Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science

6.438 ALGORITHMS FOR INFERENCE
Fall 2014

Problem Set 1

Fall 2014

Issued: Tuesday, September 9, 2014 **Due:** Thursday, September 18, 2014

Suggested Reading: Lecture notes 1 to 3

Problem 1.1

(a) Let x and y be independent identically distributed random variables with common density function

$$p(\alpha) = \begin{cases} 1 & 0 \le \alpha \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

Let $\mathsf{s} = \mathsf{x} + \mathsf{y}$.

(i) Find and sketch $p_{\mathsf{s}}(s)$.

(ii) Find and sketch $p_{\mathsf{x}|\mathsf{s}}(x|s)$ vs. x, with s viewed as a known real parameter.

(iii) The conditional mean of x given $\mathsf{s} = s$ is

$$\mathbb{E}\left[\mathsf{x} \mid \mathsf{s} = s\right] = \int_{-\infty}^{+\infty} x\, p_{\mathsf{x}|\mathsf{s}}(x|s)\, dx.$$

Find $\mathbb{E}\left[\mathsf{x} \mid \mathsf{s} = 0.5\right]$.

(iv) The conditional mean of x given s (s viewed as a random variable) is

$$m_{\mathsf{x}|\mathsf{s}} = \mathbb{E}\left[\mathsf{x}|\mathsf{s}\right] = \int_{-\infty}^{+\infty} x\, p_{\mathsf{x}|\mathsf{s}}(x|\mathsf{s})\, dx.$$

Since $m_{\mathsf{x}|\mathsf{s}}$ is a function of the random variable s, it too is a random variable. Find the probability density function for $m_{\mathsf{x}|\mathsf{s}}$.

(b) Let x and y be independent identically distributed geometric random variables with parameter p. That is, x and y have common probability mass function

$$p[n] = p(1-p)^n, \;\; n = 0, 1, \ldots.$$

Let $\mathsf{s} = \mathsf{x} + \mathsf{y}$.

(i) Find and sketch $p_{\mathsf{s}}[s]$.

(ii) Find and sketch $p_{\mathsf{x}|\mathsf{s}}[x|s]$ vs. x, with s viewed as a known integer parameter.

(iii) Find the conditional mean of x given $\mathsf{s} = s$, which is defined as

$$\mathbb{E}\left[\mathsf{x} \mid \mathsf{s} = s\right] = \sum_{x=-\infty}^{\infty} x\, p_{\mathsf{x}|\mathsf{s}}[x|s].$$

(iv) The conditional mean of x given s (s viewed as a random variable) is

$$m_{\mathsf{x}|\mathsf{s}} = \mathbb{E}\left[\mathsf{x}|\mathsf{s}\right] = \sum_{x=-\infty}^{\infty} x\, p_{\mathsf{x}|\mathsf{s}}[x|\mathsf{s}].$$

Since $m_{\mathsf{x}|\mathsf{s}}$ is a function of the random variable s, it too is a random variable. Find the probability mass function for $m_{\mathsf{x}|\mathsf{s}}$.

Problem 1.2

An important concept that arises frequently in statistical signal processing, control, and machine learning is that of conditional independence. Specifically x and y might represent random variables or vectors of interest — e.g., they might be consecutive samples of a random sequence or one might be observed and the other the quantity that we wish to estimate based on that observation. Typically, there will be some statistical dependency between x and y (there better be statistical dependency if we want to use one of them to estimate the other!), and in many cases that statistical dependency is captured through the intermediary of another random variable or vector, z. What we mean by this is that x and y are conditionally independent given z:

$$p_{\mathsf{x},\mathsf{y}|\mathsf{z}}(x, y|z) = p_{\mathsf{x}|\mathsf{z}}(x|z)p_{\mathsf{y}|\mathsf{z}}(y|z)$$

(a) Suppose that z, w_1, and w_2 are independent random variables. Use these three random variables to construct two other random variables x and y that are not independent, but are conditionally independent given z.

(b) Show that x and y are conditionally independent given z if and only if the joint distribution for the three variables factors in the following form:

$$p_{\mathsf{x},\mathsf{y},\mathsf{z}}(x, y, z) = h(x, z)g(y, z) \ .$$

Problem 1.3 (Exercise 2.9 in Koller/Friedman)

Prove or disprove (by providing a counterexample) each of the following properties of independence.

(a) $\mathsf{x} \perp\!\!\!\perp (\mathsf{y}, \mathsf{w})|\mathsf{z}$ implies $\mathsf{x} \perp\!\!\!\perp \mathsf{y}|\mathsf{z}$.

(b) $x \perp\!\!\!\perp y|z$ and $(x, y) \perp\!\!\!\perp w|y$ imply $x \perp\!\!\!\perp w|z$.

(c) $x \perp\!\!\!\perp (y, w)|z$ and $y \perp\!\!\!\perp w|z$ imply $(x, w) \perp\!\!\!\perp y|z$.

(d) $x \perp\!\!\!\perp y|z$ and $x \perp\!\!\!\perp y|w$ imply $x \perp\!\!\!\perp y|(z, w)$.

Problem 1.4 (Practice)

Determine whether each of the following separate statements is TRUE or FALSE. If you answer TRUE, be sure to provide a proof; if you answer FALSE, be sure to provide a counter-example.

