

Homework #3
(Probabilistic Inference)
Due: 10/6/11 (5:15 pm)

How to complete this HW: Either 1) type your answers in the empty spaces below each problem and print out this document, or 2) print this document and write your answers in the empty spaces on the printout. Return the homework in class, during office hours, or slip it under the door of Informatics East, Room 257 before 5:15 on Thursday, 10/6/11.

Your name: Enrique Areyan

Your email address:

Note on Honor Code: You must NOT look at previously published solutions of any of these problems in preparing your answers. You may discuss these problems with other students in the class (in fact, you are encouraged to do so) and/or look into other documents (books, web sites), with the exception of published solutions, without taking any written or electronic notes. If you have discussed any of the problems with other students, indicate their name(s) here:

N/A

Any intentional transgression of these rules will be considered an honor code violation.

General information: Justify your answers, but keep explanations short and to the point. Excessive verbosity will be penalized. If you have any doubt on how to interpret a question, tell us in advance, so that we can help you understand the question, or tell us how you understand it in your returned solution.

Grading:

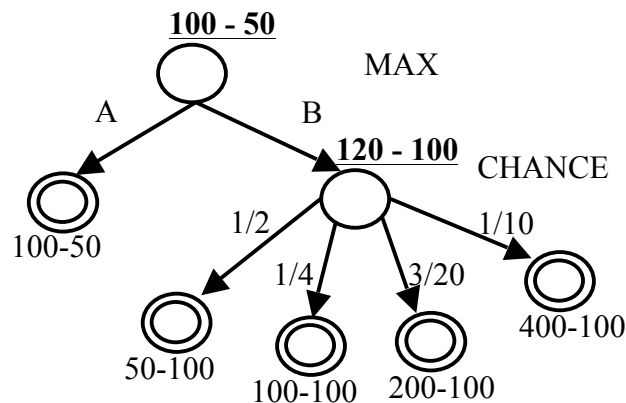
Problem#	Max. grade	Your grade
I	25	
II	25	
III	25	
IV	25	
Total	100	

I. Games of Chance (25 points)

You are the CEO of a large soda company. After decades of selling a certain brand of soda, whose recipe has been unchanged since it was invented in 1909, growth in sales has leveled off to a roughly constant level. Your product development team suggests that a new soda recipe, tailored for modern tastes, may help attract more customers.

Let action A denote the decision to keep the recipe the same. Historically, annual production costs are \$50 million and sales of the “classic” recipe are \$100 million. Let action B denote the decision to adopt the “new” recipe. Suppose that, in addition to the \$50 million in production costs, action B requires \$50 million that must be spent in advertising and changes to the manufacturing process. Your current belief is that there is a 50% chance that consumers will reject the new recipe and sales will be reduced by half, a 25% chance that sales will remain the same, a 15% chance that sales will double, and a 10% chance that sales will quadruple. (Assume that only the figures for the upcoming year matter for the purposes of this decision) Should your company change the recipe?

1. Draw the state tree representing this decision process, and label each node as MAX or CHANCE. Draw each terminal state as a double circle, labeled with ‘R-C’ where R are the rewards obtained and C are the costs incurred on the path from the root to the state (in \$million). Label edges from CHANCE nodes with the probabilities that the successor states. Back up the values of the intermediate nodes. What is the optimal action?



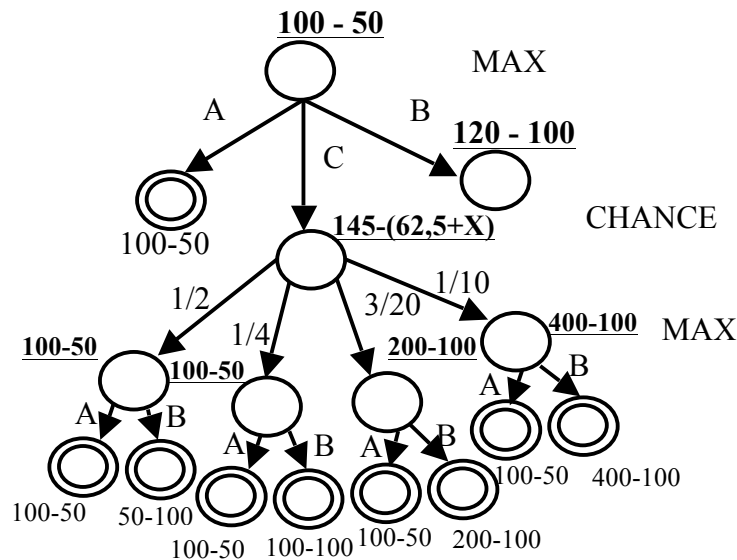
Back up values are in **bold and underline**.

