

Midterm II, Fall 2016

This test has 7 questions worth a total of **100** points, to be completed in 110 minutes. The exam is closed book, except that you are allowed to use a single two-sided hand written cheat sheet. No calculators or other electronic devices are permitted. Give your answers and show your work in the space provided.

Write the statement out below in the blank provided and sign. You may do this before the exam begins. Any plagiarism, no matter how minor, will result in an F.

“I have neither given nor received any assistance in the taking of this exam.”

Signature: _____

Name: _____ Your EdX Login: _____

SID: _____ Name of person to left: _____

Exam Room: _____ Name of person to right: _____

Primary TA: _____

- indicates that only one circle should be filled in.
- indicates that more than one box may be filled in.
- Be sure to fill in the and boxes completely and erase fully if you change your answer.
- There may be partial credit for incomplete answers. Write as much of the solution as you can, but bear in mind that we may deduct points if your answers are much more complicated than necessary.
- There are a lot of problems on this exam. Work through the ones with which you are comfortable first. **Do not get overly captivated by interesting problems or complex corner cases you’re not sure about.**
- Not all information provided in a problem may be useful.
- **There are some problems on this exam with a slow brute force approach and a faster, clever approach. Think before you start calculating with lots of numbers!**
- Write the last four digits of your SID on each page in case pages get shuffled during scanning.

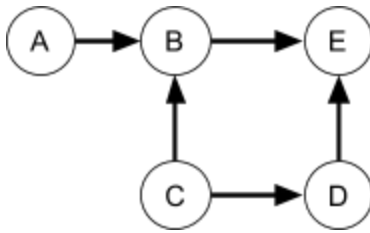
Problem	1	2	3	4	5	6	7
Points	14	9	12	20	14	13	18

Optional. Mark along the line to show your feelings on the spectrum between :(and ☺.

Before exam: [:(_____ ☺)].

After exam: [:(_____ ☺)].

1. Bayesics (14 pts) Model-Q1 is a Bayes' Net consisting of the graph and probability tables shown below:



A	P(A)
0	0.6
1	0.4

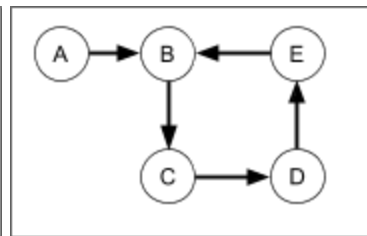
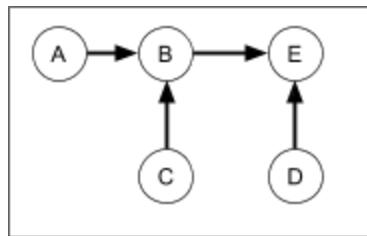
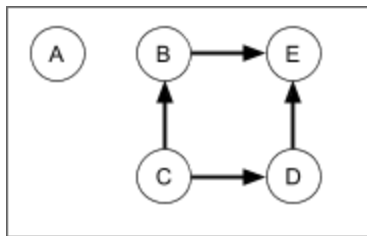
C	D	P(D C)
0	0	0.2
0	1	0.8
1	0	0.3
1	1	0.7

A	C	B	P(B A,C)
0	0	0	0.2
0	0	1	0.8
0	1	0	0.4
0	1	1	0.6
1	0	0	0.2
1	0	1	0.8
1	1	0	0.4
1	1	1	0.6

B	D	E	P(E B,D)
0	0	0	0.75
0	0	1	0.25
0	1	0	0.5
0	1	1	0.5
1	0	0	0.85
1	0	1	0.15
1	1	0	0.4
1	1	1	0.6

C	P(C)
0	0.4
1	0.6

i. (3 pts) Check the boxes above the Bayes' nets below that could also be valid for the above probability tables.



ii. (2 pts) Caryn wants to compute the distribution $P(A,C|E=1)$ using *prior sampling* on Model-Q1 (given at the top of this page). She draws a bunch of samples. The first of these is (0, 0, 1, 1, 0), given in (A, B, C, D, E) format. What's the probability of drawing this sample?

iii. (2 pts) Give an example of an inference query for Model-Q1 with **one query variable and one evidence variable** that could be estimated more efficiently (in terms of runtime) using *rejection sampling* than by using *prior sampling*. If none exist, state “not possible”.

iv. (2 pts) Give an example of an inference query for Model-Q1 with **one query variable and one evidence variable** for which *rejection sampling* provides no efficiency advantage (in terms of runtime) over using *prior sampling*. If none exist, state “not possible”.

