}$ be two intervals. Compute the joint distribution of $(Z(v),Z(v'),Z)^\top$ as a function of $c(\cdot,\cdot)$, S, v, v', Z, and $\mu(\cdot)$. What is the name of the distribution and what are the parameter functions defining it?
(6) (Bayesian Kriging) Compute the predictive stochastic process $[Z(v)\,|Z]$ at blocks $v=[a,b]\subset\mathbb{R}$ with $|v|>0$.

Hint-1:: Let $x_1 \in \mathbb{R}^{d_1}$, and $x_2 \in \mathbb{R}^{d_2}$. If

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathrm{N}_{d_1+d_2}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_1 & \Sigma_{21}^\top \\ \Sigma_{21} & \Sigma_2 \end{bmatrix}\right)$$

then it is

$$x_2|x_1 \sim \mathrm{N}_{d_2}\left(\mu_{2|1}, \Sigma_{2|1}\right)$$

where

$$\mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_1^{-1}\left(x_1 - \mu_1\right) \text{ and } \Sigma_{2|1} = \Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{21}^\top$$

Hint-2: You can use that $\int \operatorname{erf}(x)\,\mathrm{d}x = x\operatorname{erf}(x) + \frac{\exp(-x^2)}{\sqrt{\pi}} + \text{const}$, when $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp\left(-t^2\right)\mathrm{d}t$.

Solution.

(Given as Formative assessment 3)

Exercise 25. ($\star$) Assume we wish to estimate the average value in a domain V

$$Z_V = \frac{1}{|V|}\int_V Z(s)\,\mathrm{d}s$$

with the average of n sample points $\{s_i; i = 1, ..., n\}$.

$$\hat{Z} = \frac{1}{n}\sum_{i=1}^{n} Z(s_i)$$

Show that the estimation variance(or else extension variance)

$$\operatorname{Var}\left(\hat{Z} - Z_V\right) = -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(s_i - s_j) + \frac{1}{n|V|}\sum_{i=1}^{n}\int_V \gamma(s_i - x)\,\mathrm{d}x - \frac{1}{|V|^2}\int_{x\in B}\int_{y\in B}\gamma(x-y)\,\mathrm{d}x\mathrm{d}y$$

Hint:: Consider as known that

$$\operatorname{Cov}\left(Z(t) - Z(s), Z(v) - Z(u)\right) = \gamma(t-u) + \gamma(s-v) - \gamma(s-u) - \gamma(t-v)$$

Solution.

Exercise 26. ($\star$) We are interested in the mean value of some parameter $Z(s)$, say temperature, over a one-hour period segment $[0, L]$. In particular we wish to find out which of these two estimates is better: the temperature at half the hour

$$\hat{Z}_1 = Z\left(\frac{L}{2}\right)$$

or

$$\hat{Z}_2 = \frac{1}{2}\left(Z(0) + Z(L)\right)$$

the average of two consecutive measurements on the hour?

Assume that $(Z(s) : s \in \mathcal{S})$ is intrinsic random field with semivariogram $\gamma(\cdot)$. Let

$$\chi(h) = \frac{1}{h}\int_0^h \gamma(u)\,\mathrm{d}u,\ h > 0$$

$$\xi(h) = \frac{1}{h^2}\int_0^h \int_0^h \gamma(x-y)\,\mathrm{d}x\mathrm{d}y,\ h > 0$$

(1) Show that the estimation variances (or else extension variances) of $\hat{Z}_1$ and $\hat{Z}_2$ to domain segment $[0, L]$ are

$$\sigma_1^2 = \mathrm{Var}\left(\hat{Z}_1 - Z_V\right) = 2\chi\left(\frac{L}{2}\right) - \xi(L)$$

$$\sigma_2^2 = \mathrm{Var}\left(\hat{Z}_2 - Z_V\right) = 2\chi(L) - \xi(L) - \frac{\gamma(L)}{2}$$

(2) Assume semivariogram $\gamma(h) = b|h|^a$. Show that

$$\chi(h) = \frac{bh^a}{a+1}, \quad \xi(h) = \frac{2bh^a}{(a+1)(a+2)}$$

(3) Assume the power semivariogram $\gamma(h) = |h|^a$. Show that

$$\sigma_1^2 = \frac{2L^a}{(a+1)(a+2)}\frac{1}{2^a}(a+2-2^a)$$

$$\sigma_2^2 = \frac{2L^a}{(a+1)(a+2)}\frac{1}{4}\left(2+a-a^2\right)$$

Which estimate is preferable for $a = 0$, and $a = 1$,

Hint: If $Z_V = \frac{1}{|V|}\int_V Z(s)\,\mathrm{d}s$ is average value in a domain V and if $\hat{Z} = \frac{1}{n}\sum_{i=1}^n Z(s_i)$ is average of n sample points $\{s_i; i = 1, ..., n\}$ then the the estimation variance(or else extension variance) is

$$\mathrm{Var}\left(\hat{Z} - Z_V\right) = -\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \gamma(s_i - s_j) + \frac{1}{n|V|}\sum_{i=1}^n \int_V \gamma(s_i - x)\,\mathrm{d}x - \frac{1}{|V|^2}\int_{x\in B}\int_{y\in B}\gamma(x-y)\,\mathrm{d}x\mathrm{d}y$$

Solution.

(1) By using the provided formulas, I get

$$\begin{aligned}
\sigma_1^2 =& \mathrm{Var}\left(\hat{Z}_1 - Z_{[0,L]}\right) \\
=& -\cancelto{=0}{\gamma\left(\frac{L}{2} - \frac{L}{2}\right)} + \frac{1}{L}\int_0^L \gamma\left(\frac{L}{2} - x\right)\mathrm{d}x - \frac{1}{L^2}\int_0^L\int_0^L \gamma(x-y)\,\mathrm{d}x\mathrm{d}y \\
=& 2\chi\left(\frac{L}{2}\right) - \xi(L)
\end{aligned}$$

Also

$$
\begin{aligned}
\sigma_1^2 =& \operatorname{Var}\left(\hat{Z}_1 - Z_{[0,L]}\right) \\
=& -\frac{1}{2^2} 2\gamma(L) + \frac{1}{2L}\left[\int_0^L \gamma(L-x)\,\mathrm{d}x + \int_0^L \gamma(-x)\,\mathrm{d}x\right] - \frac{1}{L^2}\int_0^L\int_0^L \gamma(x-y)\,\mathrm{d}x\mathrm{d}y \\
=& -\frac{\gamma(L)}{2} + 2\chi(L) - \xi(L)
\end{aligned}
$$

(2) It is

$$
\begin{aligned}
\chi(L) =& \frac{1}{h}\int_0^h \gamma(u)\,\mathrm{d}u = \frac{1}{L}\int_0^L b\,|u|^a\,\mathrm{d}u \\
=& \frac{bh^a}{a+1}
\end{aligned}
$$

and

$$
\begin{aligned}
\xi(h) =& \frac{1}{L^2}\int_0^L\int_0^L \gamma(x-y)\,\mathrm{d}x\mathrm{d}y \\
=& \frac{2}{h^2}\int_0^h u\chi(u)\,\mathrm{d}u \\
=& \frac{2bh^a}{(a+1)(a+2)}
\end{aligned}
$$

(3) For $b = 1$, it is

$$
\begin{aligned}
\sigma_1^2 &= 2\chi\left(\frac{L}{2}\right) - \xi(L) \\
&= 2\frac{\left(\frac{L}{2}\right)^a}{a+1} - \frac{2L^a}{(a+1)(a+2)} \\
&= \frac{2L^a}{(a+1)(a+2)}\frac{1}{2^a}\left(a+2-2^a\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\sigma_2^2 &= 2\chi(L) - \xi(L) - \frac{\gamma(L)}{2} \\
&= 2\frac{(L)^a}{a+1} - \frac{2L^a}{(a+1)(a+2)} - \frac{|L|^a}{2} \\
&= \frac{2L^a}{(a+1)(a+2)}\frac{1}{4}\left(2+a-a^2\right)
\end{aligned}
$$

- For $a = 0$, $\hat{Z}_2$ is more preferable because $\sigma_1^2(a=0) = 1$ and $\sigma_2^2(a=0) = 0.5$. Perhaps this is because it is the limiting case of a pure nugget effect, and only the number of samples matters.
- For $a = 1$, $\hat{Z}_1$ and $\hat{Z}_2$ are equivalent as $\sigma_1^2(a=1) = 0.167L$ and $\sigma_2^2(a=1) = 0.167L$.

