PROBABILISTIC REASONING OVER TIME

Today

- Reading
  - AIMA Chapter 15.1-15.2, 15.5
- Goals
  - Reasoning with uncertainty over time
  - Types of inference
    - Filtering, prediction, smoothing, most likely explanation

Midterm exam details

- Next Wednesday March 12th
- Need to find a time (1.5 hours)
  - Can people come in 40 minutes before class?
- 4 sections
  - short answer
  - true/false
  - longer questions that ask you to carry out some algorithm

Midterm exam coverage

- Uninformed search
- Informed search and heuristics
- Local search
- Adversarial search
- CSPs
- Probability (subsumed by Bayesian networks)
- Bayesian networks

Modeling uncertainty over time

- Sometimes, we want to model a *dynamic* process: the value of the random variables change over time
  - Price of a stock
  - Patient stats, e.g. blood pressure, heart rate, blood sugar levels
  - Traffic on California highways
  - Pollution, humidity, temperature, rain fall, storms
  - Sensor tracking and detection

Modeling uncertainty over time

- Tracy got a new job working at the Coop. She works the late shift and doesn't get off until 2am. When she works the late shift, I often observe her eyes are red the next day. But sometimes she stays up late doing homework, and her eyes are red anyways.

- What are questions we might be interested in asking?
- How can we model this domain as a Bayesian network?

Modeling uncertainty over time

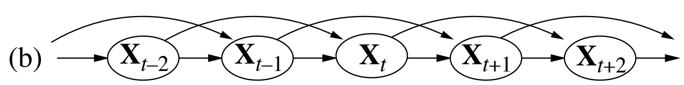
- Suppose we also know that if Tracy works the late shift one night she is less likely to work the late shift the next night.
- How does this change the Bayesian network?

States and Evidence

- Model a dynamic process as a series of time slices
- Each time slice contains a set of random variables
  - We observe the value of some random variables called the evidence. Often denoted as E_t
  - We don't observe the value of some random variables called the state. Often denoted as X_t

Transition Model

- We're often interested in reasoning about the state variables X_t given the history $X_{0:t-1}$
- Markov Assumption: the state variable X_t depends on a *bounded* subset of $X_{0:t-1}$
  - First order Markov Process: $P(X_t | X_{0:t-1}) = P(X_t | X_{t-1})$
  - Second order Markov Process: $P(X_t | X_{0:t-1}) = P(X_t | X_{t-1}, X_{t-2})$



Transition Model

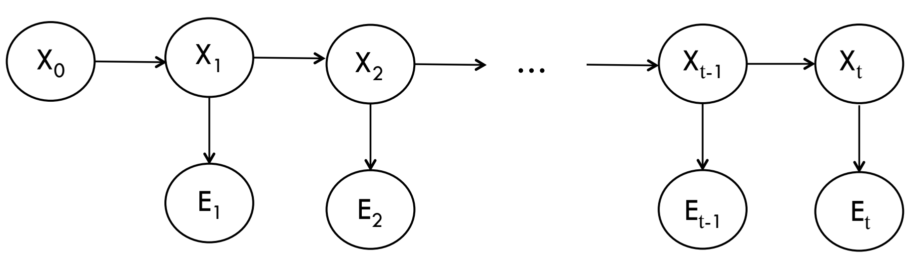
- We're often interested in reasoning about the state variables X_t given the history $X_{0:t-1}$
- Stationarity Assumption: the conditional distribution $P(X_t | X_{t-1})$ is the same for all t
  - Need to specify only one conditional distribution

Sensor (emission) model

- The state variables are responsible for generating (emitting) the evidence variables
- Sensor Markov Assumption: the evidence at time t is independent of every other random variable given the state at time t
  - As a result, your state should encompass all relevant information for specifying the evidence

Hidden Markov Model

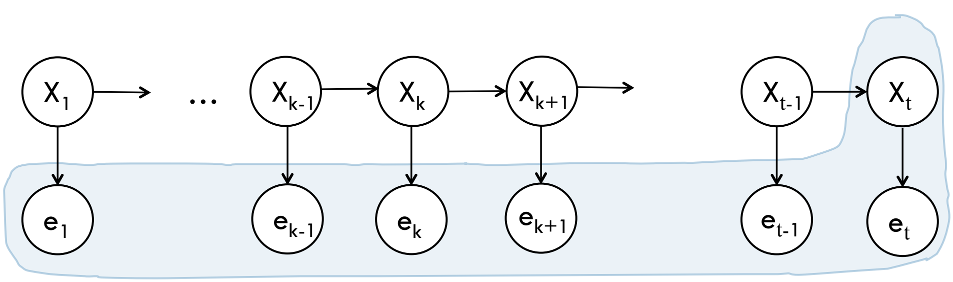
- Hidden Markov Models involve three things:
  - Transition model: $P(X_t | X_{t-1})$
  - Emission (evidence) model: $P(E_t | X_t)$
  - Prior probability: $P(X_0)$



Inference Tasks

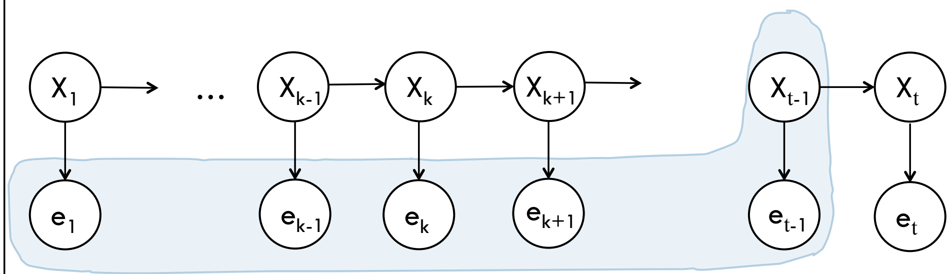
- Filtering: $P(X_t \mid e_{1:t})$
  - Decision making in the here and now
- Prediction: $P(X_{t+k} \mid e_{1:t})$
  - Trying to plan the future
- Smoothing: $P(X_k \mid e_{1:t})$ for $0 \leq k < t$
  - Gives a better (smoother) estimate than filtering by taking into account future evidence
- Most Likely Explanation (MLE): $\underset{x_{1:t}}{\operatorname{argmax}} P(x_{1:t} \mid e_{1:t})$
  - e.g., speech recognition, sketch recognition

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



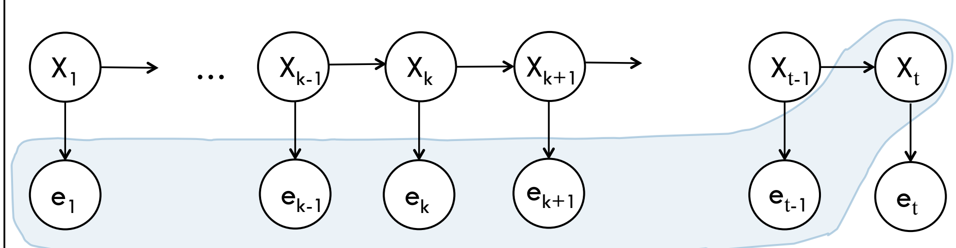
- A recursive state estimation algorithm

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



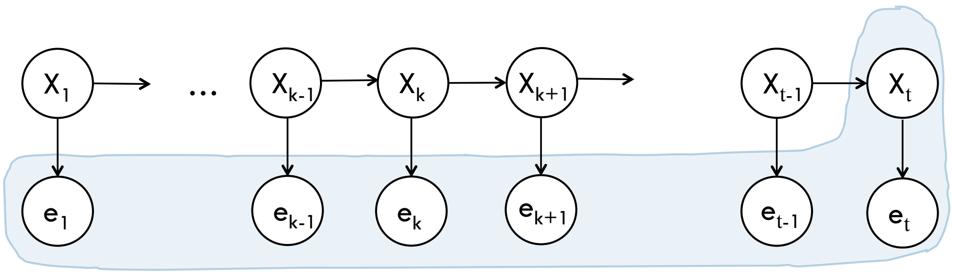
- Assume we already have $p(X_{t-1} \mid e_{1:t-1})$

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



- Update from state X_{t-1} to X_t

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



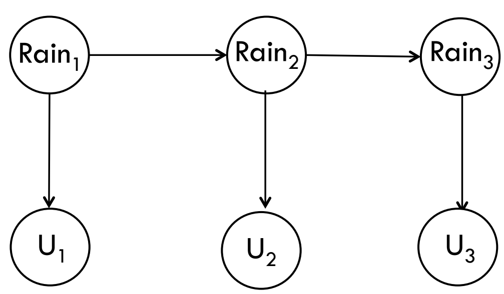
- Then incorporate the new evidence E_t

Filtering Example

$p(R_0)$ = <0.5, 0.5>

R_{t-1}	$p(R_t \mid R_{t-1})$
T	0.7
F	0.3

R_t	$p(U_t \mid R_t)$
T	0.9
F	0.2



$$p(X_t|e_{1:t}) \propto p(e_t|X_t) \sum_{X_{t-1}} p(X_t|X_{t-1})\, p(X_{t-1}|e_{1:t-1})$$

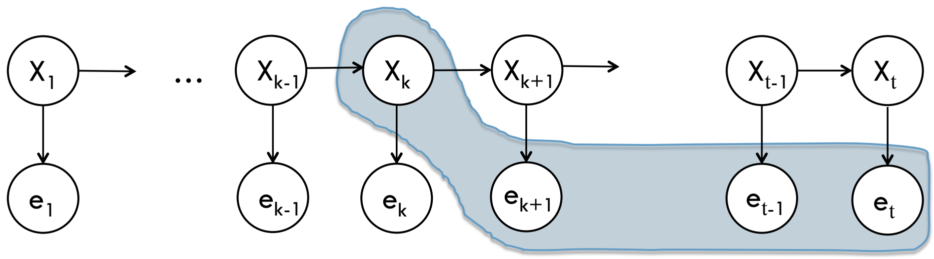
Prediction

- Compute $p(X_{t+k} \mid e_{1:t})$ for $k > 0$
- Given the equations for filtering, can you figure out how to do prediction?

Inference Tasks

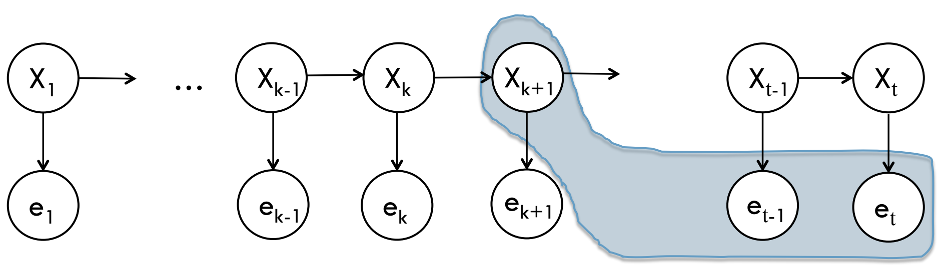
- Filtering: $P(X_t | e_{1:t})$
  - Decision making in the here and now
- Prediction: $P(X_{t+k} | e_{1:t})$
  - Trying to plan the future
- Smoothing: $P(X_k | e_{1:t})$ for $0 \leq k < t$
  - Gives a better (smoother) estimate than filtering by taking into account future evidence
- Most Likely Explanation (MLE): $\underset{x_{1:t}}{\text{argmax}}\, P(x_{1:t} | e_{1:t})$
  - e.g., speech recognition, sketch recognition

The Backward Algorithm



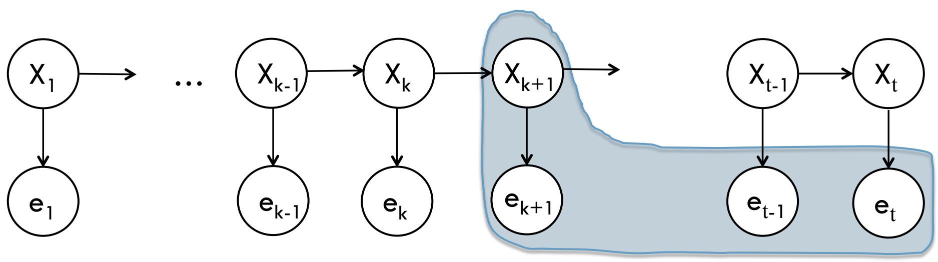
- A recursive state estimation algorithm

The Backward Algorithm



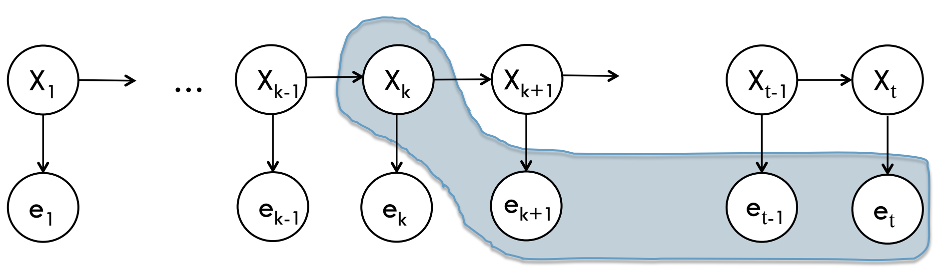
- Assume we have $p(X_{k+1} | e_{k+2:t})$

The Backward Algorithm



- Incorporate evidence via $p(e_{k+1} | X_{k+1})$

The Backward Algorithm



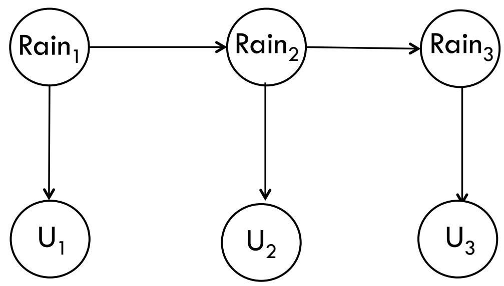
- Update the state via $p(X_{k+1} | X_k)$

Smoothing Example

$p(R_0)$ = <0.5, 0.5>

R_{t-1}	$p(R_t \mid R_{t-1})$
T	0.7
F	0.3

R_t	$p(U_t \mid R_t)$
T	0.9
F	0.2



$P(r_1 \mid u_1)$	$P(r_2 \mid u_1, u_2)$	$P(r_1 \mid u_1, u_2)$
0.818	0.883	?

Most Likely Explanation

- Find the state sequence that makes the observed evidence sequence most likely

$$\underset{X_{1:t}}{\text{argmax}}\ P(X_{1:t} \mid e_{1:t})$$

- Recursive formulation:
  - The most likely state sequence for $X_{1:t}$ is the most likely state sequence for $X_{1:t-1}$ followed by the transition to X_t
  - Equivalent to Filtering algorithm except summation replaced with max
  - Called the Viterbi Algorithm