CMSI 3300 – Classwork 5

Instructions:
This worksheet will not only provide you with practice problems whose concepts will help you understand HW3 and your final exam, and think about the types of applications in which some of our Markov models can be deployed in practice.

- Provide answers to each of the following questions and write your responses in the blanks. If you are expected to show your work in arriving at a particular solution, space will be provided for you.
- Place the names of your group members below:

Group Members:

1. Milo Fritzen
2. ______
3. ______

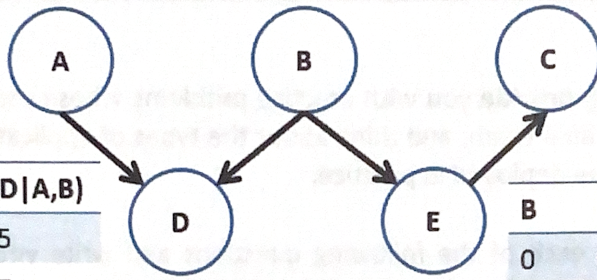
Problem 1 – Bayesian Network Approximate Inference

Consider the following Bayesian Network, and use it to answer the questions that follow:

B	P(B)
0	0.6

A	P(A)
0	0.8

E	C	P(C|E)
0	0	0.4
1	0	0.2



A	B	D	P(D|A,B)
0	0	0	0.5
0	1	0	0.7
1	0	0	0.8
1	1	0	0.1

B	E	P(E|B)
0	0	0.9
1	0	0.4

1.1. Provide a *possible* order in which variables would be sampled by *Prior Sampling:*

A, B, D, E, C

Suppose now that Prior Sampling is used to generate the following set of 10 samples.

Sample	A	B	C	D	E
0	0	1	1	0	0
1	0	0	1	1	0
2	1	0	0	1	1
3	0	1	1	1	1
4	1	1	1	0	0
5	0	0	0	1	1
6	0	1	1	1	0
7	1	0	0	1	0
8	0	0	1	1	0
9	0	0	1	1	0

1.2. Using these samples, find $\hat{P}(C = 1|D = 1)$

$\hat{P}(C=1|D=1) = \frac{5}{8}$

1.3. Using the same model, let's trace the generation of Gibbs Sampling in order to estimate $\hat{P}(B|C = 1, D = 1, E = 0)$. Suppose for the current sample we are sampling variables in the order of: $[A, B, \ldots]$ and begin with the following random state. <u>For each variable sampled thereafter, assume (for this exercise) that the most likely value of the distribution sampled-from is the one actually picked.</u>

A	B	C	D	E
0	1	1	1	0

Sampling A (Provided as an Example)

1.3.1. From what distribution will A be sampled? Show the likelihoods of sampling each value.

Distribution Sampled: $P(A|B = 1, C = 1, D = 1, E = 0)$

$$P(A = 0|B = 1, C = 1, D = 1, E = 0) = \frac{\{All\ CPTs\ Mentioning\ Query\ (A)\}}{\sum_a\{Those\ same\ CPTs\ as\ in\ numerator\}}$$

$$= \frac{P(A = 0)P(D = 1|A = 0, B = 1)}{\sum_a P(A = a)P(D = 1|A = a, B = 1)} = \frac{0.8 * 0.3}{0.8 * 0.3 + 0.2 * 0.9} \approx 0.57$$

1.3.2. Assuming our random number generator selects the value for A that is most likely, what will be the result of the above sample, and thus, the next state? (Fill in the value for A below):

A	B	C	D	E
0	1	1	1	0

Sampling B (Your Turn)

1.3.3. From what distribution will B be sampled? Show the likelihoods of sampling each value.

$$P(B=0|A=0, C=1, D=1, E=0) = \frac{P(B=0)P(D=1|A=0, B=0)\ P(E=0|B=0)}{\sum_{b \in B} P(b)\ P(D=1|A=0, b)\ P(E=0|b)}$$

$$= \frac{0.6(0.5)(0.9)}{0.6(0.5)(0.9) + (0.4)(0.3)(0.4)} \approx 0.85 \quad B=0$$

$$1-0.85 = 0.15 \quad B=1$$

1.3.4. Assuming our random number generator selects the value for B that is most likely, what will be the result of the above sample, and thus, the next state? (Fill in the value for B below):