Exercise 27. ($\star$) Consider

$$Z(s) = \sum_{p=1}^{k} a_p w_p(s)$$

where $\left\{ w_p^{(u)}(s) \right\}$ are intrinsic random fields with

$$\begin{aligned} \mathrm{E}(w_p(s)) &= 0 \\ \gamma_{p,q}(h) &= 0, \text{ for } p \neq q \end{aligned}$$

$p = 1, ..., k$ and $q = 1, ..., k$. Assume $\{a_p\}$ is a set of known constants. Let $\gamma(h)$ be the variogram function of $Z(s)$. Show that

$$\gamma(h) = \sum_{p=1}^{k} a_p^2 \gamma_{p,p}(h)$$

Solution. From

$$\begin{aligned} \mathrm{E}(w_p(s)) &= 0 \\ \gamma_{p,q}(h) &= 0, \text{ for } p \neq q \end{aligned}$$

I can see that

$$\mathrm{E}\left((w_p(s+h) - w_p(s))(w_q(s+h) - w_q(s))\right) = \begin{cases} \gamma_{p,q}(h), & p = q \\ 0 & p \neq q \end{cases}$$

It is

$$\begin{aligned} \gamma(h) =& \mathrm{E}\left((Z(s+h) - Z(s))^2 \right) \\ =& \mathrm{E}\left(\left(\sum_{p=1}^{k} a_p w_p(s+h) - \sum_{p=1}^{k} a_p w_p(s) \right)^2 \right) \\ \text{(because of the above)} =& \mathrm{E}\left(\left(\sum_{p=1}^{k} a_p (w_p(s+h) - w_p(s)) \right)^2 \right) \\ =& \sum_{p=1}^{k} a_p^2 \mathrm{E}\left((w_p(s+h) - w_p(s))^2 \right) \\ =& \sum_{p=1}^{k} a_p^2 \gamma_{p,p}(h) \end{aligned}$$

Exercise 28. ($\star$) Consider a set of random fields $\left\{ \left(Z_j^{(u)}(s) : s \in \mathcal{S} \right); j = 1, ..., k; u = 1, ..., k \right\}$ with

$$Z_j^{(u)}(s) = \sum_{p=1}^{k} a_{j,p}^{(u)} w_p^{(u)}(s),$$

where $\left\{w_p^{(u)}(s)\right\}$ are intrinsic random fields and $\left\{a_{j,p}^{(u)}\right\}$ are known constants. Let $\tilde{\gamma}_{i,j}^{(u)}(h)$ be the cross variogram function of $Z_i^{(u)}(s)$ and $Z_j^{(u)}(s)$ for $u = 1, ..., k$.

(1) Write the definition of the the cross variogram function $\tilde{\gamma}_{i,j}^{(u)}(h)$ of $Z_i^{(u)}(s)$ and $Z_j^{(u)}(s)$ for $u = 1, ..., k$

(2) Assume that

$$\mathrm{E}\left(w_p^{(u)}(s)\right) = 0$$

$$\mathrm{Cov}\left(w_p^{(u)}(s), w_q^{(v)}(s+h)\right) = \begin{cases} \gamma_{p,q}^{(u)}(h), & u = v \\ 0 & u \neq v \end{cases}$$

$u = 1, ..., k$, $p = 1, ..., k$ and $q = 1, ..., k$. Show that

$$\tilde{\gamma}_{i,j}^{(u)}(h) = \sum_{p=1}^{k} a_{i,p}^{(u)} \sum_{q=1}^{k} a_{j,q}^{(u)} \gamma_{p,q}^{(u)}(h)$$

(3) Assume that

$$\mathrm{E}\left(w_p^{(u)}(s)\right) = 0$$

$$\mathrm{Cov}\left(w_p^{(u)}(s), w_q^{(v)}(s+h)\right) = \begin{cases} \gamma^{(u)}(h), & u = v \,\mathrm{and}\, p = q \\ 0 & u \neq v \,\mathrm{or}\, p \neq q \end{cases}$$

$u = 1, ..., k$. Show that

$$\tilde{\gamma}_{i,j}^{(u)}(h) = \sum_{p=1}^{k} a_{i,p}^{(u)} \sum_{q=1}^{k} a_{j,q}^{(u)} \gamma^{(u)}(h)$$

and hence

$$\tilde{\Gamma}^{(u)}(h) = B^{(u)} \gamma^{(u)}(h)$$

where $B^{(u)} = A^{(u)} \left(A^{(u)}\right)^\top$, $u = 1, ..., k$.

(4) Consider the assumptions in previous part. Let $\{(Z_j(s) : s \in \mathcal{S}) ; j = 1, ..., k\}$ be a set of random fields on $s \in \mathcal{S}$. Let

$$Z_j(s) = \mu_j(s) + \sum_{u=0}^{m} Z_j^{(u)}(s)$$

and let

$$Z(s) = \mu(s) + \sum_{u=0}^{m} A^{(u)} w^{(u)}(s)$$

Show that the cross variogram matrix of $(Z(s) ; s \in \mathcal{S})$ is

$$\Gamma(h) = \sum_{u=0}^{m} B^{(u)} \gamma^{(u)}(h)$$

where $B^{(u)} = A^{(u)} \left(A^{(u)}\right)^\top$, $u = 1, ..., k$

Solution.

(1) It is

$$\tilde{\gamma}_{i,j}^{(u)}(h) = \frac{1}{2}\mathrm{Cov}\left(\left(Z_i^{(u)}(s+h) - Z_i^{(u)}(s)\right), \left(Z_j^{(u)}(s+h) - Z_j^{(u)}(s)\right)\right)$$

(2) From

$$\mathrm{E}\left(w_p^{(u)}(s)\right) = 0$$

$$\mathrm{Cov}\left(w_p^{(u)}(s), w_q^{(v)}(s+h)\right) = \begin{cases} \gamma_{p,q}^{(u)}(h), & u = v \\ 0 & u \neq v \end{cases}$$

I can see that

$$\mathrm{E}\left(\left(w_p^{(u)}(s+h) - w_p^{(u)}(s)\right)\left(w_q^{(v)}(s+h) - w_q^{(v)}(s)\right)\right) = \begin{cases} \gamma_{p,q}^{(u)}(h), & u = v \\ 0 & u \neq v \end{cases}$$

so

$$\begin{aligned}
\tilde{\gamma}_{i,j}^{(u)}(h) &= \frac{1}{2}\mathrm{Cov}\left(Z_i^{(u)}(s+h) - Z_i^{(u)}(s), Z_j^{(u)}(s+h) - Z_j^{(u)}(s)\right) \\
&= \frac{1}{2}\mathrm{Cov}\left(\sum_{q=1}^{k} a_{i,q}^{(u)}\left[w_q^{(u)}(s+h) - w_q^{(u)}(s)\right], \sum_{p=1}^{k} a_{j,p}^{(u)}\left[w_p^{(u)}(s+h) - w_p^{(u)}(s)\right]\right) \\
&= \frac{1}{2}\sum_{q=1}^{k} a_{i,q}^{(u)} \sum_{p=1}^{k} a_{j,p}^{(u)} \mathrm{Cov}\left(w_q^{(u)}(s+h) - w_q^{(u)}(s), w_p^{(u)}(s+h) - w_p^{(u)}(s)\right) \\
&= \sum_{q=1}^{k} a_{i,q}^{(u)} \sum_{p=1}^{k} a_{j,p}^{(u)} \gamma_{p,q}^{(u)}(h)
\end{aligned}$$

(3) Then it is

$$\tilde{\gamma}_{i,j}^{(u)}(h) = \sum_{p=1}^{k} a_{i,p}^{(u)} a_{j,p}^{(u)} \gamma^{(u)}(h)$$

The given matrix form is such that $\left[A^{(u)}\right]_{i,p} = a_{i,p}^{(u)}$.

