Exercise 3 - Spatial Statistics

Amir Ahmed

April 2020

Problem 1: Markov RF

Assume that we have observed seismic data over a domain $D \in \mathbb{R}^2$. We want to identify the underlying lithology distribution over D, the underlying lithology of a point is either sand or shale, $\{1, 0\}$ respectively.

The observations have been collected on a regular (75×75) grid L_d, with seismic data being $\{d(\mathbf{x}); \mathbf{x} \in L_d\}$. Where $d(\mathbf{x}) \in \mathbb{R}$.

We have observed the lithology distribution in a geologically comparable domain $D_c \in \mathbb{R}^2$. Assume that this was collected on a regular (66×66) grid L_{D_c}.

We assume that the underlying lithology distribution can be represented by a Mosaic RF $\{l(\mathbf{x}); \mathbf{x} \in L_D\}, l(\mathbf{x}) \in \{0, 1\}$.

Problem 1a)

We start by looking at L_d. Let the seismic data collection procedure follow the following likelihood model:

$$[d_i|\mathbf{l}] = \begin{cases} 0.02 + U_i \text{ if sand, } l_i = 0 \\ 0.08 + U_i \text{ if shale, } l_i = 1 \end{cases}$$

$i = 1, 2, \ldots, n$. With U_i being identically independently distributed $U_i \sim N(0, 0.06^2)$. This would make each observation point d_i conditionally independent on $\mathbf{l}$. That will say:

$$p(d_i|\mathbf{l}) = p(d_i|l_i) = \phi(d_i|\mu = 0.02 + 0.06l_i, \sigma^2 = 0.06^2) \tag{1}$$

Where ϕ is the pdf of the normal distribution. As all observations are independent we thus have:

$$p(\mathbf{d}|\mathbf{l}) = \prod_{i=1}^{n} p(d_i|l_i) = \prod_{i=1}^{n} \phi(d_i|\mu = 0.02 + 0.06l_i, \sigma^2 = 0.06^2) \tag{2}$$

We display the observations from L_D as a map in Figure 1, there seems to be one large gathering where $d(\mathbf{x})$ takes on relatively large values, there also seems to be some smaller gatherings of large $d(\mathbf{x})$ in areas centered around the large one.

