Bayesian inference for doubly-intractable distributions

Anne-Marie Lyne

anne-marie.lyne@curie.fr

April, 2016

Joint Work

- On Russian Roulette Estimates for Bayesian Inference with Doubly-Intractable Likelihoods

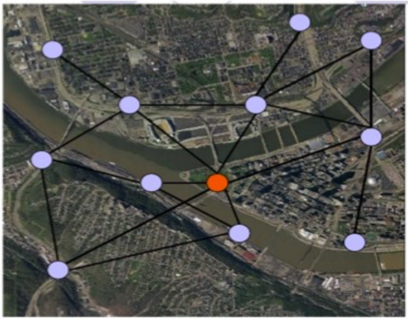
- *Anne-Marie Lyne, Mark Girolami, Yves Atchade, Heiko Strathmann, Daniel Simpson*

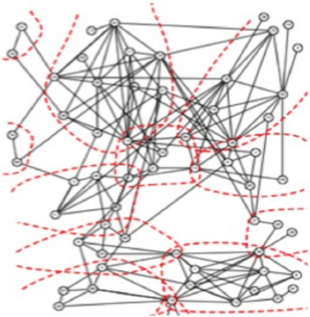
- Statist. Sci. Volume 30, Number 4 (2015)

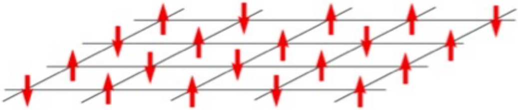
Talk overview

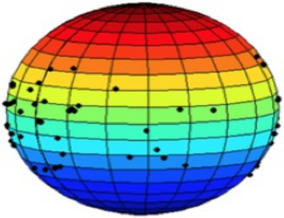
1. Doubly intractable models
2. Current Bayesian approaches
3. Our approach using Russian roulette sampling
4. Results

Motivation: Doubly-intractable models









Motivation: Exponential Random Graph Models

- Used extensively in the social networks community

Motivation: Exponential Random Graph Models

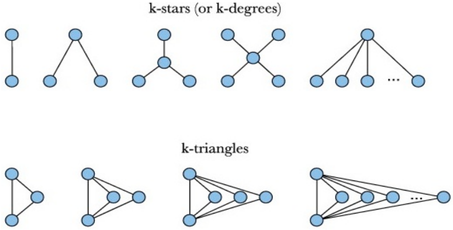
- Used extensively in the social networks community
- View the observed network as one realisation of a random variable

Motivation: Exponential Random Graph Models

- Used extensively in the social networks community
- View the observed network as one realisation of a random variable
- The probability of observing a given graph is dependent on certain 'local' graph properties

Motivation: Exponential Random Graph Models

- Used extensively in the social networks community
- View the observed network as one realisation of a random variable
- The probability of observing a given graph is dependent on certain 'local' graph properties
- For example the edge density, the number of triangles or k-stars



Motivation: Modelling social networks

$$\mathcal{P}(\mathbf{Y} = \mathbf{y}) = \mathcal{Z}(\boldsymbol{\theta})^{-1} \exp\left(\sum_k \theta_k g_k(\mathbf{y})\right)$$

- $\mathbf{g}(\mathbf{y})$ is a vector of K graph statistics
- θ is a K-dimensional parameter indicating the 'importance' of each graph statistic

Motivation: Modelling social networks

$$\mathcal{P}(\mathbf{Y} = \mathbf{y}) = \mathcal{Z}(\boldsymbol{\theta})^{-1} \exp\left(\sum_k \theta_k g_k(\mathbf{y})\right)$$

- $\mathbf{g}(\mathbf{y})$ is a vector of K graph statistics
- $\boldsymbol{\theta}$ is a K-dimensional parameter indicating the ‘importance’ of each graph statistic
- (Intractable) partition function or normalising term

$$\mathcal{Z}(\boldsymbol{\theta}) = \sum_{\mathbf{y} \in \mathcal{Y}} \exp\left(\sum_k \theta_k g_k(\mathbf{y})\right)$$

Parameter inference for doubly-intractable models

- Expectations with respect to the posterior distribution

$$\mathrm{E}_\pi[\phi(\boldsymbol{\theta})] = \int_\Theta \phi(\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta} \approx \frac{1}{N}\sum_{k=1}^{N}\phi(\boldsymbol{\theta}_k) \quad \boldsymbol{\theta}_k \sim \pi(\boldsymbol{\theta}|\mathbf{y})$$

- Simplest function of interest

$$\mathrm{E}_\pi[\boldsymbol{\theta}] = \int_\Theta \boldsymbol{\theta}\, \pi(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta} \approx \frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\theta}_k \quad \boldsymbol{\theta}_k \sim \pi(\boldsymbol{\theta}|\mathbf{y})$$

- But need to sample from the posterior distribution...

