Probabilistic Graphical Models, Spring 2013

Problem Set 1: Probability review & Bayesian networks

Due: Thursday, February 7, 2013 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 20,000 people. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

3. Show that 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).

4. This problem investigates the way in which conditional independence relationships affect the amount of information needed for probabilistic calculations. H, E_1, and E_2 are random variables and the notation $p(H)$ refers to the probability distribution for H, i.e. one number for every $h \in \mathrm{Val}(H)$.

   (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.

5. 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.

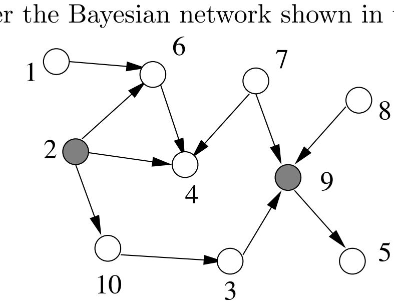
6. **Bayesian networks must be acyclic.** Suppose we have a graph $\mathcal{G} = (V, E)$ and discrete random variables $X_1, \ldots, X_n$, and define

$$f(x_1, \ldots, x_n) = \prod_{v \in V} f_v(x_v | x_{pa(v)}),$$

   where $pa(v)$ refers to the parents of variable X_v in $\mathcal{G}$ and $f_v(x_v \mid x_{pa(v)})$ specifies a distribution over X_v for every assignment to X_v's parents, i.e. $0 \leq f_v(x_v \mid x_{pa(v)}) \leq 1$ for all $x_v \in \text{Vals}(X_v)$ and $\sum_{x_v \in \text{Vals}(X_v)} f_v(x_v \mid x_{pa(v)}) = 1$. Recall that this is precisely the definition of the joint probability distribution associated with the Bayesian network $\mathcal{G}$, where the f_v are the conditional probability distributions.

   Show that if $\mathcal{G}$ has a directed cycle, f may no longer define a valid probability distribution. In particular, give an example of a cyclic graph $\mathcal{G}$ and distributions f_v such that $\sum_{x_1, \ldots, x_n} f(x_1, \ldots, x_n) \neq 1$. (A valid probability distribution must be non-negative and sum to one.) This is why Bayesian networks must be defined on *acyclic* graphs.

7. **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?