KOÇ UNIVERSITY
College of Engineering

COMP 341: INTRODUCTION TO ARTIFICIAL INTELLIGENCE

Midterm 3

Fall 2021, 26/12/2021
Duration: 120 Minutes

Name: Solutions

ID: 00110001 00110000 00110000

- This exam contains 12 pages including this cover page and 4 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	Total
Points:	24	26	26	24	100
Score:					

1. (24 points) True or False :

True ____ Gibbs sampling allows the dependent variables to influence their parents.

True ____ Approximate sampling trades-off efficiency and accuracy.

False ____ Only probability nodes are enough to make a decision in Bayesian Networks.

False ____ Value of information can be negative.

False ____ Simulating markov chains forward will result in divergence of probability values.

True ____ Hidden Markov Models can incorporate more than one emisison before incorporating "passage of time".

False ____ The transition probabilities are used to sample the weights in particle filtering.

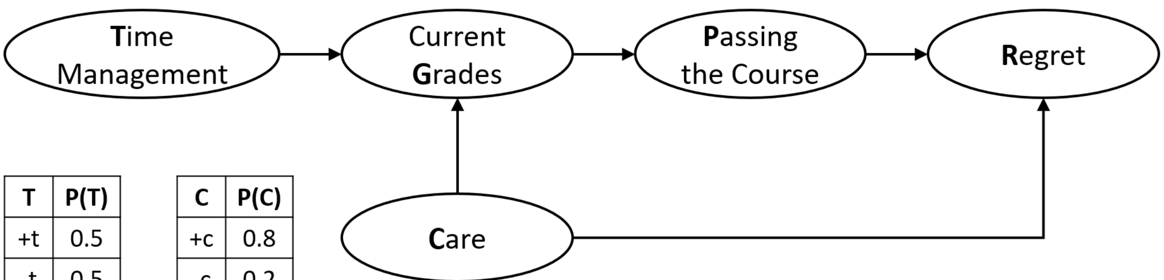
True ____ In Dynamic Bayesian Networks with more than one-hidden state, the emission probabilites are multiplied together to calculate likelihoods.

2. (26 points) BN Approximate Inference: You are given the following Bayesian Network (BN) below which is the same one from MT2. Answer the questions below based on this BN. You can leave the final answer as fractions or multiplications and sumations of values!

T	C	G	P(G|T,C)
+t	+c	+g	0.9
+t	+c	-g	0.1
+t	-c	+g	0.5
+t	-c	-g	0.5
-t	+c	+g	0.7
-t	+c	-g	0.3
-t	-c	+g	0.2
-t	-c	-g	0.8

G	P	P(P|G)
+g	+p	0.8
+g	-p	0.2
-g	+p	0.3
-g	-p	0.7

C	P	R	P(R|C,P)
+c	+p	+r	0.0
+c	+p	-r	1.0
+c	-p	+r	1.0
+c	-p	-r	0.0
-c	+p	+r	0.0
-c	+p	-r	1.0
-c	-p	+r	0.5
-c	-p	-r	0.5



T	P(T)
+t	0.5
-t	0.5

C	P(C)
+c	0.8
-c	0.2

This BN models whether a student will pass an unnamed course or not, and the associated regret. The variables, which are all binary, are:

1. **T**ime Management: Whether (+t) or not (-t) the student has good time management skills
2. **C**are: Whether the student cares about the course (+c) or not (-c)
3. Current **G**rades: The grades that the student has received so far into the semester. They are modelled as either good (+g) or bad (-g)
4. **P**assing the Course: Whether the student will pass (+p) the course or not (-p).
5. **R**egret: Whether the student feels regret (+r) or not (-r) after the semester ends

(a) (4 points) What is the most likely sample from this BN, i.e., if we were to run prior sampling, which sample would be repeated the most? If there is more than one, list all of them? What is the likelihood of the sample(s)? Show your work (i.e. the numbers you multiply) .

We calculate this by taking the argmax value of each variable, conditioned on previous selections going from parents to children (i.e. in the topological order.)

