KOÇ UNIVERSITY
College of Engineering

COMP 341: INTRODUCTION TO ARTIFICIAL INTELLIGENCE

Midterm 2

Fall 2023, 28/12/2023
Duration: 105 Minutes

Name: Solutions

ID: 00110001 00110000 00110000

- This exam contains 13 pages including this cover page and 7 questions. Check to see if any pages are missing. Put your initials on the top of every page, in case the pages become separated.
- By submitting this exam, you **agree** to fully comply with Koç University Student Code of Conduct, and accept any punishment in case of failure to comply. If you do not submit this exam on time, you will receive 0 credits.
- The exam is **closed book** and **closed notes**. You are **not** allowed to use any additional material including any electronic equipment such as computers and mobile phones.
- You are expected to be able provide clear and concise answers. Gibberish will not receive any credit. Your answers need to be readable by a human, illegible writing is not gradable.
- Read each question carefully and make sure to follow instructions. The written instructions take precedence over an answer that the instructor might give you during the exam, unless the instructor makes a class wide announcement.
- Do not write in the table below.

Question:	1	2	3	4	5	6	7	Total
Points:	16	10	8	15	14	15	22	100
Score:								

1. (16 points) True or False :

   False — Bayes Theorem is used to calculate a conditional probability distribution from a joint probability distribution.

   True — If X and Y are conditionally independent given Z, then this always holds: $P(X,Y|Z) = P(X|Z)P(Y|Z)$.

   True — Bayesian network topology has an effect on variable elimination computational complexity.

   False — In Gibbs sampling, we discard samples that do not agree with evidence.

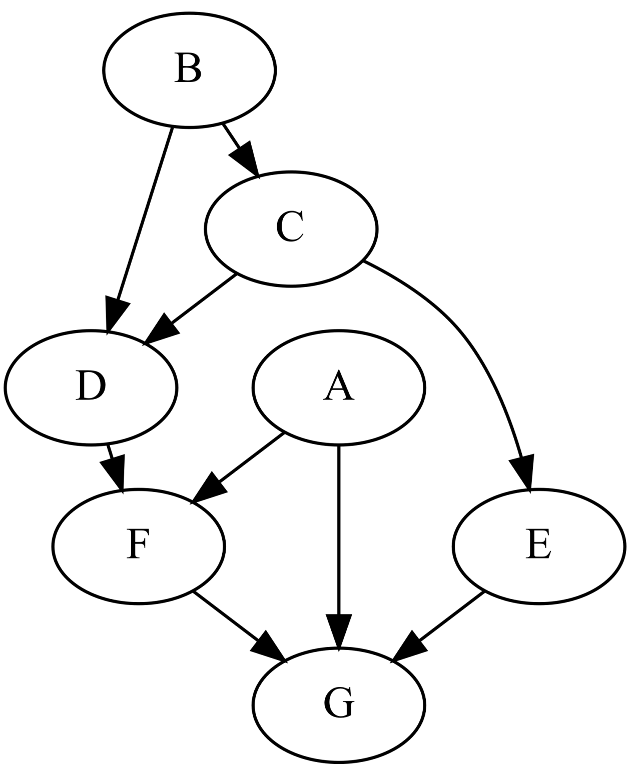
   False — In decision networks, the utilities must sum up to zero for a given utility node.

   True — In decision networks, value of information cannot be negative.

   True — For state inference over time in Markovian models, it is enough to store a constant amount of information, independent of the sequence length.

   False — Given a Hidden Markov Model and a set of sequential emissions, we can calculate the most likely state path by running the forward algorithm and taking the argmax state at each time step.

2. (10 points) Answer the conditional independence questions as True or False based on the given Bayesian Network.



   True — A and C are independent.

   False — A and C are independent given G.

   False — B and G are independent given C.

   True — B and E are independent given C.

   False — B and E are independent given C and G.

3. (8 points) You made 1000 measurements of three Boolean random variables, X_1, X_2 and X_3. The table below shows the counts of the samples. Use this table to answer the questions. You can leave the answers as fractions

