

INSTRUCTIONS

- **Due:** Wednesday, October 21, 2015 11:59 PM
- **Policy:** Can be solved in groups (acknowledge collaborators) but must be written up individually. However, we strongly encourage you to first work alone for about 30 minutes total in order to simulate an exam environment. Late homework will not be accepted.
- **Format:** Submit the answer sheet pdf containing your answers. You should solve the questions on this handout (either through a pdf annotator, or by printing, then scanning; we recommend the latter to match exam setting). **Make sure that your answers (typed or handwritten) are within the dedicated regions for each question/part. If you do not follow this format, we may deduct points.**
- **How to submit:** Go to www.gradescope.com. Log in and click on the class CS188 Fall 2015. Click on the submission titled HW 2 and upload your pdf containing your answers. If this is your first time using Gradescope, you will have to set your password before logging in the first time. To do so, click on "Forgot your password" on the login page, and enter your email address on file with the registrar's office (usually your @berkeley.edu email address). You will then receive an email with a link to reset your password.

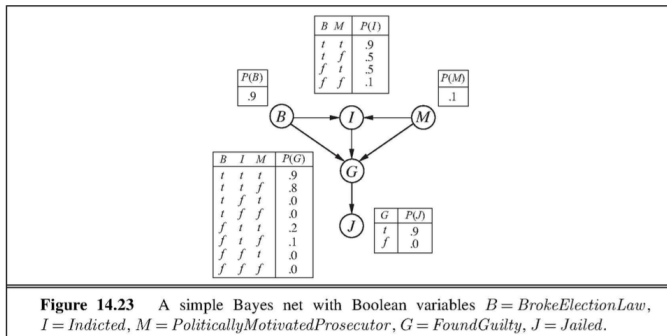
Last Name	
First Name	
SID	
Email	
Collaborators	

For staff use only

Q1	Q2	Q3	Total
/30	/40	/30	/100

1. (30 points) Bayes Net Structure and Construction

If a politician breaks election law (B), he might be indicted (I). His probability of being indicted is also influenced by whether or not the prosecuting lawyer is politically motivated (M). If he is indicted, he may be found guilty (G). The probability of being found guilty is also affected by whether he broke the law and whether his prosecutor is politically motivated. If he's found guilty, he has a high probability of going to jail (J). The Bayes net shown below describes this process:



- (a) [12 pt] For each of the following statements, write “True” if the statement must be true because of the network *structure*; else write “False”. If you write “True”, also write the conditional independence property which makes the statement true. If you write “False”, write the conditional independence assumption that would be necessary to make it true. You should not have to compute any probabilities to answer.

- i. [4 pt] $\mathbf{P}(B, I, M) = \mathbf{P}(B)\mathbf{P}(I)\mathbf{P}(M)$
False, would need $I \perp\!\!\!\perp B \perp\!\!\!\perp M$
- ii. [4 pt] $\mathbf{P}(J | G) = \mathbf{P}(J | G, I)$
True, $J \perp\!\!\!\perp I | G$.
- iii. [4 pt] $\mathbf{P}(M | G, B, I) = \mathbf{P}(M | G, B, I, J)$
True, $M \perp\!\!\!\perp J | G, B, I$.

- (b) [4 pt] Calculate the value of $P(b, i, \neg m, g, j)$.

$$\begin{aligned} P(b, i, \neg m, g, j) &= P(b)P(\neg m)P(i | b, \neg m)P(g | b, i, \neg m)P(j | g) \\ &= 0.9 \times 0.9 \times 0.5 \times 0.8 \times 0.9 \approx 0.2916 \end{aligned}$$

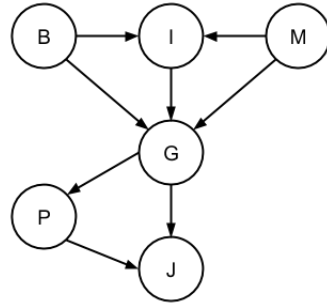
- (c) [4 pt] Calculate the probability that a politician goes to jail given that they broke the law, have been indicted, and face a politically motivated prosecutor.

$$\begin{aligned} P(j | b, i, m) &= \frac{\sum_{g'} P(j, b, i, m, g')}{\sum_{j', g'} P(j', b, i, m, g')} \\ &= \frac{\sum_{g'} P(b)P(m)P(i | b, m)P(g' | b, i, m)P(j | g')}{\sum_{j', g'} P(b)P(m)P(i | b, m)P(g' | b, i, m)P(j' | g')} \\ &= \frac{P(b)P(m)P(i | b, m) \sum_{g'} P(g' | b, i, m)P(j | g')}{P(b)P(m)P(i | b, m) \sum_{g'} P(g' | b, i, m) \sum_{j'} P(j' | g')} \\ &= \frac{\sum_{g'} P(g' | b, i, m)P(j | g')}{\sum_{g'} P(g' | b, i, m) \sum_{j'} P(j' | g')} \\ &= \frac{0.9 \times 0.9 + 0.1 \times 0}{0.9 \times 0.1 + 0.1 \times 0 + 0.9 \times 0.9 + 0.1 \times 1} \\ &\approx 0.81 \end{aligned}$$