A	B	C	D	E
0	0	1	1	0

Problem 2 – Hidden Markov Models

In bioinformatics, analyses of nucleotide sequences in DNA processing is a central, but noisy, process; determining the identity of each nucleotide (from the 4 possible: ACGT) is not only imperfect, but many nucleotides in the sequence are "non-coding," meaning they do not encode any protein sequences (and are, for many applications, ignorable). It turns out the likelihood of producing certain nucleotides (viz., GC) increases during "coding" ($X = 1$) sections, while the likelihoods of these two to express *decrease* during non-coding ($X = 0$). From lab data, we have parameterized the following HMM describing this process:

Initial

X	$P(X_0)$
0	0.50

Transitions

X_t	X_{t+1}	$P(X_{t+1}\|X_t)$
0	0	0.60
1	0	0.50

Emissions

E_t	$P(E_t\|X_t = 0)$	$P(E_t\|X_t = 1)$
A	0.30	0.20
C	0.20	0.30
G	0.20	0.30
T	0.30	0.20

2.1. Suppose we observe a (small) sequence of nucleotides: GGC. Treating our time step t as position within the sequence, use filtering to compute $B(X_3)$, our belief of the current state after observing the entire sequence, starting with the following as given:

	$t = 0$ *Init*	$t = 1$ $E_1 = G$	$t = 2$ $E_2 = G$	$t = 3$ $E_3 = C$
$B(X_t = 0)$	0.5	0.4	0.44	0.44
$B(X_t = 1)$	0.5	0.6	0.56	0.56

$$B'(X_3 = 0) = \sum_{X_2} \underbrace{P(X_3=0|X_2=x_2)}_{X_2=0}\underbrace{B(X_2=x_2)}_{X_2=1}$$

$$= 0.6(0.44) + 0.5(0.56) \approx 0.544$$

$$B'(x_3=1) = \sum_{x_2} (X_3=1|x_2)B(x_2)$$

$$= 0.4(0.44) + 0.5(0.56) \approx 0.456$$

$$B(x_3=0) = P(E_3=C|x_3=0)B'(x_3=0)\,d$$

$$= 0.2(0.544) = 0.1088d = \frac{0.1088}{0.2456} \approx 0.44$$

$$B(x_3=1) = P(E=c|x_3=1)B'(x_3=1)\,d$$

$$= 0.3(0.456) = 0.1368d = \frac{0.1368}{0.2456} \approx 0.56$$

$$d = \frac{1}{0.1088 + 0.1368} = \frac{1}{0.2456}$$

In this setting, since we really care about the *most likely state* at any given part of the sequence, it's often useful to seek an **explanation** of what led us to witness the evidence at any given time. However, to compute this *most likely explanation*, we need to amend our Forward Algorithm slightly; instead of summing over all *possible* previous states in the transitions, we focus on only the most likely, labeled $M(X_t)$, in what is known as the **Viterbi Algorithm**.

Forward Algorithm	*Viterbi Algorithm*
$B(X_{t+1}) = \alpha\, P(E_{t+1} \mid X_{t+1}) \sum_{x_t} P(X_{t+1} \mid x_t) B(x_t)$	$M(X_{t+1}) = \alpha\, P(E_{t+1} \mid X_{t+1}) \max_{x_t} P(X_{t+1} \mid x_t) M(x_t)$

2.2. Suppose we observe the sequence $GGCA$ and are using the Viterbi Algorithm to find the most likely explanation of the state at each point in its generation. Using the following partially-completed computations, find $M(X_4)$: (hint: it *is* as simple as it seems, that's why it's nice).

