CS 188 Summer 2016 Introduction to Artificial Intelligence — Midterm 2

- You have approximately 2 hours and 50 minutes.
- The exam is closed book, closed calculator, and closed notes except your one-page crib sheet.
- Mark your answers ON THE EXAM ITSELF. If you are not sure of your answer you may wish to provide a *brief* explanation. All short answer sections can be successfully answered in a few sentences AT MOST.
- For multiple choice questions with *circular bubbles*, you should only mark ONE option; for those with *checkboxes*, you should mark ALL that apply (which can range from zero to all options)

First name	
Last name	
edX username	

For staff use only:

Q1.	Potpourri	/13
Q2.	More Advanced Problems	/19
Q3.	Variable Elimination	/12
Q4.	Bayes' Nets: Representation and Independence	/16
Q5.	VPI	/13
Q6.	Sampling as an MDP	/13
Q7.	HMMs	/19
	Total	/105

THIS PAGE IS INTENTIONALLY LEFT BLANK

Q1. [13 pts] Potpourri

(a) Probability

A	$P(A)$
+a	0.8
–a	0.2

A	B	$P(B\|A)$
+a	+b	0.9
+a	–b	0.1
–a	+b	0.6
–a	–b	0.4

B	C	$P(C\|B)$
+b	+c	0.8
+b	–c	0.2
–b	+c	0.8
–b	–c	0.2

C	D	$P(D\|C)$
+c	+d	0.25
+c	–d	0.75
–c	+d	0.5
–c	–d	0.5

Using the table above and the assumptions per subquestion, calculate the following probabilities given no independence assumptions. If it is impossible to calculate without more independence assumptions, specify the least number of of independence assumptions that would allow you to answer the question (don't do any computation in this case).

(i) [1 pt] $P(+a, -b) =$

(ii) [1 pt] $P(-a, -b, +c) =$

(iii) [1 pt] Now assume C is independent of A given B and D is independent of everything else given C. Calculate $P(+a, -b, +c, +d) =$ or say what other independence assumptions are necessary.

(b) Independence

(i) [2 pts] Mark all expressions which indicate that X is independent of Y given Z.

☐ $P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$

☐ $P(X \mid Y, Z) = P(X \mid Z)$

☐ $P(X, Y, Z) = P(X, Z)P(Y)$

☐ None of the above

(ii) [2 pts] Fill in the circles of **all** expressions that are equal to $\mathbf{P(R, S, T)}$, **given no independence assumptions**:

☐ $P(R \mid S, T)\ P(S \mid T)\ P(T)$

☐ $P(T \mid R, S)\ P(R)\ P(S)$

☐ $P(R \mid S)\ P(S \mid T)\ P(T)$

☐ $P(R \mid S, T)\ P(S \mid R, T)\ P(T \mid R, S)$

☐ $P(T, S \mid R)\ P(R)$

☐ $P(T \mid R, S)\ P(R, S)$

☐ None of the above

(c) Bayes Nets

(i) [1 pt] During variable elimination, the ordering of elimination does not affect the final answer.

○ True ○ False

(ii) [1 pt] During variable elimination, the ordering of elimination does not affect the runtime.

○ True ○ False

(d) [1 pt] **Sampling** For the following descriptions, provide the sampling method that is being described.

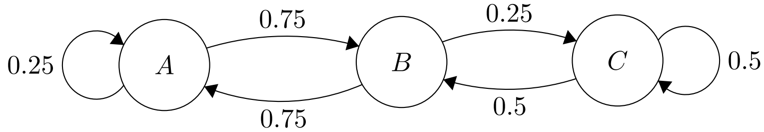
- Tally all of the values, but ignore anything that doesn't match the conditional evidence.

- Tally the values, weighting by the value of actually seeing that evidence based on the parents.

- Sample from the original joint distribution, ignoring the evidence.

(e) [3 pts] **Stationary Distributions**

Consider a Markov chain with 3 states and transition probabilities as shown below:



Compute the stationary distribution. That is, compute $P_\infty(A), P_\infty(B), P_\infty(C)$.

