Fully Bayesian binary Markov random field models: Prior specification and posterior simulation

Petter Arnesen[1] **and Håkon Tjelmeland**[1]
[1] Norwegian University of Science and Technology, Norway

Markov random fields (MRF) are frequently used as prior distributions in spatial statistics, see for instance Besag (1986) and Hurn et al. (2003). In this presentation we consider a binary MRF, $x = \{x_{ij}; (i,j) \in S\}$, defined on a rectangular lattice S, and assume the field to be stationary except for boundary effects. We assume the set of maximal cliques to consist of all $k \times l$ blocks of nodes, and the main topic of the presentation is how to build a prior distribution for the MRF model parameters in this situation. Clearly the number of free parameters becomes large even for quite small values of k and l, and it is therefore essential to define a prior which limit the effective number of parameters.

To define a prior for the model parameters we first need to formulate an identifiable parametric model. An easy alternative is to express the model in terms of interaction parameters,

$$p(x|\beta) = c(\beta) \exp \left\{ \sum_{\lambda \in \mathcal{L}} \beta^\lambda \prod_{(i,j)\in\lambda} x_{ij} \right\},$$

where $c(\beta)$ is a normalising constant, $\mathcal{L}$ is the set of all cliques, and β^λ is the interaction parameter for the clique λ. With one restriction on the interaction parameters, for example $\beta^\emptyset = 0$, this defines an identifiable model. However, the interpretation of the interaction parameters, β^λ, is not very clear and in particular the natural scale of β^λ depend on the number of elements in the clique λ. This complicates the definition of a prior for the β^λ's. We instead define a parametric form where the parameters have interpretation as potentials of clique configurations,

$$p(x|\phi) = c(\phi) \exp \left\{ \sum_{\Lambda \in \mathcal{L}_m} V_\Lambda(x_\Lambda) \right\} \quad \text{with} \quad V_\Lambda(\mathbf{1}_\Lambda^\lambda) = \phi_\Lambda^\lambda,$$

where $c(\phi)$ is a normalising constant, $\mathcal{L}_m$ is the set of all maximal cliques, $\mathbf{1}_\Lambda^\lambda$ is the colouring of the nodes in the maximal clique $\Lambda \in \mathcal{L}_m$ which assigns the value 1 to all nodes in $\lambda \subseteq \Lambda$ and 0 to the nodes in $\Lambda \setminus \lambda$. The ϕ_Λ^λ parameters have a much easier interpretation than the β^λ parameters and in particular it is reasonable to assume all the ϕ_Λ^λ parameters to be on the same scale. However, $p(x|\phi)$ is grossly overparameterised and the model is not identifiable. We identify a sufficient number of restrictions on the ϕ_Λ^λ parameters which makes the parametric model identifiable.

When defining a prior for the (restricted) ϕ_Λ^λ parameters we want to obtain a flexible prior model so that the model may adapt to the structure of any observed image. Therefore we do not put any absolute restrictions on the parameters other than the ones made to make the model identifiable. Instead we limit the effective number of parameters by assigning apriori discrete probabilities for events where groups of parameter values are exactly equal. The prior is thereby defined in two steps, first we define a prior for what group of parameter values that are exactly equal, and given such a grouping we put a prior on parameter values.

Assuming we have observed a binary image x we define a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm (Green, 1995) to sample from the associated posterior distribution. It should be noted that by a prior defined as discussed above the RJMCMC algorithm effectively serves as a model selection algorithm. When running the RJMCMC algorithm we have to cope with the computationally intractable normalising constant of the MRF, $c(\phi)$. Our approach is to adopt a previously defined approximation for binary MRFs (Tjelmeland and Austad, 2012).

We have explored our approach in a number of simulation examples. In particular we have generated simulated data from an independence model where the value in the variuos nodes are independent, and we have generated simulated data from an Ising model. In both cases, using a 2×2 maximal clique model, the posterior distribution gives a high probability to the grouping of the ϕ_Λ^λ parameters that corresponds to the model that generated the data. In a real data case, still assuming 2×2 maximal cliques, the posterior most probable grouping of the parameters had three groups, one parameter for the configuration where all nodes are zero, one parameter for the configuration where all nodes are one, and one common parameter for all the remaining clique configurations.

Keywords: Approximate inference; Ising Model; Markov random fields; Reversible jump MCMC.

References:

Besag, J. (1986). On the statistic analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48, 259-302.

Green, P. J. (1995). Reversible jump MCMC computation and Bayesian model determination.*Biometrika*, 82, 711-732.

Hurn, M., Husby, O., and Rue, H. (2003). A tutorial on image analysis. *In Spatial Statistics and Computational Methods, ed. J. Møller*, vol. 173 of Lecture Notes in Statistics, 87141. Springer Verlag.

Tjelmeland, H. and Austad, H. M. (2012). Exact and approximate recursive calculations for binary Markov random fields defined on graphs. *Journal of Computational and Graphical Statistics*, 21, 758-780.