CS 188 Fall 2017 — Introduction to Artificial Intelligence

Midterm 2

- You have approximately 110 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,
  - ☐ means mark **all options** that apply
  - ○ means mark a **single choice**
  - When selecting an answer, please fill in the bubble or square **completely** (● and ■)

First name	
Last name	
SID	
Student to your right	
Student to your left	

Your Discussion/Exam Prep* TA (fill all that apply):

☐ Brijen (Tu)	☐ Aaron (W)	☐ Aarash (W)	☐ Shea* (W)
☐ Peter (Tu)	☐ Mitchell (W)	☐ Daniel (W)	☐ Daniel* (W)
☐ David (Tu)	☐ Abhishek (W)	☐ Yuchen* (Tu)	
☐ Nipun (Tu)	☐ Caryn (W)	☐ Andy* (Tu)	
☐ Wenjing (Tu)	☐ Anwar (W)	☐ Nikita* (Tu)	

For staff use only:

Q1.	Probability Warm-Up	/9
Q2.	Dressed to Impress	/12
Q3.	Independence	/12
Q4.	Markov Model Jambalaya	/12
Q5.	Bayes Nets: Elimination	/16
Q6.	Bayes Nets: Sampling	/16
Q7.	Mini Forward	/10
Q8.	HMM: Human-Robot Interaction	/8
Q9.	A Not So Random Walk	/5
	Total	/100

THIS PAGE IS INTENTIONALLY LEFT BLANK

Q1. [9 pts] Probability Warm-Up

(a) [4 pts] You have three coins in your pocket:

- Coin 1 is a fair coin that comes up heads with probability 1/2.
- Coin 2 is a biased coin that comes up heads with probability 1/4.
- Coin 3 is a biased coin that comes up heads with probability 3/4.

Suppose you pick one of the coins uniformly at random and flip it three times. If you observe the sequence HHT (where H stands for heads and T stands for tails), what is the probability that you chose Coin 3?

Let C_i denote the event that coin i was chosen. Using Bayes' rule, we have that

$$P(C_3 \mid HHT) = \frac{P(HHT \mid C_3)P(C_3)}{P(HHT)} = \frac{P(HHT \mid C_3)P(C_3)}{\sum_{i=1}^{3} P(HHT \mid C_i)P(C_i)} = \frac{P(HHT \mid C_3)}{\sum_{i=1}^{3} P(HHT \mid C_i)},$$

where the final equality follows from the fact that $P(C_i) = 1/3$ for each i. Substituting in the values from the problem, we obtain

$$P(C_3 \mid HHT) = \frac{\frac{3}{4}\cdot\frac{3}{4}\cdot\frac{1}{4}}{\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{4}\cdot\frac{3}{4} + \frac{3}{4}\cdot\frac{3}{4}\cdot\frac{1}{4}} = \frac{\frac{9}{64}}{\frac{1}{8} + \frac{3}{64} + \frac{9}{64}} = \frac{9}{20}.$$

$P(C_3 \mid HHT) =$ ________

(b) [5 pts] Suppose X and Y are **independent** random variables over the domain $\{1, 2, 3\}$ with $P(X = 3) = 1/6$. Given the following partially specified joint distribution, what are the remaining values? Write your answers as simplified fractions in the blanks.

$P(X = 1, Y = 1) = 1/4$	$P(X = 2, Y = 1) = 1/6$	$P(X = 3, Y = 1) =$ 1/12
$P(X = 1, Y = 2) = 1/16$	$P(X = 2, Y = 2) = 1/24$	$P(X = 3, Y = 2) =$ 1/48
$P(X = 1, Y = 3) =$ 3/16	$P(X = 2, Y = 3) =$ 1/8	$P(X = 3, Y = 3) =$ 1/16

Since X and Y are independent, we have that $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all x and y, so it suffices to determine the marginal distributions $P(X)$ and $P(Y)$.

We begin with $P(X)$. First observe that

$$\frac{P(X = 1, Y = 1)}{P(X = 2, Y = 1)} = \frac{P(X = 1)P(Y = 1)}{P(X = 2)P(Y = 1)} = \frac{P(X = 1)}{P(X = 2)} = \frac{1/4}{1/6} = \frac{3}{2}.$$

Combining this with the fact that any probability distribution sums to 1, we find that

$$P(X = 1) + P(X = 2) + P(X = 3) = \frac{3}{2}\cdot P(X = 2) + P(X = 2) + \frac{1}{6} = \frac{5}{2}\cdot P(X = 2) + \frac{1}{6} = 1,$$

which implies $P(X = 2) = 1/3$. It follows that $P(X = 1) = (3/2)\cdot P(X = 2) = 1/2$.

To recover the marginal distribution of Y, we note that

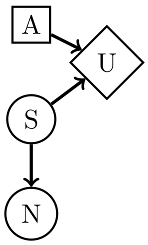
$$P(X = 1, Y = 1) = P(X = 1)P(Y = 1) = (1/2)\cdot P(Y = 1) = 1/4,$$
$$P(X = 1, Y = 2) = P(X = 1)P(Y = 2) = (1/2)\cdot P(Y = 2) = 1/16,$$

so $P(Y = 1) = 1/2$ and $P(Y = 2) = 1/8$. It follows that $P(Y = 3) = 1 - P(Y = 1) - P(Y = 2) = 3/8$.

