

Q1. [8 pts] Probability: Chain Rule Madness

Note that the set of choices is the same for all of the following questions. You may find it beneficial to read all the questions first, and to make a table on scratch paper.

- (a) [2 pts] Fill in the circles of **all** expressions that are equivalent to $\mathbf{P(A, B | C)}$,
given no independence assumptions:

- | | |
|--|--|
| <input type="radio"/> $\frac{P(C A) P(A B) P(B)}{P(C)}$ | <input type="radio"/> $\frac{P(B,C A) P(A)}{P(B,C)}$ |
| <input type="radio"/> $\frac{P(A C) P(C B) P(B)}{P(B,C)}$ | <input type="radio"/> $\frac{P(A C) P(B,C)}{P(C)}$ |
| <input type="radio"/> $\frac{P(C A,B) P(B A) P(A)}{P(B C) P(C)}$ | <input type="radio"/> $P(A B, C) P(B C)$ |
| | <input type="radio"/> None of the above. |

- (b) [2 pts] Fill in the circles of **all** expressions that are equivalent to $\mathbf{P(A, B | C)}$,
given that $\mathbf{A \perp\!\!\!\perp B | C}$:

- | | |
|--|--|
| <input type="radio"/> $\frac{P(C A) P(A B) P(B)}{P(C)}$ | <input type="radio"/> $\frac{P(B,C A) P(A)}{P(B,C)}$ |
| <input type="radio"/> $\frac{P(A C) P(C B) P(B)}{P(B,C)}$ | <input type="radio"/> $\frac{P(A C) P(B,C)}{P(C)}$ |
| <input type="radio"/> $\frac{P(C A,B) P(B A) P(A)}{P(B C) P(C)}$ | <input type="radio"/> $P(A B, C) P(B C)$ |
| | <input type="radio"/> None of the above. |

- (c) [2 pts] Fill in the circles of **all** expressions that are equivalent to $\mathbf{P(A | B, C)}$,
given no independence assumptions:

- | | |
|--|--|
| <input type="radio"/> $\frac{P(C A) P(A B) P(B)}{P(C)}$ | <input type="radio"/> $\frac{P(B,C A) P(A)}{P(B,C)}$ |
| <input type="radio"/> $\frac{P(A C) P(C B) P(B)}{P(B,C)}$ | <input type="radio"/> $\frac{P(A C) P(B,C)}{P(C)}$ |
| <input type="radio"/> $\frac{P(C A,B) P(B A) P(A)}{P(B C) P(C)}$ | <input type="radio"/> $P(A B, C) P(B C)$ |
| | <input type="radio"/> None of the above. |

- (d) [2 pts] Fill in the circles of **all** expressions that are equivalent to $\mathbf{P(A | B, C)}$,
given that $\mathbf{A \perp\!\!\!\perp B | C}$:

- | | |
|--|--|
| <input type="radio"/> $\frac{P(C A) P(A B) P(B)}{P(C)}$ | <input type="radio"/> $\frac{P(B,C A) P(A)}{P(B,C)}$ |
| <input type="radio"/> $\frac{P(A C) P(C B) P(B)}{P(B,C)}$ | <input type="radio"/> $\frac{P(A C) P(B,C)}{P(C)}$ |
| <input type="radio"/> $\frac{P(C A,B) P(B A) P(A)}{P(B C) P(C)}$ | <input type="radio"/> $P(A B, C) P(B C)$ |
| | <input type="radio"/> None of the above. |

Q2. [15 pts] Probabilities

(a) [3 pts] Fill in the circles of **all** expressions that are equal to **1**,
given no independence assumptions:

- | | |
|---|---|
| <input type="radio"/> $\sum_a P(A = a \mid B)$ | <input type="radio"/> $\sum_a \sum_b P(A = a \mid B = b)$ |
| <input type="radio"/> $\sum_b P(A \mid B = b)$ | <input type="radio"/> $\sum_a \sum_b P(A = a) P(B = b)$ |
| <input type="radio"/> $\sum_a \sum_b P(A = a, B = b)$ | <input type="radio"/> $\sum_a P(A = a) P(B = b)$ |
| | <input type="radio"/> None of the above. |

(b) [3 pts] Fill in the circles of **all** expressions that are equal to **$P(\mathbf{A}, \mathbf{B}, \mathbf{C})$** ,
given no independence assumptions:

