TMA4250 Spatial Statistics
Assignment 1: Continuous Random Fields

H. Omre IMF/NTNU

February 2006

Introduction

This assignment contains problems related to continuous Random Fields. The R-package should be used in solving the problems and relevant functions can be found in the R library `geoR` which can be loaded by the instruction `library(geoR)`

Problem 1: Gaussian Random Fields (GRF)

Consider a stationary GRF $\{R(x); x \in [0, 50] \subset \mathcal{R}^1\}$, ie a 1D GRF with

$$\begin{aligned} E\{R(x)\} &= 10 \\ Var\{R(x)\} &= 9 \\ Corr\{R(x'), R(x'')\} &= \rho(x' - x'') \end{aligned}$$

Let $\mathcal{D} = [0, 50]$ be discretised in $\mathcal{L}_{\mathcal{D}} \in \{1, 2, \ldots, 50\}$ and define the discretised GRF $\{R(x); x \in \mathcal{L}_{\mathcal{D}}\}$.

a) Display the exponential and gaussian correlation function with varying range between 5 and 20. Use the function `cov.spatial` in the `geoR` library.

Discuss the relation between features of the correlation functions and the characteristics of the random function. Develop the relation between the correlation function and the variogram function. In which situations is the latter preferable?

b) Simulate one realisation of the discretised GRF for some of the correlation functions displayed in **a)**. Discuss the relation between features of the correlation function and the characteristics of the realisations.

Select one of the realisations simulated in **b)**, and use the values in $x \in \{10, 25, 30\}$ as a set of observation $o = \{r(10), r(25), r(30)\}$.

c) Use the relation between the discretised GRF and the multivariate Gaussian pdf to specify the pdf for the conditional discretised GRF given the observations o.

Compute the expected values $\{E\{R(x)|o\}; x \in \mathcal{L}_\mathcal{D}\}$ and the variance values $\{Var\{R(x)|o\}; x \in \mathcal{L}_\mathcal{D}\}$. Display the results as an expectation function with associated 2σ intervals on either side.

Discuss the results.

d) Simulate 50 realisations of the conditional discretised GRF $\{[R(x)|o]; x \in \mathcal{L}_\mathcal{D}\}$. Display the realisations in one figure.

For each $x \in \mathcal{L}_\mathcal{D}$ compute the average and the empirical variance. Display the results as an average function with associated estimated 2σ intervals on either side.

Compare the results with the results in **c)** and comment.

Define a cut-off random function as follows:

$$i_{12}\left(R(x)\right) = \begin{cases} R(x) - 12 & \text{if } R(x) > 12 \\ 0 & \text{else} \end{cases}$$

and define the cut-off area

$$A_{12} = \int_0^{50} i_{12}\left(R(u)\right)\, du$$

which realisation can be displayed as

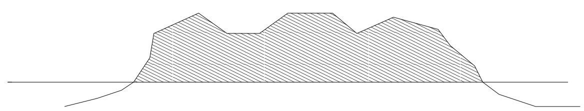

Figure 1: Schematic representation of the situation above

Note: This may correspond to the hydrocarbon volume in a petroleum reservoir with top-horizon $\{R(x); x \in \mathcal{D}\}$ and hydrocarbon-water contact at $r = 12$.

e) Estimate the expected cut-off area given the observations o, $E\{A_{12}|o\}$ in two ways

- by applying the cut-off rule on the conditional expected function in **c)**
- by the average of the areas obtained by applying the cut-off rule on the conditional realisations in **d)**

Compare the estimates and comment. Use the notation above and the Jensen's inequality to explain what is observed. How can $Var\{A_{12}|o\}$ be assessed?

Assume now that the observations $o = \{r(10), r(25), r(30)\}$ are made with observation error, ie: $o_* = \{r_*(10), r_*(25), r_*(30)\}$ where

$$R_*(i) = r(i) + U_i \ ; \ i \in \{10, 25, 30\}$$

and U_i are iid Gauss$(0,1)$. Use o and simulated realisations of the errors u_i to generate a simulated observation set o_*.

f) Repeat item **c)** and **d)** based on the observation set o_*.
Note that the observation errors causes the conditional Gaussian pdf in **c)** to be different.