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

(a) (2 points) Calculate $P(X_1 = T, X_2 = F, X_3 = T)$

$$P(X_1 = T, X_2 = F, X_3 = T) = \frac{180}{1000} = 0.18$$

(b) (2 points) Calculate $P(X_1 = T, X_3 = T)$

$$P(X_1 = T, X_3 = T) = \frac{180 + 32}{1000} = 0.212$$

(c) (2 points) Calculate $P(X_3 = T)$

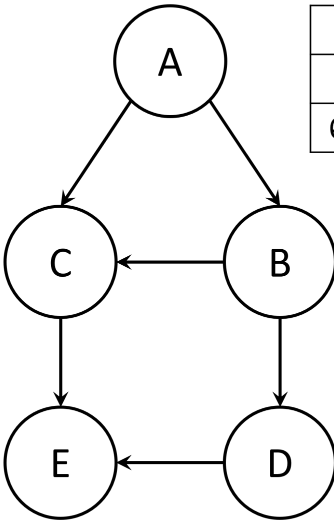
$$P(X_3 = T) = \frac{180 + 32 + 2 + 82}{1000} = 0.296$$

(d) (2 points) Calculate $P(X_2 = F|X_1 = T, X_3 = T)$

$$P(X_2 = F|X_1 = T, X_3 = T) = \frac{P(X_1 = T, X_2 = F, X_3 = T)}{X_1 = T, X_3 = T)} = \frac{0.18}{0.212} = \frac{45}{53} \approx 0.849$$

The count version also works.

4. (15 points) Answer the following questions given the Bayesian Network below.



P(A)	
+a	-a
0.4	0.6

A	P(B|A) +b	-b
+a	0.2	0.8
-a	0.4	0.6

B	P(D|B) +d	-d
+b	0.1	0.9
-b	0.7	0.3

A	B	P(C|A,B) +c	-c
+a	+b	0.4	0.6
+a	-b	0.7	0.3
-a	+b	0.8	0.2
-a	-b	0.9	0.1

C	D	P(E|C,D) +e	-e
+c	+d	0.2	0.8
+c	-d	0.7	0.3
-c	+d	0.6	0.4
-c	-d	0.1	0.9

(a) (2 points) What is the joint distribution, $P(A, B, C, D, E)$?
$P(A, B, C, D, E) = P(A)P(B|A)P(C|A, B)P(D|B)P(E|C, D)$

(b) (13 points) Calculate $P(A| + b, -e)$ using exact inference. You can leave all the numbers as fractions. If you are out of time, write your variable elimination plan for partial points.
We are going to provide the variable elimination answer. Initial factors:

- $f_1(A) = P(A)$: 2 entries
- $f_2(A, +b) = P(+b|A)$: 2 entries
- $f_3(A, +b, C) = P(C|A, +b)$: 4 entries
- $f_4(+b, D) = P(D| + b)$: 2 entries
- $f_5(C, D, -e) = P(-e|C, D)$: 4 entries

Hidden variables are C and D. We need to join their factors and sum them out. To remove C first, we need to join f_3 and f_5 which will give us 8 entries. To remove D first, we need to join f_4 and f_5 which will give us 4 entries. Thus, we start by eliminating D.

Step 1: Join $f_4(+b, D)$ and $f_5(C, D, -e)$ to get $f_6(+b, C, D, -e)$

B	E	C	D	$f_6(+b, C, D, -e)$
$+b$	$-e$	$+c$	$+d$	$0.1 \cdot 0.8 = 0.08(8/100)$
$+b$	$-e$	$+c$	$-d$	$0.9 \cdot 0.3 = 0.27(27/100)$
$+b$	$-e$	$-c$	$+d$	$0.1 \cdot 0.4 = 0.04(4/100)$
$+b$	$-e$	$-c$	$-d$	$0.9 \cdot 0.9 = 0.81(81/100)$

Step 2: Marginalize out D from f_6 to get $f_7(+b, C, -e)$

B	E	C	$f_7(+b, C, -e)$
$+b$	$-e$	$+c$	$0.08 + 0.27 = 0.35(7/20)$
$+b$	$-e$	$-c$	$0.81 + 0.04 = 0.85(17/20)$

Step 3: Join $f_3(A,+b,C)$ and $f_7(+b,C,-e)$ to get $f_8(A,+b,C,-e)$

B	E	A	C	$f_8(A,+b,C,-e)$
$+b$	$-e$	$+a$	$+c$	$0.4 \cdot 0.35 = 0.14(14/100)$
$+b$	$-e$	$+a$	$-c$	$0.6 \cdot 0.85 = 0.51(51/100)$
$+b$	$-e$	$-a$	$+c$	$0.8 \cdot 0.35 = 0.28(28/100)$
$+b$	$-e$	$-a$	$-c$	$0.2 \cdot 0.85 = 0.17(17/100)$

Step 4: Marginalize out C from f_8 to get $f_9(A,+b,-e)$

B	E	A	$f_9(A,+b,-e)$
$+b$	$-e$	$+a$	$0.14 + 0.51 = 0.65(13/20)$
$+b$	$-e$	$-a$	$0.28 + 0.17 = 0.45(9/20)$