Q2. [12 pts] Dressed to Impress

(a) [12 pts] Alice is invited to a party tonight which is said to be once-in-a-lifetime. However, this mysterious party doesn't publicize who is going and thus Alice has no idea whether the size S of the party will be large ($S = +s$) or tiny ($S = -s$). The size can affect the noise N outside the party, and it will either be noisy ($N = +n$) or quiet ($N = -n$). Alice has three dresses: blue, red and yellow. Each dress will have a different utility for her depending on the size of the party. Let's help her decide which will be best!

We have the following decision network, where circles are chance nodes, squares are decision nodes, and diamonds are utility nodes:



S	$P(S)$
+s	0.5
-s	0.5

S	N	$P(N \mid S)$
+s	+n	0.7
+s	-n	0.3
-s	+n	0.1
-s	-n	0.9

S	A	U
+s	blue	80
-s	blue	60
+s	red	40
-s	red	100
+s	yellow	60
-s	yellow	40

(i) [4 pts] What is the expected utility of wearing each dress, with both S and N unobserved?

- EU(A=blue) = $80 * 0.5 + 60 * 0.5 = 70$
- EU(A=red) = $40 * 0.5 + 100 * 0.5 = 70$
- EU(A=yellow) = $60 * 0.5 + 40 * 0.5 = 50$

What is Alice's maximum expected utility?

- MEU({}) = 70

(ii) [6 pts] Suppose Alice observes that the party is quiet, $N = -n$. Compute the following conditional probabilities with this observation:

- $P(+s|-n) = \frac{P(-n|+s)P(+s)}{P(-n|+s)P(+s)+P(-n|-s)P(-s)} = \frac{0.3 \cdot 0.5}{0.3 \cdot 0.5 + 0.9 \cdot 0.5} = 0.25$
- $P(-s|-n) =$ $P(-s|-n) = 1$ - $P(+s|-n) = 0.75$

What is the expected utility of wearing each dress?

- EU(A=blue $\mid N = -n$) = $80 * 0.25 + 60 * 0.75 = \frac{145}{2} = 65$
- EU(A=red $\mid N = -n$) = $40 * 0.25 + 100 * 0.75 = \frac{155}{2} = 85$
- EU(A=yellow $\mid N = -n$) = $60 * 0.25 + 40 * 0.75 = 45$

What is Alice's maximum expected utility given that $N = -n$?

- MEU({N=-n}) = 85

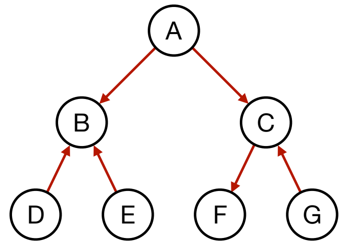
(iii) [2 pts] Construct a formula for $VPI(N)$ for the given network. To decouple this problem from your work above, use any of the symbolic terms from the following list (rather than plugging in numeric values):
$P(+n|+s)$, $P(+n|-s)$, $P(-n|+s)$, $P(-n|-s)$, $P(+n)$, $P(-n)$, $P(+s)$, $P(-s)$, $MEU(\{\})$, $MEU(\{N = +n\})$, $MEU(\{N = -n\})$

- $VPI(N) = P(+n)MEU(\{N = +n\}) + P(-n)MEU(\{N = -n\}) - MEU(\{\})$

Q3. [12 pts] Independence

In each part of this question, you are given a Bayes' net where the edges do not have a direction. Assign a direction to every edge (by adding an arrowhead at one end of each edge) to ensure that the Bayes' Net structure implies the assumptions provided. You cannot add new edges. The Bayes' nets can imply more assumptions than listed, but they *must* imply the ones listed. There may be more than one correct solution.

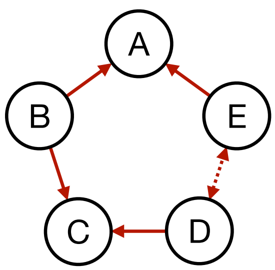
(a) [4 pts]



Assumptions:

- $A \perp\!\!\!\perp G$
- $D \perp\!\!\!\perp E$
- $E \perp\!\!\!\perp F$
- $F \perp\!\!\!\perp G \mid C$

(b) [4 pts]

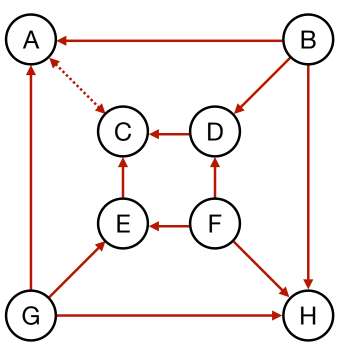


(either direction is allowed for the edge DE)

Assumptions:

- $B \perp\!\!\!\perp E$
- $E \perp\!\!\!\perp C \mid D$

(c) [4 pts]



(either direction is allowed for the edge AC)

Assumptions:

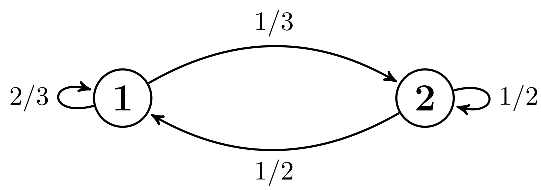
- $F \perp\!\!\!\perp G$
- $F \perp\!\!\!\perp B \mid G$
- $D \perp\!\!\!\perp E \mid F$

In order to have two nodes be independent, there must be an inactive triple along *all* paths between the two nodes.

