Exercise 1 (Random Fields and the Semivariogram)

The aim of this exercise is to get a feeling for how different model parameters affect the appearance of the simulated realisations from a random field, and to see how this in turn is reflected in the estimated semivariogram. Use the functions `grf` and `variog` from the R package `geoR`, and set the random seed to `122` for all simulations, so results are comparable and reproducible.

(a) Simulate $n = 50, 100, 200$ and 500 irregularly placed points in the unit square from a Gaussian random field with mean $\mu = 0$, variance $\sigma^2 = 1$ and Matérn correlation function with $(\kappa, \phi) = (0.5, 0.25), (1.5, 0.16)$ and $(2.5, 0.13)$. Plot the simulated data.

(b) Estimate the semivariogram for the simulated data from part (a) and plot it. Vary the bin size and investigate the effect. Overlay the true semivariogram and compare the two. Then, vary the random seed and investigate the variation in the estimated semivariogram.

(c) Add to the data from (a) three times the squared first coordinate and two times the squared second coordinate of each location. Plot the data again and repeat the semivariogram estimation from part (b). What do you notice?

Exercise 2 (The Linear Geostatistical Model)

The aim of this exercise is to apply the concepts of Box-Cox transformation, estimation and prediction using the Gaussian geostatistical model. Use the functions `boxcox.fit`, `variog`, `variofit` and `likfit` from the R package `geoR`

(a) The package `geoR` works with data sets of type `geodata`, which contains coordinates (`coords`) and observations for Y (`data`). Additional information on covariates etc. is potentially stored in `covariate`, `borders` and `other`. Read the documentation for `as.geodata`.

(b) Load the data on Swiss rainfall in the `geoR` package, `sic.all` in data `SIC`. The Swiss borders are stored in `sic.borders`. Read the documentation for the data set and plot the data.

(c) Estimate an appropriate λ for a Box-Cox transformation of the data. Use $\lambda = 0.5$ in all further analyses. Look at a histogram of the square root transformed rain fall amounts. Does it look reasonably normal?

(d) Compute the empirical semivariogram and fit a Matérn model to the semivariogram in order to estimate starting values for the parameters σ^2 and ϕ in the maximum likelihood estimation in part (e). Do the parameter estimation for fixed values $\kappa = 0.5, 1$ and 2 using the function `variofit`.

(e) Fit a Gaussian linear geostatistical model to the rainfall data (with no covariates) for fixed $\kappa = 0.5, 1$ and 2 using maximum likelihood with the initial values for σ^2 and ϕ estimated in part (d). What are the best maximum likelihood estimates for the parameters? Interpret them. Repeat the analysis using REML estimation and compare. Repeat the analysis and simultaneously estimate λ (now taking into account the correlation structure in the data), confirming the previous choice of $\lambda = 0.5$.

(f) Plot the semivariogram for the data and overlay the model-based one estimated in part (e). Do the two agree well?

(g) Now, fit the Gaussian linear geostatistical model again allowing for a linear spatial trend in both coordinates. How do the estimates of σ^2 and of the practical range change and why? Compare the likelihoods and look at the confidence intervals - is the trend significant?

(h) Compute and plot predictions and corresponding prediction standard errors (plugging in the parameter estimates from (e)) for the rainfall surface in Switzerland on a regular 7.5 km by 7.5 km grid (use the seed 120 again). What do you notice? Then, plot the map of local probabilities of exceeding the threshold of 250 milimeters of rain, given the data.