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

Problem Set 7

Issued: Wednesday, Nov. 21, 2018 **Due:** Thursday, Nov. 29, 2018

Problem 7.1

Consider the distributions $p = \text{Ber}(x)$ and $q = \text{Ber}(x + \epsilon)$. Let $\hat{q}(1) = \frac{1}{M}\sum_{i=1}^{M} X_i$ with $X_i \overset{\text{i.i.d.}}{\sim} q$. Show that for $t > \epsilon$,

$$\mathbb{P}(|\hat{q}(1) - p(1)| \geq t) \leq 2e^{-2M(t-\epsilon)^2}$$

Problem 7.2

Recall that an *independent set* of a graph $G = (\mathcal{V}, \mathcal{E})$ is a subset of vertices I such that no two vertices in I are adjacent. We can define a probability distribution over independent sets using the graph G itself as the undirected graphical model, and random variables $x_i = 1$ if vertex i is in the independent set and 0 if it is not:

$$\mu(x) = \frac{1}{Z} \prod_{(i,j)\in\mathcal{E}} \mathbb{1}_{(x_i,x_j)\neq(1,1)} \prod_{i\in\mathcal{V}} \lambda^{x_i}$$

As λ increases, the distribution places higher probability on larger independent sets. In this problem, we will use the technique of path coupling to analyze the mixing time of a Metropolis-Hastings sampler for this distribution. Throughout, we will use the following process for generating proposals:

- assume we start at configuration x
- choose a vertex $i \in G$ uniformly at random
- generate a proposal $z \sim \text{Ber}(.5)$
- accept the proposal (i.e. update to x' with $x'_i = z$ and $x'_j = x_j, j \neq i$) with probability $\min(1, \lambda^{z-x_i})$ if either:
  - $z = 0$

– $z = 1$ and $x_j = 0 \; \forall \; j \in \partial(i)$

(a) Show that the update process satisfies detailed balance, i.e. $\mu(x)P_{xx'} = \mu(x')P_{x'x}$, with $P_{xx'}$ being the probability of transitioning from configuration x to configuration x'.

(b) Write P for the 2-node complete graph. (You can omit the configuration $(x_1, x_2) = (1, 1)$ since it is not an independent set).

Recall that

$$T_{\text{mix}}(\epsilon) = \max_{x_0} \min_{t} \{t : \|\mu_{x_0}^t - \mu\|_{TV} \leq \epsilon\}$$

where $\mu_{x_0}^t$ is the distribution of the Markov chain after t iterations when starting at configuration x_0.

We saw that $\|\mu_{x_0}^t - \mu\|_{TV} \leq \mathbb{P}(X^t \neq Y^t)$ for any coupling (X^t, Y^t) where $X^t \sim \mu_{x_0}^t$ and $Y^t \sim \mu$.

(c) Let $\rho(x, x')$ be the minimum number of moves (i.e. a sequence of independent sets) needed to get from configuration x to configuration x'. Notice that $\rho(x, x') \leq 2|\mathcal{V}|$. Suppose we are able to come up with a coupling (X^t, Y^t) such that for some $\alpha > 0$,

$$\mathbb{E}[\rho(X^t, Y^t) | X^{t-1}, Y^{t-1}] \leq e^{-\alpha} \rho(X^{t-1}, Y^{t-1}) \tag{1}$$

Show that with such a coupling, we can bound the mixing time by

$$T_{mix}(\epsilon) \leq \frac{1}{\alpha} \log \frac{2|\mathcal{V}|}{\epsilon}$$

In the rest of the problem, assume that the maximum degree of any vertex is k.

(d) Define a coupling (X^{t+1}, Y^{t+1}) such that if $\rho(X^t, Y^t) = 1$, then

$$\mathbb{E}[\rho(X^{t+1}, Y^{t+1}) | X^t, Y^t] \leq e^{-\alpha}$$

for some α, and derive a condition on λ for which $\alpha > 0$. Remember that the coupling must respect the update equation of the sampler (i.e. $p(X^{t+1} = x' | X^t = x) = P_{xx'}$). *Hint: consider first drawing* $U \sim \text{Unif}([0, 1])$ *and break your analysis of* $\mathbb{E}[\rho(X^{t+1}, Y^{t+1}) | X^t, Y^t]$ *into cases.*

(e) Show that the condition in (d) guarantees (1).

Problem 7.3

Consider Figure 7.3-1.

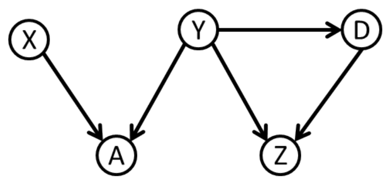


Figure 7.3-1

(a) Are X and Y conditionally independent given observation Z? Provide an explanation for your answer.

For parts (b) and (c), consider the directed graphs shown in Figure 7.3-2.