Discuss the results.

Problem 2: Spatial Prediction by Kriging

This problem is based on observations of terrain elevation which are available in the R library MASS and on the web site of the course in the file topo.dat. The 52 observations are in unit meter in a domain $\mathcal{D} \subset \mathcal{R}^2$ of size (350×350) square-meters.

a) Display the observations in various ways. The function interp in the R library akima may be useful.

Comment the results.

Let the terrain elevation over $\mathcal{D}$ be modelled by the GRF $\{R(x); x \in \mathcal{D} \subset \mathcal{R}^2\}$ with

$$\begin{aligned} E\{R(x)\} &= \sum_{l=1}^{L} \beta_l g_l(x) \\ Cov\{R(x'), R(x'')\} &= c(|x' - x''|) \end{aligned}$$

with $\boldsymbol{g} : \{[g_1(x), \ldots, g_L(x)]; x \in \mathcal{D}\}$ a set of known functions over $\mathcal{D}$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_L)$ a vector of unknown weights.

Let the vector of observations be denoted $\mathbf{r}^o = (r(x_1), \ldots, r(x_{50}))$.

b) Develop the expression for the universal kriging predictor and prediction variance in an arbitrary location $x_0 \in \mathcal{D}$.

Let the reference variable $x \in \mathcal{D} \subset \mathcal{R}^2$ be denoted $x = (x_v, x_h)$, set $L = 6$ and define the set of known functions $\boldsymbol{g}$ to be all polynomials $x_v^k x_h^l$ for $(k,l) \in \{(0,0), (1,0), (0,1), (1,1), (2,0), (0,2)\}$. Further, let the covariance function $c(\tau)$ be of exponential form with variance 700 and range parameter 100.

c)Compute the universal kriging surface with associated kriging variance in a (100×100) grid covering $\mathcal{D}$. The function krige.conv may be useful.

Display the results and comment on them.

In grid node $x_0 = (10, 10)$, compute the value for which it is 0.90 probability that $[R(x_0|\theta]$ is below it.

d) Set $L = 1$ and define the set of known functions $\boldsymbol{g}$ to only contain the constant 1. Note that the associated covariance function is different from the one in **b)**, and determine a suitable covariance function.

Repeat the procedure in **c)** and comment on the results.

Problem 3: Parameter estimation

Consider a stationary GRF $\{R(x); x \in \mathcal{D} \subset \mathcal{R}^2\}$ with $\in [(1, 30), (1, 30)]$ and

$$\begin{aligned} E\{R(x)\} &= 10 \\ Var\{R(x)\} &= 4 \\ Corr\{R(x'), R(x'')\} &= \exp\left\{-\frac{|x' - x''|}{10}\right\} \end{aligned}$$

Discretise $\mathcal{D}$ into the lattice $\mathcal{L}_\mathcal{D}$ of size 30×30.

a) Simulate one realisation of the GRF represented on $\mathcal{L}_\mathcal{D}$, ie $\{r^s(x); x \in \mathcal{L}_\mathcal{D}\}$. This may take up to 5 minutes on the computer. Store the results.

Display the realisation and an estimate of the marginal distribution of $R(x_0)$ in arbitrary $x_0 \in \mathcal{D}$. Compute the empirical variogram based on the full realisation.

Comment on the results.

b) Use the realisation $\{r^s(x); x \in \mathcal{L}_\mathcal{D}\}$. Sample the surface values in 36 locations uniformly randomly drawn in $\mathcal{L}_\mathcal{D}$.

Compute the empirical variogram estimate based on these 36 observations. Use the function `variog`.

Assume an exponential variogram function with variance σ^2 and range parameter δ. Estimate σ^2 and δ by maximum likelihood based on these 36 observations. Use the function `likfit`.

Display the variogram estimates above together with the true one.

Comment on the results.

c) Repeat the procedure in **b)** with 9, 64 and 100 observations.

Comment on the results.

d) Discuss the following questions:

- Does it make a difference if you know the variogram model type?
- Which problems occur if the GRF is not stationary, ie it has varying expectation over $\mathcal{D}$?