Pre-processing for Gibbs Random Fields

CASI 2015

Aidan Boland, Nial Friel

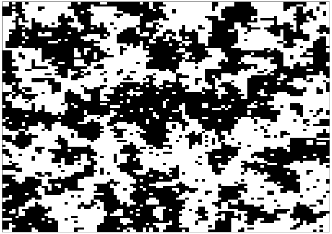
12/5/15

Outline

- Short overview of Gibbs random fields.
- Outline current techniques and their pitfalls.
- Introduce an approach to improving the computational cost of analysis.

What is a Gibbs random field?

- Type of graphical model.



- Set of nodes with corresponding random variables (y).
- $y_i \in [-1, 1]$

- Our interest is the posterior

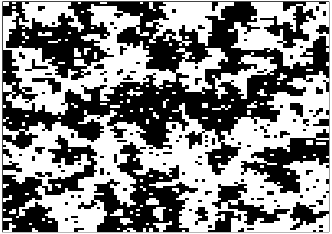
  posterior$\propto$ likelihood $\times$ prior

  $\pi(\theta|y) \propto f(y|\theta)\pi(\theta)$

- However the likelihood is intractable for GRF's

What is a Gibbs random field?

- Type of graphical model.



- Set of nodes with corresponding random variables (y).
- $y_i \in [-1, 1]$

- Our interest is the posterior

  posterior$\propto$ likelihood $\times$ prior

  $$\pi(\theta|y) \propto f(y|\theta)\pi(\theta)$$

- However the likelihood is intractable for GRF's

Insight

Intractability of GRF's

- The likelihood function is intractable
  $$f(y|\theta) = \frac{q(y|\theta)}{Z(\theta)} = \frac{\exp(\theta^T s(y))}{Z(\theta)},$$
  - θ tells us about the properties of the GRF.
  - $s(y)$ are summary statistics for the graph.
- $Z(\theta) = \sum_y \exp(\theta^T s(y))$.
- $Z(\theta)$ is intractable for all but trivial graphs.
  - 16×16 binary graph, $2^{16\times16} = 1.15 \times 10^{77}$ calculations.
  - 30×30 binary graph, $2^{30\times30} = 8.453 \times 10^{270}$ calculations!

Examples of intractable methods

- Finance: Stochastic volatility model.
- Ecology: Spatial models for biodiversity.
- Genetics: Phylogenetic trees.
- Sociology: Social Networks
- Epidemiology: Stochastic models for disease transmission
- Machine Learning: Deep architecture

Current Methods

- Want to sample from the posterior $\pi(\theta|y)$
- Can use MCMC sampling.
- Metropolis-Hastings.
  - Propose to move to $\theta' \sim h(\cdot|\theta)$,
  - Accept move with probability α,

$$\begin{aligned}\alpha &= \min\left(1, \frac{f(y|\theta')}{f(y|\theta)} \frac{\pi(\theta')}{\pi(\theta)} \frac{h(\theta'|\theta)}{h(\theta|\theta')}\right) \\ &= \min\left(1, \frac{q(y|\theta')}{q(y|\theta)} \frac{Z(\theta)}{Z(\theta')} \frac{\pi(\theta')h(\theta'|\theta)}{\pi(\theta)h(\theta|\theta')}\right)\end{aligned}$$

Current Methods (Exchange Algorithm)

- Sample instead from the augmented distribution

$$\pi(\theta', y', \theta | y) \propto f(y|\theta)\pi(\theta)h(\theta'|\theta)f(y'|\theta')$$

  - Sample $y' \sim f(\cdot|\theta')$
  - Acceptance ratio becomes

$$\alpha = \min\left(1, \frac{q(y|\theta')}{q(y|\theta)} \frac{\cancel{Z(\theta)}}{\cancel{Z(\theta')}} \frac{\pi(\theta')h(\theta'|\theta)}{\pi(\theta)h(\theta|\theta')} \frac{q(y'|\theta)}{q(y'|\theta')} \frac{\cancel{Z(\theta)}}{\cancel{Z(\theta')}}\right)$$

$$= \min\left(1, \frac{q(y|\theta')}{q(y|\theta)} \frac{\pi(\theta')h(\theta'|\theta)}{\pi(\theta)h(\theta|\theta')} \frac{q(y'|\theta)}{q(y'|\theta')}\right)$$

