Hidden Markov Models

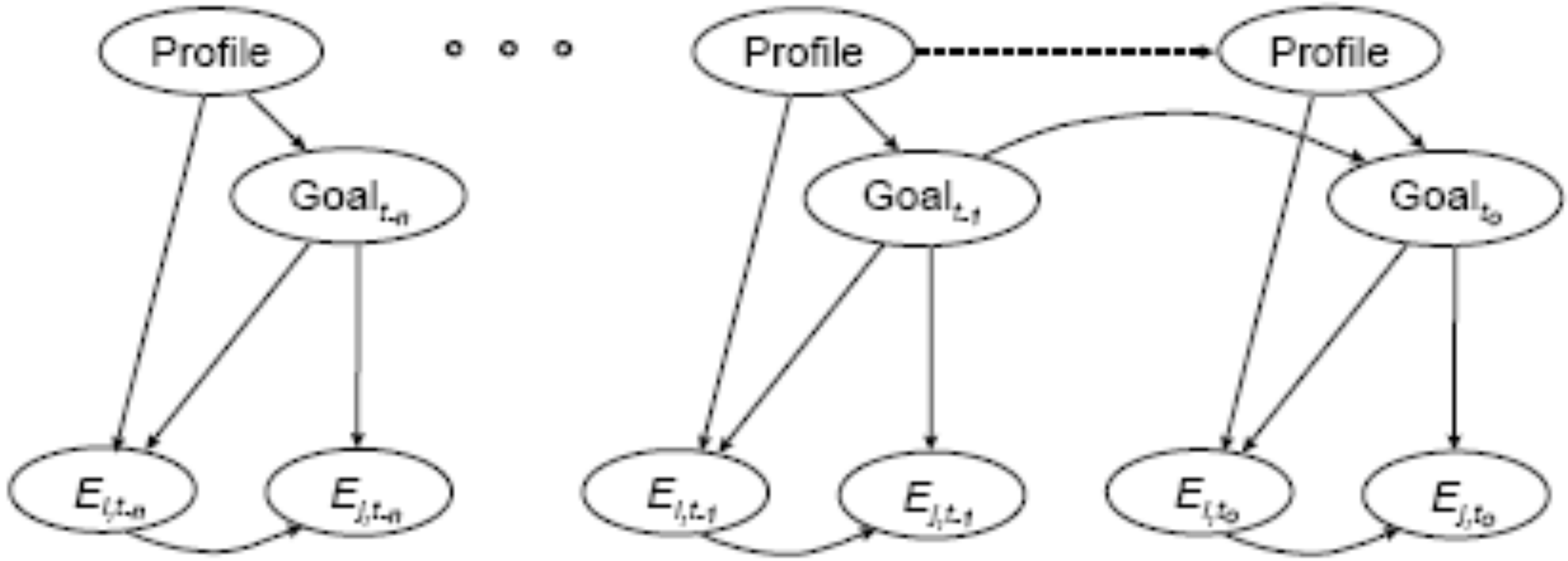
So far…

- Bayesian networks encode independence assumptions
- Inference in Bayes Nets is inherently difficult, unless the network has special structure
- Variable Elimination and Belief Propagation save time by:
  - Exploiting independence
  - Storing intermediate results
  - Eliminating irrelevant variables

Approximate Inference by Stochastic Simulation

- Basic Idea:
  - Draw N samples from a sampling distribution S
  - Compute an approximate posterior (conditional) probability P
  - Show this converges to the true probability P
- Four techniques
  - Direct Sampling
  - Rejection Sampling
  - Likelihood weighting
  - Markov chain Monte Carlo (MCMC)

Lumiere Project - Representing time…

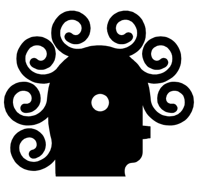


Temporal Reasoning Problem

- Static Reasoning – diagnosis of a car
- Temporal Reasoning –
  – Treating a diabetic patient
  – Speech Recognition

Question: Where is Tracy spending the night?

- Tracy has a new boyfriend, and she has been known to spend the night at his place
- When she spends the night at his place, I often observe that her hair is messy (he doesn't have a blow dryer) but sometimes she oversleeps when she is at home and her hair is mess anyway
- How can we model this as a graphical model (Bayes net)?



Question: Where is Tracy spending the night?

- Suppose I also know that where Tracy spent the night last night affects where she will spend the night tonight.
  - How does this change the model?
  - What problems start to arise?