- (d) [10 pt] After someone has been found guilty, the president can pardon them and guarantee that they cannot go to jail (though this is quite rare). Draw the new Bayes net with P = Presidential Pardon added to the network. Introduce edges from and to P to reflect its dependence on other variables in the Bayes net. Explain what conditional independence assumptions you made and why.

Then, fill in all conditional probabilities involving the new variable P with reasonable probabilities reflecting the real world. Explain how you chose the values in your probability table(s).

Add the node P and the directed edges $G \rightarrow P$ and $P \rightarrow J$:



Now $P(J | G)$ becomes $P(J | G, P)$, and we add the table $P(P | G)$.

Here are potential values for the probability tables:

G	$P(P G)$
g	0.01
$\neg g$	0

It is impossible for the president to pardon someone who is not found guilty, so $P(p | \neg g) = 0$. Given that someone has been found guilty, it's rare that the president pardons them, so $P(p | g) = 0.01$.

G	P	$P(J G, P)$
g	p	0
g	$\neg p$	0.9
$\neg g$	p	0
$\neg g$	$\neg p$	0

2. (40 points) Variable Elimination

- (a) [20 pt] Consider the variable elimination algorithm in Figure 14.11 of the textbook. Suppose you are given a Bayes net with the same variables and structure as the alarm Bayes net in Figure 14.2 (page 512) of your textbook, with the conditional probability tables given below.

$P(B)$
0.1

$P(E)$
0.1

B	E	$P(A B, E)$
b	e	0.8
b	$\neg e$	0.6
$\neg b$	e	0.6
$\neg b$	$\neg e$	0.1

A	$P(J A)$
a	0.8
$\neg a$	0.1

A	$P(M A)$
a	0.6
$\neg a$	0.1

Apply the algorithm to the query $\mathbf{P}(B | j, m)$. You will have to eliminate two variables, in any order. For *both*, include *all of the following*:

- 2 pt per variable: The name of the variable you are eliminating (e.g., A) and the variables involved in the resulting factor (e.g., $f_1(B, C)$)
- 2 pt per variable: The summation to calculate the factor (e.g., $f_1(B, C) = \sum_a \mathbf{P}(B)P(a | B)\mathbf{P}(C | a)$)
- 3 pt per variable: The values in the factor table.

Finally, you must multiply the remaining factors to get a final answer. Name the variables involved in the final resulting factor (2 pts). Find the normalizing constant to make it a probability distribution (1 pt), and write down the new probability table (3 pts).

The factors we initially begin with are $\mathbf{P}(B)$, $\mathbf{P}(E)$, $\mathbf{P}(A | B, E)$, $P(j | A)$, $P(m | A)$.

Eliminating variable E :

$$f_1(A, B) = \sum_{e'} P(e')(A | B, e') = P(e)P(A | B, e) + P(\neg e)P(A | B, \neg e)$$

A	B	$f_1(A, B)$
a	b	$0.1 \times 0.8 + 0.9 \times 0.6 = 0.08 + 0.54 = 0.62$
a	$\neg b$	$0.1 \times 0.6 + 0.9 \times 0.1 = 0.06 + 0.09 = 0.15$
$\neg a$	b	$0.1 \times 0.2 + 0.9 \times 0.4 = 0.02 + 0.36 = 0.38$
$\neg a$	$\neg b$	$0.1 \times 0.4 + 0.9 \times 0.9 = 0.04 + 0.81 = 0.85$

We now have the factors $\mathbf{P}(B)$, $f_1(A, B)$, $P(j | A)$, and $P(m | A)$ remaining.

