COMP341 Introduction to Artificial Intelligence
Assignment 3: Probability and Bayesian Networks

- This homework includes probability and bayesian networks related problems. These topics will be in the second midterm.
- 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 provide clear and concise answers. Gibberish will not receive any credit. Do not overly crowd your answers. Conciseness is a virtue. Write only what is relevant.
- Your answers need to be readable by a human, illegible writing is not gradable hence there is a strong chance that such answers will not get any credit.

Part 1: Probability (30 Points)

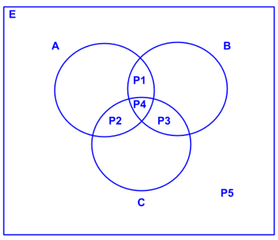
Q1 (10 Points)

Let A, B and C be three events with given probabilities where $A \cup B$ represents events occurring together or by themselves (OR) and $A \cap B$ represents the events occurring together (AND):

- $P(A) = 0.4$
- $P(B) = 0.7$
- $P(C) = 0.3$
- $P(A \cup B) = 0.8$
- $P(B \cap C) = 0.2$
- $P(C \cap (A \cup B)) = 0.2$
- $P(B \cap (A \cup C)) = 0.4$

Find the probabilities that:

1. (5 points) Exactly two of the events among A, B and C occurs.



- $P = P1 + P2 + P3$
- $P1 = P(B \cap (A \cup C)) - P(B \cap C) = 0.4 - 0.2 = 0.2$
- $P2 = P(C \cap (A \cup B)) - P(B \cap C) = 0.2 - 0.2 = 0$
- $P(A \backslash (B \cup C)) = P(A \cup B) - P(B) - P2$ $= 0.8 - 0.7 - 0 = 0.1$
- $P4 = P(A) - P(A \backslash (B \cup C)) - P1 - P2$ $= 0.4 - 0.1 - 0.2 - 0 = 0.1$
- $P3 = P(B \cap C) - P4 = 0.1$
- $P = P1 + P2 + P3 = 0.1 + 0.2 + 0 = 0.3$

2. (5 points) None of the events A, B and C occurs.

- $P5 = P(E) - P(A \cup B \cup C)$
- $P(E) = 1$
- $P(A \cup B \cup C) = P(A \cup B) + P(C) - P(C \cap (A \cup B))$ $= 0.8 + 0.3 - 0.2 = 0.9$
- $P5 = P(E) - P(A \cup B \cup C) = 1 - 0.9 = 0.1$

Q2 (20 Points)

You made 1000 measurements of three Boolean random variables, A, B and C. However, you lost the data labels,but not the counts. As a result, you gave each variable a new name X_1, X_2 and X_3. The table below shows the counts of the samples. You want to figure out the labels again using these counts, i.e. figure out which X_i corresponds to which random variable (among A, B, C). In addition to the counts, you know that:

- A and B are independent
- True value for A, is more common than the true value of B.

The table below shows the counts of the samples. Find the mapping between the original random variables and X_i's.

X_1	X_2	X_3	Counts
T	T	T	32
T	T	F	59
T	F	T	180
T	F	F	6
F	T	T	2
F	T	F	17
F	F	T	82
F	F	F	622

According to the first bulletpoint: $\forall A = a, B = b : P(A = a, B = b) = P(A = a)P(B = b)$

Hence, we need to (*at least*) calculate $P(A = false)$, $P(B = false)$ and $P(A = false, B = false)$ values and check the equality for all of X_1, X_2, X_3.

The prior probabilities are:

$$P(X_1 = false) = \frac{2 + 17 + 82 + 622}{1000} = 0.72$$

$$P(X_2 = false) = \frac{180 + 6 + 82 + 622}{1000} = 0.89$$

$$P(X_3 = false) = \frac{59 + 6 + 17 + 622}{1000} = 0.70$$

The joint probabilities are:

$$P(X_1 = false, X_2 = false) = \frac{82 + 622}{1000} = 0.70$$

$$P(X_1 = false, X_3 = false) = \frac{17 + 622}{1000} = 0.64$$

$$P(X_2 = false, X_3 = false) = \frac{6 + 622}{1000} = 0.63$$

The multiplied priors are:

$$P(X_1 = false)P(X_2 = false) = 0.64$$

$$P(X_1 = false)P(X_3 = false) = 0.51$$

$$P(X_2 = false)P(X_3 = false) = 0.63$$

The equality is only satisfied for X_2 and X_3, hence one of them should be A and the other one should be B.