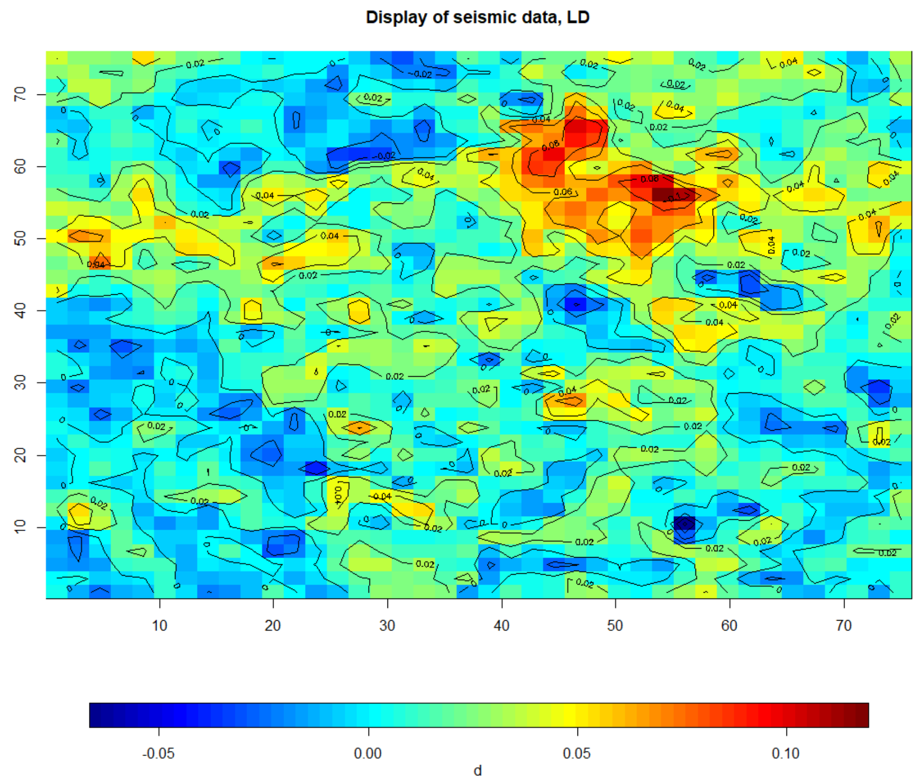

Figure 1: Display of seismic data L_D.

Problem 1b)

We now consider a uniform, independence prior model on $\mathbf{l}$. That will say:

$$p(\mathbf{l}) = const \tag{3}$$

We note that since the prior is constant we have:

$$p(\mathbf{l}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{l}) \tag{4}$$

We get the following posterior model using bayes law and the law of total probability:

$$p(\mathbf{l}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{l})}{\sum_{\mathbf{l}\in\mathbb{L}^n} p(\mathbf{d}|\mathbf{l})} \tag{5}$$

Inserting from (2) we get:

$$p(\mathbf{l}|\mathbf{d}) = \frac{\prod_{i=1}^{n} \phi(d_i|\mu = 0.02 + 0.06l_i, \sigma^2 = 0.06^2)}{\sum_{\mathbf{l}' \in \mathbb{L}^n} \prod_{i=1}^{n} \phi(d_i|\mu = 0.02 + 0.06l_i', \sigma^2 = 0.06^2)} \tag{6}$$

Where $\mathbb{L}^n$ is the n-dimensional space representing all possible values which l can take.

As the prior is independent each point would also be conditional independent, for each point we thus get the following. Let:

$$\begin{aligned} p_i &= p(l_i = 1|d_i) \\ &= \frac{p(d_i|l_i = 1)}{p(d_i|l_i = 0) + p(d_i|l_i = 1)} \\ &= \frac{\phi(d_i|\mu = 0.08, \sigma^2 = 0.06^2)}{\phi(d_i|\mu = 0.02, \sigma^2 = 0.06^2) + \phi(d_i|\mu = 0.08, \sigma^2 = 0.06^2)} \end{aligned} \tag{7}$$

As each point either is sand or shale we get:

$$1 - p_i = p(l_i = 0|d_i) \tag{8}$$

We recognize this conditioned model as something Bernoulli-distributed with probability p_i.

We thus have:

$$E(l_i|d_i) = p_i \tag{9}$$

and

$$Var(l_i|d_i) = p_i(1 - p_i) \tag{10}$$

We simulate 6 trials with the data and display the results in Figure 2.

The maximum marginal posterior predictor ($MMAP\{\mathbf{l}|\mathbf{d}\}$) is defined as:

$$MMAP\{\mathbf{l}|\mathbf{d}\} = \hat{\mathbf{l}} = argmax_{\mathbf{l} \in \mathbb{L}^n}\{p(\mathbf{l}|\mathbf{d})\} \tag{11}$$

Due to the conditional independence of the points we see:

$$\hat{l}_i = \begin{cases} 0, & \text{if } p_i < 0.5 \\ 1, & \text{if } p_i \geq 0.5 \end{cases} \tag{12}$$

Is a MMAP solution.

We plot MMAP solution, expectance and variance in Figure 3. We see that there is relatively high variance is most parts of the map, this is reflected in the large difference we see between the simulations. From both the expected value and the $MMAP$ we see that there is one large spot (top-right) where we expect a large cluster of shale as with some bands of shale on mid-left. This is somewhat reflected in the simulations. A differnce betweeen MMAP and the simulations is that the simulations tend to expect more shale than the MMAP. The map of the expected values seems to be closer to the simulations than the MMAP.

Problem 1c)

Now consider a Markov RF prior model for $\{l(\mathbf{x}); \mathbf{x} \in L_D\}$. Represented by the n-vector $\mathbf{l}$ with the clique system $\mathbf{c}_L$ consisting of two closest neighbors on the grid L_D.