- | | |
|---|--|
| <input type="radio"/> $P(A \mid B, C) P(B \mid C) P(C)$ | <input type="radio"/> $P(C \mid A, B) P(A, B)$ |
| <input type="radio"/> $P(C \mid A, B) P(A) P(B)$ | <input type="radio"/> $P(A \mid B) P(B \mid C) P(C)$ |
| <input type="radio"/> $P(A, B \mid C) P(C)$ | <input type="radio"/> $P(A \mid B, C) P(B \mid A, C) P(C \mid A, B)$ |
| | <input type="radio"/> None of the above. |

(c) [3 pts] Fill in the circles of **all** expressions that are equal to **$P(\mathbf{A} \mid \mathbf{B}, \mathbf{C})$** ,
given no independence assumptions:

- | | |
|--|---|
| <input type="radio"/> $\frac{P(A, B, C)}{\sum_a P(A = a, B, C)}$ | <input type="radio"/> $\frac{P(B \mid A, C) P(A \mid C)}{P(B, C)}$ |
| <input type="radio"/> $\frac{P(B, C \mid A) P(A)}{P(B, C)}$ | <input type="radio"/> $\frac{P(B \mid A, C) P(C \mid A, B)}{P(B, C)}$ |
| <input type="radio"/> $\frac{P(B \mid A, C) P(A \mid C)}{P(B \mid C)}$ | <input type="radio"/> $\frac{P(A, B \mid C)}{P(B \mid A, C)}$ |
| | <input type="radio"/> None of the above. |

(d) [3 pts] Fill in the circles of **all** expressions that are equal to **$P(\mathbf{A} \mid \mathbf{B})$** ,
given that **$\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C}$** :

- | | |
|--|--|
| <input type="radio"/> $\frac{P(A \mid C) P(B \mid C)}{P(B)}$ | <input type="radio"/> $\frac{P(A \mid B, C)}{P(A \mid C)}$ |
| <input type="radio"/> $\frac{P(A \mid C) P(B \mid C)}{P(B \mid C)}$ | <input type="radio"/> $\frac{\sum_c P(B \mid A, C = c) P(A, C = c)}{P(B)}$ |
| <input type="radio"/> $\frac{\sum_c P(A \mid C = c) P(B \mid C = c) P(C = c)}{\sum_{c'} P(B \mid C = c') P(C = c')}$ | <input type="radio"/> $\frac{\sum_c P(A, C = c) P(B \mid C = c)}{\sum_{c'} P(A, B, C = c')}$ |
| | <input type="radio"/> None of the above. |

(e) [3 pts] Fill in the circles of **all** expressions that are equal to **$P(\mathbf{A}, \mathbf{B}, \mathbf{C})$** ,
given that **$\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C}$** and **$\mathbf{A} \perp\!\!\!\perp \mathbf{C}$** :

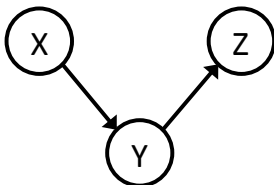
- | | |
|--|--|
| <input type="radio"/> $P(A) P(B) P(C)$ | <input type="radio"/> $P(A \mid B, C) P(B \mid A, C) P(C \mid A, B)$ |
| <input type="radio"/> $P(A) P(B, C)$ | <input type="radio"/> $P(A \mid C) P(B \mid C) P(C)$ |
| <input type="radio"/> $P(A \mid B) P(B \mid C) P(C)$ | <input type="radio"/> $P(A \mid C) P(B \mid C)$ |
| | <input type="radio"/> None of the above. |

Q3. [15 pts] Need That CPT

For each of the following questions, mark the **minimum** set of variables whose associated probability table is **needed** to answer the query.

For example, for part (a), the probability table associated with X is $P(X)$, the probability table associated with Y is $P(Y | X)$, and the probability table associated with Z is $P(Z | Y)$. The query for part (a) (i) is $P(X | Y)$.