Q2. [19 pts] More Advanced Problems

(a) [2 pts] **Probability**

Assume that Q, R, S, and T are all independent binary random variables. For the following probabilities, assume you are given a table of all values of each probability. Write down what the sum of all of the values in the table would equal (as a number). If it is impossible to tell, write down "Impossible".

$P(+r \mid S) =$

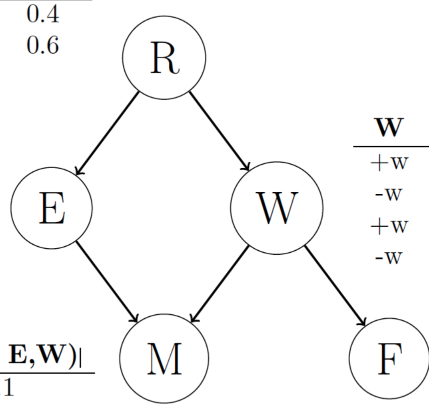
$P(R, T \mid +s) =$

$P(R \mid Q, S, T) =$

$P(R, T \mid +s, Q) =$

(b) [3 pts] **Sampling** Consider the following Bayes Net and corresponding probability tables.

R	P(R)
+r	0.4
-r	0.6



E	R	P(E | R)
+e	+r	0.3
-e	+r	0.7
+e	-r	0.6
-e	-r	0.4

W	R	P(W | R)
+w	+r	0.9
-w	+r	0.1
+w	-r	0.2
-w	-r	0.8

M	E	W	P(M | E,W)
+m	+e	+w	0.1
-m	+e	+w	0.9
+m	+e	-w	0.45
-m	+e	-w	0.55
+m	-e	+w	0.35
-m	-e	+w	0.65
+m	-e	-w	0.9
-m	-e	-w	0.1

F	W	P(F | W)
+f	+w	0.15
-f	+w	0.85
+f	-w	0.75
-f	-w	0.25

Fill in the following table with the probabilities of drawing each respective sample given that we are using each of the following sampling techniques. For rejection sampling, we say that a sample has been drawn only if it is not rejected. You may leave your answer in the form of an expression such as $.8 \cdot .4$ without multiplying it out. (Hint: $P(f, m) = .181$)

$P(sample \mid method)$	$(+r, +e, -w, +m, +f)$	$(+r, -e, +w, -m, +f)$
prior sampling		
rejection sampling		
likelihood weighting		

(c) HMM

(i) [2 pts] Consider an HMM with state variables $\{X_i\}$ and emission variables $\{Y_i\}$. Which of the following assertions are true?

☐ X_i is always conditionally independent of Y_{i+1} given X_{i+1}.

☐ There exists an HMM where X_i is conditionally independent of Y_i given X_{i+1}.

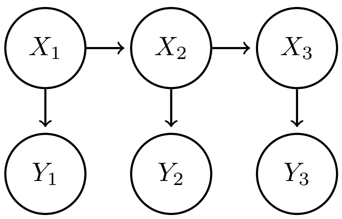
☐ If $Y_i = X_i$ with probability 1, and the state space is of size k, then the most efficient algorithm for computing $p(X_t|y_1 \cdots, y_t)$ takes $O(k)$ or less time.

☐ If we take the Bayes net below for part (ii) and reverse the vertical arrows so that we have edges from each Y_i to X_i, the result is an HMM.

☐ None of the above

(ii) [7 pts]

Likelihood weighting.