(4) It is

$$\begin{aligned}
\Gamma(h) &= \frac{1}{2}\mathrm{Cov}\left(Z(s+h) - Z(s), Z(s+h) - Z(s)\right) \\
&= \frac{1}{2}\mathrm{Cov}\left(\sum_{u=0}^{m} A^{(u)}\left(w^{(u)}(s+h) - w^{(u)}(s)\right), \sum_{v=0}^{m} A^{(v)}\left(w^{(v)}(s+h) - w^{(v)}(s)\right)\right) \\
&= \frac{1}{2}\sum_{u=0}^{m}\sum_{v=0}^{m} A^{(u)} \mathrm{Cov}\left(w^{(u)}(s+h) - w^{(u)}(s), w^{(v)}(s+h) - w^{(v)}(s)\right)\left(A^{(v)}\right)^{\top} \\
&= \frac{1}{2}\sum_{u=0}^{m}\sum_{v=0}^{m} A^{(u)} \gamma_{p,q}^{(u)}(h)\left(A^{(v)}\right)^{\top} = \frac{1}{2}\sum_{u=0}^{m} A^{(u)}\left(A^{(v)}\right)^{\top}\gamma(h) = \frac{1}{2}\sum_{u=0}^{m} B^{(u)}\gamma^{(u)}(h)
\end{aligned}$$

Part 3. **Aerial unit data / spatial data on lattices**

Exercise 29. ($\star$) Show that the conditionals $x|y \sim \mathrm{N}\left(a+by, \sigma^2+\tau^2y^2\right)$ and $y|x \sim \mathrm{N}\left(c+dx, \tilde{\sigma}^2+\tilde{\tau}^2x^2\right)$ are compatible if $\tau^2 = \tilde{\tau}^2 = 0$, $d/\tilde{\sigma}^2 = b/\sigma^2$, and $|db| < 1$. In particular see what happens if $x|y \sim \mathrm{N}\left(y, \sigma^2\right)$ and $y|x \sim \mathrm{N}\left(x, \sigma^2\right)$ namely if $\tau^2 = \tilde{\tau}^2 = 0$, $d/\tilde{\sigma}^2 = b/\sigma^2$, $\tilde{\sigma}^2 = \sigma^2$ and $d = b = 1$.

Solution. It is

$$
\begin{aligned}
\frac{g\left(x|y\right)}{q\left(y|x\right)} =& \frac{\mathrm{N}\left(x|a+by, \sigma^2+\tau^2y^2\right)}{\mathrm{N}\left(y|c+dx, \tilde{\sigma}^2+\tilde{\tau}^2x^2\right)} \\
=& \frac{\sqrt{\tilde{\sigma}^2+\tilde{\tau}^2x^2}}{\sqrt{\sigma^2+\tau^2y^2}} \exp\left(-\frac{1}{2}\left(\frac{\left(x-a-by\right)^2}{\sigma^2+\tau^2y^2} - \frac{\left(y-c-dx\right)^2}{\tilde{\sigma}^2+\tilde{\tau}^2x^2}\right)\right) \text{ (set } \tau^2=\tilde{\tau}^2=0\text{)} \\
=& \frac{\sqrt{\tilde{\sigma}^2}}{\sqrt{\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{\left(x-a-by\right)^2}{\sigma^2} - \frac{\left(y-c-dx\right)^2}{\tilde{\sigma}^2}\right)\right) \\
=& \frac{\sqrt{\tilde{\sigma}^2}}{\sqrt{\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma^2} + \frac{a^2}{\sigma^2} + \frac{b^2y^2}{\sigma^2} - \frac{2xa}{\sigma^2} - 2\frac{xby}{\sigma^2} + \frac{2aby}{\tilde{\sigma}^2}\right.\right. \\
& \left.\left. - \frac{y^2}{\tilde{\sigma}^2} - \frac{c^2}{\tilde{\sigma}^2} - \frac{d^2x^2}{\tilde{\sigma}^2} + \frac{2yc}{\tilde{\sigma}^2} + 2\frac{ydx}{\tilde{\sigma}^2} - \frac{2cdx}{\tilde{\sigma}^2}\right)\right)
\end{aligned}
$$

If $d/\tilde{\sigma}^2 = b/\sigma^2$ (and $\tau^2 = \tilde{\tau}^2 = 0$)

$$
\begin{aligned}
\frac{g\left(x|y\right)}{q\left(y|x\right)} \propto & \underbrace{\exp\left(-\frac{1}{2}\left(\left(\frac{1}{\sigma^2} - \frac{d^2}{\tilde{\sigma}^2}\right)x^2 - 2\left(\frac{a}{\sigma^2} + \frac{cd}{\tilde{\sigma}^2}\right)x\right)\right)}_{u(x)} \\
& \times \underbrace{\exp\left(+\frac{1}{2}\left(\left(\frac{1}{\tilde{\sigma}^2} - \frac{b^2}{\sigma^2}\right)y^2 - 2\left(\frac{c}{\tilde{\sigma}^2} + \frac{ab}{\sigma^2}\right)x\right)\right)}_{v(y)}
\end{aligned}
$$

for $N_g = N_q = N = \mathbb{R}$. Also it is

$$
\begin{aligned}
& \int u\left(x\right)\mathrm{d}x = \\
\int_{\mathbb{R}} \exp & \left(-\frac{1}{2}\left(\left(\frac{1}{\sigma^2} - \frac{d^2}{\tilde{\sigma}^2}\right)x^2 - 2\left(\frac{a}{\sigma^2} + \frac{cd}{\tilde{\sigma}^2}\right)x\right)\right)\mathrm{d}x < \infty
\end{aligned}
$$

when $|db| < 1$.

If $\tau^2 = \tilde{\tau}^2 = 0$, $d/\tilde{\sigma}^2 = b/\sigma^2$, $\tilde{\sigma}^2 = \sigma^2$ and $d = b = 1$, then

$$
\frac{g\left(x|y\right)}{q\left(y|x\right)} \propto \exp\left(-\frac{1}{2}\left(\left(\frac{1}{\sigma^2} - \frac{1}{\sigma^2}\right)x^2\right)\right) \exp\left(-\frac{1}{2}\left(\left(\frac{1}{\sigma^2} - \frac{1}{\sigma^2}\right)y^2\right)\right) \propto \text{const}
$$

that is $u\left(x\right)$ is constant and hence $\int u\left(x\right)\mathrm{d}x = \infty$ implying that they are not compatible.

