Q520: Third Homework: Bayesian Networks
Due: Tuesday, January 31, 2008

This homework will take more time than the last one for most people. In particular, the last one will be long because of lots of algebraic verification. So, my strong suggestion this time is to start soon. I'll plan to answer questions over the weekend, and quite a bit next week.

1. Consider the following data about children who visit a certain table at the library:

Name	Gen	Fl	Age
Pedro	m	c	3
Pierre	m	v	3
Trevor	m	v	4
Zhou	m	v	3
Sandy	f	c	3
Leila	f	c	3
Siobhan	f	v	3
Taneisha	f	v	3
Alvin	m	c	4
John	m	c	3

Name	Gen	Fl	Age
Sanjar	m	c	4
Rahim	m	v	4
Maxine	f	v	4
Cindy	f	c	4
Violet	f	c	4
Maria	f	c	3
Radha	f	c	4
Madison	f	v	4
Ibrahim	m	c	3
Gabrielle	f	v	4

We have recorded the Gender, favorite flavor, and age of the children. We regard the whole thing as a probability space, with all 20 children equally likely. We also regard Gen, Fl, and Age as random variables, with the evident value spaces.

(a) Show that Fl and Gen are not independent.

(b) Show that Fl and Gen are conditionally independent given Age.

[This one mainly is a matter of counting. If you're not able to get this one to work out, it's either a mistake in the tables or one in counting.]

2. Let X, Y, Z, and W be random variables on some space. Assume that X, Y, and W are independent given Z. Prove that X and Y are independent given Z.

[Hints: this is not as hard as it looks. We assume that all statements of the following form are true in our space:

$$\begin{aligned} & Pr(X = x, Y = y, W = w|Z = z) \\ = \; & Pr(X = x|Z = z) \cdot Pr(Y = y|Z = z) \cdot Pr(W = w|Z = z) \end{aligned} \tag{1}$$

and it's our job to show that all statements of the following form are true in our space:

$$\begin{aligned} & Pr(X = x, Y = y|Z = z) \\ = \; & Pr(X = x|Z = z) \cdot Pr(Y = y|Z = z) \end{aligned} \tag{2}$$

You may assume that the possible values of the random variables are all t and f, but this is not really necessary. (That is, it might not simplify things so much in this problem, but you can do it if you like.) You should start by saying "Fix values x, y, and z. We verify (2) for these values. " Then you will want to *use* some instances of (1). You will almost certainly need to use (1) *more than once*.

Another hint: first, forget about Z completely. This will make the notation easier. Then use subspaces to get the result when Z is present.]

3. There are several possible circumstances that might make John be late (L) for work. His alarm clock (A) could fail to go off, his children (C) could need extra time getting ready for school, and he could have delays at the subway (S). If his alarm clock goes off late, then (strangely enough) his children take care of themselves better, and so there is less of a chance that they will need extra time getting ready for school. Even if none of (A), (C), or (S) happen, there is still a small chance that John will end up late at work.

   Your task is to draw the dag for this, estimate some probabilities, and convince yourself that the conditional independence facts of your dag match your intuitions about what is happening in this example.

4. Consider the specification of Bayesian net, given below:

   Suppose that A, B and C have value spaces $\{a_1, a_2\}$, $\{b_1, b_2\}$, and $\{c_1, c_2\}$. Suppose that The tables for this Bayesian net would give the values of $\Pr(a_1)$, $\Pr(b_1|a_1)$, $\Pr(b_1|a_2)$, $\Pr(c_1|b_1)$, and $\Pr(c_1|b_2)$. The tables would also the other four values, but these may be deduced from the values above.

   (a) Your problem here is to calculate the actual Bayesian net for this example, following what we did in class. So you will need to calculate $\Pr(A = a_1)$, $\Pr(A = a_2)$, $\Pr(B = b_1)$, etc.

   (b) Recall that in class we said *how* to make an actual Bayesian net, and also that we would not *prove* that we actually got an actual Bayesian net. For a simple case, like this one, it is possible to completely check all of the conditional independence facts and therefore verify that we indeed have a Bayesian net. Do this.

   It will be helpful to adopt some abbreviations in this problem. Let's write v for $\Pr(a_1)$, so $(1 - x) = \Pr(a_2)$. Let's write w for $\Pr(b_1|a_1)$, x for $\Pr(b_1|a_2)$, y for $\Pr(c_1|b_1)$, and z for $\Pr(c_1|b_2)$.

You should write down the joint distribution explicitly. This will take some algebra.

You don't have to check *all* of the independence assertions that are part of the verification that we actually have a Bayesian net. But you should write down what facts do need to be checked, then pick some cases that you think are representative, and finally check those.