Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science

6.438 ALGORITHMS FOR INFERENCE
Fall 2014

Problem Set 2

Issued: Thursday, September 18, 2014   **Due:** Thursday, September 25, 2014

Suggested Reading: Lecture notes 3–5

Problem 2.1

Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with n vertices. Let each vertex $i \in \mathcal{V}$ be associated with a non-negative weight v_i and each edge $e \in \mathcal{E}$ be associated with a non-negative weight w_e.

(a) An independent set of the graph $\mathcal{G}$ is any subset of vertices such that no two vertices in that subset are connected by an edge in $\mathcal{E}$. For example, the empty set is a valid independent set of $\mathcal{G}$. Other examples of independent sets include singleton subsets of $\mathcal{V}$, i.e. $\{i\}$ for any $i \in V$. An independent set can be represented as a binary vector of dimension $n = |V|$ with $I = [I_i] \in \{0,1\}^{|V|}$ representing the subset of vertices $\{i \in V : I_i = 1\}$. Our interest is in the distribution over all independent sets of $\mathcal{G}$ with the probability of an independent set I proportional to:

$$P(I) \propto \exp\left(\sum_{i \in \mathcal{V}} v_i I_i\right).$$

    (i) List all of the independent sets of the graph depicted in Figure 2.1.

    (ii) Provide a representation of the distribution over independent sets for a general $\mathcal{G}$ in terms of an undirected graphical model.

(b) A matching of the graph $\mathcal{G}$ is any subset of edges such that no two edges in that subset share a vertex. For example, the empty set is a valid matching. Other examples of matchings include singleton subsets of $\mathcal{E}$, i.e. $\{e\}$ for any $e \in \mathcal{E}$. A matching can be represented as a binary vector of dimension $|\mathcal{E}|$ with $M = [M_e] \in \{0,1\}^{|\mathcal{E}|}$ representing the subset of edges $\{e \in \mathcal{E} : M_e = 1\}$. Our interest is in the distribution over all matchings of $\mathcal{G}$ with the probability of a matching M proportional to:

$$P(M) \propto \exp\left(\sum_{e \in \mathcal{E}} w_e M_e\right).$$

    (i) List all of the matchings of the graph depicted in Figure 2.1.

    (ii) For the graph given in Figure 2.1, draw an undirected graphical model representation of the distribution for matchings on that graph and give the potentials.
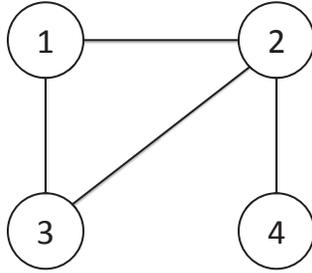
Figure 2.1

Problem 2.2

The Hammersley-Clifford theorem gives us a canonical method to turn an undirected graph into a factor graph. Specifically, assume that we are given a graph $\mathcal{G}$ and a strictly positive distribution P that is Markov with respect to $\mathcal{G}$. Then, Hammersley-Clifford tells us that P can be written as a product of potential functions for each maximal clique. Thus, to turn an undirected graph into a factor graph, we can simply define a factor node for each maximal clique in the graph. Figure 2.2 depicts an example graph and the associated factor graph.

Undirected Graph   Associated
Factor Graph
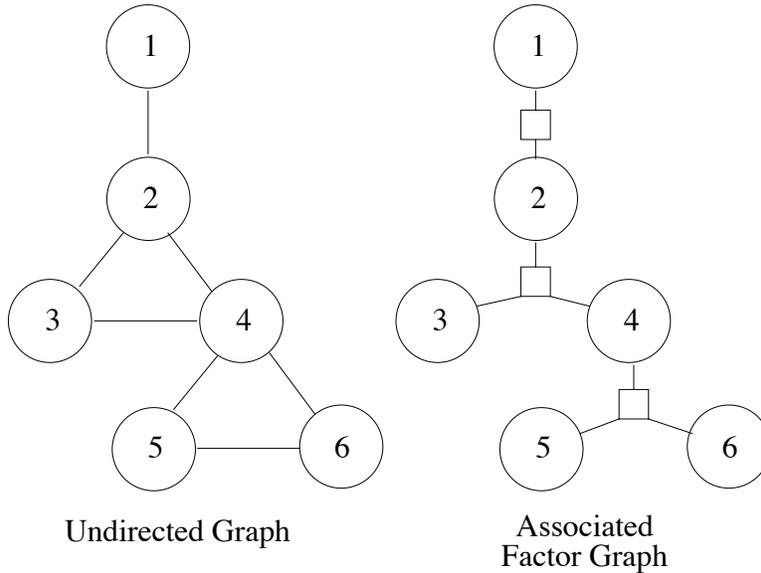
Figure 2.2

(a) Show that the factor graph produced by the canonical construction can be exponentially larger than the original graph. Specifically, show that there exists a constant $c > 1$ such that for all sufficiently large n, there exists an undirected graph with n vertices such that the associated factor graph has at least c^n vertices.