Step 5: Join $f_1(A)$, $f_2(A,+b)$ and $f_9(A,+b,-e)$ to get $f_{10}(A,+b,-e)$

B	E	A	$f_8(A,+b,C,-e)$
$+b$	$-e$	$+a$	$0.4 \cdot 0.2 \cdot 0.65 = 0.052(13/250)$
$+b$	$-e$	$-a$	$0.6 \cdot 0.4 \cdot 0.45 = 0.108(27/250)$

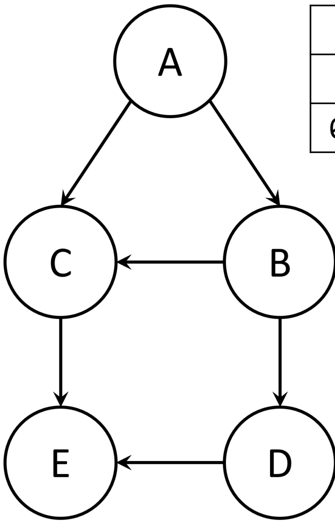
Step 6: Normalize $f_{10}(A,+b,-e)$ to get the desired probability $P(A|+b,-e)$

B	E	A	$P(A|+b,-e)$
$+b$	$-e$	$+a$	$0.052/(0.052 + 0.084) = 13/40 = 0.325$
$+b$	$-e$	$-a$	$0.084/(0.052 + 0.084) = 27/40 = 0.675$

Grading: calculation errors can be ignored as long as we can see individual math steps. If unclear, just 0 points. Overall 2 points per step, breakdown is below. If there are intermediate steps (e.g. first join f_1 and f_2, then join the resulting with f_9), we look at the entire thing (e.g. 2 points for 2 steps)

- 2 hidden variable removal (1 join and 1 sum out): 4 + 4 points
- Joining remaining 3 for A: 2 points
- Correct normalization: 2 points
- Overall legibility and clarity 1 point

5. (14 points) Answer the following questions given the same Bayesian Network as before.



P(A)	
+a	-a
0.4	0.6

A	P(B|A) +b	-b
+a	0.2	0.8
-a	0.4	0.6

B	P(D|B) +d	-d
+b	0.1	0.9
-b	0.7	0.3

		P(C|A,B)	
A	B	+c	-c
+a	+b	0.4	0.6
+a	-b	0.7	0.3
-a	+b	0.8	0.2
-a	-b	0.9	0.1

		P(E|C,D)	
C	D	+e	-e
+c	+d	0.2	0.8
+c	-d	0.7	0.3
-c	+d	0.6	0.4
-c	-d	0.1	0.9

(a) (2 points) Suppose you want to calculate $P(A|+b,-e)$ using rejection sampling. Give an example of a sample that would be rejected and another example that would not.

Rejected examples include anything that does not have $(+b,-e)$, e.g. $(+a,-b,-c,+d,-e)$
Not rejected examples include everything with $(+b,-e)$, e.g. $(+a,+b,-c,+d,-e)$
1 point per correct example

(b) (4 points) You have run rejection sampling 1000 times (total samples, before rejection) given the evidence $(+b,-e)$. You have a total of 80 samples that include $+a$. Is this information enough to calculate the probability $P(A|+b,-e)$? If so, what is the value? If not, what information do we need?

It is not enough since we do not know how many samples left after rejection. Thus we need the counts of surviving samples. 1 point for saying no, 3 points for explanation.

(c) (4 points) Suppose you want to calculate $P(A|+b,-e)$ using likelihood weighting. Write down all the possible combinations of samples and corresponding weights.

For this problem, all the non-evidence variables (A, C, D) affect the probabilities calculated in weights. Since they are binary, we have $23 = 8$ unique cases as follows (0.5 points per item):

