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 3 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.

FASIT

Problem 1 Gaussian processes

(a) Second order stationarity means that $E[Z(\boldsymbol{x})]$ is constant, μ say, for all $\boldsymbol{x}$ while $\text{Cov}[Z(\boldsymbol{x}), Z(\boldsymbol{x}+\boldsymbol{v})]$ only depend on $\boldsymbol{v}$, given through $C(\boldsymbol{v})$. Strictly stationarity means that $(Z(\boldsymbol{x}_1), ..., Z(\boldsymbol{x}_n))$ has the same distribution as $(Z(\boldsymbol{x}_1+\boldsymbol{v}), ..., Z(\boldsymbol{x}_n+\boldsymbol{v}))$ for all $n, \boldsymbol{x}_1, ..., \boldsymbol{x}_n, \boldsymbol{v}$. This is equivalent to second order stationarity in the Gaussian case.

(b) Define $\boldsymbol{\Sigma}_z$ to be the covariance matrix for $\boldsymbol{Z} = (Z(\boldsymbol{x}_1), ..., Z(\boldsymbol{x}_n))$. Then the covariance matrix for the observations is $\boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}_z + \sigma^2\boldsymbol{I}$. Further, by defining

$$\boldsymbol{c} = (C(|\boldsymbol{x}_1 - \boldsymbol{x}_0), ..., C(|\boldsymbol{x}_n - \boldsymbol{x}_0))^T,$$

we have

$$\begin{pmatrix}\boldsymbol{Y}\\ Z(\boldsymbol{x}_0)\end{pmatrix} \sim \text{MVN}\left(\mu\boldsymbol{1}_{n+1}, \begin{pmatrix}\boldsymbol{\Sigma}_y & \boldsymbol{c}\\ \boldsymbol{c}^T & C(0)\end{pmatrix}\right)$$

which leads to

$$\begin{aligned}E[Z(\boldsymbol{x}_0)|\boldsymbol{Y}] =& \mu + \boldsymbol{c}^T[\boldsymbol{\Sigma}_z + \sigma^2\boldsymbol{I}]^{-1}(\boldsymbol{Y} - \mu\boldsymbol{1}_n)\\ \text{Var}[Z(\boldsymbol{x}_0)|\boldsymbol{Y}] =& C(0) - \boldsymbol{c}^T[\boldsymbol{\Sigma}_z + \sigma^2\boldsymbol{I}]^{-1}\boldsymbol{c}\end{aligned}$$

Since the expectation is the centrepoint in the distribution for $Z(\boldsymbol{x}_0)$ given all we know, it is a reasonable predictor.

The conditional variance can be used as an uncertainty measure.

(Continued on page 2.)

(c) We have that Y_i is a binomial variable where

$$\Pr(Y_i = 1) = \Pr(Z(\boldsymbol{x}_i) > 0) = 1 - \Phi((0-\mu)/\sqrt{C(0)}) = \Phi(\mu/\sqrt{C(0)})$$

Further

$$\begin{aligned}\text{Cov}[\widetilde{Y}_i, \widetilde{Y}_j] =& E[\widetilde{Y}_i\widetilde{Y}_j] - \Phi(\mu/\sqrt{C(0)})^2\\ =& E[\widetilde{Y}_i\widetilde{Y}_j] - \Phi(\mu/\sqrt{C(0)})^2\\ =& \int_{z_i>0}\int_{z_j>0} f(z_i, z_j)dz_i dz_j - \Phi(\mu/\sqrt{C(0)})^2\end{aligned}$$

where $f(z_i, z_j)$ is the bivariate Gaussian distribution for $Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j)$

Problem 2 Extreme values

(a) We have $\Pr(M_n \leq x) \to 0$ for $x < x_{\max}$ while $\Pr(M_n \leq x_{\max}) = 1$, i.e. a degenerate distribution.

