COMP341 Introduction to Artificial Intelligence
HW4 - Probability and Bayesian Networks

- By submitting this homework, you **agree** to fully comply with Koç University Student Code of Conduct, and accept any punishment in case of failure to comply.
- You are expected to be able provide legible, clear and concise answers. Gibberish and unreadable answers will not receive any credit. Write only what is relevant.

The deadline is December 8^{th} at 16.59PM. You should hand-in your solutions to me in person (slip it under my door if I am not there) or one of the TAs **before**, or upload it to blackboard (must be readable) before the deadline. If you miss me or the TAs on Friday, upload it to blackboard. Otherwise, we are not going to accept it by hand.

Q1 (35 Points) Probability: You have done 1000 experiments and measured 3 boolean variables, namely M_1, M_2 and X. The table below summarizes the number of times you have observed a certain combination.

M_1	M_2	X	Counts
+	+	+	95
+	+	-	94
+	-	+	385
+	-	-	31
-	+	+	12
-	+	-	113
-	-	+	132
-	-	-	138

Part a (6 points) Calculate the following probabilities:

$$P(M_1 = +) = \frac{95 + 94 + 385 + 31}{1000} = 0.605$$

$$P(M_2 = +) = \frac{95 + 94 + 12 + 113}{1000} = 0.314$$

$$P(X = +) = \frac{95 + 385 + 12 + 132}{1000} = 0.624$$

Part b (20 points) Fill in the following probability tables:

M_1	M_2	$P(M_1, M_2)$
+	+	(95+94)/1000=0.189
+	-	(385+31)/1000=0.416
-	+	(12+113)/1000=0.125
-	-	(132+138)/1000=0.27

M_1	X	$P(M_2 = + \mid M_1, X)$
+	+	95/(95+385)=0.1979
+	-	94/(94+31)=0.752
-	+	12/(12+132)=0.0833
-	-	113/(113+138)=0.4502

M_1	$P(M_2 = + \mid M_1)$
+	(95+94)/(95+94+385+31)=0.3124
-	(12+113)/(12+113+132+138)=0.3165

Part c (9 points) You know that X is either the common affect of the common cause of M_1 and M_2. How can you decide which with the given probabilities? Justify your answer using probability values. Draw the corresponding Bayesian Network based on your answer.

If X is the common effect than we have $P(M_1, M_2) = P(M_1)P(M_2)$, i.e., M_1 and M_2 are independent. If X is the common cause, then $P(M_2|M_1, X) = P(M_2|X)$, i.e., M_1 and M_2 are conditionally independent given X. Let's check the common effect case first:

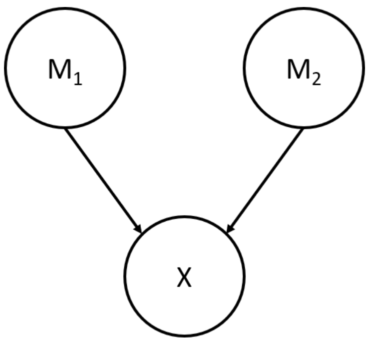
M_1	M_2	$P(M_1, M_2)$	$P(M_1)P(M_2)$
+	+	0.189	$0.605 \cdot 0.314 = 0.19$
+	-	0.416	$0.605 \cdot (1 - 0.314) = 0.415$
-	+	0.125	$(1 - 0.605) \cdot 0.314 = 0.124$
-	-	0.27	$(1 - 0.605) \cdot (1 - 0.314) = 0.271$

From this table, we can easily conclude that M_1 and M_2 are independent!.
We are already done but completeness sake, let's check the common cause case:

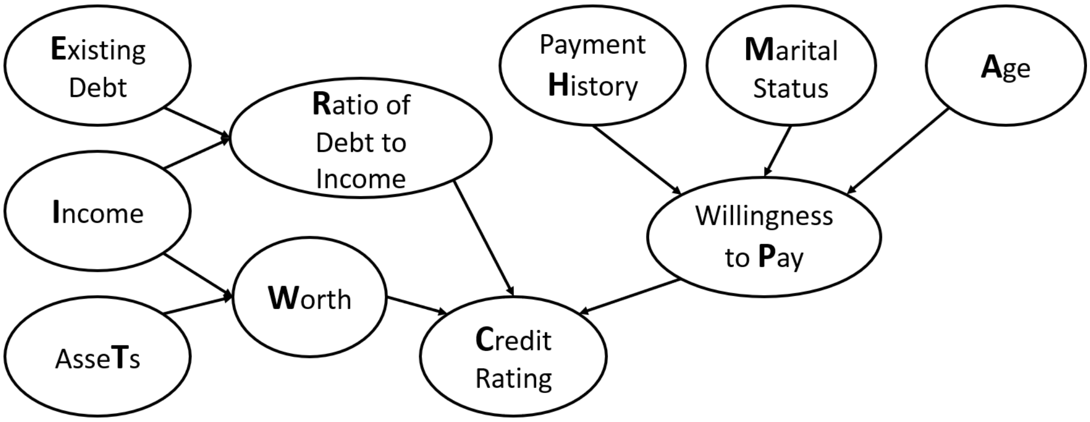
M_1	X	$P(M_2 = + \mid M_1, X)$	$P(M_2 = + \mid X)$
+	+	0.1979	0.1715
+	-	0.752	0.5505
-	+	0.0833	0.1715
-	-	0.4502	0.5505

