Scalable Inference for the Inverse Temperature of a Hidden Potts Model

Matt Moores

Department of Statistics
University of Warwick, UK

ACEMS International Workshop on Monte Carlo Methods for Spatial Stochastic Systems
July 21, 2015
Acknowledgements

Queensland University of Technology
- Kerrie Mengersen
- Tony Pettitt
- Chris Drovandi
- Clair Alston

University of Warwick & Université Paris Dauphine
- Christian Robert

Funded by an APA (Australian Department of Education & Training), QUT Postgraduate Award, UK EPSRC platform grant (ref: EP/L014165/1) and the EPSRC Network on Computational Statistics & Machine Learning (ref: EP/K009788/2)
Outline

1. Spatial Inference
   - Hidden Potts model

2. Monte Carlo methods
   - Exchange algorithm
   - Approximate Bayesian computation (ABC)
   - Indirect inference

3. Experimental Results
   - Simulation Study
   - Satellite Remote Sensing
   - Computed Tomography (CT)
Background

Image analysis often involves:

- Large datasets, with millions of pixels
- Multiple images with similar characteristics

For example: satellite remote sensing (Landsat, MODIS), medical imaging (CT scans, MRI)

<table>
<thead>
<tr>
<th>Number of pixels</th>
<th>Landsat (90m²/px)</th>
<th>CT slices (512×512)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^6$</td>
<td>0.06km²</td>
<td>...</td>
</tr>
<tr>
<td>$5^6$</td>
<td>14.06km²</td>
<td>0.1</td>
</tr>
<tr>
<td>$10^6$</td>
<td>900.00km²</td>
<td>3.8</td>
</tr>
<tr>
<td>$15^6$</td>
<td>10251.56km²</td>
<td>43.5</td>
</tr>
</tbody>
</table>
Computational cost is dominated by simulation of auxiliary variables (exchange algorithm) or pseudo-data (ABC)

(a) 2D images, $k = 3$

(b) 3D images, $k = 3$

hidden Markov random field

Joint distribution of observed pixel intensities $y_i \in y$ and latent labels $z_i \in z$:

$$
\Pr(y, z \mid \mu, \sigma^2, \beta) = \mathcal{L}(y \mid \mu, \sigma^2, z) \pi(z \mid \beta) \quad (1)
$$

Additive Gaussian noise:

$$
y_i \mid z_i = j \sim iid \mathcal{N} (\mu_j, \sigma_j^2) \quad (2)
$$

Potts model:

$$
\pi(z_i \mid z_{\setminus i}, \beta) = \frac{\exp \left\{ \beta \sum_{i \sim \ell} \delta(z_i, z_{\ell}) \right\}}{\sum_{j=1}^{k} \exp \left\{ \beta \sum_{i \sim \ell} \delta(j, z_{\ell}) \right\}} \quad (3)
$$

Potts (1952) *Proceedings of the Cambridge Philosophical Society* 48(1)
Inverse Temperature

(c) $\beta = 0.1$

(d) $\beta = 0.5$

(e) $\beta = 0.85$

(f) $\beta = 0.95$

(g) $\beta = 1.005$

(h) $\beta = 1.15$
The normalising constant has computational complexity $O(nk^n)$, since it involves a sum over all possible combinations of the labels $z \in \mathcal{Z}$:

$$C(\beta) = \sum_{z \in \mathcal{Z}} e^{\beta S(z)}$$  \hspace{1cm} (5)$$

$S(z)$ is the sufficient statistic of the Potts model:

$$S(z) = \sum_{i \sim \ell \in \mathcal{E}} \delta(z_i, z_\ell)$$  \hspace{1cm} (6)$$

where $\mathcal{E}$ is the set of all unique neighbour pairs.
Expectation of $S(z)$