The corresponding Gibbs formulation is:

$$\begin{aligned} p(\mathbf{l}) = const \times \prod_{\mathbf{c} \in \mathbf{c}_L} v_{1l}(l_i, i \in \mathbf{c}) &= const \times \prod_{<i,j> \in L_D} \beta^{I(l_i = l_j)} \\ &= const \times \beta^{\sum_{<i,j> \in L_D} I(l_i = l_j)} \end{aligned} \tag{13}$$

With $< i, j > \in L_d$ defining the set of two closest neighbors on the grid L_D.

Want to find expressions for the posterior models and want to specify the Markov formulation for the Markov RF.

Want to find the Markov formulation for the Markov RF. First see:

$$p(l_i | \mathbf{l}_{-i}) = \frac{p(\mathbf{l})}{\sum_{l'_i \in \mathbb{L}} p(l'_i, \mathbf{l}_{-i})} = \frac{p(\mathbf{l})}{p(l_i = 1, \mathbf{l}_{-i}) + p(l_i = 0, \mathbf{l}_{-i})} \tag{14}$$

$$p(l_i | \mathbf{l}_{-i}) = p(l_i | l_j, j \in n_i) \tag{15}$$

We note that the joint distribution is given by:

$$\begin{aligned} p(\mathbf{d}, \mathbf{l}) &= p(\mathbf{d}|\mathbf{l}) p(\mathbf{l}) \\ &= const \times \prod_{i=1}^{n} p(d_i | l_i) \prod_{\mathbf{c} \in \mathbf{c}_L} v_{1l}(l_i, i \in \mathbf{c}) \\ &= const \times \prod_{i=1}^{n} \phi(d_i | \mu = 0.02 + 0.06 l_i, \sigma^2 = 0.06^2) \prod_{<i,j> \in L_D} \beta^{I(l_i = l_j)} \end{aligned} \tag{16}$$

We input the above into the following:

$$p(\mathbf{l}|\mathbf{d}) = \frac{p(\mathbf{l}, \mathbf{d})}{p(\mathbf{d})} = const \times \prod_{i=1}^{n} \phi(d_i | \mu = 0.02 + 0.06 l_i, \sigma^2 = 0.06^2) \prod_{<i,j> \in L_D} \beta^{I(l_i = l_j)} \tag{17}$$

Also:

$$\begin{aligned} p(l_i, \mathbf{d} | \mathbf{l}_{-i}) &= p(\mathbf{d}|\mathbf{l}) p(l_i | \mathbf{l}_{-i}) \\ &= p(\mathbf{d}|\mathbf{l}) p(l_i | \mathbf{l}_{-i}) \\ &= \frac{p(\mathbf{l})}{p(l_i = 1, \mathbf{l}_{-i}) + p(l_i = 0, \mathbf{l}_{-i})} \prod_{i=1}^{n} \phi(d_i | \mu = 0.02 + 0.06 l_i, \sigma^2 = 0.06^2) \end{aligned} \tag{18}$$

Let n_i be the neighborhood around the ith node. Then have:

$$p(l_i, \mathbf{d} | \mathbf{l}_{-i}) = \frac{\prod_{l_j \in n_i} \beta^{I(l_i = l_j)}}{\prod_{l_j \in n_i} \beta^{I(0 = l_j)} + \prod_{l_j \in n_i} \beta^{I(1 = l_j)}} \prod_{i=1}^{n} \phi(d_i | \mu = 0.02 + 0.06 l_i, \sigma^2 = 0.06^2) \tag{19}$$

as the Markov formulation.

