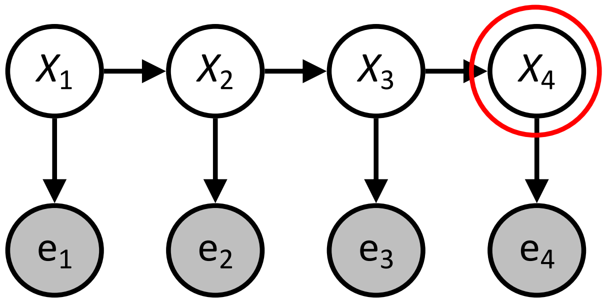
Warm-up as you walk in

- For the following Bayes net, write the query $P(X_4 \mid e_{1:4})$ in terms of the conditional probability tables associated with the Bayes net.

$P(X_4 \mid e_1, e_2, e_3, e_4) =$

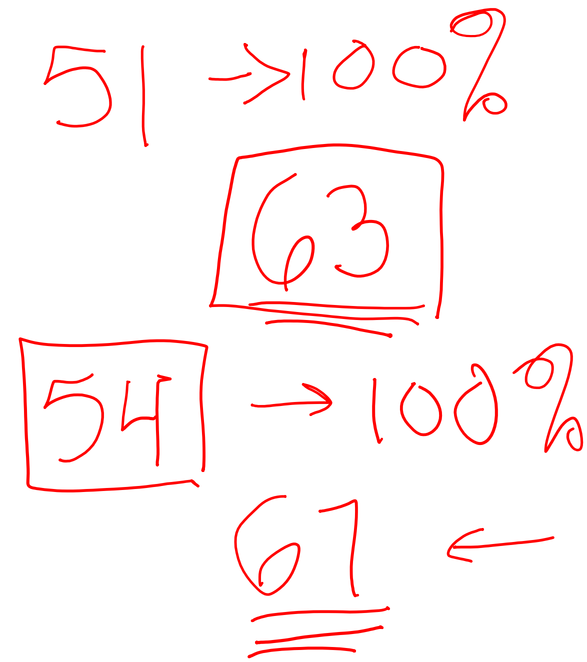


Announcements

TA applications: https://www.ugrad.cs.cmu.edu/ta/F23/

Assignments

- HW9
  - Due tonight, 10 pm
- HW10
  - Out next week, due 4/25, 10 pm
- P5
  - Out tonight, due Thursday 4/27, 10 pm



AI: Representation and Problem Solving

Hidden Markov Models

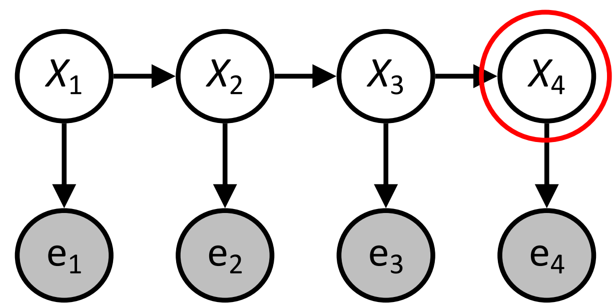
Instructors: Stephanie Rosenthal

Slide credits: CMU AI and http://ai.berkeley.edu

Warm-up as you walk in

- For the following Bayes net, write the query $P(X_4 \mid e_{1:4})$ in terms of the conditional probability tables associated with the Bayes net.

$$P(X_4 \mid e_1,e_2,e_3,e_4) = \propto P(X_4, e_1, e_2, e_3, e_4)$$



$$\sum_{X_1,X_2,X_3} P(e_1|X_1)P(e_2|X_2)P(e_3|X_3)P(X_1)\,P(X_2|X_1)P(X_3|X_2)\,P(X_4|X_3)\,P(e_4|X_4)$$

Reasoning over Time or Space

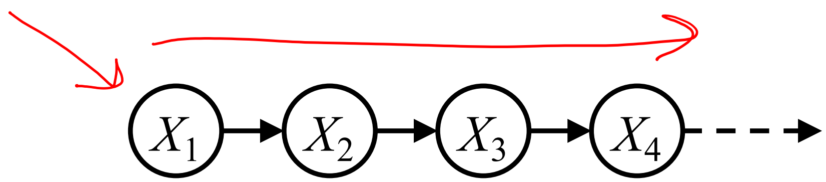
Often, we want to reason about a sequence of observations

- Speech recognition
- Robot localization
- User attention
- Medical monitoring

Need to introduce time (or space) into our models

Markov Chains

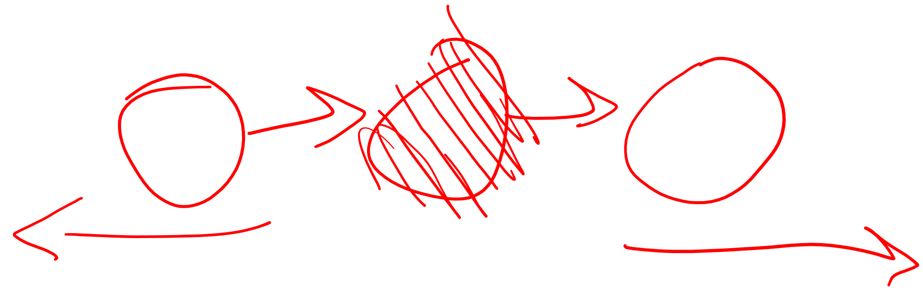
- Value of X at a given time is called the state

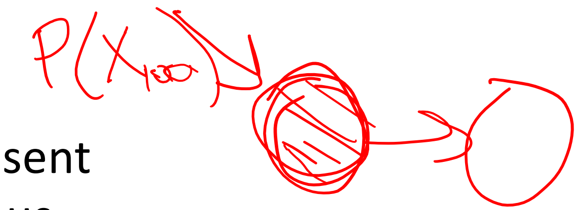


$$P(X_1) \qquad P(X_t|X_{t-1})$$

- Parameters: called transition probabilities or dynamics, specify how the state evolves over time (also, initial state probabilities)
- Stationarity assumption: transition probabilities the same at all times
- Same as MDP transition model, but no choice of action

Conditional Independence





Basic conditional independence:

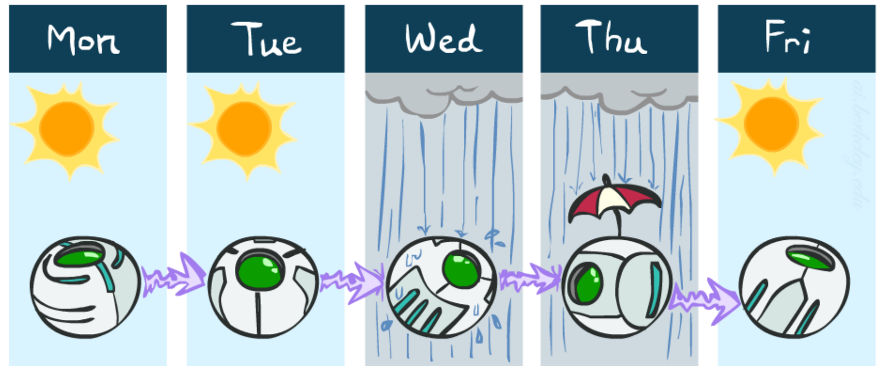
- Past and future independent given the present
- Each time step only depends on the previous
- This is called the (first order) Markov property

Note that the chain is just a (growable) BN

- We can always use generic BN reasoning on it if we truncate the chain at a fixed length

Example: Markov Chain Weather

States: X = {rain, sun}



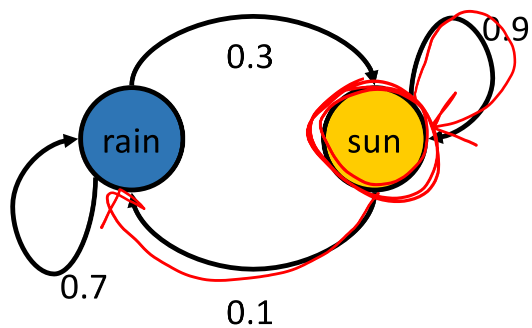
- Initial distribution: 1.0 sun

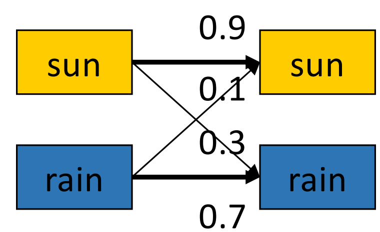
- CPT $P(X_t \mid X_{t-1})$:

X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

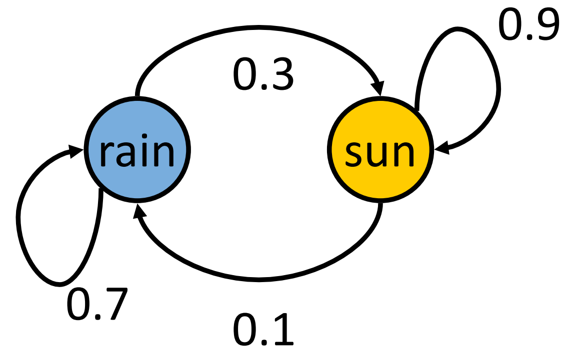
$P(X_2) \propto P(X_2, X_1)$

Two new ways of representing the same CPT





Example: Markov Chain Weather

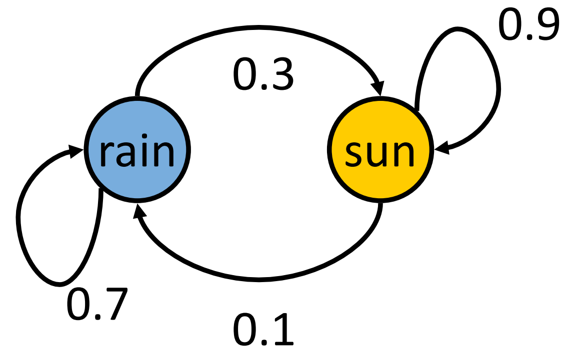


Initial distribution: $P(X_1 = sun) = 1.0$

What is the probability distribution after one step?