Current Methods (ABC)

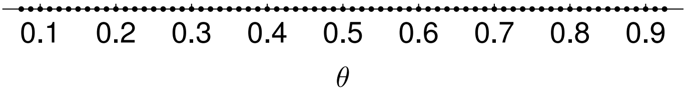
- Approximate Bayesian Computation
- Only requires the ability to sample from the likelihood
  - Sample $y' \sim f(\cdot|\theta')$
  - Accept θ' if $\rho(s(y) - s(y')) < \varepsilon$
  - Where ρ is some distance function.

- ABC MCMC
- Use acceptance ratio,

$$\alpha = \min\left(1, \frac{\pi(\theta')h(\theta'|\theta)}{\pi(\theta)h(\theta|\theta')} \mathbb{1}[\rho(s(y) - s(y')) < \varepsilon]\right)$$

Pre-processing

- Computationally expensive to sample from $f(y|\theta)$
- Prior to algorithm, sample y' at fixed set of θ values (θ_{pre}).

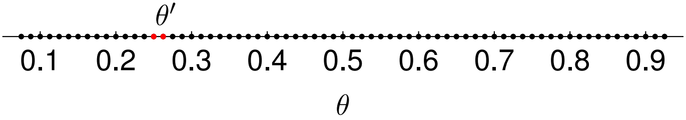


- When a sample is required, can interpolate from these pre sampled values.

$$s_{\theta'}(y') \sim N\left(\mathbb{E}[s_{\theta'_{pre}}(y)], Var(s_{\theta'_{pre}}(y))\right).$$

Pre-processing

- Computationally expensive to sample from $f(y|\theta)$
- Prior to algorithm, sample y' at fixed set of θ values (θ_{pre}).



- When a sample is required, can interpolate from these pre sampled values.

$$s_{\theta'}(y') \sim N\left(\mathbb{E}[s_{\theta'_{pre}}(y)], Var(s_{\theta'_{pre}}(y))\right).$$

Estimating the normalising ratio

- Importance sampling,

$$\frac{Z(\theta)}{Z(\theta')} = \mathbb{E}_{f(y'|\theta')} \frac{q(y'|\theta)}{q(y'|\theta')} \approx \frac{1}{N} \sum_{i=1}^{N} \frac{q(y_i'|\theta)}{q(y_i'|\theta')}$$

- Numerical integration,

$$\log \left[\frac{Z(\theta)}{Z(\theta')} \right] = \int_{\theta'}^{\theta} \mathbb{E}_{y|\theta}[s(y)] \approx$$

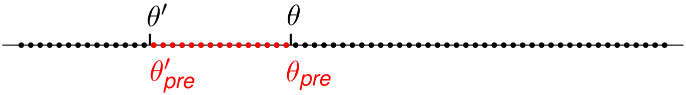
$$\sum_{i=a}^{b} \Bigg\{ (\theta_{i+1} - \theta_i) \frac{\mathbb{E}_{\theta_i}[s(y)] - \mathbb{E}_{\theta_{i+1}}[s(y)]}{2} - \frac{(\theta_{i+1} - \theta_i)^2}{12}$$

$$\Big(V_{\theta_i}(s(y)) - V_{\theta_{i+1}}(s(y)) \Big) \Bigg\}$$

UCD Dublin Insight

Estimating the normalising ratio

- Propose $\theta' \sim h(\cdot|\theta)$.
- Find nearest pre computed values,
  - $\theta \approx \theta_{pre}$
  - $\theta' \approx \theta'_{pre}$.
- Approximate $\dfrac{Z(\theta)}{Z(\theta')} \approx \widehat{\dfrac{Z(\theta_{pre})}{Z(\theta'_{pre})}}$.
- Where $\widehat{\dfrac{Z(\theta_{pre})}{Z(\theta'_{pre})}}$ is either the importance sampling or numerical integration estimate.