Now want to develop expressions for the posterior model $p(l_i|\mathbf{d}, \mathbf{l}_{-i})$ have:

$$p(l_i|\mathbf{d}, \mathbf{l}_{-i}) = p(l_i|d_i, l_j, j \in n_i)$$
$$= \frac{p(d_i|l_i)\prod_{l_j \in n_i} \beta^{I(l_i=l_j)}}{\phi(d_i|\mu = 0.02, \sigma^2 = 0.06^2)\prod_{l_j \in n_i} \beta^{I(l_j=0)} + \phi(d_i|\mu = 0.08, \sigma^2 = 0.06^2)\prod_{l_j \in n_i} \beta^{I(l_j=1)}} \tag{20}$$

We now display the observations in D_c in Figure 4.

Now want to use these observations to estimate β by a maximum pseudo-likelihood procedure. In a optimal solution we would use the assumed $p(\mathbf{l})$ distribution to do this, however we would need 2^n calculations to evaluate the normalizing constant, which in our case is infeasible. We rather use the Markov formulation to create an approximation. (here $\mathbf{d} \in \mathbb{L}^n$)

$$p(\mathbf{d}|\beta) \approx \hat{p}(\mathbf{d}|\beta) = const \times \prod_{i=1}^{n} \sum_{\{l_i', l_j | l_j \in n_i\} \in L} \prod_{l_j = l_i, l_j \in n_i} p(d_j|l_j')p(l_i'|l_j') \tag{21}$$

As we assume the observations to be exact the model reduces as:

$$p(\mathbf{d}|\mathbf{l}) = \prod_{i \in \mathbb{L}} p(d_i|l_i) \rightarrow \delta_{d_i}(l_i) = \begin{cases} 1, l_i = d_i \\ 0, \text{ else} \end{cases} \tag{22}$$

This reduces (21) to:

$$\hat{p}(\mathbf{d}|\beta) \propto \prod_{i=1}^{n} p(d_i|l_j; l_j \in n_i; \beta) = \prod_{i=1}^{n} \frac{\beta^{\sum_{l_j \in n_i} I(l_i=l_j)}}{\beta^{\sum_{l_j \in n_i} I(0=l_j)} + \beta^{\sum_{l_j \in n_i} I(1=l_j)}} \tag{23}$$

A good estimation of β would be to maximize the above giving:

$$\hat{\beta} = \text{argmax}_\beta \sum_{i=1}^{n} \left\{ \left(\sum_{l_j \in n_i} I(l_i = l_j) \right) \log \beta - \log \left(\beta^{\sum_{l_j \in n_i} I(0=l_j)} + \beta^{\sum_{l_j \in n_i} I(1=l_j)} \right) \right\} \tag{24}$$

Using the Optim function in R this value can be easily found.