	$t = 0$ Init	$t = 1$ $E_1 = G$	$t = 2$ $E_2 = G$	$t = 3$ $E_3 = C$	$t = 4$ $E_4 = A$
$M(X_t = 0)$	0.5	0.4	0.4	0.4	0.6
$M(X_t = 1)$	0.5	0.6	0.6	0.6	0.4

$$M(X_4=0) = \alpha\, P(E_4=A \mid X_4=0)\, P(X_4=0 \mid X_3=1)\, M(X_3=1)$$
$$= 0.3(0.5)(0.6)\,\alpha = 0.09\alpha$$
$$= \frac{0.09}{0.15} = 0.6$$

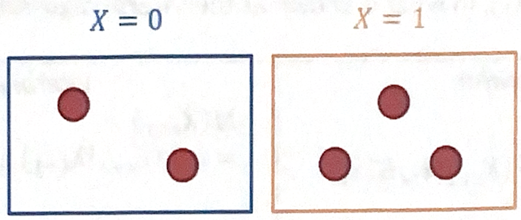
$$\alpha = \frac{1}{0.09+0.06} = \frac{1}{0.15}$$

$$M(X_4=1) = \alpha\, P(E_4=A \mid X_4=1)\, P(X_4=1 \mid X_3=1)\, M(X_3=1)$$
$$= 0.2(0.5)(0.6)\alpha = 0.06\alpha$$
$$= \frac{0.06}{0.15} = 0.4$$

2.3. Examining the completed table above, determine the *explanation* (i.e., the most likely value of the state) for the observed sequence in all steps $t = 1 \rightarrow 4$:

	$t = 1$ $E_1 = G$	$t = 2$ $E_2 = G$	$t = 3$ $E_3 = C$	$t = 4$ $E_3 = A$
X_t^*	1	1	1	0

Consider instead that we decided to use HMM Approximate Inference with a Particle Filter (doesn't make a whole lot of sense because the state is so small here but meh, practice!) with $N = 5$ particles (again, very small). Consider that our approximate belief state $\hat{B}(X_t)$ is represented by the current allocation of particles as drawn below:

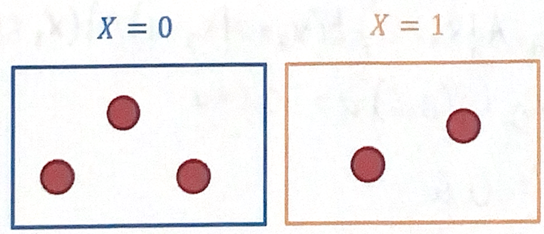


2.4. Suppose we now observe $E_{t+1} = T$ as the next nucleotide in the sequence, and we want to update our approximate belief state.

2.4.1. ***Elapse Time:*** What's the chance that each particle in the $X = 0$ state remain in this state?

$P(X_{t+1}=0 \mid X_t=0) = 0.6$

Suppose after elapsing time, we are left with the following configuration of particles:



2.4.2. ***Observe Evidence:*** Labeling the particles in the above diagram from left-to-right as $s^{(0)}, s^{(1)}, \ldots$, how are each weighted after observing $E_{t+1} = T$?

	$s^{(0)}$	$s^{(1)}$	$s^{(2)}$	$s^{(3)}$	$s^{(4)}$
$w(s^{(i)})$	0.3	0.3	0.3	0.2	0.2

2.4.3. ***Resample:*** What's the likelihood that each particle is resampled at $X = 0$ and $X = 1$?

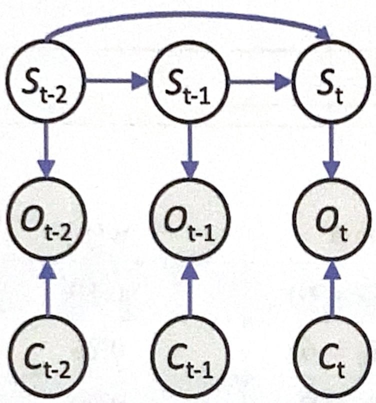
$P\left(S_{t+1}^{(i)} = 0 \mid E_{t+1} = T, S_t^{(i)} = s_t^{(i)}\right) =$ $3(0.3)\alpha = 0.9\alpha = \frac{0.9}{1.3} = 0.69$	$P\left(S_{t+1}^{(i)} = 1 \mid E_{t+1} = T, S_t^{(i)} = s_t^{(i)}\right) =$ $2(0.2)\alpha = 0.4 = \frac{0.4}{1.3} = 0.31$

In Hideo Kojima's reboot of Pacman, it is the year 2077, and the hunt between Pacman and ghosts has become more technologically dystopian. In this psychological thriller, Pacman is trying to hunt ghosts in an infinite hallway with array-like index positions labeled as in the diagram below. As it's the future, Pacman's got some new gadgets, and wants to estimate the ghost's **actual position, S_t**, using some additional flashy sensors (the entire budget of the game was used to animate them). From the sensors, Pacman can find, at each time step, **a noisy reading of the ghost's location, O_t**. However, the ghost too has gained new cloaking technology, such that if **it cloaks itself at any given time step C_t**, it adds extra noise to the sensor readings.