1. $F \perp\!\!\!\perp G$, so the path FEG must have $F \to E \leftarrow G$
2. $F \perp\!\!\!\perp G$, so the path FHG must have $F \to H \leftarrow G$
3. $F \perp\!\!\!\perp B \mid G$, so the path FDB must have $F \to D \leftarrow B$ (we must later verify that G is not a descendant of D, but there is are no other edge directions along this path that will create an inactive triple)
4. $F \perp\!\!\!\perp B \mid G$, so the path FHB must have $F \to H \leftarrow B$ (we must later verify that G is not a descendant of H)
5. $D \perp\!\!\!\perp E \mid F$, so the path DCE must have $D \to C \leftarrow E$ (we must later verify that F is not a descendant of C)
6. $D \perp\!\!\!\perp E \mid F$, and we have already assigned some edge directions along path $DBAGE$. In particular, we have $D \leftarrow B - A - G \to E$. The only possible inactive triple we can create here is $B \to A \leftarrow G$ (we must later verify that F is not a descendant of A).
7. The only remaining edge to assign is $A - C$. We can assign either direction to this edge, and then verify that all required assumptions hold for the completed Bayes Net.

Q4. [12 pts] Markov Model Jambalaya

(a) [8 pts] Consider a Markov chain for X specified by the following transition diagram. Please express all final answers as simplified fractions.



(i) [4 pts] Given that $X_0 = 1$, find $P(X_1)$ and $P(X_2)$.

$P(X_1 = 1) =$ ___2/3___

$P(X_1 = 2) =$ ___1/3___

$P(X_2 = 1) =$ ___11/18___

$P(X_2 = 2) =$ ___7/18___

$P(X_2 = 1) = P(X_1 = 1)P(X_2 = 1|X_1 = 1) + P(X_1 = 2)P(X_2 = 1|X_1 = 2) = 2/3*2/3 + 1/3*1/2 = 11/18$
$P(X_2 = 2) = P(X_1 = 1)P(X_2 = 2|X_1 = 1) + P(X_1 = 2)P(X_2 = 2|X_1 = 2) = 2/3*1/3 + 1/3*1/2 = 7/18$

(ii) [4 pts] Find $P(X_\infty)$, the stationary distribution of our Markov Chain.

$P(X_\infty = 1) =$ ___3/5___

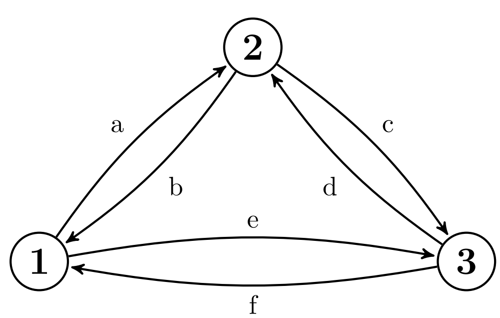
$P(X_\infty = 2) =$ ___2/5___

$$P(X_\infty = 1) = P(X_{\infty-1} = 2)P(X_\infty = 1|X_{\infty-1} = 2) + P(X_{\infty-1} = 1)P(X_\infty = 1|X_{\infty-1} = 1)$$
$$P(X_\infty = 1) = P(X_\infty = 2)P(X_i = 1|X_{i-1} = 2) + P(X_\infty = 1)P(X_i = 1|X_{i-1} = 1)$$

$$\frac{1}{3}P(X_\infty = 1) = \frac{1}{2}P(X_\infty = 2)$$
$$P(X_\infty = 1) + P(X_\infty = 2) = 1$$

$$P(X_\infty = 1) = 3/5$$
$$P(X_\infty = 2) = 2/5$$

(b) [4 pts] Consider a Markov chain for Y specified by the following transition diagram

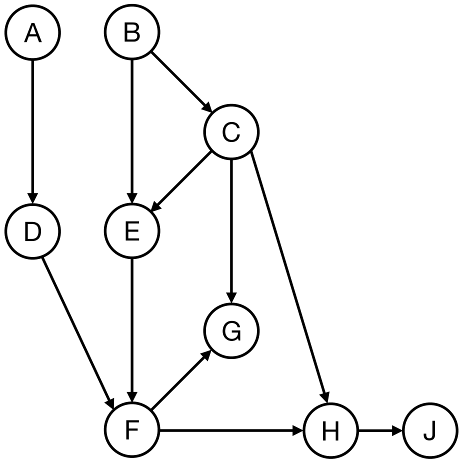


Given that $Y_0 = 1$, find $P(Y_1 = 3)$, $P(Y_2 = 3)$, $P(Y_3 = 3)$.
This problem is simplified if you just look at the possible paths from start state.

$$P(Y_1 = 3) = \underline{\quad e \quad}$$
$$P(Y_2 = 3) = \underline{\quad ac \quad}$$
$$P(Y_3 = 3) = \underline{\quad abe + efe + edc \quad}$$

Q5. [16 pts] Bayes Nets: Elimination

(a) [5 pts] Consider running variable elimination on the Bayes Net shown below.