Assume each of the variables $X_1, X_2, X_3, Y_1, Y_2, Y_3$ are binary with domains $\{\pm 1\}$. Assuming a uniform starting distribution $[.5, .5]$, and emission probabilities all equal to:

y	$P_E(y|-1)$	$P_E(y|1)$
-1	.2	.7
1	.8	.3

And transition probabilities all equal to:

x	$P_T(x|-1)$	$P_T(x|1)$
-1	.4	.6
1	.6	.4

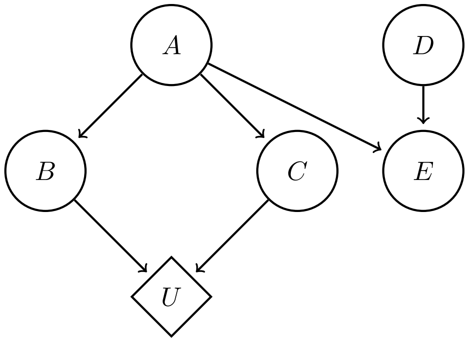
Assume that the samples are $(X_1, X_2, X_3, Y_1, Y_2, Y_3)$. Fill the following table with the samples' likelihood sampling weight (conditioning on $X_2 = 1$ and $Y_3 = 1$) and the probability of drawing the sample during likelihood weighting (You can leave the desired values as products). If a sample is invalid, say so.

Index	Sample	Weight in Likelihood Sampling	Probability of sample $P(x_1, x_2, x_3, y_1, y_2, y_3)$
1.	(1,1,1,1,1,1)		
2.	(1,1,-1,1,-1,1)		
3.	(-1,1,-1,1,-1,1)		
4.	(1,-1,1,1,-1,-1)		
5.	(1,-1,-1,-1,1,1)		

What is $P(A = 1|B = 1, F = 1)$?

Using Likelihood sampling what is $\hat{P}(A = 1|B = 1, F = 1)$?

(d) VPI Consider a decision network with the following structure, where node U is the utility:



(i) [3 pts] For each of the following, choose the most specific option that is guaranteed to be true:

○ $VPI(B) = 0$ ○ $VPI(B) \geq 0$ ○ $VPI(B) > 0$

○ $VPI(D) = 0$ ○ $VPI(D) \geq 0$ ○ $VPI(D) > 0$

○ $VPI(E) = 0$ ○ $VPI(E) \geq 0$ ○ $VPI(E) > 0$

○ $VPI(A|E) = 0$ ○ $VPI(A|E) \geq 0$ ○ $VPI(A|E) > 0$

○ $VPI(E|A) = 0$ ○ $VPI(E|A) \geq 0$ ○ $VPI(E|A) > 0$

○ $VPI(A|B,C) = 0$ ○ $VPI(A|B,C) \geq 0$ ○ $VPI(A|B,C) > 0$

(ii) [2 pts] For each of the following, fill in the blank with the most specific of $>, \geq, <, \leq, =$ to guarantee that the comparison is true, or write ? if there is no possible guarantee.

$VPI(B)$ ________ $VPI(A)$

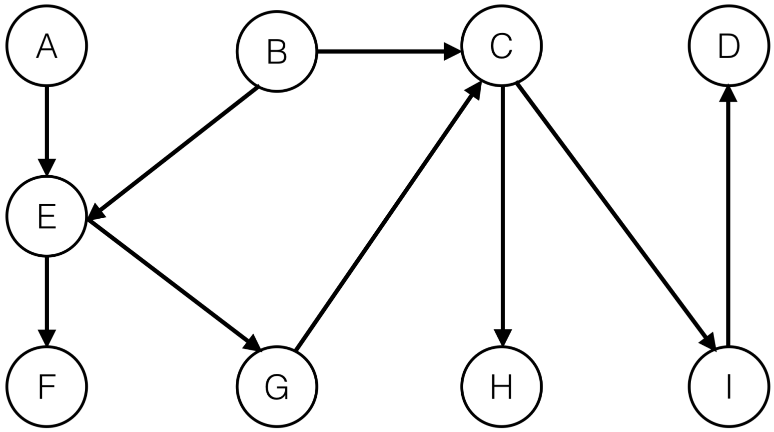
$VPI(B,C)$ ________ $VPI(A)$

$VPI(B,C)$ ________ $VPI(B) + VPI(C)$

$VPI(B|C)$ ________ $VPI(A|C)$

Q3. [12 pts] Variable Elimination

The following questions use the Bayes' net below. All variables have binary domains:



(a) [6 pts] Karthik wants to see the ocean, and so decides to compute the query $P(B, E, A, C, H)$. He wants you to help him run variable elimination to compute the answer, with the following elimination ordering: I, D, G, F. Complete the following description of the factors generated in this process:

He initially has the following factors to start out with:

$$P(A), P(B), P(C|B,G), P(D|I), P(E|A,B), P(F|E), P(G|E), P(H|C), P(I|C)$$

When eliminating I we generate a new factor f_1 as follows:

$$f_1(C, D) = \sum_i P(i|C)P(D|i)$$

This leaves us with the factors:

$$P(A), P(B), P(C|B,G), P(E|A,B), P(F|E), P(G|E), P(H|C), f_1(C,D)$$

When eliminating D we generate a new factor f_2 as follows:

This leaves us with the factors:

When eliminating G we generate a new factor f_3 as follows:

This leaves us with the factors:

When eliminating F we generate a new factor f_4 as follows:

This leaves us with the following factors. Another acceptable answer involved noting the fact that summing out the above factor yields 1, and so not appending $f_4(E)$ was fine.

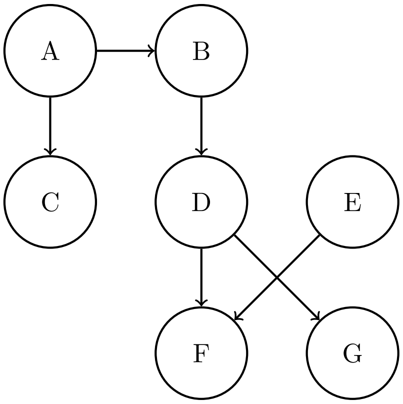
(b) [2 pts] Among f_1, f_2, f_3, f_4, which is the largest factor generated, and how large is it? Assume all variables have binary domains and measure the size of each factor by the number of rows in the table that would represent the factor.

(c) [4 pts] Given a list of all factors in a Bayes net, suppose that there exists a variable V that only occurs once in the entire list. Which of the following statements must be true when running variable elimination?

- ☐ The factor containing variable V must have precisely 2 variables.
- ☐ Eliminating V produces a factor whose size is lesser than or equal to the largest factor size during the full variable elimination process.
- ☐ Variable V must be a leaf node; that is, V cannot have any children nodes.
- ☐ The factor containing variable V must contain an even number of variables.
- ☐ The factor containing variable V must contain an odd number of variables.
- ☐ Variable V must appear on the left hand side of the conditioning bar, i.e. the |, in the factor that it appears in.
- ☐ There must also exist a different variable W that appears only once in the entire list of factors.
- ☐ There must also exist a different variable W that appears more than once in the entire list of factors.
- ☐ None of the above

Q4. [16 pts] Bayes' Nets: Representation and Independence

Parts (a), (b), and (c) pertain to the following Bayes' Net.



(a) [1 pt] Express the joint probability distribution as a product of terms representing individual conditional probabilities tables associated with the Bayes Net.

(b) [1 pt] Assume each node can take on 4 values. How many entries do the factors at A, D, and F have?

A: ________________________________

D: ________________________________

F: ________________________________

(c) [2 pts] Mark the statements that are guaranteed to be true.

- ☐ $B \perp\!\!\!\perp C$
- ☐ $A \perp\!\!\!\perp F$
- ☐ $D \perp\!\!\!\perp E|F$
- ☐ $E \perp\!\!\!\perp A|D$
- ☐ $F \perp\!\!\!\perp G|D$
- ☐ $B \perp\!\!\!\perp F|D$
- ☐ $C \perp\!\!\!\perp G$
- ☐ $D \perp\!\!\!\perp E$

Parts (d) and (e) pertain to the following probability distribution tables. The joint distribution $P(A, B, C, D)$ is equal to the product of these probability distribution tables.