$P(X_2 = sun) = ?$

Example: Markov Chain Weather



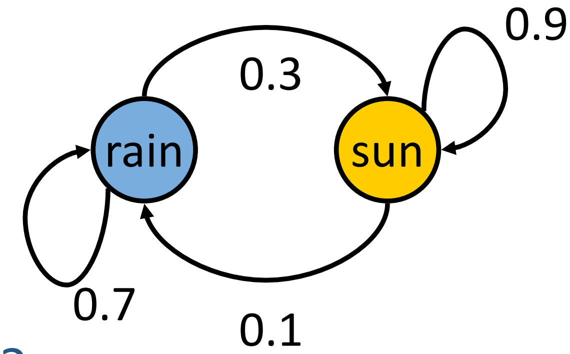
Initial distribution: $P(X_1 = sun) = 1.0$

What is the probability distribution after one step?

$P(X_2 = sun) = ?$

$$
\begin{aligned}
P(X_2 = sun) &= \sum_{x_1} P(X_1 = x_1, X_2 = sun) \\
&= \sum_{x_1} P(X_2 = sun \mid X_1 = x_1) P(X_1 = x_1) \\
&= P(X_2 = sun \mid X_1 = sun) P(X_1 = sun) + \\
&\quad P(X_2 = sun \mid X_1 = rain) P(X_1 = rain) \\
&= 0.9 \cdot 1.0 + 0.3 \cdot 0.0 = 0.9
\end{aligned}
$$

Poll 1



Initial distribution: $P(X_2 = sun) = 0.9$

What is the probability distribution after the next step?

$P(X_3 = sun) = ?$

A) 0.81

B) 0.84

C) 0.9

D) 1.0

E) 1.2

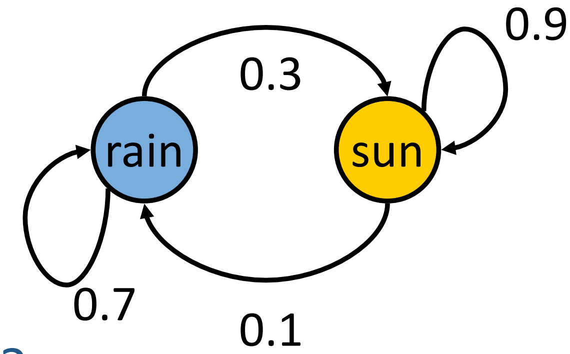
$$P(X_3) = \sum_{x_2} P(x_2, X_3)$$

$$= \sum_{x_2} P(x_2) P(X_3 \mid x_2)$$

$x_2 = s$: $.9(\langle .9, .1 \rangle) = \langle .81, .09 \rangle$

$x_2 = r$: $.1(\langle .3, .7 \rangle) = \langle .03, .07 \rangle$

Poll 1



Initial distribution: $P(X_2 = sun) \ = 0.9$

What is the probability distribution after the next step?

$P(X_3 = sun) \ = ?$

A) 0.81

B) 0.84

C) 0.9

D) 1.0

E) 1.2

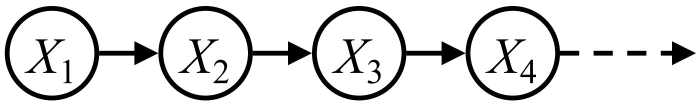
$$P(X_3 = sun) = \sum_{x_2} P(X_3 = sun, X_2 = x_2)$$

$$= \sum_{x_2} P(X_3 = sun \mid X_2 = x_2)\, P(X_2 = x_2)$$

$$= 0.9 \cdot 0.9 + 0.3 \cdot 0.1$$

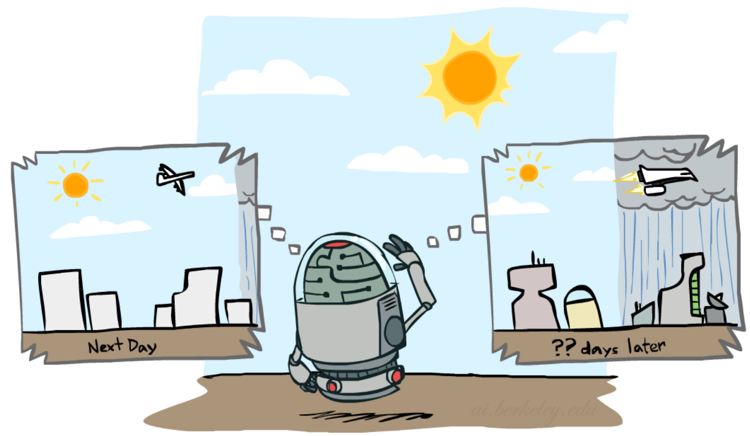
$$= 0.81 + 0.03 = 0.84$$

Markov Chain Inference

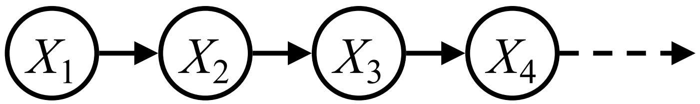


If you know the transition probabilities, $P(X_t \mid X_{t-1})$, and you know $P(X_4)$, write an equation to compute $P(X_5)$.

$$\sum_{t-1} P(X_{t-1}) P(X_t \mid X_{t-1})$$



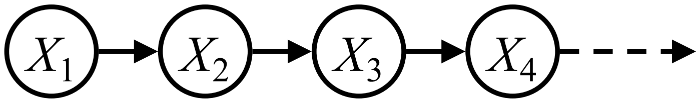
Markov Chain Inference



If you know the transition probabilities, $P(X_t \mid X_{t-1})$, and you know $P(X_4)$, write an equation to compute $P(X_5)$.

$$P(X_5) = \sum_{x_4} P(x_4, X_5)$$
$$= \sum_{x_4} P(X_5 \mid x_4)P(x_4)$$

Markov Chain Inference



If you know the transition probabilities, $P(X_t \mid X_{t-1})$, and you know $P(X_4)$, write an equation to compute $P(X_5)$.

$$
\begin{aligned}
P(X_5) &= \sum_{x_1,x_2,x_3,x_4} P(x_1, x_2, x_3, x_4, X_5) \\
&= \sum_{x_1,x_2,x_3,x_4} P(X_5 \mid x_4) P(x_4 \mid x_3) P(x_3 \mid x_2) P(x_2 \mid x_1) P(x_1) \\
&= \sum_{x_4} P(X_5 \mid x_4) \sum_{x_1,x_2,x_3} P(x_4 \mid x_3) P(x_3 \mid x_2) P(x_2 \mid x_1) P(x_1) \\
&= \sum_{x_4} P(X_5 \mid x_4) \sum_{x_1,x_2,x_3} P(x_1, x_2, x_3, x_4) \\
&= \sum_{x_4} P(X_5 \mid x_4) P(x_4)
\end{aligned}
$$

Weather prediction

States {rain, sun}

- Initial distribution $P(X_0)$

$P(X_0)$	
sun	rain
0.5	0.5

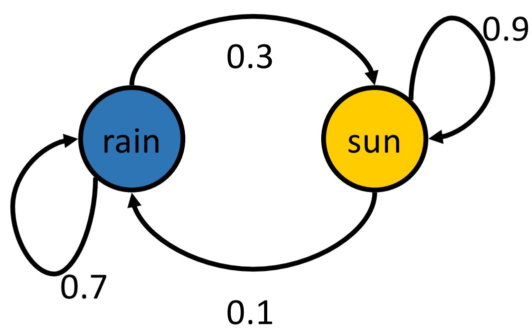
- Transition model $P(X_t|X_{t-1})$

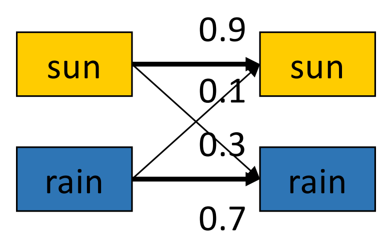
X_{t-1}	$P(X_t|X_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