First, we eliminate D to create a factor f_1
Next, we eliminate E to create a factor f_2
Next, we eliminate H to create a factor f_3

From the list below, select all factors that remain after D, E and H have been eliminated.

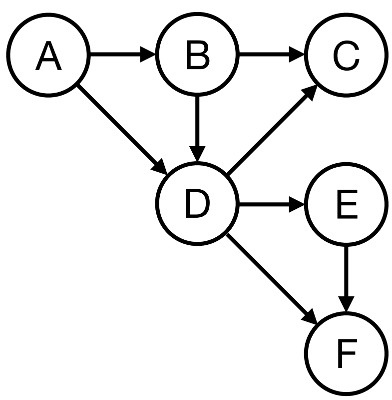
- [ ] f_1
- [x] f_2
- [x] f_3

- [x] $P(A)$
- [ ] $P(A|F)$
- [x] $P(B)$
- [ ] $P(B|A)$
- [ ] $P(B|C)$
- [ ] $P(C)$
- [x] $P(C|B)$

- [ ] $P(D|A)$
- [ ] $P(D|A,F)$
- [ ] $P(E|B)$
- [ ] $P(E|C)$
- [ ] $P(E|B,C)$
- [ ] $P(F)$
- [ ] $P(F|A,B)$
- [ ] $P(F|G)$

- [ ] $P(G|C)$
- [x] $P(G|C,F)$
- [ ] $P(H|F)$
- [ ] $P(H|C,F)$
- [ ] $P(H|J)$
- [ ] $P(J|C)$
- [ ] $P(J|F)$
- [ ] $P(J|H)$

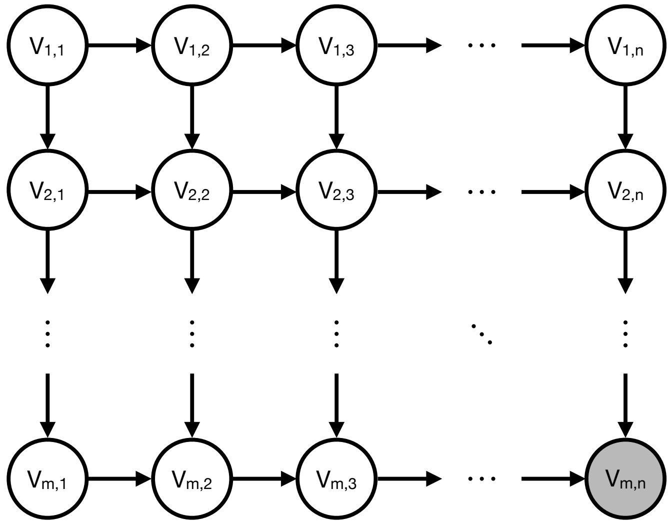
(b) [4 pts] Consider the Bayes Net shown below. Each variable in the Bayes Net can take on two possible values.



You are given the query $P(C|F)$, which you would like to answer using variable elimination. Please find a variable elimination ordering where the largest intermediate factor created during variable elimination is as small as possible.

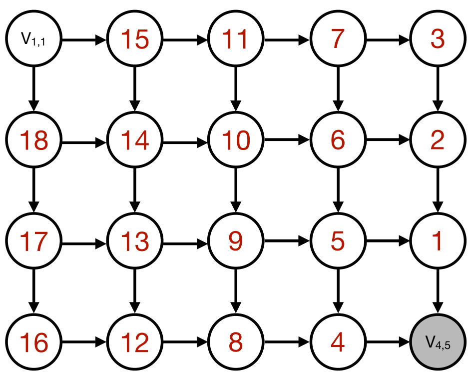
Elimination ordering: ABED, AEBD, or EABD

(c) [7 pts] Consider doing inference in an m x n lattice Bayes Net, as shown below. The network consists of mn binary variables $V_{i,j}$, and you have observed that $V_{m,n} = +v_{m,n}$.

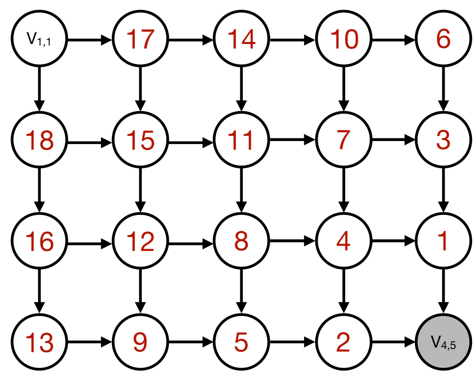


You wish to calculate $P(V_{1,1} | +v_{m,n})$ using variable elimination. To maximize computational efficiency, you wish to use a variable elimination ordering for which the size of the largest generated factor is as small as possible.

