AI: Representation and Problem Solving

Sequential Data and Hidden Markov Models

Instructors: Nihar B. Shah and Tuomas Sandholm

Slide credits: CMU AI and ai.berkeley.edu

Sequential data

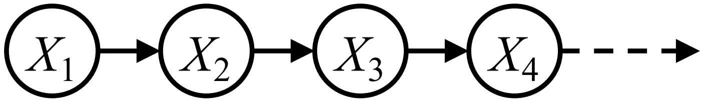
- Finance
- Speech recognition
- Robot localization
- User attention
- Medical monitoring

Need to introduce time (or space) into our models

Today

- Two popular models for sequential data
- Markov chains and hidden Markov models (HMMs)
- Used widely in many applications
- Also form building blocks for more complex models

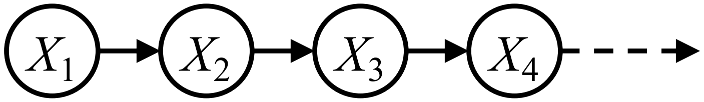
Markov Chains



- Let X denote the quantity of interest (e.g., stock price)
- Consider discrete time (e.g., days)
- Let X_t denote random variable for the value of X (stock price) at time t (i.e., day t)
- Possible values of X at a given time are called the states

- Initial state probabilities: Probability distribution of X_1
- Transition probabilities or dynamics: $P(X_t|X_{t-1})$ specify how the state evolves over time
- Stationarity assumption: transition probabilities same at all times, i.e., $P(X_t|X_{t-1}) = P(X_{t'}|X_{t'-1})$
- Same as MDP transition model, but no choice of action, no rewards

Conditional Independence

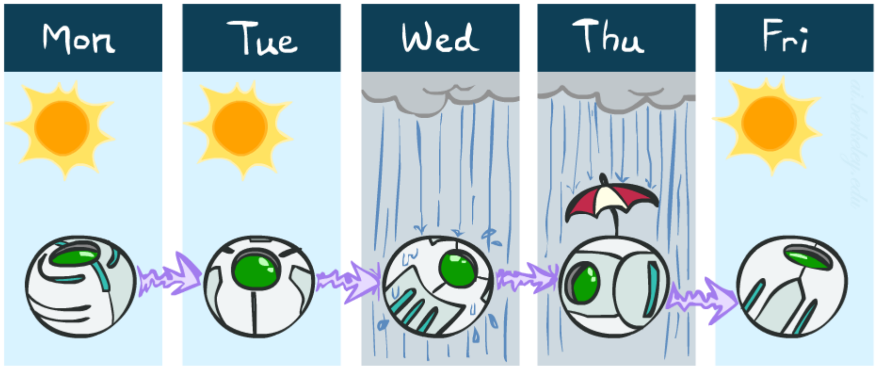


- Past and future independent given the present
- Each time step only depends on the previous
- This is called the (first order) Markov property

Note that the chain is just a (growable) Bayes net

Example: Markov Chain Weather

States: X = {rain, sun}

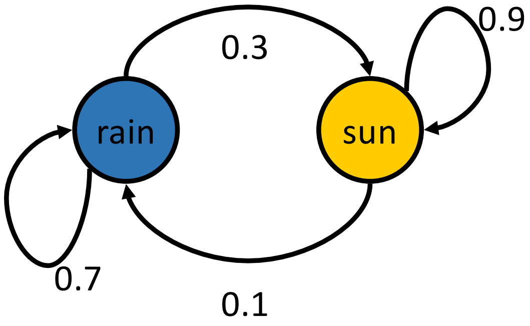


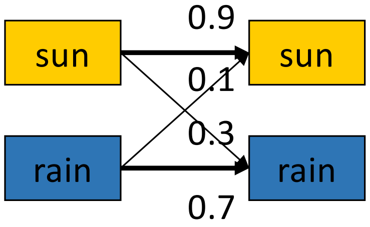
- Initial distribution: 1.0 sun

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

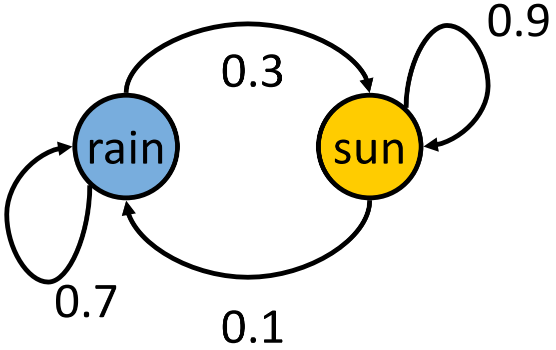
X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Two new ways of representing the same CPT





Example: Markov Chain Weather

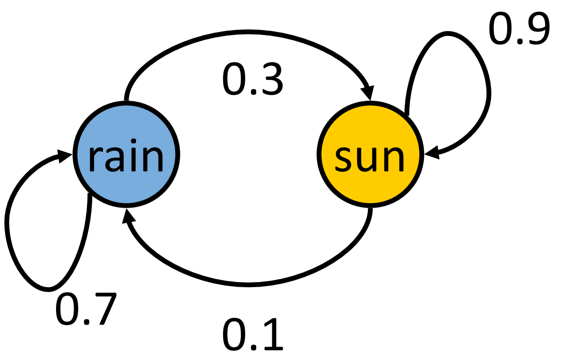


Initial distribution: $P(X_1 = sun) \ = \ 1.0$

What is the probability distribution after one step?

$P(X_2 = sun) \ = ?$

Example: Markov Chain Weather



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

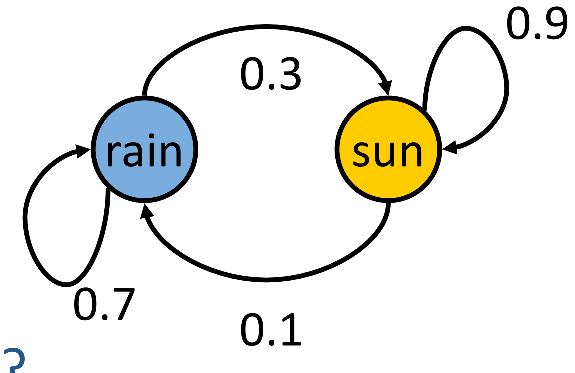
What is the probability distribution after one step?

$P(X_2 = sun) = ?$

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

Question

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

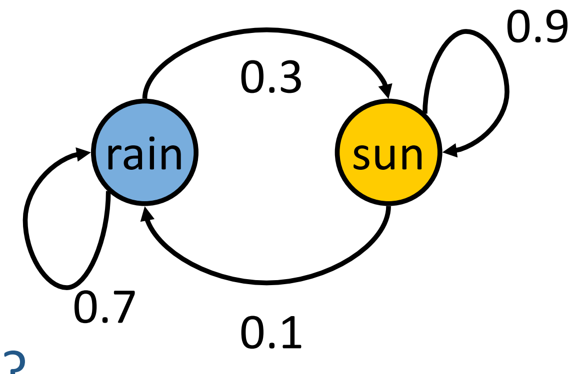


What is the probability distribution after the next step?

$P(X_3 = sun) \ = ?$

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

Question



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

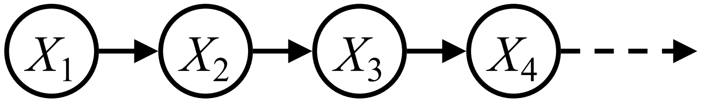
$$P(X_3 = sun) = \sum_{x_2} P(X_3 = sun, X_2 = x_2)$$

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

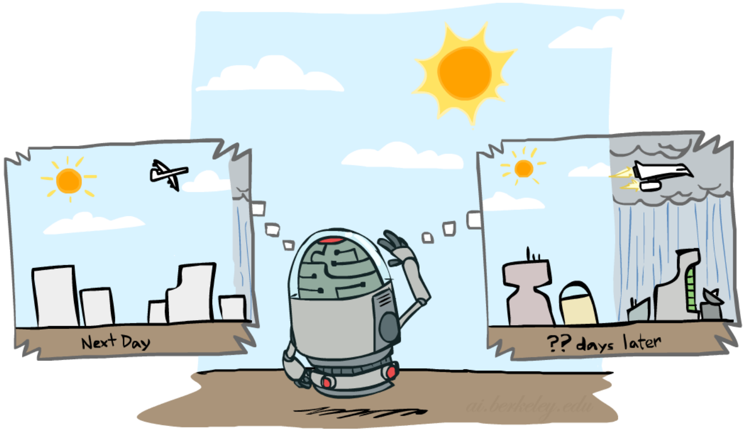
$$= 0.9 \cdot 0.9 + 0.3 \cdot 0.1$$

$$= 0.81 + 0.03 = 0.84$$

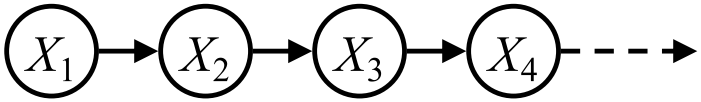
Markov Chain Inference



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



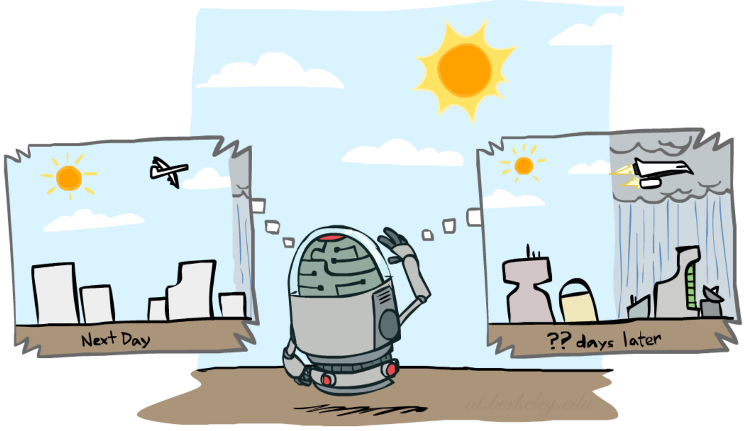
Markov Chain Inference



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

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

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



More generally

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*=1

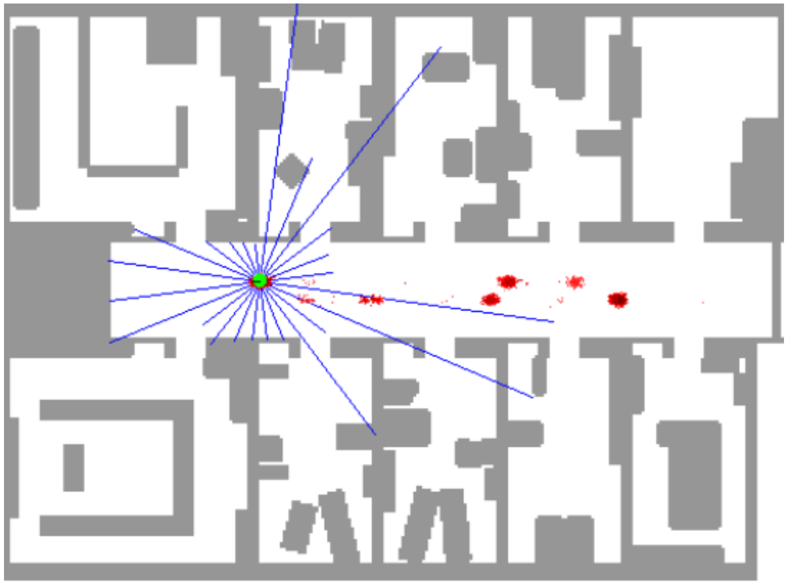
Hidden Markov Models

Hidden Markov Models

In many applications, the true state is not observed directly.
Instead, you observe some possibly noisy information.

Robot tracking:

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



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

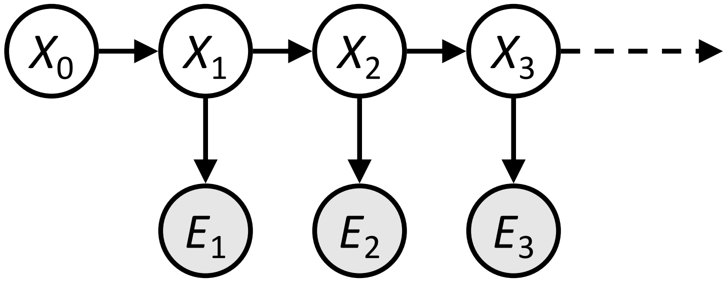
Molecular biology:

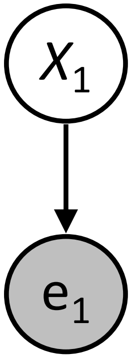
- Observations are nucleotides ACGT
- States are coding/non-coding/start/stop/splice-site etc.

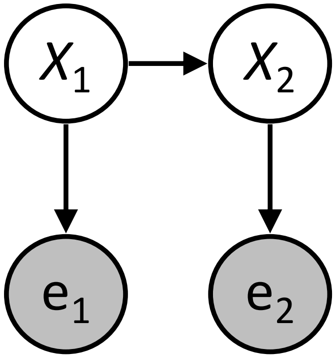
Hidden Markov Models

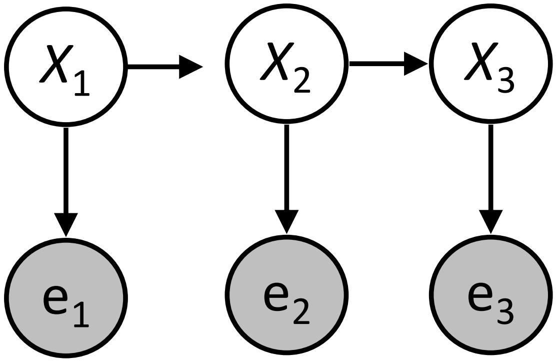
In many applications, the true state is not observed directly.
Instead, you observe some possibly noisy information.

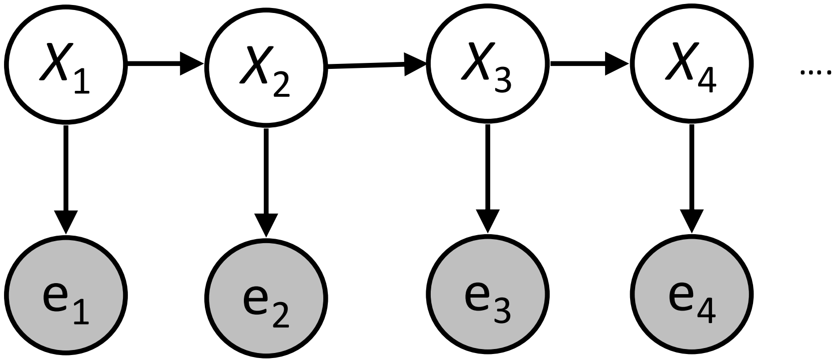
- Underlying Markov chain over states X
- You observe evidence E_t at each time step




X_1
e_1


X_1
X_2
e_1
e_2

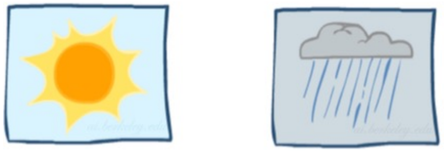

X_1
X_2
X_3
e_1
e_2
e_3


X_1
X_2
X_3
X_4
….
e_1
e_2
e_3
e_4

An HMM is defined by:

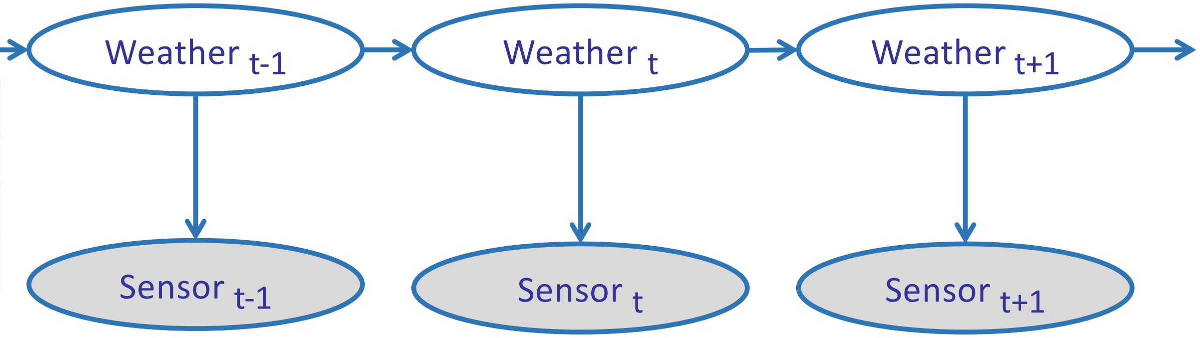
- Initial distribution: $P(X_1)$
- Transition model: $P(X_t \mid X_{t-1})$
- Sensor model: $P(E_t \mid X_t)$

Example: Weather HMM

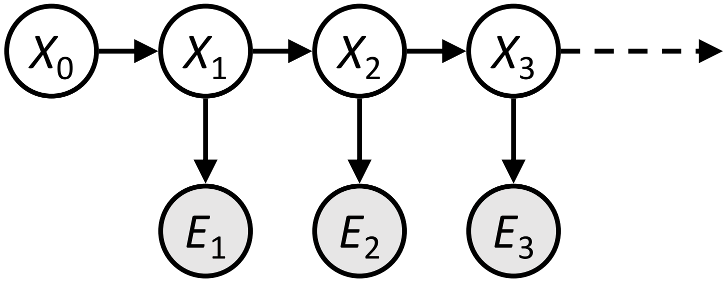


X_{t-1}	$P(X_t \mid X_{t-1})$	
	sun	rain
sun	0.9	0.1
rain	0.3	0.7

X_t	$P(E_t \mid X_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1



HMM as Probability Model



- Joint distribution for Markov model:

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

- Joint distribution for hidden Markov model:

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

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

Some useful stuff

Notation: $X_{a:b} = (X_a, X_{a+1}, \ldots, X_b)$

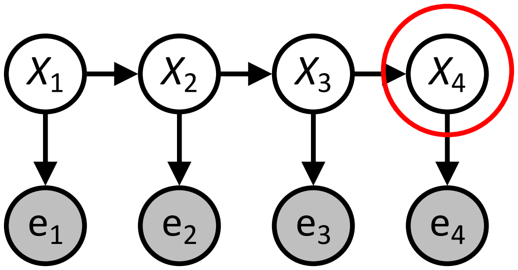
For example: $P(X_{1:2} \mid e_{1:3}) = P(X_1, X_2, \mid e_1, e_2, e_3)$

Probability: Consider a random variable B taking three possible values b_1,b_2,b_3. Suppose you know that $P(b_1)=2\alpha$, $P(b_2)=1.25\alpha$, $P(b_3)=0.75\alpha$, for some $\alpha>0$. Then what is $P(b_1)$?

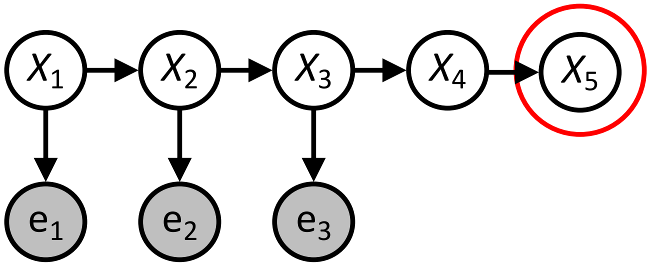
Key takeaway: If you know P(B=b) for all b up to a constant, then you can recover P(B) by normalizing.

HMM Queries

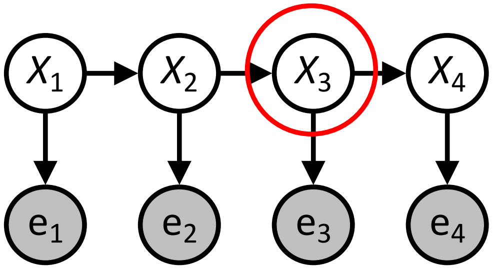
Filtering: $P(X_t | e_{1:t})$



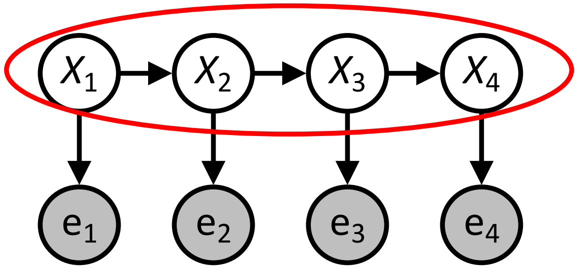
Prediction: $P(X_{t+k} | e_{1:t})$



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



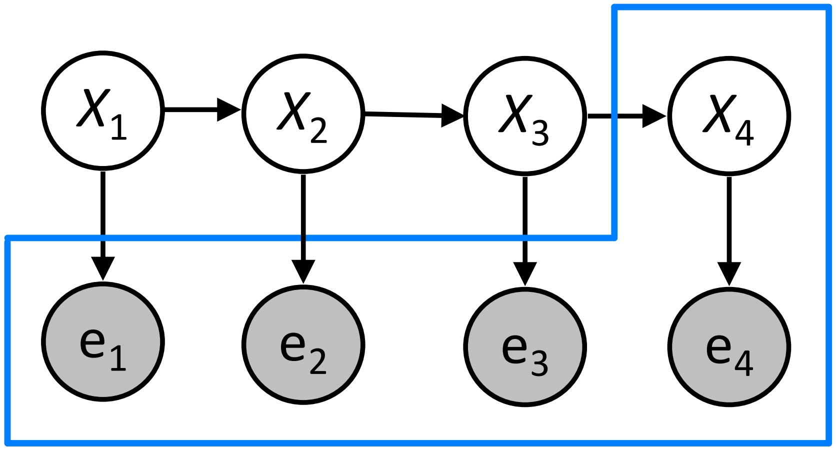
Explanation: $P(X_{1:t} | e_{1:t})$



Filtering

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

That is, what is $P(X_t|e_{1:t})$?



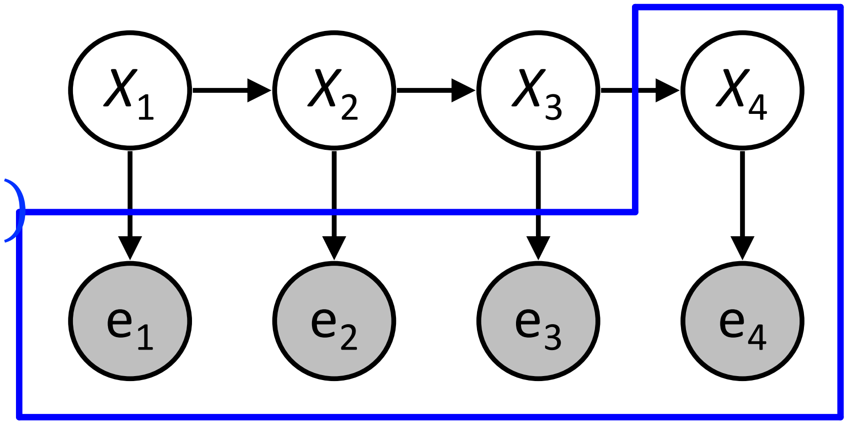
Filtering Algorithm: Exact inference

Def. of cond. probability with X_t, e_t

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

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

does not depend on X_t



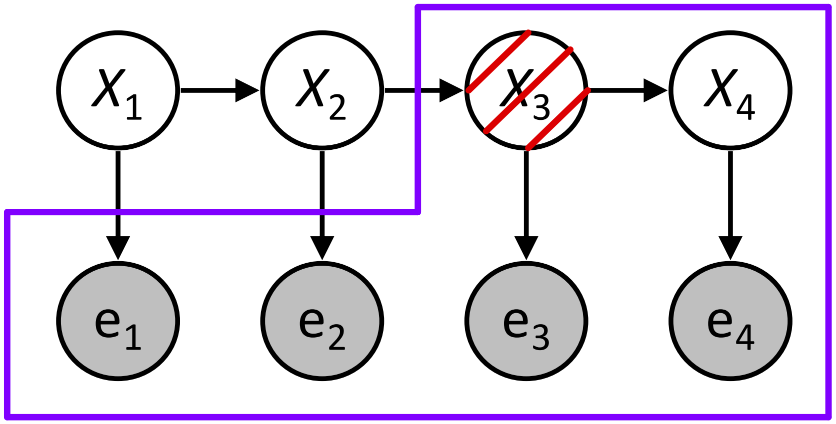
Filtering **Algorithm**

Summation over variable X_{t-1}

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

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

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



Filtering **Algorithm**

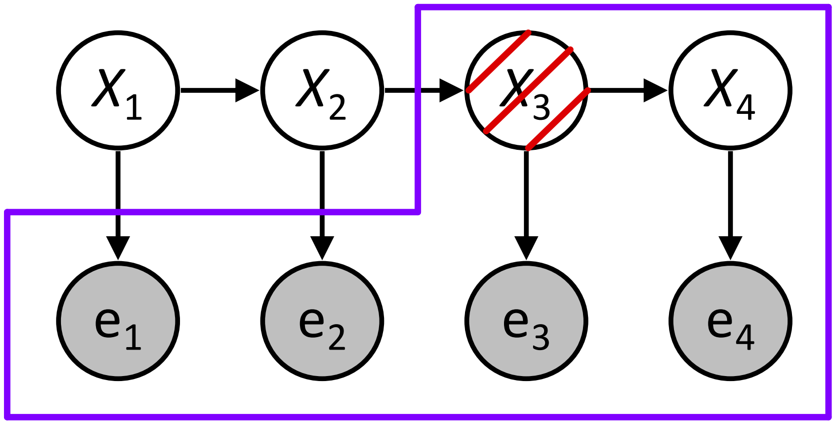
Chain rule with x_{t-1}, X_t, and e_t

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

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

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

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



Filtering Algorithm

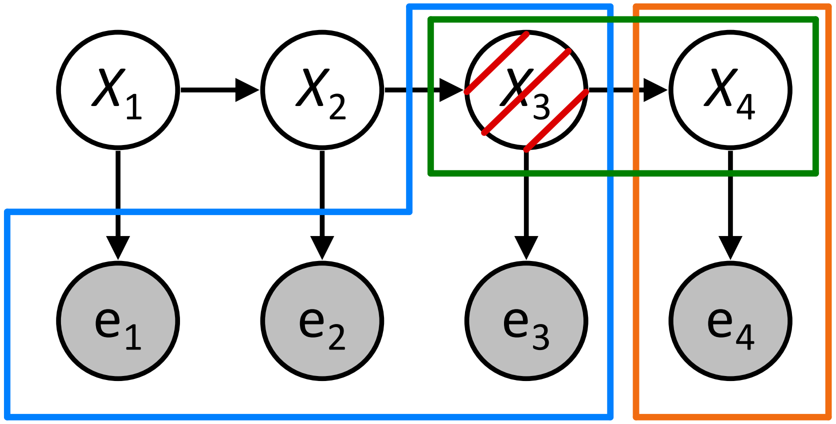
Chain rule with x_{t-1}, X_t, and e_t

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

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

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

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



Filtering **Algorithm**

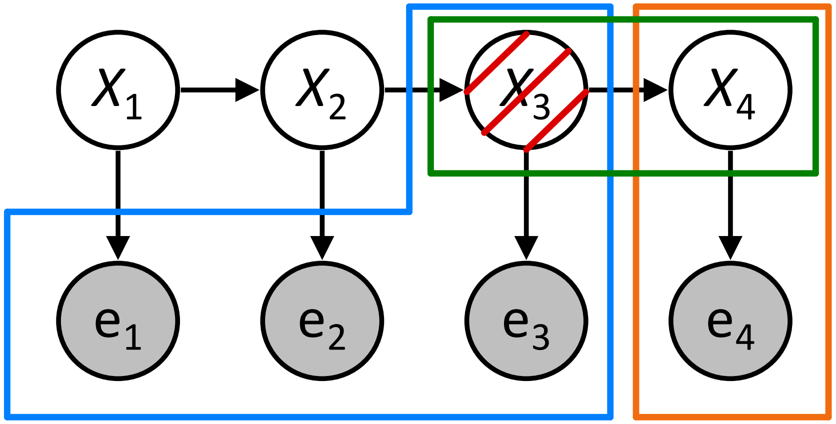
Bayes net conditional independence

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

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

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

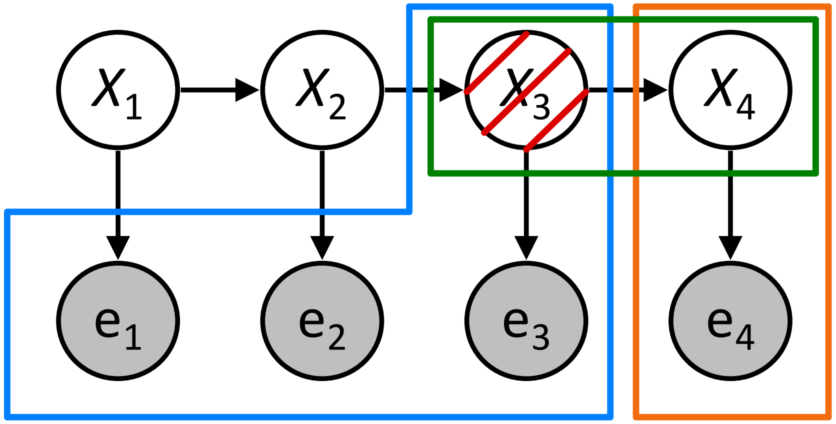
$$= \alpha \sum_{x_{t-1}} P(x_{t-1} \mid e_{1:t-1})\, P(X_t \mid x_{t-1}, e_{1:t-1})\, P(e_t \mid X_t)$$



Filtering **Algorithm**

Bayes net conditional independence

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



Filtering **Algorithm**

Pulling $P(e_t|X_t)$ out of the summation

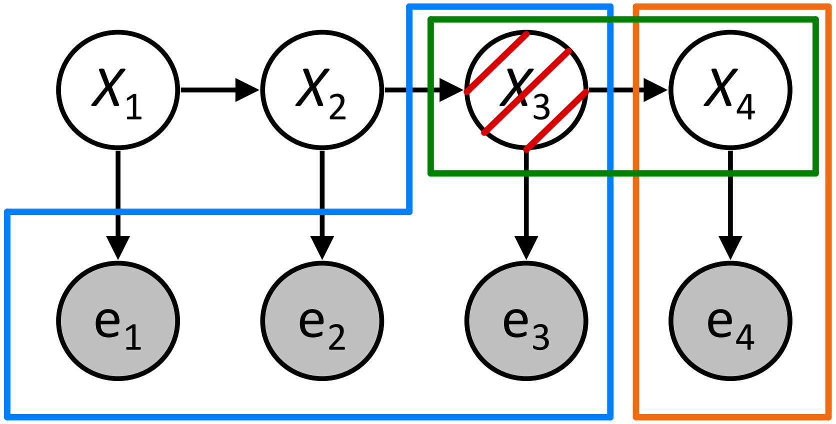
$$P(X_t \mid e_{1:t}) = P(X_t \mid e_t, e_{1:t-1})$$

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

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

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

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



Filtering **Algorithm**

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

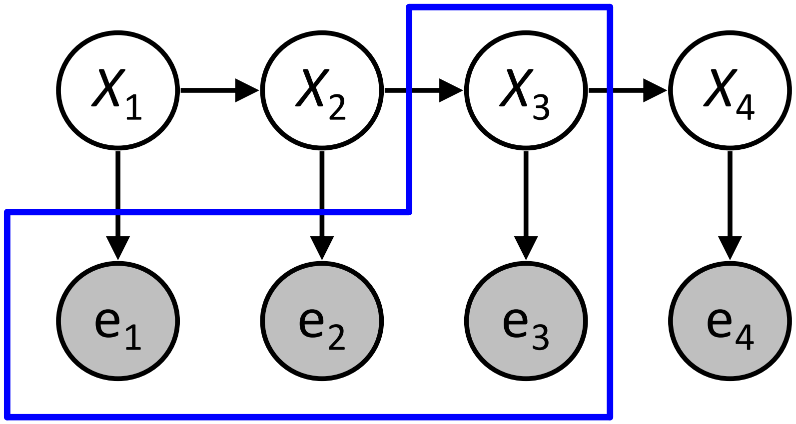
$$= \alpha \, P(X_t, e_t \mid e_{1:t-1})$$

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

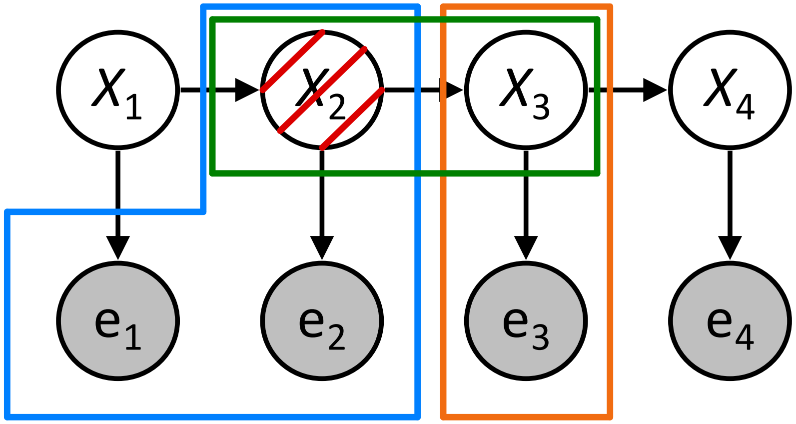
$$= \alpha \sum_{x_{t-1}} P(x_{t-1} \mid e_{1:t-1}) \, P(X_t \mid x_{t-1}) \, P(e_t \mid X_t)$$

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

Recursion!



Filtering **Algorithm**



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

Recursion!

In Class Activity: Weather HMM

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

Goal: Compute $P(X_2\text{=sun} \mid e_2\text{=}\ e_1\text{=True})$

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

X_t	$P(E_t \mid X_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Remember recursion… start with $P(X_1|e_1)$

$$P(X_1|e_1) = \frac{P(X_1,e_1)}{P(e_1)}$$
$$= \alpha\, P(e_1|X_1)P(X_1)$$

$$P(X_1 = \text{sun}|e_1 = \text{True})$$
$$= \alpha \ * 0.2 * 0.5 = 0.1\ \alpha$$

$$P(X_1 = \text{rain}|e_1 = \text{True}) = \alpha * 0.9 * 0.5 = 0.45\ \alpha$$

In Class Activity: Weather HMM

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

Goal: Compute $P(X_2=sun \mid e_2= e_1=True)$

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

X_t	$P(E_t \mid X_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Next, move to $P(X_2|e_1,e_2)$

$$P(X_2|e_1,e_2) = \alpha' P(X_2, e_2|e_1) = \alpha' P(e_2|X_2, e_1)P(X_2|e_1) = \alpha' P(e_2|X_2)P(X_2|e_1)$$

where $P(X_2|e_1) = \sum_{x_1} P(X_2|x_1)P(x_1|e_1)$

$$P(X_2 = sun|e_1 = True) = \sum_{x_1} P(X_2 = sun|x_1)P(x_1|e_1 = True)$$

$$= .9 * .1\alpha + .3 * .45\alpha = .225\alpha$$

$$P(X_2 = rain|e_1 = True) = \sum_{x_1} P(X_2 = rain|x_1)P(x_1|e_1 = True)$$

$$= .1 * .1\alpha + .7 * .45\alpha = .325\alpha$$

In Class Activity: Weather HMM

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

Goal: Compute $P(X_2\text{=sun} \mid e_2\text{=} e_1\text{=True})$

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

X_t	$P(E_t \mid X_t)$	
	true	false
sun	0.2	0.8
rain	0.9	0.1

Next, move to $P(X_2|e_1,e_2)$

$$P(X_2|e_1,e_2) = \alpha' P(X_2, e_2|e_1) = \alpha' P(e_2|X_2)P(X_2|e_1)$$

$P(X_2 = \text{sun}|e_1 = \text{True}) = .225\alpha$

$P(X_2 = \text{rain}|e_1 = \text{True}) = .325\alpha$

$P(X_2 = \text{sun}|e_1, e_2 = \text{True}) = \alpha' * 0.2 * .225\alpha = .045\ \alpha\ \alpha'$

$P(X_2 = \text{rain}|e_1, e_2 = \text{True}) = \alpha' * 0.9 * .325\ \alpha = 0.2925\ \alpha\ \alpha'$

Normalizing, we have $P(X_2 = \text{sun}|e_1, e_2 = \text{True}) = .045 / (.045 + 0.2925) = 0.133$

and $P(X_2 = \text{rain}|e_1, e_2 = \text{True}) = .2925 / (.045 + 0.2925) = 0.867$

Filtering Algorithm: Computational complexity

$$P(X_t|e_{1:t}) = = \underbrace{\alpha}_{\text{Normalize}} \underbrace{P(e_t|X_t)}_{\text{Update}} \underbrace{\sum_{x_{t-1}} P(X_t|x_{t-1})\, P(x_{t-1}|\, e_{1:t-1})}_{\text{Predict}}$$

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

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

Next lecture: Approximations!