Norges teknisk–
naturvitenskapelige universitet
Mathematical Sciences

Contact:
Jo Eidsvik
Tlf. 90127472

EXAM IN TMA4250
SPATIAL STATISTICS

Monday 23. May 2011

Tid: 09:00–13:00

Material:
Statistiscal tables and formulas, Tapir. Written yellow sheet of paper. Calculator.

Grade: 13. June 2011

Problem 1

We study a regression model with spatially correlated Gaussian noise terms. The response $y(\boldsymbol{s})$ at coordinate $\boldsymbol{s}$ is modeled by

$$y(\boldsymbol{s}) = \boldsymbol{x}^t(\boldsymbol{s})\boldsymbol{\beta} + \epsilon(\boldsymbol{s}), \qquad \boldsymbol{s} = (s(1), s(2)) \in (0,1) \times (0,1). \tag{1}$$

Here $\boldsymbol{x}(\boldsymbol{s})$ is a $p \times 1$ vector of explanatory variables at coordinate $\boldsymbol{s}$, while $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^t$ are regression parameters. The 0 mean Gaussian noise terms have covariance $\text{Cov}(\epsilon(\boldsymbol{s}), \epsilon(\boldsymbol{s}')) = \Sigma(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are model parameters in the chosen spatial covariance function.

a) Assume the noise process $\epsilon(\boldsymbol{s})$ is stationary and isotropic. What does this mean in practice and which model assumptions does it impose on the mathematical representation of $\Sigma(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta})$?

A common covariance function is the exponential $\Sigma(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}) = \theta_1 \exp(-\theta_2 h)$, where $h = ||\boldsymbol{s} - \boldsymbol{s}'||$ is the distance from $\boldsymbol{s}$ to $\boldsymbol{s}'$. Another possibility is the Cauchy type $\Sigma(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}) = \theta_1/(1 + \theta_2 h)^3$. Calculate the effective spatial correlation range in these two models, i.e. the distance h such that the correlation is 0.05.

b) We want to estimation parameters from data $y(\boldsymbol{s}_i)$ and explanatory variables $\boldsymbol{x}(\boldsymbol{s}_i)$, $i = 1, \ldots, n$. We assume an exponential covariance function for the noise process, treating θ_2 as known. Define the likelihoodfunction for $(\boldsymbol{\beta}, \theta_1)$.

Calculate the maximum likelihood estimator for $(\boldsymbol{\beta}, \theta_1)$.

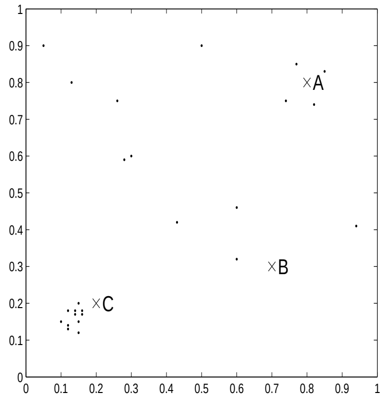

Figure 1: Data sites and three prediction sites (A, B, C).

c) We assume $\boldsymbol{\beta} = 0$. Figure 1 shows $n = 25$ measurement sites and three locations (A, B, C) where we want to predict.

Using parameter $\boldsymbol{\theta} = (1^2, 7.5)$ the three prediction variances are 0.67^2, 0.88^2 and 0.60^2. Which of the three (A, B eller C) has the smallest prediction variance (0.60^2)? Which has the largest(0.88^2)? Discuss the results.

We next study another model for the noise process:

$$\epsilon(\boldsymbol{s}) = I[\xi(\boldsymbol{s}) = 0]\epsilon_0(\boldsymbol{s}) + I[\xi(\boldsymbol{s}) = 1]\epsilon_1(\boldsymbol{s}), \qquad \boldsymbol{s} \in (0,1) \times (0,1), \tag{2}$$

where $\xi(\boldsymbol{s})$ is a random variable with outcomes 0 or 1 at each coordinate $\boldsymbol{s}$. Further, $\epsilon_0(\boldsymbol{s})$ and $\epsilon_1(\boldsymbol{s})$ are *independent* processes with 0 mean, variance 1 and covariance $\Sigma_0(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}) = \exp(-\theta_{0,2}h)$ for $\epsilon_0(\boldsymbol{s})$ and $\Sigma_1(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}) = \exp(-\theta_{1,2}h)$ for $\epsilon_1(\boldsymbol{s})$.

d) Compute $\text{Cov}(\epsilon(\boldsymbol{s}), \epsilon(\boldsymbol{s}')|\xi(\boldsymbol{s}), \xi(\boldsymbol{s}'))$ for all combinations of $\xi(\boldsymbol{s})$ and $\xi(\boldsymbol{s}')$.