The Metropolis-Hastings algorithm

To draw samples from a distribution $\pi(\theta)$:

Choose an initial θ_0, define a proposal distribution $q(\theta, \cdot)$, set n = 0.

Iterate the following for $n = 0 \ldots N_{\text{iters}}$

1. Propose new parameter value, θ', from $q(\theta_n, \cdot)$
2. Set $\theta_{n+1} = \theta'$ with probability, $\alpha(\theta_n, \theta')$, else $\theta_{n+1} = \theta_n$

$$\alpha(\theta_n, \theta') = \min\left[1, \frac{\pi(\theta')q(\theta', \theta_n)}{\pi(\theta_n)q(\theta_n, \theta')}\right]$$

3. $n = n + 1$

Doubly-intractable distributions

- Unfortunately ERGMs are example of ‘doubly-intractable’ distribution:

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\mathbf{y})} = \frac{f(\mathbf{y};\boldsymbol{\theta})}{\mathcal{Z}(\boldsymbol{\theta})}\pi(\boldsymbol{\theta})\Big/ p(\mathbf{y})$$

Doubly-intractable distributions

- Unfortunately ERGMs are example of 'doubly-intractable' distribution:

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\mathbf{y})} = \frac{f(\mathbf{y};\boldsymbol{\theta})}{\mathcal{Z}(\boldsymbol{\theta})}\pi(\boldsymbol{\theta})\Big/p(\mathbf{y})$$

- Partition function or normalising term, $\mathcal{Z}(\boldsymbol{\theta})$, is intractable and function of parameters

$$\alpha(\boldsymbol{\theta},\boldsymbol{\theta}') = \min\left(1, \frac{q(\boldsymbol{\theta}',\boldsymbol{\theta})\pi(\boldsymbol{\theta}')f(\mathbf{y};\boldsymbol{\theta}')\mathcal{Z}(\boldsymbol{\theta})}{q(\boldsymbol{\theta},\boldsymbol{\theta}')\pi(\boldsymbol{\theta})f(\mathbf{y};\boldsymbol{\theta})\mathcal{Z}(\boldsymbol{\theta}')}\right)$$

Doubly-intractable distributions

- Unfortunately ERGMs are example of 'doubly-intractable' distribution:

$$\pi(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)\pi(\theta)}{p(\mathbf{y})} = \frac{f(\mathbf{y};\theta)}{\mathcal{Z}(\theta)}\pi(\theta)\Big/p(\mathbf{y})$$

- Partition function or normalising term, $\mathcal{Z}(\theta)$, is intractable and function of parameters

$$\alpha(\theta,\theta') = \min\left(1, \frac{q(\theta',\theta)\pi(\theta')f(\mathbf{y};\theta')\mathcal{Z}(\theta)}{q(\theta,\theta')\pi(\theta)f(\mathbf{y};\theta)\mathcal{Z}(\theta')}\right)$$

- As well as ERGMs, lots of other examples... (Ising and Potts models, spatial models, phylogenetic models)

Current Bayesian approaches

- Approaches which use some kind of approximation: pseudo-likelihoods

- Exact-approximate MCMC approaches:
  - auxiliary variable methods such as Exchange algorithm (requires perfect sample if implemented correctly)
  - pseudo-marginal (requires unbiased estimate of likelihood)

$$\alpha(\theta_n, \theta') = \min\left[1, \frac{\hat{\pi}(\theta')q(\theta', \theta_n)}{\hat{\pi}(\theta_n)q(\theta_n, \theta')}\right]$$

The Exchange algorithm (Murray et al 2004 and Møller et al 2004)

- Expand the state space of our target (posterior) distribution to

$$p(\mathbf{x}, \theta, \theta' | \mathbf{y}) = \frac{f(\mathbf{y}; \boldsymbol{\theta})}{\mathcal{Z}(\boldsymbol{\theta})} \pi(\theta) q(\theta, \theta') \frac{f(\mathbf{x}; \boldsymbol{\theta}')}{\mathcal{Z}(\boldsymbol{\theta}')} \Big/ p(\mathbf{y})$$

The Exchange algorithm (Murray et al 2004 and Møller et al 2004)

- Expand the state space of our target (posterior) distribution to

$$p(\mathbf{x}, \theta, \theta' | \mathbf{y}) = \frac{f(\mathbf{y}; \boldsymbol{\theta})}{\mathcal{Z}(\boldsymbol{\theta})} \pi(\theta) q(\theta, \theta') \frac{f(\mathbf{x}; \boldsymbol{\theta}')}{\mathcal{Z}(\boldsymbol{\theta}')} \Big/ p(\mathbf{y})$$