v. (2 pts) Now Caryn wants to determine $P(A,C|E=1)$ for Model-Q1 using likelihood weighting. She draws the five samples shown below, which are given in (A, B, C, D, E) format, where the leftmost sample is “Sample 1” and the rightmost is “Sample 5”. What are the weights of the samples S1 and S3?

weight(S1): _____ weight(S3): _____

S1: (0, 0, 1, 1, 1) S2: (0, 0, 1, 1, 1) S3: (1, 0, 1, 1, 1) S4: (0, 1, 0, 0, 1) S5: (0, 1, 0, 0, 1)

vi. (1 pt) For the same samples as in part v, compute $P(A=1,C=1|E=1)$ for Model-Q1. Express your answer as a simplified fraction (e.g. $2/3$ instead of $4/6$).

vii. (2 pts) Select True or False for each of the following:

True False

- When there is no evidence, prior sampling is guaranteed to yield the exact same answer as inference by enumeration.
- When collecting a sample during likelihood weighting, evidence variables are not sampled.
- When collecting a sample during rejection sampling, variables can be sampled in any order.
- Gibbs sampling is a technique for performing approximate inference, not exact inference.

2. Pacmanian Language Modeling (9 pts)

Archaeologists uncover ancient Pacmanian ruins and discover a fragment of their rare writing. Your job is to analyze the Pacmanian language based *only* on the tiny amount of data uncovered (and your knowledge of material in this class). Specifically, the fragment contains the following 20 word sentence:

Nom Waka Pow Tuk Gah Waka Pow Tuk Bop Waka Pow Awoo Nom Waka Tuk Waka Gah
Waka Tuk Waka

Notes and hints:

- W_i represents the i^{th} word in the fragment. So $W_0 = \text{Nom}$ and $W_8 = \text{Bop}$
- The words and counts are: Awoo (1) Bop (1) Gah (2) Nom (2) Pow (3) Tuk (4) Waka (7)
- Your parameters should be computed from the 20 word fragment above!
- For those of you who are up to date on the post-MT2 material, **do not use Laplacian smoothing!**

i. (1 pt) Given a new four word sentence in Pacmanian, draw a Bayes' Network representing a **bigram language model**, where a word at position i can be dependent on the word that precedes it. You do not need to provide the conditional probability tables (CPTs).

ii. (2 pts) Given a bigram language model, write down the formula for the joint probability $P(W_0, W_1, W_2, W_3)$ in terms of the four smaller factors. You do not need to provide the CPTs.

iii. (2 pts) Given a bigram language model with parameters collected from the 20 word fragment, what is the probability of the Pacmanian sentence "Nom Waka Pow Awoo"? Hint: Only compute the parameters you need.

iv. (2 pts) Given a four word sentence in Pacmanian, draw a Bayes' Network representing a **trigram language model**, where a word at position i can be dependent on the **two** words that precede it. You do not need to provide the CPTs.

v. (2 pts) Given a trigram language model, write down the formula for the joint probability $P(W_0, W_1, W_2, W_3)$ in terms of the smaller factors.

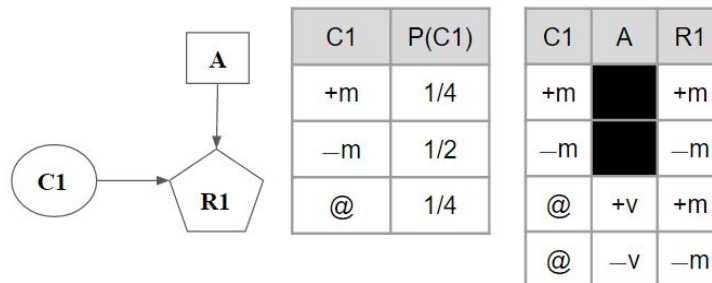
3. Smoke Weed Everyday (12 pts)

Allen is trying to decide whether to vote to legalize marijuana. He wants marijuana to be legalized, but he's pretty lazy and doesn't want to wait in line. He knows that while it's a close decision, he doesn't think his vote will actually matter.