$120 = \frac{1}{2} * 50 + \frac{1}{4} * 100 + \frac{3}{20} * 200 + \frac{1}{10} * 400$ (expected reward of action B)

The optimal action is A, given that it has the higher profit ($100 - 50 = 50$, whereas action B has $120 - 100 = 20$ profit)

2. Now consider a third action C, which performs extensive marketing research to determine what the actual change in sales would be if the company adopted the new recipe. Let the cost of this research be a variable x . Draw the new state tree representing this decision process. Label new terminal states as before (costs will a function of x).

Note: in the next state tree, node B is the same as before but omitted for simplicity.



3. Back up the values of the intermediate nodes as a function of x . How much would you be willing to spend on the marketing research? In other words, how high can you raise x before action C is no longer optimal?

The back up values can be seen **bold and underline** in the tree of question 2.

The next equation represents how much I will be willing to spend on marketing research:

$$145 - (62,5 + X) \geq 50. \text{ Solving for } X: X \geq -50 + 145 - 62,5 = 32.5$$

Spending more than 32.5 will yield action A at the top of the tree to be optimal.

II. Basic Probability (25 points)

1. Consider rolling two six-sided dice D1 and D2 where each face has a number 1-6. Each die is fair.

- a. Give a description of the space of possible worlds.

$$\Omega = \{(x,y): 1 \leq x \leq 6, 1 \leq y \leq 6\}, \text{ with } |\Omega| = 36$$

- b. What is the probability of the event that the total value of the two dice is 9 or more?

Let $E_{9\text{orMore}}$ denote the event that the total value of the two dice is 9 or more. Then,

$$E_{9\text{orMore}} = \{(x,y): x+y \geq 9\} \subset \Omega. \text{ Therefore,}$$

$$E_{9\text{orMore}} = \{(3,6),(6,3),(4,5),(5,4),(4,6),(6,4),(5,5),(6,5),(5,6),(6,6)\}$$

$$P(E_{9\text{orMore}}) = |E_{9\text{orMore}}| / |\Omega| = 10/36$$

- The probability that the total value is 9 or more AND D1 shows a 3 or less?

- The probability that the total value exceeds 9 or more OR D1 shows a 3 or less?

$$D1 = \{(x,y): 1 \leq x \leq 3\} \subset \Omega. \text{ Thus, } D1 = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),$$

$$(2,1),(2,2),(2,3),(2,4),(2,5),(2,6), (3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}. \text{ We have that } |D1| = 18.$$

$$P(D1) = |D1| / |\Omega| = 18/36.$$

From the following equation we can obtain answers to both questions:

$$P(E_{9\text{orMore}} \text{ or } D1) = P(E_{9\text{orMore}}) + P(D1) - P(E_{9\text{orMore}}, D1)$$

We know that there is only one case in which the first dice shows 3 or less and the sum is 9 or more, namely, (3,6). Therefore, $P(E_{9\text{orMore}}, D1) = 1/36$. Solving for:
 $P(E_{9\text{orMore}} \text{ or } D1) = 10/36 + 18/36 - 1/36 = 27/36$

2. Prove that the event that D1 shows a 3 or less is independent of the event that D2 shows a 5 or more.

We say that two events A and B are independent iff $P(A,B) = P(A)P(B)$.

We already calculated $P(D1)$. Now, $D2 = \{(x,y): y=5 \text{ or } y=6\} \subset \Omega$. Thus, $D2 = \{(1,5),(1,6), (2,5),(2,6), (3,5),(3,6), (4,5),(4,6), (5,5),(5,6), (6,5),(6,6)\}$. $P(D2) = 12/36$.

Looking at both D1 and D2, we can see that $D1 \cap D2 = (\{(1,5),(1,6), (2,5),(2,6), (3,5),(3,6)\})$. Therefore $P(D1 \cap D2) = 6/36$. On the other hand $P(A)P(B) = 18/36 \times 12/36 = 6/36$. The events are indeed independent.

