CSCD84: Artificial Intelligence

Worksheet Solution: Variable Elimination, HMM, and Particle Filtering

Q1. Bayes Nets: Variable Elimination

$\mathbb{P}(A)$	
$+a$	0.25
$-a$	0.75

$\mathbb{P}(B \mid A)$	$+b$	$-b$
$+a$	0.5	0.5
$-a$	0.25	0.75

$\mathbb{P}(C \mid A)$	$+c$	$-c$
$+a$	0.2	0.8
$-a$	0.6	0.4

$\mathbb{P}(D \mid B)$	$+b$	$-b$
$+b$	0.6	0.4
$-b$	0.8	0.2

$\mathbb{P}(E \mid B)$	$+c$	$-c$
$+b$	0.25	0.75
$-b$	0.1	0.9

1.a. Using the Bayes' Net and conditional probability tables above, calculate the following quantities:

1.a. i. $\mathbb{P}(+b \mid +a)$

Answer. $\mathbb{P}(+b \mid +a) = 1/2$

1.a. ii. $\mathbb{P}(+a, +b)$

Answer. $\mathbb{P}(+a, +b) = 1/2 \times 1/4 = 1/8$

1.a. iii. $\mathbb{P}(+a \mid +b)$

Answer. $\mathbb{P}(+a \mid +b) = \frac{\mathbb{P}(+a,+b)}{\mathbb{P}(+b)} = \frac{\mathbb{P}(+a,+b)}{\mathbb{P}(+b,+a)+\mathbb{P}(+b,-a)} = \frac{\mathbb{P}(+a,+b)}{\mathbb{P}(+b|+a)\mathbb{P}(+a)+\mathbb{P}(+b|-a)\mathbb{P}(-a)} = \frac{1/8}{1/2\times1/4+1/4\times3/4} = 2/5$

1.b. Now we are going to consider variable elimination in the Bayes' Net above.

1.b. i. Assume we have the evidence $+c$ and wish to calculate $\mathbb{P}(E \mid +c)$. What factors do we have initially?

Answer. $\mathbb{P}(A), \mathbb{P}(B \mid A), \mathbb{P}(+c \mid A), \mathbb{P}(D \mid B)$, and $\mathbb{P}(E \mid B)$.

1.b. ii. If we eliminate variable B, we create a new factor. What probability does that factor correspond to and what is the equation to calculate that factor?

Answer. $F_1(A, D, E) \sum_{\hat{b}\in\{+b,-b\}} \mathbb{P}(\hat{b} \mid A)\mathbb{P}(D \mid \hat{b})\mathbb{P}(E \mid \hat{b}) = \mathbb{P}(D, E \mid A)$

1.b. iii. After eliminating variable B, what are the new set of factors? As in **1.b. ii.**, write the probabilities that the factors represent. For each factor, also provide its size.

Answer. The new set of factors is $\mathbb{P}(A), \mathbb{P}(+c \mid A)$, and $\mathbb{P}(D, E \mid A)$.

- $\mathbb{P}(A)$ with size 2.
- $\mathbb{P}(+c \mid A)$ with size 2.
- $\mathbb{P}(D, E \mid A)$ with size $2 \times 2 \times 2 = 8$.

1.b. iv. Now assume we have the evidence $-c$ and are trying to calculate $\mathbb{P}(A \mid -c)$. What is the most memory efficient elimination ordering? If more than one ordering is most memory efficient, provide any one of them.

Answer. Starting from B, as we saw in the previous part, would generate a factor with size 8. Furthermore, if B is the second variable that we eliminate, the maximum factor size would be 4. Finally, if B is the third variable that we eliminate, the maximum factor size would be 4., as well (which is the input size). Hence, (D, B, E), (E, B, D), (D, E, B), and (E, D, B) are equally memory efficient.

1.b. v. Once we have run variable elimination and have $f(A, -c)$, how do we calculate $\mathbb{P}(+a \mid -c)$?