(a) [3 pts]

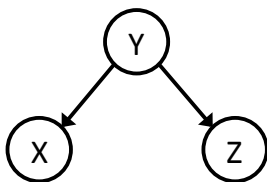


(i) [1 pt] $P(X | Y)$ ☐ X ☐ Y ☐ Z

(ii) [1 pt] $P(X | Z)$ ☐ X ☐ Y ☐ Z

(iii) [1 pt] $P(Y | Z)$ ☐ X ☐ Y ☐ Z

(b) [3 pts]

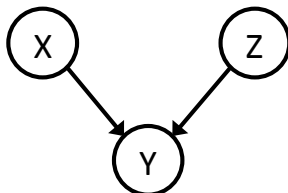


(i) [1 pt] $P(X | Y)$ ☐ X ☐ Y ☐ Z

(ii) [1 pt] $P(X | Z)$ ☐ X ☐ Y ☐ Z

(iii) [1 pt] $P(Z | X)$ ☐ X ☐ Y ☐ Z

(c) [3 pts]

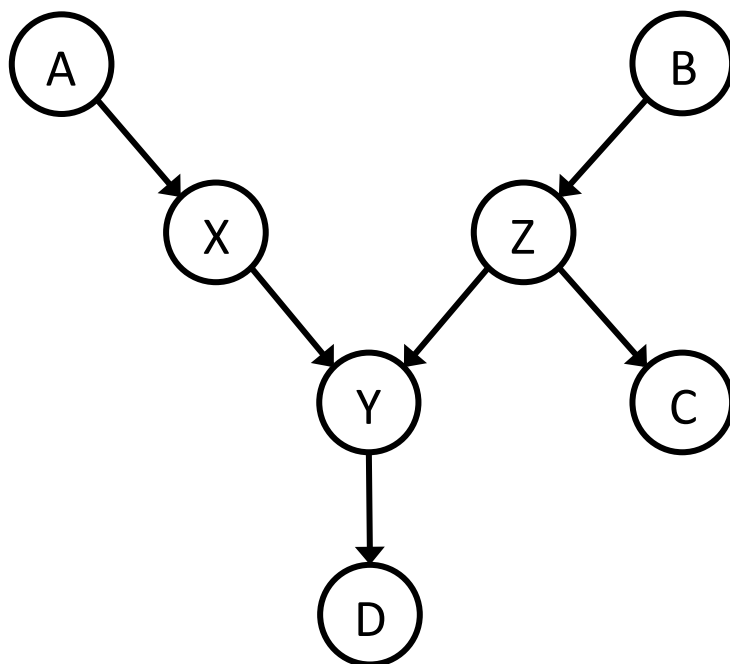


(i) [1 pt] $P(X | Y)$ ☐ X ☐ Y ☐ Z

(ii) [1 pt] $P(X | Z)$ ☐ X ☐ Y ☐ Z

(iii) [1 pt] $P(Y | Z)$ ☐ X ☐ Y ☐ Z

(d) [6 pts]