3. Let D1 and D2 now denote the random variables indicating the value of the rolled die.
- What are $\text{Val}(D1)$ and $\text{Val}(D2)$? $\text{Val}(D1) = \text{Val}(D2) = \{1,2,3,4,5,6\}$
 - What is $P(D1,D2)$? You do not need to write it out explicitly, but give enough information to describe it *exactly*. $P(D1,D2) = 1/36$ for each value of D1 and D2. In other words $D1,D2 = (d1,d2)$, $1 \leq d1 \leq 6, 1 \leq d2 \leq 6$. $P(D1,D2) = |d1,d2|/|\Omega|$. But there is only one way in which $D1 = d1$ and $D2 = d2$ at the same time. Thus, $P(D1,D2) = 1/36$
 - Let T denote the random variable indicating the total values of the dice. What is $P(D1,D2,T)$? You do not need to write it out explicitly, but give enough information to describe it *exactly*. $P(D1,D2,T) = \langle \text{value } x \text{ for each } (d1,d2,t): 1 \leq d1 \leq 6, 1 \leq d2 \leq 6, 2 \leq t \leq 12 \rangle$

4.

- a. Prove that D1 and D2 are independent.

$P(D1,D2) = 1/36$ for every value of D1 and D2, and $P(D1) = P(D2) = 1/6$ for each value of D1 and D2. These are independent because $P(D1,D2) = P(D1)P(D2)$.

- b. Prove that D1 and T are not independent.

D1 can have values from 1 to 6 and T from 2 to 12. There are not enough values to make them coincide each time. Moreover, if we have a 1 in D1, we cannot have a 12 on T. Therefore, these events are not independent of each other.

- c. Prove that D1 and D2 are not independent given T.

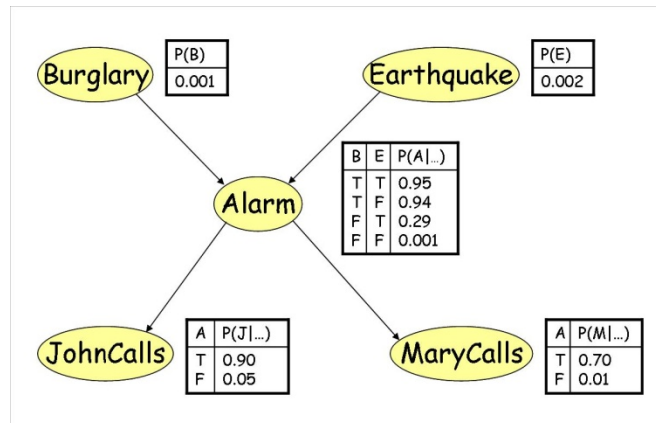
Two random variables A and B are conditionally independent given C, if

$$P(A, B|C) = P(A|C) P(B|C)$$

But, for values $D1=1, D2=1, T=3$, we have: $P(D1=1,D2=1 | T=3) = 0$ and $P(D1=1|T=3) \neq 0$ and $P(D2=1|T=3) \neq 0$. Thus, the conditional independence condition does not hold.

III. Inference in Bayesian Networks (25 points)

You are given the following Bayes net (it is the same example as used in class):



1. Before you make any observation, how would you compute the prior probability of MaryCalls? [We don't ask you to do the numerical computation. We ask you to express the probability of MaryCalls, $P(M)$, using quantities that are given in the belief net. To simplify notations, replace MaryCalls by M , Earthquake by E , etc... and let $P(M)$, $P(-M)$, $P(E)$, ... denote the probability of M , not M , E , etc...]

To compute the probability of Mary Calls, we need to marginalize over Alarm:

$$P(M) = P(M|A)P(A) + P(M|-A)P(-A)$$

But we will have to find $P(A)$ (and we will have $P(-A)$), which is given again by marginalization:

$$P(A) = P(A|B,E)P(B)P(E) + P(A|B,-E)P(B)P(-E) + P(A|-B,E)P(-B)P(E) + P(A|-B,-E)P(-B)P(-E)$$

All of these figures are in the belief net.

2. Now, assume that you observe that JohnCalls is True (i.e., $\Pr(J)=1$). How would you compute the new (posterior) probability of MaryCalls? [Again, we don't ask you for numerical values. Explain how you would do the computation. In this explanation, you will need to use the prior probability of JohnCalls, $P(J)$. No need to explain how you compute it, since its computation is very similar to that of $P(M)$ in Question 1.]