  - Gibbs sample $q(\theta, \theta') \frac{f(\mathbf{x}; \boldsymbol{\theta}')}{\mathcal{Z}(\boldsymbol{\theta}')}$
  - Propose to swap $\theta \leftrightarrow \theta'$ using Metropolis-Hastings

$$\alpha(\theta, \theta') = \frac{f(y; \theta') f(x; \theta) \pi(\theta') q(\theta', \theta)}{f(y; \theta) f(x; \theta') \pi(\theta) q(\theta, \theta')}$$

Pseudo-marginal MCMC (Roberts and Andrieu, 2009)

- Need an unbiased positive estimate of the target distribution $\hat{p}(y|\theta, u)$ such that

$$\int \hat{p}(y|\theta, u) p_\theta(u) \mathrm{d}u = p(y|\theta)$$

Pseudo-marginal MCMC (Roberts and Andrieu, 2009)

- Need an unbiased positive estimate of the target distribution $\hat{p}(y|\theta, u)$ such that

$$\int \hat{p}(y|\theta, u) p_\theta(u) \mathrm{d}u = p(y|\theta)$$

- Define joint distribution

$$\pi(\theta, u|y) = \frac{\pi(\theta)\hat{p}(y|\theta, u)p_\theta(u)}{p(y)}$$

Pseudo-marginal MCMC (Roberts and Andrieu, 2009)

- Need an unbiased positive estimate of the target distribution $\hat{p}(y|\theta, u)$ such that

$$\int \hat{p}(y|\theta, u) p_\theta(u) \mathrm{d}u = p(y|\theta)$$

- Define joint distribution

$$\pi(\theta, u|y) = \frac{\pi(\theta)\hat{p}(y|\theta, u)p_\theta(u)}{p(y)}$$

- This integrates to one, and has the posterior as its marginal

Pseudo-marginal MCMC (Roberts and Andrieu, 2009)

- Need an unbiased positive estimate of the target distribution $\hat{p}(y|\theta, u)$ such that

$$\int \hat{p}(y|\theta, u)p_\theta(u)\mathrm{d}u = p(y|\theta)$$

- Define joint distribution

$$\pi(\theta, u|y) = \frac{\pi(\theta)\hat{p}(y|\theta, u)p_\theta(u)}{p(y)}$$

- This integrates to one, and has the posterior as its marginal
- We can sample from this distribution!

$$\alpha(\theta, \theta') = \frac{\hat{p}(y|\theta', u')\pi(\theta')p_{\theta'}(u')}{\hat{p}(y|\theta, u)\pi(\theta)p_\theta(u)} \times \frac{q(\theta', \theta)p_\theta(u)}{q(\theta, \theta')p_{\theta'}(u')}$$

What's the problem?

- we need unbiased estimate of

$$\pi(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)\pi(\theta)}{p(\mathbf{y})} = \frac{f(\mathbf{y};\theta)}{\mathcal{Z}(\theta)}\pi(\theta)\Big/p(\mathbf{y})$$

What's the problem?

- we need unbiased estimate of

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\mathbf{y})} = \frac{f(\mathbf{y};\boldsymbol{\theta})}{\mathcal{Z}(\boldsymbol{\theta})}\pi(\boldsymbol{\theta}) \Big/ p(\mathbf{y})$$

- We can unbiasedly estimate $\mathcal{Z}(\theta)$ but if we take the reciprocal then the estimate is no longer unbiased.

$$\mathrm{E}[\hat{\mathcal{Z}}(\theta)] = \mathcal{Z}(\theta)$$

$$\mathrm{E}\left[\frac{1}{\hat{\mathcal{Z}}(\theta)}\right] \neq \frac{1}{\mathcal{Z}(\theta)}$$

Our approach

- Construct an unbiased estimate of the likelihood, based on a series expansion of the likelihood and stochastic truncation.
- Use pseudo-marginal MCMC to sample from the desired posterior distribution.

Proposed methodology

- Construct random variables $\{V_\theta^j, j \geq 0\}$ such that the series

$$\hat{\pi}(\theta|y, \{V^j\}) = \sum_{j=0}^{\infty} V_\theta^j \quad \text{has} \quad \mathbb{E}\left[\hat{\pi}(\theta|y, \{V^j\})\right] = \pi(\boldsymbol{\theta}|y).$$

Proposed methodology

- Construct random variables $\{V_\theta^j, j \geq 0\}$ such that the series

$$\hat{\pi}(\theta|y, \{V^j\}) = \sum_{j=0}^{\infty} V_\theta^j \quad \text{has} \quad \mathbb{E}\left[\hat{\pi}(\theta|y, \{V^j\})\right] = \pi(\boldsymbol{\theta}|y).$$

- This infinite series then needs to be truncated unbiasedly.
- This can be achieved via a number of Russian roulette schemes.