(i) [4 pts] First consider the special case where $m = 4$ and $n = 5$. A reproduction of the lattice is shown below, with variable names for non-query variables omitted. Please provide your optimal elimination ordering for this example by numbering the nodes below in the order they will be eliminated (i.e. write a number such as $1, 2, 3, \ldots$ inside every node that will be eliminated.)

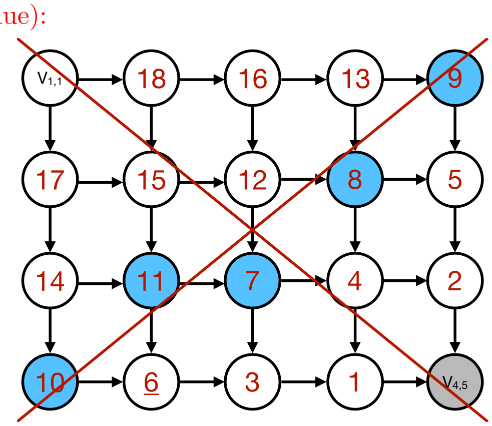


(or)



Note that there is actually more than one correct ordering, and that a few minor variations on the orderings given above are possible. However, it's important to start near the same corner as the evidence variable and to never create a factor that involves more than 4 non-evidence variables.

However, the ordering shown below is suboptimal (eliminating node 6 will create a size 2^5 factor involving the five nodes highlighted in blue):

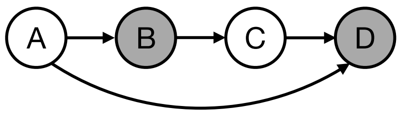


(ii) [3 pts] Now consider the general case (assume $m > 2$ and $n > 2$). What is the size of the largest factor generated under the most efficient elimination ordering? Your answer should be the number of rows in the factor's table, expressed in terms of m and n.

Size (number of rows) of the largest factor: $2^{\min(m,n)}$

Q6. [16 pts] Bayes Nets: Sampling

Consider the following Bayes Net, where we have observed that $B = +b$ and $D = +d$.



$P(A)$	
$+a$	0.5
$-a$	0.5

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

$P(C\|B)$		
$+b$	$+c$	0.1
$+b$	$-c$	0.9
$-b$	$+c$	0.7
$-b$	$-c$	0.3

$P(D\|A,C)$			
$+a$	$+c$	$+d$	0.6
$+a$	$+c$	$-d$	0.4
$+a$	$-c$	$+d$	0.1
$+a$	$-c$	$-d$	0.9
$-a$	$+c$	$+d$	0.2
$-a$	$+c$	$-d$	0.8
$-a$	$-c$	$+d$	0.5
$-a$	$-c$	$-d$	0.5

(a) [4 pts] Consider doing Gibbs sampling for this example. Assume that we have initialized all variables to the values $+a, +b, +c, +d$. We then unassign the variable C, such that we have $A = +a$, $B = +b$, $C = ?$, $D = +d$. Calculate the probabilities for new values of C at this stage of the Gibbs sampling procedure.

$$P(C = +c \text{ at the next step of Gibbs sampling}) = \frac{0.1 \cdot 0.6}{0.1 \cdot 0.6 + 0.9 \cdot 0.1} = \boxed{\frac{2}{5}}$$

$$P(C = -c \text{ at the next step of Gibbs sampling}) = \frac{0.9 \cdot 0.1}{0.1 \cdot 0.6 + 0.9 \cdot 0.1} = \boxed{\frac{3}{5}}$$

(b) [8 pts] Consider a sampling scheme that is a hybrid of rejection sampling and likelihood-weighted sampling. Under this scheme, we first perform rejection sampling for the variables A and B. We then take the sampled values for A and B and extend the sample to include values for variables C and D, using likelihood-weighted sampling.

(i) [2 pts] Below is a list of candidate samples. Mark the samples that would be rejected by the rejection sampling portion of the hybrid scheme.

- ■ $-a$ $-b$
- □ $+a$ $+b$
- ■ $+a$ $-b$
- □ $-a$ $+b$

(ii) [4 pts] To decouple from part (i), you now receive a *new* set of samples shown below. Fill in the weights for these samples under our hybrid scheme.

				Weight
$-a$	$+b$	$-c$	$+d$	$\boxed{0.5}$
$+a$	$+b$	$-c$	$+d$	$\boxed{0.1}$
$+a$	$+b$	$-c$	$+d$	$\boxed{0.1}$
$-a$	$+b$	$+c$	$+d$	$\boxed{0.2}$
$+a$	$+b$	$+c$	$+d$	$\boxed{0.6}$

(iii) [2 pts] Use the weighted samples from part (ii) to calculate an estimate for $P(+a| + b, +d)$.

