Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
6.438 Algorithms for Inference
Fall 2018

Problem Set 6

Issued: Thursday, Nov. 08, 2018 **Due:** Tuesday, Nov. 20, 2018

Problem 6.1
Decide whether each of the following random processes are Markov (i.e., a sequence $x[0], x[1], \ldots$ can be represented by a Markov chain). Be sure to justify your answers.

(a) Let $x[n] = a \sin(\omega n + \phi)$, where a, ω and ϕ are independent Gaussian random variables with mean 0 and variance 1. Is $x[n]$ a Markov process?

(b) Let $x_1[n]$ and $x_2[n]$ be independent Markov processes. Is $x_1[2n]$ a Markov process? Also, define the process $y[n]$ defined by $y[2n] = x_1[n]$ and $y[2n+1] = x_2[n]$. Is $y[n]$ a Markov process?

(c) (**Optional**) Consider a sequence of i.i.d. random variables $x[0], x[1], \ldots$, where each $x[n]$ is geometrically distributed with parameter p, i.e., $p_{x[n]}(x) = p(1-p)^{x-1}$ for $n = 1, 2, \ldots$. Define $y[n]$ to be the largest k such that $\sum_{i=0}^{k-1} x[i] < n$. Let $z[0], z[1], \ldots$ be a sequence of i.i.d. Gaussian random variables with 0 mean and variance σ^2. Finally, consider the process $w[n] = z[y[n]]$. Is $w[n]$ a Markov process?

Problem 6.2
In this problem, we develop an efficient algorithm for sampling from a two-dimensional Ising model. In particular, suppose all variables x_{ij} take values in $\{-1, +1\}$. Using the graph structure $\mathcal{G}$ shown in Figure 6.2-1, define the distribution

$$p(x; \theta) \propto \exp\left\{ \sum_{(i,j)\in\mathcal{E}} \theta x_i x_j \right\}.$$

(a) Derive the update rules for a node-by-node Gibbs sampler for this model.

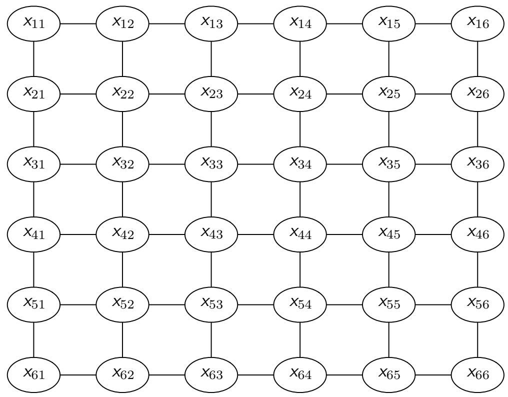


Figure 6.2-1

(b) Suppose we are given a tree-structured undirected graphical model $\mathcal{T}$ with variables $\mathsf{y} = (\mathsf{y}_1, \ldots, \mathsf{y}_N)$. Give an efficient procedure for sampling from the joint distribution $p_{\mathsf{y}}(\mathbf{y})$.

Hint: One way to generate an exact sample from a distribution over random variables $\mathsf{y} = (\mathsf{y}_1, \ldots, \mathsf{y}_N)$ is to sample each y_i from the distribution $p_{\mathsf{y}_i|\mathsf{y}_1,\ldots,\mathsf{y}_{i-1}}(\cdot|y_1, \ldots, y_{i-1})$ in sequence. Show that with a suitable variable ordering, all of these conditional distributions can be computed by running belief propagation once.

(c) In *block Gibbs sampling*, we partition a graph into r subsets $A_1, \ldots, A_r$. In each iteration, for each A_i, we sample x_{A_i} from the conditional distribution $p_{\mathsf{x}_{A_i}|\mathsf{x}_{\mathcal{V}\setminus A_i}}$.

For the Ising model $\mathcal{G}$ described above, consider the two comb-shaped subsets A and B shown in Figure 6.2-2. Describe how to use your sampler from part (b) to perform the block Gibbs updates. (For this part, you may assume a black-box implementation of your sampling procedure from part (b).)

Problem 6.3
Computational Exercise

Implement the node-by-node sampler and the block Gibbs sampler from the last question. Run it for 1000 iterations (sweeps over all the variables)

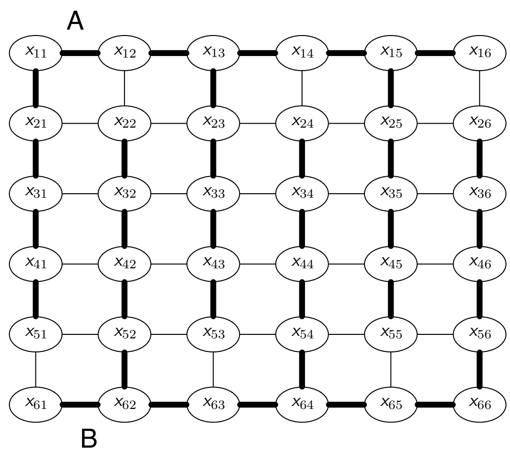


Figure 6.2-2

on an Ising model of size 60×60 with coupling parameter $\theta = 0.45$. Show the state of the variables after every 100 iterations.