Proposed methodology

- Construct random variables $\{V_\theta^j, j \geq 0\}$ such that the series

$$\hat{\pi}(\theta|y, \{V^j\}) = \sum_{j=0}^{\infty} V_\theta^j \quad \text{has} \quad \mathbb{E}\left[\hat{\pi}(\theta|y, \{V^j\})\right] = \pi(\boldsymbol{\theta}|y).$$

- This infinite series then needs to be truncated unbiasedly.
- This can be achieved via a number of Russian roulette schemes.
- Define random time, τ_θ, such that $u := (\tau_\theta, \{V_\theta^j, 0 \leq j \leq \tau_\theta\})$

$$\pi(\theta, u|y) = \sum_{j=0}^{\tau_\theta} V_\theta^j \quad \text{which satisfies}$$

$$\mathbb{E}\left[\pi(\theta, u|y)|\{V_\theta^j, j \geq 0\}\right] = \sum_{j=0}^{\infty} V_\theta^j$$

Implementation example

- Rewrite the likelihood as an infinite series. Inspired by the work of Booth (2005).

Implementation example

- Rewrite the likelihood as an infinite series. Inspired by the work of Booth (2005).
- Simple manipulation:

$$\frac{f(\mathbf{y};\boldsymbol{\theta})}{\mathcal{Z}(\boldsymbol{\theta})} = \frac{f(\mathbf{y};\boldsymbol{\theta})}{\tilde{\mathcal{Z}}(\boldsymbol{\theta})} \frac{1}{1-\left[1-\frac{\mathcal{Z}(\boldsymbol{\theta})}{\tilde{\mathcal{Z}}(\boldsymbol{\theta})}\right]} = \frac{f(\mathbf{y};\boldsymbol{\theta})}{\tilde{\mathcal{Z}}(\boldsymbol{\theta})} \sum_{n=0}^{\infty} \kappa(\boldsymbol{\theta})^n$$

where

$$\kappa(\theta) = 1 - \frac{\mathcal{Z}(\boldsymbol{\theta})}{\tilde{\mathcal{Z}}(\boldsymbol{\theta})}$$

Implementation example

- Rewrite the likelihood as an infinite series. Inspired by the work of Booth (2005).
- Simple manipulation:

$$\frac{f(\mathbf{y};\boldsymbol{\theta})}{\mathcal{Z}(\boldsymbol{\theta})} = \frac{f(\mathbf{y};\boldsymbol{\theta})}{\tilde{\mathcal{Z}}(\boldsymbol{\theta})} \frac{1}{1-\left[1-\frac{\mathcal{Z}(\boldsymbol{\theta})}{\tilde{\mathcal{Z}}(\boldsymbol{\theta})}\right]} = \frac{f(\mathbf{y};\boldsymbol{\theta})}{\tilde{\mathcal{Z}}(\boldsymbol{\theta})} \sum_{n=0}^{\infty} \kappa(\boldsymbol{\theta})^n$$

where

$$\kappa(\theta) = 1 - \frac{\mathcal{Z}(\boldsymbol{\theta})}{\tilde{\mathcal{Z}}(\boldsymbol{\theta})}$$

- The series converges for $|\kappa(\boldsymbol{\theta})| < 1$.

Implementation example continued

- We can unbiasedly estimate each term in the series using n independent estimates of $\mathcal{Z}(\boldsymbol{\theta})$.

$$\frac{f(\mathbf{y};\boldsymbol{\theta})}{\mathcal{Z}(\boldsymbol{\theta})} = \frac{f(\mathbf{y};\boldsymbol{\theta})}{\tilde{\mathcal{Z}}(\boldsymbol{\theta})} \sum_{n=0}^{\infty} \left[1 - \frac{\mathcal{Z}(\boldsymbol{\theta})}{\tilde{\mathcal{Z}}(\boldsymbol{\theta})}\right]^{n}$$

$$\approx \frac{f(\mathbf{y};\boldsymbol{\theta})}{\tilde{\mathcal{Z}}(\boldsymbol{\theta})} \sum_{n=0}^{\infty} \prod_{i=1}^{n} \left[1 - \frac{\hat{\mathcal{Z}}_i(\boldsymbol{\theta})}{\tilde{\mathcal{Z}}(\boldsymbol{\theta})}\right]$$