The estimate of $P(+a| + b, +d)$ is $\frac{0.1 + 0.1 + 0.6}{0.5 + 0.1 + 0.1 + 0.2 + 0.6} = \boxed{\frac{8}{15}}$

(c) [4 pts] We now attempt to design an alternative hybrid sampling scheme that combines elements of likelihood-weighted and rejection sampling. For each proposed scheme, indicate whether it is valid, i.e. whether the weighted samples it produces correctly approximate the distribution $P(A, C| + b, +d)$.

(i) [2 pts] *First collect a likelihood-weighted sample for the variables A and B. Then switch to rejection sampling for the variables C and D. In case of rejection, the values of A and B and the sample weight are* ***thrown away****. Sampling then restarts from node* ***A****.*

● Valid ○ Invalid

(ii) [2 pts] *First collect a likelihood-weighted sample for the variables A and B. Then switch to rejection sampling for the variables C and D. In case of rejection, the values of A and B and the sample weight are* ***retained****. Sampling then restarts from node* ***C****.*

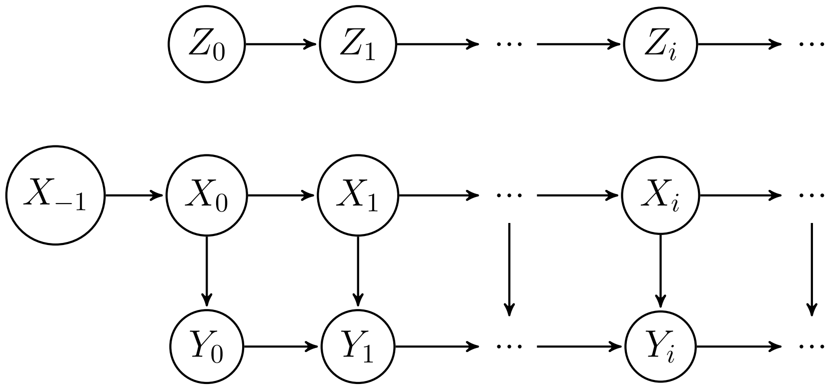
○ Valid ● Invalid

The sampling procedure in part (i) is the correct way of combining likelihood-weighted and rejection sampling: any time a node gets rejected, the sample must be thrown out in its entirety. In part (ii), however, the evidence that $D = +d$ has no effect on which values of A are sampled or on the sample weights. This means that values for A would be sampled according to $P(A| + b)$, not $P(A| + b, +d)$.

As an extreme case, suppose node D had a different probability table where $P(+d| + a) = 0$. Following the procedure from part (ii), we might start by sampling $(+a, +b)$ and assigning a weight according to $P(+b| + a)$. However, when we move on to rejection sampling we will be forced to continuously reject all possible values because our evidence $+d$ is inconsistent with our the assignment of $A = +a$. This means that the procedure from part (ii) is flawed to the extent that it might fail to generate a sample altogether!

Q7. [10 pts] Mini Forward

(a) [10 pts] Let $Z_0, Z_1, Z_2, ...$ be a Markov Chain. You believe that any state Z_i is represented as the union of two states X_i and Y_i. Therefore $P(Z_i) = P(X_i, Y_i)$. The following diagrams illustrate the independence assumptions of the problem.



Recall that the mini forward algorithm allows us to recursively calculate the forward simulation of our model when we are given the initial state and transition dynamics. In other words, it gives us a way to find $P(Z_i)$.

Suppose we are given $X_{-1} = x$ and the tables $P(Y_0 \mid X_0)$, $P(X_i \mid X_{i-1})$, and $P(Y_i \mid Y_{i-1}, X_i)$. Derive the mini forward algorithm that tells us $P(Z_i)$ in terms of given distributions.

(i) [3 pts] Write out the base case of the mini forward algorithm for $P(Z_i)$.

Because we are given X_{-1}, we know $P(X_{-1})$.
We want $P(Z_0)$, so that we have a base case for our recursion.

$P(Z_0) = P(X_0|X_{-1} = x)P(Y_0 \mid X_0)$

(ii) [7 pts] Write out the recursive component of the mini forward algorithm for $P(Z_i)$. It is not necessary to solve the problem, but you can consider Z_i to be equal to (X_i, Y_i).

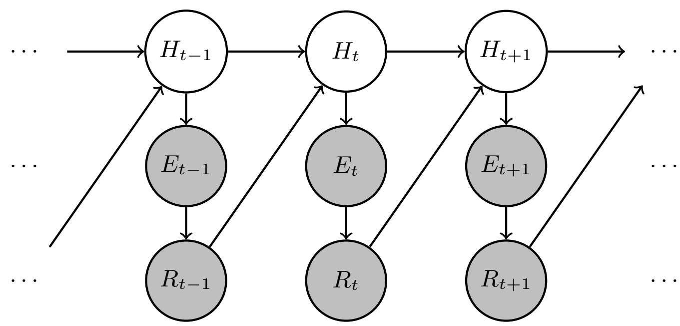
The first thing to realize in this question is that $P(Z_i) = P(X_i, Y_i)$ and $P(Z_{i-1}) = P(X_{i-1}, Y_{i-1})$. You'll be recursing over the right hand sides of those equalities. Then you also need to see that $P(X_{i-1} \mid X_{i-2}) = P(X_i \mid X_{i-1})$