A	$P(A)$
+a	0.8
−a	0.2

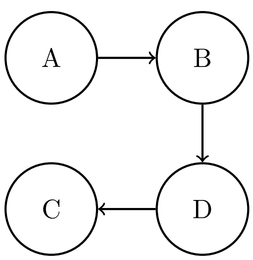
A	B	$P(B\|A)$
+a	+b	0.9
+a	−b	0.1
−a	+b	0.6
−a	−b	0.4

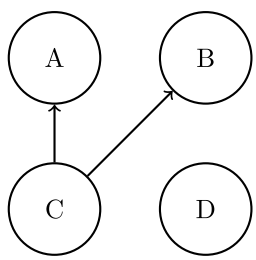
B	C	$P(C\|B)$
+b	+c	0.8
+b	−c	0.2
−b	+c	0.8
−b	−c	0.2

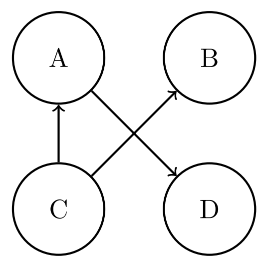
C	D	$P(D\|C)$
+c	+d	0.25
+c	−d	0.75
−c	+d	0.5
−c	−d	0.5

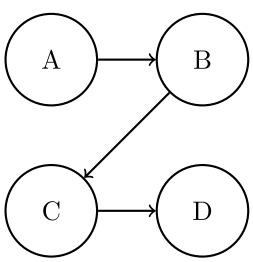
(d) [2 pts] State all non-conditional independence assumptions that are implied by the probability distribution tables.

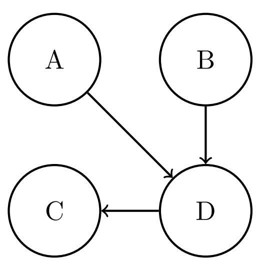
(e) [3 pts] Circle all the Bayes net(s) that can represent a distribution that is consistent with the tables given.

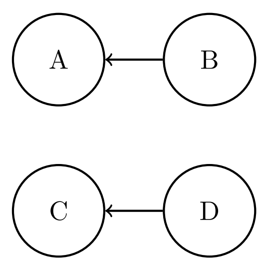










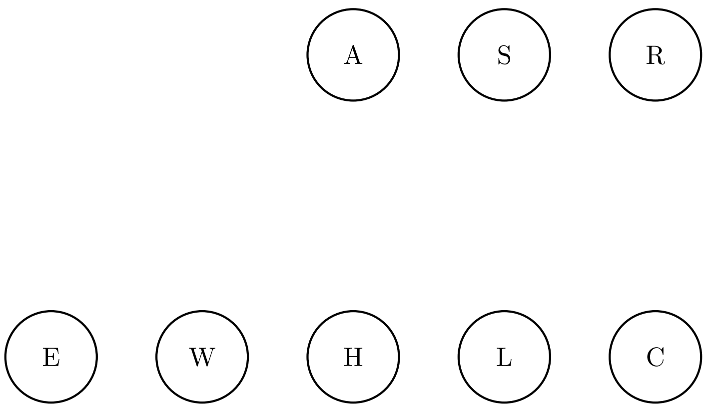


You are building advanced safety features for cars that can warn a driver if they are falling asleep (A) and also calculate the probability of a crash (C) in real time. You have at your disposal 6 sensors (random variables):

- E: whether the driver's eyes are open or closed
- W: whether the steering wheel is being touched or not
- L: whether the car is in the lane or not
- S: whether the car is speeding or not
- H: whether the driver's heart rate is somewhat elevated or resting
- R: whether the car radar detects a close object or not

A influences $\{E, W, H, L, C\}$. C is influenced by $\{A, S, L, R\}$.

(f) [2 pts] Draw the Bayes Net associated with the description above by adding edges between the provided nodes where appropriate.



(g) [2 pts] Mark all the independence assumptions that must be true.

☐ $E \perp\!\!\!\perp S$	☐ $L \perp\!\!\!\perp R \mid C$
☐ $W \perp\!\!\!\perp H \mid A$	☐ $W \perp\!\!\!\perp R$
☐ $S \perp\!\!\!\perp R$	☐ $A \perp\!\!\!\perp C$
☐ $E \perp\!\!\!\perp L$	☐ $E \perp\!\!\!\perp C \mid L$

(h) [2 pts] The car's sensors tell you that the car is in the lane ($L = +l$) and that the car is not speeding ($S = -s$). Now you would like to calculate the probability of crashing, $P(C \mid +l, -s)$. We will use the variable elimination ordering R, A, E, W, H. Write down the largest factor generated during variable elimination. Box your answer.

