CS 510/610 Topics on probabilistic graphical models

Homework 2

Portland State U, Winter 2021 — *02/09/21*

Lecturer: Fang Song — *Due: 02/25/21*

Instructions. This problem set contains 12 pages (including this cover page) and 4 questions. A random subset of problems will be graded.

- Your solutions will be graded on *correctness* and *clarity*. You should only submit work that you believe to be correct, and you will get significantly more partial credit if you clearly identify the gap(s) in your solution. It is good practice to start any long solution with an informal (but accurate) summary that describes the main idea. You may opt for the "I take 15%" option.
- You need to submit a PDF file before the deadline. Either a clear scan of you handwriting or a typeset document is accepted. You will get 5 bonus points for typing in LaTeX (Download and use the accompany TeX file).
- You may collaborate with others on this problem set. However, you must **write up your own solutions** and **list your collaborators and any external sources** for each problem. Be ready to explain your solutions orally to a course staff if asked.

1. (Normalized importance sampling) A common situation in real-world applications is that we only know a probability up to a normalizing constant Z. Namely, we can only compute an unnormalized distribution $\tilde{p}(X)$ of $p(X)$ such that $\tilde{p}(X) = Zp(X)$. In this case, if we run importance sampling with a proposal distribution q, we can only compute the weight function with respect to $\tilde{p}$:

$$w(X) = \frac{\tilde{p}(X)}{q(X)}.$$

   (a) (5 points) Show that $\mathbb{E}_q[w(X)] = Z$.

(b) (5 points) Let f be an arbitrary function on $\text{Val}(X)$. Show that

$$\mathbb{E}_p(f(X)) = \frac{\mathbb{E}_q[f(X)w(X)]}{\mathbb{E}_q[w(x)]}.$$

(N.B. As a result, we can form an empirical estimation $\hat{I}_T(f)$ by that of both the numerator and denominator, i.e., $\hat{I}_T(f) := \frac{\sum_{t=1}^T f(x^t)w(x^t)}{\sum_{t=1}^T w(x^t)} \approx \mathbb{E}_p[f(X)]$, where $x^1, \ldots, x^T$ are i.i.d samples from q.)

(c) (5 points) Show that $\mathbb{E}_p[\frac{\tilde{p}(X)}{q(X)}] \geq 1$ and the equality holds iff. $p = q$.

(d) (10 points (bonus)) Now consider $X = (X_1, \ldots, X_D)$, and assume that both p and q can be factorized, i.e., $p(x) = \Pi_{i=1}^{D} p_i(x_i)$, and $q(x) = \Pi_{i=1}^{D} q_i(x_i)$. Denote $w^i := w(x^i) = \frac{\tilde{p}(x^i)}{q(x^i)}$ (note that each sample $x^i = (x_1^i, \ldots, x_D^i)$ is D-dimensional). A measure of the variability of two components in vector $w = (w^1, \ldots, w^T)$ is given by $\mathbb{E}_q[(w^i - w^j)^2]$. Show that $\mathbb{E}_q[(w^i - w^j)^2]$ has exponential growth with respect to D.

(N.B. This tells us that the standard importance sampling would blow up in high-dimensional cases.)

2. (15 points (bonus)) Consider the following heuristics of picking the elimination order in the variable elimination algorithm (or the corresponding graph elimination procedure).

   - *Min-neighbors*. At each point, choose a vertex that has the minimum number of neighbors in the current graph.
   - *Min-Weight*. Choose a vertex of the minimum product of the domain cardinality of its neighbors.
   - *Min-Fill*. Choose a vertex that causes the minimum number of edges to be added due to its elimination.

   Show that none of them dominate the others; that is, for any pair of strategies there is always a graph where the ordering produced by one is better than that produced by the other. As our measure of performance, use the computational cost of full variable elimination, i.e., computing the partition function.

3. (Markov Chains) Consider the following two conditions on a Markov chain T:

   I. It is possible to get from any state to any state using a positive probability path in the state graph.

   II. For each state x, there is a positive probability of transitioning directly from x to x (a self-loop).

   Answer the following.

   (a) (5 points) Show that, for a finite-state Markov chain, these two conditions together imply that T is regular.

   (b) (5 points) Show that regularity of the Markov chain implies condition I.

   (c) (5 points) Show an example of a regular Markov chain that does not satisfy the condition II.

4. (MCMC)

   (a) (10 points) Show that any distribution π that satisfies the detailed balance equation below must be a stationary distribution of T.

$$\forall x, x' \in \mathrm{Val}(X), \pi(x)T(x \to x') = \pi(x')T(x' \to x)\,.$$

(b) (10 points) Consider the Markov chain T induced by the Metropolis-Hasting algorithm:

$$T(x \to x') := \min \left\{ 1, \frac{\pi(x')q(x' \to x)}{\pi(x)q(x \to x')} \right\},$$

where q is a proposal distribution (in class we denoted $q(x' \to x)$ by the conditional probability $q(x|x')$). Verify that π and T satisfy the detailed balance equation. Conclude that π is the stationary distribution of A (assuming that A is regular).

(c) (10 points) Consider a simple Bayesian network p with two ternary variables ($\text{Val}(A) = \text{Val}(B) = \{0, 1, 2\}$). Let the CPDs (conditional probability distributions) be as follows.

A	$p(A)$
0	0.1
1	0.3
2	0.6

A	$p(B = 0 \mid A)$	$p(B = 1 \mid A)$	$p(B = 2 \mid A)$
0	0.2	0.3	0.5
1	0.3	0.3	0.4
2	0.1	0.5	0.4

Suppose that we want to run Gibbs sampling to sample from p, construct the corresponding Markov chain by describing the graph of the state space and the transition probabilities.

$$A \longrightarrow B$$

Figure 1: Simple Bayesian network

(d) (Exercise. Do not turn in.) Show that Gibbs sampling is a special case of the Metropolis-Hastings algorithm. Namely, provide a particular proposal distribution Q_i for each local transition T^{Q_i} that induces precisely the same distribution over the transitions taken as the associated Gibbs transition distribution T_i.

(e) (15 points) Consider an unnormalized distribution $\tilde{p}(X)$ which is hard to sample from, and a proposal distribution q from which we can draw independent samples. Consider a Markov chain where we define

$$T(x \to x') = q(x') \min\left[1, \frac{w(x')}{w(x)}\right] ,$$

for $x' \neq x$, where $w(x) := \frac{\tilde{p}(x)}{q(x)}$. And we further define $T(x \to x) := 1 - \sum_{x \neq x'} T(x \to x')$. Intuitively, the transition from x to x' selects an independent sample x' from q, and then moves toward it, depending on whether its importance weight is better than that of our current point x. Show that T defines a legal Markov chain and p ($\tilde{p}$ normalized) is its stationary distribution.