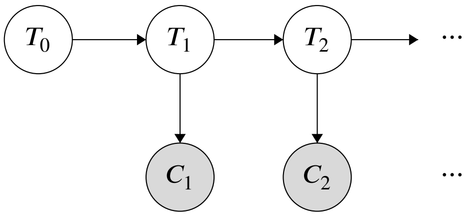
Q3. [12 pts] HMMs: Head TA Kirby

Kirby is serving as a TA. He wants to evaluate his teaching performance after each of his weekly discussion sections, and he does so based on how much his students collaborate during his section. He models this situation using an HMM.



T_0	$P(T_0)$
+t	0.6
-t	0.4

T_{i+1}	T_i	$P(T_{i+1}\|T_i)$
+t	+t	0.7
-t	+t	0.3
+t	-t	0.4
-t	-t	0.6

C_i	T_i	$P(C_i\|T_i)$
+c	+t	0.5
-c	+t	0.5
+c	-t	0.1
-c	-t	0.9

T_i is a binary random variable representing whether Kirby taught sufficiently well during Week i. C_i is another binary random variable representing whether students collaborated during Week i. He does not see his students' collaboration during Week 0.

(a) [8 pts] Using the two steps of the forward algorithm, calculate the distribution of $P(T_1|C_1 = +c)$. For the sake of organization, you may use the first box for the **time elapse** update and the second box for the **observation** update. You may also leave your final answers unsimplified (for example, as fractions).

Time elapse update:

$$\begin{aligned}P(T_1 = +t) &= P(T_1 = +t|T_0 = +t)P(T_0 = +t) + P(T_1 = +t|T_0 = -t)P(T_0 = -t)\\ &= 0.7 * 0.6 + 0.4 * 0.4\\ &= 0.58\\ P(T_1 = -t) &= P(T_1 = -t|T_0 = +t)P(T_0 = +t) + P(T_1 = -t|T_0 = -t)P(T_0 = -t)\\ &= 0.3 * 0.6 + 0.6 * 0.4\\ &= 0.42\end{aligned}$$

Observation update:

$$
\begin{aligned}
P(T_1 = +t, C_1 = +c) &= P(C_1 = +c | T_1 = +t) P(T_1 = +t) \\
&= 0.5 * 0.58 = \mathbf{0.29} \\
P(T_1 = -t, C_1 = +c) &= P(C_1 = +c | T_1 = -t) P(T_1 = -t) \\
&= 0.1 * 0.42 = \mathbf{0.042} \\
P(C_1 = +c) &= P(C_1 = +c, C_1 = +c) + P(T_1 = -t, C_1 = +c) \\
&= 0.542 + 0.21 = \mathbf{0.332} \\
P(T_1 = +t | C_1 = +c) &= \frac{P(T_1 = +t, C_1 = +c)}{P(C_1 = +c)} \\
P(T_1 = -t | C_1 = +c) &= \frac{P(T_1 = -t, C_1 = +c)}{P(C_1 = +c)}
\end{aligned}
$$

$P(T_1 = +t | C_1 = +c) =$ $\boxed{\frac{0.29}{0.332}}$

$P(T_1 = -t | C_1 = +c) =$ $\boxed{\frac{0.042}{0.332}}$

(b) [4 pts] In order to save computational resources, Kirby turns to particle filtering to analyze this HMM.

(i) [2 pts] At timestep $t = 3$, Kirby has observed the following evidence: $C_1 = +c$, $C_2 = -c$, and $C_3 = +c$. Following the particle filtering algorithm, assign weights to particles in the following states at $t = 3$:

Particles in state $+t$ will have weight: $\boxed{0.5}$

Particles in state $-t$ will have weight: $\boxed{0.1}$

The weight of a particle in some arbitrary state t is $P(C_3 = +c | T_3 = t)$.

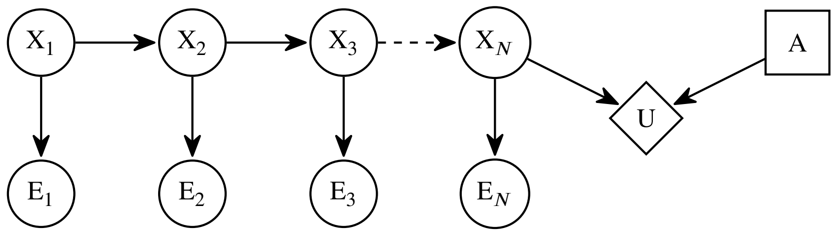
(ii) [2 pts] At timestep $t = 6$, we observe 3 particles in state $+t$ and 5 particles in state $-t$, and $C_6 = -c$. Fill in the table describing the distribution that we resample our new particles from for $t = 7$. Show any work in the box on the left.

T_7	$P(T_7)$
+t	0.25
-t	0.75

The 3 particles in state +t each have weight $P(C_6 = -c | T_6 = +t) = 0.5$ (totaling $0.5 * 3 = 1.5$), while the 5 particles in state -t each have weight $P(C_6 = -c | T_6 = -t) = 0.9$ (totaling $0.9 * 5 = 4.5$). Normalizing the weights to form a probability distribution across the domain of T, we arrive at $\mathbf{P(T_7 = +t) = 0.25}$ and $\mathbf{P(T_7 = -t) = 0.75}$.

Q6. [13 pts] HMMs and VPI: Markov Menagerie

Consider the following hidden Markov model (HMM). All random variables are binary.



We would like to compute $\text{VPI}(E_N) = \text{MEU}(E_N) - \text{MEU}(\emptyset)$.

(a) [1 pt] Which of these distributions is needed to compute $\text{MEU}(\emptyset)$?