Define $a_n = x_{\max}$ and $b_n = n^{-1}$. We have

$$\begin{aligned}\Pr(Z_n \leq x) =& \Pr(x_{\max} - M_n \leq x/n)\\ =& \Pr(M_n \geq x_{\max} - x/n)\\ =& 1 - \Pr(M_n < x_{\max} - x/n)\\ =& 1 - F[x_{\max} - x/n]^n\\ =& 1 - \exp\{n \log F[x_{\max} - x/n]\}\\ =& 1 - \exp\{\tfrac{\log F[x_{\max} - x/n]}{n^{-1}}\}\end{aligned}$$

This gives

$$\begin{aligned}\lim_{n\to\infty} \Pr(Z_n \leq x) =& 1 - \lim_{n\to\infty} \exp\{\tfrac{\log F[x_{\max} - x/n]}{n^{-1}}\}\\ =& 1 - \lim_{n\to\infty} \exp\{\tfrac{\frac{f(x_{\max} - x/n)xn^{-2}}{F[x_{\max} - x/n]}}{-n^{-2}}\}\\ =& 1 - \lim_{n\to\infty} \exp\{-\tfrac{f(x_{\max} - x/n)x}{F[x_{\max} - x/n]}\}\\ =& 1 - \exp\{-xf(x_{\max})\}\end{aligned}$$

(b) We have

$$\Pr(M_n \geq x) = \Pr(Z_n \leq n(x_{\max} - x)) \approx 1 - \exp\{-n(x_{\max} - x)f(x_{\max})\}$$

This gives

$$1 - \exp\{-n(x_{\max} - x)f(x_{\max})\} \approx N^{-1}$$

which implies

$$x \approx x_{\max} - \log(N/(N-1))/[nf(x_{\max})]$$

This gives the value for which the maximum will exceed every Nth year.

(Continued on page 3.)

Problem 3 Markov random fields

(a) The Markov property is that

$$\Pr(X_k|X_j, j \neq i) = \Pr(X_k|X_j, j \in N_k)$$

where N_k denotes the set of neighbors of k.

We have that

$$\begin{aligned}
&\Pr(x_k|x_j, j \neq i)\\
&\quad\propto \exp\{-\beta \sum_{i<j;|i-j|=1} x_i x_j\}\\
&\quad\propto \exp\{-\beta \sum_{|k-j|=1} x_k x_j\}
\end{aligned}$$

showing the Markov property with $N_k = \{j; |k-j| = 1\}$.

(b) We have

$$\begin{aligned}
\Pr(\boldsymbol{X} = \boldsymbol{x}|\boldsymbol{Y} = \boldsymbol{y}) &\propto \Pr(\boldsymbol{X} = \boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{X} = \boldsymbol{x})\\
&\propto \exp\{-\beta \sum_{i<j;|i-j|=1} x_i x_j\} \prod_{i=1}^n f(y_i - a x_i - \tfrac{b}{M_i} \sum_{j;|j-i|=1} x_j)\\
&\propto \exp\{-\beta \sum_{i<j;|i-j|=1} x_i x_j + \sum_{i=1}^n \log f(y_i - a x_i - \tfrac{b}{M_i} \sum_{j;|j-i|=1} x_j)\}
\end{aligned}$$

Then

$$\begin{aligned}
\Pr(X_k =&x_k|X_j = x_j, j \neq k, \boldsymbol{Y} = \boldsymbol{y})\\
&\propto \exp\{-\beta \sum_{i<j;|i-j|=1} x_i x_j + \sum_{i=1}^n \log f(y_i - a x_i - \tfrac{b}{M_i} \sum_{j;|j-i|=1} x_j)\}\\
&\propto \exp\{-\beta \sum_{|k-j|=1} x_k x_j + \log f(y_k - a x_k - \tfrac{b}{M_k} \sum_{j;|j-k|=1} x_j)\}\\
&\quad \exp\{\sum_{j;|k-j|=1} \log f(y_j - a x_j - \tfrac{b}{M_j} \sum_{l;|l-j|=1} x_l)\}
\end{aligned}$$

showing that we still have a Markov random field, but now with $N_k = \{j; |k-j| = 1 \text{ or } |k-j| = \sqrt{2} \text{ or } |k-j| = 2\}$, i.e the first nearest neighbors, the 4 neighbors on the diagonal and the four grid points two units apart either in the horizontal or the verctical direction.

(c) Define $\{X_{t,i}\}$ to be the space-time process. Direct extension:

$$\Pr(\boldsymbol{X} = \boldsymbol{x}) \propto \exp\{-\beta \sum_{t=1}^T \sum_{i<j;|i-j|=1} x_{t,i} x_{t,j} - \gamma \sum_i \sum_{t=2}^T x_{t,i} x_{t-1,i}\} \quad (*)$$