Implementation example continued

- We can unbiasedly estimate each term in the series using n independent estimates of $\mathcal{Z}(\theta)$.

$$\frac{f(\mathbf{y};\theta)}{\mathcal{Z}(\theta)} = \frac{f(\mathbf{y};\theta)}{\tilde{\mathcal{Z}}(\theta)} \sum_{n=0}^{\infty} \left[1 - \frac{\mathcal{Z}(\theta)}{\tilde{\mathcal{Z}}(\theta)}\right]^n$$

$$\approx \frac{f(\mathbf{y};\theta)}{\tilde{\mathcal{Z}}(\theta)} \sum_{n=0}^{\infty} \prod_{i=1}^{n} \left[1 - \frac{\hat{\mathcal{Z}}_i(\theta)}{\tilde{\mathcal{Z}}(\theta)}\right]$$

- Computed using importance sampling (IS) or sequential Monte Carlo (SMC), for example.

Implementation example continued

- We can unbiasedly estimate each term in the series using n independent estimates of $\mathcal{Z}(\theta)$.

$$\frac{f(\mathbf{y};\theta)}{\mathcal{Z}(\theta)} = \frac{f(\mathbf{y};\theta)}{\tilde{\mathcal{Z}}(\theta)} \sum_{n=0}^{\infty} \left[1 - \frac{\mathcal{Z}(\theta)}{\tilde{\mathcal{Z}}(\theta)}\right]^n$$

$$\approx \frac{f(\mathbf{y};\theta)}{\tilde{\mathcal{Z}}(\theta)} \sum_{n=0}^{\infty} \prod_{i=1}^{n} \left[1 - \frac{\hat{\mathcal{Z}}_i(\theta)}{\tilde{\mathcal{Z}}(\theta)}\right]$$

- Computed using importance sampling (IS) or sequential Monte Carlo (SMC), for example.
- But can't compute an infinite number of them...

Unbiased estimates of infinite series

- Take an infinite convergent series $\mathcal{S} = \sum_{k=0}^{\infty} a_k$ which we would like to estimate unbiasedly.

Unbiased estimates of infinite series

- Take an infinite convergent series $\mathcal{S} = \sum_{k=0}^{\infty} a_k$ which we would like to estimate unbiasedly.
- *The simplest:* draw integer k with probability $p(K = k)$ where $\sum_{k=0}^{\infty} p(K = k) = 1$, then $\hat{\mathcal{S}} = a_k / p(K = k)$

Unbiased estimates of infinite series

- Take an infinite convergent series $\mathcal{S} = \sum_{k=0}^{\infty} a_k$ which we would like to estimate unbiasedly.
- *The simplest:* draw integer k with probability $p(K = k)$ where $\sum_{k=0}^{\infty} p(K = k) = 1$, then $\hat{\mathcal{S}} = a_k / p(K = k)$
- $\mathbb{E}[\hat{\mathcal{S}}] = \sum_k \frac{p(K=k) a_k}{p(K=k)} = \mathcal{S}$

Unbiased estimates of infinite series

- Take an infinite convergent series $\mathcal{S} = \sum_{k=0}^{\infty} a_k$ which we would like to estimate unbiasedly.
- *The simplest:* draw integer k with probability $p(K = k)$ where $\sum_{k=0}^{\infty} p(K = k) = 1$, then $\hat{\mathcal{S}} = a_k / p(K = k)$
- $\mathbb{E}[\hat{\mathcal{S}}] = \sum_k \frac{p(K=k) a_k}{p(K=k)} = \mathcal{S}$
- (This is essentially importance sampling)
- Variance: $\sum_{n=0}^{\infty} \left[\frac{a_n^2}{p(N=n)} \right] - \mathcal{S}^2$

Russian roulette

- *Alternative:* Russian roulette.
- Choose series of probabilities, $\{q_n\}$, and draw sequence of i.i.d. uniform random variables, $\{U_n\} \sim \mathcal{U}[0,1]$, $n = 1, 2, 3 \ldots$

Russian roulette

- *Alternative:* Russian roulette.
- Choose series of probabilities, $\{q_n\}$, and draw sequence of i.i.d. uniform random variables, $\{U_n\} \sim \mathcal{U}[0, 1]$, $n = 1, 2, 3 \ldots$
- Define time $\tau = \inf\{k \geq 1 : \; U_k \geq q_k\}$

Russian roulette

- *Alternative:* Russian roulette.
- Choose series of probabilities, $\{q_n\}$, and draw sequence of i.i.d. uniform random variables, $\{U_n\} \sim \mathcal{U}[0,1]$, $n = 1, 2, 3 \ldots$
- Define time $\tau = \inf\{k \geq 1 : \; U_k \geq q_k\}$
- $S_\tau = \sum_{j=0}^{\tau-1} \frac{a_j}{\prod_{i=1}^{j} q_i}$, is an unbiased estimate of $\mathcal{S}$.