- Initial distribution $P(X_0)$

- Transition model $P(X_t|X_{t-1})$

Two new ways of representing the same CPT

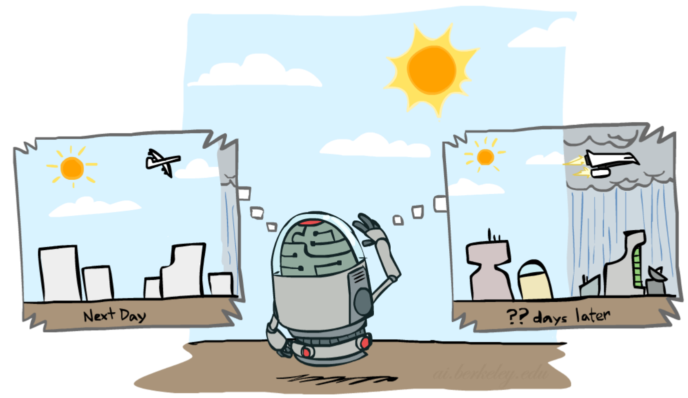




Weather prediction

Time 0: $P(X_0) = \langle 0.5, 0.5 \rangle$

X_{t-1}	$P(X_t \mid X_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7



What is the weather like at time 1?

$$P(X_1) =$$

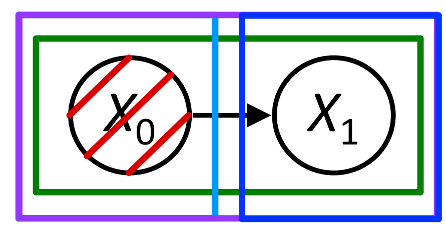
$$\sum_{x_0} P(X_0 = x_0, X_1)$$

$$= \sum_{x_0} P(X_1 \mid X_0 = x_0) P(X_0 = x_0)$$

$$= 0.5\langle 0.9, 0.1 \rangle + 0.5\langle 0.3, 0.7 \rangle$$

$$= \langle 0.6, 0.4 \rangle$$

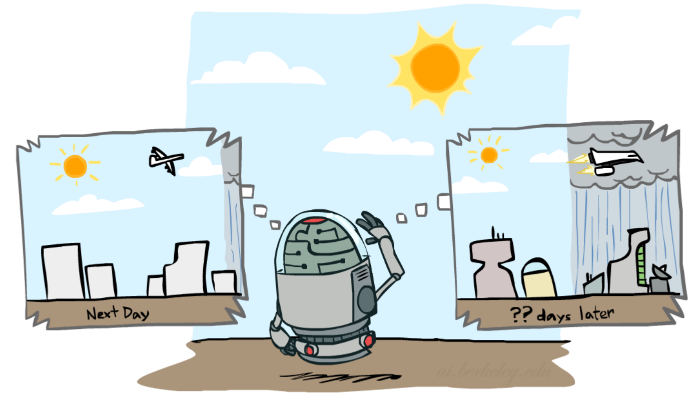
s	.6
r	.4



Weather prediction, contd.

Time 1: $P(X_1) = \langle 0.6, 0.4 \rangle$

X_{t-1}	$P(X_t \mid X_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7



What is the weather like at time 2?

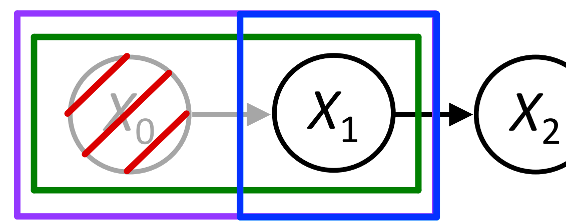
$P(X_2) =$

$$\sum_{x_1} P(X_1 = x_1, X_2)$$

$$= \sum_{x_1} P(X_2 \mid X_1 = x_1) P(X_1 = x_1)$$

$$= 0.6\langle 0.9, 0.1 \rangle + 0.4\langle 0.3, 0.7 \rangle$$

$$= \langle 0.66, 0.34 \rangle$$



Weather prediction, contd.

Time 2: $P(X_2) = \langle 0.66, 0.34 \rangle$

X_{t-1}	$P(X_t \mid X_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

Time 2: $P(X_2) = \langle 0.66, 0.34 \rangle$

What is the weather like at time 3?

$P(X_3) =$

What is the weather like at time 3?

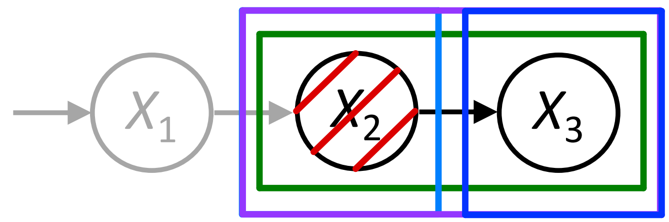
$P(X_3) =$

$$\sum_{x_2} P(X_2 = x_2, X_3)$$

$$= \sum_{x_2} P(X_3 \mid X_2 = x_2) P(X_2 = x_2)$$

$$= 0.66\langle 0.9, 0.1 \rangle + 0.34\langle 0.3, 0.7 \rangle$$

$$= \langle 0.696, 0.304 \rangle$$



Forward algorithm (simple form)

What is the state at time t?

$$P(X_t) = \sum_{x_{t-1}} P(X_{t-1} = x_{t-1}, X_t)$$

$$= \sum_{x_{t-1}} P(X_t | X_{t-1} = x_{t-1}) P(X_{t-1} = x_{t-1})$$

Transition model

Probability from previous iteration

Iterate this update starting at t=0

Prediction with Markov chains

As time passes, uncertainty "accumulates"

(Transition model: ghosts usually go clockwise)

<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
<0.01	<0.01	1.00	<0.01	<0.01	<0.01
<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

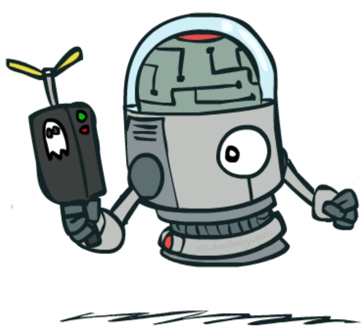
T = 1

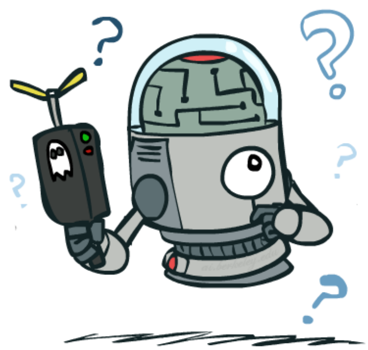
<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
<0.01	<0.01	0.06	<0.01	<0.01	<0.01
<0.01	0.76	0.06	0.06	<0.01	<0.01
<0.01	<0.01	0.06	<0.01	<0.01	<0.01

T = 2

0.05	0.01	0.05	<0.01	<0.01	<0.01
0.02	0.14	0.11	0.35	<0.01	<0.01
0.07	0.03	0.05	<0.01	0.03	<0.01
0.03	0.03	<0.01	<0.01	<0.01	<0.01

T = 5





Observations Reduce Uncertainty

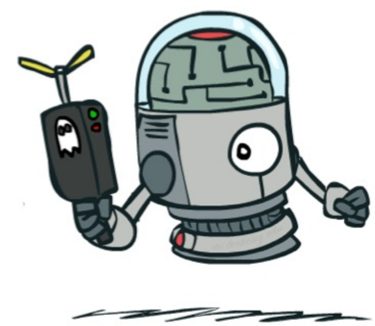
As we get observations, beliefs get reweighted, uncertainty "decreases"

0.05	0.01	0.05	<0.01	<0.01	<0.01
0.02	0.14	0.11	0.35	<0.01	<0.01
0.07	0.03	0.05	<0.01	0.03	<0.01
0.03	0.03	<0.01	<0.01	<0.01	<0.01

Before observation

<0.01	<0.01	<0.01	<0.01	0.02	<0.01
<0.01	<0.01	<0.01	0.83	0.02	<0.01
<0.01	<0.01	0.11	<0.01	<0.01	<0.01
<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

After observation



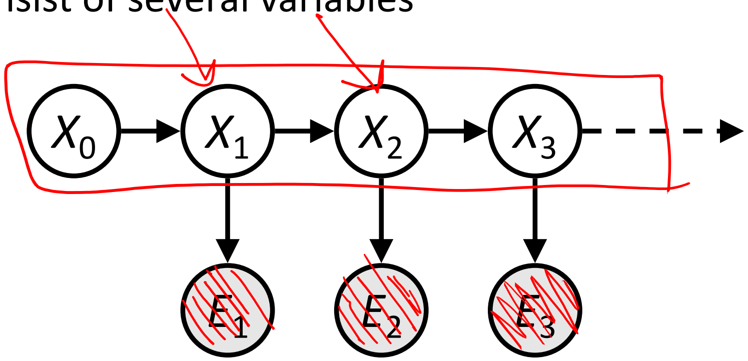
Hidden Markov Models

Hidden Markov Models

Usually the true state is not observed directly

Hidden Markov models (HMMs)

- Underlying Markov chain over states X
- You observe evidence E at each time step
- X_t is a single discrete variable; E_t may be continuous and may consist of several variables



Real HMM Examples

Speech recognition HMMs:

- Observations are acoustic signals (continuous valued)
- States are specific positions in specific words (so, tens of thousands)

Machine translation HMMs:

- Observations are words (tens of thousands)
- States are translation options

Robot tracking:

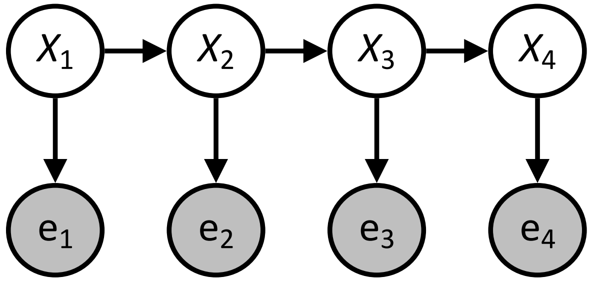
- Observations are range readings (continuous)
- States are positions on a map (continuous)

Molecular biology:

- Observations are nucleotides ACGT
- States are coding/non-coding/start/stop/splice-site etc.