Part A: Allen lives in some County C1, e.g Alameda. Our model has three possible scenarios for C1:

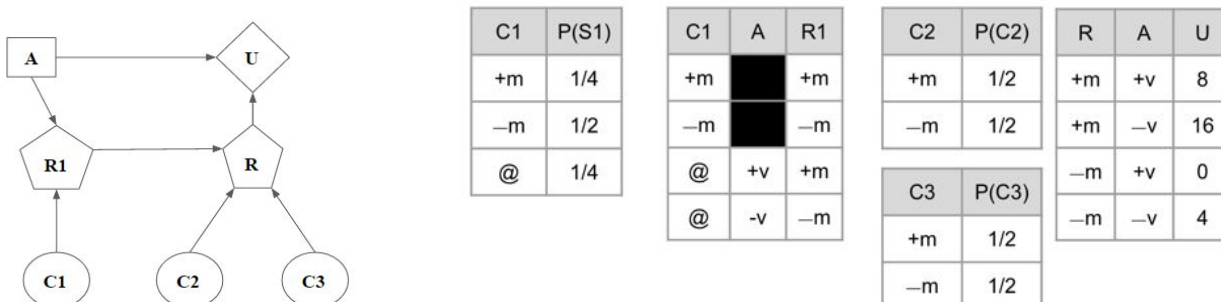
- Scenario **-m**: marijuana is not legalized regardless of Allen's vote (probability 1/2)
- Scenario **+m**: marijuana is legalized regardless of Allen's vote (probability 1/4)
- Scenario **@**: marijuana is legalized but only if Allen votes; otherwise it is not (probability 1/4)

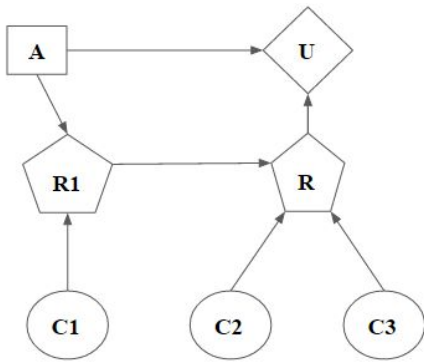
We can model this interaction between Allen's action A (**+v** for vote, **-v** for not vote) and C1 using a new type of node called a variable node. Its value is deterministically given by its inputs. The resulting network is shown below. The table on the right lists the result R1 of the county(C1)'s decision on marijuana, given the possible values for C1 and A. R1 is either (**+m**)arijuana legalized or (**-m**)arijuana not legalized. The black squares mean that Allen's vote doesn't matter in those cases.



- (0.5 pt) What is the chance of marijuana legalization assuming Allen votes (i.e. $P(R1 = +m \mid A = +v)$)?
- (0.5 pt) What is the chance of marijuana legalization assuming Allen doesn't vote (i.e. $P(R1 = +m \mid A = -v)$)?

Part B: Allen lives in a small state with only three counties, C1, C2, and C3. In counties C2 and C3, (**+m**)arijuana legalized or (**-m**)arijuana not legalized have a 50% chance each of occurring and Allen's vote does not affect those counties. Let R represent the outcome of the vote for the state, which can either be **R = +m** for legalization, or **R = -m** for not legalizing marijuana. The majority result of the counties determines the result of the state, e.g. if $R1 = +m, C2 = +m, C3 = -m$, then $R = +m$. The resulting model (including Allen's utility function) is given in the figure at the top of the next page. For space reasons, no table is shown for R in the figure. Larger version on next page.





C1	P(S1)
+m	1/4
-m	1/2
@	1/4

C1	A	R1
+m		+m
-m		-m
@	+v	+m
@	-v	-m

C2	P(C2)
+m	1/2
-m	1/2

C3	P(C3)
+m	1/2
-m	1/2

R	A	U
+m	+v	8
+m	-v	16
-m	+v	0
-m	-v	4

Make sure to show your work **clearly** on this page, as it will be used for determining partial credit!
 iii. (6 pts) **What is $MEU(\emptyset)$** , i.e. the maximum expected utility if Allen doesn't have any evidence? To maximize his utility, **should Allen vote**? Note that Allen's utility is a function of the result AND his action.

$MEU(\emptyset)$: _____ Should Allen vote? _____

iv. (3 pts) Suppose Allen could somehow know that $C1 = +m$. How much would he value additional information about the results of $C2$? In other words, what is $VPI(C2 | C1 = +m)$?

v. (2 pts) Is $VPI(C2|C1=@)$ greater than, less than, or equal to $VPI(C2|C1=+m)$?