Exercise 30. ($\star$) Consider the hard core lattice gas Z on a finite grid $\emptyset \neq \mathcal{S} \subset \mathbb{Z}^2$ with value set $\mathcal{Z} = \{0, 1\}$. Write $i \sim j$ whenever $0 < \|i - j\| \leq 1$ so that sites i and j are neighbours when they are horizontally or vertically adjacent. The probability mass function is, for $z \in \mathcal{Z}^S$, defined by

$$\Pr_Z(z) = \begin{cases} \frac{1}{C} \prod_{i \in S} \alpha^{z_i}, & \text{if } z_i z_j = 0 \text{ whenever } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

with C normalizing constant and $\alpha > 0$.

(1) Compute the local characteristics

(2) Order the sites in S lexicographically. Show that there exist $x = (x_i; i \in S) \in \mathcal{Z}^S$ and $y = (y_i; i \in S) \in \mathcal{Z}^S$ and $i' \in \mathcal{S}$ such that

$$\Pr_{i'}\left(y_{i'} | x_{\{j:j<i'\}}, y_{\{j:j>i'\}}\right) = 0$$

but $\Pr_Z(x) > 0$ and $\Pr_Z(y) > 0$.

Solution.

(1) Suppose that $(z_j; j \neq i)$ is feasible in the sense that $z_j z_k = 0$ whenever $j \sim k$. Then

$$\frac{\Pr_i(1|z_j, j \neq i)}{\Pr_i(0|z_j, j \neq i)} = \begin{cases} \alpha & \text{if } x_j = 0 \text{ whenever } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

depends on the neighbours of i only, so Z is Markov with respect to $\sim$.

Furthermore, if $z_j = 0$ for all $j \sim i$ then

$$\Pr_i(1|z_j, j \neq i) = 1 - \Pr_i(0|z_j, j \neq i) = \frac{\alpha}{1+\alpha}$$

If $z_j = 1$ for some $j \sim i$ then

$$\Pr_i(0|z_j, j \neq i) = 1$$

(2) Consider a lattice that consists of two adjacent sites labelled 1 and 2and take $y = (0,1)$, $x = (1,0)$. Then both x and y have positive probability of occurring, but

$$\Pr_2(y_2|x_1) = \Pr_2(1|1) = 0$$

Exercise 31. ($\star\star$) Show that

(1) ... any positive-definite covariance matrix Σ can be expressed as the covariance matrix of a CAR model $\Sigma_{\text{CAR}} = (I-B)^{-1} K$, for a unique pair of matrices B and K where $(I-B)$ is non-singular and K is diagonal.

(2) ... any positive-definite covariance matrix Σ can be expressed as the covariance matrix of a SAR model $\Sigma_{\text{SAR}} = \left(I - \tilde{B}\right)^{-1} \Lambda \left(I - \tilde{B}^\top\right)^{-1}$ for a (non-unique) pair of matrices $\tilde{B}$ and Λ where $\left(I - \tilde{B}\right)$ is non-singular, $\left[\tilde{B}\right]_{i,i} = 0$, and Λ is diagonal.

(3) ... any SAR model can be written as a unique CAR model.

Solution.

(1) Express

$$\Sigma^{-1} = D - R$$

for

$$[D]_{i,j} = \begin{cases} \left[\Sigma^{-1}\right]_{i,i} & i = j \\ 0 & i \neq j \end{cases}, \text{ and } [R]_{i,j} = \begin{cases} 0 & i = j \\ -\left[\Sigma^{-1}\right]_{i,j} & i \neq j \end{cases}$$

then

$$\Sigma = (D - R)^{-1} = \left(D\left(I - D^{-1}R\right)\right)^{-1} = \left(I - D^{-1}R\right)^{-1} D^{-1}$$

Now define $B = D^{-1}R$ and $K = D^{-1}$, and you get $\Sigma = \Sigma_{\text{CAR}}$. Now regarding the uniqueness, assume there is another pair of $\mathring{B}$, and $\mathring{K}$ such that $\Sigma_{\text{CAR}} = \left(I - \mathring{B}\right)^{-1} \mathring{K}$. Then

$$\text{diag}\left(\Sigma^{-1}\right) = \text{diag}\left(\Sigma_{\text{CAR}}^{-1}\right) = \text{diag}\left(\mathring{K}^{-1}\left(I - \mathring{B}\right)\right) = \text{diag}\left(\mathring{K}^{-1}\right)$$

and similarly $\text{diag}\left(\Sigma^{-1}\right) = \text{diag}\left(K^{-1}\right)$. Hence it has to be $\mathring{K} = K$ because both are diagonal matrices. Then it is

$$\left(I - \mathring{B}\right)^{-1} \mathring{K} = (I - B)^{-1} K \overset{\mathring{K}=K}{\Longleftrightarrow} \mathring{B} = B.$$

So the representation is unique.

(2) Express

$$\Sigma^{-1} = LL^{\top}$$

where L is a lower triangular matrix with $[L]_{i,i} > 0$. Such matrix decomposition can be done by Cholesky decomposition, square-matrix decomposition, etc... and hence it is not always unique. Then

$$\Sigma = \left(LL^{\top}\right)^{-1} = L^{-\top}L^{-1}$$

Now express, $L = D - C$ for

$$[D]_{i,j} = \begin{cases} [L]_{i,i} & i = j \\ 0 & i \neq j \end{cases}, \text{ and } [C]_{i,j} = \begin{cases} 0 & i = j \\ -[L]_{i,j} & i \neq j \end{cases}$$

then

$$\begin{aligned} \Sigma &= (D - C)^{-\top}(D - C)^{-1} = \left(I - D^{-1}C\right)^{-\top} D^{-\top}D^{-1}\left(I - D^{-1}C\right)^{-1} \\ &= \left(I - C^{\top}D^{-\top}\right)^{-1} D^{-\top}D^{-1}\left(I - \left(C^{\top}D^{-\top}\right)^{\top}\right)^{-1} \end{aligned}$$

Set $\tilde{B} = C^{\top}D^{-\top}$ and $\Lambda = D^{-\top}D^{-1}$ and you get $\Sigma_{\text{SAR}} = \Sigma$ for non-unique pairs of $\tilde{B}$ and Λ.

(3) SAR and CAR are both Gaussian's with the same mean. SAR's variance matrix is positive definite, and hence it can be written in a unique manner as a CAR's variance matrix

---

(Given as Formative assessment 4)

Exercise 32. ($\star$) Show that the local characteristics

$$\Pr_1 (x_1|x_2) = \sqrt{\frac{1}{2\pi}} \exp\left(-\frac{1}{2}(x_1 - x_2)^2\right)$$
$$\Pr_2 (x_2|x_1) = \sqrt{\frac{1}{2\pi}} \exp\left(-\frac{1}{2}(x_2 - x_1)^2\right)$$

do not define a proper joint distribution on $\mathbb{R}^{\{1,2\}}$.

Solution.

Exercise 33. ($\star\star$) Consider the model

$$Z = BZ + (I - B) X\beta + E$$

where $E \sim \mathrm{N}\left(0, \sigma^2 I\right)$, X is a $n \times p$ design matrix X, $\beta \in \mathbb{R}^p$, B is an $n \times n$ matrix with $[B]_{i,i} = 0$ and $(I - B)$ is non-singular.

(1) Show that

$$\mathrm{E}(Z) = X\beta$$
$$\mathrm{Var}(Z) = \sigma^2 (I - B)^{-1} \left(I - B^\top\right)^{-1}$$

(2) Show that the above model is SAR for $Z - \mathrm{E}(Z)$
(3) Assume that $((I - B) X)^\top (I - B) X$ is non-singular. Compute the Maximum Likelihood Estimators (MLE) $\hat{\beta}$ and $\hat{\sigma}^2$ of β and σ^2.
(4) Derive the sampling distribution of $\hat{\beta}$ given X.

Solution.