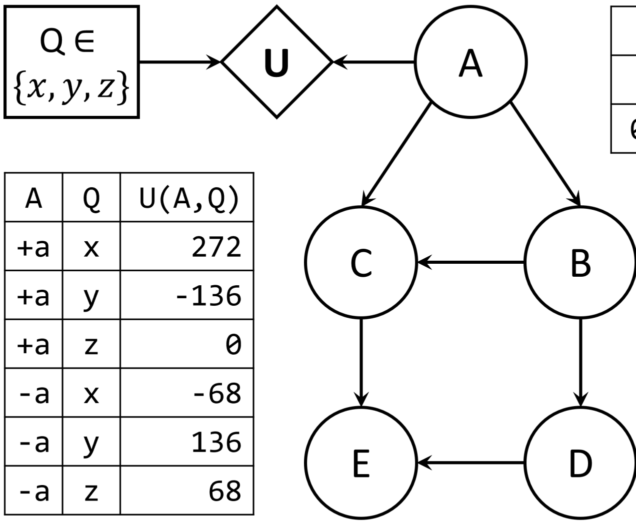
- Sample $(+a, +b, +c, +d, -e)$ with weight $P(+b|+a)P(-e|+c,+d) = 0.2 \cdot 0.8 = 0.16$
- Sample $(+a, +b, +c, -d, -e)$ with weight $P(+b|+a)P(-e|+c,-d) = 0.2 \cdot 0.3 = 0.06$
- Sample $(+a, +b, -c, +d, -e)$ with weight $P(+b|+a)P(-e|-c,+d) = 0.2 \cdot 0.4 = 0.08$
- Sample $(+a, +b, -c, -d, -e)$ with weight $P(+b|+a)P(-e|-c,-d) = 0.2 \cdot 0.9 = 0.18$
- Sample $(-a, +b, +c, +d, -e)$ with weight $P(+b|-a)P(-e|+c,+d) = 0.4 \cdot 0.8 = 0.32$
- Sample $(-a, +b, +c, -d, -e)$ with weight $P(+b|-a)P(-e|+c,-d) = 0.4 \cdot 0.3 = 0.12$
- Sample $(-a, +b, -c, +d, -e)$ with weight $P(+b|-a)P(-e|-c,+d) = 0.4 \cdot 0.4 = 0.16$
- Sample $(-a, +b, -c, -d, -e)$ with weight $P(+b|-a)P(-e|-c,-d) = 0.4 \cdot 0.9 = 0.36$

Wrong samples (e.g. involving -b,+e) get -0.25 points. Multiplying together all possible probabilities (i.e. joint likelihood) gets 0 points

(d) (4 points) You have run sampling for likelihood weighting 1000 times given the evidence $(+b, -e)$. You have a total of 276 samples that include $+a$. Is this information enough to calculate the probability $P(A|+b,-e)$? If so, what is the value? If not, what information do we need?

It is not enough since the sampling procedure for likelihood weighting not consistent (see the definition in the slides). That is why we calculate the weights. Thus we need the counts of the individual samples and not just the count with $+a$. 1 point for saying no, 3 points for explanation.

6. (15 points) Answer the following questions given the Decision Network (DN) below. This DN is an augmented version of the previous Bayesian Networks. (Hint: $4 \cdot 17 = 68$)



A	Q	U(A,Q)
+a	x	272
+a	y	-136
+a	z	0
-a	x	-68
-a	y	136
-a	z	68

P(A)	
+a	-a
0.4	0.6

	P(B|A)	
A	+b	-b
+a	0.2	0.8
-a	0.4	0.6

	P(D|B)	
B	+d	-d
+b	0.1	0.9
-b	0.7	0.3

		P(C|A,B)	
A	B	+c	-c
+a	+b	0.4	0.6
+a	-b	0.7	0.3
-a	+b	0.8	0.2
-a	-b	0.9	0.1

		P(E|C,D)	
C	D	+e	-e
+c	+d	0.2	0.8
+c	-d	0.7	0.3
-c	+d	0.6	0.4
-c	-d	0.1	0.9

You are additionally given several probabilities below. You may directly use them in your answers. However, some of them may not be needed and you may need to calculate more! So be careful

- $P(A=+a, B=+b) = 0.08, P(A=+a, B=-b) = 0.32,$ $P(A=-a, B=+b) = 0.24, P(A=-a, B=-b) = 0.36$
- $P(B=+b|C=+c) = 0.29, P(B=-b|C=+c) = 0.71,$ $P(B=+b|C=-c) = 0.42, P(B=-b|C=-c) = 0.58$
- $P(C=+c) = 0.77, P(C=-c) = 0.23$

(a) (4 points) Calculate the maximum expected utility (MEU) and the optimal action. You can leave the numbers as fractions and you do not have to calculate the final expressions.

$$EU(Q=x) = \sum_{a \in \{+a,-a\}} P(A=a)U(A=a,Q=x) = 0.4 \cdot 272 + 0.6 \cdot (-68) = 68$$

$$EU(Q=y) = \sum_{a \in \{+a,-a\}} P(A=a)U(A=a,Q=x) = 0.4 \cdot (-136) + 0.6 \cdot 136 = 27.2$$

$$EU(Q=z) = \sum_{a \in \{+a,-a\}} P(A=a)U(A=a,Q=x) = 0.4 \cdot 0 + 0.6 \cdot 68 = 40.8$$

$MEU() = 68$ with the optimal action $Q = x$.