The most likely sample is $\{+t, +c, +g, +p, -r\}$ The corresponding likelihood is $0.5 \times 0.8 \times 0.9 \times 0.8 \times 1.0 = 0.288$. 2 points (only 1 point if you do not show your work but full points even if the final number is not calculated)

(b) (6 points) You saw your friend having regret. Last time you checked, he had good grades. He probably failed the course but you want to estimate his time management skills (i.e. $P(T = +t|G = +g, R = +r)$) with approximate inference. You run likelihood weighting and get the samples given below. Calculate the weight of each sample and show your work.

Sample	Count	Weights
$+t, -c, +g, +p, +r$	8	$P(+g\| +t, -c) \times P(+r\| -c, +p) = 0.5 \times 0 = 0$
$-t, -c, +g, -p, +r$	3	$0.2 \times 0.5 = 0.1$
$-t, +c, +g, +p, +r$	31	$0.7 \times 0 = 0$
$+t, +c, +g, +p, +r$	29	$0.9 \times 0 = 0$
$+t, +c, +g, -p, +r$	12	$0.9 \times 1 = 0.9$
$-t, +c, +g, -p, +r$	7	$0.7 \times 1 = 0.7$
$-t, -c, +g, +p, +r$	6	$0.2 \times 0 = 0$
$+t, -c, +g, -p, +r$	4	$0.5 \times 0.5 = 0.25$

Only weights that are not 1 are from the evidence CPTs. No partial grades for multiplying everything together. Half points for no work but correct answer

(c) (4 points) Calculate $P(T = +t|G = +g, R = +r)$ using the counts and the weights from the previous part.

$P(T = +t|G = +g, R = +r) =$

$$\frac{8 \times 0 + 29 \times 0 + 12 \times 0.9 + 4 \times 0.25}{8 \times 0 + 29 \times 0 + 12 \times 0.9 + 4 \times 0.25 + 3 \times 0.1 + 31 \times 0 + 7 \times 0.7 + 6 \times 0}$$

$= \frac{11.8}{11.8 + 5.2} = 11.8/17 = 59/85 \approx 0.694$

I will accept answers who drop the numbers involving zero multiplications. Calculating the final number is not needed but the form must be correct (i.e. correct counts multiplied with correct weights and correct numerator, denominator)

(d) (2 points) Now you want to do Gibbs sampling. Your current state is $+t, -c, +g, -p, +r$. You are going to sample the variable **P** next. Which probability distribution will you use, i.e., what is the probabilistic inference query? (e.g. just a single short statement as $P(A = +a|B = -b, C = +c)$). Simplify as much as you can (no partial for un-simplified).

The query is $P(P|T = +t, C = -c, G = +g, R = +r)$. From the BN, we can see that P is independent of T given the other variables. Thus we simplify it to $P(P|C = -c, G = +g, R = +r)$. Un-simplified will only get 1 point. Wrong answers will get 0

(e) (4 points) How do you get to the distribution in the previous part using the conditional probability tables of the Bayesian Network? (Multiply and divide them).

With the conditional independence it is easy to see that $P(P, C = -c, G = +g, R = +r) = P(P| + g)P(+r|P, -c)$. We can normalize this to get $P(P| + g)P(+r|P, -c)$, i.e. $P(P|C = -c, G = +g, P = -p, R = +r) \propto P(P, C = -c, G = +g, P = -p, R = +r)$. The way to normalize is (dropping the variable naming):

$P(P| + t, -c, +g, - + r) = \dfrac{P(P| + g)P(+r|P, -c)}{\sum_P P(P| + g)P(+r|P, -c)}$

This can also be done by starting from the joint and getting to the result as well. I will accept this answer too. Dropping the variable naming:

$$\begin{aligned} P(P| + t, -c, +g, +r) &= \frac{P(+t, -c, +g, P, +r)}{P} \\ &= \frac{P(+t)P(-c)P(+g| + t, -c)P(P| + g)P(+r|P, -c)}{\sum_P P(+t)P(-c)P(+g| + t, -c)P(P| + g)P(+r|P, -c)} \\ &= \frac{P(P| + g)P(+r|P, -c)}{\sum_P P(P| + g)P(+r|P, -c)} \\ &\propto P(P| + g)P(+r|P, -c) \end{aligned}$$

(f) (6 points) Finally, calculate the conditional probability table you would use to sample **P** when the previous state is $+t, -c, +g, -p, +r$.