Please hand in a short report (**no longer than 2 pages**, and you are encouraged to print double-sided) with:

1) a list of key routines in your program and brief description of their functionalities;

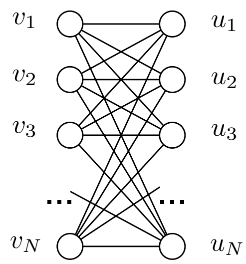
2) required samples visualized as images;

3) a brief discussion of results, e.g., which of the two samplers mixes faster? runs faster per sweep? Does initialization affect the final sample? (Try initializing with all +1s and all -1s)

4) a screenshot of key routines of your code (it is sufficient to include only the most important ones. You can shrink the images to fit into the two-page limit).

Problem 6.4

Matchings in a graph are subsets of edges such that no two edges share a vertex. Here we focus on the special case of a complete bipartite graph with vertices $v_1, \ldots, v_N$ on the left and $u_1, \ldots, u_N$ on the right, as shown:



In such a graph, a *perfect matching* is a matching which includes N edges. We are interested in sampling from a distribution over perfect matchings. We can denote a perfect matching using the variables $\sigma = [\sigma_{ij}] \in \{0, 1\}^{N \times N}$, where $\sigma_{ij} = 1$ if v_i and u_j are matched and $\sigma_{ij} = 0$ otherwise. Observe that σ is a perfect matching if and only if

$$\begin{aligned} \sum_{k=1}^{N} \sigma_{ik} &= 1 \qquad \text{for all } 1 \leq i \leq N \\ \sum_{k=1}^{N} \sigma_{kj} &= 1 \qquad \text{for all } 1 \leq j \leq N. \end{aligned}$$

A perfect matching σ can also be thought of as a permutation $\sigma : \{1, \ldots, N\} \to \{1, \ldots, N\}$. E.g., if $\sigma_{12} = \sigma_{21} = \sigma_{33} = 1$, this would correspond to the permutation $\sigma(1) = 2$, $\sigma(2) = 1$, and $\sigma(3) = 3$.

Consider the distribution defined by:

$$\begin{aligned} p(\sigma) &\propto \exp\left(\sum_{i,j} w_{ij}\sigma_{ij}\right) \mathbb{1}_{\{\sigma \text{ is a perfect matching}\}} \\ &= \exp\left(\sum_{i} w_{i\sigma(i)}\right) \mathbb{1}_{\{\sigma \text{ is a perfect matching}\}} \end{aligned}$$

where $w_{ij} \geq 0$ for all i, j.

(a) In this part, consider the uniform distribution over perfect matchings, i.e. $w_{ij} = 0$ for all i, j. Describe a simple procedure to sample σ from this uniform distribution. (Hint: Think about permutations.)

Consider the Metropolis-Hastings rule defined by: choose $i, j \in \{1, \ldots, N\}$ uniformly at random. If $i = j$, do nothing, otherwise with probability

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

swap $\sigma(i)$ and $\sigma(j)$, i.e., define the new permutation to be $\sigma'(k) = \sigma(k)$ for $k \neq i, j$ and $\sigma'(i) = \sigma(j)$ and $\sigma'(j) = \sigma(i)$.

(b) Show that for any perfect matching σ

$$p(\sigma) \geq \frac{1}{N! \exp(Nw^*)}, \tag{1}$$

where $w^* = \max_{i,j} w_{ij}$.
Hint: Bound the partition function of $p(\sigma)$.

(c) Show that under the above Markov chain, for any valid transition $\sigma \to \sigma'$

$$\mathbf{P}_{\sigma\sigma'} \geq \exp(-2w^*) \frac{1}{N^2}. \tag{2}$$

We can upper bound the mixing time (or "burn in time", i.e., number of steps until the sample is ϵ close to a sample from the true distribution) of the Markov chain with the following expression:

$$T_{mix}(\epsilon) \leq \frac{1}{\Phi^2} \left(\log \frac{1}{\min_\sigma p(\sigma)} + \log \frac{1}{\epsilon} \right), \tag{3}$$

where Φ is the conductance of the Markov chain defined by

$$\Phi = \min_S \frac{\sum_{\sigma \in S, \sigma' \in S'} p(\sigma) \mathbf{P}_{\sigma\sigma'}}{p(S)p(S')}.$$

(d) Argue using (1) and (2) that

$$\Phi \geq \frac{1}{N! \exp(Nw^*)} \frac{1}{N^2} \exp(-2w^*). \tag{4}$$

(e) Using (1), (4), and (3), obtain a bound on the mixing time of the Markov chain.