(a) $x \perp\!\!\!\perp (y, z)$ implies $x \perp\!\!\!\perp y|z$.
In other words, if x, y, z are three random variables such that x is independent of (y,z), then x is conditionally independent of y given z.

(b) $u \perp\!\!\!\perp v$ and $u \perp\!\!\!\perp w$ and $v \perp\!\!\!\perp w$ imply $u \perp\!\!\!\perp v|w$.
In other words, if u, v, w are a collection of pairwise-independent random variables, then u and v are conditionally independent given w.

Problem 1.5

Consider the following random variables. x_1, x_2 and x_3 represent the outcomes of three (independent) fair coin tosses. x_4 is the indicator function of the event that $x_1 = x_2$, and x_5 is the indicator function of the event that $x_2 = x_3$.

(a) Specify a directed graphical model (give the directed acyclic graph and local conditionals) that describes the joint probability distribution.

(b) List all conditional independencies that are implied by the graph.

(c) List any additional conditional independencies that are displayed by this probability distribution but are not implied by the graph.

(d) If the coins were biased, would your answer to part (c) change?

Problem 1.6

Consider the directed graphs shown in Figure 1.6.

(a) Determine the maximal set B for which $x_1 \perp\!\!\!\perp x_B|x_2$ for the graph $\mathcal{G}_1$.

(b) Determine the maximal set B for which $x_1 \perp\!\!\!\perp x_B|x_2$ for the graph $\mathcal{G}_2$.

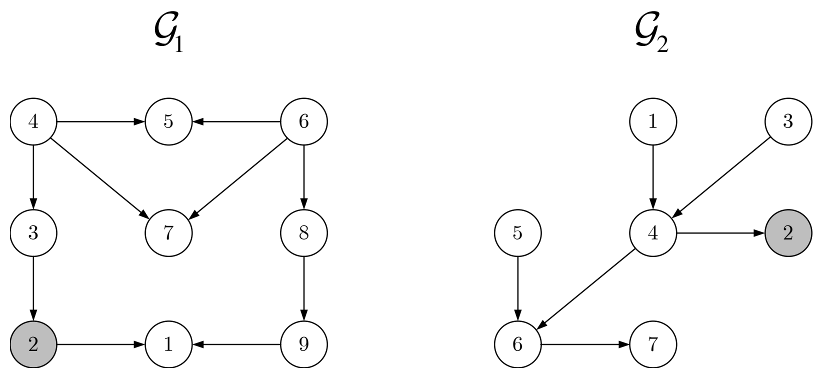
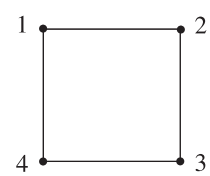


Figure 1.6

Problem 1.7
In this problem we'll show by example that the distribution of a graphical model need not have a factorization of the form in the Hammersley–Clifford Theorem if the distribution is not strictly positive. In particular, we'll take a look at a distribution on the following simple 4-node cycle: where at each node we have a binary random



variable, x_i, $i = 1, 2, 3, 4$. Consider a distribution $p(x_1, x_2, x_3, x_4)$ which assigns a probability of 1/8 to each of the following set of values of (x_1, x_2, x_3, x_4):

$$
\begin{array}{cccc}
(0,0,0,0) & (1,0,0,0) & (1,1,0,0) & (1,1,1,0) \\
(0,0,0,1) & (0,0,1,1) & (0,1,1,1) & (1,1,1,1)
\end{array}
$$

and a value of 0 to all other sets of values of (x_1, x_2, x_3, x_4).

(a) We first need to show that this distribution is Markov on our graph. To do this, it shouldn't be difficult to see that what we need to show are the following conditions:

- The pair of variables x_1 and x_3 are conditionally independent given (x_2, x_4)
- The pair of variables x_2 and x_4 are conditionally independent given (x_1, x_3)

We'll do this as follows:

(i) First, show that if we interchange x_1 and x_4 and interchange x_2 and x_3 we obtain the same distribution, i.e., $p(x_1, x_2, x_3, x_4) = p(x_4, x_3, x_2, x_1)$. This implies that if we can show the first of the conditions listed above, then the other is also true.

(ii) Now, show that whatever pair of values you choose for (x_2, x_4), we then know either x_1 or x_3 with certainty (e.g., what is the conditional distribution for x_1 conditioned on $(x_2, x_4) = (0, 1)$). Since we know either x_1 or x_3 with certainty, then conditioning on the other one of these obviously provides no additional information, trivially proving conditional independence.

(b) What we now need to show is that the distribution can't be factored in the way stated in the Hammersley–Clifford Theorem. We'll do this by contradiction. Noting that the maximal cliques in our graph are just the edges and absorbing the single node compatibility functions into the maximal clique terms and the proportionality constant $1/Z$ into any of these terms, we know that if our distribution has the factorization implied by Hammersley–Clifford, we can write it in the following form:

$$p(x_1, x_2, x_3, x_4) = \phi_{12}(x_1, x_2)\phi_{23}(x_2, x_3)\phi_{34}(x_3, x_4)\phi_{41}(x_4, x_1)$$

Show that assuming that our distribution has such a factorization leads to a contradiction by examining the values of $p(0, 0, 0, 0)$, $p(0, 0, 1, 0)$, and $p(0, 0, 1, 1)$, and $p(1, 1, 1, 0)$.