Just plug everything in

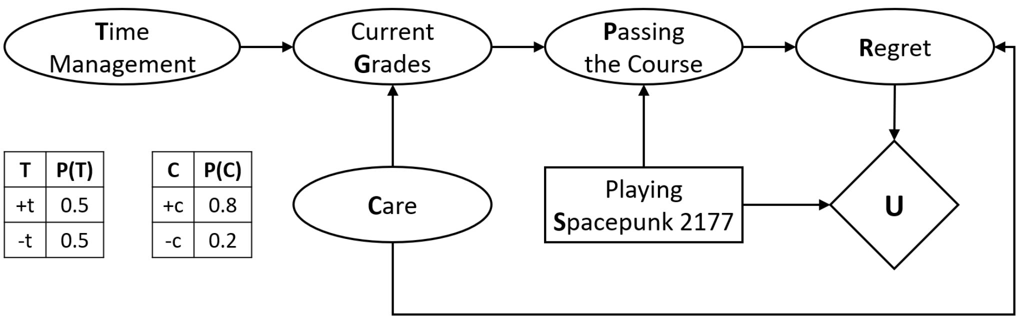
P	$P(P\| + g)P(+r\|P, -c)$	$P(P\| - c, +g, -p, +r)$ (Normalized)
$+p$	$0.8 \times 0 = 0$	0.0
$-p$	$0.2 \times 0.5 = 0.1$	1.0

3. (26 points) There is a new game that is about to be released. It is highly anticipated and a lot of students want to play the game. However, doing so affects their chances of passing the course since the game can be addictive, even with good time management skills. The previous Bayesian Network is modified to reflect this and get the below Decision Network.

T	C	G	P(G|T,C)
+t	+c	+g	0.9
+t	+c	-g	0.1
+t	-c	+g	0.5
+t	-c	-g	0.5
-t	+c	+g	0.7
-t	+c	-g	0.3
-t	-c	+g	0.2
-t	-c	-g	0.8

G	S	P	P(P|G,S)
+g	+s	+p	0.4
+g	+s	-p	0.6
+g	-s	+p	0.8
+g	-s	-p	0.2
-g	+s	+p	0.1
-g	+s	-p	0.9
-g	-s	+p	0.3
-g	-s	-p	0.7

C	P	R	P(R|C,P)
+c	+p	+r	0.0
+c	+p	-r	1.0
+c	-p	+r	1.0
+c	-p	-r	0.0
-c	+p	+r	0.0
-c	+p	-r	1.0
-c	-p	+r	0.5
-c	-p	-r	0.5



T	P(T)
+t	0.5
-t	0.5

C	P(C)
+c	0.8
-c	0.2

R	S	U
+r	+s	-100
+r	-s	-80
-r	+s	100
-r	-s	-10

The added nodes are:

1. Playing **S**pacefunk 2177: Whether the student decides to play the game (+s) or not (-s)
2. **U**: The utility student receives at the end of the semester based on their regret status and whether they played the game or not

Hint: Whever you are evaluating an action, treat it as an evidence variable for the corresponding Conditional Probability Tables. You are given the following probabilities to help with your calculations:

1. $P(P = +p|C = +c, S = +s) = 0.34$
2. $P(P = +p|C = +c, S = -s) = 0.7$
3. $P(R = +r|G = +g, C = +c, S = +s) = 0.6$
4. $P(R = +r|G = +g, C = +c, S = -s) = 0.2$
5. $P(R = +r|G = -g, C = +c, S = +s) = 0.9$
6. $P(R = +r|G = -g, C = +c, S = -s) = 0.7$
7. $P(G = +g|C = +c) = 0.8$
8. $P(G = -g|C = +c) = 0.2$

(a) (8 points) You care about the course but you also want to play the game. To decide whether to play or not, you want to calculate the expected utility of each action and pick the optimal one. Calculate $EU(S = +s|C = +c)$ and $EU(S = -s| + c)$. Show your work.

