TMA4250 Spatial Statistics
Assignment 3: Mosaic Spatial Variables

IMF/NTNU/HO

March 2020

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 on the course homepage 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 (66×66) 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 on the course homepage 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 $U_i; i = 1, 2, \dots, n$ 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 6 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 E$\{\mathbf{l}|\mathbf{d}\}$ and the posterior variances in the diagonal terms of the matrix Var$\{\mathbf{l}|\mathbf{d}\}$, and display them as maps.

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

Comment on the results.

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 clique system $\mathbf{c}_{\mathtt{L}}$ consisting of two closest neighbors on the grid $\mathtt{L_D}$.

The corresponding Gibbs formulation is,

$$p(\mathbf{l}) = const \times \prod_{\mathbf{c} \in \mathbf{c}_{\mathtt{L}}} \nu_{1l}(l_i; i \in \mathbf{c}) = const \times \prod_{<i,j> \in \mathtt{L_D}} \beta^{I(l_i = l_j)}$$

where $< i, j > \in \mathtt{L_D}$ defines the set of two-closest neighbors on the grid $\mathtt{L_D}$, parameter $\beta \in \mathbb{R}_{[1,\infty)}$, and $I(A)$ equal 1 if A is true and equal 0 else.

Specify the associated Markov formulation for the Markov RF.

Develop expressions for the posterior models $p(\mathbf{l}|\mathbf{d})$ and $p(l_i|\mathbf{d}, \mathbf{l}_{-i}); i = 1, 2, \dots, n$.

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 $\hat{\beta}$.

Focus is on realizations from $p(\mathbf{l}|\mathbf{d})$; with related predictions E$\{\mathbf{l}|\mathbf{d}\}$, variances in the diagonal terms of Var$\{\mathbf{l}|\mathbf{d}\}$, and alternative predictions MMAP$\{\mathbf{l}|\mathbf{d}\}$.

Set the model parameter $\beta = \hat{\beta}$ and use a MCMC/Gibbs algorithm to generate realizations from the posterior model $p(\mathbf{l}|\mathbf{d})$. Specify the McMC procedure on Algorithm format. Use torus/wrapping boundary conditions to avoid border problems. One simulation sweep corresponds to one visit per node in expectation. Consider carefully the number of sweeps required to obtain approximate convergence. Document that the algorithm has approximately converged by displaying convergence plots of sand proportion as convergence indicator, and explain how approximately independent realizations can be obtained. Display 6 approximately independent realizations as maps.

Comment on the results.

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 predictor $\mathrm{MMAP}\{\mathbf{l}|\mathbf{d}\}$. Explain the procedure for estimating them. Display the results from the estimations as maps.

Comment on the results.

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