Eliminating variable A :

$$f_2(j, m, B) = \sum_{a'} f_1(a', B)P(m | a')P(j | a') = f_1(a, B)P(m | a)P(j | a) + f_1(\neg a, B)P(m | \neg a)P(j | \neg a)$$

B	$f_2(j, m, B)$
b	$0.62 \times 0.6 \times 0.8 + 0.38 \times 0.1 \times 0.1 \approx 0.30$
$\neg b$	$0.15 \times 0.6 \times 0.8 + 0.85 \times 0.1 \times 0.1 \approx 0.08$

Now the remaining factors are $\mathbf{P}(B)$ and $f_2(j, m, B)$. Multiplying them together we get

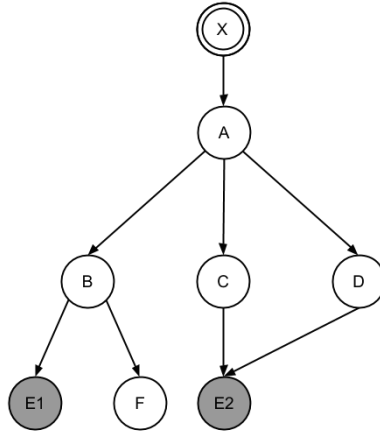
$$f_3(j, m, B) = f_2(j, m, B)\mathbf{P}(B)$$

B	$f_3(j, m, B)$
b	$0.1 \times 0.3 = 0.03$
$\neg b$	$0.9 \times 0.08 = 0.072$

We normalize this factor by dividing by the sum of its table entries: $0.03 + 0.072 = 0.102$. This gives a final table of

B	$\mathbf{P}(B j, m)$
b	0.29
$\neg b$	0.71

- (b) [20 pt] Consider the Bayes net below. Suppose we are trying to compute the query $\mathbf{P}(X \mid e_1, e_2)$. Assume all variables are binary.



- i. [4 pt] Suppose we choose to eliminate variables in the order A, B, C, D, F . What is the largest factor resulting from summing out over one of these variables? How many entries are in its table? Assume that we separately store pairs of numbers which we know sum to one.

There are two possible answers: $f_1(X, B, C, D)$, from eliminating variable A , or $f_2(C, X, D, F, e_1)$, from summing out B . Both have $2^4 = 16$ entries in their tables.

- ii. [16 pt] An optimal variable elimination ordering is one which minimizes the sum of the sizes of factors generated. What is one optimal variable elimination ordering? For each variable eliminated in this order, include the resulting factor and its size. The optimal ordering will be worth 6 points. Each variable's entry will be worth 2 points.

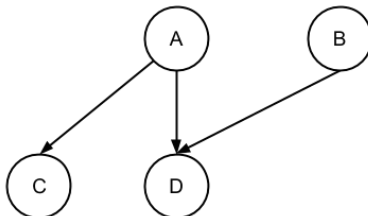
An optimal variable elimination order is any one which eliminates A last. For example, F, B, C, D, A :

- $F \rightarrow f_1(B) = 1$: size 1 (or size 0)
- $B \rightarrow f_2(A, e_1)$: size 2
- $C \rightarrow f_3(A, e_2, D)$: size 4
- $D \rightarrow f_4(A, e_2)$: size 2
- $A \rightarrow f_5(X, e_1, e_2)$: size 2

3. (30 points) Irrelevance Criteria

A variable in a Bayes net is said to be *irrelevant* to a given query if we could remove the variable from the Bayes net without changing the answer to the query. For example, in the Bayes net $A \rightarrow B$, the variable B is irrelevant to the query $\mathbf{P}(A)$, because we already have the $\mathbf{P}(A)$ table. On the other hand, A is *not* irrelevant to the query $\mathbf{P}(B)$, because $\mathbf{P}(A)$ is needed to compute $\mathbf{P}(B)$.

(a) [10 pt] Consider the Bayes net below:



i. [5 pt] Suppose we are making the query $\mathbf{P}(A \mid D = d)$. Prove that C is irrelevant to this query using the following steps:

- 1 pt: Write the full joint distribution as a product of the CPTs in the Bayes net
- 1 pt: Sum over this product and normalize to get $\mathbf{P}(A \mid D = d)$
- 3 pt: Show algebraically that this expression does not depend on the variable C .

The full joint distribution is given by

$$\mathbf{P}(A, B, C, d) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C \mid A)\mathbf{P}(d \mid A, B)$$

We know that $\mathbf{P}(A \mid d) = \alpha \mathbf{P}(A, d)$:

$$\mathbf{P}(A \mid d) = \frac{1}{\alpha} \sum_{b', c'} \mathbf{P}(A)P(b')P(c' \mid A)\mathbf{P}(d \mid A, b')$$

We can pull c' out of the summation:

$$\mathbf{P}(A \mid d) = \frac{1}{\alpha} \sum_{c'} P(c' \mid A) \sum_{b'} P(A)P(b')P(D \mid A, b')$$