(i) [1 pt] Write down a more efficient variable elimination ordering, i.e. one whose largest factor is smaller than the one generated in the previous question.

Q5. [13 pts] VPI

You are the latest contestant on Monty Hall's game show, which has undergone a few changes over the years.
In the game, there are n closed doors: behind one door is a car ($U(car) = 1000$), while the other $n-1$ doors each have a goat behind them ($U(goat) = 10$). You are permitted to open exactly one door and claim the prize behind it.

You begin by choosing a door uniformly at random.

(a) [2 pts] What is your expected utility?

Answer:

(b) [4 pts] After you choose a door but before you open it, Monty offers to open k other doors, each of which are guaranteed to have a goat behind it.
If you accept this offer, should you keep your original choice of a door, or switch to a new door?

$EU(keep)$:

$EU(switch)$:

Action that achieves MEU:

(c) [2 pts] What is the value of the information that Monty is offering you?

Answer:

(d) [2 pts] Monty is changing his offer!
After you choose your initial door, you are given the offer to choose any other door and open this second door. If you do, after you see what is inside the other door, you may switch your initial choice (to the newly opened door) or keep your initial choice.
What is the value of this new offer?

Answer:

(e) [3 pts] Monty is generalizing his offer: you can pay $\$d^3$ to open d doors as in the previous part. (Assume that $U(\$x) = x$.) You may now switch your choice to any of the open doors (or keep your initial choice). What is the largest value of d for which it would be rational to accept the offer?

Answer:

Q6. [13 pts] Sampling as an MDP

(a) **(i)** [1 pt] You are given a Bayes net with binary random variables A, B, C, D, and E. You want to estimate $P(A, B, C, E| + d)$ using rejection sampling. Which of the following quantities denotes the probability that a sample will be *rejected*? (Mark all that apply.)

- ☐ $P(+d)$
- ☐ $P(-d)$
- ☐ $1 - P(+d)$
- ☐ $1 - P(-d)$

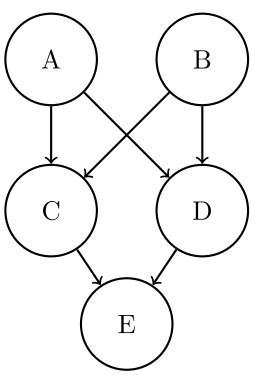
(ii) [1 pt] For the same Bayes net, you would like to estimate $P(A, B, C| + d, +e)$ using rejection sampling. Which of the following quantities denotes the probability that a sample will be *rejected*? (Mark all that apply.)

- ☐ $P(+d, +e)$
- ☐ $P(-d, -e)$
- ☐ $1 - P(+d, +e)$
- ☐ $1 - P(-d, -e)$
- ☐ $1 - P(+d)P(+e)$
- ☐ $1 - P(-d)P(-e)$

(iii) [1 pt] For the same Bayes net, suppose additionally that $D \perp\!\!\!\perp E$. Which of the following quantities denotes the probability that a sample will be *rejected*?

- ☐ $P(+d, +e)$
- ☐ $P(-d, -e)$
- ☐ $1 - P(+d, +e)$
- ☐ $1 - P(-d, -e)$
- ☐ $1 - P(+d)P(+e)$
- ☐ $1 - P(-d)P(-e)$

(b) [2 pts] Use the following Bayes Net for this question *only*:



In how many different orders could I sample from the random variables in this Bayes Net? (You may use simple arithmetic operations in your answer.)

(c) **(i)** [1 pt] In a general Bayes net over N random variables, what is the largest possible number of orderings in which I could sample?

(ii) [1 pt] What is the smallest possible number of orderings in which I could sample?