Expected utility is probability times utility. The query variables are the ones connected to the utility node (evidence and hidden variables are as usual).

$$EU(S = +s|C = +c) = \sum_{r=\{+r,-r\}} P(R = r|C = +c, S = +s)U(R = r, S = +s)$$
$$EU(S = -s|C = +c) = \sum_{r=\{+r,-r\}} P(R = r|C = +c, S = -s)U(R = r, S = -s)$$

We need the following distribution for this: $P(R|C = +c, S)$. We have $P(R|C, P)$ as a CPT. We are also given $P(P|C = +c, S)$ (note that the $P = -p$ versions are calcualted by subtracting from 1). With these two we can write (including the conditional independence relations):
$P(R, P| + c, S) = P(R| + c, P)P(P| + c, S)$
$\Rightarrow P(R| + c, S) = \sum_{p=\{+p,-p\}} P(R| + c, p)P(p| + c, S)$

S	R	$P(R\| + c, S)$
$+s$	$+r$	$0 \cdot 0.34 + 1 \cdot (1 - 0.34) = 0.66$
$+s$	$-r$	$1 \cdot 0.34 + 0 \cdot (1 - 0.34) = 0.34$
$-s$	$+r$	$0 \cdot 0.70 + 1 \cdot (1 - 0.7) = 0.30$
$-s$	$-r$	$1 \cdot 0.70 + 0 \cdot (1 - 0.7) = 0.70$

Plug in the values:
$EU(S = +s|C = +c) = 0.66 \cdot -100 + 0.34 \cdot 100 = -32$
$EU(S = -s|C = +c) = 0.3 \cdot -80 + 0.7 \cdot -10 = -31$

(b) (2 points) Based on your answer to the previous part, do you play the game or not?

$EU(S = -s|C = +c) > EU(S = +s|C = +c)$, thus the best decision is not to play.

(c) (12 points) You care about the course but the instructor still has not announced the grades yet. You want to calculate the value of the grades information to see if it is worth e-mailing the instructor. Towards this end, you need to calculate $MEU(C = +c, G)$ (note that G is a random variable still!).

The question is asking $VPI(G|+c) = MEU(+c, G) - MEU(+c)$. From the previous question, we have calculated $MEU(+c) = -31$.

Since we do not know the value of the variable of interest, we have to calculate the probabilities of the potential outcomes. We then calculate the expected utilities for each potential outcome, as if they happened, multiply them with outcome probabilites and sum them up. To get the maximum expected utility, we need to calculate it for both actions and take the max. All of this together:

$$MEU(+c, G) = \sum_{g=\{+g,-g\}} P(g|+c) \max_s \left(\sum_{r=\{+r,-r\}} P(r|+c, g, s) U(r, s) \right)$$

We need $P(G|+c)$ and $P(R|+c, G, S)$. Both of these are given to us! Let's calculate $EU(S|+c, G) = \sum_{r=\{+r,-r\}} P(r|+c, G, S) U(r, S)$ first :

S	G	$EU(S\|+c, G)$
$+s$	$+g$	$0.6 \cdot -100 + 0.4 \cdot 100 = -20$
$+s$	$-g$	$0.9 \cdot -100 + 0.1 \cdot 100 = -80$
$-s$	$+g$	$0.2 \cdot -80 + 0.8 \cdot -10 = -24$
$-s$	$-g$	$0.7 \cdot -80 + 0.3 \cdot -10 = -59$

The best move for $+g$ is $+s$ with -20 and for $-g$ it is $-s$ with -59.

$MEU(+c, G) = 0.8 \cdot -20 + 0.2 \cdot -59 = -27.8$

$\Rightarrow VPI(G|+c) = MEU(+c, G) - MEU(+c) = -27.8 - (-31) = 3.2$

(d) (4 points) At this point do you have enough information to make the decision to contact the instructor? If so what do you do? If not, why not?

We need the "cost" of contacting the instructor to make a decision. If the cost is below 3.2, we would contact. If you have assumed a cost and made the comparison you get 3 points (since no cost was given). Any "philosophical" answers would get 0 points.

4. (24 points) You are trying to find your friend in a building with 4 floors that is arranged as given below. F1 is the first floor and F4 is the fourth floor. Your friend randomly moves between the floors. The probability of going up and down at each floor and the prior probability of being at a floor is given below. In addition, there is a sound sensor on the first floor. The corresponding emission probabilities are also given below.