(a) $n = 12 \& k \in \{2, 3, 4\}$

(b) $k = 3 \& n \in \{4, 6, 9, 12\}$

**Figure:** Distribution of $\mathbb{E}_{z|\beta}[S(z)]$
Standard deviation of $S(z)$

(a) $n = 12$ & $k \in \{2, 3, 4\}$

(b) $k = 3$ & $n \in \{4, 6, 9, 12\}$

Figure: Distribution of $\sigma_{z|\beta}[S(z)]$
Distribution of $S(z)$

In terms of the normalising constant:

$$
\mathbb{E}_{z|\beta}[S(z)] = \frac{d}{d\beta} \log\{C(\beta)\} \quad (7)
$$

Score function:

$$
\frac{d}{d\beta} \log\{p(S(z) | \beta)\} = S(z) - \mathbb{E}_{z|\beta}[S(z)] \quad (8)
$$

Fisher information:

$$
\mathcal{I}(\beta) = \mathbb{E}_{z|\beta} \left[ \left( \frac{d}{d\beta} \log\{p(S(z) | \beta)\} \right)^2 \right] \quad (9)
$$

$$
= \mathbb{E}_{z|\beta} \left[ \frac{d^2}{d\beta^2} \log\{C(\beta)\} \right] \quad (10)
$$
Spatial Inference

Monte Carlo methods

Experimental Results

Conclusion

Special cases

When $\beta = 0$ the labels are independent, hence:

$$E_0 = E_{z|\beta=0}[S(z)] = \frac{1}{k}|\mathcal{E}|$$

$$V_0 = V_{z|\beta=0}[S(z)] = |\mathcal{E}| \left( \frac{1}{k} \right) \left( 1 - \frac{1}{k} \right)$$

where $|\mathcal{E}|$ is the total number of edges in the image lattice.

As $\beta \to \infty$,

$$E_\infty = \lim_{\beta \to \infty} E_{z|\beta}[S(z)] = |\mathcal{E}|$$

$$V_\infty = \lim_{\beta \to \infty} V_{z|\beta}[S(z)] = 0$$

For a 2D lattice with asymptotically large $n$:

$$\beta_{crit} = \log\{1 + \sqrt{k}\}$$

$$V_{max} = V_{z|\beta=\beta_{crit}}[S(z)] = \frac{2}{\pi} |\mathcal{E}| \log\{|\mathcal{E}|\}$$

Pickard (1987) *JASA* 82(397)

**Exchange Algorithm**

**Algorithm 1** Exchange Algorithm

1: for all iterations $t = 1, \ldots, T$ do
2: Draw proposed parameter value $\beta' \sim q(\beta'|\beta_{t-1})$
3: Generate $w|\beta'$ by (perfect) sampling from Eq. (3)
4: Calculate the Radon-Nikodym derivative:

$$
\rho = \frac{q(\beta_{t-1}|\beta')\pi(\beta')C(\beta_{t-1})e^{\beta'S(z)}C(\beta')e^{\beta_{t-1}S(w)}}{q(\beta'|\beta_{t-1})\pi(\beta_{t-1})C(\beta')e^{\beta_{t-1}S(z)}C(\beta_{t-1})e^{\beta'S(w)}}
$$

5: Draw $u \sim \text{Uniform}[0, 1]$
6: if $u < \min(1, \rho)$ then
7: \hspace{1cm} $\beta_t \leftarrow \beta'$ else $\beta_t \leftarrow \beta_{t-1}$
8: end if
9: end for

Murray, Ghahramani & MacKay (2006) *Proc. 22nd Conf. UAI*
Approximate Bayesian Computation

Algorithm 2 ABC-MCMC

1: for all iterations $t = 1, \ldots, T$ do
2: Draw proposed parameter value $\beta' \sim q(\beta' | \beta_{t-1})$
3: Generate $w | \beta'$ by sampling from Eq. (3)
4: Draw $u \sim \text{Uniform}[0, 1]$
5: if $u < \min \left(1, \frac{\pi(\beta')q(\beta_{t-1} | \beta')}{\pi(\beta_{t-1})q(\beta' | \beta_{t-1})} \right)$ and $\|S(w) - S(z)\| < \epsilon$ then
6: $\beta_t \leftarrow \beta'$ else $\beta_t \leftarrow \beta_{t-1}$
7: end if
8: end for

Marjoram, Molitor, Plagnol & Tavaré (2003) PNAS 100(26)
Grelaud, Robert, Marin, Rodolphe & Taly (2009) Bayesian Analysis 4(2)
Precomputation Step

The distribution of the summary statistics \( f(S(w) | \beta) \) is independent of the observed data \( y \) and the labels \( z \)

- By simulating pseudo-data for values of \( \beta \), we can create a binding function \( \phi(\beta) \) for an auxiliary model \( f_A(S(w) | \phi(\beta)) \)
- This binding function can be reused across multiple datasets, amortising its computational cost

By replacing \( S(w) \) with our auxiliary model, we avoid the need to simulate pseudo-data or auxiliary variables during model fitting.