States and Evidence

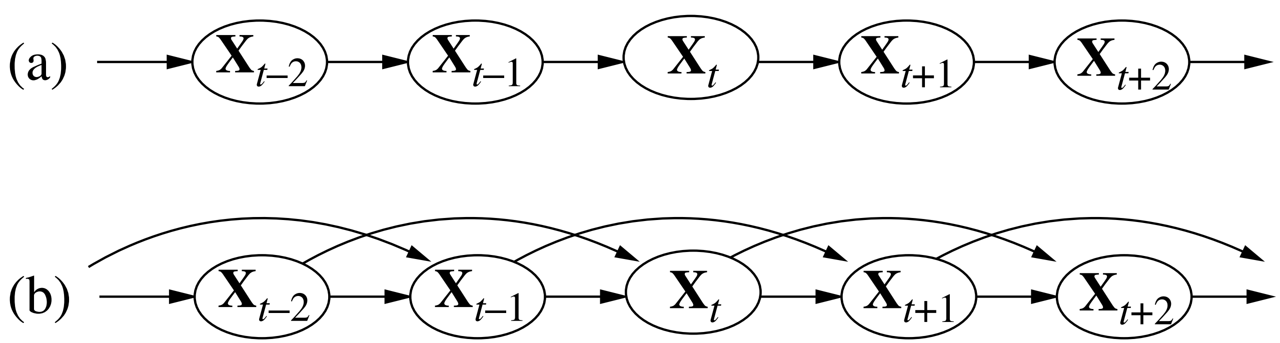
- Reason about a variable X_t given the history of the variable (at times 0:t-1)
  - e.g. Where did Tracy spend the night last (*time 3*) night given she spent Friday (*time 0*) at home and Sat (*time 1*) and Sun (*time 2*) at her boyfriend's place?

Concept: Markov Chains

- In a Markov chain a variable X_t depends on a bounded subset of $X_{0:t-1}$
- First order Markov Process:

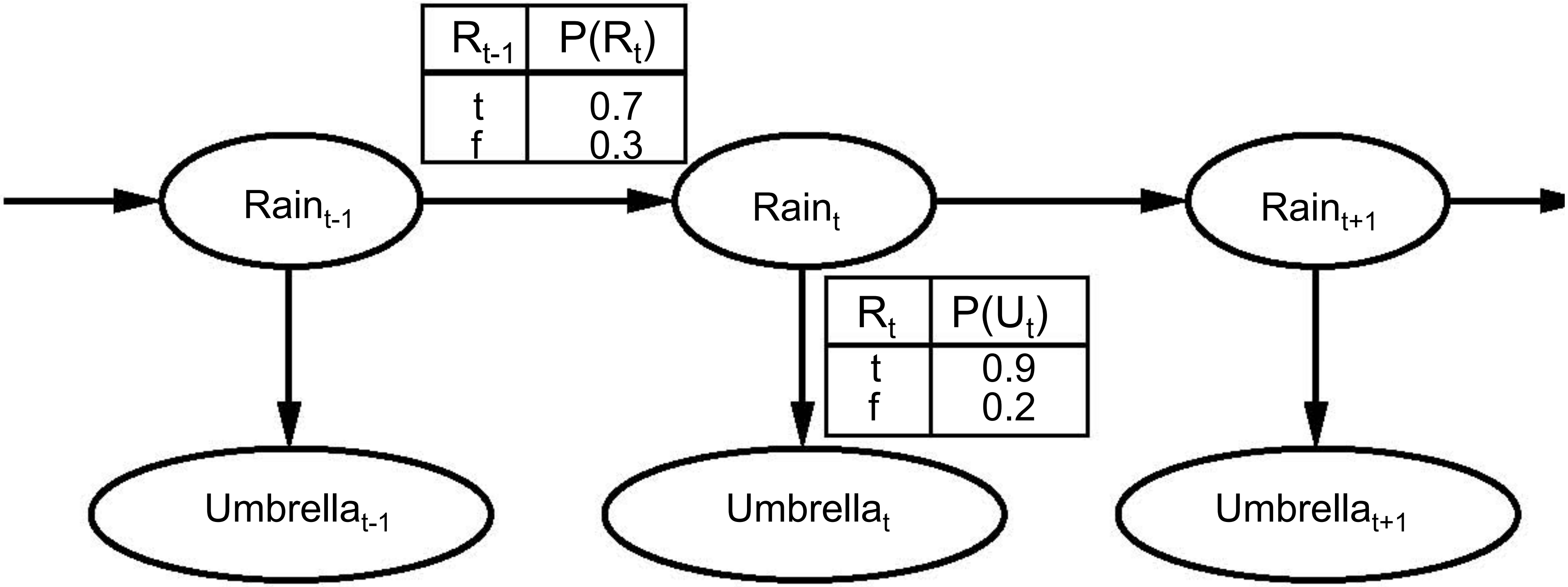
  $P(X_t|X_{0:t-1}) = P(X_t|X_{t-1})$
- Second order Markov Process:

  $P(X_t|X_{0:t-1}) = P(X_t|X_{t-1},X_{t-2})$



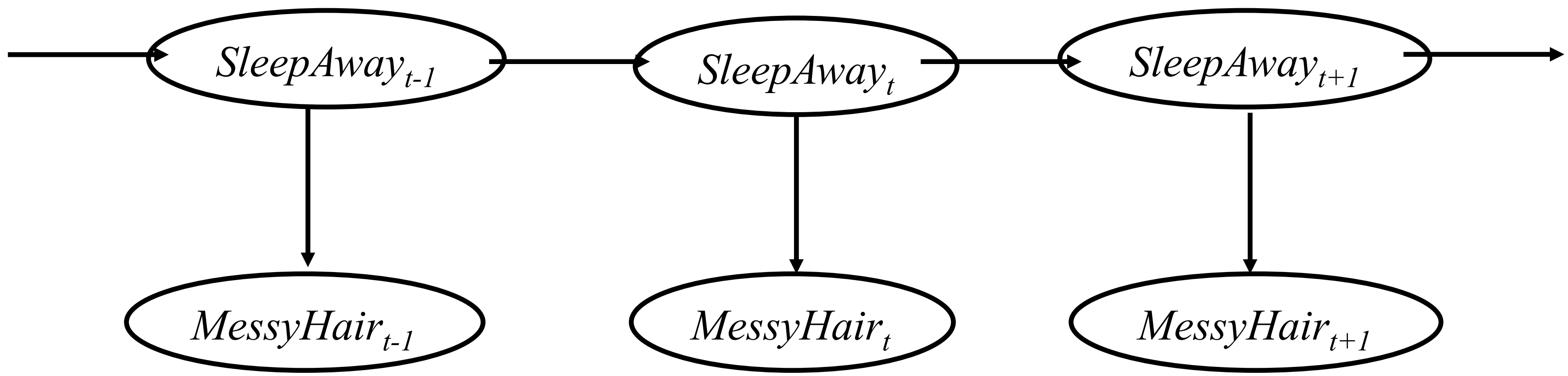
Markov chains: Adding Evidence

- Hidden Markov Models involve three things:
  - transition model: $P(X_t|X_{t-1})$
  - evidence (also called sensor or observation) model: $P(E_t|X_t)$
  - Prior probability: $P(X_0)$
- Sensor Markov assumption: $P(E_t|X_{0:t},E_{0:t-1})=P(E_t|X_t)$
  - e.g., if I know Tracy spent the night at home last night, then the state of her hair does not depend on what her hair looked like on Saturday



R_{t-1}	$P(R_t)$
t	0.7
f	0.3

R_t	$P(U_t)$
t	0.9
f	0.2



$SleepAway_{t-1}$
$SleepAway_t$
$SleepAway_{t+1}$
$MessyHair_{t-1}$
$MessyHair_t$
$MessyHair_{t+1}$

Modeling a dynamic world

- We need to track and predict signals that change
- Real-world example:
  - Speech Recognition
  - Sketch Recognition

- Issue: what is a "step"?

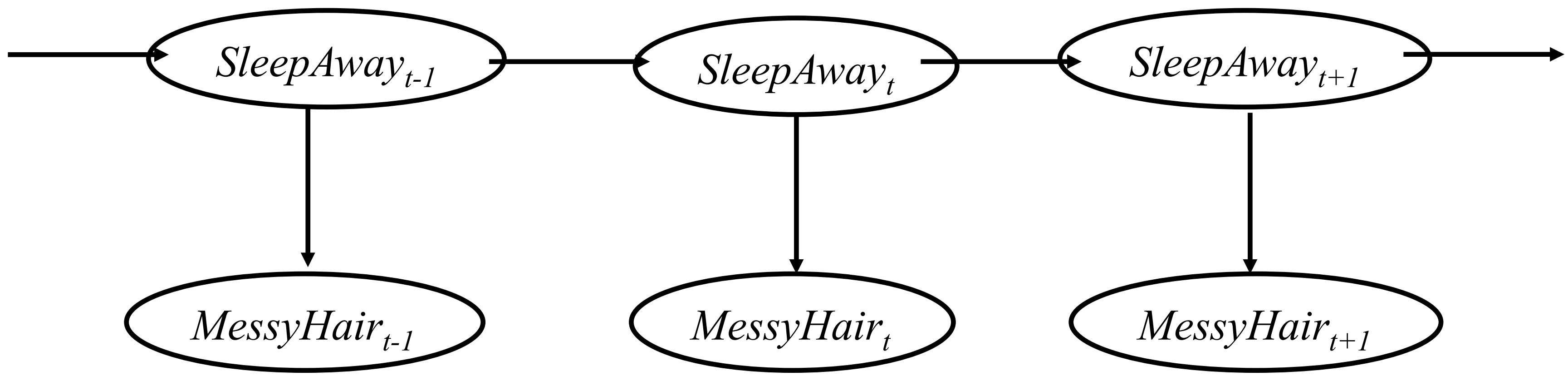
Inference Tasks

- What might we want to do with this model?

