PROBABILISTIC REASONING OVER TIME

Today

- Reading
  - AIMA Chapter 15.1-15.2, 15.5
- Goals
  - Introduce dynamic Bayesian networks
  - Forward algorithm
  - HW5 Ghostbusters

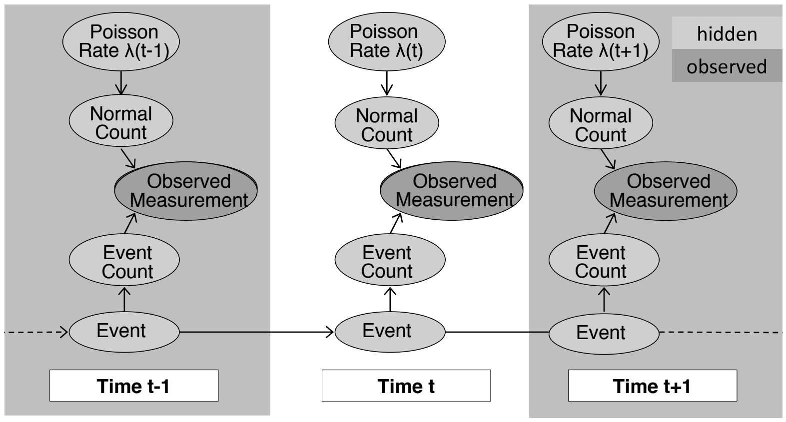
Modeling uncertainty over time

- 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

How to model a dynamic process?

- Model a dynamic process as a series of time slices
- Each time slice contains a set of random variables
  - Evidence variables whose value we observe (E_t)
  - State variables whose value we don't observe (X_t)
- Variables within a time slice are connected
- Variables between time slices are connected

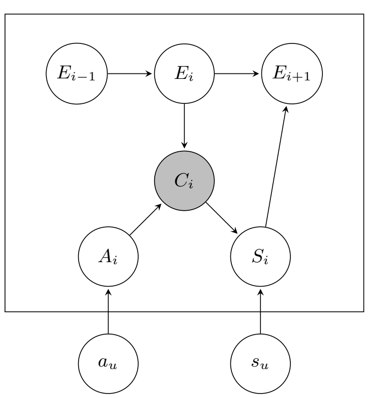
Examples of DBN



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

Examples of DBN

Track: Data Mining / Session: Click Models



- E_i: did the user *examine* the url?
- A_i: was the user *attracted* by the url?
- S_i: was the user *satisfied* by the landing page?

Figure 1: The DBN used for clicks modeling. C_i is the the only observed variable.

Examples of DBN

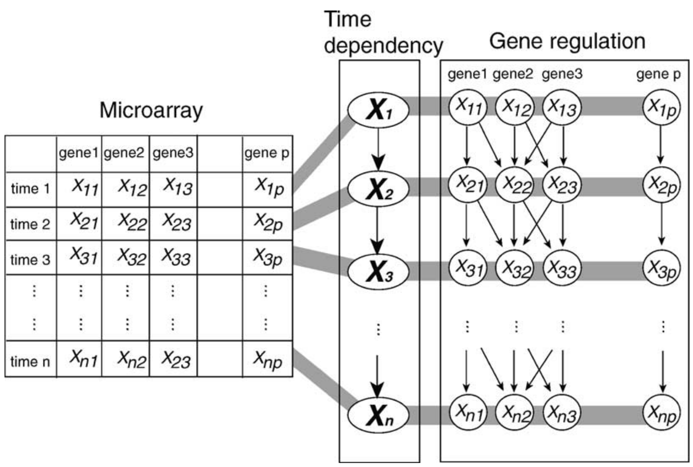
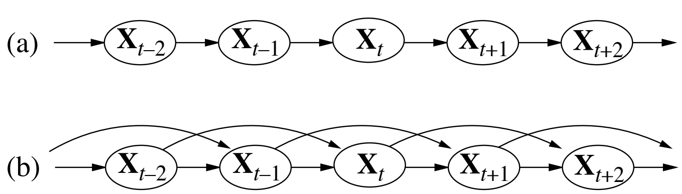


Fig. 2. Graphical view of a dynamic Bayesian network model.

