IE 598 HW5

Tanmay Gupta, Peter Maginnis

April 26, 2015

5.1a) Sampling from a uniform distribution over the set of perfect matchings is identical to uniformly sampling a random permutation. Here we may represent a perfect matching as a set of index pairs $\sigma = \{(i, j)\}_{i=1}^{n}$ such that the collection is bijective. One simple way to do this is

Algorithm 1 Uniformly sampling a random permutation

for $i = 1$ to N **do**
  sample $j \sim \text{Unif}\left(\{1, \ldots, n\} \setminus \{j' : (i', j') \in \sigma \text{ and } i' < i\}\right)$
  $\sigma \leftarrow \sigma \cup \{(i, j)\}$
end for

(b) The weighted measure μ over the set of all perfect matchings P is given by

$$\mu(\sigma) = \frac{1}{Z(w)} \exp\left\{\sum_i w_{i\sigma(i)}\right\}$$

where

$$\begin{aligned} Z(w) &= \sum_{\sigma \in P} \mu(\sigma) \\ &= \sum_{\sigma \in P} \exp\left\{\sum_i w_{i\sigma(i)}\right\} \\ &\leq \sum_{\sigma \in P} \exp\left\{\sum_i w^*\right\} \\ &= \exp\{Nw^*\} \sum_{\sigma \in P} 1 \\ &= N! \exp\{Nw^*\}. \end{aligned}$$

Thus we may derive the lower bound

$$
\begin{aligned}
\mu(\sigma) &\geq \frac{1}{N! \exp\{Nw^*\}} \exp\left\{\sum_i w_{i\sigma(i)}\right\} \\
&\geq \frac{1}{N! \exp\{Nw^*\}},
\end{aligned}
$$

since $w_{ij} \geq 0$ for every $i, j \in \{1, \dots, N\}$.

(c) Given the Metropolis-Hastings rule that samples $i, i' \sim \mathrm{Unif}(\{1, \dots, N\})$ and swaps $\sigma(i)$ and $\sigma(i')$ with probability

$$
R = \min\left\{1, \exp(-w_{i\sigma(i)} - w_{i'\sigma(i')} + w_{i\sigma(i')} + w_{i'\sigma(i)})\right\},
$$

we may lower bound the probability of a valid transition $\mathbb{P}(\text{ next state is } \sigma' | \text{ current state is } \sigma)$ by noting

$$
\begin{aligned}
\mathbb{P}_{\sigma,\sigma'} &= \mathbb{P}(i, i' \text{ are both sampled })\mathbb{P}(\text{ swap is accepted }) \\
&= \frac{2}{N^2} R \\
&\geq \frac{2}{N^2} \frac{1}{\exp 2w^*} \\
&\geq \frac{1}{N^2 \exp 2w^*},
\end{aligned}
$$

where the second to last inequality follows from the fact that $0 \leq w_{i,j} \leq w^*$ for every i, j.

(d) We proceed by first using the bound from (c) and then the one from (b) as follows: For any set of matchings S

$$
\begin{aligned}
\frac{\sum_{\sigma \in S, \sigma' \in S^{\mathsf{c}}} \mu(\sigma)\mathbb{P}_{\sigma,\sigma'}}{\mu(S)\mu(S^{\mathsf{c}})} &\geq \frac{1}{N^2 \exp 2w^*} \frac{\sum_{\sigma \in S, \sigma' \in S^{\mathsf{c}}} \mu(\sigma)}{\mu(S)\mu(S^{\mathsf{c}})} \\
&\geq \frac{1}{N^2 \exp 2w^*} \frac{1}{N! \exp\{Nw^*\}} \frac{\sum_{\sigma \in S, \sigma' \in S^{\mathsf{c}}} 1}{\mu(S)\mu(S^{\mathsf{c}})} \\
&\geq \frac{1}{N! N^2 \exp((N+2)w^*)},
\end{aligned}
$$

where we use the fact that $\mu(S) \leq |S| = \sum_{\sigma \in S} 1$ for any S, since $\mu(\sigma) \leq 1$ for every σ. Thus, we have the lower bound on the conductance

$$
\begin{aligned}
\Phi &= \min_S \frac{\sum_{\sigma \in S, \sigma' \in S^{\mathsf{c}}} \mu(\sigma)\mathbb{P}_{\sigma,\sigma'}}{\mu(S)\mu(S^{\mathsf{c}})} \\
&\geq \frac{1}{N! N^2 \exp((N+2)w^*)}.
\end{aligned}
$$

(e) Combining our bounds with the bound provided in class derived from Cheeger's inequality

$$
\begin{aligned}
T_{\text{mix}}(\epsilon) &\leq \frac{2\log\frac{2}{\epsilon\sqrt{\mu_{\min}}}}{\Phi^2} \\
&\leq \left(N!N^2\exp\left((N+2)w^*\right)\right)^2 2\log\frac{2}{\epsilon\sqrt{\mu_{\min}}} \\
&\leq \left(N!N^2\exp\left((N+2)w^*\right)\right)^2 2\log\frac{2\sqrt{N!\exp\left\{Nw^*\right\}}}{\epsilon}
\end{aligned}
$$

5.2a) Update rules for node-by-node Gibbs sampler are as follows-
1. Sample (i, j) uniformly randomly from the 60×60 grid
2. Sample the value of x_{ij} from the conditional distribution of x_{ij} given all the other variables. The conditional distribution is given by

$$\mathbb{P}(x_{ij}|x_{\text{rest}}) \propto \exp\left[\theta x_{ij}\left(x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1}\right)\right]$$

5.2b) An efficient procedure for sampling from tree structured undirected graphical model is as follows:
1. Run belief propagation on the tree to get messages from the root downwards. These messages represent $\mathbb{P}(x_{child}|x_{parent})$.

2. Sample from the single variable distribution of the root, then sample from the distribution of the children conditioned on the root and so on.

5.2c) Block gibbs sampling can be done using our sampling procedure for tree structured undirected graphical models from (b) by attaching a unary term to each node such that the messages incorporate conditioning on the nodes not in that block, resulting in a standalone tree distribution from which we can sample.

5.2d) Result of block gibbs sampling

Figure 1: Two samples generated by block gibbs sampler

Figure 2: Two samples generated by our gibbs sampler

From the above observations it seems that the block gibbs sampler has a lower mixing time. This become even more apparent when we look at extreme values of θ such as $\theta = 100$ where the block gibbs sampler converges in just 3 iteration while our gibbs sampler takes 6.

Figure 3: Comparison of mixing time for block (left) and our Gibbs sampler (right) for $\theta = 100$