Answer. $\mathbb{P}(+a \mid -c) = \frac{f(+a,-c)}{f(+a,-c)+f(-a,-c)}$

Q2. Consider an HMM with the following Initial Distribution, Transition Probabilities, and Emission (Hidden Variable) Conditional Probability Table (CPT).

Initial Distribution

State	
$A_1 =$ true	0.99

Transition Probabilities

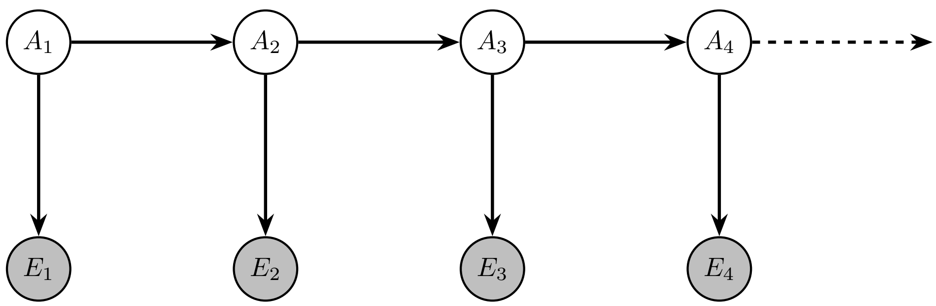
	$A_{t+1} =$ true	$A_{t+1} =$ false
$A_t =$ true	0.99	0.01
$A_t =$ false	0.01	0.99

Emission Probabilities

State	E_t	$\mathbb{P}(E_t \mid A_t)$
$A_t =$ true	1	0.2
$A_t =$ true	0	0.8
$A_t =$ false	1	0.9
$A_t =$ false	0	0.1

2.a. Draw the HMM implied by the CPTs above.

Answer.



2.b. Calculate $\mathbb{P}(A_4 = \text{true} \mid E_1 = 0,\ E_2 = 1,\ E_3 = 0)$.

Answer. For a better understanding of inference in HMM, we provide multiple solutions to this question.

Answer I. (Prediction, filtering, and linear algebra) As we discussed during the lectures, this is known as prediction task in HMM, *i.e.*, predicting the probability of future state given the current and past observations. As the HMM graph suggests, due to the Markov property of HMMs, given the current state, the future is independent from all past states and the past and current observations. Intuitively speaking, this fact hints us that we should first find the current state's distribution given the current and past observations, and then use it to predict the future state's distribution. Mathematically speaking,

$$\begin{aligned}
\mathbb{P}(A_4 \mid E_1 = 0, E_2 = 1, E_3 = 0) &= \sum_{a_3} \mathbb{P}(A_4, a_3 \mid E_1 = 0, E_2 = 1, E_3 = 0) \\
&\overset{(*)}{=} \sum_{a_3} \mathbb{P}(A_4 \mid a_3, E_1 = 0, E_2 = 1, E_3 = 0)\mathbb{P}(a_3 \mid E_1 = 0, E_2 = 1, E_3 = 0) \\
&\overset{(**)}{=} \sum_{a_3} \mathbb{P}(A_4 \mid a_3)\mathbb{P}(a_3 \mid E_1 = 0, E_2 = 1, E_3 = 0) \\
&\overset{(***)}{=} \sum_{a_3} \mathbb{P}(A_4 \mid a_3) f_{1:3}
\end{aligned}$$

where $(*)$ is due to law of probability, $(**)$ is due to the Markov property of HMM, and $(***)$ is derived by replacing the second term with the filtering notation we introduced in the lecture.

In lecture, we had seen that

$$f_{1:t+1} = \alpha_{t+1}\mathbb{P}(e_{t+1} \mid A_{t+1}) \sum_{a_t} \mathbb{P}(a_t \mid e_{1:t})\mathbb{P}(A_{t+1} \mid a_t) = \alpha_{t+1} O_{t+1} T^{\mathsf{T}} f_{1:t},$$

where

$$T = \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix}, O_{t+1} = \begin{bmatrix} \mathbb{P}(e_{t+1} \mid A_{t+1} = \text{true}) & 0 \\ 0 & \mathbb{P}(e_{t+1} \mid A_{t+1} = \text{false}) \end{bmatrix},$$

and

$$f_{1:1} = \mathbb{P}(A_1 \mid E_1 = 0) = \alpha_1 \mathbb{P}(A_1)\mathbb{P}(E_1 = 0 \mid A_1) = \alpha_1 O_1 \mathbb{P}(A_1),$$

and α_t's are just some normalizing scalars.