Here is a possible solution:

$$P(Z_i) = P(X_i, Y_i)$$
$$P(X_i, Y_i) = \sum_{x_{i-1}} \sum_{y_{i-1}} P(X_i \mid x_{i-1}) P(Y_i \mid y_{i-1}, X_i) P(x_{i-1}, y_{i-1})$$

Q8. [8 pts] HMM: Human-Robot Interaction

In the near future, autonomous robots would live among us. Therefore, it is important for the robots to know how to properly act in the presence of humans. In this question, we are exploring a simplified model of this interaction. Here, we are assuming that we can observe the robot's actions at time t, R_t, and an evidence observation, E_t, directly caused by the human action, H_t. Humans actions and Robots actions from the past time-step affect the Human's and Robot's actions in the next time-step. In this problem, we will remain consistent with the convention that capital letters (H_t) refer to random variables and lowercase letters (h_t) refer to a particular value the random variable can take. The structure is given below:



You are supplied with the following probability tables: $P(R_t \mid E_t)$, $P(H_t \mid H_{t-1}, R_{t-1})$, $P(H_0)$, $P(E_t \mid H_t)$.

Let us derive the forward algorithm for this model. We will split our computation into two components, a **time-elapse update** expression and a **observe update** expression.

(a) [2 pts] We would like to incorporate the evidence that we observe at time t. Using the time-lapse update expression we will derive separately, we would like to find the **observe update** expression:

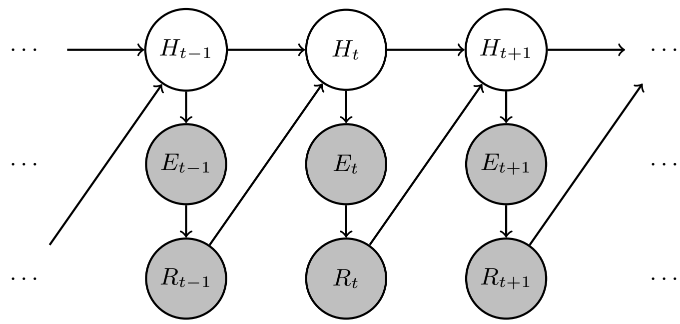
$$O(H_t) = P(H_t|e_{0:t}, r_{0:t})$$

In other words, we would like to compute the distribution of potential human states at time t given all observations up to and including time t. In addition to the conditional probability tables associated with the network's nodes, we are given $T(H_t) = P(H_t \mid e_{0:t-1}, r_{0:t-1})$, which we will assume is correctly computed in the time-elapse update that we will derive in the next part. From the options below, select *all* the options that **both** make valid independence assumptions and would evaluate to the observe update expression.

- ■ $\frac{P(H_t|e_{0:t-1},r_{0:t-1})P(e_t|H_t)P(r_t|e_t)}{\sum_{h_t} P(h_t|e_{0:t-1},r_{0:t-1})P(e_t|h_t)P(r_t|e_t)}$
- ■ $\frac{P(H_t|e_{0:t-1},r_{0:t-1})P(e_t|H_t)}{\sum_{h_t} P(h_t|e_{0:t-1},r_{0:t-1})P(e_t|h_t)}$
- □ $\frac{\sum_{e_t} P(H_t|e_{0:t-1},r_{0:t-1})P(e_t|H_t)}{\sum_{h_t} P(h_t|e_{0:t-1},r_{0:t-1})P(e_t|r_{t-1},H_{t-1})}$
- □ $\sum_{r_{t-1}} P(H_t \mid e_{0:t-1}, r_{0:t-1})P(r_{t-1} \mid e_{t-1})$
- □ $\sum_{r_t} P(H_t \mid e_{0:t-1}, r_{0:t-1})P(r_t \mid r_{t-1}, e_t)$
- □ $\sum_{h_{t+1}} P(H_t \mid e_{0:t-1}, r_{0:t-1})P(h_{t+1} \mid r_t)$

$$P(H_t|e_{0:t}, r_{0:t}) = \frac{P(H_t, e_{0:t}, e_{0:t})}{\sum_{h_t} P(h_t, e_{0:t}, e_{0:t})} = \frac{P(H_t \mid e_{0:t-1}, r_{0:t-1})P(e_t \mid H_t)P(r_t \mid e_t)}{P(r_t \mid e_t)\sum_{h_t} P(h_t \mid e_{0:t-1}, r_{0:t-1})P(e_t \mid h_t)}$$

The structure below is identical to the one in the beginning of the question and is repeated for your convenience.



(b) [6 pts] We are interested in predicting what the state of human is at time t (H_t), given all the observations through $t-1$. Therefore, the **time-elapse update** expression has the following form:

$$T(H_t) = P(H_t|e_{0:t-1}, r_{0:t-1})$$

Derive an expression for the given time-elapse update above using the probability tables provided in the question and the observe update expression, $O(H_{t-1}) = P(H_{t-1}|e_{0:t-1}, r_{0:t-1})$. Write your final expression in the space provided at below. You may use the function O in your solution if you prefer.