HMM as a Bayes Net Warm-up

- For the following Bayes net, write the query $P(X_4 \mid e_{1:4})$ in terms of the conditional probability tables associated with the Bayes net.

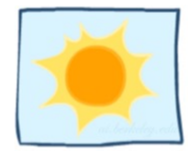


$P(X_4 \mid e_1, e_2, e_3, e_4) =$

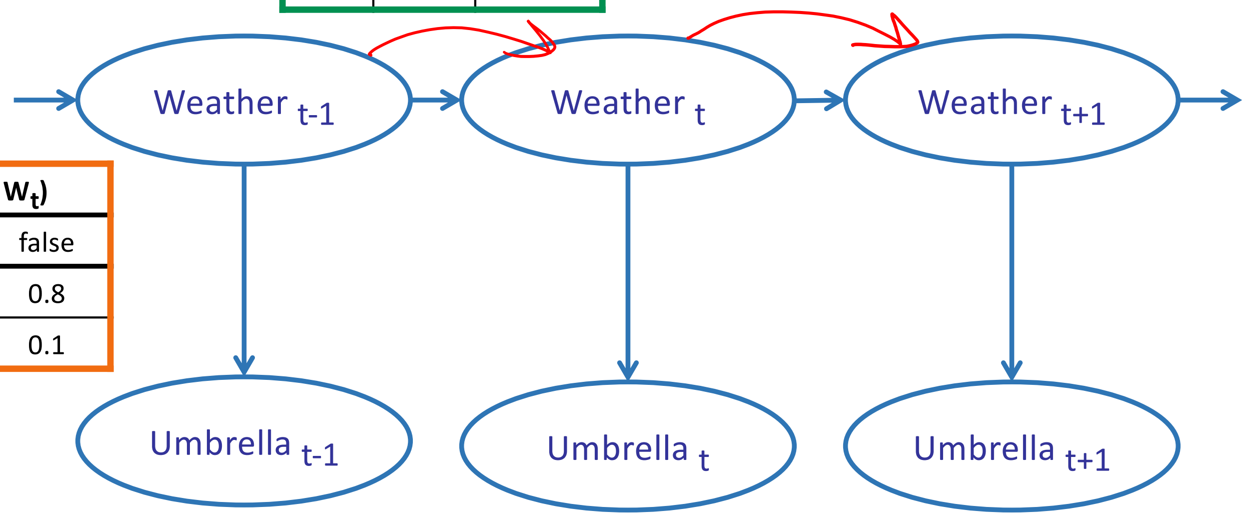
Example: Weather HMM

An HMM is defined by:

- Initial distribution: $P(X_0)$
- Transition model: $P(X_t \mid X_{t-1})$
- Sensor model: $P(E_t \mid X_t)$



W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7



W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

HMM as Probability Model

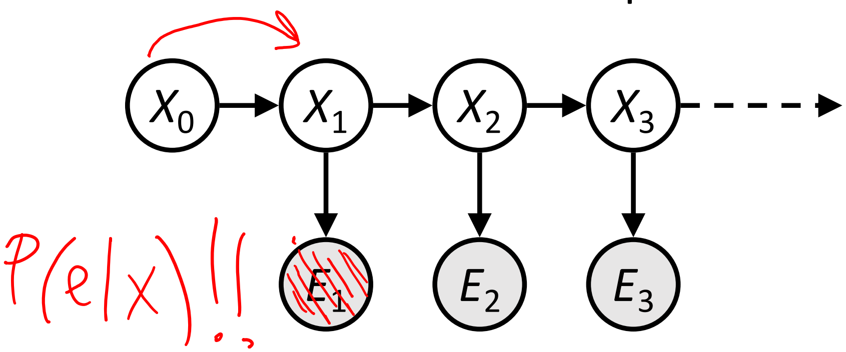
- Joint distribution for Markov model:

$$P(X_0,\ldots, X_T) = P(X_0) \prod_{t=1:T} P(X_t \mid X_{t-1})$$

- Joint distribution for hidden Markov model:

$$P(X_0, X_1,E_1, \ldots, X_T,E_T) = P(X_0) \prod_{t=1:T} P(X_t \mid X_{t-1})\, P(E_t \mid X_t)$$

- Future states are independent of the past given the present
- Current evidence is independent of everything else given the current state
- Are evidence variables independent of each other?



Useful notation: $X_{a:b} = X_a , X_{a+1}, \ldots, X_b$

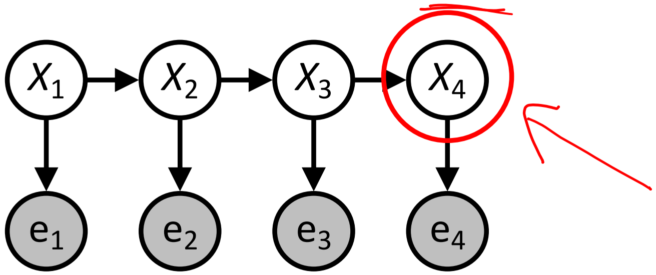
For example: $P(X_{1:2} \mid e_{1:3}) = P(X_1, X_2, \mid e_1 , e_2, e_3)$

HMM Queries

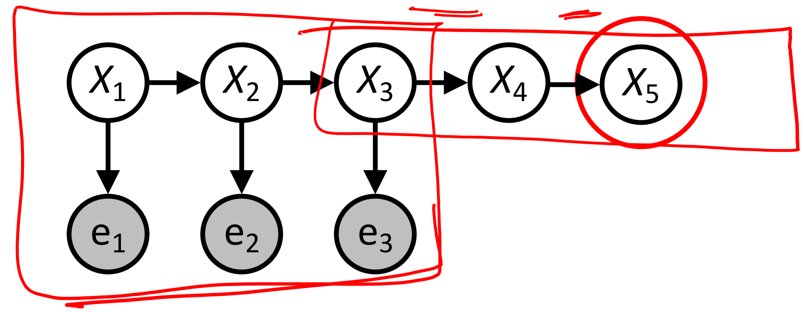
e_1, e_2, e_3, e_4 e_5

$e_{1:5}$

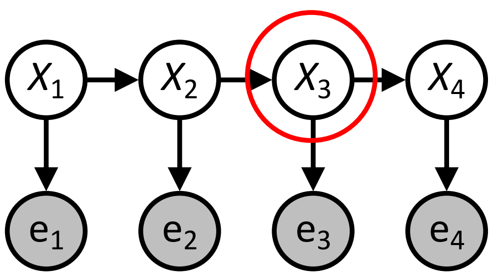
Filtering: $P(X_t | e_{1:t})$



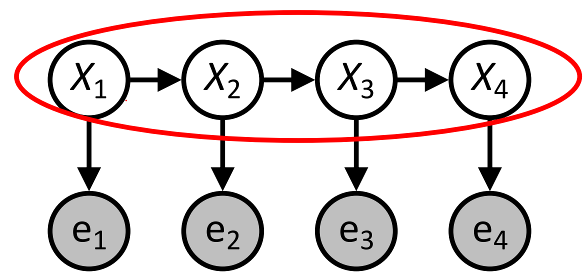
Prediction: $P(X_{t+k} | e_{1:t})$



Smoothing: $P(X_k | e_{1:t})$, $k<t$



Explanation: $P(X_{1:t} | e_{1:t})$



Filtering **Algorithm**

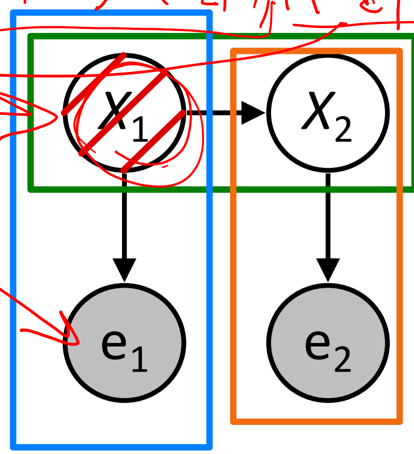
Query: What is the current state, given all of the current and past evidence?