1.5 points for correct equation, 0.5 point per individual value (1.5 total), 0.5 point for highlighting the MEU and 0.5 points for highlighting the optimal action.

(b) (11 points) Calculate the value of observing B. (i.e. $VPI(B)$). You can leave the numbers as fractions and you do not have to calculate the final expressions.

Inference: We are given $P(A, B)$ but we need $P(B)$ and $P(A|B)$. We do not need the rest. Calculating $P(B)$ is straightforward. Sum out A to get $P(B = +b) = 0.08 + 0.24 = 0.32$ and $P(B = -b) = 0.32 + 0.36 = 0.68$. For $P(A|B)$, we can either normalize $P(A, B)$ over A or use $P(A|B) = P(A, B)/P(B)$. The latter:

A	B	$P(A\|B)$
$+a$	$+b$	$0.08/0.32 = 1/4$
$+a$	$-b$	$0.32/0.68 = 8/17$
$-a$	$+b$	$0.24/0.32 = 3/4$
$-a$	$-b$	$0.36/0.68 = 9/17$

$P(B)$ 1 point, $P(A|B)$, 2 points. Any other choice would lead to 0. Total: 3 points. If the approach is correct but the math is wrong, deduct at most 1 point.

EUs for each action and MEU for values of B: Then, we need to calculate MEU for each possible value of B and each possible action:

$$EU(Q = x, B = +b) = \sum_{a\in\{+a,-a\}} P(A = a|B = +b)U(A = a, Q = x) = 1/4 \cdot 272 + 3/4 \cdot (-68) = 17$$
$$EU(Q = y, B = +b) = \sum_{a\in\{+a,-a\}} P(A = a|B = +b)U(A = a, Q = x) = 1/4 \cdot (-136) + 3/4 \cdot 136 = 68$$
$$EU(Q = z, B = +b) = \sum_{a\in\{+a,-a\}} P(A = a|B = +b)U(A = a, Q = x) = 1/4 \cdot 0 + 3/4 \cdot 68 = 51$$
$$EU(Q = x, B = -b) = \sum_{a\in\{+a,-a\}} P(A = a|B = -b)U(A = a, Q = x) = 8/17 \cdot 272 + 9/17 \cdot (-68) = 92$$
$$EU(Q = y, B = -b) = \sum_{a\in\{+a,-a\}} P(A = a|B = -b)U(A = a, Q = x) = 8/17 \cdot (-136) + 9/17 \cdot 136 = 8$$
$$EU(Q = z, B = -b) = \sum_{a\in\{+a,-a\}} P(A = a|B = -b)U(A = a, Q = x) = 8/17 \cdot 0 + 9/17 \cdot 68 = 36$$

We can see that $MEU(B = +b) = 68$ with $Q = y$ and $MEU(B = -b) = 92$ with $Q = x$
Correct equation (The $P(A|B)$ part) 1 point, doing it for each value of B 1 point, 0.25 points per correct value and 1 for picking the correct MEUs (per B value). Total 1+1 $6 \cdot 0.25$ + 1 = 4.5 points).