(b) **(Practice)** Show that there is a polynomial time algorithm, whose run time is no

2

greater than a polynomial function of the number of vertices n, that given as input an undirected graph $\mathcal{G}$, determines whether or not the factor graph associated to $\mathcal{G}$ is a tree.

Note that the naive algorithm that computes the factor graph and then checks if it is a tree is not polynomial time, because from part (a) we know that just writing down the factor graph can take exponential time.

In your solution, you may want to use the following concepts and results:

- Recall that a graph is *chordal* if any cycle of length 4 or more nodes has a chord, which is an edge joining two nodes that are not adjacent in the cycle. Testing the chordality of a graph can be done in linear time (for fun, try to come up with the algorithm).

- We use K_4 to denote a clique containing 4 nodes, and say a graph contains K_4 if it has a clique of 4 nodes. We also use $\widetilde{K_4}$ to denote the graph generated from deleting one edge from K_4, and say a graph contains $\widetilde{K_4}$ if it has such a subgraph. Testing whether a graph contains $\widetilde{K_4}$ can be done in polynomial time (e.g. a brute-force search would take as little time as $\binom{n}{4} = \mathcal{O}(n^4)$).

Problem 2.3

Consider the directed graphical model shown in Figure 2.3:
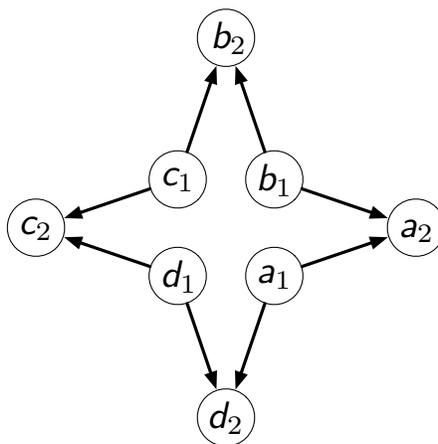
Figure 2.3

(a) You have four coins, labeled A, B, C, and D that have different biases. The outcome of any coin flip is Heads or Tails. You flip all four coins. Call these random outcomes a_1, b_1, c_1, and d_1.

Consider the following procedure:

Step 1. If $a_1 = b_1$, you flip subset $\mathcal{X}$ of the coins again.

Step 2. If $b_1 = c_1$, you flip subset $\mathcal{Y}$ of the coins again.

Step 3. If $c_1 = d_1$, you flip subset $\mathcal{Z}$ of the coins again.

Step 4. If $d_1 = a_1$, you flip subset $\mathcal{W}$ of the coins again.

Let a_2, b_2, c_2, and d_2 be states of coins A, B, C, and D, respectively, after the above procedure. Which coins are flipped at each step if the given directed graph is a perfect map for a_1, b_1, c_1, d_1, a_2, b_2, c_2, and d_2? For each step, determine the corresponding subset ($\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$) that needs to be flipped.

(b) List all conditional and unconditional independencies among a_2, b_2, c_2, and d_2. NOTE: You do not need to list any conditional or unconditional independencies that involve any of a_1, b_1, c_1, or d_1, and you should assume that those variables are unobserved.

(c) Draw a four-node, undirected graphical model that is a minimal I-map over the variables a_2, b_2, c_2, and d_2, assuming that outcomes a_1, b_1, c_1, and d_1 are unobserved.

Problem 2.4

Let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ be an undirected graph and let x be a collection of random variables defined on its nodes. Recall from class, that a distribution over x is *(globally) Markov* with respect to $\mathcal{G}$ if, for any disjoint subsets of nodes A, B, C such that B separates A from C, the statement $x_A \perp\!\!\!\perp x_C | x_B$ is satisfied. Now, here are two other notions of Markovianity.

A distribution is *pairwise Markov* with respect to $\mathcal{G}$ if, for any two nodes α and β not directly linked by an edge in $\mathcal{G}$, the corresponding variables x_α and x_β are independent conditioned on all of the remaining variables, i.e.

$$\forall \alpha, \beta \in \mathcal{V}, (\alpha, \beta) \notin \mathcal{E} : x_\alpha \perp\!\!\!\perp x_\beta | x_{\mathcal{V} \setminus \{\alpha, \beta\}}$$

A distribution is *locally Markov* with respect to $\mathcal{G}$ if any x_α, when conditioned on the variables on the neighbors of α, is independent of the remaining variables, i.e.

$$\forall \alpha \in \mathcal{V} : x_\alpha \perp\!\!\!\perp x_{\mathcal{V} \setminus \{\alpha\} \setminus \mathcal{N}(\alpha)} | x_{\mathcal{N}(\alpha)}$$

Decide if each of the following statements is true or false and provide your reasoning.