Piecewise linear model

Figure: Binding functions for $S(w) \mid \beta$ with $n = 5^6, k = 3$
Parametric auxiliary model for $V_{z|\beta}[S(z)]$

$$\hat{\phi}_{\sigma^2}(\beta) = \begin{cases} V_0 + (V_{max} - V_0)e^{-\phi_1\sqrt{\beta_{crit} - \beta}} & : 0 \leq \beta < \beta_{crit} \\ V_{max}e^{-\phi_2\sqrt{\beta - \beta_{crit}}} & : \beta \geq \beta_{crit} \end{cases}$$

(11)
The binding function for the expectation is available in closed form:

\[ \hat{\phi}_\mu(\beta) = \begin{cases} 
\mathbb{E}_0 + \beta V_0 + \int_0^\beta (V_{max} - V_0) e^{-\phi_1 \sqrt{\beta_{crit} - \beta}} d\beta & : 0 \leq \beta < \beta_{crit} \\
\mathbb{E}_{\beta_{crit}} + \int_{\beta_{crit}}^\beta V_{max} e^{-\phi_2 \sqrt{\beta - \beta_{crit}}} d\beta & : \beta \geq \beta_{crit}
\end{cases} \]
Algorithm 3 Bayesian Indirect Inference

1: Generate \( w_s | \beta_s \) for sample points \( \beta_s \), where \( s = 1, \ldots, S \)
2: Fit the binding functions \( \hat{\phi}_{\sigma^2}(\beta) \) & \( \hat{\phi}_{\mu}(\beta) \)
3: **for all** iterations \( t = 1, \ldots, T \) **do**
4: Draw proposed parameter value \( \beta' \sim q(\beta'|\beta_{t-1}) \)
5: Approximate the Radon-Nikodym derivative:
\[
\rho = \frac{q(\beta_{t-1}|\beta')\pi(\beta') f_A \left( S(z) \mid \hat{\phi}_{\mu}(\beta'), \hat{\phi}_{\sigma^2}(\beta') \right)}{q(\beta'|\beta_{t-1})\pi(\beta_{t-1}) f_A \left( S(z) \mid \hat{\phi}_{\mu}(\beta_{t-1}), \hat{\phi}_{\sigma^2}(\beta_{t-1}) \right)}
\]
6: Draw \( u \sim \text{Uniform}[0, 1] \)
7: **if** \( u < \min(1, \rho) \) **then**
8: \( \beta_t \leftarrow \beta' \) **else** \( \beta_t \leftarrow \beta_{t-1} \)
9: **end if**
10: **end for**
Simulation Study

20 images, $n = 125 \times 125$, $k = 3$:

- $\beta \sim \mathcal{U}(0, 1.307)$
- $\mathbf{z} \sim f(\cdot | \beta)$ using 2000 iterations of Swendsen-Wang
- $\mu_j \in \{\mathcal{N}(-0.15, 0.05^2), \mathcal{N}(0.05, 0.05^2), \mathcal{N}(0.25, 0.05^2)\}$
- $\frac{1}{\sigma_j^2} \sim \Gamma\left(\frac{3}{2}, \frac{0.015}{2}\right)$

Comparison of 3 algorithms:
- **exchange** approximate exchange algorithm
  (using 500 iters of Gibbs sampling)
- **ABC-MCMC** with pseudo-data
  **auxiliary** indirect inference using $f_A(S(\mathbf{w}) | \phi(\beta))$

