75563 Spatial Statistics,
Assignment 1: Gaussian Random Fields and Kriging

Introduction This exercise contain problems concerning Gaussian random fields, Kriging, covariance functions, variograms etc. The assignment is designed for the use of the Spatial library of Splus. Chapter 15 of Venables & Ripley: Modern Applied Statistics with Splus, as well as web links from the home page may be useful. The S-plus library `Spatial` can be loaded by `library(spatial)`.

Problem 1: 1D Gaussian Random Fields

(a) Create a plot of the exponential covariance functions of order 1 and 2, with various choices of sill and range. Choose the parameter values such that it is possible to see their relative influence. Give an interpretation of the various parameters.

Note: Use functions available in the S-plus library `spatial`: `expcov` and `gaucov`.

What is the theoretical relation between covariance function and variogram function? What will make you chose one instead of the other?

(b) Sample a normalized 1D Gaussian random field on a grid of size 50 by using the multi-Gaussian pdf. Use various covariance functions and discuss the influence of parameter choices on the realizations.

Note: A function is given in "help" under the heading "Examples" for the function `rnorm`.

(c) Select one of the samples with exponential covariance function of order one in (b), and use the values in node 10, 25 and 30 as observations. Derive the expressions for the conditional pdf in the nodes of the grid not observed given the three observations. Plot conditional expectation and conditional variance (or standard deviation)for this example. Plot also the sample from which the observations were collected. Discuss the results.

(d) Sample the conditional Gaussian random field given the three observations in (c) 25 times and plot the results. Compare with the results of (b) and (c) and discuss.

(e) Compute the average and variance of the 25 samples in (d) in every grid node. Plot and comment the results.

(f) Define a cut-off value at 1.0 and consider the area below your selected realization and above this cut-off. To predict this area, given the three observations, two predictors for this can be imagined:

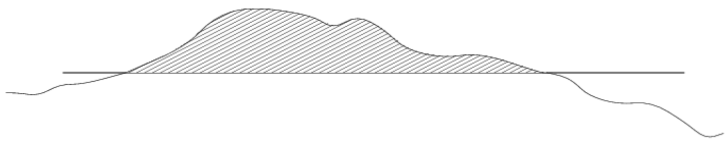
1. the area above the cut-off in the conditional expectation field generated in (c).
2. the average of the areas above the cut-off in many repeated samples of the conditional field in (d).

Use both these predictors to obtain a prediction of the area above the cut-off. Which predictor gives the largest value? Is there a theoretical reason for this? How would you represent the uncertainty of prediction using strategy 2? Compute the true area above the cut-off from the sample from which the observations were collected in (c). How does the true area relate to the predictions and the uncertainty for strategy 2 above?

Note: The splus syntax `sum(y[y>0])` may simplify your program.

(g) Repeat item (c)-(f) with an exponential covariance function of second order. Compare the results. Can any deviations be caused by the difference in covariance function?

Note: This schematic example can be seen as prediction of the oil in place under a geologic horizon and above the oil-water contact.



Figur 1: Schematic presentation of the situation above

Problem 2: Trend Surfaces and Kriging

Functions and data for this problem are in the library `spatial`. The data set `topo` contains 52 observations of the terrain elevation in a 310×310 feet square, using 50 feet units..

(a) Consider first a trend surface on the polynomial form:

$$f((x_1, x_2)) = \sum_{r+s \le p} a_{rs} x_1^r x_2^s$$

with p being the polynomial order.

Adapt the trend surface with $p = 2, 3, 4, 6$ to the data. Use the splus function `surf.ls`. Present the various surfaces with the plotting options ;`contour`, `image` and `persp`, include the data points in the plots. The function `trmat` might be useful. Discuss the features of the different plotting options. Comment on the results. What number of parameters must be estimated in a surface with $p = 6$? Can this cause problems ?

(b) In Universal Kriging, the surface has the following form:

$$Z(x) = f(x)^T \beta + \epsilon(x)$$

with x being the 2D spatial reference; $f(x)$ being given functions; β being unknown coefficients; and $\epsilon(x)$ being a Gaussian random field with expectation zero and spatial covariance function:

$$\mathrm{cov}\{Z(x), Z(y)\} = \mathrm{cov}\{\epsilon(x), \epsilon(y)\}$$

Generate the universal Kriging surface , using `surf.gls`. Use variance 1.0 and exponential covariance function with parameter $d = 0.7$. Let $f(x)$ be of polynomial form as in (a) and use different values of p, ie. different orders on the polynomial. Use in particular $p = 0$. Plot and comment the results. The function `prmat` might be useful. Try also different values for d.

Change the covariance function to second order exponential and rerun. Plot and comment the results.

(c) Plot the trend surfaces in the kriging models in (b),use `trmat`.

What is the difference between this estimate and the one obtained in (a), that is: what is the difference between the functions, `surf.gls` and `surf.ls`.

What is the difference of the trend surfaces and the Kriging surfaces?

(d) Plot the prediction standard deviation for the Kriging surfaces in (b). Comment on the results. Use the splus function `semat`.

Problem 3: Simulation and estimation.

Simulation of Gaussian random fields for dimension 2 and larger is computer intensive. In this problem we will use a program for simulation of 2D Gaussian random fields. The program simGauss simulates a normalized Gaussian random field with exponential correlation function, $c(r) = \exp(-3r/d)$, with $r = \|x_1 - x_2\|$. Input to the program is grid size and the parameter d in the covariance function. Note that the definition of 1. order exponential covariance function is is slightly different from Splus. To start the program type `/local/kurs/romstat/simGauss $RANDOM` The term `$RANDOM` gives the seed to the random generator, If this is left out a default value is used. The program writes the results to a file as a matrix of values, to represent the Gaussian random field at the grid nodes.

(a) Simulate a Gaussian random field with the program simGauss, using grid size 50×50 and parameter $d = 20$. Try out other parameter values as well. Plot the realisations by Splus. Use `sim <- read.table(filnavn ")` to read the data file into splus. Store the sampled fields for later use.

(b) Assume you observe the Gaussian Random field in 36 locations Randomly distributed across the grid. Use the function `correlogram` in splus to estimate the empirical correlation function. Plot the estimate and compare with the true correlation function.

(c) Estimate the correlation function using 9, 25, 49, 64 and 100 randomly distributed observations. How many observations do you need in order to get a satisfactory result.

(d) Now let the observations be placed on a regular grid. estimate the covariance using 9, 25, 49, 64 and 100 observations. How does this compare with (c).

(e) Identify some problems when estimating the covariance function if the function form is unknown.