● $P(X_N)$ ○ $P(X_N|X_{N-1})$ ○ $P(X_N|E_1, \ldots, E_N)$ ○ $P(X_N|E_N)$

The utility U depends on X_N, so we need a distribution over X_N to compute the expected utility of each action.
$\text{MEU}(\emptyset)$ means that we are given no evidence, so the distribution over X_N should also be given no evidence.

(b) [2 pts] Which of these computations can be used to derive the distribution in part **(a)**?

○ Run the forward algorithm with no modifications.
○ Run the forward algorithm, skipping all time elapse updates.
● Run the forward algorithm, skipping all observation updates.
○ Read the conditional probability table under X_N.

We need $P(X_N)$. This is not in the Bayes' net; the CPT under X_N is $P(X_N|X_{N-1})$.
To obtain $P(X_N)$, we need to run the forward algorithm, skipping all observation updates, because we have no evidence.

(c) [3 pts] Write an expression that can be used to compute $\text{MEU}(E_N)$.

$$\text{MEU}(E_N) = \sum_{e_N} \textbf{(i)} \left[\max_a \sum_{x_N} \textbf{(ii)} \;\; \textbf{(iii)} \right]$$

(i) ● $P(E_N)$ ○ $P(E_N|E_{N-1})$ ○ $P(E_N|X_N)$ ○ $P(E_N|X_1, \ldots, X_N)$

(ii) ○ $P(X_N|X_{N-1})$ ○ $P(X_N|X_1, \ldots, X_{N-1})$ ● $P(X_N|E_N)$ ○ $P(X_N|E_1, \ldots, E_N)$

(iii) ○ 1 ○ $U(x_N)$ ○ $U(a)$ ● $U(x_N, a)$

$$\sum_{e_N} P(e_N) \left[\max_a \sum_{x_N} P(x_N|e_N) U(x_N, a) \right]$$

We need $P(X_N|E_N)$ to compute the expected utility of each action.
Then, we also need $P(E_N)$ so that we can weight each expected utility by the probability of that specific evidence occurring.
Finally, we need the utility, which depends on both the value of X_N and the action selected.

(d) [2 pts] Consider the distribution in part **(a)**, which you computed in part **(b)**.
From the distribution in **(a)**, which of the following additional computations results in the distribution in blank **(ii)**?

- ● Apply an additional observation update.
- ○ Apply an additional time elapse update.
- ○ Apply an additional time elapse and observation update.
- ○ Re-run the forward algorithm from the beginning, with no modifications.

From the previous subpart, we have $P(X_N)$, but we need $P(X_N|E_N)$. This requires one extra evidence update in the forward algorithm: we need to weight every value in the table $P(X_N)$ by $P(E_N|X_N)$, and then normalize.

In equations: We have $P(X_N)$ (from the earlier subparts) and $P(E_N|X_N)$ (from the Bayes' net). We want $P(X_N|E_N)$, so we can apply Bayes' rule:

$$P(X_N|E_N) = \frac{P(X_N)P(E_N|X_N)}{P(E_N)}$$

In other words, in the numerator, we're multiplying every value in the $P(X_N)$ table by $P(E_N|X_N)$, and then we normalize (since the denominator is a constant).

(e) [2 pts] Which of the following computations can be used to derive the distribution in blank **(i)**?

- ○ $P(x_N)P(E_N|x_N)$
- ● $\sum_{x_N} P(x_N)P(E_N|x_N)$
- ○ $\sum_{e_N} P(x_N)P(e_N|x_N)$
- ○ $\sum_{x_N} P(E_N)P(x_N|E_N)$

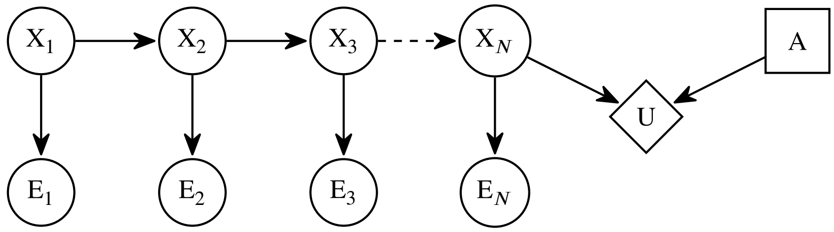
We're given $P(X_N)$ (from earlier subparts), and $P(E_N|X_N)$ (from the Bayes' net). Using the chain rule, we can get:

$$P(X_N)P(E_N|X_N) = P(X_N, E_N)$$

Then, you can sum out X_N to get:

$$\sum_{x_N} P(x_N)P(E_N|x_N) = P(E_N)$$

The diagram, repeated for your convenience:



For the rest of the question, suppose we would like to compute $\text{VPI}(E_1, \ldots, E_N) = \text{MEU}(E_1, \ldots, E_N) - \text{MEU}(\emptyset)$.

(f) [2 pts] To compute $\text{MEU}(E_1, \ldots, E_N)$, we'll need one or more distribution(s) over X_N.
Which of these computations will generate the necessary distribution(s) over X_N?

- ○ Run the forward algorithm once, with no modifications.
- ● Run the forward algorithm 2^N times, with no modifications.
- ○ Run the forward algorithm once, skipping all observation updates.
- ○ Run the forward algorithm 2^N times, skipping all observation updates.

We don't know what the evidence is, so we have to run the forward algorithm once for every possible setting of the evidence $P(X_N | e_1, \ldots, e_N)$.

(g) [1 pt] To compute $\text{MEU}(E_1, \ldots, E_N)$, which other distribution do we need?

○ $P(E_1)$ ○ $P(E_N)$ ● $P(E_1, \ldots, E_N)$ ○ $P(E_1, \ldots, E_N | X_N)$

We need to weight each $\text{MEU}(e_1, \ldots, e_N)$ by the corresponding probability of evidence, $P(E_1, \ldots, E_N)$.