UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Examination in: STK4150 — Environmental and spatial statistics

Day of examination: Friday, June 4, 2010.

Examination hours: 09.00 – 12.00.

This problem set consists of 2 pages.

Appendices: None.

Permitted aids: Accepted calculator.

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1 Gaussian processes

Assume $\{Z(\boldsymbol{x})\}$ is a Gaussian process in $\mathcal{R}^2$.

(a) Define what we mean about second order and strictly stationarity for the process.

(b) Assume we at positions $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ observe

$$Y_i = Z(\boldsymbol{x}_i) + \varepsilon_i$$

where $\varepsilon_1, ..., \varepsilon_n$ are iid and $N(0, \sigma^2)$.

Assume we want to predict $Z(\boldsymbol{x}_0)$ at some point $\boldsymbol{x}_0$. Given μ, the expected value of $Z(\boldsymbol{x})$ and $C(\boldsymbol{v})$, the covariance function for $\{Z(\boldsymbol{x})\}$, derive an expression for the conditional expectation of $Z(\boldsymbol{x}_0)$ given the observations and argue why this is a reasonable predictor.

Also derive an uncertainty measure for the predictor.

(c) Assume now that we instead observe

$$\widetilde{Y}_i = I(Z(\boldsymbol{x}_i) > 0)$$

where $I(A)$ is equal to one if A is true and zero otherwise.

Derive the distribution for $\widetilde{Y}_i$.

Also derive an expression for $\mathrm{Cov}[\widetilde{Y}_i, \widetilde{Y}_j]$.

(Continued on page 2.)

Problem 2 Extreme values

Assume $X_1, ..., X_n$ are iid and from some continuous distribution with distribution function F having the property that $F(x) < 1$ for $x < x_{\max}$ and $F(x_{\max}) = 1$ with $x_{\max} < \infty$. We are interested in the distribution for $M_n = \max\{X_1, ..., X_n\}$ for large n.

(a) Explain why looking at the limiting distribution for M_n as $n \to \infty$ does not help us much in this situation.

Assuming $b_n = n^{-1}$, show that $Z_n = (x_{\max} - M_n)/b_n$ for large n can be approximated by a non-degenerate distribution.

Use this to find an approximate distribution for M_n for large n.

(b) Assume now X_i is a daily measurement of some pollution quantity at a specific site and M_n corresponds to yearly maxima.

Define x_N by the solution to $\Pr(M_n \geq x_N) = N^{-1}$. Find an approximation for x_N.

Discuss the meaning of this value.

Discuss possible problems in using standard extreme value theory in this context.

Problem 3 Markov random field

Assume $\boldsymbol{X} = (X_1, ..., X_n)$ are a set of binary random variables defined on a regular lattice with probability distribution given by

$$\Pr(\boldsymbol{X} = \boldsymbol{x}) \propto \exp\{-\beta \sum_{i<j; |i-j|=1} x_i x_j\} \qquad (*)$$

(a) Show that $\boldsymbol{X}$ is a Markov random field and specify the neighborhood system related to this.

(b) Assume now that we observe

$$Y_i = aX_i + \tfrac{b}{M_i} \sum_{j; |j-i|=1} X_j + \varepsilon_i$$

where $\varepsilon_1, ..., \varepsilon_n$ are iid with density $f(\cdot)$ and M_i is the number of j's such that $|j - i| = 1$.

Derive an expression for the conditional distribution $\Pr(\boldsymbol{X} = x | \boldsymbol{Y} = y)$ and show that also this conditional distribution is a Markov random field.

Specify the neighborhood system in this case (for instance by a draw).

(c) Suggest possible ways of extending model (*) to a spatio-temporal setting (assuming discrete time points).

END