15-381	Artificial Intelligence	HW11

INSTRUCTIONS

- **Due: Thursday, 25 April 2019 at 10:00 PM EDT.** Remember that you have NO slip days for Written Homework, but you may turn it in up to 24 hours late with 50% Penalty.
- **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). 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:** Submit a pdf with your answers on Gradescope. Log in and click on our class 15-381 and click on the submission titled HW11 and upload your pdf containing your answers.
- **Policy:** See the course website for homework policies and Academic Integrity.

Name	
Andrew ID	

For staff use only

Q1	Q2	Q3	Q4	Total
/25	/20	/35	/20	/100

Q1. [25 pts] Variable Elimination

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.

B	$P(B)$
$+b$	0.1

E	$P(E)$
$+e$	0.1

A	B	E	$P(A \mid B, E)$
$+a$	$+b$	$+e$	0.8
$+a$	$+b$	$-e$	0.6
$+a$	$-b$	$+e$	0.6
$+a$	$-b$	$-e$	0.1

J	A	$P(J \mid A)$
$+j$	$+a$	0.8
$+j$	$-a$	0.1

M	A	$P(M \mid A)$
$+m$	$+a$	0.6
$+m$	$-a$	0.1

Apply the algorithm to the query $P(B \mid +j,\ +m)$. You will have to eliminate the variables E and A, in that order. For each variable, write:

(i) the variables involved in the resulting factor (e.g., for eliminating X the factor might be $f_1(Y, Z)$, so the variables involved are Y, Z),

(ii) the summation to calculate the factor (e.g., $f_1(Y, Z) = \sum_x P(Y)P(x \mid Y)P(Z \mid x)$), and

(iii) the numeric values in the factor table.

(a) [6 pts] Eliminating E:

(i) Variables:

(ii) Summation:

(iii) Factor table:

(b) [6 pts] Eliminating A:

(i) Variables:

(ii) Summation:

(iii) Factor table:

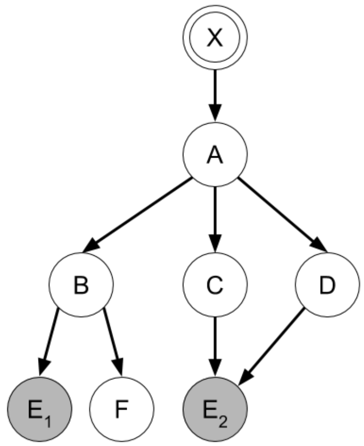
(c) [6 pts] Finally, you must multiply the remaining factors to get a final table to answer the query. Multiply the remaining factors to get the final resulting factor, and fill out the values in the corresponding factor table.

(i) Final factor:	**(ii) Factor table:**

Now, find the normalizing constant to make this factor the probability distribution we want, $P(B \mid j, m)$. Write out the values of this normalized probability table.

(iii) Constant:	**(iv) Probability table:**

Now consider the Bayes net below. Suppose we are trying to compute the query $P(X \mid e_1, e_2)$. Assume all variables are binary.



(d) [4 pts] Suppose we choose to eliminate variables in the order A, B, C, D, F.
Of the factors resulting from summing out over each of these variables, which factor has the most entries in its corresponding table? How many entries are in its table? Assume that we have separate entries for pairs of numbers even if we know sum to one (e.g., we would store both $P(X = +)$ and $P(X = -)$).

(i) Factor:	**(ii) # entries:**

(e) [3 pts] An optimal variable elimination ordering is one which minimizes the sum of the sizes of factors generated.

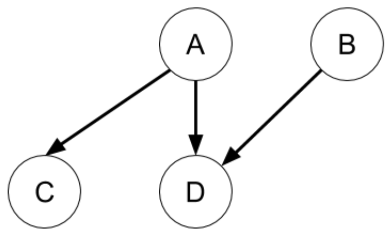
Fill in the table below with an optimal variable elimination ordering. For each variable, include the resulting factor and the number of entries in its table, again assuming that we separately store pairs of numbers which sum to one.

Variable	Factor	# Entries

Q2. [20 pts] 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 $A \to B$, 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 $\mathbf{P}(B)$.

Consider the Bayes net below:



(a) [9 pts] Suppose we are making the query $\mathbf{P}(A \mid D = d)$. Prove that C is irrelevant to this query using the following steps:

(i) Write the full joint distribution as a product of the CPTs in the Bayes net.

(ii) Sum over this product and normalize to get $\mathbf{P}(A \mid d)$.

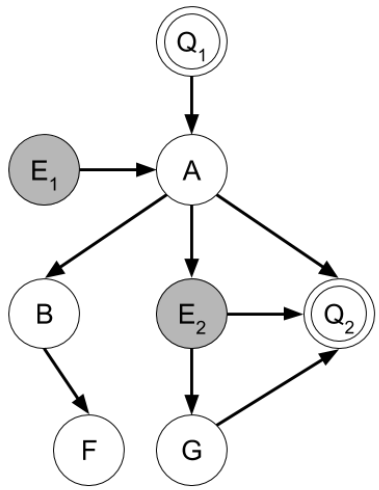
(iii) Explain why this expression does not depend on the variable C.

(b) [3 pts] Now suppose we are making the query $\mathbf{P}(C \mid D)$. Execute the first two steps in part (a) for this query, and then argue why B is *not* irrelevant.

The *ancestor criterion* says that any node which is not an ancestor of a query or evidence variable is irrelevant.