F4
F3
F2
F1

X_0	$P(X_0)$
F4	0.2
F3	0.3
F2	0.1
F1	0.4

X_T	Stay	Down	Up
F4	0.6	0.4	-
F3	0.5	0.3	0.2
F2	0.5	0.2	0.3
F1	0.6	-	0.4

S	P(S|X)
-s	0.1+0.2d
+s	0.9-0.2d

d: floor difference from the first floor
X = F1, d = 0
X = F2, d = 1
X = F3, d = 2
X = F4, d = 3

I have fixed the text to avoid confusion in the future (i.e. fixed it for myself).

(a) (6 points) Initial ($t = 0$) distribution is given to you as the prior distribution. What is his state distribution at $t = 1$ after moving once? Show your work.

$B'_1(X) = \sum_{x_0} P(X|x_0)P(x_0)$

$$\begin{aligned}
B'_1(X = F_1) &= P(F_1|F_1)P(F_1) + P(F_1|F_2)P(F_2) + P(F_1|F_3)P(F_3) + P(F_1|F_4)P(F_4) \\
(given) &= 0.6 \cdot 0.4 + 0.2 \cdot 0.1 + 0.0 \cdot 0.3 + 0.0 \cdot 0.2 = 0.26 \\
(uniform) &= 0.6 \cdot 0.25 + 0.2 \cdot 0.25 + 0.0 \cdot 0.25 + 0.0 \cdot 0.25 = 0.20 \\
B'_1(X = F_2) &= P(F_2|F_1)P(F_1) + P(F_2|F_2)P(F_2) + P(F_2|F_3)P(F_3) + P(F_2|F_4)P(F_4) \\
(given) &= 0.4 \cdot 0.4 + 0.5 \cdot 0.1 + 0.3 \cdot 0.3 + 0.0 \cdot 0.2 = 0.30 \\
(uniform) &= 0.4 \cdot 0.25 + 0.5 \cdot 0.25 + 0.3 \cdot 0.25 + 0.0 \cdot 0.25 = 0.30 \\
B'_1(X = F_3) &= P(F_3|F_1)P(F_1) + P(F_3|F_2)P(F_2) + P(F_3|F_3)P(F_3) + P(F_3|F_4)P(F_4) \\
(given) &= 0.0 \cdot 0.4 + 0.3 \cdot 0.1 + 0.5 \cdot 0.3 + 0.4 \cdot 0.2 = 0.26 \\
(uniform) &= 0.0 \cdot 0.25 + 0.3 \cdot 0.25 + 0.5 \cdot 0.25 + 0.4 \cdot 0.25 = 0.30 \\
B'_1(X = F_4) &= P(F_4|F_1)P(F_1) + P(F_4|F_2)P(F_2) + P(F_4|F_3)P(F_3) + P(F_4|F_4)P(F_4) \\
(given) &= 0.0 \cdot 0.4 + 0.0 \cdot 0.1 + 0.2 \cdot 0.3 + 0.6 \cdot 0.2 = 0.18 \\
(uniform) &= 0.0 \cdot 0.25 + 0.0 \cdot 0.25 + 0.2 \cdot 0.25 + 0.6 \cdot 0.25 = 0.20
\end{aligned}$$

(b) (6 points) After your friend moves, the sound sensor is activated ($S_1 = +s$). What is his state distribution now? Show your work.

$B_1(X) = P(s|X)B'_1(X)$

$$
\begin{aligned}
B_1(X = F_1) \propto P(F_1| + s)B'_1(X = F_1) &= 0.9 \cdot 0.26 = 0.234(given) \\
&= 0.9 \cdot 0.2 = 0.18(uniform) \\
B_1(X = F_2) \propto P(F_2| + s)B'_1(X = F_2) &= 0.7 \cdot 0.3 = 0.21(given) \\
&= 0.7 \cdot 0.3 = 0.21(uniform) \\
B_1(X = F_3) \propto P(F_3| + s)B'_1(X = F_3) &= 0.5 \cdot 0.26 = 0.13(given) \\
&= 0.5 \cdot 0.3 = 0.15(uniform) \\
B_1(X = F_4) \propto P(F_4| + s)B'_1(X = F_4) &= 0.3 \cdot 0.18 = 0.054(given) \\
&= 0.3 \cdot 0.2 = 0.06(uniform)
\end{aligned}
$$