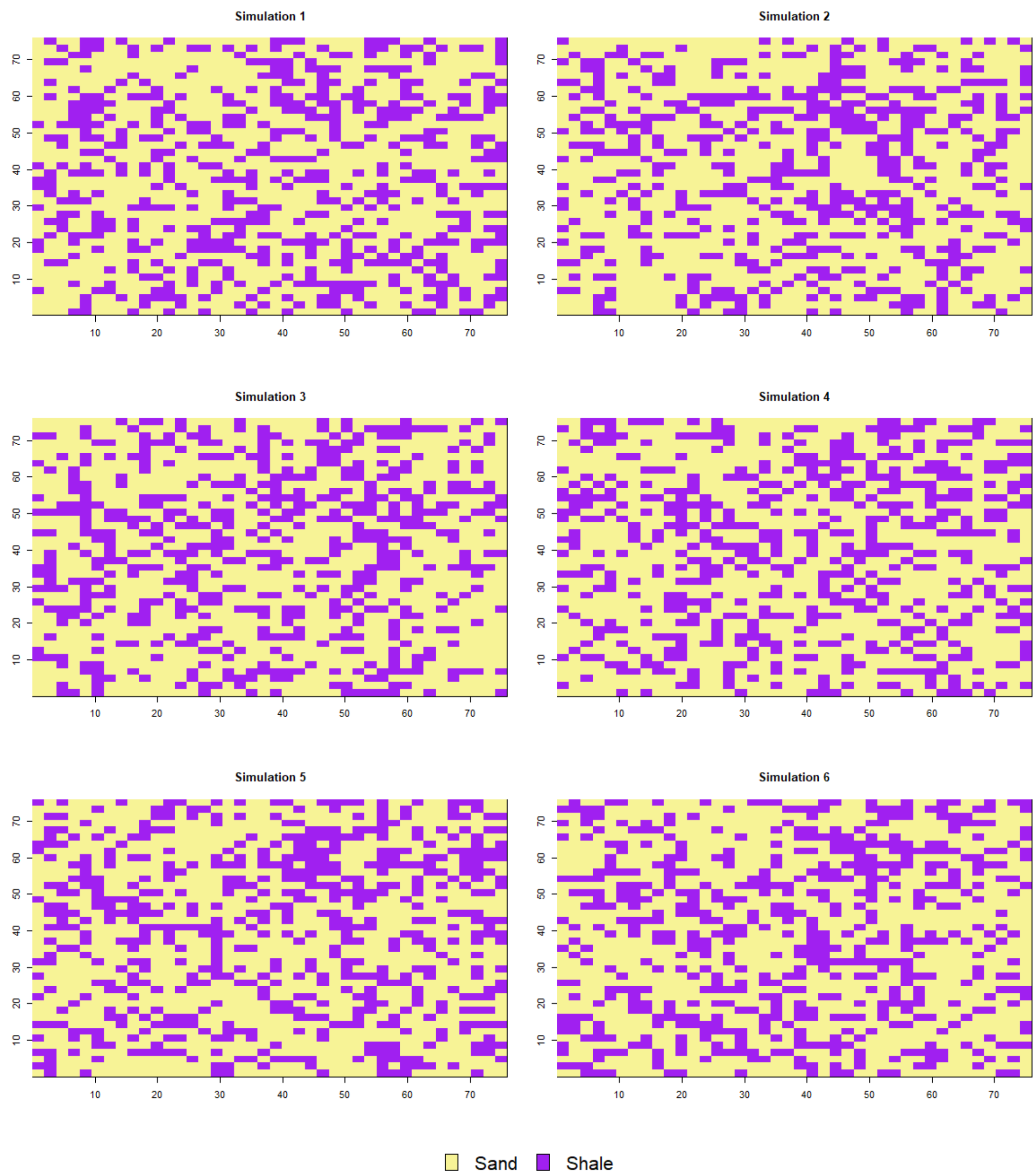

Figure 2: Display of six posterior realizations of L_D.