(i) [2 pts] $P(X \mid Z)$

☐ X ☐ Y ☐ Z ☐ A ☐ B ☐ C ☐ D

(ii) [2 pts] $P(X \mid D)$

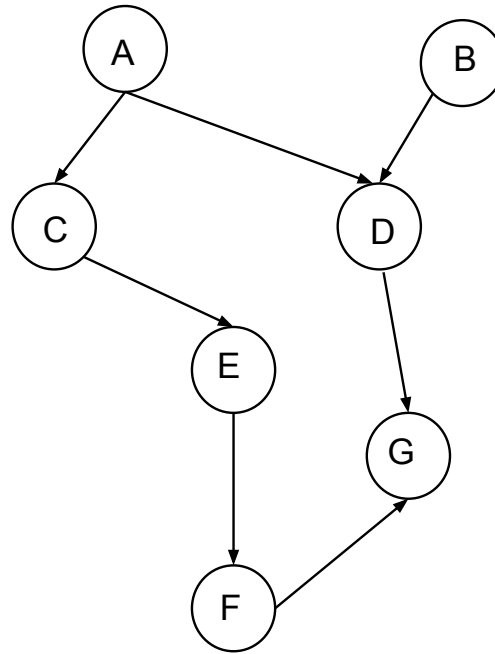
☐ X ☐ Y ☐ Z ☐ A ☐ B ☐ C ☐ D

(iii) [2 pts] $P(X \mid Z, D)$

☐ X ☐ Y ☐ Z ☐ A ☐ B ☐ C ☐ D

Q4. [14 pts] Fun with Bayes Nets

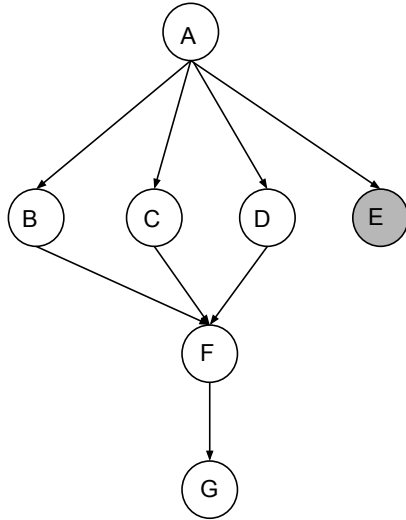
(a) [6 pts] Consider a Bayes Net with the following graph:



Which of the following are guaranteed to be true without making any additional conditional independence assumptions, other than those implied by the graph? (Mark all true statements)

- ☐ $P(A \mid C, E) = P(A \mid C)$
- ☐ $P(A, E \mid G) = P(A \mid G) * P(E \mid G)$
- ☐ $P(A \mid B = b) = P(A)$
- ☐ $P(A \mid B, G) = P(A \mid G)$
- ☐ $P(E, G \mid D) = P(E \mid D) * P(G \mid D)$
- ☐ $P(A, B \mid F) = P(A \mid F) * P(B \mid F)$

(b) [8 pts] Now consider a Bayes Net with the following graph:



The factors associated with the Bayes Net are $P(A)$, $P(B | A)$, $P(C | A)$, $P(D | A)$, $P(E | A)$, $P(F | B, C, D)$ and $P(G | F)$. We will consider variable elimination to answer the query $P(G | +e)$.

(i) [2 pts] Suppose the first variable we eliminate is A . Provide an expression for the resulting factor as a function of the original factors.

(ii) [2 pts] Suppose the first variable we eliminate is B . Provide an expression for the resulting factor generated as a function of the original factors.

(iii) [2 pts] Suppose the first variable we eliminate is F . Provide an expression for the resulting factor as a function of the original factors.

(iv) [2 pts] Suppose we eliminated the variables A, B, C, D, F in that order, and the single remaining factor is $f(+e, G)$. How do we obtain $P(G | +e)$ from this remaining factor? (Your answer should be in the form of an equation.)

$$P(G | +e) = \underline{\hspace{2cm}}$$

Q5. [12 pts] Value of Gambling and Bribery

The local casino is offering a new game. There are two biased coins that are indistinguishable in appearance. There is a *head-biased* coin, which yields head with probability 0.8 (and tails with probability 0.2). There is a *tail-biased* coin, which yields tail with probability 0.8 (and head with probability 0.2).

At the start of the game, the dealer gives you one of the two coins at random, with equal probability. You get to flip that coin once. Then you decide if you want to stop or continue. If you choose to continue, you flip it 10 more times. In those 10 flips, each time it yields head, you get \$1, and each time it yields tail, you lose \$1.

- (a) [1 pt] What is the expected value of your earnings if continuing to play with a head-biased coin?

- (b) [1 pt] What is the expected value of your earnings if continuing to play with a tail-biased coin?

- (c) [3 pts] Suppose the first flip comes out head.

- (i) [1 pt] What is the posterior probability that the coin is *head-biased*? _____

- (ii) [1 pt] What is the expected value of your earnings for continuing to play? _____

- (iii) [1 pt] Which is the action that maximizes the expected value of your earnings? ☐ Continue ☐ Stop

- (d) Suppose the first flip comes out tail.

- (i) [1 pt] What is the posterior probability that the coin is *tail-biased*? _____

- (ii) [1 pt] What is the expected value of your earnings for continuing to play? _____