Marching **forward** through the HMM network

$$\sum_{x_1} P(X_1 \mid e_1) P(X_2 \mid X_1) P(e_2 \mid X_2) = P(X_2 \mid e_1) P(e_2 \mid X_2) = P(X_2 \mid e_1, e_2)$$

$P(X_1)$

$P(e_1 \mid X_1)$

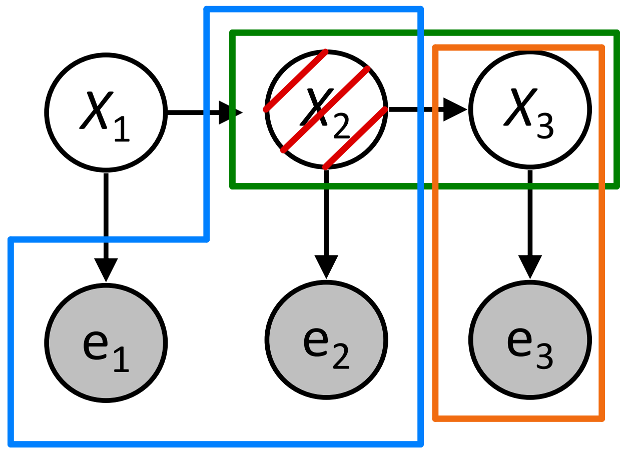


$P(X_1 \mid e_1)$ ← query

Filtering **Algorithm**

Query: What is the current state, given all of the current and past evidence?

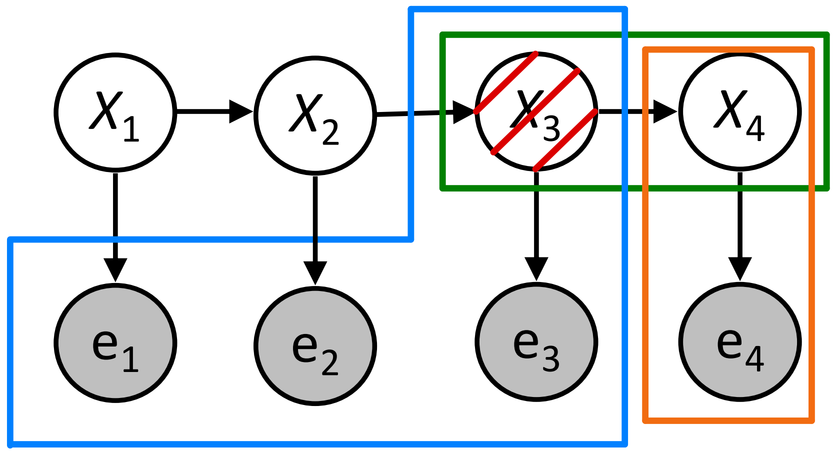
Marching **forward** through the HMM network



Filtering **Algorithm**

Query: What is the current state, given all of the current and past evidence?

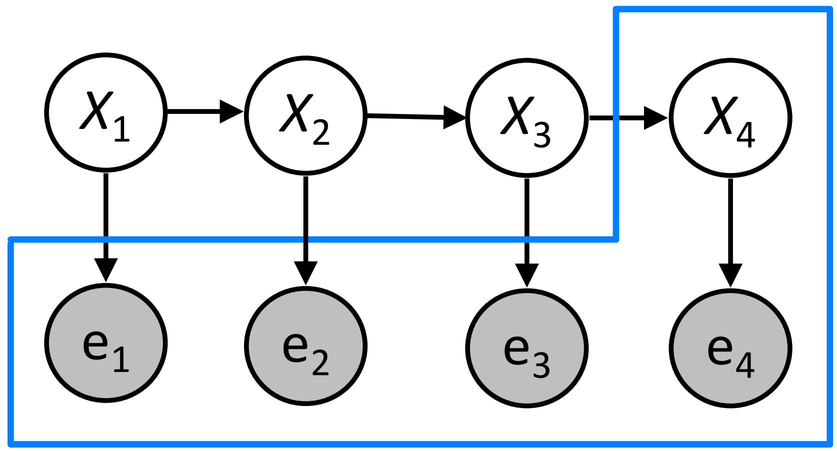
Marching **forward** through the HMM network



Filtering **Algorithm**

Query: What is the current state, given all of the current and past evidence?

Marching **forward** through the HMM network



Filtering **Algorithm** = Forward Algorithm

$$P(X_{t+1} | e_{1:t+1}) = \alpha\, P(e_{t+1} | X_{t+1}) \sum_{x_t} P(X_{t+1} | x_t)\, P(x_t | e_{1:t})$$

Normalize — α

Update — $P(e_{t+1} | X_{t+1})$

Predict — $\sum_{x_t} P(X_{t+1} | x_t)\, P(x_t | e_{1:t})$

$$\boldsymbol{f}_{1:t+1} = \text{FORWARD}(\boldsymbol{f}_{1:t}, e_{t+1})$$

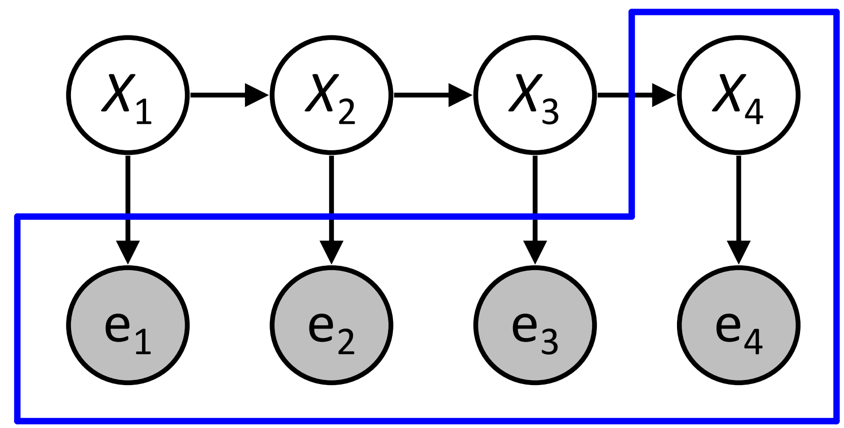
Filtering **Algorithm**

Query: What is the current state, given all of the current and past evidence?

Matching math with Bayes net

$$P(X_t \mid e_{1:t}) = P(X_t \mid e_t, e_{1:t-1})$$
$$= \alpha\, P(X_t, e_t \mid e_{1:t-1})$$

Def. of cond. probability with X_t, e_t



Filtering **Algorithm**

Query: What is the current state, given all of the current and past evidence?

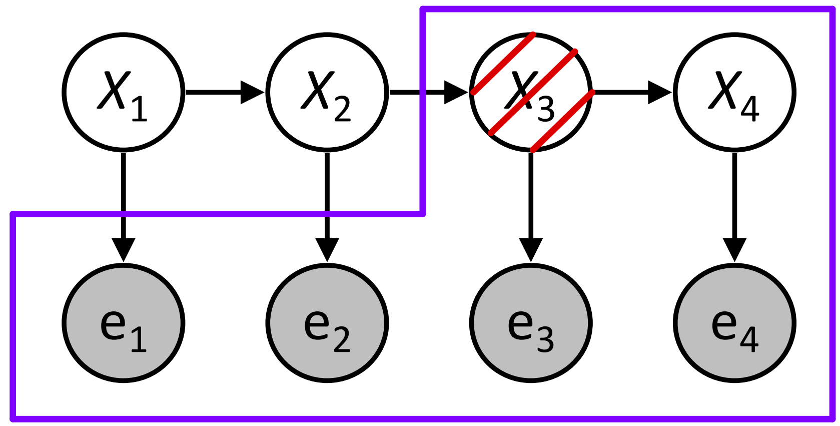
Matching math with Bayes net

$$P(X_t \mid e_{1:t}) = P(X_t \mid e_t, e_{1:t-1})$$

$$= \alpha\, P(X_t, e_t \mid e_{1:t-1})$$

$$= \alpha \sum_{x_{t-1}} P(x_{t-1}, X_t, e_t \mid e_{1:t-1})$$

Summation over variable X_{t-1}



Filtering **Algorithm**

Query: What is the current state, given all of the current and past evidence?

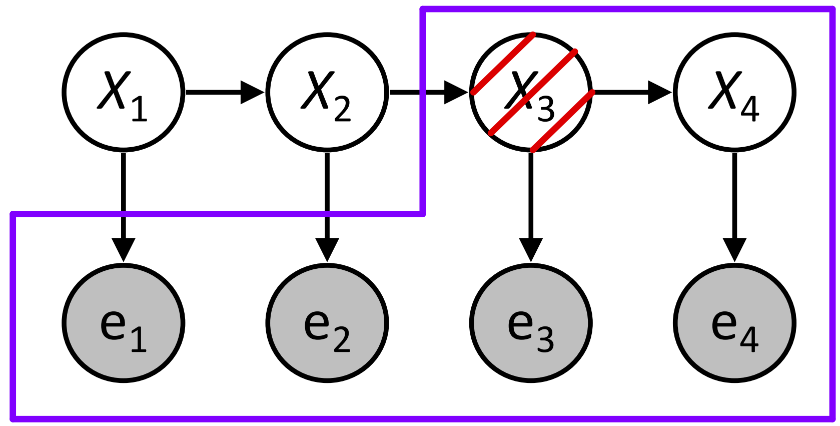
Matching math with Bayes net

$$P(X_t \mid e_{1:t}) = P(X_t \mid e_t, e_{1:t-1})$$

$$= \alpha\, P(X_t, e_t \mid e_{1:t-1})$$

$$= \alpha \sum_{x_{t-1}} P(x_{t-1}, X_t, e_t \mid e_{1:t-1})$$

$$= \alpha \sum_{x_{t-1}} P(x_{t-1} \mid e_{1:t-1})\, P(X_t \mid x_{t-1}, e_{1:t-1})\, P(e_t \mid X_t, x_{t-1}, e_{1:t-1})$$



Chain rule with x_{t-1}, X_t, and e_t

Filtering **Algorithm**

Query: What is the current state, given all of the current and past evidence?

