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

Petter Arnesen and Håkon Tjelmeland
Norwegian University of Science and Technology, Trondheim, Norway

Abstract: Construct a prior distribution for the number of parameters and the parameter values in higher order interaction binary MRFs. Define an MCMC scheme to simulate from the posterior.

Theory

Likelihood-MRF

Consider an MRF x on a lattice $S = \{1, ..., nm\}$

$$p(x|\theta) = c(\theta) \exp\left(\sum_{\Lambda \in \mathcal{L}_m} U(x_\Lambda, \theta)\right),$$

where we assume the set of all maximal cliques $\mathcal{L}_m$ to be all $k \times l$ block of nodes on a torus. A naive parametrization of the the potential function is

$$U(x_\Lambda, \theta) = \sum_{y \in \{0,1\}^{k \times l}} \theta^y I(x_\Lambda = y) = \theta^{x_\Lambda},$$

however this is grossly overparametrized. One can established an identifiable parametrization by constraining some of the θ parameters to be equal. For instance when $k = l = 2$ the 16 configurations of size 2×2 can be defined to belong to the configuration sets

$$c_0 = \left\{ \binom{00}{00} \right\},\ c_1 = \left\{ \binom{10}{00}, \binom{01}{00}, \binom{00}{10}, \binom{00}{01} \right\},$$

$$c_2 = \left\{ \binom{11}{00}, \binom{00}{11} \right\}, c_3 = \left\{ \binom{10}{10}, \binom{01}{01} \right\},$$

$$c_4 = \left\{ \binom{10}{01} \right\},\ c_5 = \left\{ \binom{01}{10} \right\},\ c_6 = \left\{ \binom{11}{10} \right\},$$

$$c_7 = \left\{ \binom{11}{01} \right\},\ c_8 = \left\{ \binom{10}{11} \right\},\ c_9 = \left\{ \binom{01}{11} \right\},$$

$$c_{10} = \left\{ \binom{11}{11} \right\},$$

where each set is constrained to have the same parameter $\phi_0, ..., \phi_{10}$. To obtain a fully identifiable parametrization we need one more restriction, for instance a sum-to-zero constrain. The number of configurations and free parameters for some values of $k \times l$ is given below

$k \times l$	2^{kl}	N_{kl}
1×2	4	2
2×2	16	10
2×3	64	44
3×3	512	400
3×4	4096	3392
4×4	65536	57856

where N_{kl} is the number of free parameters in the identifiable parametrization.

Prior

Since $\phi_0, ..., \phi_{N_{kl}}$ are parameters on the same scale we reduce the number of parameters by assuming a positive prior probability for the event of groups of parameters to have exactly the same value.

In particular we let $z = \{(C_i, \varphi_i)\}_{i=1}^r$ where $C_1, ..., C_r$ is a partition of the set of all configuration sets, $C_i \neq \emptyset$, $C_i \cap C_j = \emptyset$, $C_1 \cup ... \cup C_r = \{c_0, ..., c_{N_{kl}}\}$, and where φ_i is the common parameter value for set C_i. We let r denote the number of groups, and define the prior

$$p(z) = p(\{C_1, ..., C_r\}) p(\{\varphi_1, ..., \varphi_r\}|r),$$

where $p(\{C_1, ..., C_r\}$ is defined by assuming $r \sim U(1, ..., N_{kl} + 1)$, and that given the number of groups the possible groupings are uniformly distributed. For the group parameters $\{\varphi_1, ..., \varphi_r\}$ we assume independently that $\varphi_i \sim N(0, \sigma^2)$ under the restriction $\sum_i \varphi_i = 0$.

MCMC proposals

A reversible jump MCMC algorithm is defined with proposals:

- Random walk for parameter values.
- Switching group membership.
- Creating/deleting groups.

Intractable normalizing constant

For MRF, exact calculations of the likelihood is limited by the intractable normalizing constant $c(\theta)$. We adopt an approximation to the MRF using pseudo-Boolean functions (Tjelmeland and Austad 2012), where the conditional distribution of one variable given all previous variables is allowed to depend on maximally ν previous variables.

Examples

We investigate two examples using 2×2 cliques, i.e. $k = l = 2$.