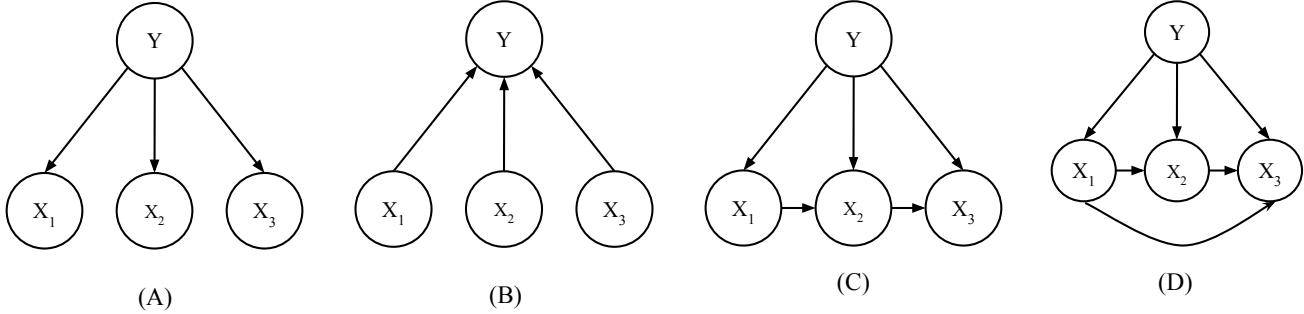
- (iii) [1 pt] Which is the action that maximizes the expected value of your earnings? ☐ Continue ☐ Stop

- (e) [1 pt] What is the expected value of your earnings after playing the game optimally one time?

- (f) [3 pts] Suppose again that the first flip yields head. The dealer knows which coin you picked. How much are you willing to pay the dealer to find out the type of the coin?

Q6. [12 pts] Bayes Net Modeling

Abhishek has been getting a lot of spam recently(!) and is not satisfied with his email client's Naive Bayes spam classifier. Thankfully he knows about Bayes Nets and has decided to implement his own spam classifier. It is your job to help him model his Bayes Nets and train them. The following are 4 Bayes Nets he is considering. Each variable can take on the values $\{0, 1\}$.



- (a) [6 pts] Abhishek wants to know how much memory he needs to store each of these Bayes Nets on disk. The amount of memory depends on the number of values he would need to store in the CPTs for that Bayes Net. For each of the nets above give the **least** number of parameters he would need to store to completely specify the Bayes Net.

A = _____

C = _____

B = _____

D = _____

- (b) It's now time to train the Bayes Nets. Abhishek has training datasets D_l, D_m, D_s and a test dataset D_t . The number of training examples in each set vary as $|D_l| > |D_m| > |D_s|$. e_{tr} and e_{te} represent train and test errors and their arguments represent which dataset a model was trained on. For example, $e_{tr}(A, D_l)$ refers to the training error of model A on dataset D_l .

- (i) [4 pts] Abhishek tries a bunch of experiments using model D. In a typical scenario¹, which of the following can be expected to be true? (Mark all that apply)

- ☐ $e_{tr}(D, D_l) \geq e_{tr}(D, D_m) \geq e_{tr}(D, D_s)$
- ☐ $e_{tr}(D, D_s) \geq e_{tr}(D, D_m) \geq e_{tr}(D, D_l)$
- ☐ $e_{te}(D, D_l) \geq e_{te}(D, D_m) \geq e_{te}(D, D_s)$
- ☐ $e_{te}(D, D_s) \geq e_{te}(D, D_m) \geq e_{te}(D, D_l)$

- ☐ $e_{tr}(D, D_l) \geq e_{te}(D, D_l)$
- ☐ $e_{tr}(D, D_l) \leq e_{te}(D, D_l)$
- ☐ $e_{tr}(D, D_s) \geq e_{te}(D, D_s)$
- ☐ $e_{tr}(D, D_s) \leq e_{te}(D, D_s)$

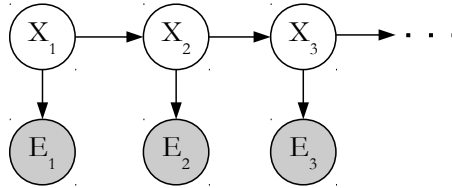
- (ii) [2 pts] Abhishek is now trying to compare performance across different models. In a typical scenario, which of the following can be expected to be true? (Mark all that apply)

- ☐ $e_{tr}(A, D_l) \geq e_{tr}(B, D_l) \geq e_{tr}(D, D_l)$
- ☐ $e_{tr}(A, D_l) \leq e_{tr}(B, D_l) \leq e_{tr}(D, D_l)$

- ☐ $e_{te}(A, D_l) \geq e_{te}(B, D_l) \geq e_{te}(D, D_l)$
- ☐ $e_{te}(A, D_l) \leq e_{te}(B, D_l) \leq e_{te}(D, D_l)$