(1) It is

$$\mathrm{E}(Z) = \mathrm{E}(BZ + (I - B) X\beta + E) \Longleftrightarrow$$
$$\mathrm{E}(Z) = \mathrm{E}(BZ) + (I - B) X\beta + \mathrm{E}(E) \Longleftrightarrow$$
$$(I - B)\,\mathrm{E}(Z) = (I - B) X\beta + \overset{=0}{\cancel{\mathrm{E}(E)}} \Longleftrightarrow$$
$$\mathrm{E}(Z) = X\beta$$

also

$$\mathrm{Var}((I - B) Z) = \mathrm{Var}((I - B) X\beta + E)$$
$$\mathrm{Var}((I - B) Z) = \mathrm{Var}(E)$$
$$(I - B)\,\mathrm{Var}(Z)\,(I - B)^\top = \sigma^2 I$$
$$\mathrm{Var}(Z) = (I - B)^{-1} \sigma^2 I \,(I - B)^{-\top}$$

(2) It is

$$
\begin{aligned}
Z - \mathrm{E}(Z) = B(Z - X\beta) + E &\iff \\
(Z - X\beta) = B(Z - X\beta) + E &\iff \\
\tilde{Z} = B\tilde{Z} + E &
\end{aligned}
$$

where $E \sim \mathrm{N}\left(0, \sigma^2 I\right)$, hence $\tilde{Z} := Z - \mathrm{E}(Z)$ is a SAR model given the assumptions taken.

(3) The likelihood of Z given the parameters β, and σ^2 is

$$
\begin{aligned}
L\left(Z; \beta, \sigma^2\right) &= \mathrm{N}\left(Z | \mathrm{E}(Z), \mathrm{Var}(Z)\right) \\
&= \mathrm{N}\left(Z | X\beta, (I - B)^{-1} \sigma^2 I (I - B)^{-\top}\right)
\end{aligned}
$$

Hence

$$
\begin{aligned}
-2\log\left(L\left(Z; \beta, \sigma^2\right)\right) =& -2\log\left(\mathrm{N}\left(Z | X\beta, (I - B)^{-1} \sigma^2 I (I - B)^{-\top}\right)\right) \\
=& \log\left(\det\left((I - B)^{-1} \sigma^2 I (I - B)^{-\top}\right)\right) \\
&+ (Z - X\beta)^\top \left((I - B)^{-1} \sigma^2 I (I - B)^{-\top}\right)^{-1} (Z - X\beta) \\
=& \log\left(\det\left((I - B)^{-1} \sigma^2 I (I - B)^{-\top}\right)\right) \\
&+ \frac{1}{\sigma^2} (Z - X\beta)^\top (I - B)^\top (I - B)(Z - X\beta)
\end{aligned}
$$

The likelihood equations are

$$
\begin{aligned}
0 &= \nabla_{(\beta,\sigma^2)} \left(-2\log\left(L\left(Z; \beta, \sigma^2\right)\right)\right)\Big|_{(\beta,\sigma^2) = \left(\hat{\beta}, \hat{\sigma}^2\right)} \\
&= \begin{bmatrix} \frac{\partial}{\partial \beta} \left(-2\log\left(L\left(Z; \beta, \sigma^2\right)\right)\right) \\ \frac{\partial}{\partial \sigma^2} \left(-2\log\left(L\left(Z; \beta, \sigma^2\right)\right)\right) \end{bmatrix}_{(\beta,\sigma^2) = \left(\hat{\beta}, \hat{\sigma}^2\right)} \\
&= \begin{bmatrix} -\frac{1}{\sigma^2} X^\top 2 (I - B)^\top (I - B)(Z - X\beta) \\ -\frac{n}{\sigma^2} + \frac{1}{\sigma^4} (Z - X\beta)^\top (I - B)^\top (I - B)(Z - X\beta) \end{bmatrix}_{(\beta,\sigma^2) = \left(\hat{\beta}, \hat{\sigma}^2\right)}
\end{aligned}
$$

This is because

$$
\begin{aligned}
\frac{\partial}{\partial \beta} \left(-2\log\left(L\left(Z; \beta, \sigma^2\right)\right)\right) =& \frac{\partial}{\partial \beta} \left(\frac{1}{\sigma^2} (Z - X\beta)^\top (I - B)^\top (I - B)(Z - X\beta)\right) \\
=& \left[\frac{\partial}{\partial \beta} (Z - X\beta)\right] \frac{\partial}{\partial \xi} \left(\frac{1}{\sigma^2} \xi^\top (I - B)^\top (I - B)\xi\right)\Big|_{\xi = Z - X\beta} \\
=& -\frac{1}{\sigma^2} X^\top 2 (I - B)^\top (I - B)(Z - X\beta)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial}{\partial \sigma^2}\left(-2\log\left(L\left(Z;\beta,\sigma^2\right)\right)\right) =& \frac{\partial}{\partial \sigma^2}\left(\log\left(\det\left(\left(I-B\right)^{-1}\sigma^2 I\left(I-B\right)^{-\top}\right)\right)\right)\\
&+\frac{\partial}{\partial \sigma^2}\left(\frac{1}{\sigma^2}\left(Z-X\beta\right)^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\beta\right)\right)\\
=& \frac{\partial}{\partial \sigma^2}\left(-n\log\left(\sigma^2\right)\right)+\frac{\partial}{\partial \sigma^2}\left(\frac{1}{\sigma^2}\right)\left(Z-X\beta\right)^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\beta\right)\\
=& -\frac{n}{\sigma^2}+\frac{1}{\sigma^4}\left(Z-X\beta\right)^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\beta\right)
\end{aligned}
$$

So the likelihood equations are

$$
\begin{aligned}
0 =& X^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\hat{\beta}\right)\\
0 =& -\frac{n}{\hat{\sigma}^2}+\frac{1}{\hat{\sigma}^4}\left(Z-X\hat{\beta}\right)^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\hat{\beta}\right)
\end{aligned}
$$

Solving the first equation wrt $\hat{\beta}$ I get

$$
\begin{aligned}
0 =& X^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\hat{\beta}\right)\iff\\
0 =& X^\top\left(I-B\right)^\top\left(I-B\right)Z-X^\top\left(I-B\right)^\top\left(I-B\right)X\hat{\beta}\iff\\
X^\top\left(I-B\right)^\top\left(I-B\right)X\hat{\beta} =& X^\top\left(I-B\right)^\top\left(I-B\right)Z\iff\\
\hat{\beta} =& \left(X^\top\left(I-B\right)^\top\left(I-B\right)X\right)^{-1}X^\top\left(I-B\right)^\top\left(I-B\right)Z
\end{aligned}
$$

provided that $X^\top\left(I-B\right)^\top\left(I-B\right)X$ is non-singular (this is given, anyway).