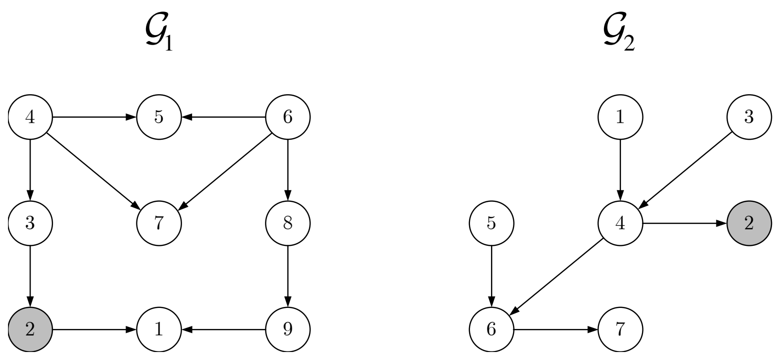


Figure 7.3-2

(b) Determine the maximal set $\mathcal{B}$ for which $X_1 \perp\!\!\!\perp X_{\mathcal{B}} | X_2$ for the graph $\mathcal{G}_1$.

(c) Determine the maximal set $\mathcal{B}$ for which $X_1 \perp\!\!\!\perp X_{\mathcal{B}} | X_2$ for the graph $\mathcal{G}_2$.

Problem 7.4

Let $\mathcal{I}(p)$ represent the set of conditional independence statements that are true the distribution p. For example, if $p(x_1, x_2, x_3) = f(x_1, x_2)g(x_2, x_3)$, then $\mathcal{I}(p) = \{x_1 \perp\!\!\!\perp x_3 | x_2\}$.

Similarly, let $\mathcal{I}(\mathcal{G})$ represent the set of conditional independence statement implied by the graph $\mathcal{G}$, which can be either directed or undirected. For example, if $\mathcal{G}$ is the complete graph, $\mathcal{I}(\mathcal{G}) = \varnothing$, while if $\mathcal{G}$ has no edges, $\mathcal{I}(\mathcal{G}) = \{x_i \perp\!\!\!\perp x_j | x_S \; \forall \; i \neq j, S \subseteq \mathcal{V} \setminus \{i, j\}\}$

We say that $\mathcal{G}$ is an *I-map* of p if $\mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}(p)$ (we have seen an equivalent way of saying this before: p is Markov with respect to $\mathcal{G}$). Conversely, we say that $\mathcal{G}$ is a *D-map* of p if $\mathcal{I}(p) \subseteq \mathcal{I}(\mathcal{G})$. Finally, we call $\mathcal{G}$ a *minimal I-map* of p if it is an I-map of p and any proper subgraph $\mathcal{G}'$ of $\mathcal{G}$ is not an I-map of p.

(a) Argue that the following procedure for constructing a directed graph results in a minimal I-map of $p(x_1, \ldots, x_n)$:

- Choose any permutation σ of $1, \ldots, n$
- For $i = 1, \ldots, n$, let the parents of σ_i be the smallest set π_{σ_i} such that $x_{\sigma_i} \perp\!\!\!\perp x_{\{\sigma_1, \ldots, \sigma_{i-1}\} \setminus \pi_{\sigma_i}} | \pi_{\sigma_i}$

(b) Let p be a distribution and $\mathcal{G}$ be a DAG such that $\mathcal{I}(\mathcal{G}) = \mathcal{I}(p)$ (in general, such a DAG is not always possible to obtain, but assume that we can for the given distribution). Let $\mathcal{I}_\sigma$ denote exactly those conditional independence statements found in the second step of the procedure, using distribution p and permutation σ. Prove that all conditional independence statements involving a single node are captured by one of these sets. (Formally, if $x_i \perp\!\!\!\perp x_A | x_B \in \mathcal{I}(p)$, then $x_i \perp\!\!\!\perp x_A | x_B \in \cup_\sigma \mathcal{I}_\sigma$)

Problem 7.5

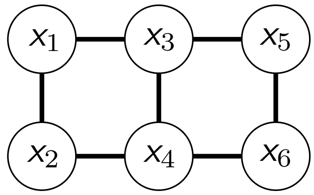
We define a graph as a *maximal D-map* for a family of distributions if adding even a single edge makes the graph no longer a D-map for that family.

(a) Consider a family of probability distributions defined on a set of random variables $\{x_1, x_2, x_3, x_4\}$ such that $x_k \perp\!\!\!\perp x_l$ for all $k \neq l$.

Using the given independence statements, draw an undirected maximal D-map for this family.

(b) Find a directed graphical model that is a maximal D-map of the family of distributions represented by the following undirected graphical model:

No additional variables are allowed.
Hint: Consider the effects of adding V-structures to your graph.



(c) Is it possible to find a family of distributions whose undirected minimal I-map has fewer edges than its undirected maximal D-map? If so, give an example distribution and the undirected graphs for the minimal I-map and maximal D-map. If not, explain why not.

Recall that a graph is a minimal I-map for a family of distributions if removing even a single edge makes the graph no longer an I-map for that family.