Using the second bulletpoint and the prior values, we can conclude that: X_1 is C, X_2 is B and, X_3 is A.

Part 2: BN Representation and Exact Inference (40 Points)

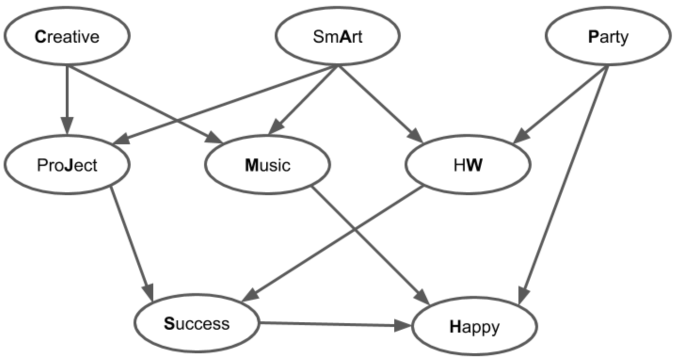
As part of a study regarding the role of COMP 341 on student happiness, we collected important data from *fictional* students that were previously enrolled in this class. In an entirely optional *fictional* survey that all *fictional* students were required to complete, we asked the following highly objective questions:

- Do you party frequently? [**Party**: Yes/No]
- Are you wicked smart? [**Smart**: Yes/No]
- Are you creative? [**Creative**: Yes/No]
- Did you do well on all your written homework assignments? [**HW**: Yes/No]
- Do you play any kind of musical instrument? [**Music**: Yes/No]
- Did you succeed in your Pacman Projects? [**Project**: Yes/No]
- Did you succeed in your most important class (which is COMP 341)? [**Success**: Yes/No]
- Are you currently Happy? [**Happy**: Yes/No]

Based on the *fictional* survey results and after consulting a *fictional* behavioral psychologist we obtained the following complete set of *fictional* conditional relationships:

- **HW** depends only on **Party** and **Smart**
- **Music** depends only on **Smart** and **Creative**
- **Project** depends only on **Smart** and **Creative**
- **Success** depends only on **HW** and **Project**
- **Happy** depends only on **Party**, **Music**, and **Success**

Hence, we acquired the following *fictional* Bayesian Network(BN):



Q3 (12 Points)

Using the BN above, answer the following true/false questions.

1. **Success** is independent of **Party** given **HW**. False.
   There is an active path: $party \rightarrow (hw) \leftarrow smart \rightarrow project \rightarrow success$
   because of the v-structure at hw is observed.

2. **Smart** is independent of **Party** given **Success**. False.
   There is an active path: $smart \rightarrow project \rightarrow (success) \leftarrow hw \leftarrow party$
   because of the v-structure at hw activated by success.

3. **Creative** is independent of **Party** given **Happy**. False.
   There is an active path: $party \rightarrow (happy) \leftarrow music \leftarrow creative$
   because of the v-structure at happy is observed.

4. **Happy** is independent of **Smart** given **Music**. False.
   There is an active path: $smart \rightarrow project \rightarrow success \rightarrow happy$

5. **Success** is independent of **Smart** given **Project** and **HW**. True.
   There are no active paths.

6. **Success** is independent of **Creative** given **Project**. False.
   There is an active path: $creative \rightarrow (project) \leftarrow smart \rightarrow hw \rightarrow success$

Q4 (28 Points)

Using the Bayesian Network given above, answer the following questions. From now on, the node names will be shortened as follows: **C**reative, Sm**A**rt, **P**arty, Pro**J**ect, **M**usic, Home**W**ork, **S**uccess and **H**appy.