Russian roulette

- *Alternative:* Russian roulette.
- Choose series of probabilities, $\{q_n\}$, and draw sequence of i.i.d. uniform random variables, $\{U_n\} \sim \mathcal{U}[0,1]$, $n = 1, 2, 3 \ldots$
- Define time $\tau = \inf\{k \geq 1 : \; U_k \geq q_k\}$
- $S_\tau = \sum_{j=0}^{\tau-1} \frac{a_j}{\prod_{i=1}^{j} q_i}$, is an unbiased estimate of $\mathcal{S}$.
- Must choose $\{q_n\}$ to minimise variance of estimator.

Debiasing estimator, McLeish (2011), Rhee Glynn (2012)

- Want unbiased estimate of $\mathrm{E}[Y]$, but can't generate Y. Can generate approximations, Y_n, s.t. $\lim_{n\to\infty} \mathrm{E}[Y_n] \to \mathrm{E}[Y]$.

Debiasing estimator, McLeish (2011), Rhee Glynn (2012)

- Want unbiased estimate of $\mathrm{E}[Y]$, but can't generate Y. Can generate approximations, Y_n, s.t. $\lim_{n\to\infty} \mathrm{E}[Y_n] \to \mathrm{E}[Y]$.

$$\widehat{\mathrm{E}[Y]} = Y_0 + \sum_{i=1}^{\infty} Y_i - Y_{i-1}$$

Debiasing estimator, McLeish (2011), Rhee Glynn (2012)

- Want unbiased estimate of $\mathrm{E}[Y]$, but can't generate Y. Can generate approximations, Y_n, s.t. $\lim_{n\to\infty} \mathrm{E}[Y_n] \to \mathrm{E}[Y]$.

$$\widehat{\mathrm{E}[Y]} = Y_0 + \sum_{i=1}^{\infty} Y_i - Y_{i-1}$$

- Define probability distribution for N, non-negative, integer-valued random variable, then

Debiasing estimator, McLeish (2011), Rhee Glynn (2012)

- Want unbiased estimate of $\mathrm{E}[Y]$, but can't generate Y. Can generate approximations, Y_n, s.t. $\lim_{n\to\infty} \mathrm{E}[Y_n] \to \mathrm{E}[Y]$.

$$\widehat{\mathrm{E}[Y]} = Y_0 + \sum_{i=1}^{\infty} Y_i - Y_{i-1}$$

- Define probability distribution for N, non-negative, integer-valued random variable, then

$$Z = Y_0 + \sum_{i=1}^{N} \frac{Y_i - Y_{i-1}}{P(N \geq i)}$$

- is an unbiased estimator of $\mathrm{E}[Y]$ and has finite variance.

Negative estimates

- Often we cannot guarantee the overall estimate will be positive

Negative estimates

- Often we cannot guarantee the overall estimate will be positive
- Can use a trick from the Physics literature...
- Recall we have an unbiased estimate of the likelihood, $\hat{p}(y|\theta, u)$

$$\mathrm{E}_\pi[\phi(\theta)] = \int \phi(\theta)\pi(\theta|y)\mathrm{d}\theta = \int\int \phi(\theta)\pi(\theta, u|y)\ \mathrm{d}\theta\mathrm{d}u$$

$$= \frac{\int\int \phi(\theta)\ \hat{p}(y|\theta, u)\pi(\theta)p_\theta(u)\ \mathrm{d}\theta\mathrm{d}u}{\int\int \hat{p}(y|\theta, u)\pi(\theta)p_\theta(u)\ \mathrm{d}\theta\mathrm{d}u}$$

Negative estimates

- Often we cannot guarantee the overall estimate will be positive
- Can use a trick from the Physics literature...
- Recall we have an unbiased estimate of the likelihood, $\hat{p}(y|\theta, u)$

$$\mathrm{E}_\pi[\phi(\theta)] = \int \phi(\theta)\pi(\theta|y)\mathrm{d}\theta = \int\int \phi(\theta)\pi(\theta, u|y)\ \mathrm{d}\theta\mathrm{d}u$$

$$= \frac{\int\int \phi(\theta)\ \hat{p}(y|\theta, u)\pi(\theta)p_\theta(u)\ \mathrm{d}\theta\mathrm{d}u}{\int\int \hat{p}(y|\theta, u)\pi(\theta)p_\theta(u)\ \mathrm{d}\theta\mathrm{d}u}$$

$$= \frac{\int\int \phi(\theta)\sigma(\hat{p})\ |\hat{p}(y|\theta, u)|\pi(\theta)p_\theta(u)\ \mathrm{d}\theta\mathrm{d}u}{\int\int \sigma(\hat{p})\ |\hat{p}(y|\theta, u)|\pi(\theta)p_\theta(u)\ \mathrm{d}\theta\mathrm{d}u}$$

Negative estimates cont.