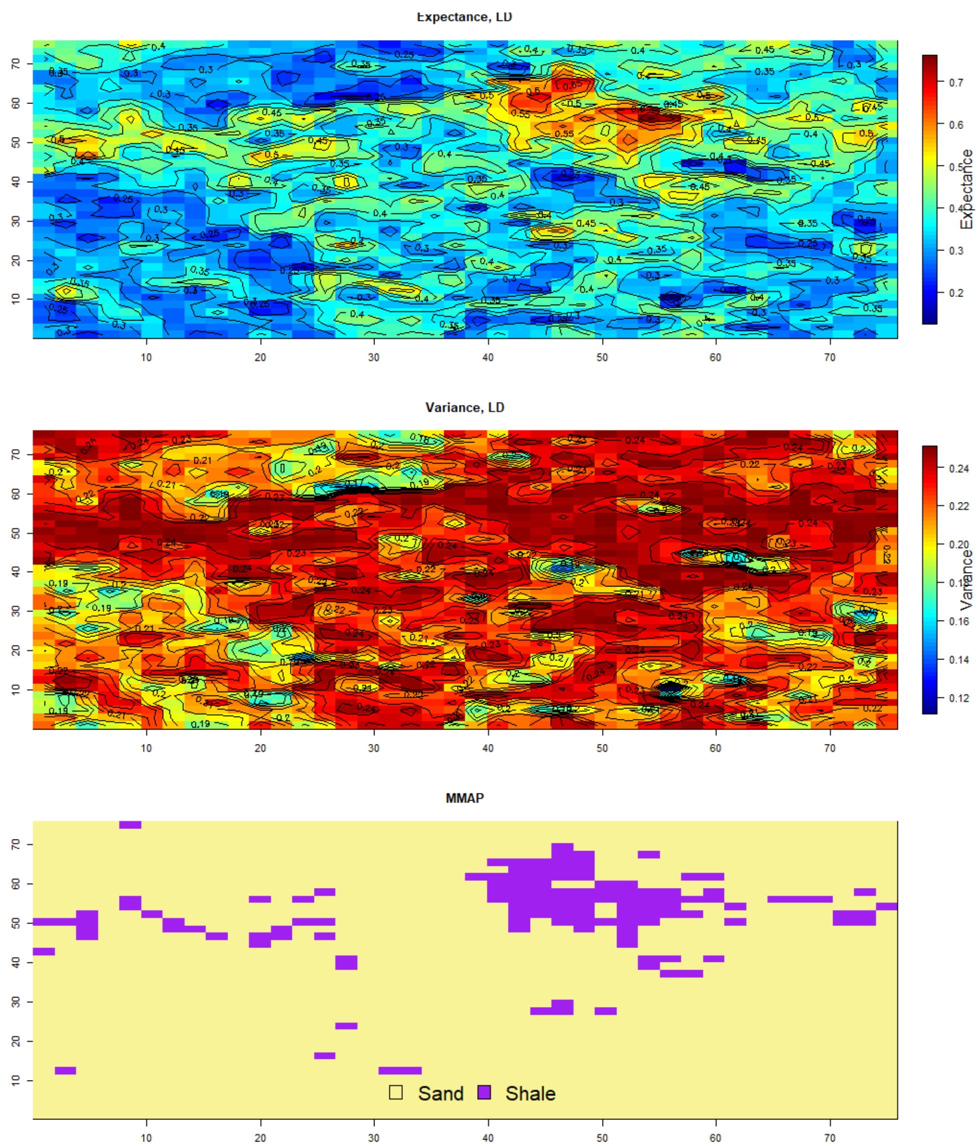

Figure 3: Display of six posterior realizations of L_D.

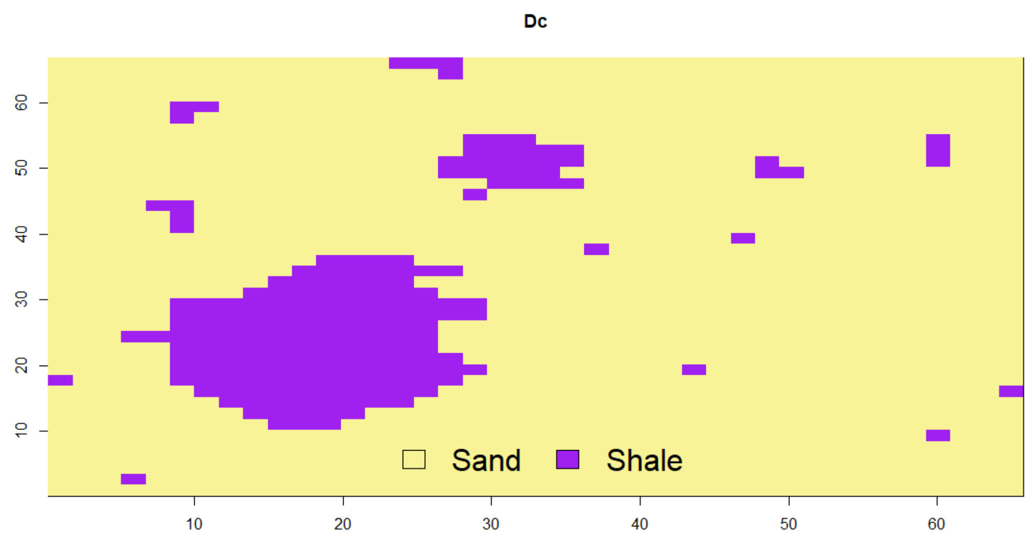

Figure 4: Display of observed data in D_c