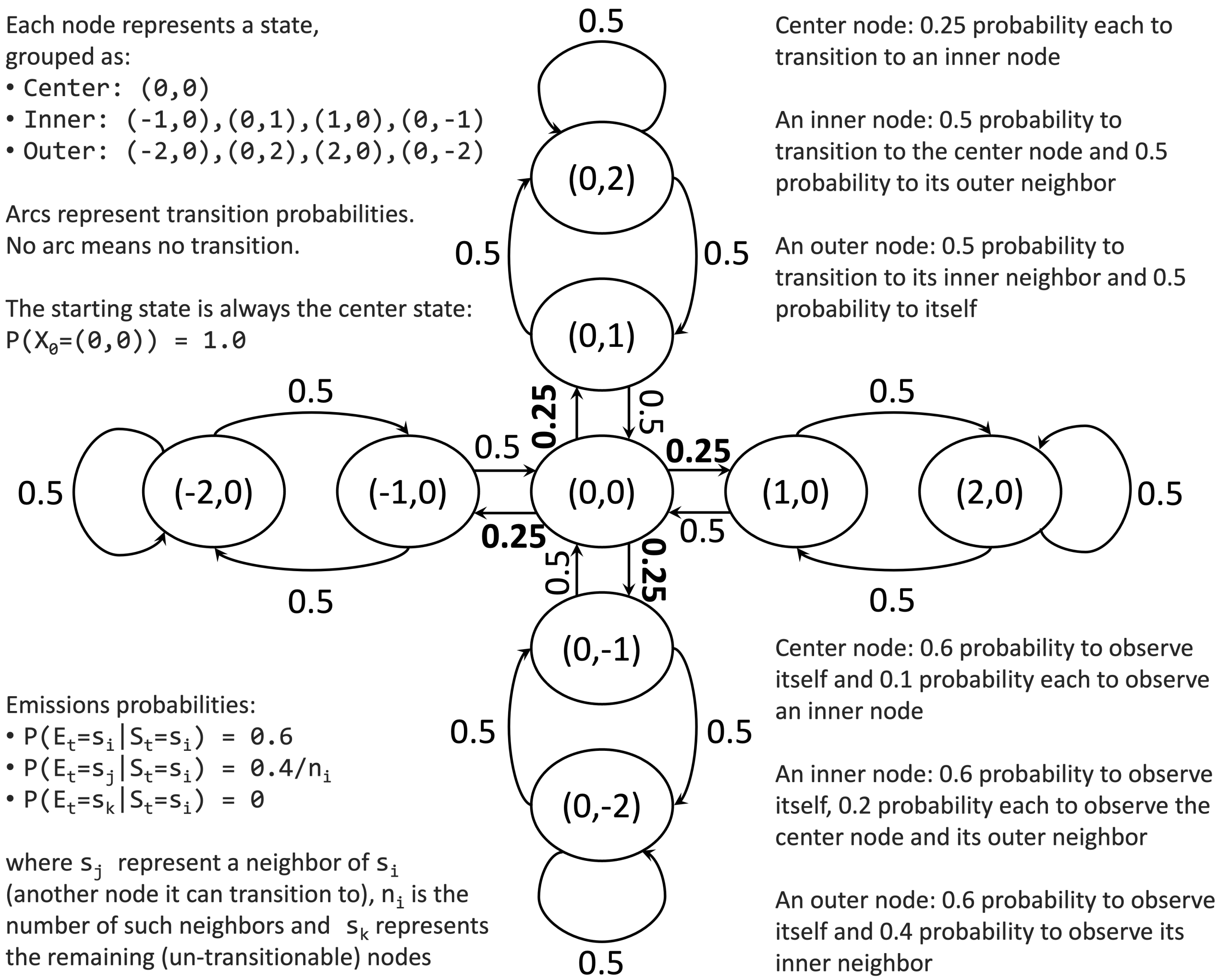
MEU for B: B is not observed yet so we weight the MEUs with $P(B)$

$$\begin{aligned} MEU(B) &= \sum_{b\in\{+b,-b\}} P(B = b)MEU(B = b) \\ &= 0.32 \cdot 68 + 0.68 \cdot 92 = 68(0.32 + 0.92) = 84.32 \end{aligned}$$

Correct equation 1 point (0 if they include any actions), correct value 0.5 point. Total 1.5

VPI of B: The last step is to calculate the VPI (1 point for the equation)
$VPI(B) = MEU(B) - MEU() = 1.24 \cdot 68 - 68 = 0.24 \cdot 68 = 16.32$.
1 point for overall legibility and clarity.

7. (22 points) Answer the questions based on the given Hidden Markov Model (HMM) below.



(a) (2 points) The HMM transitions but you have not made an observation yet. What is the state distribution ($P(X_1)$)? You do not have to write any 0 probabilities.
The correct answer will get full points but need the equation for partial.
Equation $P(X_1 = x_1) = \sum_{x_0} P(X_1 = x_1|X_0 = x_0)P(X_0 = x_0)$. For an inner state:
$P(X_1 = (-1, 0)) = P(X_1 = (-1, 0)|X_0 = (0, 0))P(X_0 = (0, 0)) = 0.25 \cdot 1 = 0.25$
Similarly: $P(X_1 = (1, 0)) = P(X_1 = (0, -1)) = P(X_1 = (0, 1)) = 0.25$