Ising model

We generated one realization from the Ising model

$$p(x|\omega) = \frac{1}{c(\omega)} \exp\left\{ -\omega \sum_{i \sim j} I(x_i \neq x_j) \right\},$$

with $\omega = 0.4$ to use as data. For this model it is possible to calculate that the correct grouping with 2×2 cliques must be $C_0 = \{c_0, c_{10}\}$, $C_1 = \{c_1, c_2, c_3, c_6, c_7, c_8, c_9\}$, and $C_2 = \{c_4, c_5\}$. The posterior estimate shows that this grouping is in fact the posterior most probable grouping with probability 0.94. The estimated probability of two configuration sets to be grouped together is shown below

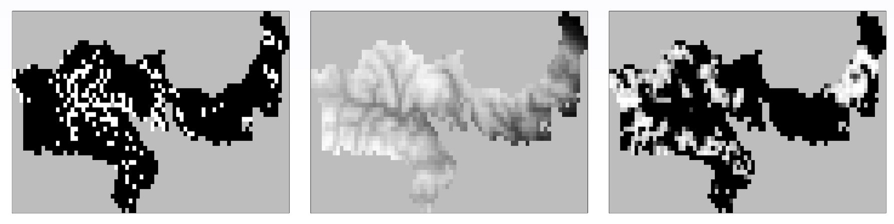
	c_0	c_{10}	c_1	c_2	c_3	c_6	c_7	c_8	c_9	c_4	c_5
c_0	1	1									
c_{10}	1	1									
c_1			1	0.96	0.97	0.97	0.97	0.97	0.96		
c_2			0.96	1	0.96	0.96	0.96	0.96	0.96		
c_3			0.97	0.96	1	0.97	0.96	0.96	0.95		
c_6			0.97	0.96	0.97	1	0.96	0.96	0.96		
c_7			0.97	0.96	0.96	0.96	1	0.97	0.97		
c_8			0.97	0.96	0.96	0.96	0.97	1	0.96		
c_9			0.96	0.96	0.95	0.96	0.97	0.96	1		
c_4										1	1
c_5										1	1

Ising model example: Estimated posterior probabilities for two configuration sets to be grouped together. The true grouping is shown in grey, and only probabilities larger than 0.05 are given. Note the permutation done to the ordering of c_i.

The parameter estimates for the posterior most probable grouping is given below

Parameter	Estimate	95 % cred.int.	True value
φ_0	0.385	(0.357,0.412)	0.400
φ_1	-0.001	(-0.038,0.036)	0.000
φ_2	-0.384	(-0.425,-0.320)	-0.400

Red deer census count data



Red deer example: Presence/absence of red deer, altitude, and mires.

Four covariates, altitude, mires, northing, and easting are included in this example. The estimated posterior most probable grouping becomes $C_0 = \{c_0\}$, $C_1 = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9\}$, and $C_2 = \{c_{10}\}$, with probability 0.33. Over 2500 different groupings are visited in this example, but except for the most probable grouping all other groupings have a probability less than 0.05. The estimated probability of two configuration sets to be grouped together is shown below

	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
c_0	1					0.12	0.06				
c_1		1	0.69	0.68	0.58	0.67	0.75	0.7	0.68	0.6	
c_2		0.69	1	0.8	0.69	0.62	0.64	0.7	0.69	0.65	
c_3		0.68	0.8	1	0.69	0.61	0.64	0.7	0.69	0.65	
c_4		0.58	0.69	0.69	1	0.55	0.58	0.64	0.64	0.69	0.05
c_5	0.12	0.67	0.62	0.61	0.55	1	0.62	0.61	0.61	0.57	
c_6	0.06	0.75	0.64	0.64	0.58	0.62	1	0.65	0.64	0.58	
c_7		0.7	0.7	0.7	0.64	0.61	0.65	1	0.67	0.63	
c_8		0.68	0.69	0.69	0.64	0.61	0.64	0.67	1	0.63	
c_9		0.6	0.65	0.65	0.69	0.57	0.58	0.63	0.63	1	0.06
c_{10}					0.05					0.06	1

Red deer example: Estimated posterior probabilities for two configuration sets to be grouped together. The true grouping is shown in grey, and only probabilities larger than 0.05 are given.

The parameter estimate for the most probable grouping is given below

Parameter	Estimate	95 % cred.int.
φ_0	1.323	(0.768,1.914)
φ_1	0.489	(-0.212,1.161)
φ_2	-1.812	(-3.338,-0.708)

Further work

Put a prior also on the clique types that is included in the MRF, and thereby remove the $k \times l$ assumption.