Exercises on Probabilistic Inference - CSE 473 – Spring 2006

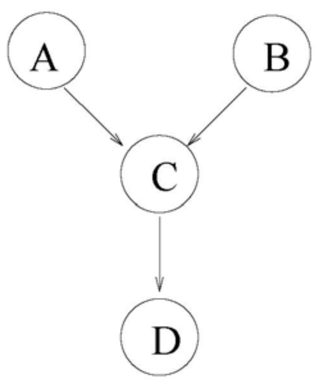
You should do these exercises, but they will not be turned in or graded. Instead, a solution sheet will be given out so you can check your own work. **You are responsible for understanding how to solve problems like these, and similar questions *will* appear on the final exam.**

1. The city of Metropolis is served by two taxi cab companies, the Blue Cab Company and the Green Cab Company. The Blue Cab Company has 100 cabs and the Green Cab Company has 10 cabs. One evening a pedestrian is injured by a hit-and-run taxi cab. He believes the cab that hit him was green, so he sues the Green Cab Company.

During the trial, the defense lawyer for the Green Cab Company produces an expert witness who testifies that in the evening twilight an eyewitness will confuse the colors blue and green of a moving vehicle about 25% of the time.

If the defense lawyer is correct, what is the probability that the Green Cab Company is in fact guilty?

2. Write down an expression for the joint probability distribution represented by the following Bayesian network. Next, write down *all* the independence and conditional independence statements over all sets of variables that are revealed by the following Bayesian network. Use the notation I(X,Y|Z) to indicate that variables X and Y are independent given the (possibly empty) set of evidence variables Z. For example, P(A,B|C,D) might be one such statement. Note that if your list contains I(X,Y|Z) it need not also contain I(Y,X|Z), since they mean the same thing.



3. Four generation of laboratory mice have been bred as follows:

Initial generation: Alice, Bob, Cindy, Dave, Ellen

Second generation:

- Fred has parents Alice and Bob
- Gwen has parents Cindy and Bob
- Henry has parents Cindy and Dave
- Iona has parents Ellen and Bob

Third generation:

- John has parents Gwen and Fred
- Katherine has parents Henry and Iona

Fourth generation:

- Louis has parents Katherine and John

Review of (simplified) genetics: Every individual has two copies of each gene, one inherited from each parent. Furthermore, the individual will pass one of the copies to any offspring with equal probability. Suppose there are two varieties of a particular gene, call them P and D. Then the genotype of an individual (considering just this gene) may be PP, PD, DP, or DD. In general there is no distinction in function between the genotypes PD and DP, so they may be considered to be a single type. Genes control the *traits* an individual exhibits as it grows.

A trait associated with a gene variety is *recessive* if in order for it to be exhibited the individual has to have two copies of that gene variety, *i.e.*, one from each parent. For example, if a particular trait associated with D is recessive, then the trait only shows up in individuals with genotype DD. Individuals with only one copy of the gene, *i.e.*, PD, are called *carriers*.

Being *albino* is a recessive trait.

Your task:

(a) Write a Bayesian network that corresponds to the information above about how the mice were bred, where each node corresponds to the genotype of a particular mouse. (Note that you do not need to distinguish PD from DP.) Assume that none of the initial generation of mice are albino, and for each mouse in the initial generation, the probability that it is a carrier is 1%. Fill out the CPTs (conditional probability tables) of the network.

(b) Suppose that Bob is the only carrier in the initial generation. What is the probability that John is albino? That John is a carrier? That Louis is albino?

(c) Suppose (instead) that you learn that Louis is albino. What is the probability that Bob is the only carrier in the initial generation? Note that you can compute this easily using Bayes' rule and your answer to the previous question.

(c) Suppose (instead) you start performing genetic testing on all of the mice to precisely determine their genotypes. Unfortunately, John escapes before he has been tested. So far you have gotten back the genetic results for Fred, Gwen, and Louis. In order to do the best possible job in *estimating* the probability distribution of John's genotype, do you need to get back the results on any more of the mice? If not, why? If so, which mouse or mice do you still need to get results about?