Let $\alpha_G = 0.234 + 0.21 + 0.13 + 0.054 = 0.628$, $\alpha_U = 0.18 + 0.21 + 0.15 + 0.06 = 0.6$

After normalization (2 points):

$$
\begin{aligned}
B_1(X = F_1) &= 0.234/0.628 = 117/314 \approx 0.373(given) \\
&= 0.18/0.6 = 0.3(uniform) \\
B_1(X = F_2) &= 0.21/0.628 = 105/314 \approx 0.334(given) \\
&= 0.21/0.6 = 0.35(uniform) \\
B_1(X = F_3) &= 0.13/0.628 = 65/314 \approx 0.207(given) \\
&= 0.15/0.6 = 0.25(uniform) \\
B_1(X = F_4) &= 0.054/0.628 = 27/314 \approx 0.086(given) \\
&= 0.06/0.6 = 0.1(uniform)
\end{aligned}
$$

I will accept the fractional results

(c) (12 points) Your friend moves again and this time the sound sensor is not activated ($S_2 = -s$). What is his most likely state sequence for $t = 0, 1, 2$? Show your work.

(First rows given, second rows uniform)

$m_t(x_t) = P(e_t|x_t) \max_{x_{t-1}} (P(x_t|x_{t-1}) m_{t-1}(x_{t-1}))$ and m_0 is the priors

$$\begin{aligned}
m_1(F_1) &= P(+s|F_1) \max (P(F_1|F_1)m_0(F_1), P(F_1|F_2)m_0(F_2), P(F_1|F_3)m_0(F_3), P(F_1|F_4)m_0(F_4)) \\
&= 0.9 \max (0.6 \cdot 0.4, 0.2 \cdot 0.1, 0 \cdot 0.3, 0 \cdot 0.2) = 0.9 \cdot 0.24, \text{ argmax} = F_1 \\
&= 0.9 \max (0.6 \cdot 0.25, 0.2 \cdot 0.25, 0 \cdot 0.25, 0 \cdot 0.25) = 0.9 \cdot 0.15, \text{ argmax} = F_1 \\
m_1(F_2) &= P(+s|F_2) \max (P(F_2|F_1)m_0(F_1), P(F_2|F_2)m_0(F_2), P(F_2|F_3)m_0(F_3), P(F_2|F_4)m_0(F_4)) \\
&= 0.7 \max (0.4 \cdot 0.4, 0.5 \cdot 0.1, 0.3 \cdot 0.3, 0 \cdot 0.2) = 0.7 \cdot 0.16, \text{ argmax} = F_1 \\
&= 0.7 \max (0.4 \cdot 0.25, 0.5 \cdot 0.25, 0.3 \cdot 0.25, 0 \cdot 0.25) = 0.7 \cdot 0.125, \text{ argmax} = F_2 \\
m_1(F_3) &= P(+s|F_3) \max (P(F_3|F_1)m_0(F_1), P(F_3|F_2)m_0(F_2), P(F_3|F_3)m_0(F_3), P(F_3|F_4)m_0(F_4)) \\
&= 0.5 \max (0 \cdot 0.4, 0.3 \cdot 0.1, 0.5 \cdot 0.3, 0.4 \cdot 0.2) = 0.5 \cdot 0.15, \text{ argmax} = F_3 \\
&= 0.5 \max (0 \cdot 0.25, 0.3 \cdot 0.25, 0.5 \cdot 0.25, 0.4 \cdot 0.25) = 0.5 \cdot 0.125, \text{ argmax} = F_3 \\
m_1(F_4) &= P(+s|F_4) \max (P(F_4|F_1)m_0(F_1), P(F_4|F_2)m_0(F_2), P(F_4|F_3)m_0(F_3), P(F_4|F_4)m_0(F_4)) \\
&= 0.3 \max (0 \cdot 0.4, 0 \cdot 0.1, 0.2 \cdot 0.3, 0.6 \cdot 0.2) = 0.3 \cdot 0.12, \text{ argmax} = F_4 \\
&= 0.3 \max (0 \cdot 0.25, 0 \cdot 0.25, 0.2 \cdot 0.25, 0.6 \cdot 0.25) = 0.3 \cdot 0.15, \text{ argmax} = F_4
\end{aligned}$$