Assume that $s(1) = 0.5$, and set $\xi(0.5, s(2)) = 0$ for $s(2) < 0.25$, $\xi(0.5, s(2)) = 1$ for $0.25 \leq s(2) < 0.35$, $\xi(0.5, s(2)) = 0$ for $0.35 \leq s(2) < 0.75$, and $\xi(0.5, s(2)) = 1$ for $s(2) \geq 0.75$. Sketch 5 possible realizations of the $\epsilon(0.5, s(2))$ process with parameters $\theta_{0,2} = 3$ and $\theta_{1,2} = 30$.

Assume that the $\xi(\boldsymbol{s})$ are defined from a Gaussian process $z(\boldsymbol{s})$, $\boldsymbol{s} \in (0,1) \times (0,1)$ such that:

$$\xi(\boldsymbol{s}) = 0 \ \text{ if } z(\boldsymbol{s}) < 0, \quad \xi(\boldsymbol{s}) = 1 \ \text{ if } z(\boldsymbol{s}) \geq 0. \tag{3}$$

Here we set $E(z(\boldsymbol{s})) = 0$, $Var(z(\boldsymbol{s})) = 1$ and with covariance $\Sigma_z(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}_z) = \exp(-\theta_{z,2}h)$, $h = ||\boldsymbol{s} - \boldsymbol{s}'||$.

e) Use the bivariate normal distribution to show that $P(\xi(\boldsymbol{s}) = 0 \cap \xi(\boldsymbol{s}') = 0) = \frac{\arctan(-\sqrt{1-\rho^2}/\rho)}{2\pi}$, where $\rho = \exp(-\theta_{z,2}h)$, $h = ||\boldsymbol{s} - \boldsymbol{s}'||$.

Take for granted that $\int_{-\infty}^{0} \phi(x)\Phi(bx)dx = \frac{\arctan(1/b)}{2\pi}$, where $\phi(x)$ is the density function of the standard normal and $\Phi(x)$ is the cumulative function of the standard normal.

f) Use the result from 1e) to compute $\text{Cov}(\epsilon(\boldsymbol{s}), \epsilon(\boldsymbol{s}'))$.

Problem 2

a) We consider a homogeneous Poisson process with intensity λ on the unit square $(0,1) \times (0,1)$. What is the probability of 0 points in $(0, 0.1) \times (0, 0.1)$?

We split the domain $(0,1) \times (0,1)$ in 100 disjoint cells of equal area 0.1^2. What is the probability that at least one of the cells contains 0 points? Find an approximate value for λ such that there is 0.05 probability for this event.

b) We next consider an inhomogeneous Poisson process with intensity $\lambda(x, y) = \lambda_0 xy$, $0 < x < 1$, $0 < y < 1$. Use the same split in 100 disjoint cells of equal size. What is the probability of 0 points in $((i-1)/10, i/10) \times ((j-1)/10, j/10)$, $i, j = 1, \ldots, 10$?

Describe a Monte Carlo algorithm for estimating the probability that at least one cell contains 0 points.

Problem 3

Data $\boldsymbol{x}$ has been acquired on a regular two dimensional grid of size $m \times n$. We model $\boldsymbol{x} = (x_1, \ldots, x_{mn})'$ as a binary Markov random field, with cell outcomes $x_i \in \{0, 1\}$, $i = 1, \ldots, mn$. We use a second order neighborhood.

a) Draw a second order neighborhood on a regular grid. What is a clique? Define the cliques of this second order neighborhoood.

Draw all clique configurations (16 in total). Use symmetry relations to split the configurations into 4 classes: 1 'one-colored', 2 'corner', 3 'row/column', 4 'diagonal'.

Each of these four classes of configurations are assigned a potential $\phi(\boldsymbol{x}_k)$, where $\boldsymbol{x}_k$ are the combinations of variables in the clique. $\phi(\boldsymbol{x}_k)$ takes one of the following values; ϕ_1, ϕ_2, ϕ_3, ϕ_4.

b) Define the pseudolikelihood as the product of the full conditionals of all single cells.

Express some of the terms in the pseudolikelihood using the model from above.

Describe a method for computing the maximum pseudolikelihood estimator of $(\phi_1, \phi_2, \phi_3, \phi_4)$.