Transition Model

- The transition model specifies the probability of X_t given the history $X_{0:t-1}$
- Markov Assumption: the state variable X_t depends on a *finite and fixed* 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

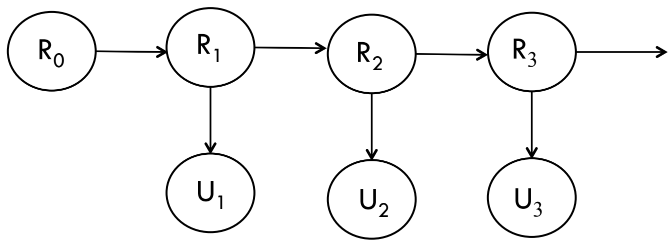
- Stationarity Assumption: the conditional distribution $P(X_t | X_{t-1})$ is the same for all t
- Only need to specify one conditional distribution for all edges

Sensor (emission) model

- The sensor model specifies the probability of the evidence E_t given the states $X_{0:t}$ and the evidence $E_{0:t-1}$
- Sensor Markov Assumption: the evidence E_t is independent of every other random variable given X_t
  - The state encompasses all relevant information for generating the evidence

Hidden Markov Model

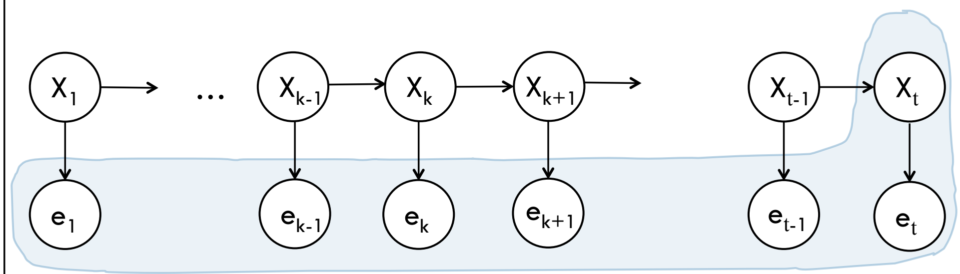
- A Hidden Markov Model is the simplest type of DBN
  - State is represented by a single discrete random variable



Inference Tasks

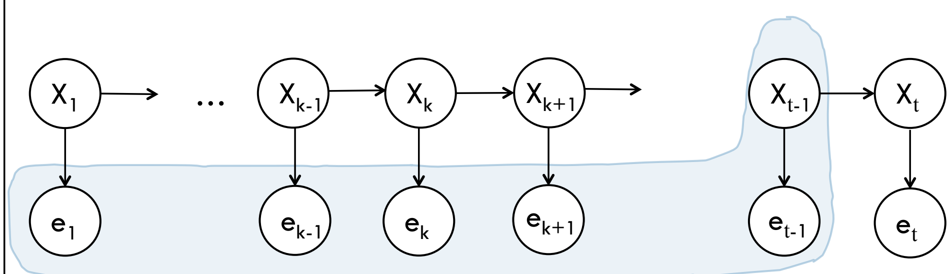
- Filtering: $P(X_t | e_{1:t})$
  - Decision making in the here and now
- Prediction: $P(X_{t+k} | e_{1:t})$
  - Trying to plan the future
- Smoothing: $P(X_k | e_{1:t})$ for $0 \leq k < t$
  - Gives a better (smoother) estimate than filtering by taking into account future evidence
- Most Likely Explanation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

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



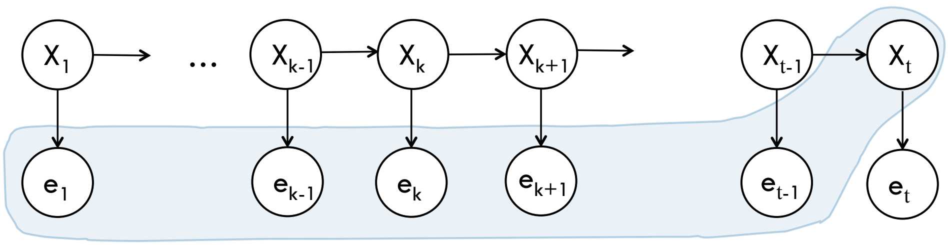
- A recursive state estimation algorithm

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



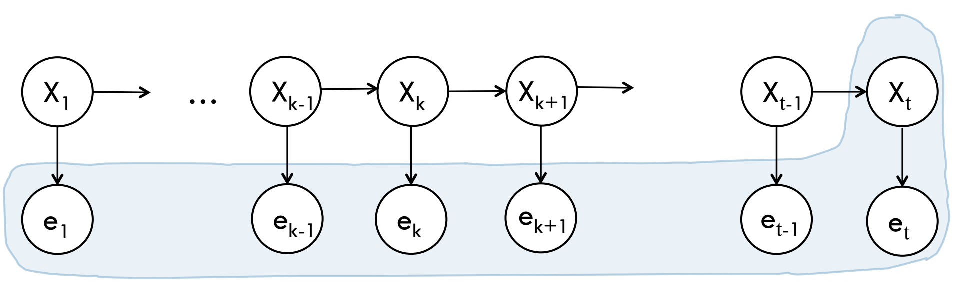
- Step Zero: Assume we already have $p(X_{t-1} | e_{1:t-1})$

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



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

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



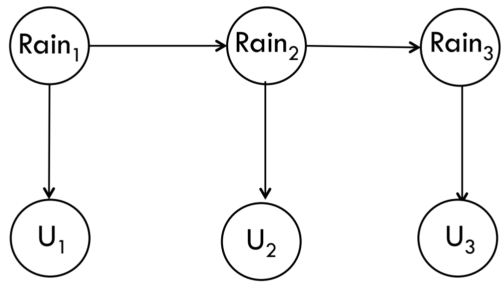
- Step Two: Then incorporate the new evidence E_t