Let's put all pieces together.

$$\begin{aligned}
\mathbb{P}(A_4 \ \mid E_1 = 0, E_2 = 1, E_3 = 0) &= \sum_{a_3} \mathbb{P}(A_4 \mid a_3) f_{1:3} = T^\mathsf{T} f_{1:3} = T^\mathsf{T} \alpha_3 O_3 T^\mathsf{T} f_{1:2} \\
&= T^\mathsf{T} \alpha_3 O_3 T^\mathsf{T} \alpha_2 O_2 T^\mathsf{T} f_{1:1} = T^\mathsf{T} \alpha_3 O_3 T^\mathsf{T} \alpha_2 O_2 T^\mathsf{T} \alpha_1 O_1 \mathbb{P}(A_1) = \alpha T^\mathsf{T} O_3 T^\mathsf{T} O_2 T^\mathsf{T} O_1 \mathbb{P}(A_1) \\
&= \alpha \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} 0.8 & 0 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} 0.2 & 0 \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} 0.8 & 0 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} 0.99 \\ 0.01 \end{bmatrix} \\
&= \alpha \begin{bmatrix} 0.12303087 \\ 0.00218383 \end{bmatrix} = \frac{1}{0.12303087 + 0.00218383} \begin{bmatrix} 0.12303087 \\ 0.00218383 \end{bmatrix} = \begin{bmatrix} 0.9825593 \\ 0.0174407 \end{bmatrix}.
\end{aligned}$$

Note that we do not need to directly find the value of α by calculating the multiplication $\alpha_1\alpha_2\alpha_3$. α's job is to transform $\begin{bmatrix} 0.12303087 \\ 0.00218383 \end{bmatrix}$ into a valid distribution. Hence, α must be equal to $\frac{1}{0.12303087+0.00218383}$.

Therefor, $\mathbb{P}(A_4 = \text{true} \mid E_1 = 0, E_2 = 1, E_3 = 0) = 0.9825593$.

Answer II. (VE)

$$\begin{aligned}
\mathbb{P}(A_4 \ \mid E_1 = 0, E_2 = 1, E_3 = 0) &= \alpha \mathbb{P}(A_4, E_1 = 0, E_2 = 1, E_3 = 0) \\
&= \alpha \sum_{A_1, A_2, A_3} \mathbb{P}(A_4, A_3, A_2, A_1, E_1 = 0, E_2 = 1, E_3 = 0) \\
&= \alpha \sum_{A_3} \mathbb{P}(A_4 \mid A_3)\mathbb{P}(E_3 = 0 \mid A_3) \sum_{A_2} \mathbb{P}(E_2 = 1 \mid A_2)\mathbb{P}(A_3 \mid A_2) \sum_{A_1} \mathbb{P}(A_1)\mathbb{P}(E_1 = 0 \mid A_1)\mathbb{P}(A_2 \mid A_1)
\end{aligned}$$

We have two options:

Option 1 Rewriting the above equation as matrix multiplication:

$$\begin{aligned}
&\alpha \sum_{A_3} \mathbb{P}(A_4 \mid A_3)\mathbb{P}(E_3 = 0 \mid A_3) \sum_{A_2} \mathbb{P}(E_2 = 1 \mid A_2)\mathbb{P}(A_3 \mid A_2) \sum_{A_1} \mathbb{P}(A_1)\mathbb{P}(E_1 = 0 \mid A_1)\mathbb{P}(A_2 \mid A_1) \\
&= \alpha T^\mathsf{T} O_3 T^\mathsf{T} O_2 T^\mathsf{T} O_1 \mathbb{P}(A_1),
\end{aligned}$$

which would be exactly the same as what we had in the previous approach.