¹Train and test data sampled from same underlying distribution

Q7. [20 pts] Most Likely Estimates in HMMs



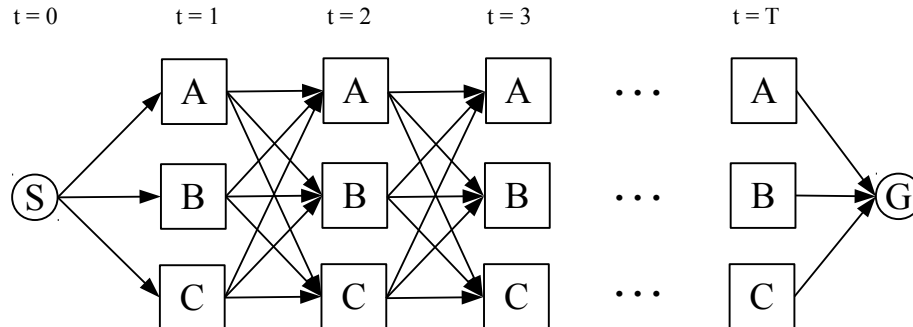
The Viterbi algorithm finds the most probable sequence of hidden states $X_{1:T}$, given a sequence of observations $e_{1:T}$. Throughout this question you may assume there are no ties. Recall that for the canonical HMM structure, the Viterbi algorithm performs the following **dynamic programming**² computations:

$$m_t[x_t] = P(e_t|x_t) \max_{x_{t-1}} P(x_t|x_{t-1})m_{t-1}[x_{t-1}]$$

- (a) [3 pts] For the HMM structure above, which of the following probabilities are maximized by the sequence of states returned by the Viterbi algorithm? Mark **all** the correct option(s).

- ☐ $P(X_{1:T})$
☐ $P(X_T|e_T)$
☐ $P(X_{1:T}|e_{1:T})$
☐ $P(X_{1:T}, e_{1:T})$
☐ $P(X_1)P(e_1|X_1) \prod_{t=2}^T P(e_t|X_t)P(X_t|X_{t-1})$
☐ $P(X_1) \prod_{t=2}^T P(X_t|X_{t-1})$
☐ None of the above

- (b) Consider an HMM structure like the one in part (a) above. Say for all time steps t , the state X_t can take on one of the three values $\{A, B, C\}$. Then, we can represent the state transitions through the following directed graph, also called a *Trellis Diagram*.



We wish to formulate the most probable sequence of hidden state query as a **graph search problem**.

Note in the diagram above, dummy nodes S and G have been added to represent the *start* state and the *goal* state respectively. Further, the transition from the starting node S to the first state X_1 occurs at time step $t = 0$; transition from X_T (the last HMM state) to the goal state G occurs at time step $t = T$.

(The questions for this section are on the following page)

² If you're not familiar with dynamic programming, it is essentially a recursive relation in which the current value is defined as a function of previously computed values. In this case, the value at time t is defined as a function of the values at time $t - 1$

Definition : Let $w_{Y \rightarrow Z}^t$, be the cost of the edge for the transition from state Y at time t to state Z at time $t + 1$. For example, $w_{A \rightarrow B}^1$ is the cost of the edge for transition from state A at time 1 to state B at time 2.

(i) [3 pts] For which **one** of the following values for the weights $w_{Y \rightarrow Z}^t$, $1 \leq t < T$, would the minimum cost path be exactly the same as most likely sequence of states computed by the Viterbi algorithm.

- ☐ $w_{Y \rightarrow Z}^t = -P(X_{t+1} = Z | X_t = Y)$
☐ $w_{Y \rightarrow Z}^t = -P(e_{t+1} | X_{t+1} = Z)P(X_{t+1} = Z | X_t = Y)$
☐ $w_{Y \rightarrow Z}^t = -\log(P(X_{t+1} = Z | X_t = Y))$
☐ $w_{Y \rightarrow Z}^t = -\log(P(e_{t+1} | X_{t+1} = Z)P(X_{t+1} = Z | X_t = Y))$
☐ $w_{Y \rightarrow Z}^t = \frac{1}{P(X_{t+1} = Z | X_t = Y)}$
☐ $w_{Y \rightarrow Z}^t = \frac{1}{P(e_{t+1} | X_{t+1} = Z)P(X_{t+1} = Z | X_t = Y)}$

(ii) [3 pts] The initial probability distribution of the state at time $t = 1$ is given $P(X_1 = Y), Y \in \{A, B, C\}$.