Clearly, M_1 and M_2 are not conditionally independent given X.

Thus the BN is as follows:



Q2 (20 Points) You are given the following Bayesian Network to decide on a customer's credit rating.



Part a (2 points) Write down the joint probability of this network with the highlighted characters as the corresponding variable names, i.e., $P(E, I, T, R, W, H, C, M, A, P)$.
$P(E, I, T, R, W, H, C, M, A, P)$
$= P(E)P(I)P(T)P(H)P(M)P(A)P(R|E, I)P(W|I, T)P(P|H, M, A)P(C|W, R, P)$

Part b (3 points) There are 10 variables, thus we need to specify 10 probability tables. Assume that all the variables can take k values (e.g. $k = 2$ for a binary variable). How many parameters do we need to fully specify the Bayesian Network i.e., how many total entries should there be in the tables, ignoring that some probabilities need to sum up to 1?
Look at the parents of each node. For a given node with r parents and k values, we have k^{r+1} entries in its CPT. Thus $6k + 2k^3 + 2k^4$.
Probabilities sum up to one, so we actually need lower number of values which is $k^{r+1} - k^r$ for a node with r parents and k values.

Part c (5 points) A customer walks in and asks for a credit. He discloses his Income, Marital Status and Age. You first want to check his credit rating. You want to use variable elimination to get the answer. What are your initial factors?
$f_1(E) = P(E)$, $f_2(I = i) = P(I = i)$, $f_3(T) = P(T)$, $f_4(H) = P(H)$, $f_5(M = m) = P(M = m)$, $f_6(A = a) = P(A = a)$, $f_7(R, E, I = i) = P(R|E, I = i)$, $f_8(W, I = i, T) = P(W|I = i, T)$, $f_9(P, H, M = m, A = a) = P(P|H, M = m, A = a)$, $f_{10}(C, W, R, P) = P(C|W, R, P)$

Part d (2 points) You suddenly get a call from your manager. Somehow he knows the customer and says that he is very willing to pay his credit back and that he has no debts. If you were to calculate his credit rating with the newly acquired information using variable elimination, what would be the **minimum** size of the largest factor, in terms of table length assuming k possible values for all variables, you get during the process?
Given P, the largest factor (in terms of number of non-evidence variables) would be $f(C, R, W, P = p)$ (there are others of the same size). The size of the corresponding would be k^3 (or $k^3 - k^2$).

Part e (8 points) Carry out the variable elimination steps conceptually with the latest information. Make sure to name/enumerate your factors and use he words "join", "sum out" and "normalize" appropriately.

For example you have $P(B|+a), P(C|+a), P(C|D)$ and the aim is to find $P(D|+a)$:
Initial factors: $f_1(+a, B) = P(B|+a)$, $f_2(+a, C) = P(C|+a)$ and $f_3(C, D) = P(C|D)$
Step 1: Join f_2 and f_3 to get $f_4(+a, C, D)$
Step 2: Sum out C from f_4 to get $f_5(+a, D)$
Step 3: Normalize f_5 to get the desired answer

Your answer:
I am only going to use the factors that will be needed. Initial factors:
$f_1(I), f_2(T), f_3(R, I, E = e), f_4(W, I, T), f_5(C, R, W, P = p)$

Step 1: Join f_2 and f_4 to get $f_6(W, I, T)$
Step 2: Sum out T from f_6 to get $f_7(W, I)$
Step 3: Join f_1, f_3 and f_7 to get $f_8(R, W, I)$ Step 4: Sum out I from f_8 to get $f_9(R, W)$
Step 5: Join f_5 and f_9 to get $f_{10}(C, R, W, P = p)$ Step 6: Sum out R from f_{10} to get $f_{11}(C, W, P = p)$
Step 6: Sum out W from f_{11} to get $f_{12}(C, P = p)$
Step 7: Normalize f_{12} to get the desired probability