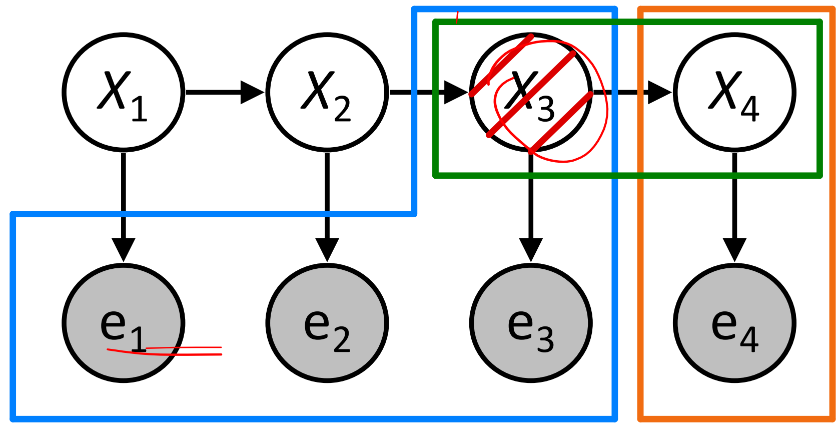
Matching math with Bayes net

$$P(X_t \mid e_{1:t}) = P(X_t \mid e_t, e_{1:t-1})$$

$$= \alpha\, P(X_t, e_t \mid e_{1:t-1})$$

$$= \alpha \sum_{x_{t-1}} P(x_{t-1}, X_t, e_t \mid e_{1:t-1})$$

$$= \alpha \sum_{x_{t-1}} P(x_{t-1} \mid e_{1:t-1})\, P(X_t \mid x_{t-1}, e_{1:t-1})\, P(e_t \mid X_t, x_{t-1}, e_{1:t-1})$$



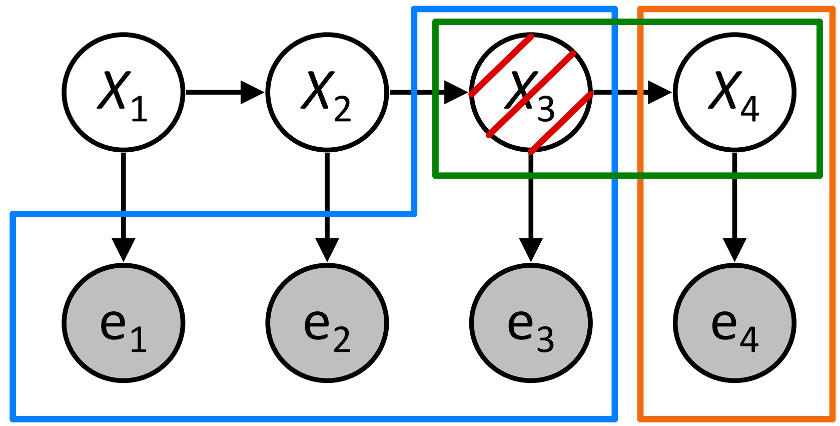
Chain rule with x_{t-1}, X_t, and e_t

Filtering **Algorithm**

Query: What is the current state, given all of the current and past evidence?

Matching math with Bayes net

$$P(X_t \mid e_{1:t}) = P(X_t \mid e_t, e_{1:t-1})$$
$$= \alpha\, P(X_t, e_t \mid e_{1:t-1})$$
$$= \alpha \sum_{x_{t-1}} P(x_{t-1}, X_t, e_t \mid e_{1:t-1})$$
$$= \alpha \sum_{x_{t-1}} P(x_{t-1} \mid e_{1:t-1})\, P(X_t \mid x_{t-1})\, P(e_t \mid X_t)$$



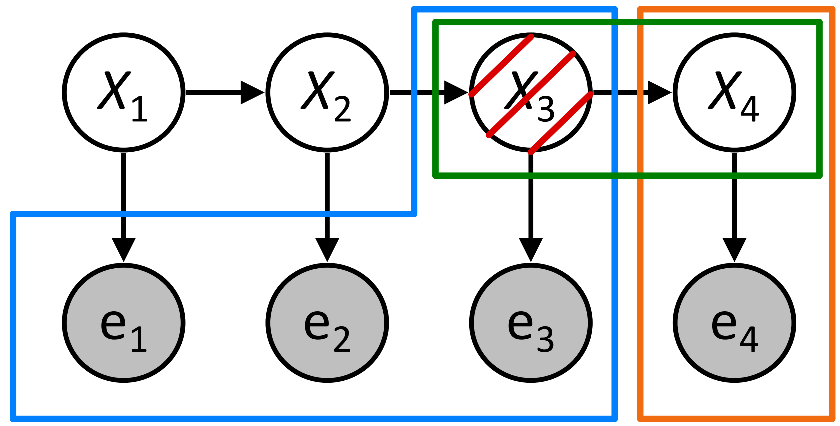
Bayes net conditional independence

Filtering **Algorithm**

Query: What is the current state, given all of the current and past evidence?

Matching math with Bayes net

$$P(X_t \mid e_{1:t}) = P(X_t \mid e_t, e_{1:t-1})$$
$$= \alpha\, P(X_t, e_t \mid e_{1:t-1})$$
$$= \alpha \sum_{x_{t-1}} P(x_{t-1}, X_t, e_t \mid e_{1:t-1})$$
$$= \alpha \sum_{x_{t-1}} P(x_{t-1} \mid e_{1:t-1})\, P(X_t \mid x_{t-1})\, P(e_t \mid X_t)$$
$$= \alpha\, P(e_t \mid x_t) \sum_{x_{t-1}} P(x_t \mid x_{t-1})\, P(x_{t-1} \mid e_{1:t-1})$$

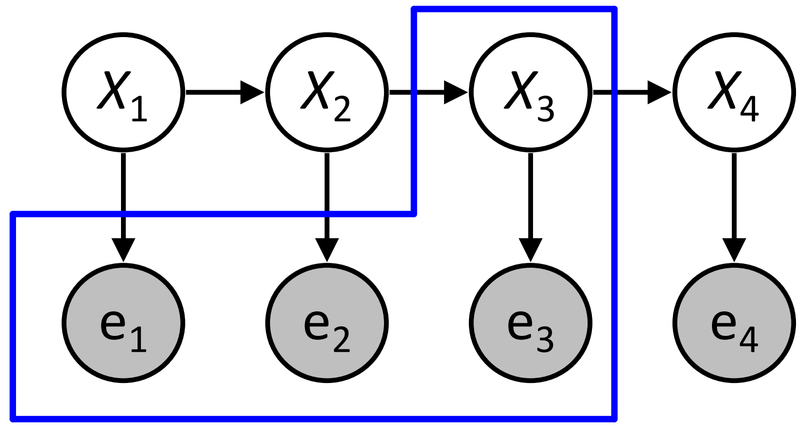


Pulling $P(e_t \mid X_t)$ out of the summation

Filtering **Algorithm**

Query: What is the current state, given all of the current and past evidence?

Matching math with Bayes net



Recursion!

$$P(X_t \mid e_{1:t}) = P(X_t \mid e_t, e_{1:t-1})$$

$$= \alpha\, P(X_t, e_t \mid e_{1:t-1})$$

$$= \alpha \sum_{x_{t-1}} P(x_{t-1}, X_t, e_t \mid e_{1:t-1})$$

$$= \alpha \sum_{x_{t-1}} P(x_{t-1} \mid e_{1:t-1})\, P(X_t \mid x_{t-1})\, P(e_t \mid X_t)$$

$$= \alpha\, P(e_t \mid x_t) \sum_{x_{t-1}} P(x_t \mid x_{t-1})\, P(x_{t-1} \mid e_{1:t-1})$$

Filtering **Algorithm**

Query: What is the current state, given all of the current and past evidence?

