Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
6.438 Algorithms for Inference
Fall 2020

QUIZ 1

Thursday, October 15, 2020
9:00 am – 11:00 am / 7:00 pm – 9:00 pm

- This is a closed book exam, but two $8\frac{1}{2}$" $\times$11" sheets of notes (4 sides total) are allowed.
- Calculators are allowed, but probably won't be useful.
- There are **3** problems of approximately equal value.
- The problems are not necessarily in order of difficulty. We recommend that you read through all the problems first, then do the problems in whatever order suits you best.
- Quite often later parts of a problem can be done independently of earlier parts, so if you get stuck in one part, proceed to the next and come back to the earlier part later.
- Record all your solutions either on a printed copy of the answer booklet, or blank sheets of paper with each problem part on a different page. You may want to first work things through on the scratch paper, and then carefully transfer to the solution that you would like us to look at.
- A correct answer does not guarantee full credit, and a wrong answer does not guarantee loss of credit. You should clearly but concisely indicate your reasoning and **show all relevant work**. Your grade on each problem will be based on our best assessment of your level of understanding as reflected by what you have written in the answer booklet. **No credit will be given for an answer without valid justification!**
- Please be neat—we can't grade what we can't decipher!

Problem 1

(a) Suppose there is a student enrolled in a university class. There are 5 random variables we consider in this setting: D, I, G, S, L.

- D denotes the difficulty of the class. It takes values 0 (easy) or 1 (hard).
- I denotes the IQ of the student.
- G denotes the grade that the student gets from the class. It takes values A, B, or C.
- S denotes the SAT score of the student.
- L denotes the quality of recommendation letter the student gets from the professor after completing the course. It takes values 0 (not a strong letter) or 1 (a strong letter).

The grade is determined by the student's IQ and the difficulty of the class, and it affects the quality of the recommendation letter. Also, the IQ of the student influences their SAT score.

(i) Draw a DAG consisting of the 5 random variables that is consistent with the causal relationships described above. Make sure that your DAG contains as few edges as possible, while being consistent.

(ii) From the DAG, determine whether the following conditional independence statements hold. Justify why.

- $D \perp\!\!\!\perp I$
- $D \perp\!\!\!\perp S \mid L$
- $S \perp\!\!\!\perp L \mid I$

(b) Consider the following DAG $\mathcal{G}$.

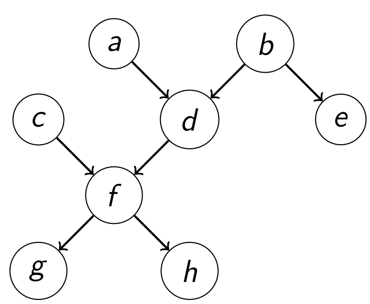


Figure 1: A DAG $\mathcal{G}$.

(i) Draw a DAG that is I-equivalent to the DAG $\mathcal{G}$. Explain why this is the only DAG that is I-equivalent to $\mathcal{G}$.

(ii) Draw an undirected graph that is a minimal I-map for the DAG $\mathcal{G}$.

(c) Suppose there are 5 random variables a, b, c, d, and e, which satisfy the following conditional indepedence relations:

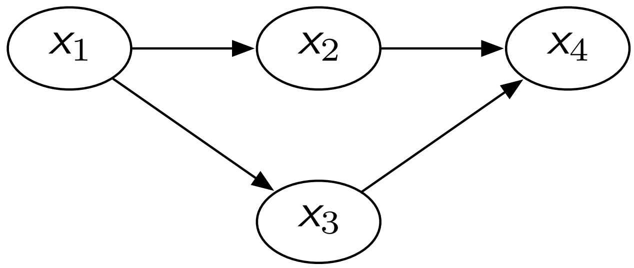
$$\begin{aligned}
&a \perp\!\!\!\perp c \mid \mathcal{A}, \quad \text{for } \mathcal{A} = \{b\}, \{b, d\}, \{b, e\}, \{b, d, e\}, \\
&c \perp\!\!\!\perp d \mid \mathcal{A}, \quad \text{for } \mathcal{A} = \{b\}, \{a, b\}, \{a, e\}, \{b, e\}, \{a, b, e\}, \\
&c \perp\!\!\!\perp e \mid \mathcal{A}, \quad \text{for } \mathcal{A} = \{b\}, \{a, b\}, \{b, d\}, \{a, b, d\}, \\
&a \perp\!\!\!\perp e \mid b, d, \quad a \perp\!\!\!\perp e \mid b, c, d, \\
&b \perp\!\!\!\perp d \mid a, e, \quad b \perp\!\!\!\perp d \mid a, c, e.
\end{aligned}$$

(i) Draw an undirected graph that is the P-map for the 5 random variables. Justify your answer.