Option 2 Finding the factors:

$$\begin{aligned}
&\alpha \sum_{A_3} \mathbb{P}(A_4 \mid A_3)\mathbb{P}(E_3 = 0 \mid A_3) \sum_{A_2} \mathbb{P}(E_2 = 1 \mid A_2)\mathbb{P}(A_3 \mid A_2) \sum_{A_1} \mathbb{P}(A_1)\mathbb{P}(E_1 = 0 \mid A_1)\mathbb{P}(A_2 \mid A_1) \\
&= \alpha \sum_{A_3} \mathbb{P}(A_4 \mid A_3)\mathbb{P}(E_3 = 0 \mid A_3) \sum_{A_2} \mathbb{P}(E_2 = 1 \mid A_2)\mathbb{P}(A_3 \mid A_2) F_1(A_2, E_1 = 0) \\
&= \alpha \sum_{A_3} \mathbb{P}(A_4 \mid A_3)\mathbb{P}(E_3 = 0 \mid A_3) F_2(A_3, E_1 = 0, E_2 = 1) \\
&= \alpha F_3(A_4, E_1 = 0, E_2 = 1, E_3 = 0),
\end{aligned}$$

where,

$$F_1(A_2, E_1 = 0) = \begin{bmatrix} 0.78409 \\ 0.00891 \end{bmatrix},$$

$$F_2(A_3, E_1 = 0, E_2 = 1) = \begin{bmatrix} 0.15533001 \\ 0.00950699 \end{bmatrix},$$

$$F_3(A_4, E_1 = 0, E_2 = 1, E_3 = 0) = \begin{bmatrix} 0.12303087 \\ 0.00218383 \end{bmatrix}.$$

Hence,

$$\mathbb{P}(A_4 \mid E_1 = 0, E_2 = 1, E_3 = 0) = \alpha \begin{bmatrix} 0.12303087 \\ 0.00218383 \end{bmatrix} = \frac{1}{0.12303087 + 0.00218383} \begin{bmatrix} 0.12303087 \\ 0.00218383 \end{bmatrix}$$
$$= \begin{bmatrix} 0.9825593 \\ 0.0174407 \end{bmatrix}.$$

2.c. What is the probability of observing the emission sequence ($E_1 = 0,\ E_2 = 1,\ E_3 = 0$)?

Answer. There are multiple ways to find $\mathbb{P}(E_1 = 0, E_2 = 1, E_3 = 0)$.

Answer I. (Be lazy) $\mathbb{P}(E_1 = 0, E_2 = 1, E_3 = 0)$ is the inverse of the normalizing factor, α, in the previous part. Hence, $\mathbb{P}(E_1 = 0, E_2 = 1, E_3 = 0) = 0.12303087 + 0.00218383 \approx 0.125$.

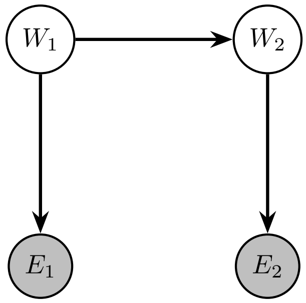
Answer II. (VE)

$$\mathbb{P}(E_1 = 0, E_2 = 1, E_3 = 0) = \sum_{A_1, A_2, A_3, A_4} \mathbb{P}(A_1, A_2, A_3, A_4, E_1 = 0, E_2 = 1, E_3 = 0)$$
$$= \sum_{A_1, A_2, A_3, A_4} \mathbb{P}(A_1)\mathbb{P}(E_1 = 0 \mid A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(E_2 = 1 \mid A_2)\mathbb{P}(A_3 \mid A_2)\mathbb{P}(E_3 = 0 \mid A_3)\mathbb{P}(A_4 \mid A_3)$$
$$= \sum_{A_4}\sum_{A_3} \mathbb{P}(A_4 \mid A_3)\mathbb{P}(E_3 = 0 \mid A_3) \sum_{A_2} \mathbb{P}(E_2 = 1 \mid A_2)\mathbb{P}(A_3 \mid A_2) \sum_{A_1} \mathbb{P}(A_1)\mathbb{P}(E_1 = 0 \mid A_1)\mathbb{P}(A_2 \mid A_1)$$
$$= [1, 1] \times T^\intercal O_3 T^\intercal O_2 T^\intercal O_1 \mathbb{P}(A_1).$$