- greater than
 less than
 equal

4. Probability Potpourri Pain Problem (20 pts)

i. (6 pts) Suppose we know that $A \perp B$ and $B \perp C | D$. Which of the following statements (if any) are **definitely true**? Which are **sometimes** true? Which are **never** true? Here \perp represents the independence symbol.

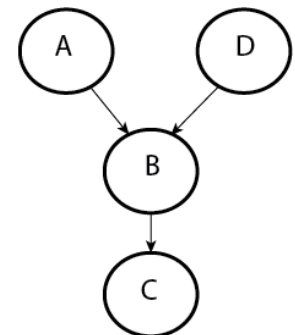
DEFINITELY SOMETIMES NEVER

- | | | | |
|-----------------------|-----------------------|-----------------------|------------------------------------------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | $P(A, B, C, D) = P(A) P(B) P(C) P(D)$ |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | $P(A, B, C, D) = P(A) P(B) P(C D)$ |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | $P(A, B, C, D) = P(A) P(B A) P(C A, B) P(D A, B, C)$ |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | $P(B, C D) = P(B) P(C D)$ |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | $P(A, C D) = P(C) P(A D)$ |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | $P(X, Y Z) = P(X Y) P(Y Z)$ |

ii. (3 pt) Draw a Bayes' Net that is consistent with the assumptions that $A \perp B$ and $B \perp C | D$. Your Bayes' Net may include additional independence assumptions other than these.

iii. (1 pt) True False: If X is independent of Y , then X is independent of Y given Z , i.e. $X \perp Y \rightarrow X \perp Y | Z$

iv. (4 pts) Given the Bayes' Net below, prove algebraically that A is independent of D . Briefly (less than five words per step) explain or justify each step of your proof. Use the lines given for the steps of your proof. You may not need all lines.

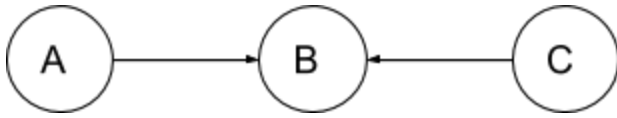


Step

Explanation

_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

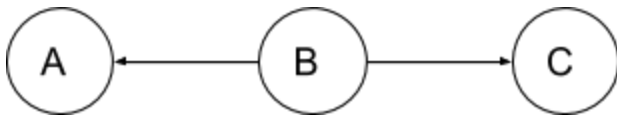
v. (2 pts) Suppose we have the Bayes' Net model below:



Which of the following statements regarding independence **must be true** given the model?

- $A \perp B$ $A \perp C \mid B$
- $A \perp C$ $A \perp B \mid C$
- $B \perp C$ $B \perp C \mid A$

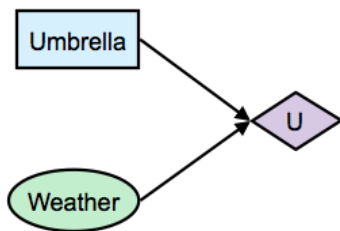
vi. (2 pts) Consider the Bayes' Net model below.



Which of the following independence assumptions **must be true** given the model?

- $A \perp B$ $A \perp C \mid B$
- $A \perp C$ $A \perp B \mid C$
- $B \perp C$ $B \perp C \mid A$

vii. (2 pts) Consider the simple model shown below:



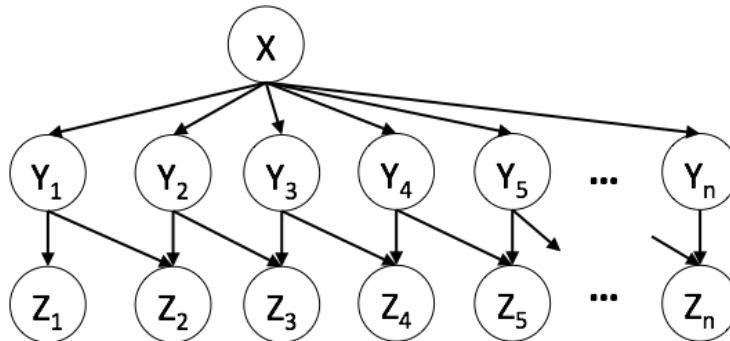
W	P(W)
rain	0.1
sun	0.9

Umbrella	W	U(A, W)
take	rain	0
take	sun	100
leave	rain	0
leave	sun	100

Trivially, we can calculate that the maximum expected utility in the absence of evidence is $MEU(\emptyset) = 90$. Suppose we want to calculate the value of knowing that it is raining $VPI(w = \text{rain})$. Calculating, we find that $MEU(\text{rain}) = 0$. This implies that the value of knowing that it is raining is -90 . In class, we said that VPI is always non-negative. Explain where this argument fails.