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

Part b (2 points) 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?
For a given node with r parents and n values, we have $n^{(r+1)}$ entries in its CPT. As we have given that all variables are binary in the question, $n = 2$, and the answer is:

$3n + 4n^{(2+1)} + n^{(3+1)} = 3 * 2 + 4 * 2^3 + 2^4 = 6 + 32 + 16 = 54$

Part c (4 points) An alumni visits one of the course TAs for a small chat. The TA knows that the alumni is smart and creative, however the alumni does not want to provide the details of his night life (i.e. if he is partying or not). The TA wants to understand if he is currently happy or not. He uses variable elimination after the alumni leaves. What are his initial factors?

- $f_1(C = c) = P(C = c)$,
- $f_2(A = a) = P(A = a)$,
- $f_3(P) = P(P)$,
- $f_4(J, C = c, A = a) = P(J|C = c, A = a)$,
- $f_5(M, C = c, A = a) = P(M|C = c, A = a)$,
- $f_6(W, A = a, P) = P(W|A = a, P)$,
- $f_7(S, J, W) = P(S|J, W)$,
- $f_8(H, S, M, P) = P(H|S, M, P)$

Part d (4 points) TA suddenly receives a call from the professor. Apparently he remembers the alumni and says that he was very successful in his projects and homeworks. If the TA was to calculate his level of happiness with the newly acquired information using variable elimination, what would be the **minimum** size of the largest factor in terms of table length?
Given J and P the largest factor (in terms of number of non-evidence variables) would be $f_8(H, S, M, P) = P(H|S, M, P)$. The size of the corresponding would be $2^4 = 16$ or $2^4 - 2^3 = 8$

Part e (16 points) Carry out the variable elimination steps conceptually with starting from the situation of part d. Make sure to name/enumerate your factors and use the 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: $f1(+a, B) = P(B|+a)$, $f2(+a, C) = P(C|+a)$ and $f3(C, D) = P(C|D)$
Step 1: Join $f2$ and $f3$ to get $f4(+a, C, D)$
Step 2: Sum out C from $f4$ to get $f5(+a, D)$
Step 3: Normalize $f5$ to get the desired answer

Your answer:

As we know $C = c$, $A = a$, $J = j$ and $W = w$, the factors we need are:

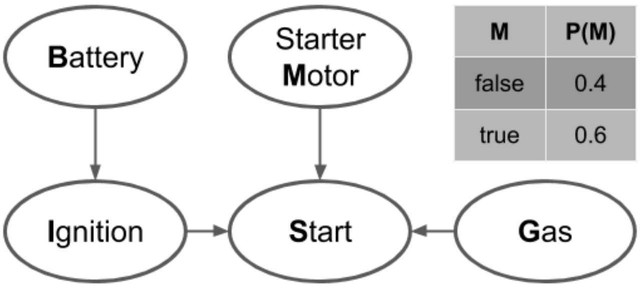
- $f_1(P) = P(P)$,
- $f_2(M, C = c, A = a) = P(M|C = c, A = a)$,
- $f_3(S, J = j, W = w) = P(S|J = j, W = w)$,
- $f_4(H, S, M, P) = P(H|S, M, P)$

Step 1: Join f_3 and f_4 to get $f_5(H, S, M, P, J = j, W = w)$
Step 2: Sum out S from f_5 to get $f_6(H, M, P, J = j, W = w)$
Step 3: Join f_2 and f_6 to get $f_7(H, M, P, J = j, W = w, C = c, A = a)$
Step 4: Sum out M from f_7 to get $f_8(H, P, J = j, W = w, C = c, A = a)$
Step 5: Join f_1 and f_8 to get $f_9(H, P, J = j, W = w, C = c, A = a)$
Step 6: Sum out P from f_9 to get $f_{10}(H, J = j, W = w, C = c, A = a)$
Step 7: Normalize f_{10} to get the desired probability.

Part 3: BN Approximate Inference (30 Points)

Q5 (20 Points)

You are trying to fix a car that is not starting and to follow the basic principles behind how a car operates, you come up with the following bayesian network:



B	P(B)
empty	0.3
full	0.7

M	P(M)
false	0.4
true	0.6

B	I	P(I|G)
empty	false	0.95
	true	0.05
full	false	0.15
	true	0.85

G	P(G)
empty	0.8
full	0.2

I	M	G	S	P(S|I,M,G)
false	false	empty	true	0.01
		full	true	0.05
	true	empty	true	0.05
		full	true	0.1
true	false	empty	true	0.05
		full	true	0.1
	true	empty	true	0.2
		full	true	0.9

Part a (2 points) Mandatory Joint Distribution Question: What is the expression of $P(B, M, G, I, S)$?