Guess what? In the previous part, we had already found $T^\intercal O_3 T^\intercal O_2 T^\intercal O_1 \mathbb{P}(A_1)$. Substituting that in the equation above, we get

$$\mathbb{P}(E_1 = 0, E_2 = 1, E_3 = 0) = [1, 1] \begin{bmatrix} 0.12303087 \\ 0.00218383 \end{bmatrix} = 0.12303087 + 0.00218383 \approx 0.125.$$

Q3. Consider the following Hidden Markov Model. E_1 and E_2 are supposed to be shaded.



W_1	$\mathbb{P}(W_1)$
0	0.3
1	0.7

W_t	W_{t+1}	$\mathbb{P}(W_{t+1} \mid W_t)$
0	0	0.4
0	1	0.6
1	0	0.8
1	1	0.2

W_t	E_t	$\mathbb{P}(E_t \mid W_t)$
0	a	0.9
0	b	0.1
1	a	0.5
1	b	0.5

Suppose that we observe $E_1 = a$ and $E_2 = b$. Using the forward algorithm, compute the probability distribution $\mathbb{P}(W_2 \mid E_1 = a, E_2 = b)$ one step at a time.

3.a. Compute $\mathbb{P}(W_1, E_1 = a)$.

Answer. $\mathbb{P}(W_1, E_1 = a) = \mathbb{P}(E_1 = a \mid W_1)\mathbb{P}(W_1)$. Hence, $\mathbb{P}(W_1 = 0, E_1 = a) = 0.3 \times 0.9 = 0.27$ and $\mathbb{P}(W_1 = 1, E_1 = a) = 0.7 \times 0.5 = 0.35$.

3.b. Using the previous calculation, compute $\mathbb{P}(W_2, E_1 = a)$.

Answer.

$$\begin{aligned}\mathbb{P}(W_2, E_1 = a) &= \sum_{w_1} \mathbb{P}(W_2, w_1, E_1 = a) = \sum_{w_1} \mathbb{P}(W_2 \mid w_1, E_1 = a)\mathbb{P}(w_1, E_1 = a) \\ &= \sum_{w_1} \mathbb{P}(W_2 \mid w_1)\mathbb{P}(w_1, E_1 = a).\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{P}(W_2 = 0, E_1 = a) = 0.4 \times 0.27 + 0.8 \times 0.35 = 0.388 \\ \mathbb{P}(W_2 = 1, E_1 = a) = 0.6 \times 0.27 + 0.2 \times 0.35 = 0.232\end{aligned}$$

3.c. Using the previous calculation, compute $\mathbb{P}(W_2, E_1 = a, E_2 = b)$.

Answer.

$$\mathbb{P}(W_2, E_1 = a, E_2 = b) = \mathbb{P}(E_2 = b \mid W_2, E_1 = a)\mathbb{P}(W_2, E_1 = a) = \mathbb{P}(E_2 = b \mid W_2)\mathbb{P}(W_2, E_1 = a).$$

Hence,

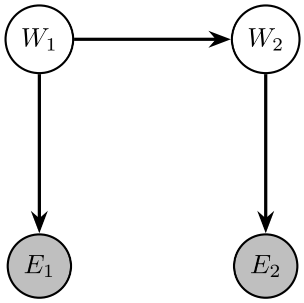
$$\begin{aligned}\mathbb{P}(W_2 = 0, E_1 = a, E_2 = b) = 0.1 \times 0.388 = 0.0388 \\ \mathbb{P}(W_2 = 1, E_1 = a, E_2 = b) = 0.5 \times 0.232 = 0.116\end{aligned}$$

3.d. Finally, compute $\mathbb{P}(W_2 | E_1 = a, E_2 = b)$.