5. Bayes' Nets Inference (14 pts)

Consider the following Bayes' Net where all variables are binary.



i. (2 pts) Suppose we somehow calculate the distribution $P(Y_n | Z_1 = z_1, Z_2 = z_2, \dots, Z_n = z_n)$. What is the size of this factor (number of rows)? Note that the Z variables are all observed evidence!

Number of rows in this factor: _____

ii. (3 pts) Suppose we try to calculate $P(Y_n | Z_1 = z_1, Z_2 = z_2, \dots, Z_n = z_n)$ using variable elimination, and start by eliminating the variable Y_1 . What new factor is generated? **Give your answer in standard probability distribution notation**, as opposed to using notation that involves f, e.g. use $P(A, B | C)$, not $f_1(A, B, C)$.

What is the size of the probability table for this new factor (number of rows)? As before, assume all the Z variable are observed evidence.

New factor generated: _____ Size of this new factor: _____



iii. (2 pts) Suppose we decide to perform variable elimination to calculate $P(Y_n | Z_1 = z_1, Z_2 = z_2, \dots, Z_n = z_n)$. Suppose that we eliminate the variables in the order $Y_1, Y_2, \dots, Y_{n-1}, X$. What factor is generated when we finally eliminate X ? **Give your answer in standard probability distribution notation.** Note Y_n is not eliminated!

New factor generated: _____

iv. (4 pts) Find the best and worst variable elimination orderings for calculating $P(Y_n | Z_1 = z_1, Z_2 = z_2, \dots, Z_n = z_n)$. An ordering is considered better than another if the the sum of the sizes of the factors that are generated is smaller. You do not need to calculate the precise value of this sum to answer this question. If there are orderings that are tied, give just one ordering.

Best ordering: _____

Worst ordering: _____

v. (2 pts) Assume now we want to use variable elimination to calculate **a new query** $P(Z_n | Y_1, Y_2, \dots, Y_{n-1})$. Mark all of the following variables that produce a *constant* factor after being eliminated for all possible elimination orderings. Note that for this problem, the set of evidence variables is different than in previous parts.

Z_1 Z_2 Z_{n-1} X

vi. (1 pt) True False: We can simply delete any factor that will produce a constant factor from our feature set before computing any distribution.

6. The Homebody Ladybug (13 pts)

Consider a ladybug who lives on a number line where, at each step, the ladybug changes her position x by flying one step to the right (setting $x = x + 1$) or one step to the left (setting $x = x - 1$) with probabilities:

$$P_{left} = \frac{1}{2} + \frac{1}{2} \frac{x}{c+|x|}, \text{ where } c \text{ is a constant greater than } 0.$$

$$P_{right} = 1 - P_{left}$$

For example, if the constant c equals 1, the probabilities of a move to the left at positions $x = -2, -1, 0, 1, 2$ are given by $\frac{1}{6}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4},$ and $\frac{5}{6}$, respectively.

i. (1.5 pts) Draw the Bayes' Net for the ladybug's position, using X_i to represent the ladybug's position at time step i . Do not assume the ladybug starts at the origin.

ii. (1.5 pts) What is the domain of the variable X_i ? Do not assume the ladybug starts at the origin.

iii. (3 pts) Assume that at time 0 the lady bug starts at the origin ($X_0 = 0$). What is $P(X_{10} = 5 \mid X_0 = 0)$ if $c = 1$?

iv. (2 pts) You'd like to estimate the probability that the ladybug is at the origin at time $t=4$, given that you observe that the ladybug is at $+1$ in timestep $t = 1$, i.e. $P(X_4 = 0 \mid X_1 = 1)$. Consider the following possible samples for $\{X_0, \dots, X_4\}$. Put a check **in the box to the left the samples that are valid for computing this query and could have come from the ladybug model**.