The derivation of the time-elapse update for this setup is similar to the one we have seen in lecture; however, here, we have additional observations and dependencies.

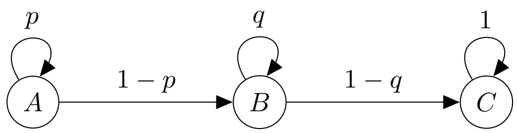
$$\begin{aligned} P(H_t|e_{0:t-1}, r_{0:t-1}) &= \sum_{h_{t-1}} P(H_t, h_{t-1} \mid e_{0:t-1}, r_{0:t-1}) \\ &= \sum_{h_{t-1}} P(H_t \mid h_{t-1}, r_{t-1}) P(h_{t-1} \mid e_{0:t-1}, r_{0:t-1}) \end{aligned}$$

$P(H_t|e_{0:t-1}, r_{0:t-1}) =$ $\sum_{h_{t-1}} P(H_t \mid h_{t-1}, r_{t-1}) P(h_{t-1} \mid e_{0:t-1}, r_{0:t-1})$

SID: ________

Q9. [5 pts] A Not So Random Walk

Pacman is trying to predict the position of a ghost, which he knows has the following transition graph:



Here, $0 < p < 1$ and $0 < q < 1$ are arbitrary probabilities. It is known that the ghost always starts in state A. For this problem, we consider time to begin at 0. For example, at time 0, the ghost is in A with probability 1, and at time 1, the ghost is in A with probability p or in B with probability $1 - p$.

In all of the following questions, you may assume that n is large enough so that the given event occurs with non-zero probability.

Note: For full credit, you may refer to the correct answer of a **previous** part using the notation $f_k(t)$, where k is a part number and t is a time. For instance, to refer to the correct answer for part (ii) for time $n - 1$, you could write $f_{(\text{ii})}(n-1)$. You may only refer to previous parts, not future parts.

Please note the low point value for each subproblem and allocate time accordingly. Answers should be simplified to the extent possible.

(i) [1 pt] Suppose $p \neq q$. What is the probability that the ghost is in A at time n?
For the ghost to be in A at time n, it must have stayed in A for n steps, which occurs with probability

$$p^n.$$

(ii) [1 pt] Suppose $p \neq q$. What is the probability that the ghost first reaches B at time n?
For the ghost to first reach B at time n, it must have stayed in A for $n - 1$ steps, then transitioned to B. This occurs with probability

$$f_{(\text{i})}(n-1) \cdot (1-p) = p^{n-1}(1-p).$$

(iii) [1 pt] Suppose $p \neq q$. What is the probability that the ghost is in B at time n?
For the ghost to be in B at time n, it must have first reached B at time i for some $1 \leq i \leq n$, then stayed there for $n - i$ steps. Summing over all values of i gives

$$\sum_{i=1}^{n} f_{(\text{ii})}(i) \cdot q^{n-i} = \sum_{i=1}^{n} p^{i-1}(1-p)q^{n-i} = \frac{(1-p)q^n}{p} \sum_{i=1}^{n} \left(\frac{p}{q}\right)^i = \frac{(1-p)q^n}{p} \cdot \frac{p}{q} \cdot \frac{1 - \left(\frac{p}{q}\right)^n}{1 - \frac{p}{q}} = (1-p)\frac{q^n - p^n}{q - p}.$$

(iv) [1 pt] Suppose $p \neq q$. What is the probability that the ghost first reaches C at time n?
For the ghost to first reach C at time n, it must have been in B at time $n - 1$, then transitioned to C. This occurs with probability

$$f_{(\text{iii})}(n-1) \cdot (1-q) = (1-p)\frac{q^{n-1} - p^{n-1}}{q - p}(1-q).$$

(v) [1 pt] Suppose $p \neq q$. What is the probability that the ghost is in C at time n?
For the ghost to be in C at time n, it must not be in A or B at time n. This occurs with probability

$$1 - f_{(\text{i})}(n) - f_{(\text{iii})}(n) = 1 - p^n - (1-p)\frac{q^n - p^n}{q - p}.$$

Alternatively, for the ghost to be in C at time n, it must have first reached C at time i for some $2 \le i \le n$, then stayed there for $n - i$ steps. Note that we can equivalently range over $1 \le i \le n$ for computational convenience, since $f_{(\text{iv})}(1) = 0$. Summing over all values of i gives

$$\begin{aligned}\sum_{i=2}^{n} f_{(\text{iv})}(i) \cdot 1^{n-i} = \sum_{i=1}^{n} f_{(\text{iv})}(i) \cdot 1^{n-i} &= \sum_{i=1}^{n} (1-p)\frac{q^{i-1} - p^{i-1}}{q-p}(1-q) \\ &= \frac{(1-p)(1-q)}{q-p}\left(\frac{1-q^n}{1-q} - \frac{1-p^n}{1-p}\right) = \frac{(1-p)(1-q^n) - (1-q)(1-p^n)}{q-p},\end{aligned}$$

which is equivalent to the previous expression.