$\sum_{c'} P(c' \mid A) = 1$, so this just simplifies to

$$P(A \mid d) = \frac{1}{\alpha} \sum_{b'} P(A)P(b')P(D \mid A, b')$$

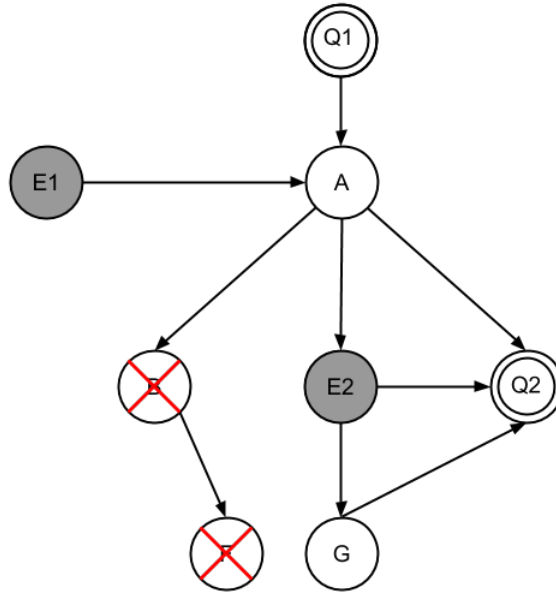
Once we have built the $P(A, d)$ table, we can simply sum over the entries and normalize by that value to get $P(A \mid d)$, without ever using the table $P(C \mid A)$.

ii. [5 pt] Suppose we are making the query $\mathbf{P}(C \mid D)$. Execute the first two steps in part i) for this query, and then argue why B is not irrelevant.

$$\mathbf{P}(C \mid D) = \frac{1}{\alpha} \sum_{a', b'} P(a')P(b')\mathbf{P}(C \mid a')\mathbf{P}(D \mid a', b')$$

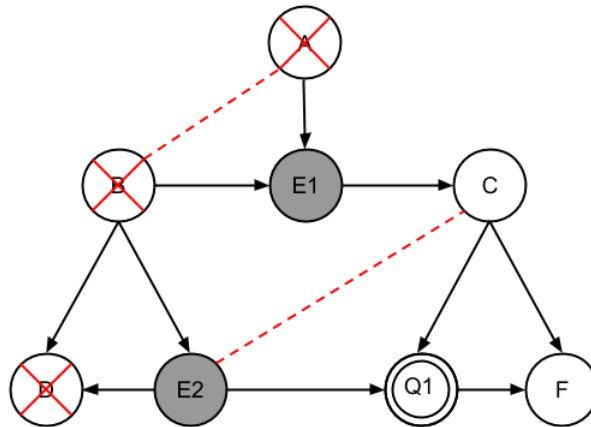
We can't pull out a factor involving B out of the summation that cancels to 1, so B is not irrelevant to the query.

(b) [5 pt] The *ancestor criterion* says that any node which is not an ancestor of a query or evidence variable is irrelevant. In the Bayes net below, query variables are indicated by a double circle and evidence variables are shaded in. Cross out all the nodes that are irrelevant to this query according to the ancestor criterion.



- (c) [10 pt] The moral graph of a Bayes net is an undirected graph containing all of the same connections as the original Bayes net, plus edges that connect variables which shared a child in the original Bayes net. Another criterion of irrelevance says that X is irrelevant to the query $\mathbf{P}(Q_1 \dots Q_n \mid e_1 \dots e_n)$ if in the moral graph, every path between a query variable Q_i and X goes through some evidence variable e_j (i.e., X is *m-separated* from the query variables given the evidence variables).

For the following Bayes net, draw in the additional edges found in the moral graph. Then cross out all the variables that are irrelevant to the query according to the m-separation criterion. Finally, list on the side the variables that *are not* considered irrelevant by the ancestor criterion but *are* considered irrelevant by the m-separation criterion.



A and B are considered irrelevant by the m-separation criterion but are not considered irrelevant by the ancestor criterion.

- (d) [5 pt] The *Markov blanket* of a variable X in a Bayes net is the set of X 's parents, children, and its children's parents. Explain why the moral graph can also be defined as an undirected graph in which an edge (X, Y) exists if and only if X is in the Markov blanket of Y and vice versa. (Be sure to clearly explain *both implications*.)

$(X, Y) \in \text{moral graph} \implies X, Y \text{ in each other's Markov blankets}$: If (X, Y) is in the moral graph, it must be the case that either

- $X \rightarrow Y$ or $Y \rightarrow X$ existed in the original Bayes net, or
- $X \rightarrow A$ and $Y \rightarrow A$ existed in the original Bayes net

In both of these cases X and Y would be in each other's Markov blankets by the definition given above.

$X \text{ in } Y\text{'s Markov blanket} \implies (X, Y) \text{ in moral graph}$: If X is an element of the Markov blanket of Y , then X is either a parent or child of Y , or is a parent of one of Y 's children. If X is a parent or child of Y , (X, Y) is present in the moral graph because all the original Bayes net edges were present. If X is a parent of one of Y 's children, the edge is present because all nodes which share a child are "married" in the moral graph (hence the name).