Correction

- We can use the pre computed values to improve the accuracy of the estimate.

$$\frac{Z(\theta_{pre})}{Z(\theta)} \approx \frac{\widehat{Z(\theta_{pre})}}{Z(\theta)} = \frac{1}{N}\sum_{i=1}^{N}\frac{q(y_i'|\theta_{pre})}{q(y_i'|\theta)}$$

- Where y_i' is the i th pre computed sample at θ_{pre}.

- Finally $\frac{Z(\theta)}{Z(\theta')} \approx \frac{\widehat{Z(\theta'_{pre})}}{Z(\theta')} \times \frac{\widehat{\widehat{Z(\theta_{pre})}}}{Z(\theta'_{pre})} \times \frac{\widehat{Z(\theta)}}{Z(\theta_{pre})}$

Theoretical Guarantee

- Want $|\alpha(\theta,\theta') - \hat{\alpha}(\theta,\theta',y')| \leq \delta(\theta,\theta')$
- Letting $\hat{R}_i = \widehat{\frac{Z(\theta_{i+1})}{Z(\theta_i)}}$.
- We get the following bound

$$\mathbb{P}\left(\left|\prod_{i=1}^{M}\hat{R}_i - \prod_{i=1}^{M}\mathbb{E}\left[\hat{R}_i\right]\right| \leq \gamma\right) \geq 1-\varepsilon$$

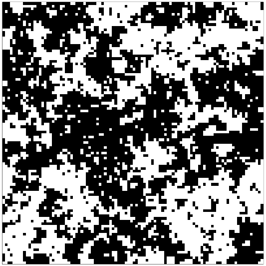
with $\gamma = \alpha(1+(M-1)MK^{M-1})$

and $\alpha = (b-a)\sqrt{\frac{1}{2N}\log\left(\frac{2M}{\varepsilon}\right)}$.

- where
  - M is the number of pre processed θ's between θ and θ'.
  - N is the number of samples at each θ_{pre}
  - K is the upper bound on the ratio of likelihoods at any two consequtive values of θ_{pre}.

Ising

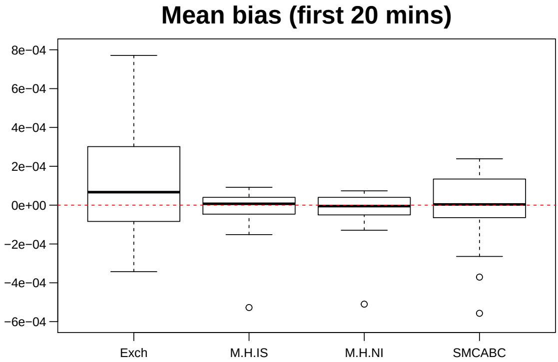
- Single parameter model.
- Defined on a rectangular lattice.
- Models spatial distribution of binary variables $(-1, 1)$.
- $f(y|\theta) = \frac{1}{Z(\theta)} \exp(\theta s(y))$
  - $s(y) = \sum_{j=1}^{N} \sum_{i \sim j} y_i y_j$
  - $i \sim j$ denotes that i and j are neighbours.
  - θ determines the degree of association between neighbours.



Ising example

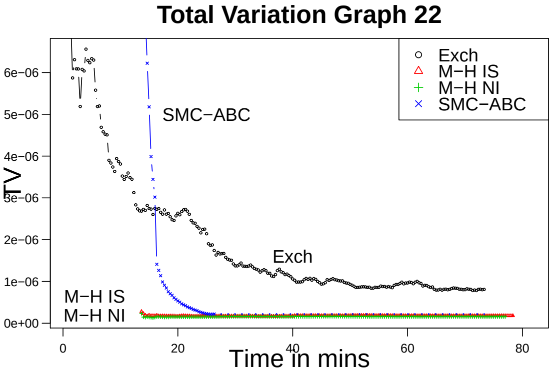
Simulation Study

- Simulated 24 80×80 Ising graphs.
- Ran an exchange algorithm for 24 hours to use as a 'ground truth'.
- Pre computation took 16 minutes per graph.



Simulation Study

- Simulated 24 graphs of size 80 $\times$ 80.
- Ran an exchange algorithm for 24 hours to use as a ‘ground truth’.
- Pre computation took 16 minutes per graph.



END

- Any questions?

Funded under the programme for Research in Third-level Institutions and co-funded under the European Regional Development fund.

Investing In Your Future