We want to compute $P(M|J)$, which from the network we can see that:

$$P(M|J) = P(M|A,J)P(A|J) + P(M|-A,J)P(-A|J), \text{ but } M,J \text{ are conditionally independent given } A, \text{ so:}$$

$$P(M|J) = P(M|A)P(A|J) + P(M|-A)P(-A|J). \text{ Both } P(M|A) \text{ and } P(M|-A) \text{ are in the network.}$$

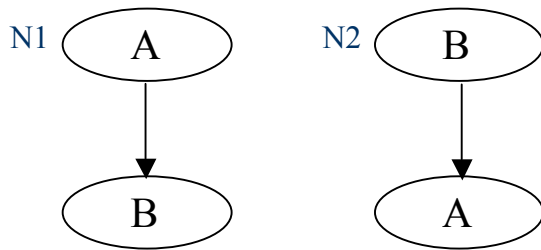
Using Bayes' rule: $P(A|J) = P(J|A)P(A)/P(J)$. The term $P(J|A)$ is on the network. $P(A)$ we already got from question 1.

By means of a reasoning analog to the calculation of $P(A)$ above:

$$P(J) = P(J|A)P(A) + P(J|-A)P(-A). \text{ Finally, we can use the fact that } P(-A|J) = 1 - P(A|J), \text{ which can be calculated from the network. Thus, we got all the components to calculate } P(M|J).$$

IV. Properties of Bayesian Networks (25pts)

1. Here you will prove that the following two Bayes nets $N1$ and $N2$ are equivalent; that is, they can represent the same joint probability distribution $P(A,B)$.



Assume A and B are Boolean variables with joint probability distribution:

	A,B	A,¬B	¬A,B	¬A,¬B
P(A,B)	a	b	c	d

Define the conditional probability tables of N1 and N2 so that their product equals P(A,B). *Hint: write down the entries of the CPTs as variables, and compute the product of the CPTs in algebraic form. Then solve for the variables as expressions of a,b,c,d.*

The idea is to find the unknown values the following CPTs in terms of a,b,c,d:

N1:

A	P(B A)
T	x
F	y

N2:

B	P(A B)
T	z
F	w

We have only two random variables, so:

$$P(A) = P(A,B) + P(A, -B) = a + b$$

$$P(-A) = P(-A,B) + P(-A,-B) = c + d$$

$$P(B) = P(B,A) + P(B, -A) = a + c$$

$$P(-B) = P(-B,A) + P(-B,-A) = b + d$$

Also, the following conditional probabilities hold:

For network N1:

$$P(A,B) = P(A)P(B|A) \Rightarrow a = (a+b)x \Rightarrow x = a/(a+b)$$

$$P(-A,B) = P(-A)P(B|-A) \Rightarrow c = (c+d)y \Rightarrow y = c/(c+d)$$

For network N2:

$$P(A,B) = P(A|B)P(B) \Rightarrow a = z(a+c) \Rightarrow z = a/(a+c)$$

$$P(A,-B) = P(A|-B)P(-B) \Rightarrow b = w(b+d) \Rightarrow w = b/(b+d)$$

The CPTs are:

N1:

A	P(B A)
T	a/(a+b)
F	c/(c+d)

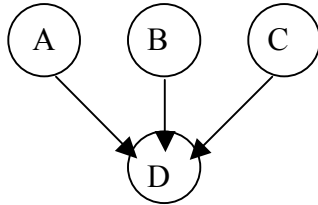
N2:

B	P(A B)
T	a/(a+c)
F	b/(b+d)

2. What are all the independence relationships can be inferred from the following Bayesian network? Write a list of statements in the form:

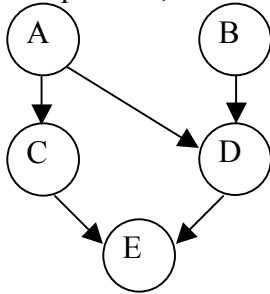
$$X \perp Y \mid Z_1, \dots, Z_n, [W_1, \dots, W_m]$$

where this means X is independent of Y , given ALL the evidence Z_1, \dots, Z_n , and optionally any of the evidence W_1, \dots, W_m .



$A \perp B \mid [C]; B \perp A \mid [C]; B \perp C \mid [A]; C \perp B \mid [A]; A \perp C \mid [B]; C \perp A \mid [B]$

3. Same question, but for the following Bayes net.



$A \perp B \mid [C]$	$B \perp A \mid [C]$
$E \perp A \mid C, D, [B]$	$A \perp E \mid C, D, [B]$
$E \perp B \mid D, A, [C]$	$B \perp E \mid D, A, [C]$
$C \perp D \mid A, [B, E]$	$D \perp C \mid A, [B, E]$
$B \perp C \mid [A, D, E]$	$C \perp B \mid [A, D, E]$