(ii) Draw a DAG that is a minimal I-map for the 5 random variables. Justify your answer.

Problem 2

Consider the following directed graphical model on 4 vertices:



We assume that the data is generated by the following process:

$$\begin{aligned}
x_1 &= \varepsilon_1 \\
x_2 &= \beta_{12}x_1 + \varepsilon_2 \\
x_3 &= \beta_{13}x_1 + \varepsilon_3 \\
x_4 &= \beta_{24}x_2 + \beta_{34}x_3 + \varepsilon_4,
\end{aligned}$$

where $\varepsilon_i \sim \mathcal{N}(0,1)$ are independent. We are given N i.i.d samples from this distribution $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$, where $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, x_4^{(i)})$. Our goal is to estimate the β parameters, which we will do via maximum likelihood estimation.

(a) Show that we can write the likelihood of the data $p(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}; \beta_{12}, \beta_{13}, \beta_{24}, \beta_{34})$ as the following product of terms that only depend on subsets of the parameters:

$$\begin{aligned}
&p(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}; \beta_{12}, \beta_{13}, \beta_{24}, \beta_{34}) \\
&\qquad \propto f(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}; \beta_{12}) g(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}; \beta_{13}) h(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}; \beta_{24}, \beta_{34}).
\end{aligned}$$

Compute each of the functions f, g, and h.

(b) Use the above factorization to compute the maximum likelihood estimates for all of the β parameters in the graphical model.

(c) Express the joint distribution of one sample $p(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, x_3, x_4)$, as an exponential family. What are the natural parameters, and what are the sufficient statistics?

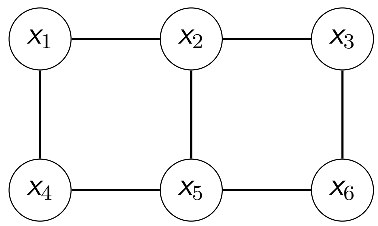
(d) Now, let $\mathcal{G}$ be a directed graph on n vertices. We can define a Gaussian graphical model on $\mathcal{G}$, given by

$$x_i = \beta_i^T x_{\pi_i} + \varepsilon_i,$$

where $\beta_i \in \mathbb{R}^{|\pi_i|}$ and $\varepsilon_i \sim \mathcal{N}(0,1)$ are independent. If we express the joint distribution of $x_1, \ldots, x_n$ as an exponential family, prove that $x_i x_j$ is a sufficient statistic for $i \neq j$ if and only if i and j are connected by an edge in every undirected I-map for the DAG $\mathcal{G}$.

Problem 3

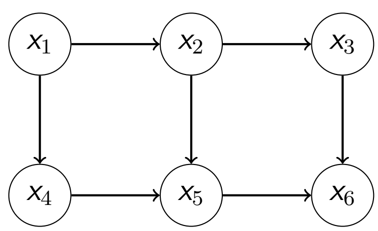
(a) Consider the following undirected graph. We would like to compute the marginal distribution of x_6.



(i) Suppose you run the variable elimination algorithm with elimination order $(1, 2, 3, 4, 5, 6)$. Draw the sequence of graphs that you obtain as the algorithm proceeds, with the edges introduced by the elimination procedure. Draw the reconstituted graph.

(ii) Identify an elimination order with smaller maximum clique size in the reconstituted graph than the order in Part (a)(i). Repeat Part (a)(i) with that order.

(b) Consider the following causal DAG. We would like to compute the interventional distribution of $x_6 \mid \mathrm{do}(x_4 = x_4)$



(i) Draw the interventional graph.

(ii) Assume that we know each of the factors $p(x_i \mid x_{\pi_i})$. Write the distribution $p(x_1, x_2, x_3, x_5, x_6 \mid \mathrm{do}(x_4 = x_4))$ in terms of these factors.

(iii) We can naively sum over all variables to compute the interventional distribution $p(x_6 \mid \mathrm{do}(x_4 = x_4))$ in $O(|\mathcal{X}|^6)$ time. Describe how you can use the elimination algorithm to compute this distribution in $O(|\mathcal{X}|^4)$ time.