$P(B, M, G, I, S) = P(B)P(M)P(G)P(I|B)P(S|I, M, G)$

Part b (4 points) You tried to start the engine and it worked. You know that the car's battery is full and the starter motor is functional. Since your gas gauge is not accurate, you are not sure whether your car has enough fuel to start the next time. For some reason, you have recorded your previous observations given in the appendix of this document. Calculate $P(S = true)$ according to this list.
We need to utilize *prior sampling*. There are 14 times $S = true$ occurs among 40 samples. Hence:
$P(S = true) = \frac{14}{39} = 0.36$

Part c (4 points) You realized that you also need to include evidence in your calculations to have a more accurate estimate of the probability. Calculate $P(S = true|B = full, M = true)$ according to the list provided in the appendix of this document.
We need to use *rejection sampling* to utilize the evidence. There are 6 times $S = true$ occurs among 13 samples where $B = full$ and $M = true$. Hence:
$P(S = true|B = full, M = true) = \frac{6}{13} = 0.46$

Part d (10 points) You are still not satisfied with your calculations and want to have a more accurate idea. You decide to use likelihood weighting. Fill in the table below with the weight of each sample (not the total weight!). Then, calculate $P(S = true|B = full, M = true)$.

B	M	I	G	S	Counts	Weights Per Sample
$B = full$	$M = true$	$I = true$	$G = full$	$S = true$	66	0.7*0.6=0.42
$B = full$	$M = true$	$I = true$	$G = empty$	$S = true$	1	0.7*0.6=0.42
$B = full$	$M = true$	$I = false$	$G = full$	$S = true$	31	0.7*0.6=0.42
$B = full$	$M = true$	$I = false$	$G = empty$	$S = true$	0	0.7*0.6=0.42
$B = full$	$M = true$	$I = false$	$G = full$	$S = false$	14	0.7*0.6=0.42
$B = full$	$M = true$	$I = false$	$G = empty$	$S = false$	19	0.7*0.6=0.42
$B = full$	$M = true$	$I = true$	$G = full$	$S = false$	25	0.7*0.6=0.42
$B = full$	$M = true$	$I = true$	$G = empty$	$S = false$	44	0.7*0.6=0.42

All the weights are the same as the evidence variables are not dependent on anything else than themselves. Hence, we can get rid of the weights and calculate only by samples:

$$P(S = true|B = full, M = true) = \frac{66 + 1 + 31 + 0}{66 + 1 + 31 + 0 + 14 + 19 + 25 + 44} = \frac{98}{200} = 0.49$$

Q6 (10 Points)

Answer the below questions based on the BN from the question 5 for Gibbs sampling.

Part a (2 points) Given $\{B = full, M = true, I = true, G = empty\}$, calculate the probability distribution which will be used to sample S.
This is a trivial question, we can directly use the CPT of S. (Due to the BN topology, S is conditionally independent of B given M, I, G):
$P(S = true|I = true, M = true, G = empty) = 0.2$
$P(S = false|I = true, M = true, G = empty) = 0.8$

Part b (6 points) Given $\{B = full, I = true, G = empty, S = false\}$, calculate the probability distribution which will be used to sample M.

$$P(M|B, I, G, S) = \frac{P(M, B, I, G, S)}{P(B, I, G, S)} = \frac{P(B)P(M)P(G)P(I|B)P(S|I, M, G)}{\sum_m P(B)P(M = m)P(G)P(I|B)P(S|I, M = m, G)}$$

$$P(M|B, M, I, S) = \frac{P(M)P(S|I, M, G)}{\sum_m P(M = m)P(S|I, M = m, G)}$$

$$P(M = true|B = full, I = true, G = empty, S = false)$$
$$= \frac{P(M = true)P(S = false|I = true, M = true, G = empty)}{\sum_m P(M = m)P(S = false|I = true, M = m, G = empty)}$$
$$= \frac{0.6 * 0.8}{0.6 * 0.8 + 0.4 * 0.95} = 0.558$$

