AI: Representation and Problem Solving

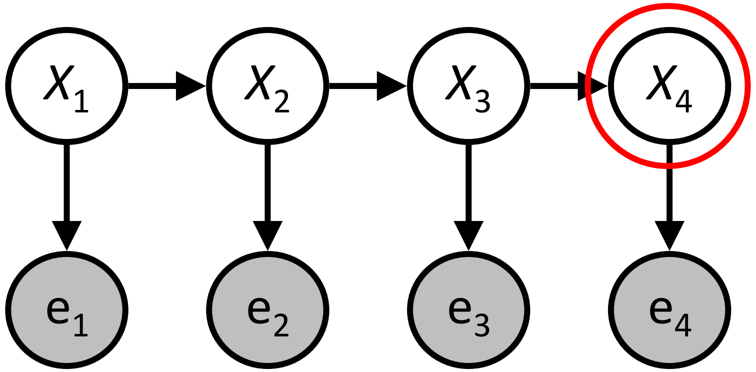
Hidden Markov Models

Instructors: Pat Virtue

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

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.



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

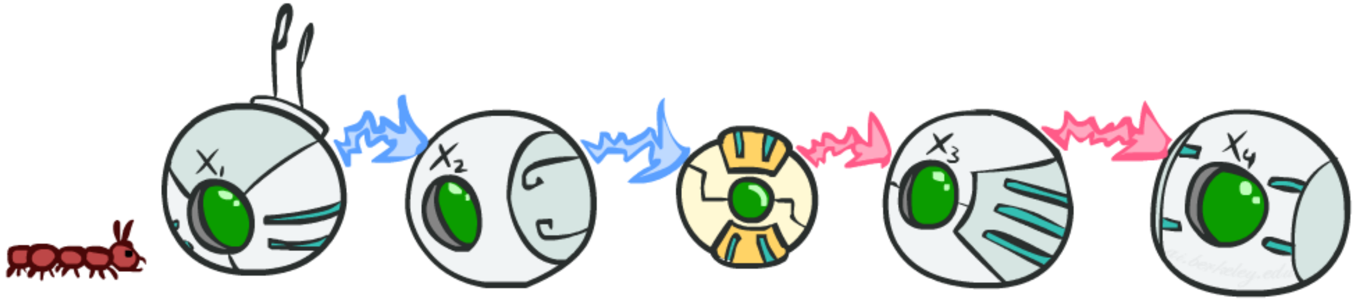
Reasoning over Time or Space

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

Conditional Independence



Basic conditional independence:

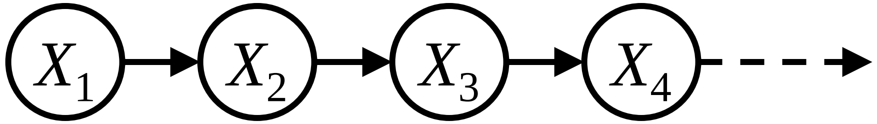
- Past and future independent given the present
- Each time step only depends on the previous
- This is called the (first order) Markov property

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

Markov Chain Models

Value of X at a given time is called the state



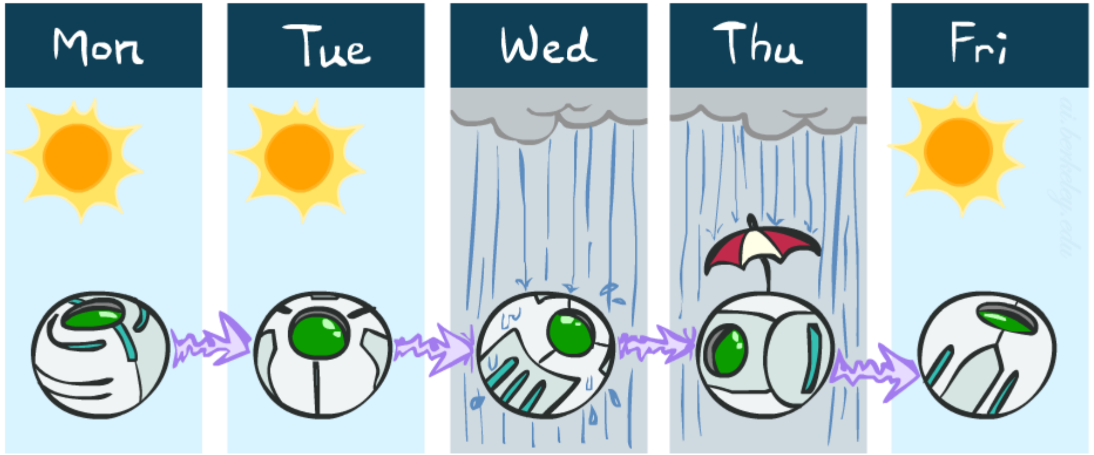
$P(X_1)$

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

- 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

Example: Markov Chain Weather

States: X = {rain, sun}

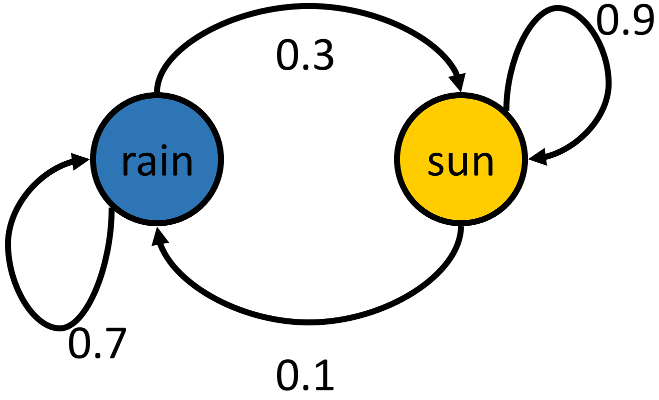


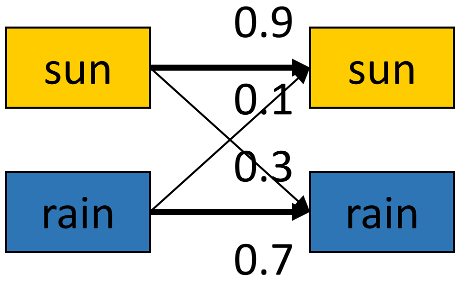
- Initial distribution: 1.0 sun

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

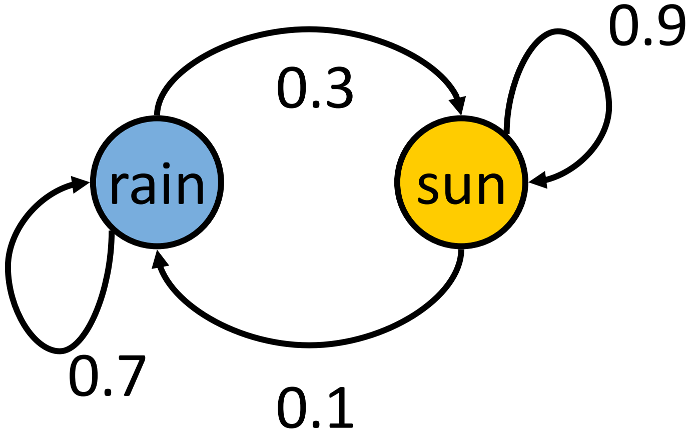
X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Two new ways of representing the same CPT





Example: Markov Chain Weather

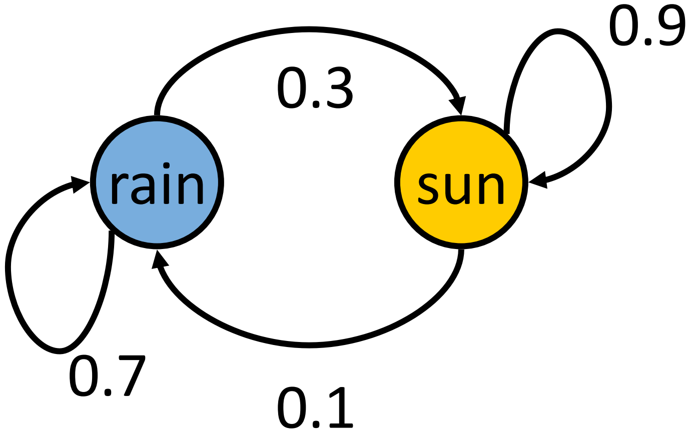


Initial distribution: $P(X_1 = sun) = 1.0$

What is the probability distribution after one step?

$P(X_2 = sun) = ?$

Example: Markov Chain Weather



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

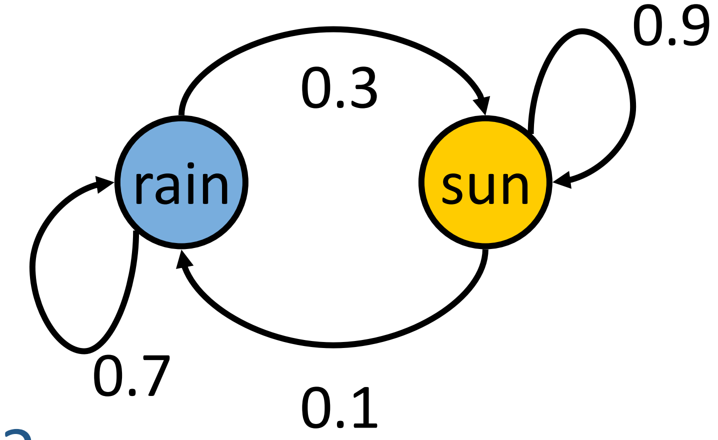
What is the probability distribution after one step?

$P(X_2 = sun) = ?$

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

Poll 1

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

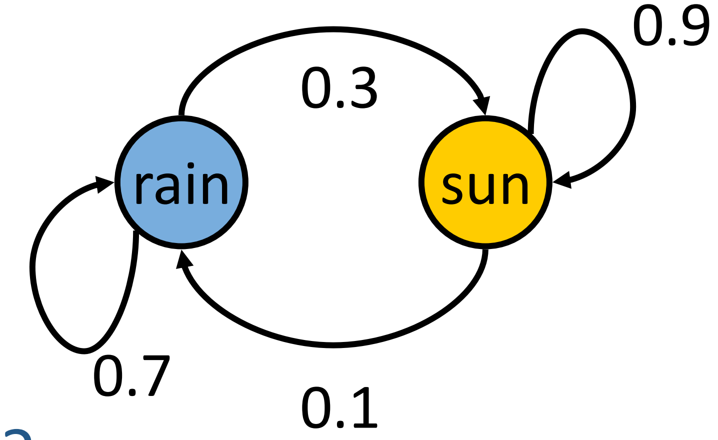


What is the probability distribution after the next step?

$P(X_3 = sun) \ = ?$

A) 0.81
B) 0.84
C) 0.9
D) 1.0
E) 1.2

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

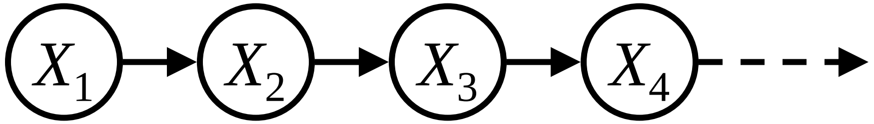
$$P(X_3 = sun) = \sum_{x_2} P(X_3 = sun, X_2 = x_2)$$

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

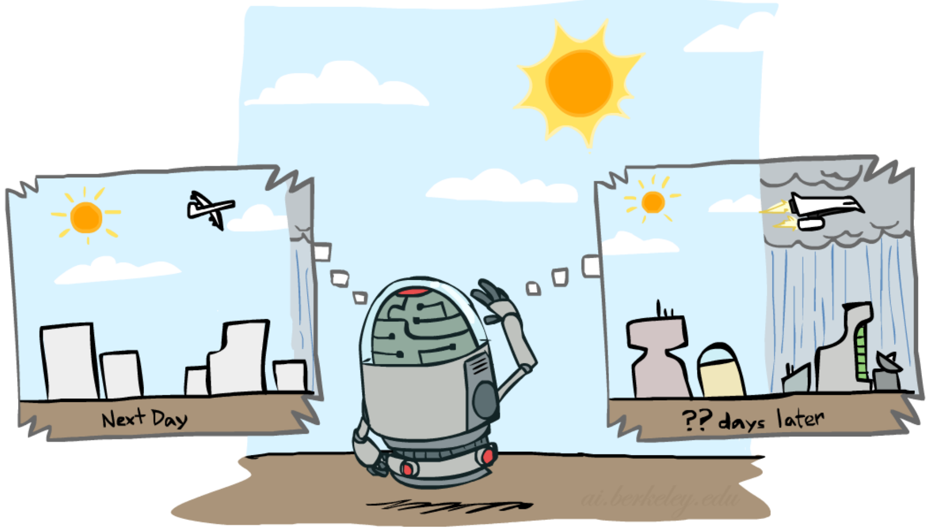
$$= 0.9 \cdot 0.9 + 0.3 \cdot 0.1$$

$$= 0.81 + 0.03 = 0.84$$

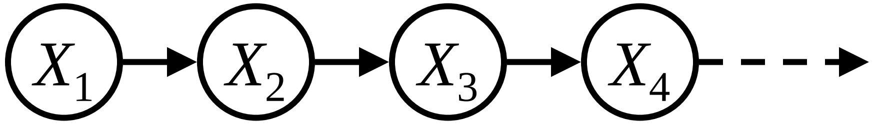
Markov Chain Inference



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



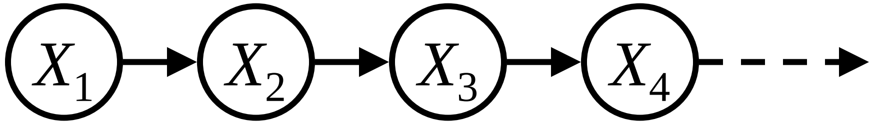
Markov Chain Inference



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 $P(X_5)$.

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

Weather prediction

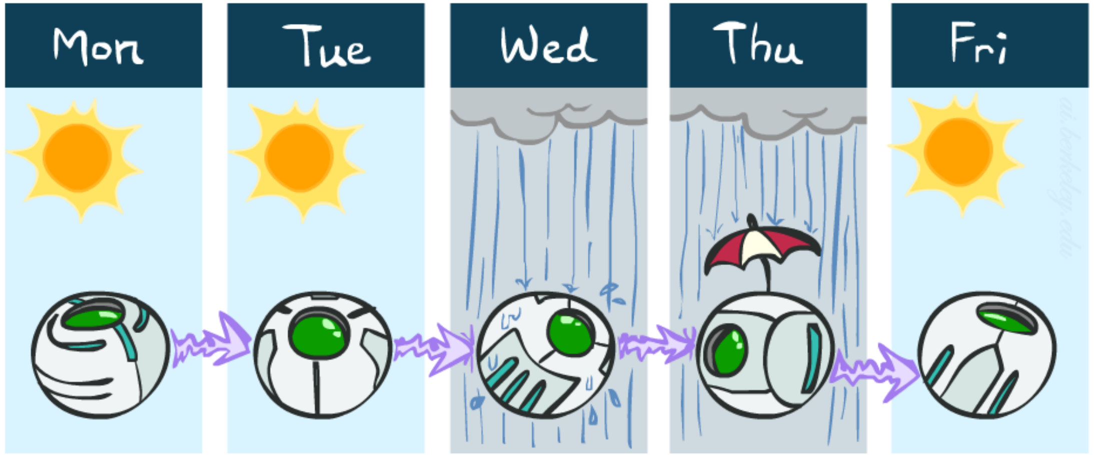
States {rain, sun}

- Initial distribution $P(X_0)$

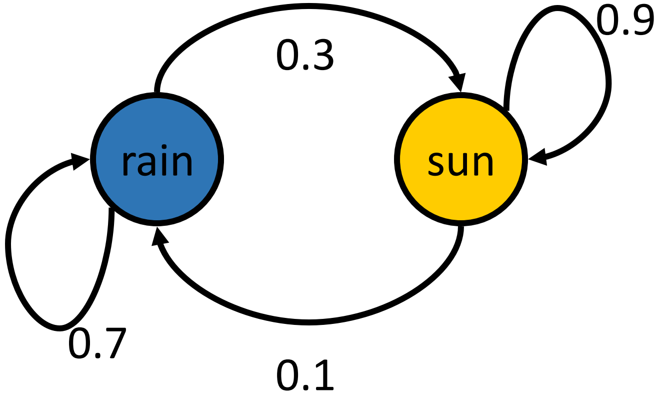
P(X_0)	
sun	rain
0.5	0.5

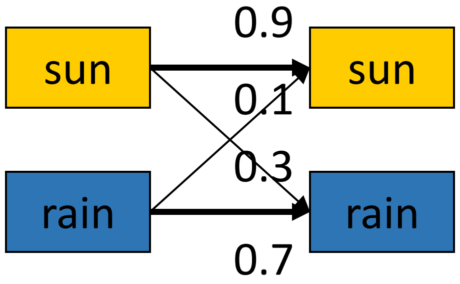
- Transition model $P(X_t \mid X_{t-1})$

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



Two new ways of representing the same CPT

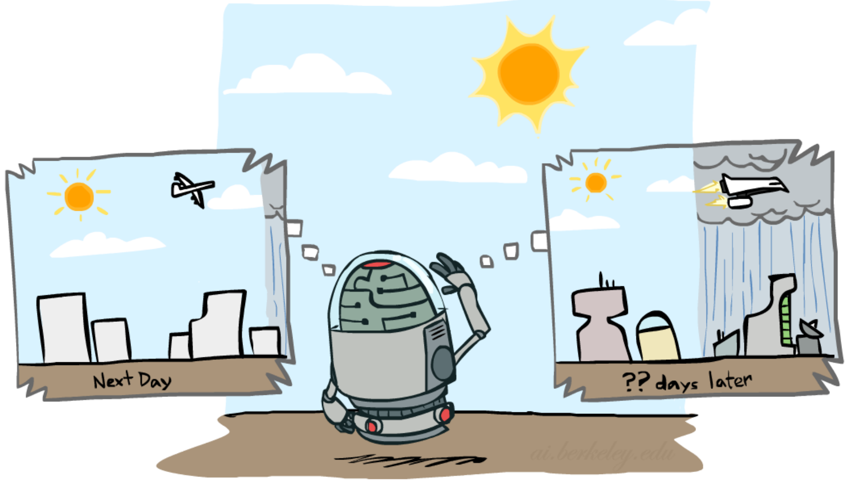




Weather prediction

Time 0: $P(X_0)$ =<0.5,0.5>

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



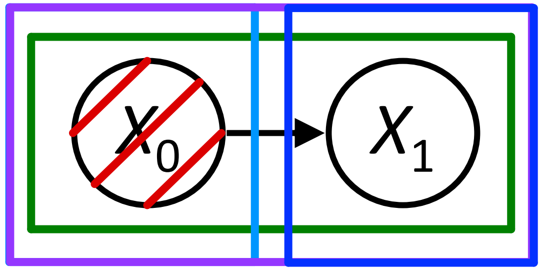
What is the weather like at time 1?

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

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

$$= 0.5<0.9,0.1> + 0.5<0.3,0.7>$$

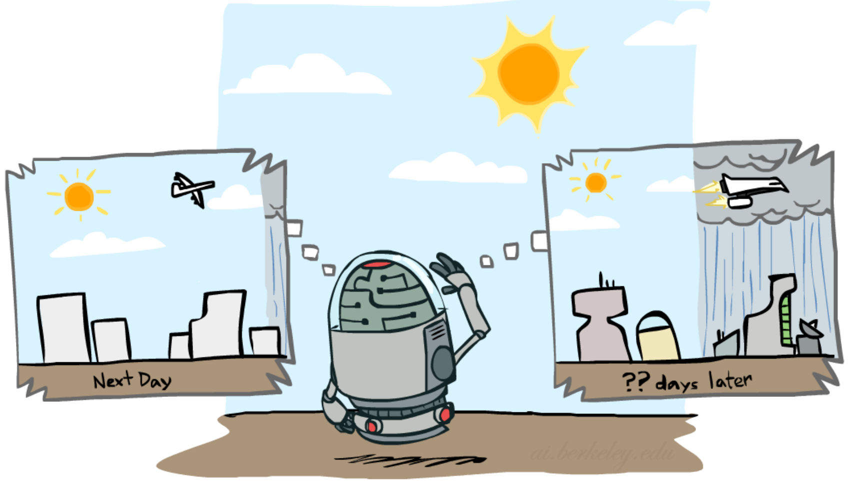
$$= <0.6,0.4>$$



Weather prediction, contd.

Time 1: $P(X_1)$ =<0.6,0.4>

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



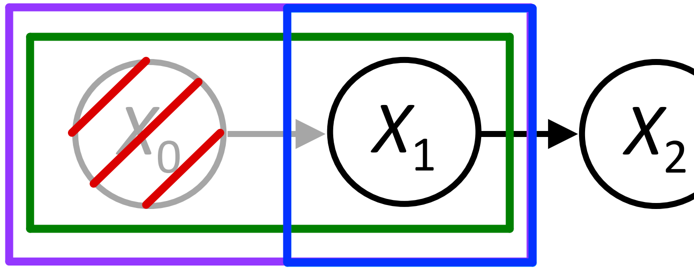
What is the weather like at time 2?

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

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

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

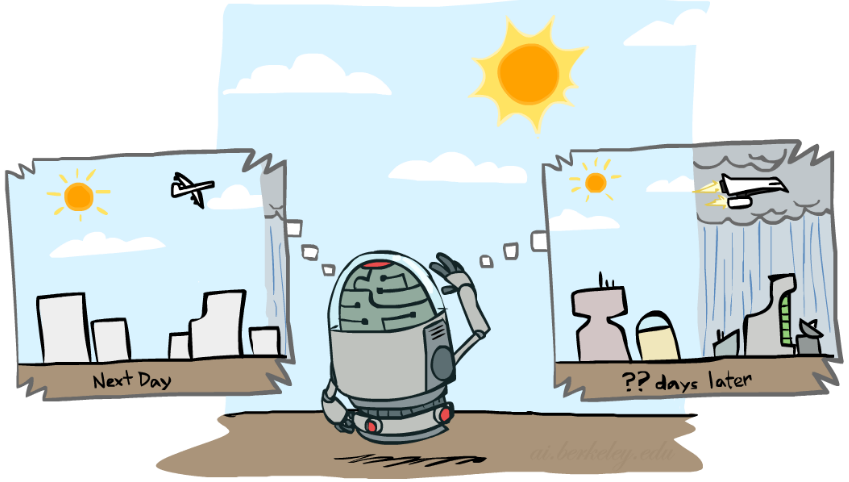
$$= \langle 0.66,0.34\rangle$$



Weather prediction, contd.

Time 2: $P(X_2)$ =<0.66,0.34>

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



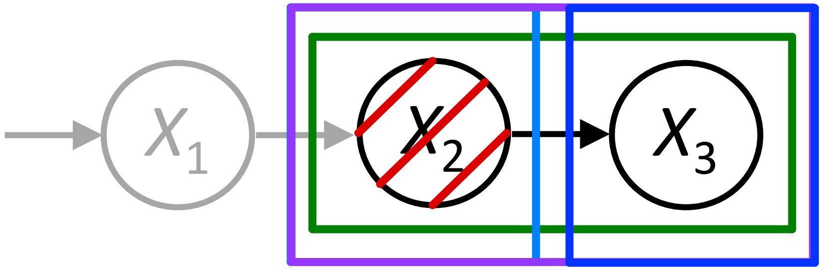
What is the weather like at time 3?

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

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

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

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



Forward algorithm (simple form)

Transition model

Probability from previous iteration

What is the state at time *t*?

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

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

Iterate this update starting at *t*=0

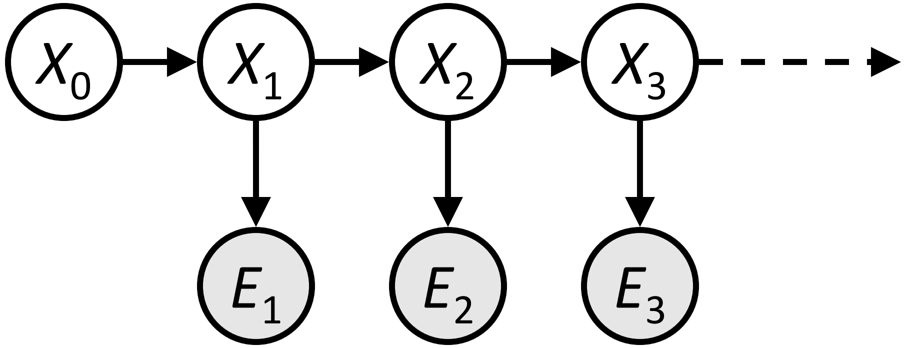
Hidden Markov Models

Hidden Markov Models

Usually the true state is not observed directly

Hidden Markov models (HMMs)

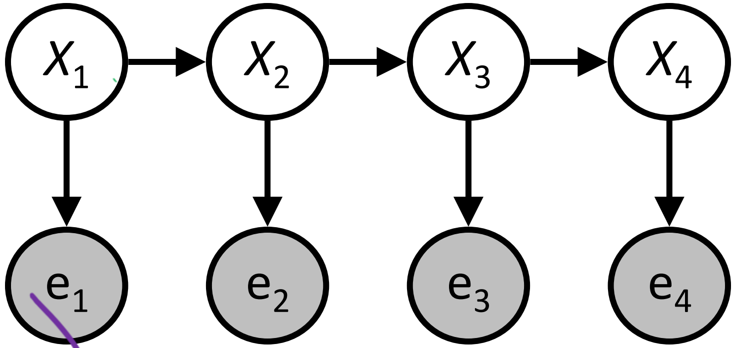
- Underlying Markov chain over states X
- You observe evidence E at each time step
- X_t is a single discrete variable; E_t may be continuous and may consist of several variables



HMM as a Bayes Net Warm-up

$$\alpha = \frac{1}{P(e_1 e_2 e_3 e_4)}$$

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.



$$P(X_4 \mid e_1, e_2, e_3, e_4) = \alpha P(X_4, e_{1:4})$$

$$= \alpha \sum_{x_1} \sum_{x_2} \sum_{x_3} P(x_1, x_2, x_3, X_4, e_1, e_2, e_3, e_4)$$

$$= \alpha \sum \sum \sum P(x_1) P(e_1 \mid x_1) P(x_2 \mid x_1) P(e_2 \mid x_1) P(x_3 \mid x_2) \ldots$$

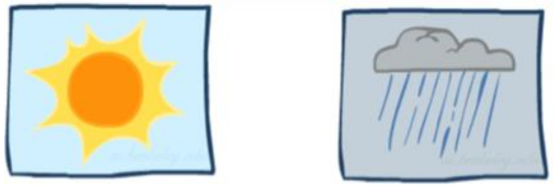
Useful notation: $X_{a:b} = X_a, X_{a+1}, \ldots, X_b$

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

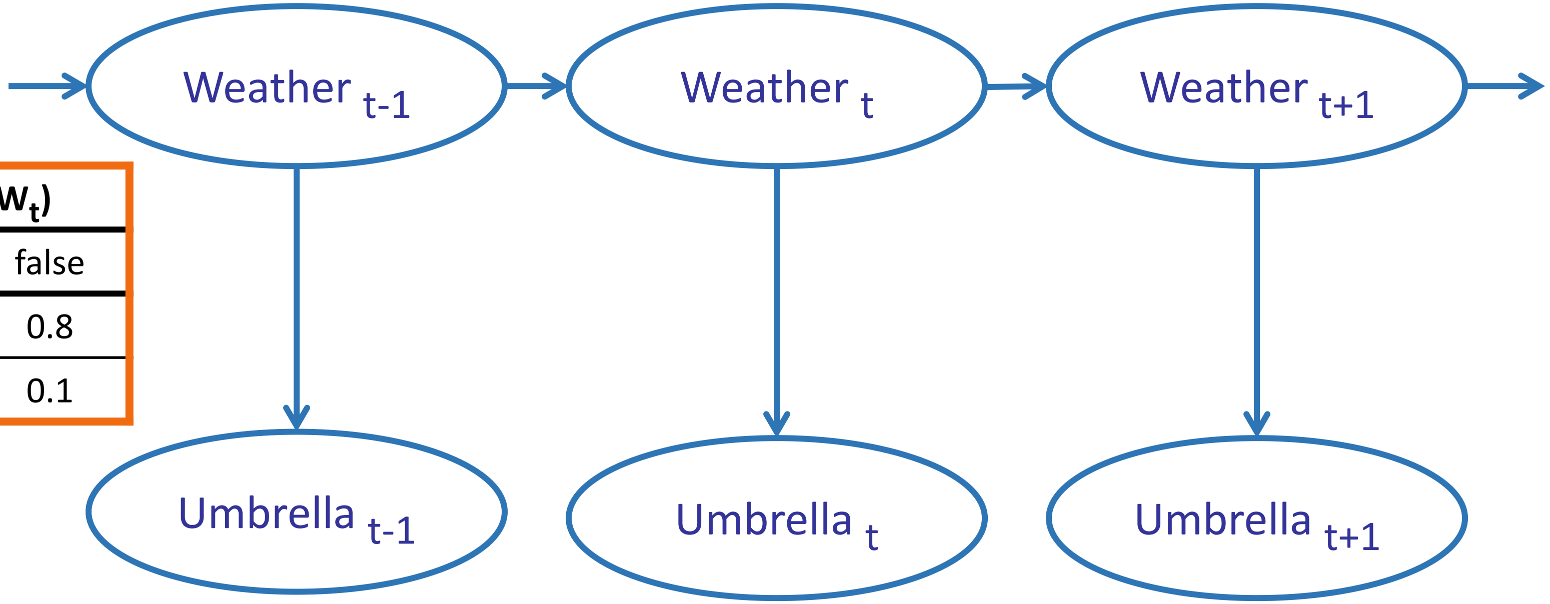
Example: Weather HMM

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



Weather $_{t-1}$

Weather $_t$

Weather $_{t+1}$

Umbrella $_{t-1}$

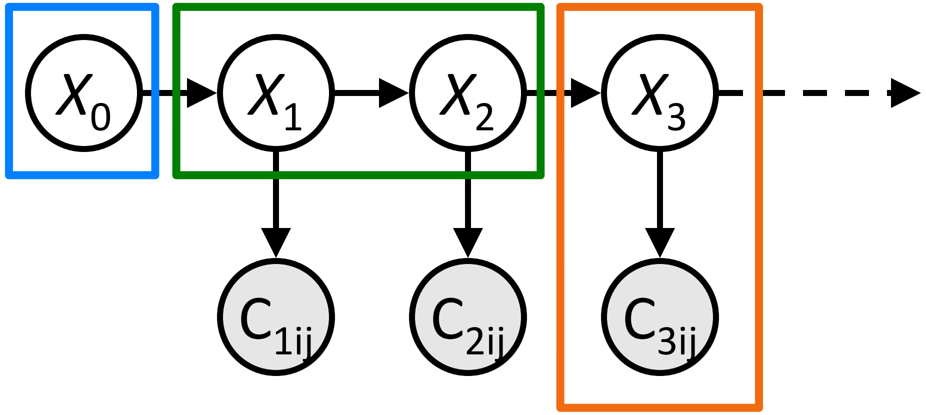
Umbrella $_t$

Umbrella $_{t+1}$

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

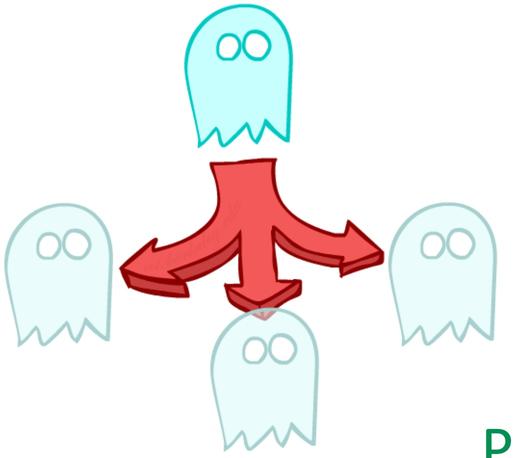
Example: Ghostbusters HMM

- State: location of moving ghost
- Observations: Color recorded by ghost sensor at clicked squares
- $P(X_0)$ = uniform
- $P(X_t \mid X_{t-1})$ = usually move clockwise, but sometimes move randomly or stay in place
- $P(C_{tij} \mid X_t)$ = same sensor model as before: red means close, green means far away.



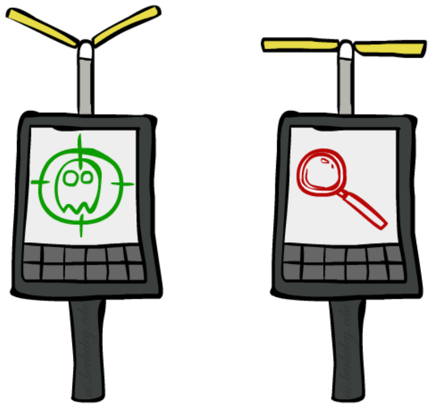
1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

$P(X_0)$



1/6	1/6	1/2
0	1/6	0
0	0	0

$P(X_2 \mid X_1=(2,3))$

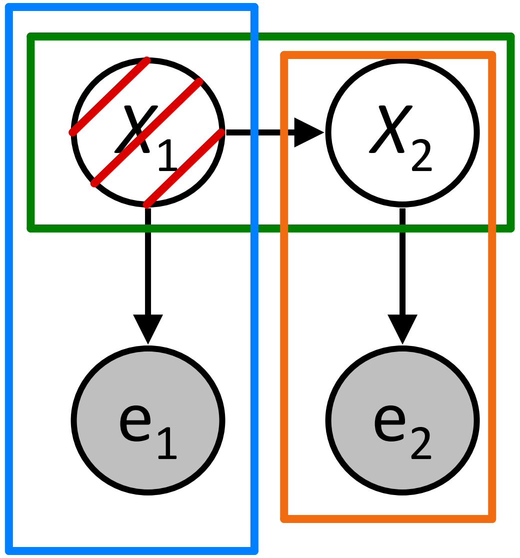


[Demo: Ghostbusters – Circular Dynamics – HMM (L14D2)]

Filtering Algorithm

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

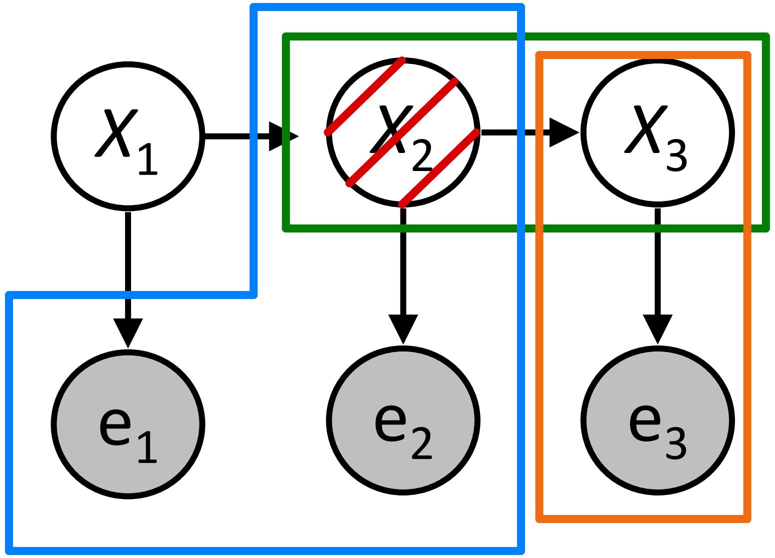
Marching **forward** through the HMM network



Filtering Algorithm

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

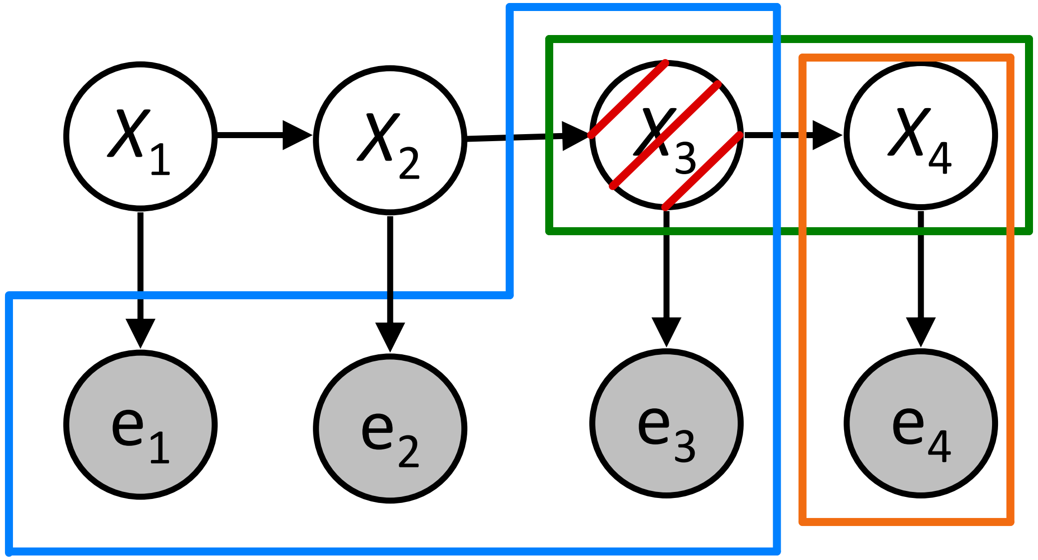
Marching **forward** through the HMM network



Filtering Algorithm

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

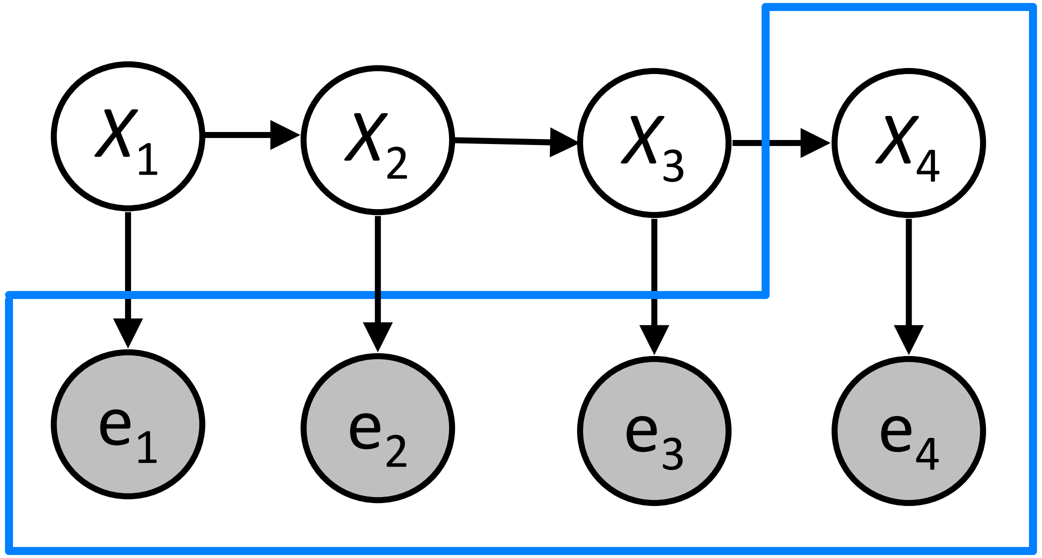
Marching **forward** through the HMM network



Filtering Algorithm

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

Marching **forward** through the HMM network



Filtering Algorithm

$$P(X_{t+1} \mid e_{1:t+1}) = \underbrace{\alpha}_{\text{Normalize}} \underbrace{P(e_{t+1} \mid X_{t+1})}_{\text{Update}} \underbrace{\textstyle\sum_{x_t} P(X_{t+1} \mid x_t)\, P(x_t \mid e_{1:t})}_{\text{Predict}}$$

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

HMM as Probability Model

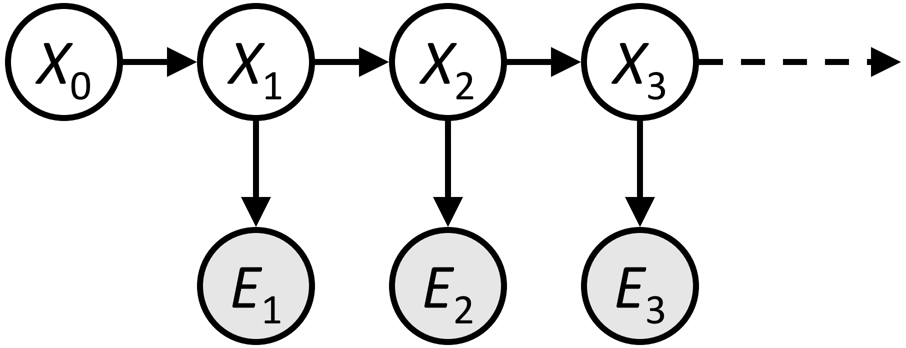
- Joint distribution for Markov model:

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

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

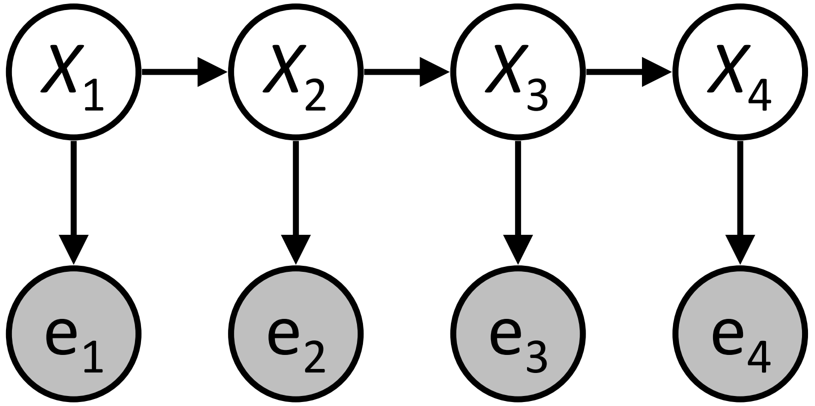
Real HMM Examples

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:

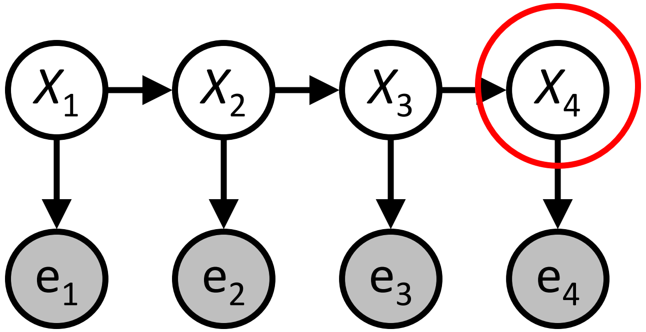
- Observations are range readings (continuous)
  - Or even images
- States are positions on a map (continuous)

Molecular biology:

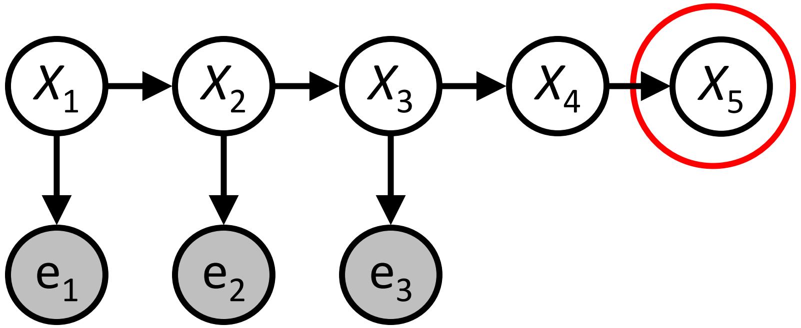
- Observations are nucleotides ACGT
- States are coding/non-coding/start/stop/splice-site etc.

Other HMM Queries

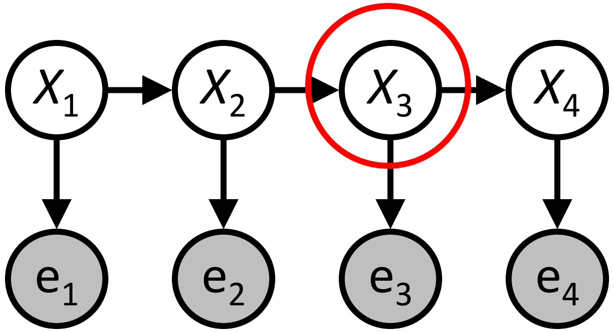
Filtering: $P(X_t | e_{1:t})$



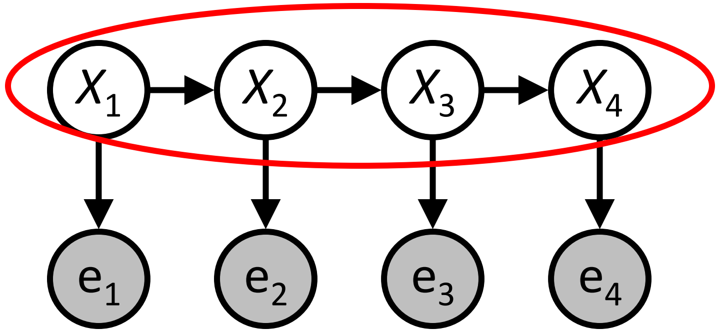
Prediction: $P(X_{t+k} | e_{1:t})$



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



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

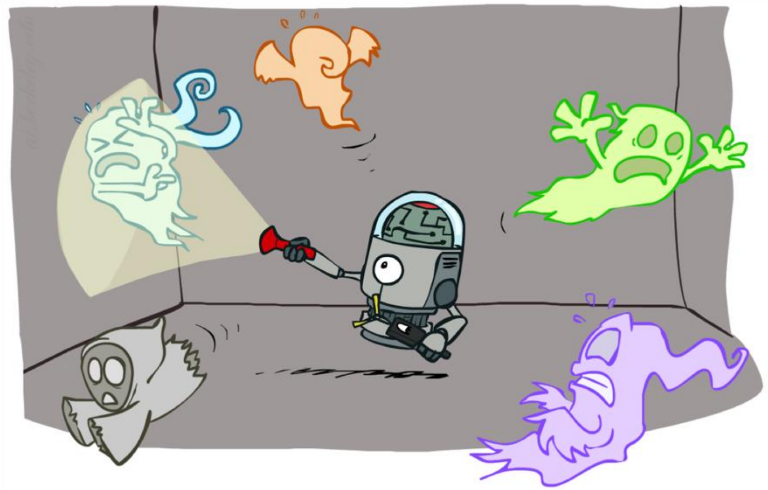


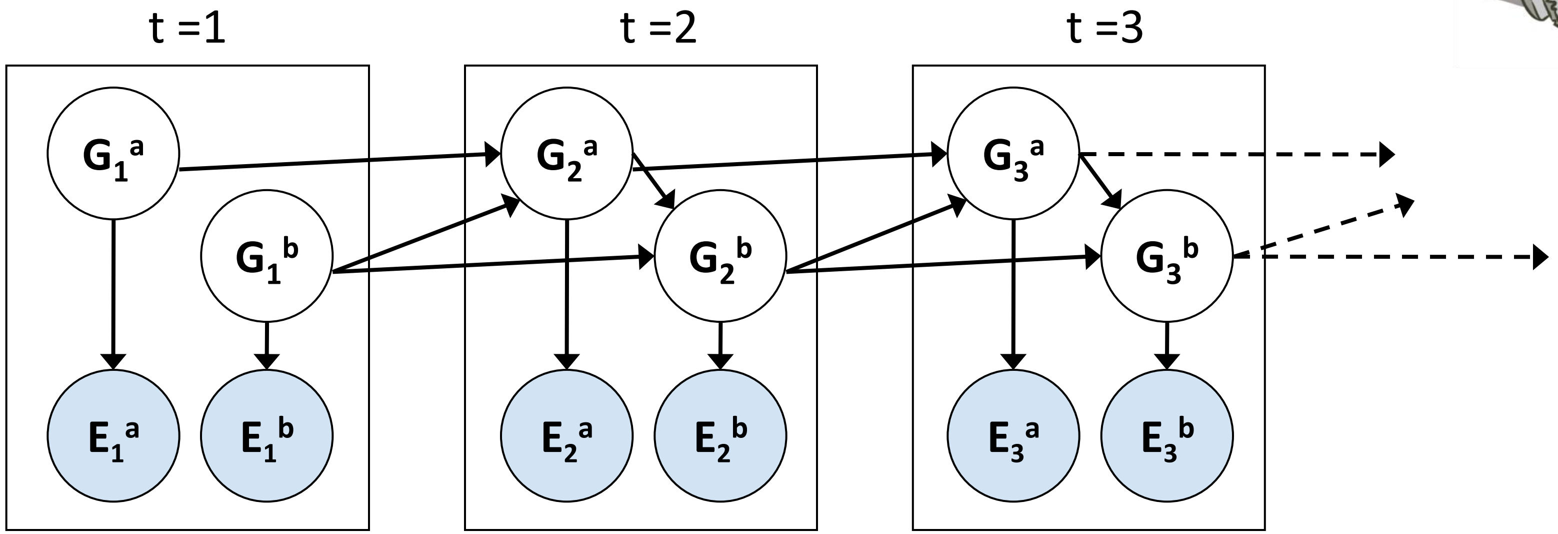
Dynamic Bayes Nets (DBNs)

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*





Filtering Algorithm

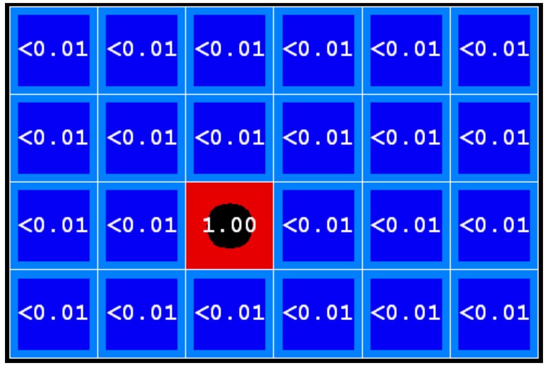
$$P(X_{t+1} \mid e_{1:t+1}) = \underbrace{\alpha}_{\text{Normalize}} \; \underbrace{P(e_{t+1} \mid X_{t+1})}_{\text{Update}} \; \underbrace{\sum_{x_t} P(X_{t+1} \mid x_t) \, P(x_t \mid e_{1:t})}_{\text{Predict}}$$

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

Example: Prediction step

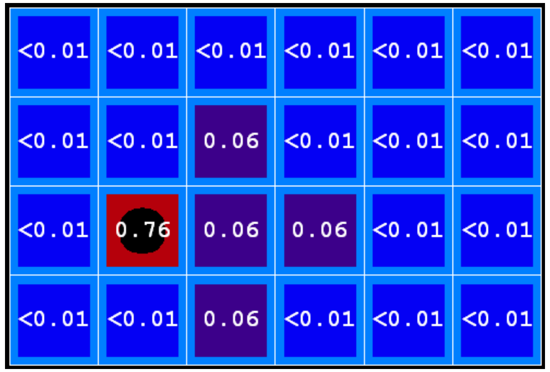
As time passes, uncertainty “accumulates”

(Transition model: ghosts usually go clockwise)



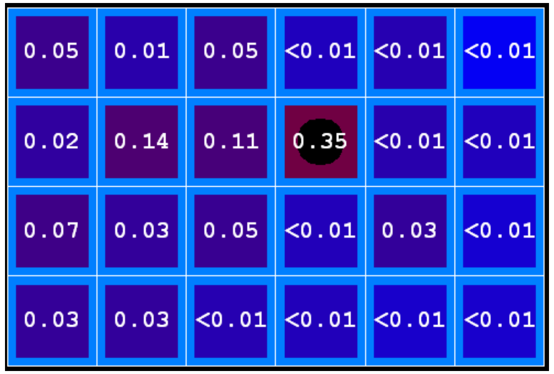
<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
<0.01	<0.01	1.00	<0.01	<0.01	<0.01
<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

T = 1



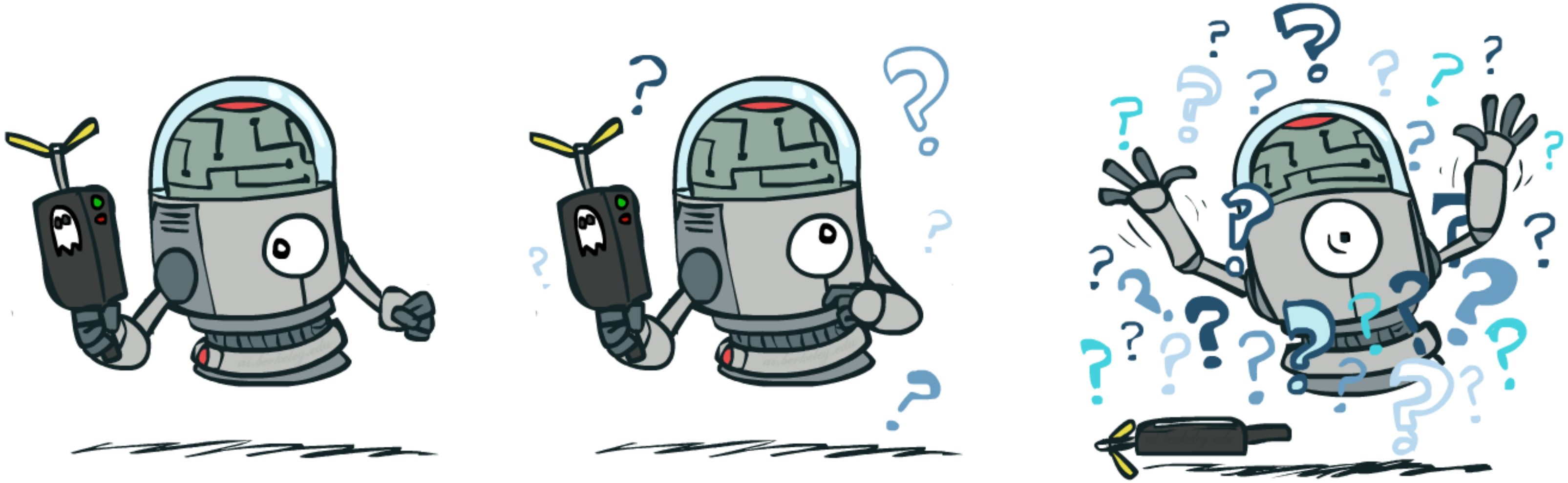
<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
<0.01	<0.01	0.06	<0.01	<0.01	<0.01
<0.01	0.76	0.06	0.06	<0.01	<0.01
<0.01	<0.01	0.06	<0.01	<0.01	<0.01

T = 2



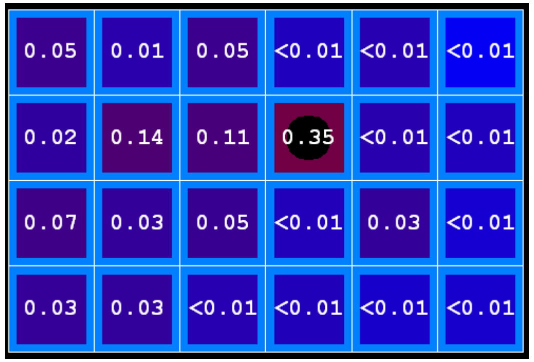
0.05	0.01	0.05	<0.01	<0.01	<0.01
0.02	0.14	0.11	0.35	<0.01	<0.01
0.07	0.03	0.05	<0.01	0.03	<0.01
0.03	0.03	<0.01	<0.01	<0.01	<0.01

T = 5



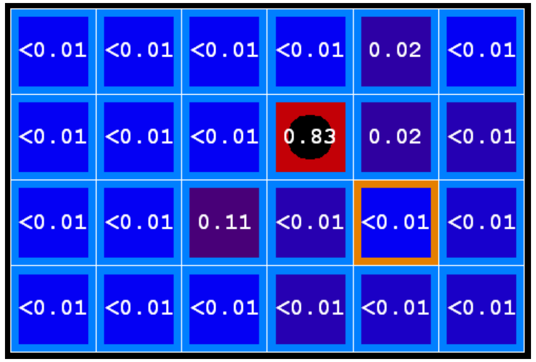
Example: Update step

As we get observations, beliefs get reweighted, uncertainty “decreases”



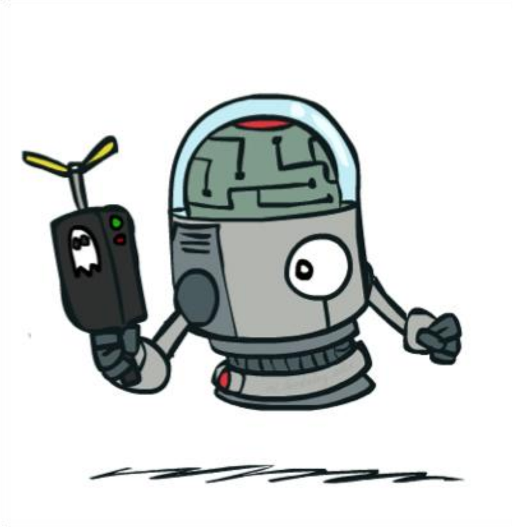
0.05	0.01	0.05	<0.01	<0.01	<0.01
0.02	0.14	0.11	0.35	<0.01	<0.01
0.07	0.03	0.05	<0.01	0.03	<0.01
0.03	0.03	<0.01	<0.01	<0.01	<0.01

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



Demo Ghostbusters – Circular Dynamics -- HMM

Filtering Algorithm

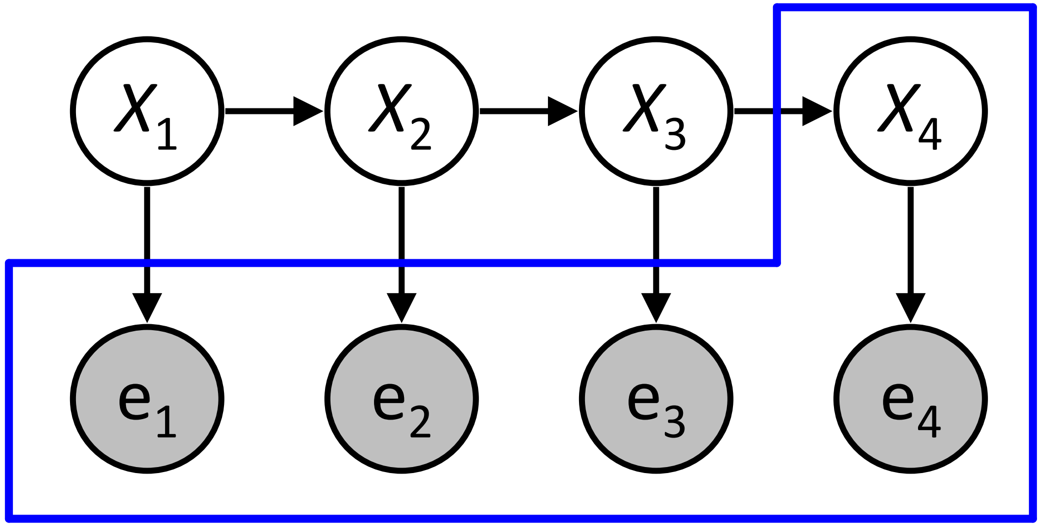
Query: What is the current state, given all of the current and past evidence?

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?

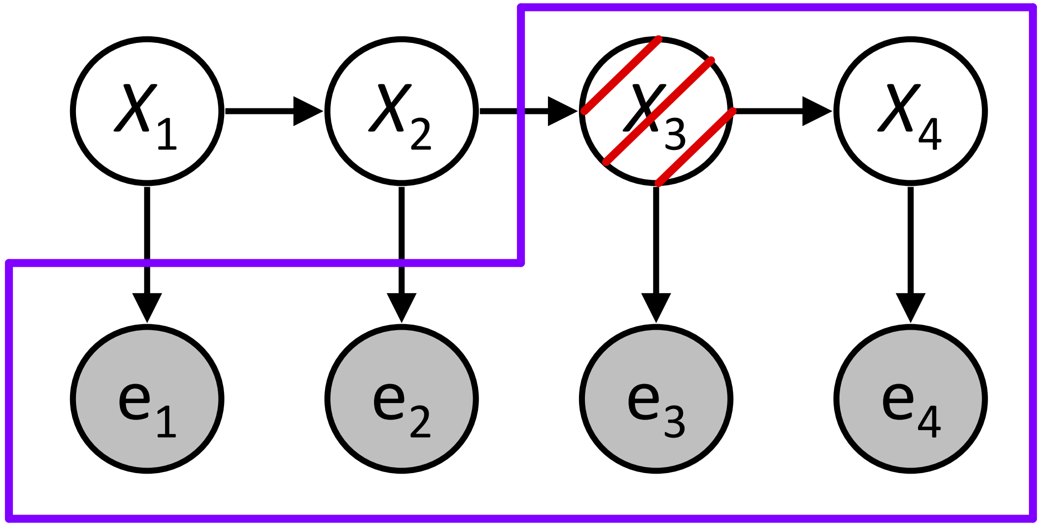
Matching math with Bayes net

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

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

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

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

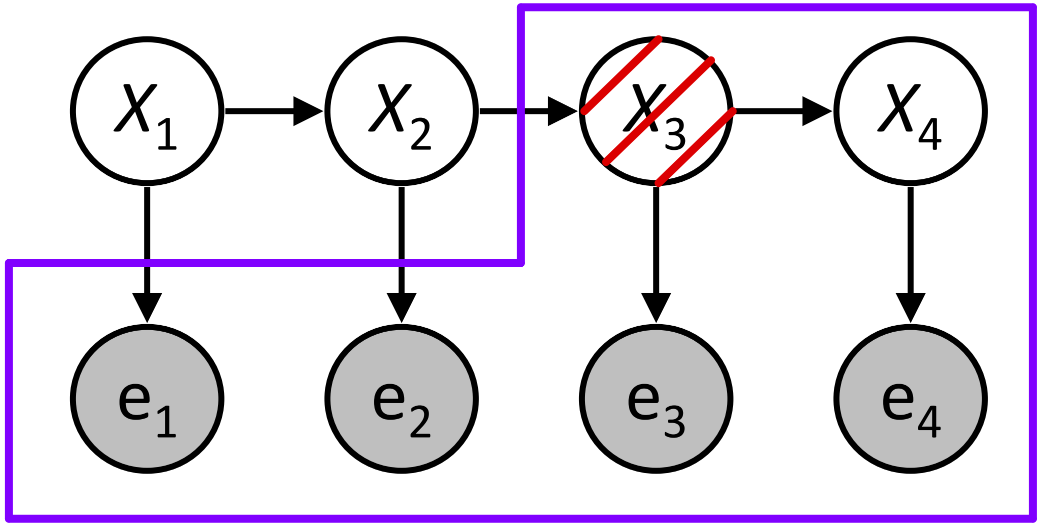
Chain rule with x_{t-1}, X_t, and e_t

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

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

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

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



Filtering Algorithm

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

Matching math with Bayes net

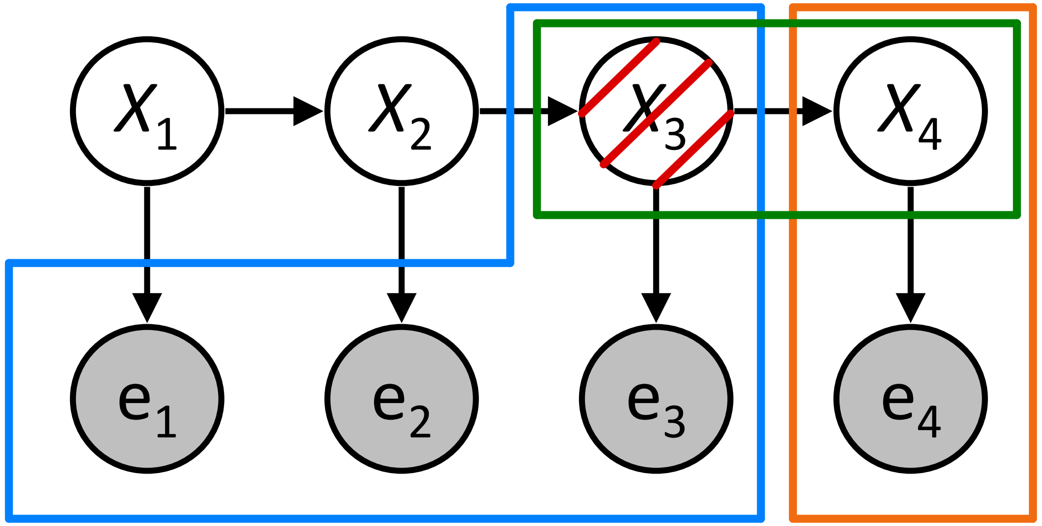
$$P(X_t \mid e_{1:t}) = P(X_t \mid e_t, e_{1:t-1})$$

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

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

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

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?

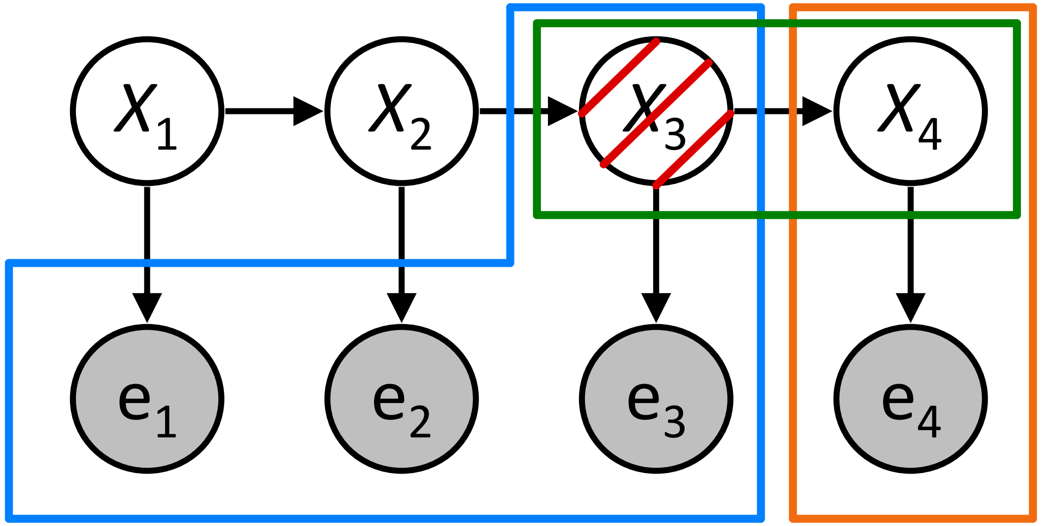
Matching math with Bayes net

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

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

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

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



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

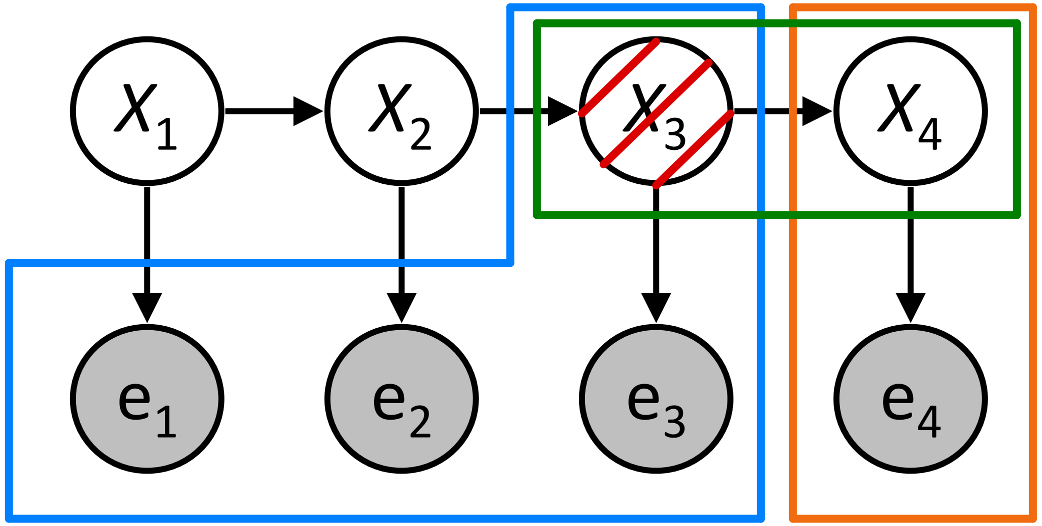
$$P(X_t \mid e_{1:t}) = P(X_t \mid e_t, e_{1:t-1})$$

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

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

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

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



Pulling $P(e_t|X_t)$ 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

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

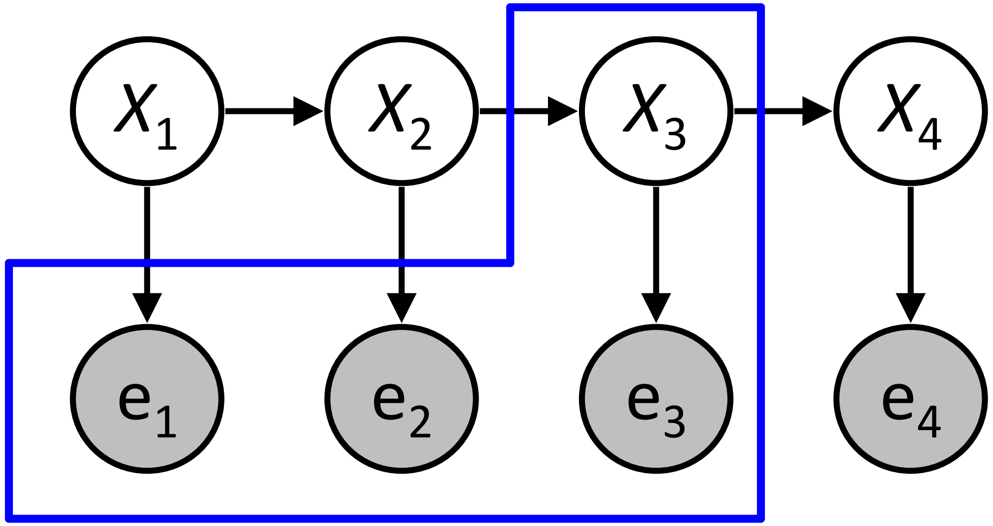
$$= \alpha\, P(X_t, e_t \mid e_{1:t-1})$$

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

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

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

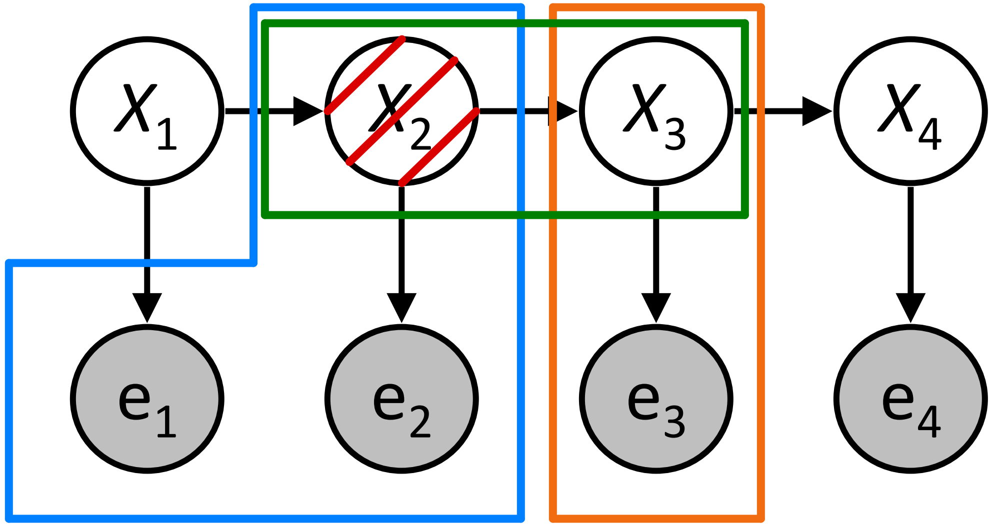
Recursion!



Filtering Algorithm

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

Matching math with Bayes net



Recursion!

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

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

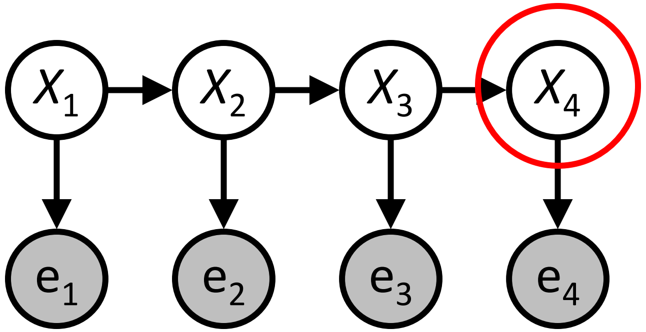
$$= \alpha \sum_{x_{t-1}} P(x_{t-1}, X_t, e_t \mid e_{1:t-1})$$

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

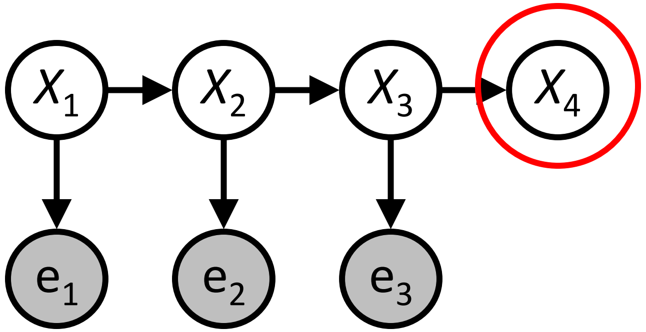
$$= \alpha\, P(e_t \mid x_t) \sum_{x_{t-1}} P(x_t \mid x_{t-1})\, P(x_{t-1} \mid e_{1:t-1})$$

Other HMM Queries

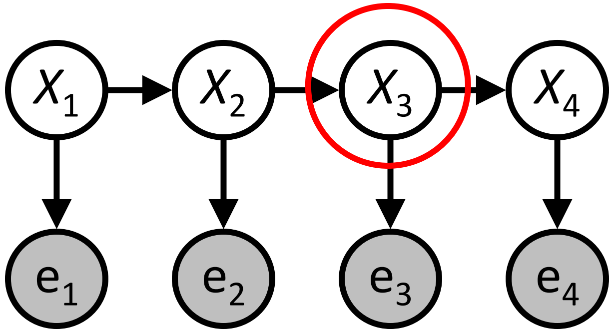
Filtering: $P(X_t | e_{1:t})$



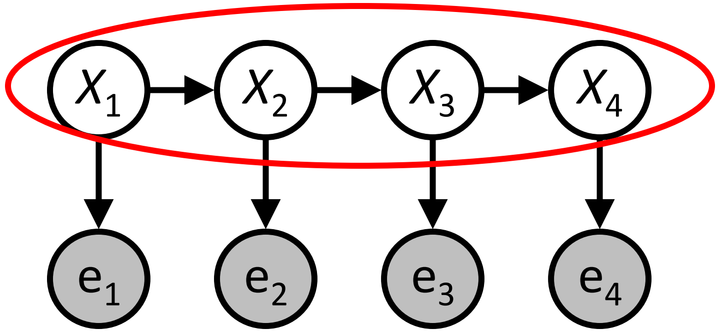
Prediction: $P(X_t | e_{1:t-1})$



Smoothing: $P(X_t | e_{1:N})$, $t<N$



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



Demo: Pacman Ghostbusters

[Demos: ghostbusters (L14D1,3)]

Filtering Algorithm

$$P(X_{t+1} \mid e_{1:t+1}) = \underbrace{\alpha}_{\text{Normalize}} \; \underbrace{P(e_{t+1} \mid X_{t+1})}_{\text{Update}} \; \underbrace{\sum_{x_t} P(X_{t+1} \mid x_t)\, P(x_t \mid e_{1:t})}_{\text{Predict}}$$

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

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