Solving the second equation wrt $\hat{\sigma}^2$ I get

$$
\begin{aligned}
0 =& -\frac{n}{\hat{\sigma}^2}+\frac{1}{\hat{\sigma}^4}\left(Z-X\hat{\beta}\right)^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\hat{\beta}\right)\iff\\
0 =& -\frac{n}{1}+\frac{1}{\hat{\sigma}^2}\left(Z-X\hat{\beta}\right)^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\hat{\beta}\right)\iff\\
\hat{\sigma}^2 =& \frac{1}{n}\left(Z-X\hat{\beta}\right)^\top\left(I-B\right)^\top\left(I-B\right)\left(Z-X\hat{\beta}\right)
\end{aligned}
$$

(4) It is Normal as a linear combination of Normally distributed random variables. Its moments (mean and variance) are

$$
\begin{aligned}
\mathrm{E}\left(\hat{\beta}|X\right) =& \mathrm{E}\left(\left(X^\top\left(I-B\right)^\top\left(I-B\right)X\right)^{-1}X^\top\left(I-B\right)^\top\left(I-B\right)Z|X\right)\\
=& \left(X^\top\left(I-B\right)^\top\left(I-B\right)X\right)^{-1}X^\top\left(I-B\right)^\top\left(I-B\right)\mathrm{E}\left(Z|X\right)\\
=& \left(X^\top\left(I-B\right)^\top\left(I-B\right)X\right)^{-1}X^\top\left(I-B\right)^\top\left(I-B\right)X\beta\\
=& \beta
\end{aligned}
$$

and

$$
\begin{aligned}
\operatorname{Var}\left(\hat{\beta}|X\right) =&\operatorname{Var}\left(\left(X^{\top}(I-B)^{\top}(I-B)X\right)^{-1}X^{\top}(I-B)^{\top}(I-B)Z|X\right)\\
=&\left(X^{\top}(I-B)^{\top}(I-B)X\right)^{-1}X^{\top}(I-B)^{\top}(I-B)\operatorname{Var}(Z|X)\\
&\left(\left(X^{\top}(I-B)^{\top}(I-B)X\right)^{-1}X^{\top}(I-B)^{\top}(I-B)\right)^{\top}\\
=&\left(X^{\top}(I-B)^{\top}(I-B)X\right)^{-1}X^{\top}\cancel{(I-B)^{\top}(I-B)}\\
&\sigma^2\cancel{(I-B)^{-1}\left(I-B^{\top}\right)^{-1}}\\
&\left(\left(X^{\top}(I-B)^{\top}(I-B)X\right)^{-1}X^{\top}(I-B)^{\top}(I-B)\right)^{\top}\\
=&\sigma^2\cancel{\left(X^{\top}(I-B)^{\top}(I-B)X\right)^{-1}}\cancel{X^{\top}(I-B)^{\top}(I-B)X}\left(X^{\top}(I-B)^{\top}(I-B)X\right)^{-1}\\
=&\sigma^2\left(X^{\top}(I-B)^{\top}(I-B)X\right)^{-1}
\end{aligned}
$$

- Notice that, in Frequentist Statistical framework, once we have computed the sampling distributions (those above), we can produce inference tools in the similar manner to Normal Linear regression.

Exercise 34. ($\star\star\star$) Suppose that $\mathcal{S}$ is a finite set that contains at least two elements and is equipped with a symmetric relation $\sim$. Consider the Poisson auto-regression model defined as

$$
\begin{cases}
y_i|y_{\mathcal{S}\setminus\{i\}}\sim & \text{Poisson}(\mu_i)\\
\log(\mu_i)= & \theta\sum_{i\sim j, j\neq i} y_j
\end{cases}
$$

for $y\in\mathbb{N}^{\mathcal{S}}$.

Hint: You can use that if $X\sim\text{Poisson}(\mu)$ then X has PMF

$$
\Pr_X(x|\mu)=\frac{1}{x!}\exp(-\mu)\,\mu^x 1(x\in\{0,1,2,...\})
$$

(1) Show that the above model is well-defined if and only if $\theta\leq 0$.

(2) Find the canonical potential with respect to $\zeta=0$.

Solution. It is

$$
\Pr_i\left(y_i|y_{\mathcal{S}\setminus\{i\}}\right)=\frac{1}{y_i!}\exp(-\mu_i)\,\mu_i^{y_i}1(y_i\in\mathbb{N})
$$

(1) It is

$$
\Pr_i\left(y_i=0|y_{\mathcal{S}\setminus\{i\}}\right)=\exp(-\mu_i)
$$

and

$$
\Pr_i\left(y_i=\ell|y_{\mathcal{S}\setminus\{i\}}\right)=\frac{1}{\ell!}\exp(-\mu_i)\,\mu_i^{\ell}
$$

for $\ell \in \mathbb{N}$. Then by the Besag's factorization theorem wrt reference 0 it is

$$\begin{aligned}\frac{\mathrm{Pr}_Y(y)}{\mathrm{Pr}_Y(0)} &= \prod_{i\in\mathcal{S}} \frac{\mathrm{Pr}_i(y_i|y_1,...,y_{i-1},0,...,0)}{\mathrm{Pr}_i(0|y_1,...,y_{i-1},0,...,0)} \\ &= \prod_{i\in\mathcal{S}} \frac{\frac{1}{y_i!}\cancel{\exp(-\mu_i)}\mu_i^{y_i}}{\cancel{\exp(-\mu_i)}} = \prod_{i\in\mathcal{S}} \frac{1}{y_i!}\mu_i^{y_i} \\ &= \prod_{i\in\mathcal{S}} \frac{1}{y_i!}\exp\left(\theta\sum_{i\sim j, j\neq i} y_j\right)^{y_i} \\ &= \exp\left(\theta\sum_{i\sim j, j\neq i} y_i y_j - \sum_{i\in\mathcal{S}}\log(y_i!)\right)\end{aligned}$$

That is

$$\Pr_Y(y) = \exp\left(\theta\sum_{i\sim j, j\neq i} y_i y_j - \sum_{i\in\mathcal{S}}\log(y_i!)\right)\Pr_Y(0)$$

Now, if $\theta \le 0$ then $\theta\sum_{i\sim j, j\neq i} y_i y_j \le 0$ hence the constant is

$$\begin{aligned}\sum_{y\in\mathbb{N}^{\mathcal{S}}} \frac{\mathrm{Pr}_Y(y)}{\mathrm{Pr}_Y(0)} &= \sum_{y\in\mathbb{N}^{\mathcal{S}}} \exp\left(\theta\sum_{i\sim j, j\neq i} y_i y_j - \sum_{i\in\mathcal{S}}\log(y_i!)\right) \\ &\le \sum_{y\in\mathbb{N}^{\mathcal{S}}} \exp\left(-\sum_{i\in\mathcal{S}}\log(y_i!)\right) \\ &= \sum_{y\in\mathbb{N}^{\mathcal{S}}}\prod_{i\in\mathcal{S}}\frac{1}{y_i!} \\ &= \left(\sum_{\ell=0}^{\infty}\frac{1}{\ell!}\right)^{\mathrm{Card}(\mathcal{S})} = \exp(\mathrm{Card}(\mathcal{S})) < \infty\end{aligned}$$

If $\theta > 0$ without loss of generality consider the first two sites and suppose that $1 \sim 2$, then

$$\frac{\mathrm{Pr}_Y\left((y_1, y_2, 0, ..., 0)^\top\right)}{\mathrm{Pr}_Y(0)} = \frac{\exp(\theta y_1 y_2)}{y_1! y_2!}$$

should be summable. However, the series

$$\sum_{y_1=1}^{\infty}\sum_{y_2=1}^{\infty}\frac{\exp(\theta y_1 y_2)}{y_1! y_2!} = \infty$$

diverges as the general term $\frac{\exp(\theta y_1 y_2)}{y_1! y_2!}$ does not go to zero.

(2) By definition, $V_\emptyset = 0$.

Then I will use Theorem **??** from Handout **??**, and $\zeta = 0$.

For $\mathcal{A} = \{i\}$, it is

$$V_{\{i\}}(y) = \log\left(\Pr_i(y_i|0,...,0)\right) - \log\left(\Pr_i(0|0,...,0)\right) = -\log(y_i!)$$

For $\mathcal{A}=\{i,j\}$, it is

$$
\begin{aligned}
V_{\{i,j\}}(y) =& \log\left(\Pr_i(y_i|y_j,0,...,0)\right)\\
&-\log\left(\Pr_i(y_i|0,...,0)\right)-\log\left(\Pr_i(y_j|0,...,0)\right)\\
&+\log\left(\Pr_i(0|0,...,0)\right)\\
=&-y_iy_j1(i\sim j)
\end{aligned}
$$

So

$$V_{\{i,j\}}(y)=-y_iy_j1(i\sim j)$$

Since the joint distribution is proportional such as

$$
\begin{aligned}
\Pr_Y(y) &= \exp\left(\theta\sum_{i\sim j,j\neq i}y_iy_j-\sum_{i\in\mathcal{S}}\log(y_i!)\right)\Pr_Y(0)\\
&\propto \exp\left(\theta\sum_{i\sim j,j\neq i}y_iy_j-\sum_{i\in\mathcal{S}}\log(y_i!)\right)
\end{aligned}
$$

all the other potentials are zero.