(a) If a distribution is globally Markov with respect to $\mathcal{G}$, then it is locally Markov with respect to $\mathcal{G}$.

(b) If a distribution is locally Markov with respect to $\mathcal{G}$, then it is globally Markov with respect to $\mathcal{G}$

(c) If a distribution is locally Markov with respect to $\mathcal{G}$, then it is pairwise Markov with respect to $\mathcal{G}$.

(d) If a distribution is pairwise Markov with respect to $\mathcal{G}$, then it is locally Markov with respect to $\mathcal{G}$.

Problem 2.5 (Practice)
Recall that a perfect map for a distribution P is a graph $\mathcal{G}$ (directed or undirected) such that P satisfies a conditional independence relationship if and only if this conditional independence relationship is implied by $\mathcal{G}$.

(a) Construct a distribution which has no perfect map. Specifically, no undirected or directed graph should be a perfect map for the distribution.

(b) Construct a directed graph and a distribution such that the directed graph is a perfect map for the distribution, but no undirected graph is a perfect map for the distribution.

(c) Do the reverse of part (b), i.e., construct an undirected graph and a distribution such that the undirected graph is a perfect map for the distribution, but no directed graph is a perfect map for the distribution.

(d) Prove that for any undirected graph $\mathcal{G}$, there exists some probability distribution P such that $\mathcal{G}$ is a perfect map for P.
    (*Hint:* Start with a collection of independent random variables $y_1, y_2, \ldots, y_N$, where N may be much larger than n, the number of vertices of $\mathcal{G}$. Then, for each vertex of $\mathcal{G}$, try to associate a cleverly chosen subset of the y_i.)

(e) Prove that for any DAG (directed acyclic graph) $\mathcal{G}$, there exists some probability distribution P such that $\mathcal{G}$ is a perfect map for P.

(f) This part is concerned with the opposite question to parts (b) and (c) – when can a distribution P have a perfect map that is an undirected graph and a perfect map that is a directed graph? Two equivalent answers to this question are stated below:

  (i) For an arbitrary undirected graph $\mathcal{G}$, there exists a directed graph that implies exactly the same conditional independencies as $\mathcal{G}$ if and only if $\mathcal{G}$ is chordal.

  (ii) For an arbitrary DAG $\mathcal{G}$, there exists an undirected graph that implies exactly the same conditional independencies as $\mathcal{G}$ if and only if moralizing $\mathcal{G}$ does not add any edges.

  Prove either statement (i) or (ii), i.e., you can prove whichever version you find easier to prove.

(g) Parts (b) and (c) show that there are distributions for which directed graphs can be perfect maps but for which no undirected graphs can be perfect maps, and vice versa. From parts (d) and (e), we also know that no undirected or directed graphs are "useless", i.e., there is always some distribution for which a given graph is a perfect map.

  Part (f) characterizes when a distribution can have a directed graph and an undirected graph as perfect maps. So, the final cases to consider are the following:

  (i) Can two different undirected graphs be perfect maps for the same distribution?
  (ii) Can two different directed graphs be perfect maps for the same distribution?

Show that the answer to part (i) is no, but that the answer to part (ii) is yes. Thus, we can throw away some directed graphs without changing the set of distributions that have directed graphs as perfect maps, but throwing away *any* undirected graph will shrink the set of distributions that have undirected graphs as perfect maps.

Problem 2.6
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:
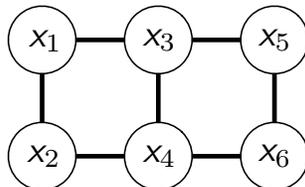
   No additional variables are allowed.
   Hint: Consider the effects of adding V-structures to your graph.

(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.

Problem 2.7 (Practice)
Recall that the factor graph associated with a directed graph has one factor for each local conditional defined on the graph. Similarly, the factor graph associated with an undirected graph has one factor for each potential defined on the graph. (To ensure that the undirected graph completely specifies the associated factor graph, let us assume that there are no potentials associated with non-maximal cliques.)

(a) Let $\mathcal{G}$ be a polytree, and let $\mathcal{G}_M$ be its moral graph. Let $\mathcal{F}$ denote the factor graph associated with $\mathcal{G}$, and let $\mathcal{F}_M$ denote the factor graph associated with $\mathcal{G}_M$. For every vertex i in $\mathcal{G}$ with no parents, add a factor f_i to $\mathcal{F}_M$ that is connected to the variable node i. Prove that $\mathcal{F}$ and $\mathcal{F}_M$ are identical. i.e., the factor graph associated with the

moral graph of a polytree is the same as the factor graph associated with the polytree, modulo the single-variable factors.

(*Hint:* Use induction. Work through the nodes in a topological ordering, building $\mathcal{G}_M$, $\mathcal{F}_M$ and $\mathcal{F}$.)

(b) Prove that the factor graph associated with a polytree is a factor tree.

6.438 Algorithms for Inference

Fall 2014