(d) Recall that rejection sampling is most efficient when we reject as *early* as possible. In general, it might be hard to determine which sample ordering will make this possible. We'd like to formulate the problem as an MDP, and use policy iteration to select an optimal ordering.

(i) [1 pt] The state space of this MDP will either be some collection of random variables, or (variable, value) pairs. More specifically, which of the following is an appropriate minimal state representation for this MDP? (Mark one.) Hint: it may be helpful to refer to the transition function described below.

- ○ Set of variables that have been sampled so far (e.g. $\{A, B, D, \cdots\}$).
- ○ Set of (variable, value) pairs that have been sampled so far (e.g. $\{(A, +a), (B, -b), (D, +d), \cdots\}$).
- ○ Ordered list of variables pairs that have been sampled so far (e.g. $\{[A, B, D, \cdots]\}$).
- ○ Ordered list of (variable, value) pairs that have sampled so far (e.g. $\{[(A, +a), (B, -b), (D, +d), \cdots]\}$).

(ii) [1 pt] If the Bayes net has N binary random variables, how big is this state space? (Choose the tightest upper bound out of the answers given.)

- ○ $O(n)$
- ○ $O(2^n)$
- ○ $O(3^n)$
- ○ $O(n!)$
- ○ $O(2^n n!)$
- ○ $O(3^n n!)$

The action space and transition function of this MDP are as follows: Every random variable corresponds to an action. When we select a random variable, we sample a value from the corresponding distribution. If this value causes the sampler to reject, we immediately transition to a terminal "sink state". Otherwise, we add the variable (or (variable, value) pair) to the collection chosen above.

(iii) [2 pts] If $\gamma = 0.5$, which of the following is an appropriate reward? Recall that we want to reward the sampler for rejecting as quickly as possible. (Mark all that apply.)

- ☐ -1 per turn, and 0 if the sample is rejected
- ☐ 1 per turn, and 0 if the sample is rejected
- ☐ 0 per turn, and 1 if the sample is rejected
- ☐ 0 per turn, and -1 if the sample is rejected

(iv) [2 pts] If $\gamma = 1.0$, which of the following is an appropriate reward? (Mark all that apply.)

- ☐ -1 per turn, and 0 if the sample is rejected
- ☐ 1 per turn, and 0 if the sample is rejected
- ☐ 0 per turn, and 1 if the sample is rejected
- ☐ 0 per turn, and -1 if the sample is rejected

Q7. [19 pts] HMMs

Consider a process where there are transitions among a finite set of states $s_1, \cdots, s_k$ over time steps $i = 1, \cdots, N$. Let the random variables $X_1, \cdots, X_N$ represent the state of the system at each time step and be generated as follows:

- Sample the initial state s from an initial distribution $P_1(X_1)$, and set $i = 1$
- Repeat the following:
  1. Sample a duration d from a duration distribution P_D over the integers $\{1, \cdots, M\}$, where M is the maximum duration.
  2. Remain in the current state s for the next d time steps, i.e., set

$$x_i = x_{i+1} = \cdots = x_{i+d-1} = s \tag{1}$$

  3. Sample a successor state s' from a transition distribution $P_T(X_t|X_{t-1} = s)$ over the other states $s' \neq s$ (so there are no self transitions)
  4. Assign $i = i + d$ and $s = s'$.

This process continues indefinitely, but we only observe the first N time steps.

(a) [2 pts] Assuming that all three states s_1, s_2, s_3 are different, what is the probability of the sample sequence $s_1, s_1, s_2, s_2, s_2, s_3, s_3$? Write an algebraic expression. Assume $M \geq 3$.

At each time step i we observe a noisy version of the state X_i that we denote Y_i and is produced via a conditional distribution $P_E(Y_i|X_i)$.

(b) [1 pt] Only in this subquestion assume that $N > M$. Let $X_1, \cdots, X_N$ and $Y_1, \cdots, Y_N$ random variables defined as above. What is the maximum index $i \leq N - 1$ so that $X_1 \perp\!\!\!\perp X_N | X_i, X_{i+1}, \cdots, X_{N-1}$ is guaranteed?