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



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

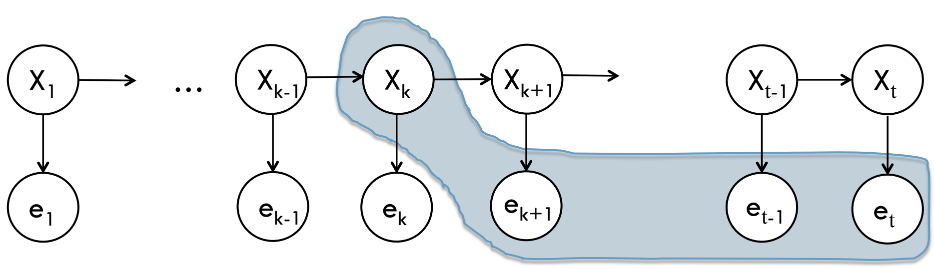
Prediction

- 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

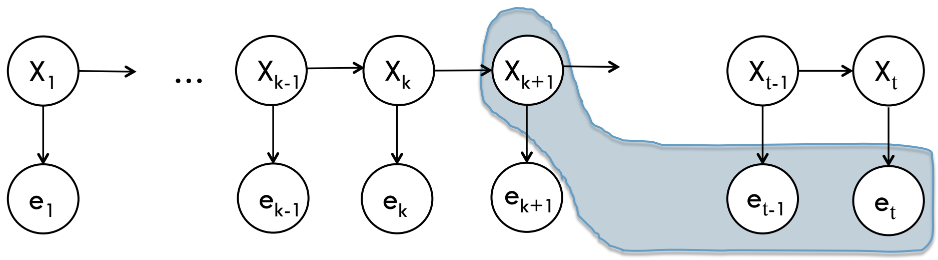
- Filtering: $P(X_t | e_{1:t})$
  - Decision making in the here and now
- Prediction: $P(X_{t+k} | e_{1:t})$
  - Trying to plan the future
- Smoothing: $P(X_k | e_{1:t})$ for $0 \leq k < t$
  - Gives a better (smoother) estimate than filtering by taking into account future evidence
- Most Likely Explanation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

The Backward Algorithm



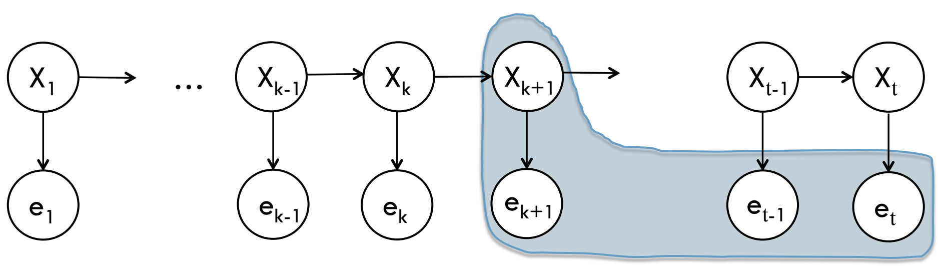
- A recursive state estimation algorithm

The Backward Algorithm



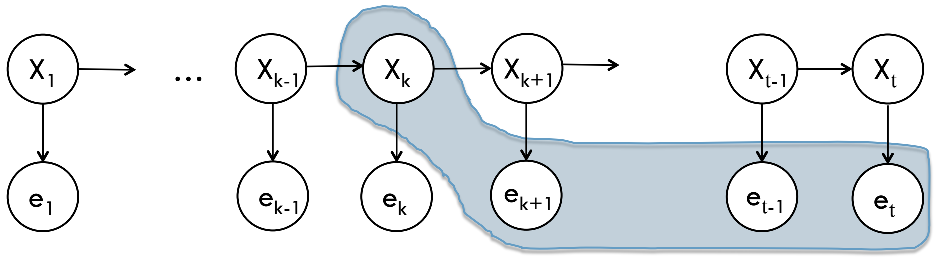
- Step Zero: Assume we have $p(X_{k+1} | e_{k+2:t})$

The Backward Algorithm



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

The Backward Algorithm



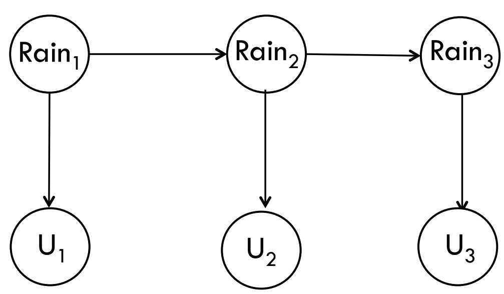
- Step Two: Update the state via $p(X_{k+1} | X_k)$

Smoothing 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