The maxes are easy.

$$\begin{aligned}
m_2(F_1) &= P(-s|F_1) \max (P(F_1|F_1)m_1(F_1), P(F_1|F_2)m_1(F_2), P(F_1|F_3)m_1(F_3), P(F_1|F_4)m_1(F_4)) \\
&= 0.1 \max (0.6 \cdot 0.9 \cdot 0.24, 0.2 \cdot 0.7 \cdot 0.16, 0, 0) = 0.1 \cdot 0.6 \cdot 0.9 \cdot 0.24 \\
&\text{argmax} = F_1 \\
&= 0.1 \max (0.6 \cdot 0.9 \cdot 0.15, 0.2 \cdot 0.7 \cdot 0.125, 0, 0) = 0.1 \cdot 0.6 \cdot 0.9 \cdot 0.15 \\
&\text{argmax} = F_1 \\
m_2(F_2) &= P(-s|F_2) \max (P(F_2|F_1)m_1(F_1), P(F_2|F_2)m_1(F_2), P(F_2|F_3)m_1(F_3), P(F_2|F_4)m_1(F_4)) \\
&= 0.3 \max (0.4 \cdot 0.9 \cdot 0.24, 0.5 \cdot 0.7 \cdot 0.16, 0.3 \cdot 0.5 \cdot 0.15, 0) = 0.3 \cdot 0.4 \cdot 0.9 \cdot 0.24 \\
&\text{argmax} = F_1 \\
&= 0.3 \max (0.4 \cdot 0.9 \cdot 0.15, 0.5 \cdot 0.7 \cdot 0.125, 0.3 \cdot 0.5 \cdot 0.125, 0) = 0.3 \cdot 0.4 \cdot 0.9 \cdot 0.15 \\
&\text{argmax} = F_1 \\
m_2(F_3) &= P(-s|F_3) \max (P(F_3|F_1)m_1(F_1), P(F_3|F_2)m_1(F_2), P(F_3|F_3)m_1(F_3), P(F_3|F_4)m_1(F_4)) \\
&= 0.5 \max (0, 0.3 \cdot 0.7 \cdot 0.16, 0.5 \cdot 0.5 \cdot 0.15, 0.4 \cdot 0.3 \cdot 0.12) = 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.15 \\
&\text{argmax} = F_3 \\
&= 0.5 \max (0, 0.3 \cdot 0.7 \cdot 0.125, 0.5 \cdot 0.5 \cdot 0.125, 0.4 \cdot 0.3 \cdot 0.15) = 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.125 \\
&\text{argmax} = F_3 \\
m_2(F_4) &= P(-s|F_4) \max (P(F_4|F_1)m_1(F_1), P(F_4|F_2)m_1(F_2), P(F_4|F_3)m_1(F_3), P(F_4|F_4)m_1(F_4)) \\
&= 0.7 \max (0, 0, 0.2 \cdot 0.5 \cdot 0.15, 0.6 \cdot 0.3 \cdot 0.12) = 0.7 \cdot 0.6 \cdot 0.3 \cdot 0.12 \\
&\text{argmax} = F_4 \\
&= 0.7 \max (0, 0, 0.2 \cdot 0.5 \cdot 0.125, 0.6 \cdot 0.3 \cdot 0.15) = 0.7 \cdot 0.6 \cdot 0.3 \cdot 0.15 \\
&\text{argmax} = F_4
\end{aligned}$$

The maxes are not too difficult to get without computation.

For the given priors, the overall max among m_2 is also not difficult to figure out without computation. It is $\text{argmax}(m_2) = F_2$. Its argmax is F_1. The argmax at $t = 1$ for F_1 is also F_1. Thus the most likely state path is F_1, F_1, F_2.

For the uniform priors, the overall max among m_2 is more difficult to figure out without computation but still doable. It is $\text{argmax}(m_2) = F_4$. Its argmax is also F_4. The argmax at $t = 1$ for F_4 is yet again F_4. Thus the most likely state path is F_4, F_4, F_4.

Non-Viterbi Solutions: Up to half points