TMA4250 Spatial Statistics
Assignment 3: Mosaic Spatial Variables

IMF/NTNU/HO&TR

April 2018

Introduction

This assignment contains problems related to mosaic spatial variables or more specifically Markov random fields (RF). The R-package should be used in solving the problems and relevant functions can be found in the R library `spatial` which can be loaded by the instruction `library(spatial)`.

Problem 1: Markov RF

This problem is based on observations of seismic data over a domain $\mathtt{D} \subset \mathbb{R}^2$. The objective is to identify the underlying {sand, shale} lithology distribution over $\mathtt{D}$, represented by $\{0, 1\}$ respectively.

The observations are collected on a regular (75×75) grid $\mathtt{L}_\mathtt{D}$, and the seismic data are denoted $\{d(\mathbf{x}); \mathbf{x} \in \mathtt{L}_\mathtt{D}\}; d(\mathbf{x}) \in \mathbb{R}$, represented by the n-vector $\mathbf{d}$. The observations are available in the R library MASS in the file *seismic.dat*.

Moreover, observations of the lithology distribution {sand, shale} in a geologically comparable domain $\mathtt{D}_c \subset \mathbb{R}^2$ is available. The lithology distribution is collected on a regular (50×50) grid $\mathtt{L}_{\mathtt{D}_c}$, with the same spacing as $\mathtt{L}_\mathtt{D}$, over $\mathtt{D}_c$. The observations with code $\{0, 1\}$ for {sand, shale} is available in the R library MASS in the file *complit.dat*.

Assume that the underlying lithology distribution can be represented by a Mosaic RF $\{l(\mathbf{x}); \mathbf{x} \in \mathtt{L}_\mathtt{D}\}; l(\mathbf{x}) \in \{0, 1\}$ represented by the n-vector $\mathbf{l}$.

a) The seismic data collection procedure defines the likelihood model:

$$[d_i|\mathbf{l}] = \begin{cases} 0.02 + U_i \text{ if } l_i = 0 - \text{sand} \\ 0.08 + U_i \text{ if } l_i = 1 - \text{shale} \end{cases} \quad ; i = 1, 2, \ldots, n$$

with $U_i; i = 1, 2, \ldots, n$ iid Gauss$\{0, 0.06^2\}$.

Specify the expression for the likelihood model $p(\mathbf{d}|\mathbf{l})$.

Display the observations $\{d(\mathbf{x}); \mathbf{x} \in \mathtt{L_D}\}$ as a map.

b) Consider a uniform, independence prior model on $\mathbf{l}$, i.e.

$$p(\mathbf{l}) = const$$

Develop an expression for the posterior model $p(\mathbf{l}|\mathbf{d})$, simulate 10 realizations of the posterior Mosaic RF $\{l(\mathbf{x}); \mathbf{x} \in \mathtt{L_D}|\mathbf{d}\}$, and display them as maps.

Develop expressions for the posterior expectation $\mathrm{E}\{\mathbf{l}|\mathbf{d}\}$ and the variances in the diagonal terms of the matrix $\mathrm{Var}\{\mathbf{l}|\mathbf{d}\}$, and display them as maps.

Develop expressions for the maximum marginal posterior predictor $\mathrm{MMAP}\{\mathbf{l}|\mathbf{d}\}$, and display the result as map.

Comment on the results, and give an interpretation of $\mathrm{E}\{\mathbf{l}|\mathbf{d}\}$.

c) Consider a Markov RF prior model for $\{l(\mathbf{x}); \mathbf{x} \in \mathtt{L_D}\}$, represented by the n-vector $\mathbf{l}$, with neighborhood system $\mathbf{n_L}$ consisting of four closest neighbors of each grid-node.

The Markov formulation is,

$$p(l_i|l_j; j \in \mathbf{n}_i) = const \times \exp\{\beta \sum_{j \in \mathbf{n}_i} I(l_i = l_j)\}; i = 1, 2, \ldots, n$$

with $I(A)$ equal 1 if A is true and equal 0 else.

Specify the associated Gibbs formulation for the Markov RF, i.e. $p(\mathbf{l})$.

Develop expressions for the posterior models $p(\mathbf{l}|\mathbf{d})$ and $p(l_i|\mathbf{d}, \mathbf{l}_{-i}); i = 1, 2, \ldots, n$. Focus is on realizations from $p(\mathbf{l}|\mathbf{d})$ and the prediction $\mathrm{E}\{\mathbf{l}|\mathbf{d}\}$, the variances in the diagonal terms of $\mathrm{Var}\{\mathbf{l}|\mathbf{d}\}$, and the alternative prediction $\mathrm{MMAP}\{\mathbf{l}|\mathbf{d}\}$. Explain a procedure for assessing them.

Display the observations from the geologically comparable domain $\mathtt{D}_c$ as a map.

Use these observations to estimate β by a maximum pseudo-likelihood procedure. Explain the procedure for doing so, and denote the estimate $\hat{\beta}$.

Set the model parameter $\beta = \hat{\beta}$ and use a MCMC/Gibbs algorithm to generate realizations from the posterior model $p(\mathbf{l}|\mathbf{d})$. Use torus/wrapping boundary conditions to avoid border problems. Document that the algorithm has converged by dispaying convergence plots of convergence indicators. Display 10 independent realizations as maps.

Estimate the predictor $\mathrm{E}\{\mathbf{l}|\mathbf{d}\}$, the prediction variances in the diagonal terms of $\mathrm{Var}\{\mathbf{l}|\mathbf{d}\}$, and the alternative prediction $\mathrm{MMAP}\{\mathbf{l}|\mathbf{d}\}$. Display the results from the estimations as maps.

Comment on the results.

d) Compare the results in **b)** and **c)** and comment on them.