Which **one** of the following should be the value of $w_{S \rightarrow Y}^0, Y \in \{A, B, C\}$ — these are the cost on the edges connecting S to the states at time $t = 1$?

- ☐ $w_{S \rightarrow Y}^0 = -P(X_1 = Y)$
☐ $w_{S \rightarrow Y}^0 = -P(e_1 | X_1 = Y)P(X_1 = Y)$
☐ $w_{S \rightarrow Y}^0 = -\log(P(X_1 = Y))$
☐ $w_{S \rightarrow Y}^0 = -\log(P(e_1 | X_1 = Y)P(X_1 = Y))$
☐ $w_{S \rightarrow Y}^0 = \frac{1}{P(X_1 = Y)}$
☐ $w_{S \rightarrow Y}^0 = \frac{1}{P(e_1 | X_1 = Y)P(X_1 = Y)}$
☐ $w_{S \rightarrow Y}^0 = \alpha, \alpha \in \mathbb{R} : (\text{some constant})$

(iii) [3 pts] Which **one** of the following should be the value of $w_{Y \rightarrow G}^T, Y \in \{A, B, C\}$ — these are the cost on the edges connecting the states at last time step $t = T$ to the goal state G ?

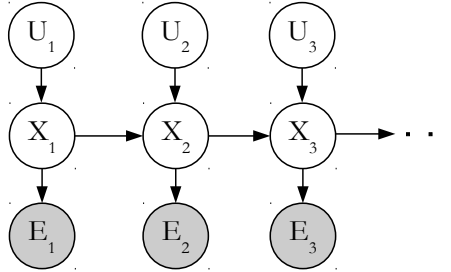
- ☐ $w_{Y \rightarrow G}^T = -P(X_T = Y)$
☐ $w_{Y \rightarrow G}^T = -P(e_T | X_T = Y)P(X_T = Y)$
☐ $w_{Y \rightarrow G}^T = -\log(P(X_T = Y))$
☐ $w_{Y \rightarrow G}^T = -\log(P(e_T | X_T = Y)(P(X_T = Y)))$
☐ $w_{Y \rightarrow G}^T = \frac{1}{P(X_T = Y)}$
☐ $w_{Y \rightarrow G}^T = \frac{1}{P(e_T | X_T = Y)P(X_T = Y)}$
☐ $w_{Y \rightarrow G}^T = \alpha, \alpha \in \mathbb{R} : (\text{some constant})$

(c) We consider extending the Viterbi algorithm for finding the most likely sequence of states in modified HMMs.

For your convenience, the computations performed by the Viterbi algorithm for the canonical HMM structure, like in part (a), are repeated below:

$$m_t[x_t] = P(e_t|x_t) \max_{x_{t-1}} P(x_t|x_{t-1})m_{t-1}[x_{t-1}]$$

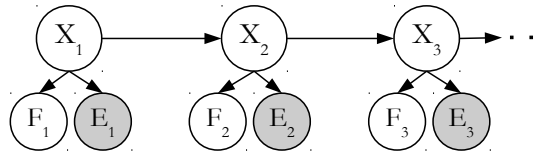
- (i) [4 pts] Consider the HMM below with additional variables U_t . The HMM can be interpreted as : The state X_t at time t is caused due to some action U_t and previous state X_{t-1} . The state X_t emits an observation E_t . Both U_t and X_t are unobserved.



We want to find the most likely sequence of states $X_{1:T}$ and actions $U_{1:T}$, given the sequence of observations $e_{1:T}$. Write a dynamic programming update for $t > 1$ **analogous** to the one for the canonical HMM structure.

$$m_t[x_t, u_t] = \underline{\hspace{10cm}}$$

- (ii) [4 pts] Consider the HMM below with two emission variables at each time step F_t and E_t . E_t is observed while X_t and F_t are unobserved.



We want to find the most likely sequence of states $X_{1:T}$ and the unobserved emissions $F_{1:T}$, given the sequence of observations $e_{1:T}$. Write a dynamic programming update for $t > 1$ **analogous** to the one for the canonical HMM structure.

$$m_t[x_t, f_t] = \underline{\hspace{10cm}}$$