Probabilistic Graphical Models, Spring 2012

Problem Set 1: Probability review & Bayesian networks

Due: Thursday, February 9, 2012 at 5pm

---

1. A fair coin is tossed 4 times. Define X to be the number of heads in the first 2 tosses, and Y to be the number of heads in all 4 tosses.
   - Calculate the table of the joint probability $p(X, Y)$.
   - Calculate the tables of marginal probabilities $p(X)$ and $p(Y)$.
   - Calculate the tables of conditional probabilities $p(X \mid Y)$ and $p(Y \mid X)$.
   - What is the distribution of $Z = Y - X$?

2. You go for your yearly checkup and have several lab tests performed. A week later your doctor calls you and says she has good and bad news. The bad news is that you tested positive for a marker of a serious disease, and that the test is 99% accurate (i.e. the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative given that you don't have the disease). The good news is that this is a rare disease, striking only 1 in 10,000 people. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

3. Show that using a conditional parameterization based on cyclic graphs can lead to improper probability distributions. This is why Bayesian networks are defined on *acyclic* graphs.

   (Hint: recall that a probability distribution is defined by an outcome space and an assignment of non-negative numbers to each outcome that sum to 1.)

4. Show that the statement
   $$p(A, B|C) = p(A|C)p(B|C)$$
   is equivalent to the statement
   $$p(A|B, C) = p(A|C)$$
   and also to
   $$p(B|A, C) = p(B|C)$$
   (you need to show both directions, i.e. that each statement implies the other).

5. This problem investigates the way in which conditional independence relationships affect the amount of information needed for probabilistic calculations.

   (a) Suppose we wish to calculate $p(H|E_1, E_2)$, and we have no conditional independence information. Which of the following sets of numbers are sufficient for the calculation?

      i. $p(E_1, E_2), p(H), p(E_1|H), p(E_2|H)$.

      ii. $p(E_1, E_2), p(H), p(E_1, E_2|H)$.

      iii. $p(E_1|H), p(E_2|H), p(H)$.

   Provide justification for your answer.

(b) Suppose we know that E_1 and E_2 are conditionally independent given H. Now which of the above three sets are sufficient? Explain why.

6. Consider the following distribution over 3 binary variables X, Y, Z:

$$p(x,y,z) \quad = \quad \begin{cases} 1/12 & x \oplus y \oplus z = 0 \\ 1/6 & x \oplus y \oplus z = 1 \end{cases}$$

where $\oplus$ denotes the XOR function.

Show that there is no directed acyclic graph G such that $I_{d-sep}(G) = I(p)$.

7. Let X, Y, Z be binary random variables with a joint distribution that factorizes over the directed graph $X \to Z \leftarrow Y$ (v-structure). We define the following quantities:

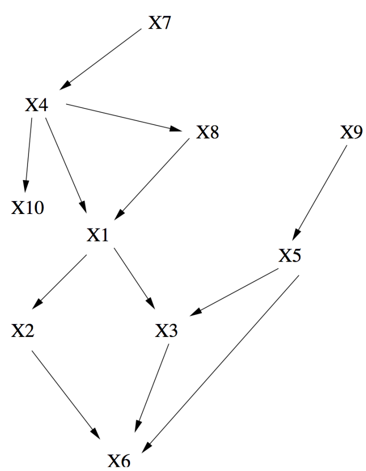
$$\begin{aligned} a =& p(X = 1) \\ b =& p(X = 1 \mid Z = 1) \\ c =& p(X = 1 \mid Z = 1, Y = 1) \end{aligned}$$

(a) For all the following cases, provide examples of conditional probability tables (table CPDs), and compute a, b, c, such that:

- $a > c$
- $a < c < b$
- $b < a < c$

(b) Think of X, Y as causes of Z, and for all the above cases summarize (in a sentence or two) why the statements are true for your examples.

8. **Markov blanket.** Let $\mathcal{X} = \{X_1, ..., X_n\}$ be a set of random variables with distribution p given by the following graph.

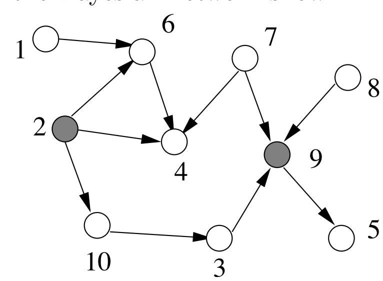


(a) Consider the variable X_1. What is the minimal subset of the variables, $A \subseteq \mathcal{X} - \{X_1\}$, such that $(X_1 \perp \mathcal{X} - A - \{X_1\} | A)$? Justify your answer.

(b) Now, generalize this to any BN defined by (G, p). Specifically, consider variable X_i. What is the *Markov blanket* of X_i? Namely, the minimal subset of variables $A \subseteq \mathcal{X} - \{X_i\}$ such that $(X_i \perp \mathcal{X} - A - \{X_i\} \mid A)$? Prove that this subset is necessary and sufficient.

(Hint: Think about the variables that X_i cannot possibly be conditionally independent of, and then think some more).

9. **D-separation.** Consider the Bayesian network shown in the below figure:



(a) For what pairs (i, j) does the statement $X_i \perp X_j$ hold? (Do not assume any conditioning in this part.)

(b) Suppose that we condition on $\{X_2, X_9\}$, shown shaded in the graph. What is the largest set A for which the statement $X_1 \perp X_A \mid \{X_2, X_9\}$ holds? The algorithm for d-separation given in Section 3.3.3 of Koller & Friedman may be helpful.

(c) What is the largest set B for which $X_8 \perp X_B \mid \{X_2, X_9\}$ holds?

10. Exercise 3.11 from Koller & Friedman.

In addition: (c) What happens if the algorithm that you gave in part (b) is used to remove the class variable in the naive Bayes model?

11. Exercise 3.15 from Koller & Friedman. Justify your answer.

12. Exercise 3.2 from Koller & Friedman.

13. Consider the Markov model given by $X_1 \to X_2 \to \ldots \to X_{n-1} \to X_n$, where $X_i \in \{0, 1\}$. The distributions $p(X_1), p(X_2 \mid X_1), \ldots, p(X_n \mid X_{n-1})$ are provided to us as tables.

(a) Give an algorithm to compute $p(X_i = 1)$ for all $i = 1 \ldots n$

(b) Give an algorithm to compute $p(X_i = 1 \mid X_1 = 1)$ for all $i = 1 \ldots n$

(c) Give an algorithm to compute $p(X_1 = 1 \mid X_i = 1)$ for all $i = 1 \ldots n$
(Hint: combine the results of (a) and (b))

All algorithms should have a running time that is $O(n)$. This is our first example of a probabilistic inference algorithm! Notice how you were able to take advantage of the graphical model structure to come up with a more efficient algorithm than naive marginalization.

Additional recommended exercises from Koller & Friedman book: 2.14, 2.15, 2.16, 2.20, 3.3, 3.6, 3.8, 3.16. Do not hand these in – they will not be graded.