Feng & Tierney (2014) PottsUtils, *R package version 0.3-2*
Posterior Samples

(a) exchange

(b) ABC-MCMC

(c) auxiliary

(d) error
Satellite image of southwest Brisbane

Figure: Normalised difference vegetation index (NDVI)
Piecewise linear model

(a) $\hat{\phi}_\mu(\beta)$

(b) $\hat{\phi}_\sigma(\beta)$

Figure: Linear interpolation for $n = 978380, k = 6$
Parametric auxiliary model

Figure: Binding functions for $n = 978380, k = 6$
Results

<table>
<thead>
<tr>
<th>Method</th>
<th>95% CI for $\beta$</th>
<th>Iterations</th>
<th>Elapsed</th>
<th>CPU time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange alg.</td>
<td>[1.275; 1.278]</td>
<td>5K+5K</td>
<td>50.8h</td>
<td>399.5h</td>
</tr>
<tr>
<td>ABC-MCMC</td>
<td>[1.316; 1.323]</td>
<td>5K+5K</td>
<td>50.6h</td>
<td>398.2h</td>
</tr>
<tr>
<td>Piecewise linear</td>
<td>[1.249; 1.255]</td>
<td>5K+55K</td>
<td>1.3h</td>
<td>10.0h</td>
</tr>
<tr>
<td>Parametric aux.</td>
<td>[1.259; 1.265]</td>
<td>5K+55K</td>
<td>1.8h</td>
<td>7.6h</td>
</tr>
</tbody>
</table>

Table: Results for the satellite image of Brisbane, Australia.

Precomputation of $f_A(S(w) \mid \phi(\beta))$ took 13h 23m for 987 values of $\beta$. 
Electron Density phantom

(a) CIRS Model 062 ED phantom

(b) CT scan
Figure: Linear interpolation for $n = 898366, k = 9$
Parametric auxiliary model

(a) $\hat{\phi}_\mu(\beta)$

(b) $\hat{\phi}_\sigma(\beta)$

Figure: Binding functions for $n = 898366$, $k = 9$
Results

(a) Pooled posterior for $\beta$

(b) Elapsed runtime (hours)

Figure: Results for 28 CT scans of the ED phantom
Summary

Scalability of Bayesian computation for intractable likelihoods can be improved by pre-computing an auxiliary model $f_A(S(w) \mid \phi(\beta))$:

- Pre-computation took 1.4 hours on a 16 core Xeon server for 987 values of $\beta$ with 15,625 pixels (13.4 hours for 978,380 pixels).
- Average runtime for model fitting improved from 107 hours (exchange algorithm) or 115 hours (ABC-MCMC) to only 4 hours using the parametric auxiliary model.

The tractable, parametric approximation for $\hat{\phi}_{\sigma^2}(\beta)$ could be used to design more efficient MCMC proposals, as well as to select the design points for $\beta$.

This method could be extended to multivariate applications, such as estimating both $\beta$ and $k$ for the hidden Potts model, or estimating $\theta$ for an exponential random graph model (ERGM).
For Further Reading I

M. Moores, A. N. Pettitt & K. Mengersen
Scalable Bayesian inference for the inverse temperature of a hidden Potts model.

M. Moores, C. C. Drovandi, K. Mengersen & C. P. Robert
Pre-processing for approximate Bayesian computation in image analysis.

M. Moores & K. Mengersen
Bayesian approaches to spatial inference: modelling and computational challenges and solutions.

M. Moores & K. Mengersen
bayesImageS: Bayesian methods for image segmentation using a hidden Potts model.
R package version 0.3-1
https://researchdatafinder.qut.edu.au/display/n11957
For Further Reading II

- C. C. Drovandi, A. N. Pettitt & A. Lee

- C. C. Drovandi, A. N. Pettitt & M. J. Faddy

- R. G. Everitt


- D. K. Pickard
Appendix

For Further Reading III

D. Feng & L. Tierney
PottsUtils: Utility Functions of the Potts Models.
R package version 0.3-2
http://CRAN.R-project.org/package=PottsUtils

I. Murray, Z. Ghahramani & D. J. C. MacKay
MCMC for Doubly-intractable Distributions.

P. Marjoram, J. Molitor, V. Plagnol & S. Tavaré
Markov chain Monte Carlo without likelihoods.

R. B. Potts
Some generalized order-disorder transformations.

R. H. Swendsen & J.-S. Wang
Nonuniversal critical dynamics in Monte Carlo simulations.