Noting that, at each time step, the sensors must calibrate to not only the previous step, but the one before that as well, Pacman generates the following modified HMM:

State S_t: Corridor Index

...	4	5	6	7	8	...



Transition Model

D_1	D_2	$P(S_t \mid S_{t-1}, S_{t-2})$
0	0	0.6
0	1	0.2
0	2	0.0
1	0	0.3
1	1	0.3
1	2	0.5

$D_1 = |S_t - S_{t-1}|$
$D_2 = |S_t - S_{t-2}|$

Emission Model

C_t	D_t	$P(O_t \mid S_t, C_t)$
0	0	0.4
0	1	0.2
0	2	0.1
1	0	0.6
1	1	0.2
1	2	0.0

$D_t = |O_t - S_t|$

Given that our belief $B(S_t)$ relies on two previous states, were we to use *Particle Filtering* in this problem, we would need each particle to hold two states. For example, the particle denoting a vote that the ghost was in index 6 at time $t = 6$ and then at index 7 at $t = 7$ is: $(S_6 = 6, S_7 = 7)$

2.5 Suppose we have a particle $(S_6 = 6, S_7 = 7)$; what is the probability that, when accounting for passage of time, the particle transitions to $(S_7 = 7, S_8 = 8)$? (Show how you arrive at your answer using the model above).

$P((S_7 = 7, S_8 = 8)) = P(S_8 = 8 \mid S_7 = 7, S_6 = 6) = 0.5$

$D_1 = |S_8 - S_7| = |8 - 7| = 1$

$D_2 = |S_8 - S_6| = |8 - 6| = 2$

2.6 Independent from the previous step, suppose we observe the following:

$$C_6 = 1, \quad O_6 = 6, \quad C_7 = 0, \quad O_7 = 8$$

When we update the likelihood of particles for these observations, what would be the weight w associated with a particle $(S_6 = 7, S_7 = 8)$? [Hint: examine what we are given in the emission model, and observe that we would have to incorporate essentially 2 updates rolled into 1 given that our particles consist of 2 states; this translates to a simple product of each step's emission].

$w((S_6=7, S_7=8)) = P(O_7=8 \mid S_7=8, C_7=0)\, P(O_6=6 \mid S_8=7, C_6=1)$ $D_7 = \lvert O_7 - S_7\rvert = \lvert 8-8\rvert = 0 \qquad = 0.4\,(0.2)$ $D_6 = \lvert O_6 - S_6\rvert = \lvert 6-7\rvert = 1 \qquad = 0.08$
$w((S_6 = 7, S_7 = 8)) =$ 0.08

Ignoring the previous tracking of a single particle, suppose we now have the following set of weighted particles, having already updated for time and observations, but have yet to resample:

Particle p	$w(p)$
$(S_7 = 5, S_8 = 6)$	0.10
$(S_7 = 7, S_8 = 6)$	0.25
$(S_7 = 7, S_8 = 7)$	0.30
	0.65

2.7 *Prior to resampling*, suppose we renormalize the weights associated with each state the "old-fashioned" way like in exact inference, what would be Pacman's belief state at $t = 8$? Complete the table below to indicate your answer, showing work for each cell [Hint: just look at S_8 in the particles above and the weights attached to each]:

Position, S_8	$\hat{B}(S_8)$
$S_8 = 5$	$\hat{B}(S_8=5) = \frac{0}{0.65} = 0$
$S_8 = 6$	$\hat{B}(S_8=6) = \frac{0.35}{0.65} \approx 0.54$
$S_8 = 7$	$\hat{B}(S_8=7) = \frac{0.3}{0.65} \approx 0.46$
$S_8 = 8$	$\hat{B}(S_8=8) = \frac{0}{0.65} \approx 0$