$\{0, 1, 2, 3, 4\}$ $\{0, -1, 0, 1, 0\}$ $\{0, -1, 0, 2, -2\}$ $\{0, 1, 0, -1, 0\}$
 $\{0, 1, 2, 1, 0\}$ $\{0, -1, -2, -1, 0\}$ $\{0, -1, -2, -1, 0\}$ $\{0, 1, 1, 2, 0\}$

v. (1 pt) From the valid samples in part iv, what is $P(X_4 = 0 \mid X_1 = 1)$? _____

vi. (2 pts) How would you expect $P(X_4 = 0 \mid X_0 = 0)$ to change as c increases?

vii. (2 pts) Suppose you are trying to use Gibbs Sampling to estimate $P(X_4 = 0 \mid X_0 = 0)$. If we start by initializing $X_1, X_2, X_3,$ and X_4 to random 32 bit integers, will the Gibbs Sampling process **yield a reasonable estimate**? Will the sampling procedure complete in a **reasonable amount of time**? Explain your answer.

7. Ms. Pac-Man's Spicy Psychic Stake Out (18 pts)

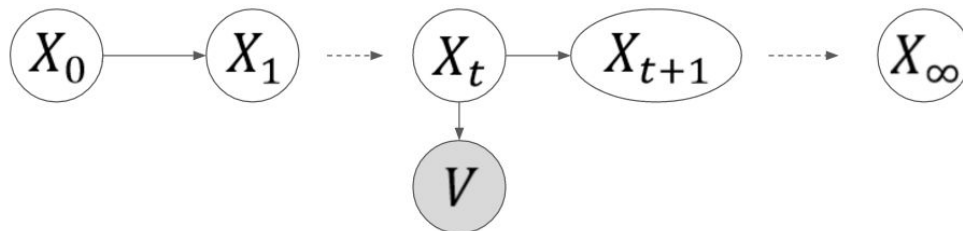
The parts of this problem are independent. **You can do any of them without doing the rest.**

i) (4 pts) Ms. Pac-Man is training herself to understand the habits of ghosts. She has captured a ghost named Clyde, and forced it to walk down a hallway to train her model. She builds the model below, and calculates the conditional probability tables (CPTs) for this model, i.e. $P(X_0)$, $P(X_1 | X_0)$, $P(X_2 | X_1)$, etc. where X_i is the location of the ghost at time i .



Give an expression for computing $P(X_i)$ in terms of these CPTs.

ii) (4 pts) Ms. Pac-Man has been experimenting with SPICE to gain psychic abilities. When she takes SPICE, she sees a psychic vision v that provides her with a belief about the ghost's future position at some specified time t , shown below. In other words, she now knows v , and thus also knows the distribution $P(v | X_t)$.



Suppose Ms. Pac-Man wants to figure out where the ghost will go next in the timestep after t . Give an expression for computing $P(X_{t+1} | v)$ in terms of distributions either given as part of her model, or calculated in part i. In total this includes the CPTs, $P(v | X_t)$, and $P(X_i)$. You may not need all of them.

iii) (6 pts) Now suppose Ms. Pac-Man wants to figure out where the ghost was the timestep **before** t given her SPICE vision. Give an expression for computing $\mathbf{P}(\mathbf{X}_{t-1} \mid \mathbf{v})$ in terms of distributions either given as part of her model, or calculated in part i or ii of Question 7. You may not need all of them.

iv) (2 pts) Suppose Ms. Pac-Man takes SPICE at time 0, which provides her a vision \mathbf{v} corresponding to time $t = 188$, i.e. she knows \mathbf{v} and $\mathbf{P}(\mathbf{v} \mid \mathbf{X}_{188})$. Explain how she would calculate Clyde's **most likely current position (i.e. at time 0)** in terms of the distributions given in parts i, ii, and iii. You may not need all of them.

v) (1 pt) True False: If we assume our model is stationary (i.e. $\mathbf{P}(\mathbf{X}_n \mid \mathbf{X}_{n-1}) = \mathbf{P}(\mathbf{X}_{n-1} \mid \mathbf{X}_{n-2})$), then the limit without evidence as time goes to infinity must be a uniform distribution., i.e. $\mathbf{P}(\mathbf{X}_\infty)$ is uniform.

vi) (1 pt) True False: If we assume our model is stationary (i.e. $\mathbf{P}(\mathbf{X}_n \mid \mathbf{X}_{n-1}) = \mathbf{P}(\mathbf{X}_{n-1} \mid \mathbf{X}_{n-2})$), then the limit with evidence as time goes to infinity must be a uniform distribution., i.e. $\mathbf{P}(\mathbf{X}_\infty \mid \mathbf{v})$ is uniform.

THIS PAGE INTENTIONALLY LEFT BLANK. WORK ON THIS PAGE WILL NOT BE GRADED.