Perhaps not the most elegant derivation. Proposition **??** in Handout **??** provides a more elegant tool to compute such stuff which is based on the “exponential distribution family”.

Exercise 35. ($\star\star\star$) Consider the model

$$Z=X\beta+B(Z-X\beta)+E$$

where X is a $n\times p$ design matrix X, $\beta\in\mathbb{R}^p$, B is an $n\times n$ symmetric positive definite matrix with $[B]_{i,i}=0$, $E\sim\mathrm{N}\left(0,\sigma^2(I-B)\right)$, and $\sigma^2>0$.

(1) This is a multiple choice question, choose any number of correct answers.
   (a) Z follows a simultaneous autoregressive (SAR) with Gaussian joint distribution with mean $X\beta$ and covariance matrix $\sigma^2(I-B)^{-1}$
   (b) Ising model
   (c) Conditional autoregressive (CAR) with Gaussian joint distribution with mean $X\beta$ and covariance matrix $\sigma^2 I$
   (d) Convolutional neural network (CNN)

(2) Compute the Maximum Likelihood Estimators (MLE) $\hat{\beta}$, and $\hat{\sigma}^2$ of β, and σ^2, as

$$
\begin{aligned}
\hat{\beta} &= \left(X^\top(I-B)X\right)^{-1}X^\top(I-B)Z\\
\hat{\sigma}^2 &= \frac{1}{n}\left(Z-X\hat{\beta}\right)^\top(I-B)\left(Z-X\hat{\beta}\right)
\end{aligned}
$$

Solution.

(Given as Formative assessment 4)

Exercise 36. ($\star\star$) Let $Z \in \mathcal{Z}^{\mathcal{S}}$ where $\mathcal{S} = \{1, ..., n\}$ and $\mathcal{Z} = \mathbb{R}$. Consider the model

$$Z = X\beta + B\left(Z - X\beta\right) + E$$

where X is a $n \times p$ design matrix X, $\beta \in \mathbb{R}^p$, $I - B$ is an $n \times n$ symmetric positive definite matrix with $[B]_{i,i} = 0$, $E \sim \mathrm{N}\left(0, \sigma^2\left(I - B\right)\right)$, and $\sigma^2 > 0$.

Hint: The following formulas are provided for your information

- $\partial\left(XY\right) = \left(\partial X\right)Y + X\left(\partial Y\right)$
- $\partial\left(X^\top\right) = \left(\partial X\right)^\top$
- $\frac{\partial}{\partial x}\left(x^\top Bx\right) = \left(B + B^\top\right)x$
- $\frac{\partial}{\partial x}\left(\left(s - Ax\right)^\top W\left(s - Ax\right)\right) = -2AW\left(s - Ax\right)$

(1) This is a multiple choice question, choose any number of correct answers.
  (a) Z follows a simultaneous autoregressive (SAR) with Gaussian joint distribution with mean $X\beta$ and covariance matrix $\sigma^2\left(I - B\right)^{-1}$
  (b) Ising model
  (c) Conditional autoregressive (CAR) with Gaussian joint distribution with mean $X\beta$ and covariance matrix $\sigma^2 I$
  (d) Bernoulli regression

(2) Show that the minus two log Pseudo-Likelihood is such as

$$-2\log\left(\text{pseudo-}L\left(Z; \beta, \sigma^2\right)\right) = n\log\left(\sigma^2\right) + \frac{1}{\sigma^2}\left(Z - X\beta\right)^\top\left(I - B\right)^2\left(Z - X\beta\right) + \text{const.}$$

(3) Compute the Maximum Pseudo-Likelihood Estimators (MPLE) $\tilde{\beta}$ and $\tilde{\sigma}^2$ of β and σ^2

Solution.

Exercise 37. ($\star\star$) Let B be a symmetric matrix with $[B]_{s,t} = 0$ and such that $(I - B)$ is positive definite. Consider the conditional autoregression model on a finite family $\mathcal{S} = \emptyset$ of sites defined by Gaussian local characteristics with

$$\mathrm{E}\left(Z_t | Z_{\mathcal{S}\setminus t}\right) = \mu + \sum_{s \neq t}[B]_{s,t}\left(Z_s - \mu\right)$$

and $\mathrm{Var}\left(Z_t | Z_{\mathcal{S}\setminus t}\right) = 1$ for $s \in \mathcal{S}$ for some unknown parameter $\mu \in \mathbb{R}$.

(1) Compute the joint distribution of $Z = \left(Z_1, ..., Z_n\right)^\top$
(2) Compute the MLE $\hat{\mu}$ of μ.
(3) Compute the sampling distribution of $\hat{\mu}$.
(4) Compute an $(1 - a)\,100\%$ confidence interval for μ based on the sampling distribution of $\hat{\mu}$ and with the minimum length. State any assumptions you take.
(5) Compute the rejection area $\mathcal{R}_a\left(\{Z_i\}\right)$ of the likelihood ratio test with null hypothesis $\mathrm{H}_0 : \mu = 0$ and alternative hypothesis $\mathrm{H}_1 : \mu \neq 0$ at significance level a.

Solution.

(1) I use Besag's theorem as in Proposition **??** in Handout **??**: Aerial unit data / spatial data on lattices.

Without lose of generality, consider zero mean $\mu = 0$ (or equivalently set $Z := Z - \mu 1$). The full conditionals $Z_i | z_{\mathcal{S}-i}$ are compatible with the joint distribution $\Pr_Z(z)$. By using Besag's factorization theorem with reference state $z^* = 0$ we get

$$
\begin{aligned}
\Pr_Z(z) &= \prod_{i=1}^{n} \frac{\Pr_i\left(z_i | z_1, ..., z_{i-1}, z^*_{i+1} = 0, ..., z^*_n = 0\right)}{\Pr_i\left(z^*_i = 0 | z_1, ..., z_{i-1}, z^*_{i+1} = 0, ..., z^*_n = 0\right)} \Pr_Z(z^* = 0) \\
&= \prod_{i=1}^{n} \frac{\mathrm{N}\left(z_i | \sum_{j<i} b_{i,j} z_j + 0, 1\right)}{\mathrm{N}\left(0 | \sum_{j<i} b_{i,j} z_j + 0, \kappa_i\right)} \Pr_Z(z^* = 0) \\
&\propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\left(z_i - \sum_{j<i} b_{i,j} z_j\right)^2 + \frac{1}{2}\left(0 - \sum_{j<i} b_{i,j} z_j\right)^2\right) \\
&= \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\left(z_i^2 - 2 z_i \sum_{j<i} b_{i,j} z_j\right)\right) \Pr_Z(z^* = 0) \\
&= \exp\left(-\sum_i \frac{z_i^2}{2} + \frac{1}{2} 2 \sum_i \sum_{j<i} b_{i,j} z_i z_j\right) \Pr_Z(z^* = 0) \\
&= \exp\left(-\frac{1}{2} z^\top I z + \frac{1}{2} z^\top B z\right) \Pr_Z(z^* = 0) = \exp\left(-\frac{1}{2} z^\top (I - B) z\right) \Pr_Z(z^* = 0) \\
&= \mathrm{N}\left(z | 0, (I - B)^{-1}\right) \qquad (5)
\end{aligned}
$$