Matching math with Bayes net

$$P(X_t \mid e_{1:t}) = P(X_t \mid e_t, e_{1:t-1})$$

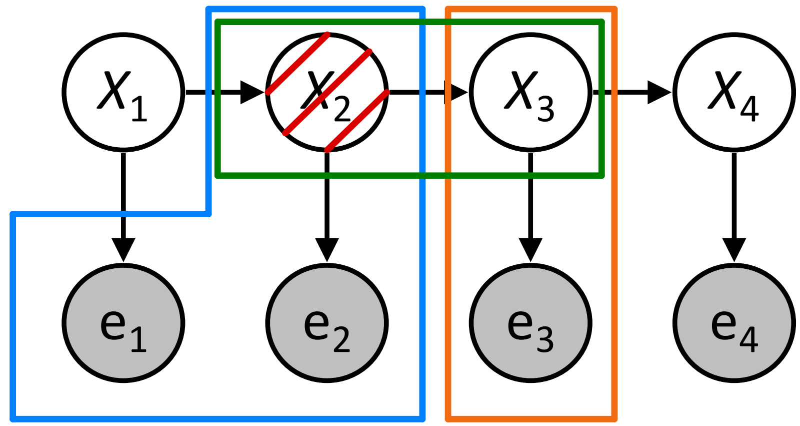
$$= \alpha \, P(X_t, e_t \mid e_{1:t-1})$$

$$= \alpha \sum_{x_{t-1}} P(x_{t-1}, X_t, e_t \mid e_{1:t-1})$$

$$= \alpha \sum_{x_{t-1}} P(x_{t-1} \mid e_{1:t-1}) \, P(X_t \mid x_{t-1}) \, P(e_t \mid X_t)$$

$$= \alpha \, P(e_t \mid x_t) \sum_{x_{t-1}} P(x_t \mid x_{t-1}) \, P(x_{t-1} \mid e_{1:t-1})$$

Recursion!



Poll 2

$$P(X_{t+1} | e_{1:t+1}) = \alpha \, P(e_{t+1} | X_{t+1}) \sum_{x_t} P(X_{t+1} | x_t) \, P(x_t | e_{1:t})$$

Normalize — Update — Predict

What is the runtime of the forward algorithm in terms of the number of states |X| and time t? Assume all 3 CPTs are given.

A) O(|X|^2 * t)

B) O(|X| * t)

C) O(|X|^2)

D) O(|X|)

Filtering **Algorithm**

$$P(X_{t+1} | e_{1:t+1}) = \alpha\, P(e_{t+1} | X_{t+1}) \sum_{x_t} P(X_{t+1} | x_t)\, P(x_t | e_{1:t})$$

Normalize — Update — Predict

$\boldsymbol{f}_{1:t+1}$ = FORWARD($\boldsymbol{f}_{1:t}$, e_{t+1})

Cost per time step: $O(|X|^2)$ where $|X|$ is the number of states

Time and space costs are ***constant***, independent of t

$O(|X|^2)$ is infeasible for models with many state variables

We get to invent really cool approximate filtering algorithms

In Class Activity: Weather HMM

An HMM is defined by:

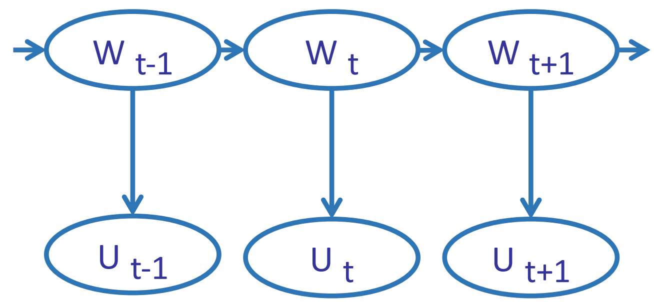
- Initial distribution: $P(X_1)$
- Transition model: $P(X_t \mid X_{t-1}) = P(W_t \mid W_{t-1})$
- Sensor model: $P(E_t \mid X_t) = P(U_t \mid W_t)$

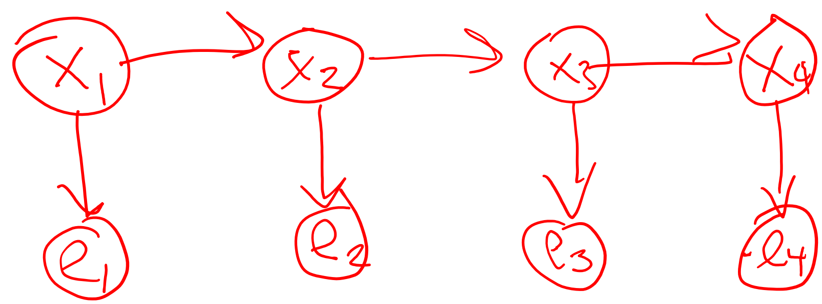
W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Given $P(X_1)$ = {sun:0.5, rain:0.5}

Compute $P(X_4\text{=sun} \mid e_4\text{=} e_3\text{=} e_2\text{=} e_1\text{=True})$





In Class Activity: Weather HMM

An HMM is defined by:

- Initial distribution: $P(X_1)$
- Transition model: $P(X_t \mid X_{t-1})$
- Sensor model: $P(E_t \mid X_t)$

W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Compute $P(X_4\text{=sun} \mid e_4= e_3= e_2= e_1\text{=True})$ and $P(X_1) = \{\text{sun:0.5, rain:0.5}\}$

$P(X_1, e_1) = P(e_1|X_1)P(X_1)$ #OBSERVE (chain rule)

$P(X_1|e_1) = \alpha P(X_1, e_1) \rightarrow \alpha = 1/\sum_{x_1} P(e_1|x_1)P(x_1)$ #Don't forget to NORMALIZE

$P(X_2|e_1) = \sum_{x \in X_1} P(X_2|x)P(x|e_1)$ #PREDICT

In Class Activity: Weather HMM

An HMM is defined by:

- Initial distribution: $P(X_1)$
- Transition model: $P(X_t \mid X_{t-1})$
- Sensor model: $P(E_t \mid X_t)$

W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Compute $P(X_4=\text{sun} \mid e_4= e_3= e_2= e_1=\text{True})$ and $P(X_1) = \{\text{sun}:0.5, \text{rain}:0.5\}$

$P(X_2|e_1) = \sum_{x \in X_1} P(X_2|x)P(x|e_1)$ #PREDICT

$$P(X_2|e_1, e_2) = \alpha P(X_2, e_2|e_1) = \alpha P(e_2|X_2)P(X_2|e_1);\ \alpha = 1/\sum_{x \in X_2} P(e_2|x)P(x|e_1)$$

$P(X_3|e_1, e_2) = \sum_{x_2 \in X_2} P(X_3|x_2)P(x_2|e_1, e_2)$ #PREDICT

In Class Activity: Weather HMM

An HMM is defined by:

- Initial distribution: $P(X_1)$
- Transition model: $P(X_t \mid X_{t-1})$
- Sensor model: $P(E_t \mid X_t)$

W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Compute $P(X_4\text{=sun} \mid e_4\text{= }e_3\text{= }e_2\text{= }e_1\text{=True})$ and $P(X_1) = \{\text{sun:0.5, rain:0.5}\}$

$P(X_3|e_1,e_2) = \sum_{x_2 \in X_2} P(X_3|x_2)P(x_2|e_1,e_2)$ #PREDICT

$$P(X_3|e_1,e_2,e_3) = \alpha P(X_3,e_3|e_1,e_2) = \alpha P(e_3|X_3)P(X_3|e_1,e_2);$$

$$\alpha = 1/\sum_{x \in X_3} P(e_3|x)P(x|e_1,e_2)$$

$P(X_4|e_1,e_2,e_3) = \sum_{x \in X_3} P(X_4|x)P(x|e_1,e_2,e_3)$ #PREDICT

In Class Activity: Weather HMM

An HMM is defined by:

- Initial distribution: $P(X_1)$
- Transition model: $P(X_t \mid X_{t-1})$
- Sensor model: $P(E_t \mid X_t)$

W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Compute $P(X_4=\text{sun} \mid e_4= e_3= e_2= e_1=\text{True})$ and $P(X_1) = \{\text{sun}:0.5, \text{rain}:0.5\}$

$P(X_4|e_1,e_2,e_3) = \sum_{x \in X_3} P(X_4|x)P(x|e_1,e_2,e_3)$ #PREDICT

$$P(X_4|e_1,e_2,e_3,e_4) = \alpha P(X_4,e_4|e_1,e_2,e_3) = \alpha P(e_4|X_4)P(X_4|e_1,e_2,e_3);$$

$$\alpha = 1/\sum_{x \in X_4} P(e_4|x)P(x|e_1,e_2,e_3)$$

In Class Activity: Weather HMM

An HMM is defined by:

- Initial distribution: $P(X_1)$
- Transition model: $P(X_t \mid X_{t-1})$
- Sensor model: $P(E_t \mid X_t)$

W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Compute $P(X_4\text{=sun} \mid e_4\text{= }e_3\text{= }e_2\text{= }e_1\text{=True})$ and $P(X_1)$ = {sun:0.5, rain:0.5}

$P(X_1, e_1) = P(e_1|X_1)P(X_1)$ #OBSERVE (chain rule)

$P(e_1 = True|X_1 = sun)P(X_1 = sun) = .2 * .5 = .1$

$P(e_1 = True|X_1 = rain)P(X_1 = rain) = .9 * .5 = .45$

$P(X_1|e_1) = \frac{P(X_1,e_1)}{P(e_1)} = P(e_1|X_1)P(X_1)/\sum_{x\in X_1} P(e_1|x)P(x)$ #NORMALIZE USING BAYES RULE

$P(X_1 = sun|e_1 = True) = \frac{.1}{.1 + .45} = .18$

$P(X_1 = rain|e_1 = True) = \frac{.45}{.1 + .45} = .82$

In Class Activity: Weather HMM

An HMM is defined by:

- Initial distribution: $P(X_1)$
- Transition model: $P(X_t \mid X_{t-1})$
- Sensor model: $P(E_t \mid X_t)$

W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Compute $P(X_4{=}sun \mid e_4{=}\ e_3{=}\ e_2{=}\ e_1{=}True)$ and $P(X_1) = \{sun{:}0.5,\ rain{:}0.5\}$

$P(X_2|e_1) = \sum_{x \in X_1} P(X_2|x)P(x|e_1)$ #PREDICT

$$P(X_2 = sun|e_1 = True) = \sum_{x \in X_1} P(X_2 = sun|x)P(x|e_1 = True) = .9 * .18 + .3 * .82 = .41$$

$$P(X_2 = rain|e_1 = True) = \sum_{x \in X_1} P(X_2 = rain|x)P(x|e_1 = True) = .1 * .18 + .7 * .82 = .59$$

In Class Activity: Weather HMM

An HMM is defined by:

- Initial distribution: $P(X_1)$
- Transition model: $P(X_t \mid X_{t-1})$
- Sensor model: $P(E_t \mid X_t)$

W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Compute $P(X_4=sun \mid e_4= e_3= e_2= e_1=True)$ and $P(X_1) = \{sun:0.5, rain:0.5\}$

$$P(X_2|e_1,e_2) = \alpha P(X_2,e_2|e_1) = \alpha P(e_2|X_2)P(X_2|e_1);\ \alpha = 1/\sum_{x\in X_2} P(e_2|x)P(x|e_1)$$

$$P(X_2 = sun|e_1,e_2 = True) = \alpha P(e_2|X_2 = sun)P(X_2 = sun|e_1) = \alpha(.2)(.41) = .13$$

$$P(X_2 = rain|e_1,e_2 = True) = \alpha P(e_2|X_2 = rain)P(X_2 = rain|e_1) = \alpha(.9)(.59) = .87$$

$P(X_3|e_1,e_2) = \sum_{x\in X_2} P(X_3|x)P(x|e_1,e_2)$ #PREDICT

$P(X_3 = sun|e_1,e_2) = P(X_3 = sun|x = sun)P(x = sun|e_1,e_2) + P(X_3|x = rain)P(x = rain|e_1,e_2) = 0.38$

$P(X_3 = rain|e_1,e_2) = P(X_3 = rain|x = sun)P(x = sun|e_1,e_2) + P(X_3|x = rain)P(x = rain|e_1,e_2) = 0.62$

In Class Activity: Weather HMM

An HMM is defined by:

- Initial distribution: $P(X_1)$
- Transition model: $P(X_t \mid X_{t-1})$
- Sensor model: $P(E_t \mid X_t)$

W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Compute $P(X_4\text{=sun} \mid e_4 = e_3 = e_2 = e_1 = \text{True})$ and $P(X_1) = \{\text{sun:0.5, rain:0.5}\}$

$$P(X_3|e_1,e_2,e_3) = \alpha P(X_3, e_3|e_1,e_2) = \alpha P(e_3|X_3)P(X_3|e_1,e_2);$$

$$\alpha = 1/\sum_{x \in X_3} P(e_3|x)P(x|e_1,e_2)$$

$$P(X_3 = sun|e_1,e_2,e_3) = \alpha P(e_3 = True|X_3 = sun)P(X_3 = sun|e_1,e_2) = \alpha(.2)(.38) = .12$$

$$P(X_3 = rain|e_1,e_2,e_3) = \alpha P(e_3 = True|X_3 = rain)P(X_3 = rain|e_1,e_2) = \alpha(.9)(.62) = .88$$

In Class Activity: Weather HMM

An HMM is defined by:

- Initial distribution: $P(X_1)$
- Transition model: $P(X_t \mid X_{t-1})$
- Sensor model: $P(E_t \mid X_t)$

W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Compute $P(X_4\text{=sun} \mid e_4\text{= }e_3\text{= }e_2\text{= }e_1\text{=True})$ and $P(X_1)$ = {sun:0.5, rain:0.5}

$P(X_4|e_1, e_2, e_3) = \sum_{x \in X_3} P(X_4|x)P(x|e_1, e_2, e_3)$ #PREDICT

$$P(X_4 = sun|e_1, e_2, e_3) = \sum_{x \in \{sun, rain\}} P(X_4 = sun|x)P(x|e_1, e_2, e_3) = .9 * .12 + .3 * .88 = .37$$

$$P(X_4 = rain|e_1, e_2, e_3) = \sum_{x \in \{sun, rain\}} P(X_4 = rain|x)P(x|e_1, e_2, e_3) = .1 * .12 + .7 * .88 = .63$$

In Class Activity: Weather HMM

An HMM is defined by:

- Initial distribution: $P(X_1)$
- Transition model: $P(X_t \mid X_{t-1})$
- Sensor model: $P(E_t \mid X_t)$

W_{t-1}	$P(W_t \mid W_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

W_t	$P(U_t \mid W_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Compute $P(X_4\text{=sun} \mid e_4\text{= }e_3\text{= }e_2\text{= }e_1\text{=True})$ and $P(X_1)$ = {sun:0.5, rain:0.5}

$$P(X_4|e_1,e_2,e_3,e_4) = \alpha P(X_4,e_4|e_1,e_2,e_3) = \alpha P(e_4|X_4)P(X_4|e_1,e_2,e_3);$$
$$\alpha = 1/\sum_{x\in X_4} P(e_4|x)P(x|e_1,e_2,e_3)$$

$$\alpha P(e_4 = True|X_4 = sun)P(X_4 = sun|e_1,e_2,e_3) = \alpha(.2{*}.37) = .115$$
$$\alpha P(e_4 = True|X_4 = rain)P(X_4 = rain|e_1,e_2,e_3) = \alpha(.9{*}.63) = .885$$

Poll 3

Suppose we are given P(X4=sun | e4= e3= e2= e1=True), along with the same CPT tables as the activity example, and we want to compute P(X5=sun | e5= e4= e3= e2= e1=True).

What is the first step we would perform?

Predict

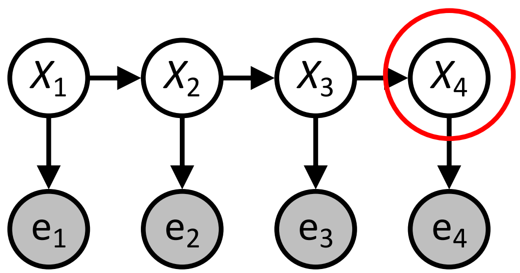
Observe

Forward

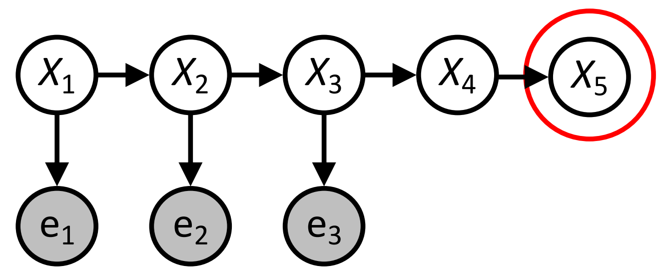
Smoothing

Other HMM Queries

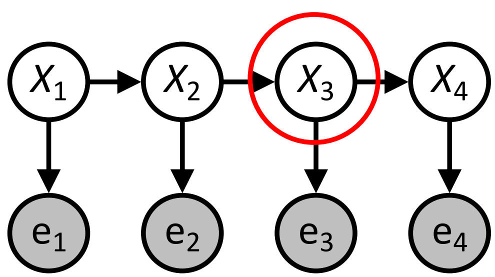
Filtering: $P(X_t | e_{1:t})$



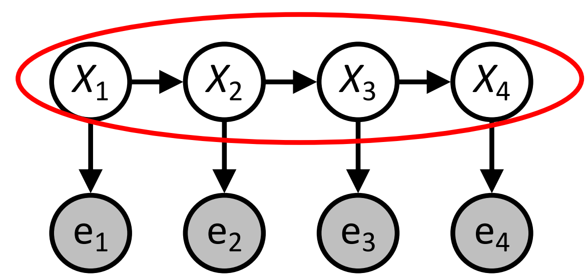
Prediction: $P(X_{t+k} | e_{1:t})$



Smoothing: $P(X_k | e_{1:t})$, $k<t$



Explanation: $P(X_{1:t} | e_{1:t})$



Inference Tasks

Filtering: $P(X_t | e_{1:t})$

- belief state—input to the decision process of a rational agent

Prediction: $P(X_{t+k} | e_{1:t})$ for $k > 0$

- evaluation of possible action sequences; like filtering without the evidence

Smoothing: $P(X_k | e_{1:t})$ for $0 \leq k < t$

- better estimate of past states, essential for learning

Most likely explanation: $\text{argmax}_{x_{1:t}} P(x_{1:t} \mid e_{1:t})$

- speech recognition, decoding with a noisy channel

Dynamic Bayes Nets (DBNs)

We want to track multiple variables over time, using multiple sources of evidence

Idea: Repeat a fixed Bayes net structure at each time

Variables from time *t* can condition on those from *t-1*