1 point per step, 1 point for the initial factors

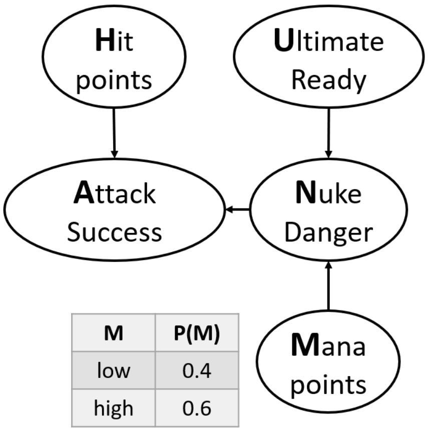
Q3 (30 Points) A friend of yours has introduced you to the a Multiplayer Online Battle Arena (MOBA) game. After your first few games, you realize that you have been struggling to make correct attack decisions. Your friend, who has taken the Comp341 course before, decides to give you some tips:

- When your opponent has high mana points - MP - (i.e. the points necessary to use their attack moves), then he might nuke you (i.e. cause high amount of damage to you).
- When your opponent has his ultimate ready (i.e. an attack move that causes high damage; but has a high mana cost and a high cooldown period), then he might nuke you with his ultimate.
- If your opponent has low hit points - HP - (i.e. wounded), then you should consider attacking.
- If you are not under high nuke danger, then you should consider attacking.

Then you build the following Bayesian Network together where H is your opponents' HP, U is whether his ultimate is ready or not, M is his MP, N is the nuke danger you are under and A is the estimate of your attack success:

H	P(H)
low	0.3
high	0.7

U	P(U)
no	0.5
yes	0.5



H	N	A	P(A|H,N)
low	low	no	0.05
		yes	0.95
	high	no	0.3
		yes	0.7
high	low	no	0.6
		yes	0.4
	high	no	0.9
		yes	0.1

M	P(M)
low	0.4
high	0.6

M	U	N	P(N|M,U)
low	no	low	0.9
		high	0.1
	yes	low	0.7
		high	0.3
high	no	low	0.4
		high	0.6
	yes	low	0.01
		high	0.99

Answer the following questions using this network (Knowledge of MOBA games are not needed to solve the question).

Part a (2 point) Mandatory Joint Distribution Question: What is the expression of $P(A, H, M, N, U)$?

$P(A, H, M, N, U) = P(H)P(U)P(M)P(N|M, U)P(A|H, N)$

Part b (10 points) You are not good at keeping track of the opponent's ultimate availability. This is also reflected with the corresponding distribution. However, you can see his HP and MP. He has low HP but high MP. If you decide to attack, what is your chance of success, i.e., calculate $P(A = yes|H = low, M = high)$ using variable elimination. Calculate the final expression.

Initial factors:

$P(H = low), P(M = high), P(U), P(N|M = high, U), P(A|H = low, N)$

Step 1: Join $P(U)$ and $P(N|M = high, U)$:

$f_1(U, N, M = high) = P(U)P(N|M = high, U)$

U	N	$f_1(U, N, M = high)$
no	low	$0.5 \cdot 0.4 = 0.2$
no	high	$0.5 \cdot 0.6 = 0.3$
yes	low	$0.5 \cdot 0.01 = 0.005$
yes	high	$0.5 \cdot 0.99 = 0.495$

Step 2: Sum out U from f_1 to get $f_2(N, M = high)$

N	$f_2(N, M = high)$
low	$0.2 + 0.005 = 0.205$
high	$0.3 + 0.495 = 0.795$

Step 3: Join $P(A|H = low, N)$ and $f_2(N, M = high)$ to get $f_3(A, N, M = high, H = low)$

A	N	$f_3(A, N, M = high, H = low)$
no	low	$0.205 \cdot 0.05 = 0.01025$
no	high	$0.795 \cdot 0.30 = 0.2385$
yes	low	$0.205 \cdot 0.95 = 0.19475$
yes	high	$0.795 \cdot 0.70 = 0.5565$

Step 4: Sum out N from f_3 to get $f_4(A, M = high, H = low)$

A	$f_4(A, M = high, H = low)$
no	$0.01025 + 0.2385 = 0.24875$
yes	$0.19475 + 0.5565 = 0.75125$

Step 5: Normalize f_4 (no need in this case since, it is already normalized.)