Recovering the mean from (5), it is

$$\Pr_Z(z) = \mathrm{N}\left(z - \mu 1 | 0, (I - B)^{-1}\right) = \mathrm{N}\left(z | \mu 1, (I - B)^{-1}\right)$$

So

$$Z \sim \mathrm{N}\left(\mu 1, (I - B)^{-1}\right)$$

Alternatively, I just remember that this is a CAR model with joint distribution

$$Z \sim \mathrm{N}\left(\mu 1, (I - B)^{-1}\right)$$

(2) The -2 log likelihood function is

$$
\begin{aligned}
-2 \log\left(\Pr(Z | \mu)\right) &= -2 \log\left(\mathrm{N}\left(Z | \mu 1, (I - B)^{-1}\right)\right) \\
&= (Z - \mu 1)^\top (I - B)(Z - \mu 1) + \text{const.}
\end{aligned}
$$

The likelihood equations are

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \mu}\left(-2\log\left(L\left(Z;\mu\right)\right)\right)\Big|_{\mu=\hat{\mu}} \\
&= 1^{\top}2\left(I-B\right)\left(Z-\mu 1\right)\Big|_{\mu=\hat{\mu}}
\end{aligned}
$$

Hence the MLE is

$$
\begin{aligned}
\hat{\mu} &= \left(1^{\top}\left(I-B\right)1\right)^{-1}1^{\top}\left(I-B\right)Z \\
&= \frac{1}{1^{\top}\left(I-B\right)1}1^{\top}\left(I-B\right)Z \\
&= \frac{1}{1^{\top}I1-1^{\top}B1}\left(1^{\top}IZ-1^{\top}BZ\right) \\
&= \frac{1}{1^{\top}I1-1^{\top}B1}\left(\sum_{t\in\mathcal{S}}Z_t-\sum_{t\in\mathcal{S}}Z_t\sum_{s\in\mathcal{S}}B_{s,t}\right) \\
&= \frac{1}{|\mathcal{S}|-\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{S}}B_{s,t}}\sum_{t\in\mathcal{S}}\left(1-\sum_{s\in\mathcal{S}}B_{s,t}\right)Z_t
\end{aligned}
$$

(3) Hence the sampling distribution of $\hat{\mu}$ is

$$
\hat{\mu}\sim \mathrm{N}\left(\mu,\frac{1}{1^{\top}\left(I-B\right)1}\right)
$$

or

$$
\frac{\hat{\mu}-\mu}{\sqrt{\frac{1}{1^{\top}(I-B)1}}}\sim \mathrm{N}\left(0,1\right)
$$

This is because $\hat{\mu}$is normal as a linear combination of Normal random variables, and it has moments

$$
\begin{aligned}
\mathrm{E}\left(\hat{\mu}\right) &= \mathrm{E}\left(\frac{1}{1^{\top}\left(I-B\right)1}1^{\top}\left(I-B\right)Z\right)=\frac{1}{1^{\top}\left(I-B\right)1}1^{\top}\left(I-B\right)\mathrm{E}\left(Z\right) \\
&= \frac{1}{1^{\top}\left(I-B\right)1}1^{\top}\left(I-B\right)\mu 1=\mu
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}\left(\hat{\mu}\right) &= \frac{1}{1^{\top}\left(I-B\right)1}1^{\top}\left(I-B\right)\mathrm{Var}\left(Z\right)\left(I-B\right)^{\top}1\frac{1}{1^{\top}\left(I-B\right)1} \\
&= \frac{1}{1^{\top}\left(I-B\right)1}1^{\top}\left(I-B\right)\left(I-B\right)^{-1}\left(I-B\right)^{\top}1\frac{1}{1^{\top}\left(I-B\right)1} \\
&= \frac{1}{1^{\top}\left(I-B\right)1}
\end{aligned}
$$

(4) Because

$$
\frac{\hat{\mu}-\mu}{\sqrt{\frac{1}{1^{\top}(I-B)1}}}\sim \mathrm{N}\left(0,1\right)
$$

an $(1-a)\,100\%$ confidence interval for μ with the minimum length is

$$\left\{\hat{\mu} \pm z_{1-\frac{a}{2}}\sqrt{\frac{1}{1^{\top}(I-B)\,1}}\right\}$$

where $z_{1-\frac{a}{2}}$ is the $1-\frac{a}{2}$ quantile of the standard Normal distribution.

(5) The pair of hypotheses is

$$\mathrm{H}_0 : \mu = 0$$
$$\mathrm{H}_1 : \mu \neq 0$$

The rejection area is

$$\mathcal{R}_a(\{Z_i\}) = \left\{\{Z_i\} : -2\log\left(\frac{\sup_{\mu=0}\Pr(Z|\mu)}{\sup_{\mu\neq 0}\Pr(Z|\mu)}\right) > \lambda\right\} = \left\{\{Z_i\} : -2\log\left(\frac{\Pr(Z|0)}{\Pr(Z|\hat{\mu})}\right) > \lambda\right\}$$

and

$$\begin{aligned}
-2\log\left(\frac{\Pr(Z|0)}{\Pr(Z|\hat{\mu})}\right) &= -2\log(\Pr(Z|0)) + 2\log(\Pr(Z|\hat{\mu})) \\
&= Z^{\top}(I-B)\,Z - (Z-\hat{\mu}1)^{\top}(I-B)\,(Z-\hat{\mu}1) \\
&= \frac{\left(\underline{1}^{\top}(I-B)\,Z\right)^2}{\underline{1}^{\top}(I-B)\,\underline{1}} = \frac{(\hat{\mu})^2}{\mathrm{Var}(\hat{\mu})}
\end{aligned}$$

So

$$\mathcal{R}_a(\{Z_i\}) = \left\{\{Z_i\} : -2\log\left(\frac{\sup_{\mu=0}\Pr(Z|\mu)}{\sup_{\mu\neq 0}\Pr(Z|\mu)}\right) > \lambda\right\} = \left\{\{Z_i\} : \frac{(\hat{\mu})^2}{\mathrm{Var}(\hat{\mu})} > \lambda\right\}$$

It is

$$\sqrt{\frac{(\hat{\mu})^2}{\mathrm{Var}(\hat{\mu})}} \overset{\mathrm{H}_0:\mu=0}{=} \frac{\hat{\mu}-\mu}{\sqrt{\frac{1}{1^{\top}(I-B)1}}} \overset{\mathrm{H}_0:\mu=0}{\sim} \mathrm{N}(0,1)$$

so

$$\frac{(\hat{\mu})^2}{\mathrm{Var}(\hat{\mu})} \sim \chi^2_1$$

To compute λ at significance level a, it is

$$a = \Pr\left(\left\{\{Z_i\} : \frac{(\hat{\mu})^2}{\mathrm{Var}(\hat{\mu})} > \lambda\right\} |\mathrm{H}_0\right) = \Pr_{X^2\sim\chi^2_1}\left(X^2 > \lambda\right) = 1 - \Pr_{X^2\sim\chi^2_1}\left(X^2 \leq \lambda\right)$$

and

$$\Pr_{X\sim\chi^2_1}\left(X^2 \leq \lambda\right) = 1-a$$

so $\lambda = \chi^2_{1,1-a}$

$$\mathcal{R}_a(\{Z_i\}) = \left\{\{Z_i\} : \frac{(\hat{\mu})^2}{\mathrm{Var}(\hat{\mu})} > \chi^2_{1,1-a}\right\} = \left\{\{Z_i\} : \frac{\left(\underline{1}^{\top}(I-B)\,Z\right)^2}{\underline{1}^{\top}(I-B)\,\underline{1}} > \chi^2_{1,1-a}\right\}$$