(b) (2 points) After the first transition, you observed $E_1 = (0, 0)$. What is the state distribution ($P(X_1|E_1 = (0, 0))$)? You do not have to write any 0 probabilities.
Emission probabilities are equal, the distribution does not change. Correct answer will get full but need the equation for partial. Wrong emission probabilities will be ignored if they are equal.
Equation: $P(X_1 = x_1|E_1 = e_1) \propto P(E_1 = e_1|X_1 = x_1)P(X_1 = x_1))$. For an inner state:
$P(X_1 = (-1, 0)|E_1 = (0, 0)) \propto P(E_1 = (0, 0)|X_1 = (-1, 0))P(X_1 = (-1, 0))) = 0.2 \cdot 0.25 = 0.05$
$P(X_1 = (1, 0)|E_1 = (0, 0)) = P(X_1 = (0, -1)|E_1 = (0, 0)) = P(X_1 = (0, 1)|E_1 = (0, 0)) \propto 0.05$
Normalize: $P(X_1 = (-1, 0)|E_1 = (0, 0)) = P(X_1 = (1, 0)|E_1 = (0, 0)) = P(X_1 = (0, -1)|E_1 = (0, 0)) = P(X_1 = (0, 1)|E_1 = (0, 0)) = 0.25$

(c) (2 points) Another transition happens and you observe $E_2 = (-1, 0)$. What are the possible states that you can be in?

We can only observe $(-1, 0)$ in states $(-1, 0)$, $(0, 0)$ and $(-2, 0)$. However, we cannot be in state $(-1, 0)$ after two transitions starting from $(0, 0)$. Thus the answer is $(0, 0)$ and $(-2, 0)$. Just the states without the explanation will get full points.

(d) (2 points) Calculate the probabilities for the states in the previous part (essentially $P(X_2|E_1 = (0, 0), E_2 = (-1, 0))$ but you can take short-cuts).

We can be in the inner states at $t = 1$ and $(0, 0)$ or $(-2, 0)$ at $t = 2$. We can transitionto $(0, 0)$ from all the inner states. We can only transition to $(-2, 0)$ from $(-1, 0)$. With this info:

Let $B_t(X) = P(X|e_{1:t})$ ($B_1(X)$ from part b).
$B_2(X = (0, 0)) \propto P(E_2 = (-1, 0)|X_2 = (0, 0)) \sum_s P(X_2 = (0, 0)|X_1 = s)B_1(X = s)$
$= 0.1 \cdot (0.5 \cdot 0.25 + 0.5 \cdot 0.25 + 0.5 \cdot 0.25 + 0.5 \cdot 0.25) = 0.05$
$B_2(X = (-2, 0)) \propto P(E_2 = (-1, 0)|X_2 = (-2, 0)) \sum_s P(X_2 = (-2, 0)|X_1 = s)B_1(X = s)$
$= 0.4 \cdot (0.5 \cdot 0.25) = 0.05$

We normalize these to get $B_2(X = (0, 0)) = B_2(X = (-2, 0)) = 0.5$.

(e) (6 points) This time two transitions happen, but you forget the observe the first one (E_3 is missing). The last observation you have is $E_4 = (0, -1)$. What is the distribution of the states now ($P(X_4|E_1 = (0, 0), E_2 = (-1, 0), E_4 = (0, -1))$)?

We need to first just do a "passage of time" step. Then we will do full forward update (or two passage of time steps followed by an observation step).

At $t = 2$, we can only be in $(0, 0)$ and $(-2, 0)$ as found above. At $t = 3$, we can only be in the inner nodes or $(-2, 0)$. We can transition to $(-1, 0)$ from both $(0, 0)$ and $(-2, 0)$. The other inner nodes are only reached from $(0, 0)$:
$B'_3(X) = P(X|e_1, e_2) = \sum_{X_2} B_2(X_2)P(X_3|X_2)$
$B'_3((-2, 0)) = B_2((-2, 0))P((-2, 0)|(-2, 0)) = 0.5 \cdot 0.5 = 0.25$
$B'_3((-1, 0)) = B_2((-2, 0))P((-1, 0)|(-2, 0)) + B_2((0, 0))P((-1, 0)|(0, 0))$
$= 0.5 \cdot 0.5 + 0.5 \cdot 0.25 + = 0.375$
$B'_3((1, 0)) = B'_3((0, 1)) = B'_3((0, -1)) = B_2((0, 0))P(\cdot|(0, 0)) = 0.5 \cdot 0.25 = 0.125$

In short: $(-2, 0) : 0.250, (-1, 0) : 0.375, (1, 0) : 0.125, (0, 1) : 0.125, (0, -1) : 0.125$ This part is 3 points (1 point for only results)