(c) [3 pts] Only in this subquestion, assume the max duration $M = 2$, and P_D uniform over $\{1, 2\}$ and each x_i is in an alphabet $\{a, b\}$. For $(X_1, X_2, X_3, X_4, X_5, Y_1, Y_2, Y_3, Y_4, Y_5)$ draw a Bayes Net over these 10 random variables with the property that removing any of the edges would yield a Bayes net inconsistent with the given distribution.

(d) In this part we will explore how to write the described process as an HMM with an extended state space. Write the states $z = (s, t)$ where s is a state of the original system and t represents the time elapsed in that state. For example, the state sequence $s_1, s_1, s_1, s_2, s_3, s_3$ would be represented as $(s_1, 1), (s_1, 2), (s_1, 3), (s_2, 1), (s_3, 1), (s_3, 2)$.

Answer all of the following in terms of the parameters $P_1(X_1), P_D(d), P_T(X_{j+1}|X_j), P_E(Y_i|X_i), k$ (total number of possible states), N and M (max duration).

(i) [1 pt] What is $P(Z_1)$?

$P(x_1, t_1) =$

(ii) [3 pts] What is $P(Z_{i+1}|Z_i)$? Hint: You will need to break this into cases where the transition function will behave differently.

$P(X_{i+1}, t_{i+1} \mid X_i, t_i) =$

(iii) [1 pt] What is $P(Y_i|Z_i)$?

$P(Y_i \mid X_i, t_i) =$

(e) In this question we explore how to write an algorithm to compute $P(X_N|y_1, \cdots, y_N)$ using the particular structure of this process.

Write $P(X_t|y_1, \cdots, y_{t-1})$ in terms of other factors. Construct an answer by checking the correct boxes below:

$$P(X_t|y_1, \cdots, y_{t-1}) = \underline{\quad\textbf{(i)}\quad} \; \underline{\quad\textbf{(ii)}\quad} \; \underline{\quad\textbf{(iii)}\quad}$$

(i) [1 pt]

- ◯ $\sum_{i=1}^{k} \sum_{d=1}^{M} \sum_{d'=1}^{M}$
- ◯ $\sum_{i=1}^{k}$
- ◯ $\sum_{i=1}^{k} \sum_{d=1}^{M}$
- ◯ $\sum_{d=1}^{M}$

(ii) [1 pt]

- ◯ $P(Z_t = (X_t, d)|Z_{t-1} = (s_i, d))$
- ◯ $P(X_t|X_{t-1} = s_d)$
- ◯ $P(X_t|X_{t-1} = s_i)$
- ◯ $P(Z_t = (X_t, d')|Z_{t-1} = (s_i, d))$

(iii) [1 pt]

- ◯ $P(Z_{t-1} = (s_d, i)|y_1, \cdots, y_{t-1})$
- ◯ $P(Z_{t-1} = (s_i, d)|y_1, \cdots, y_{t-1})$
- ◯ $P(X_{t-1} = s_d|y_1, \cdots, y_{t-1})$
- ◯ $P(X_{t-1} = s_i|y_1, \cdots, y_{t-1})$

(iv) [1 pt] Now we would like to include the evidence y_t in the picture. What would be the running time of each update of the **whole table** $P(X_t|y_1, \cdots, y_t)$?. Assume tables corresponding to any factors used in **(i)**, **(ii)**, **(iii)** have already been computed.

- ◯ $O(k^2)$
- ◯ $O(k^2M^2)$
- ◯ $O(k^2M)$
- ◯ $O(kM)$

Note: Computing $P(X_N|y_1, \cdots, y_N)$ will take time $N\times$ your answer in **(iv)**.

(v) [4 pts] Describe an update rule to compute $P(X_t|y_1, \cdots, y_{t-1})$ that is faster than the one you discovered in parts **(i)**, **(ii)**, **(iii)**. **Specify its running time**. Hint: Use the structure of the transitions $Z_{t-1} \to Z_t$.

THIS PAGE IS INTENTIONALLY LEFT BLANK