Answer.

$$\mathbb{P}(W_2 | E_1 = a, E_2 = b) = \alpha \mathbb{P}(W_2, E_1 = a, E_2 = b) = \alpha \begin{bmatrix} 0.0388 \\ 0.116 \end{bmatrix} = \frac{1}{0.0388 + 0.116} \begin{bmatrix} 0.0388 \\ 0.116 \end{bmatrix} \approx \begin{bmatrix} 0.25 \\ 0.75 \end{bmatrix}.$$

Q4. Let's use Particle Filtering to estimate the distribution of $\mathbb{P}(W_2|E_1 = a, E_2 = b)$. Consider the following Hidden Markov Model. E_1 and E_2 are supposed to be shaded.



W_1	$\mathbb{P}(W_1)$
0	0.3
1	0.7

W_t	W_{t+1}	$\mathbb{P}(W_{t+1} \mid W_t)$
0	0	0.4
0	1	0.6
1	0	0.8
1	1	0.2

W_t	E_t	$\mathbb{P}(E_t \mid W_t)$
0	a	0.9
0	b	0.1
1	a	0.5
1	b	0.5

We start with two particles representing our distribution for W_1.

- $P_1 : W_1 = 0$
- $P_2 : W_1 = 1$

Use the following random numbers to run particle filtering:

$$[0.22, 0.05, 0.33, 0.20, 0.84, 0.54, 0.79, 0.66, 0.14, 0.96]$$

4.a. Observe: Compute the weight of the two particles after evidence $E_1 = a$.

Answer.

$$w(P_1) = P(E_1 = a|W_1 = 0) = 0.9,$$
$$w(P_2) = P(E_1 = a|W_1 = 1) = 0.5.$$

4.b. Resample: Using the random numbers, resample P_1 and P_2 based on the weights.

Answer. After normalizing the weights, the probability of picking P_1 and P_2 would be $\frac{0.9}{1.4} \approx 0.642$ and $\frac{0.5}{1.4} \approx 0.358$, respectively. We randomly sample two numbers from the uniform distribution. According to the given random sequence, the two numbers are 0.22 and 0.05. Since, both of them are less than 0.642, we pick P_1 for both of them, *i.e.*,

$$P_1 = \text{sample(weights}, 0.22) = 0,$$
$$P_2 = \text{sample(weights}, 0.05) = 0.$$

4.c. Predict: Sample P_1 and P_2 from applying the time update.

Answer. Given the transitional probabilities, $\mathbb{P}(W_2 \mid W_1 = 0) = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}$. We randomly sample two numbers from the uniform distribution. According to the given random sequence, the two numbers are 0.33 and 0.20. Since, both of them are less than 0.4, we pick $W_2 = 0$ for both of them, *i.e.*

$$P_1 = \text{sample}(\mathbb{P}(W_2|W_1 = 0), 0.33) = 0,$$
$$P_2 = \text{sample}(\mathbb{P}(W_2|W_1 = 0), 0.20) = 0.$$

4.d. Update: Compute the weight of the two particles after evidence $E_2 = b$.

Answer.

$$w(P_1) = \mathbb{P}(E_2 = b|W_2 = 0) = 0.1,$$
$$w(P_2) = \mathbb{P}(E_2 = b|W_2 = 0) = 0.1.$$

4.e. Resample: Using the random numbers, resample P_1 and P_2 based on the weights.

Answer. Since both particles have $W_2 = 0$, resampling will still leave us with two particles with $W_2 = 0$.

$$P_1 = 0,$$
$$P_2 = 0.$$

4.f. What is our estimated distribution for $\mathbb{P}(W_2 \mid E_1 = a, E_2 = b)$?

Answer. $\mathbb{P}(W_2 = 0 \mid E_1 = a, E_2 = b) = 2/2 = 1.$