We can only observe $(0, -1)$ from $(0, 0)$, $(0, -1)$ and $(0, -2)$. We can transition into $(0, 0)$ from all the inner nodes. We can transition into $(0, -2)$ from $(0, -1)$. However, we cannot transition into $(0, -1)$ given the information! (cannot be in $(0, 0)$ or $(0, -2)$ at $t = 3$). Thus, we will run the forward equations only for $(0, 0)$ and $(0, -2)$. $B_4(X) = P(X|e_1, e_2, e_4) \propto P(e_4|X) \sum_{X_3} B'_3(X_3)P(X|X_3)$
$B_4((0, 0)) = P((-1, 0)|(0, 0)) \left(\sum_{X \in inner} B'_3(X)P((0, 0)|X)\right)$
$= 0.1(0.5(0.125 \cdot 3 + 0.375)) = 0.0375$
$B_4((0, -2)) = 0.4(0.125 \cdot 0.5) = 0.025$

We normalize this to get $P(X_4 = (0, 0)|E_1 = (0, 0), E_2 = (-1, 0), E_4 = (0, -1)) = 0.6$ and $P(X_4 = (0, -2)|E_1 = (0, 0), E_2 = (-1, 0), E_4 = (0, -1)) = 0.4$. This part is 3 points (1 point for only results)

(f) (8 points) Starting from scratch, you observe $E_1 = (0,0), E_2 = (-1,0), E_3 = (-1,0), E_4 = (0,0)$ (the first observation comes after a transition). What is the most likely state path?
Correct state path without any reasoning: 2 points and no partial. Correct state path with verbal reasoning: 4 points Using a state trellis in addition to verbal reasoning but lacking exact numbers: 6 points Any type of solution (visual e.g. state trellis or directly the equations) with correct numbers/equations: 8 points (deduct few points for mathematical errors)

Mathematical Solution (with shortcuts and unwritten values being 0):
$m_t(X) = P(E = e_t|X)\max_{x_{t-1}}(m_{t-1}(x_{t-1})P(X|x_{t-1})), m_0(0,0) = 1.$

Only inner states at $t = 1$ with the same value:
$m_1(X) = 0.2 \cdot 1 \cdot 0.25 = 0.05, X \in$ Inner States (Argmax is (0,0) for all)

With the given e_2, we can be in either (-2,0) or (0,0).
$m_2((0,0)) = 0.1 \cdot 0.05 \cdot 0.5 = 0.0025$ (same value from all inner states (argmax is tied) so didn't write the max).
$m_2((-2,0)) = 0.4 \cdot 0.05 \cdot 0.5 = 0.01$ (only possiblity is $(-1,0)$)

With the given e_3 and the past, we can only be in $(-1,0)$ or $(-2,0)$
$m_3((-1,0)) = 0.6 \cdot \max(0.0025 \cdot 0.25, 0.01 \cdot 0.5) = 0.003$ with argmax $(-2,0)$
$m_3((-2,0)) = 0.4 \cdot 0.01 \cdot 0.5 = 0.002$ with argmax $(-2,0)$

With the given e_4 and the past, we can only be in $(0,0)$ or $(-1,0)$
$m_4((0,0)) = 0.6 \cdot 0.003 \cdot 0.5 = 0.0009$ with argmax $(-1,0)$ (only option)
$m_4((-1,0)) = 0.2 \cdot 0.002 \cdot 0.5 = 0.0002$ with argmax $(-2,0)$ (only option)

At this point argmax of m_4 is (0,0). The argmax (only option) was (-1,0). For $t = 3$ and (-1,0), argmax was (-2,0). We have only single options going back for the rest as (-1,0) and (0,0). Note that there was only two points where we had alternatives Reversing the order, we get the state path:
$(0,0) \to (-1,0) \to (-2,0) \to (-1,0) \to (0,0)$

An example with just reasoning (it is possible to only have a forward reasoning for this problem.):

- t=0: Initially we are at (0,0)
- t=1: All inner states have the same likelihood.
- t=2: We can be in either (-2,0) or (0,0). Their state likelihoods are the same, however, for comparing state paths, (-2,0) is more likely (but (0,0) is in more possible paths)
- t=3: We can be in the "left branch" (cannot be in (0,0)).
- t=4: Based on the emission, we can be in (0,0) or (-1,0). With the emission, (0,0) is more likely which is our choice for $t = 4$. Now let's go backwards.
- t=3: We can only transition into (0,0) from (-1,0) so this is what we pick.
- t=2: Now there are two choices. We pick (-2,0) both due to the reasoning in bullet point 3 and that transitioning to (-1,0) from (-2,0) is more likely than from (0,0) (0.5 vs 0.25).
- t=1: We can only transition to (-2,0) from (-1,0) so this is what we pick.
- t=0: (0,0) is already given

In short: $(0,0) \to (-1,0) \to (-2,0) \to (-1,0) \to (0,0)$