- From last slide

$$\mathrm{E}_{\pi}[\phi(\theta)] = \frac{\int\int \phi(\theta)\sigma(\hat{p})\ |\hat{p}(y|\theta,u)|\pi(\theta)p_{\theta}(u)\ \mathrm{d}\theta\mathrm{d}u}{\int\int \sigma(\hat{p})\ |\hat{p}(y|\theta,u)|\pi(\theta)p_{\theta}(u)\ \mathrm{d}\theta\mathrm{d}u}$$

$$= \frac{\int\int \phi(\theta)\sigma(\hat{p})\ q(\theta,u|y)\ \mathrm{d}\theta\mathrm{d}u}{\int\int \sigma(\hat{p})\ q(\theta,u|y)\ \mathrm{d}\theta\mathrm{d}u}$$

Negative estimates cont.

- From last slide

$$\mathrm{E}_\pi[\phi(\theta)] = \frac{\int\int \phi(\theta)\sigma(\hat{p})\ |\hat{p}(y|\theta,u)|\pi(\theta)p_\theta(u)\ \mathrm{d}\theta\mathrm{d}u}{\int\int \sigma(\hat{p})\ |\hat{p}(y|\theta,u)|\pi(\theta)p_\theta(u)\ \mathrm{d}\theta\mathrm{d}u}$$

$$= \frac{\int\int \phi(\theta)\sigma(\hat{p})\ q(\theta,u|y)\ \mathrm{d}\theta\mathrm{d}u}{\int\int \sigma(\hat{p})\ q(\theta,u|y)\ \mathrm{d}\theta\mathrm{d}u}$$

- Can get a Monte Carlo estimate of $\phi(\theta)$ wrt the posterior using samples from the ‘absolute’ distribution

$$\mathrm{E}_\pi[\phi(\theta)] \approx \frac{\sum_k \phi(\theta_k)\sigma(\hat{p}_k)}{\sum_k \sigma(\hat{p}_k)}$$

Summary

- So, we can get an unbiased estimate of doubly-intractable distribution.

Summary

- So, we can get an unbiased estimate of doubly-intractable distribution.
  - Draw a random integer, $k \sim p(k)$
  - Compute k-th term in infinite series
  - Compute overall unbiased estimate of likelihood

Summary

- So, we can get an unbiased estimate of doubly-intractable distribution.
  - Draw a random integer, $k \sim p(k)$
  - Compute k-th term in infinite series
  - Compute overall unbiased estimate of likelihood
- Plug the *absolute* value into a pseudo-marginal MCMC scheme to sample from the posterior (or close to...).
- Compute expectations with respect to the posterior using importance sampling identity.

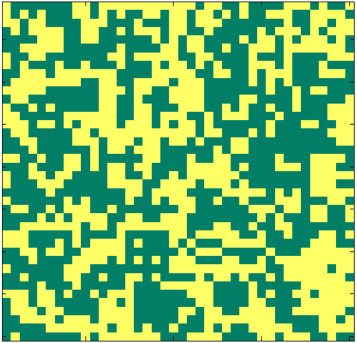
Summary

- So, we can get an unbiased estimate of doubly-intractable distribution.
  - Draw a random integer, $k \sim p(k)$
  - Compute k-th term in infinite series
  - Compute overall unbiased estimate of likelihood
- Plug the *absolute* value into a pseudo-marginal MCMC scheme to sample from the posterior (or close to...).
- Compute expectations with respect to the posterior using importance sampling identity.
- However, the methodology is computationally costly as need many low-variance estimates of partition function.

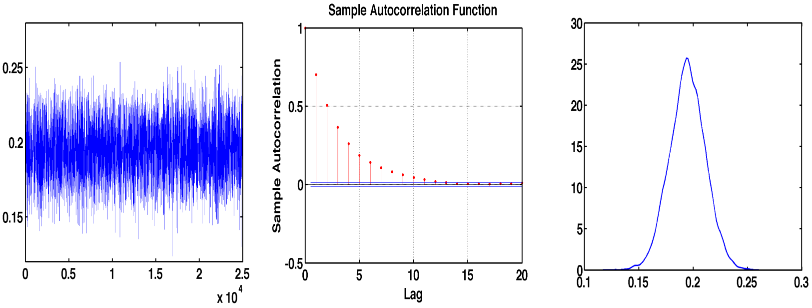
Summary

- So, we can get an unbiased estimate of doubly-intractable distribution.
  - Draw a random integer, $k \sim p(k)$
  - Compute k-th term in infinite series
  - Compute overall unbiased estimate of likelihood
- Plug the *absolute* value into a pseudo-marginal MCMC scheme to sample from the posterior (or close to...).
- Compute expectations with respect to the posterior using importance sampling identity.
- However, the methodology is computationally costly as need many low-variance estimates of partition function.
- But, we can compute estimates in parallel...