Thus $P(A = yes|H = low, M = high) = 0.75125$

Part c (3 points) Given this probability, $P(A = yes|H = low, M = high)$, would you attack and if so why and if not, why not? Do you need more information to make this decision?
I would attack since the chance of a successful attack is pretty good. However, more information would be useful since the question does not specify what would happen if the attack succeeds, fails or if decide not to attack.

Part d (15 points) Your opponent has not been seen for a while. You assume that he has not used his ultimate yet. However, you are not sure of his HP and MP. He suddenly appears. If he nukes you, you will probably die (not in the BN) but you might also defeat him. If you run, you live but then there is no reward. Do you attack or run? First calculate $P(A = yes|U = yes)$ and $P(N = high|U = yes)$ in any way you want (I suggest variable elimination), and base your answer on these values.
Let's start with $P(A = yes|U = yes)$ Initial factors:
$P(H), P(A|H, N), P(N|M, U = yes), P(M)$

Step 1: Join $P(N|M, U = yes)$ and $P(M)$ to get $f_1(N, M, U = yes)$

M	N	$f_1(N, M, U = yes)$
low	low	$0.7 \cdot 0.4 = 0.28$
low	high	$0.3 \cdot 0.4 = 0.12$
high	low	$0.01 \cdot 0.6 = 0.006$
high	high	$0.99 \cdot 0.6 = 0.594$

Step 2: Sum out M from f_1 to get $f_2(N, U = yes)$

N	$f_2(N, U = yes)$
low	$0.28 + 0.006 = 0.286$
high	$0.12 + 0.594 = 0.714$

From f_2, we can directly get $P(N = high|U = yes) = 0.714$ (notice that we do not need to normalize).
Step 3: Join $P(H)$ and $P(A|H, N)$ to get $f_3(H, N, A)$

H	N	A	$f_3(H, N, A)$
low	low	no	$0.05 \cdot 0.3 = 0.015$
low	low	yes	$0.95 \cdot 0.3 = 0.285$
low	high	no	$0.3 \cdot 0.3 = 0.09$
low	high	yes	$0.7 \cdot 0.3 = 0.21$
high	low	no	$0.6 \cdot 0.7 = 0.42$
high	low	yes	$0.4 \cdot 0.7 = 0.28$
high	high	no	$0.9 \cdot 0.7 = 0.63$
high	high	yes	$0.1 \cdot 0.7 = 0.07$

Step 4: Sum out H from f_3 to get $f_4(N, A)$

N	A	$f_4(N, A)$
low	no	$0.015 + 0.42 = 0.435$
low	yes	$0.285 + 0.28 = 0.565$
high	no	$0.09 + 0.63 = 0.72$
high	yes	$0.21 + 0.07 = 0.28$

Step 5: Join f_2 and f_4 to get $f_5(N, A, U = yes)$

N	A	$f_5(N, A, U = yes)$
low	no	$0.286 \cdot 0.435 = 0.12441$
low	yes	$0.286 \cdot 0.565 = 0.16159$
high	no	$0.714 \cdot 0.72 = 0.51408$
high	yes	$0.714 \cdot 0.28 = 0.19992$

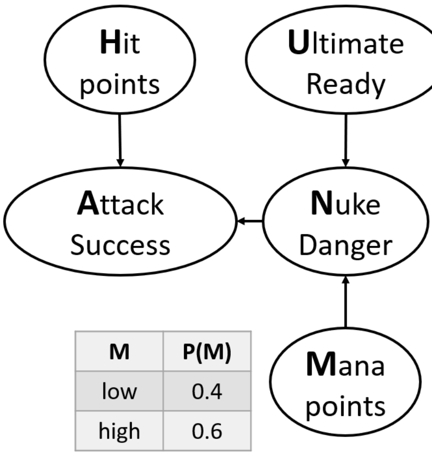
Step 6: For $P(A|U = yes)$, sum out N from f_5 and normalize (ne need in this case)

A	$P(A|U = yes)$
no	$0.12441 + 0.51408 = 0.63849$
yes	$0.16159 + 0.19992 = 0.36151$

We have $P(A = yes|U = yes) = 0.36151$ and $P(N = high|U = yes) = 0.714$. I would not attack since the chance of success is low and the nuke danger is high. It is likely that I would die.