$P(M = false|B = full, I = true, G = empty, S = false) = 1 - P(M = true|B = full, I = true, G = empty, S = false) = 0.442$

Part c (2 points) The next uniform sample you get in the range $[0, 1)$ is 0.4. Based on your answer to part b, what would be the sample you chose for M?
$0 \leq u < 0.558 \rightarrow M = true$ and $0.558 \leq u < 1 \rightarrow M = false$; then when $u = 0.4 \rightarrow M = true$

Submission

- You are going to submit a **SINGLE PDF FILE** with your solutions on the blackboard site OR give your written homework to the TA, Onur Demiray. DO NOT DO BOTH!
- You can use scanners, apps similar to CamScanner, your cameras or any other possible way to convert your homework submissions to a digital format.
- Do not send your solutions in a picture format (e.g. JPG, PNG, BMP, TIFF, GIF, etc.). They PDF must have A4 paper size.
- Any submission other than **PDF format** is going to receive a very large grade deduction.
- While you are submitting, please make sure your scanned assignment is birght (i.e. that you are not submitting a dark page), this is important because we should be able to see what you wrote to grade your submission.

Appendix: Samples for Q5

$B = full$	$M = false$	$I = false$	$S = false$	$G = full$
$B = full$	$M = false$	$I = true$	$S = false$	$G = full$
$B = empty$	$M = true$	$I = false$	$S = false$	$G = full$
$B = empty$	$M = true$	$I = true$	$S = false$	$G = empty$
$B = full$	$M = true$	$I = true$	$S = true$	$G = full$
$B = empty$	$M = true$	$I = true$	$S = true$	$G = empty$
$B = empty$	$M = true$	$I = false$	$S = false$	$G = empty$
$B = full$	$M = true$	$I = true$	$S = true$	$G = empty$
$B = empty$	$M = true$	$I = false$	$S = true$	$G = full$
$B = empty$	$M = true$	$I = true$	$S = false$	$G = full$
$B = full$	$M = false$	$I = true$	$S = true$	$G = empty$
$B = full$	$M = true$	$I = true$	$S = false$	$G = empty$
$B = empty$	$M = false$	$I = true$	$S = false$	$G = empty$
$B = empty$	$M = false$	$I = true$	$S = false$	$G = empty$
$B = full$	$M = false$	$I = true$	$S = false$	$G = empty$
$B = empty$	$M = true$	$I = false$	$S = false$	$G = full$
$B = empty$	$M = true$	$I = true$	$S = false$	$G = empty$
$B = full$	$M = true$	$I = true$	$S = true$	$G = full$
$B = full$	$M = true$	$I = false$	$S = true$	$G = full$
$B = empty$	$M = false$	$I = true$	$S = true$	$G = empty$
$B = empty$	$M = true$	$I = false$	$S = false$	$G = empty$
$B = empty$	$M = false$	$I = false$	$S = false$	$G = full$
$B = full$	$M = false$	$I = true$	$S = false$	$G = empty$
$B = full$	$M = true$	$I = false$	$S = false$	$G = full$
$B = full$	$M = true$	$I = true$	$S = false$	$G = empty$
$B = full$	$M = true$	$I = false$	$S = true$	$G = full$
$B = full$	$M = true$	$I = false$	$S = false$	$G = empty$
$B = full$	$M = true$	$I = false$	$S = false$	$G = empty$
$B = full$	$M = true$	$I = true$	$S = false$	$G = full$
$B = empty$	$M = false$	$I = true$	$S = true$	$G = full$
$B = empty$	$M = true$	$I = true$	$S = false$	$G = empty$
$B = empty$	$M = false$	$I = false$	$S = true$	$G = full$
$B = full$	$M = true$	$I = true$	$S = false$	$G = full$
$B = empty$	$M = true$	$I = true$	$S = true$	$G = full$
$B = empty$	$M = false$	$I = false$	$S = false$	$G = full$
$B = empty$	$M = true$	$I = true$	$S = false$	$G = empty$
$B = empty$	$M = true$	$I = false$	$S = true$	$G = full$
$B = full$	$M = true$	$I = false$	$S = true$	$G = full$
$B = full$	$M = false$	$I = true$	$S = false$	$G = empty$