Results: Ising models



We simulated a 40x40 grid of data points from a Gibbs sampler with $J\beta = 0.2$



- Geometric construction with Russian roulette sampling
- AIS
- Parallel implementation using Matlab.

Example: Florentine business network

- ERGM model, 16 nodes

Example: Florentine business network

- ERGM model, 16 nodes
- Graph statistics in model exponent are number of edges, number of 2- and 3-stars and number of triangles

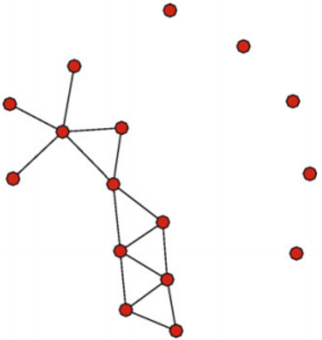
Example: Florentine business network

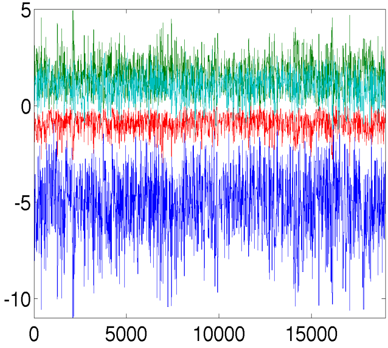
- ERGM model, 16 nodes
- Graph statistics in model exponent are number of edges, number of 2- and 3-stars and number of triangles
- Estimates of the normalising term were computed using SMC

Example: Florentine business network

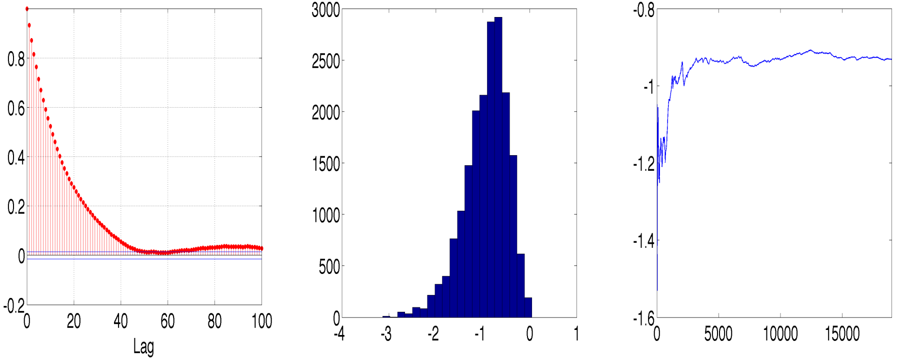
- ERGM model, 16 nodes
- Graph statistics in model exponent are number of edges, number of 2- and 3-stars and number of triangles
- Estimates of the normalising term were computed using SMC
- Series truncation was carried out using Russian roulette

Example: Florentine business network

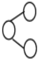
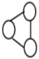




Example: Florentine business network



Example: Florentine business network

Parameter	Configuration	Estimate (standard error)
θ		−4.27 (1.13)
σ_2		1.09 (0.65)
σ_3		−0.67 (0.41)
τ		1.32 (0.65)

	Mean	Standard error
Edges	-5.1629	1.6645
2-stars	1.5532	0.8078
3stars	-0.9313	0.4684
Triangles	0.9891	0.6778

Further work

- Optimise various parts of the methodology, stopping probabilities etc.
- Compare with approximate approaches in terms of variance and computation

Further work

- Optimise various parts of the methodology, stopping probabilities etc.
- Compare with approximate approaches in terms of variance and computation

Thank you for listening!

References

- MCMC for doubly-intractable distributions. Murray, Ghahramani, MacKay (2004)
- An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. Møller, Pettitt, Berthelsen, Reeves (2004).
- Unbiased Estimation with Square Root Convergence for SDE Models. Rhee, Glynn (2012)
- A general method for debiasing a Monte Carlo estimator. McLeish (2011)
- Unbiased Monte Carlo Estimation of the Reciprocal of an Integral. Booth (2007)