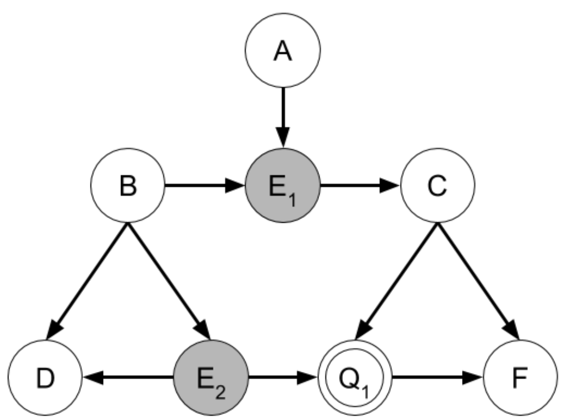
In the Bayes net below, query variables are indicated by a double circle and evidence variables are shaded in.

(c) [3 pts] Cross out all the nodes that are irrelevant to this query according to the ancestor criterion.



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 $\mathbf{P}(Q_1 \ldots Q_n \mid e_1 \ldots e_n)$ 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).

(d) [3 pts] For the following Bayes net, **draw in** the additional edges found in the moral graph **and cross out** all the variables that are irrelevant to the query according to the m-separation criterion.



(e) [2 pts] Which variables *are not* considered irrelevant by the ancestor criterion but *are* considered irrelevant by the m-separation criterion?

Variables:

Q3. [35 pts] Dynamic Bayes Net and Hidden Markov Model

A professor wants to know if students are getting enough sleep. Each day, the professor observes whether the students sleep in class, and whether they have red eyes. Let S_t be the random variable of the student having enough sleep, R_t be the random variable for the student having red eyes, and C_t be the random variable of the student sleeping in class on day t. The professor has the following domain theory:

- The prior probability of getting enough sleep at time t, with no observations, is 0.6
- The probability of getting enough sleep on night t is 0.9 given that the student got enough sleep the previous night, and 0.2 if not
- The probability of having red eyes is 0.1 if the student got enough sleep, and 0.7 if not
- The probability of sleeping in class is 0.2 if the student got enough sleep, and 0.4 if not

(a) [5 pts] The professor wants to formulate this information as a Dynamic Bayesian network. Provide a diagram and complete probability tables for the model.

DBN diagram:

Probability tables:

(b) [30 pts] Using the DBN you defined and for the evidence values

$-r_1, -c_1$ = not red eyes, not sleeping in class

$+r_2, -c_2$ = red eyes, not sleeping in class

$+r_3, +c_3$ = red eyes, sleeping in class

perform the following computations:

(i) State estimation: Compute $P(S_t|r_{1:t}, c_{1:t})$ for each of $t = 1, 2, 3$

$P(S_1 \mid -r_1, -c_1)$

$P(S_2 \mid r_{1:2}, c_{1:2})$

$P(S_3 \mid r_{1:3}, c_{1:3})$

(ii) Smoothing: Compute $P(S_t|r_{1:3}, c_{1:3})$ for $t = 2, 3$. (Hint: AIMA pg. 574)

We can build upon part (b) above, starting with $P(S_3|r_{1:3}, c_{1:3})$.

$P(S_2|r_{1:3}, c_{1:3})$

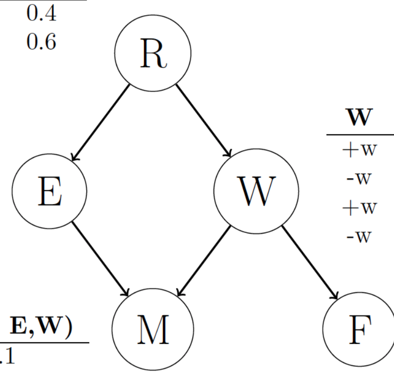
At every time step t for t = 1 to n, you observe a tuple (r_t, c_t) telling you whether the student had red eyes and whether they were sleeping in class. Given these observations and $P(S_k|r_{1:k}, c_{1:k})$, find an expression for $P(S_k|r_{1:n}, c_{1:n})$, where $0 \leq k \leq n$. You may only use the probability tables in the DBN and $P(S_k|r_{1:k}, c_{1:k})$.

$P(S_k|r_{1:n}, c_{1:n})$

Q4. [20 pts] Sampling

Consider the following Bayes Net and corresponding probability tables.

R	P(R)
+r	0.4
-r	0.6



E	R	P(E | R)
+e	+r	0.3
-e	+r	0.7
+e	-r	0.6
-e	-r	0.4

W	R	P(W | R)
+w	+r	0.9
-w	+r	0.1
+w	-r	0.2
-w	-r	0.8

M	E	W	P(M | E,W)
+m	+e	+w	0.1
-m	+e	+w	0.9
+m	+e	-w	0.45
-m	+e	-w	0.55
+m	-e	+w	0.35
-m	-e	+w	0.65
+m	-e	-w	0.9
-m	-e	-w	0.1

F	W	P(F | W)
+f	+w	0.15
-f	+w	0.85
+f	-w	0.75
-f	-w	0.25

After sampling to approximate the query $P(R|+f,+m)$, we have the following 3 samples:

$$(+r,+e,-w,+m,+f) \quad (+r,-e,+w,-m,+f) \quad (+r,+e,-w,+m,+f)$$

(a) [12 pts] Fill in the following table with the probabilities of *drawing* each respective sample given that we are using each of the following sampling techniques. (Hint: $P(+f,+m) = .181$)

Method	$P(+r,+e,-w,+m,+f \mid method)$	$P(+r,-e,+w,-m,+f \mid method)$
prior sampling		
rejection sampling		
likelihood weighting		

(Continued on next page)

(b) [8 pts] We are going to use Gibbs sampling to estimate the probability of getting the third sample$(+r, +e, -w, +m, +f)$. We will start from the sample $(-r, -e, -w, +m, +f)$ and resample E first then R. What is the probability of drawing sample $(+r, +e, -w, +m, +f)$?

Answer: