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

Dynamic Bayesian Network

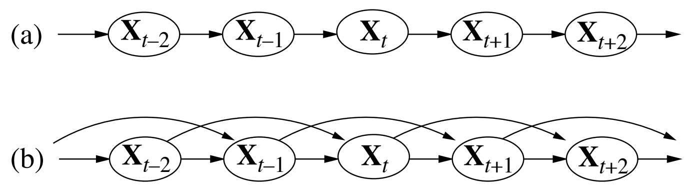
- Any BN that represents a temporal probability distribution using state variables and evidence variables is called a Dynamic Bayesian Network
- A Hidden Markov Model is the simplest type of DBN
  - State is represented by a single variable
  - Evidence is represented by a single variable
  - Applications
    - speech recognition
    - handwriting recognition
    - gesture recognition

Dynamic Bayesian Network

- 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

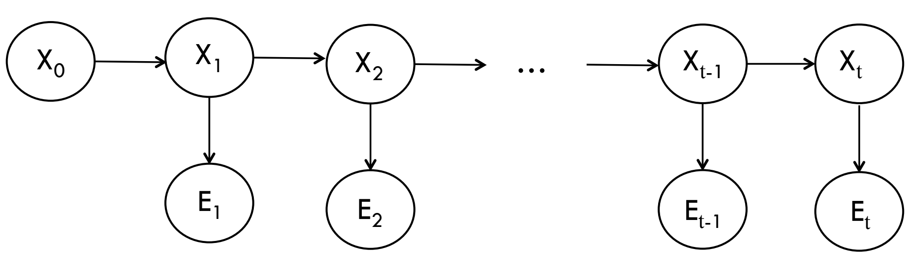
- We're often interested in reasoning about the state variables X_t given the history $X_{0:t-1}$
- Stationarity Assumption: the conditional distribution $P(X_t|X_{t-1})$ is the same for all t
  - Need to specify only one conditional distribution

Sensor (emission) model

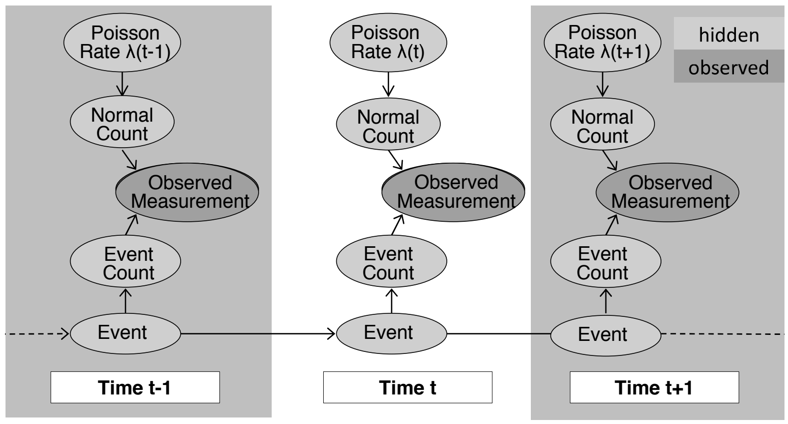
- The state variables are responsible for generating (emitting) the evidence variables
- Sensor Markov Assumption: the evidence at time t is independent of every other random variable given the state at time t, i.e. $p(E_t | X_{0:N}, E_{0:t-1}, E_{t+1:N}) = p(E_t | X_t)$
  - As a result, your state should 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: $P(X_0)$



Examples of DBN

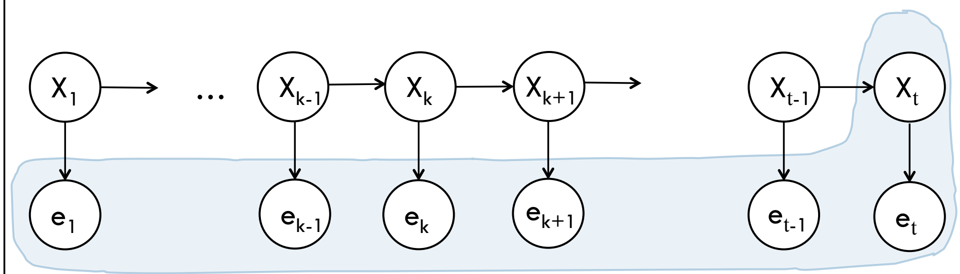


J. Hutchins et al. Probabilistic analysis of a large-scale urban traffic sensor data set.

Inference Tasks

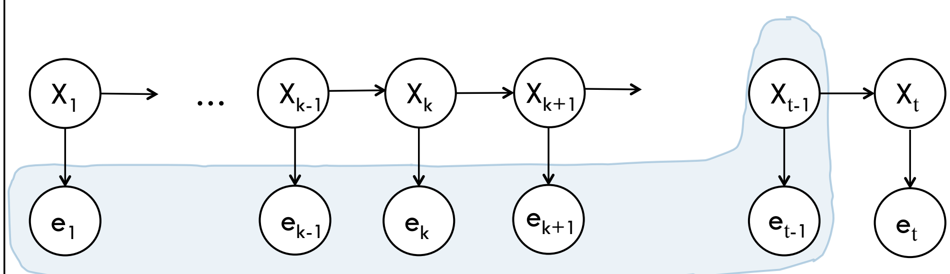
- Filtering: $P(X_t | e_{1:t})$
  - Decision making in the here and now
- Prediction: $P(X_{t+k} | e_{1:t})$
  - Trying to plan the future
- Smoothing: $P(X_k | e_{1:t})$ for $0 \leq k < t$
  - Gives a better (smoother) estimate than filtering by taking into account future evidence
- Most Likely Explanation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

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



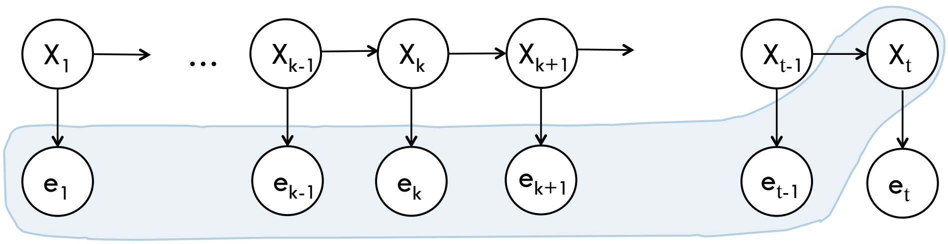
- A recursive state estimation algorithm

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



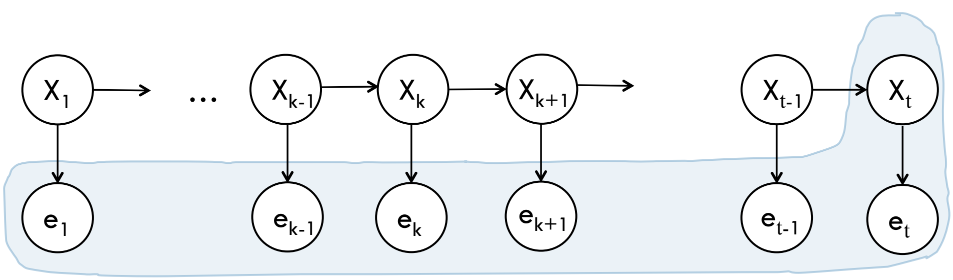
- Step Zero: Assume we already have $p(X_{t-1} | e_{1:t-1})$

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



- Step One: Update from state X_{t-1} to X_t

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



- Step Two: Then incorporate the new evidence E_t

The Forward Algorithm

$$p(X_t|e_{1:t}) = p(X_t|e_{1:t-1}, e_t)$$
$$\propto p(e_t|X_t, e_{1:t-1})\ p(X_t|e_{1:t-1})$$
$$= \underbrace{p(e_t|X_t)}_{\text{Incorporate evidence}}\ \underbrace{p(X_t|e_{1:t-1})}_{\text{Update state}}$$

The Forward Algorithm

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

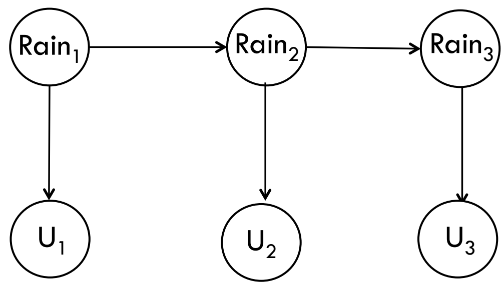
$$= p(e_t|X_t) \sum_{X_{t-1}} p(X_t, X_{t-1}|e_{1:t-1})$$
$$= p(e_t|X_t) \sum_{X_{t-1}} p(X_t|X_{t-1}, e_{1:t-1})\ p(X_{t-1}|e_{1:t-1})$$
$$= \underbrace{p(e_t|X_t)}_{\text{Emission}} \sum_{X_{t-1}} \underbrace{p(X_t|X_{t-1})\ p(X_{t-1}|e_{1:t-1})}_{\text{Transmission + recursion}}$$

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?

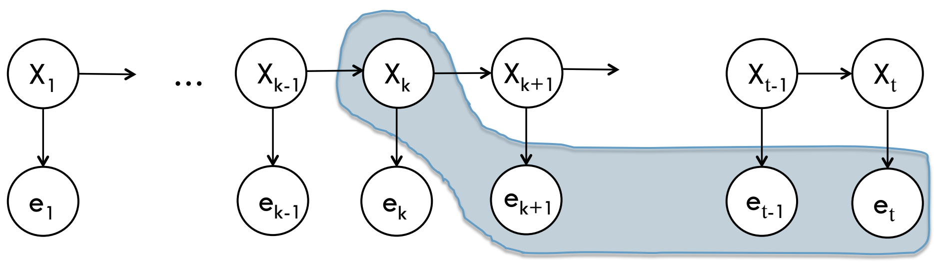
Inference Tasks

- 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): $\underset{x_{1:t}}{\operatorname{argmax}} P(x_{1:t} | e_{1:t})$
  - e.g., speech recognition, sketch recognition

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

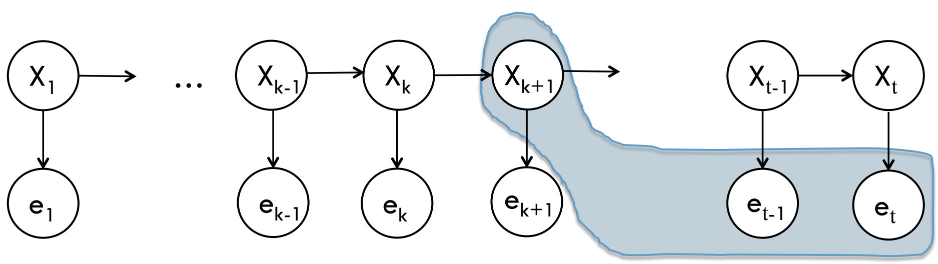
$$
\begin{aligned}
p(X_k|e_{1:t}) &= p(X_k|e_{1:k}, e_{k+1:t}) \\
&\propto p(X_k, e_{k+1:t}|e_{1:k}) \\
&= p(e_{k+1:t}|X_k, e_{1:k})\ p(X_k|e_{1:k}) \\
&= p(e_{k+1:t}|X_k)\ \underbrace{p(X_k|e_{1:k})}_{\text{Forward Algorithm}}
\end{aligned}
$$

The Backward Algorithm



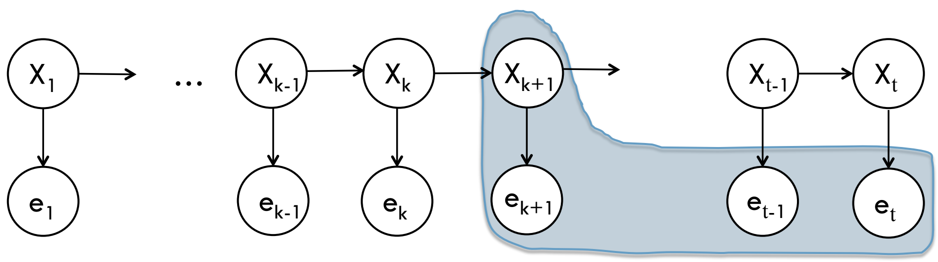
- A recursive state estimation algorithm

The Backward Algorithm



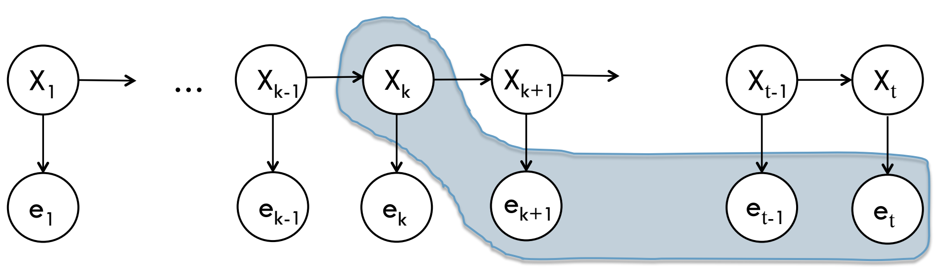
- Step Zero: Assume we have $p(X_{k+1} | e_{k+2:t})$

The Backward Algorithm



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

The Backward Algorithm



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

Smoothing: $p(X_k \mid e_{1:t})$ for $1 \leq k < t$

The Forward Backward Algorithm

$$p(X_k|e_{1:t}) = p(X_k|e_{1:k}, e_{k+1:t})$$
$$\propto p(X_k, e_{k+1:t}|e_{1:k})$$
$$= p(e_{k+1:t}|X_k, e_{1:k})\ p(X_k|e_{1:k})$$
$$= p(e_{k+1:t}|X_k)\ \underbrace{p(X_k|e_{1:k})}_{\text{Forward Algorithm}}$$

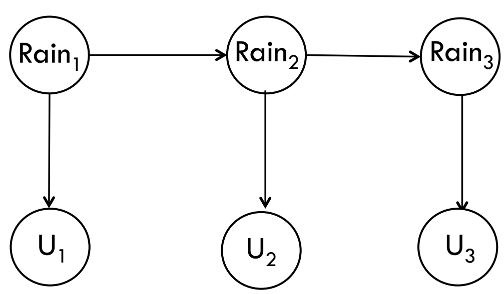
$$p(e_{k+1:t}|X_k) = \sum_{X_k+1} p(e_{k+1:t}, X_{k+1}|X_k)$$
$$= \sum_{X_k+1} p(e_{k+1:t}|X_{k+1})\ p(X_{k+1}|X_k)$$
$$= \sum_{X_k+1} \underbrace{p(e_{k+1}|X_{k+1})}_{\text{Emission}}\ \underbrace{p(e_{k+2:t}|X_{k+1})}_{\text{Recursion}}\ \underbrace{p(X_{k+1}|X_k)}_{\text{Transmission}}$$

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

$$\underset{X_{1:t}}{\operatorname{argmax}} P(X_{1:t} | e_{1:t})$$

- 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