Q4 (15 Points) Consider the same BN (repeated for your convenience). You want to try approximate inference. (Since the prior and rejection sampling questions will be analogous to Q1, we go straight into rejection sampling and Gibbs sampling)

H	P(H)
low	0.3
high	0.7



U	P(U)
no	0.5
yes	0.5

H	N	A	P(A|H,N)
low	low	no	0.05
		yes	0.95
	high	no	0.3
		yes	0.7
high	low	no	0.6
		yes	0.4
	high	no	0.9
		yes	0.1

M	P(M)
low	0.4
high	0.6

M	U	N	P(N|M,U)
low	no	low	0.9
		high	0.1
	yes	low	0.7
		high	0.3
high	no	low	0.4
		high	0.6
	yes	low	0.01
		high	0.99

Part a (10 points) You attacked your opponent and succeeded. He did not use his Ultimate. In the heat of the moment you did not check his HP but only checked his MP which was high. For when he comes back, you are curious about whether he had his Ultimate ready or not, i.e., you are after $P(U = yes|A = yes, M = high)$. You decide to use likelihood weighting and draw the samples below. Fill in the table below with the weight of each sample (not the total weight!). Then, calculate $P(U = yes|A = yes, M = high)$.

A	H	M	N	U	Counts	Weight Per Sample
$A = yes$	$H = high$	$M = high$	$N = high$	$U = yes$	332	$0.6 \cdot 0.1 = 0.06$
$A = yes$	$H = high$	$M = high$	$N = low$	$U = yes$	6	$0.6 \cdot 0.4 = 0.24$
$A = yes$	$H = low$	$M = high$	$N = high$	$U = yes$	155	$0.6 \cdot 0.7 = 0.42$
$A = yes$	$H = low$	$M = high$	$N = low$	$U = yes$	1	$0.6 \cdot 0.95 = 0.57$
$A = yes$	$H = low$	$M = high$	$N = low$	$U = no$	68	$0.6 \cdot 0.95 = 0.57$
$A = yes$	$H = low$	$M = high$	$N = high$	$U = no$	93	$0.6 \cdot 0.7 = 0.42$
$A = yes$	$H = high$	$M = high$	$N = low$	$U = no$	123	$0.6 \cdot 0.4 = 0.24$
$A = yes$	$H = high$	$M = high$	$N = high$	$U = no$	222	$0.6 \cdot 0.1 = 0.06$

You can notice that some of the different sample combinations result in the same weight. Think about why would this be. In fact, we can have a smaller table.

$P(U = yes|A = yes, M = high)$

$$= \frac{(332 \cdot 0.06 + 6 \cdot 0.24 + 155 \cdot 0.42 + 1 \cdot 0.57)}{(332 \cdot 0.06 + 6 \cdot 0.24 + 155 \cdot 0.42 + 1 \cdot 0.57) + (68 \cdot 0.57 + 93 \cdot 0.42 + 123 \cdot 0.24 + 222 \cdot 0.06)}$$

$$= \frac{87.03}{87.03 + 120.66} = \frac{87.03}{207.69} = 0.419$$

Part b (5 points) You want to try Gibbs sampling. You are at a point where you have $A = yes, H = high, M = high, N = high, U = yes$. You want to re-sample N. Calculate $P(N|A, H, M, N)$. Hint: You can make the problem easier by using N's Markov Blanket. Once you get this distribution, you want to sample from it. What would be the sampled value of N in this case if the uniform distribution returns 0.3. Justify your answer.

This is an annoying question, markov blanket for N includes all the variables. However, they still simplify:

$$P(N|A,H,M,N) = \frac{P(A,H,M,N,U)}{P(A,H,M,U)} = \frac{P(H)P(U)P(M)P(N|M,U)P(A|H,N)}{\sum_n P(H)P(U)P(M)P(N=n|M,U)P(A|H,N=n)}$$
$$= \frac{P(H)P(U)P(M)P(N|M,U)P(A|H,N)}{P(H)P(U)P(M)\sum_n P(N=n|M,U)P(A|H,N=n)} = \frac{P(N|M,U)P(A|H,N)}{\sum_n P(N=n|M,U)P(A|H,N=n)}$$

Given the current state, we want $P(N|A = yes, H = high, M = high, U = yes)$.

$$P(N = low|A = yes, H = high, M = high, U = yes)$$
$$= \frac{P(N = low|M = high, U = yes)P(A = yes|H = high, N = low)}{\sum_n P(N = n|M = high, U = yes)P(A = yes|H = high, N = n)}$$
$$= \frac{0.01 \cdot 0.4}{0.01 \cdot 0.4 + 0.99 \cdot 0.1} = \frac{4}{4+99} = 0.03883$$
$$P(N = high|A = yes, H = high, M = high, U = yes) = 1 - \frac{99}{4+99} = 0.96117$$

Note that you did not need to calculate the denominator. You could have just calculated $P(N = high|M = high, U = yes)P(A = yes|H = high, N = high)$ and re-normalized.