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 $0<=k<t$
  - "Revisionist history" (essential for learning)
- Most Likely Explanation (MLE): $\text{argmax}_{x1:t} P(x_{1:t}|e_{1:t})$
  - e.g., speech recognition, sketch recognition

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



- a recursive state estimation algorithm

$$P(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) = f(\mathbf{e}_{1:t+1}, P(\mathbf{X}_t \mid \mathbf{e}_{1:t}))$$

Filtering

$$P(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) = f(\mathbf{e}_{1:t+1}, P(\mathbf{X}_t \mid \mathbf{e}_{1:t}))$$

$$= P(\mathbf{X}_{t+1} \mid \mathbf{e}_{t+1}, \mathbf{e}_{1:t})$$

$$= \alpha P(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}, \mathbf{e}_{1:t}) P(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t})$$

$$= \alpha P(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) P(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t})$$

Filtering

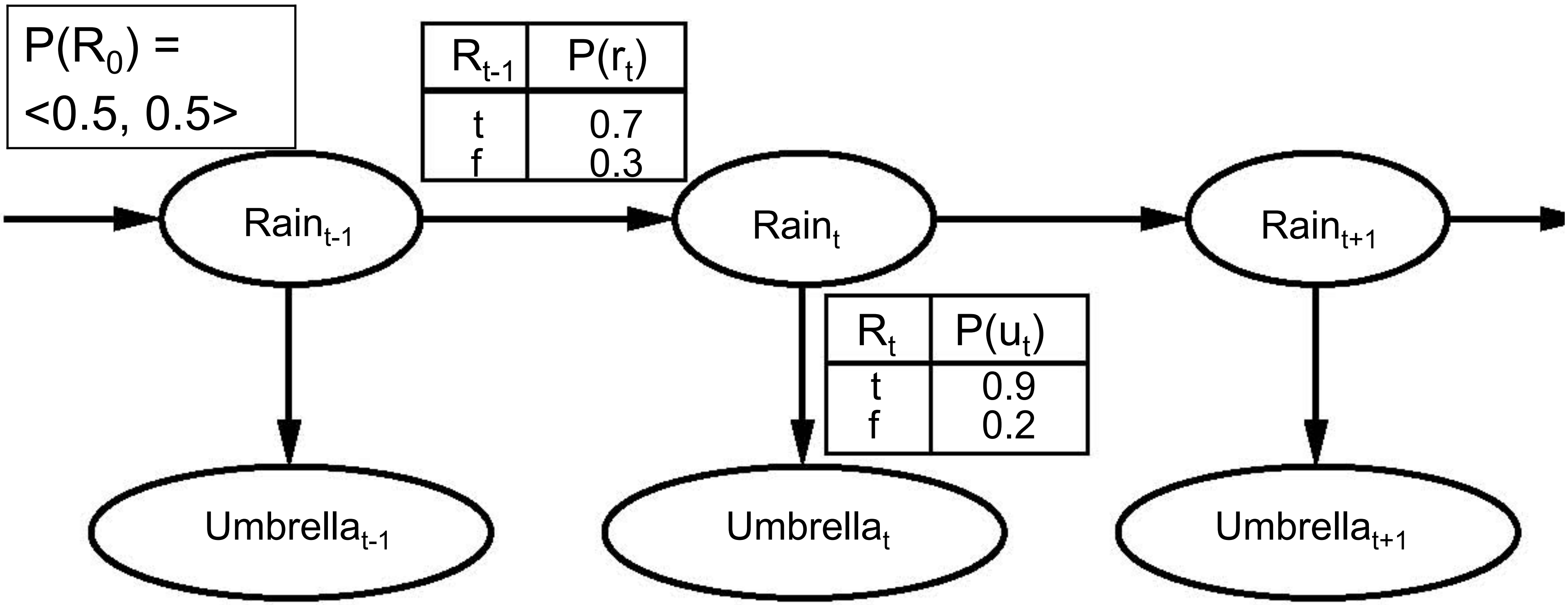
$$P(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha P(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) P(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t})$$

$$= \alpha P(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} P(\mathbf{X}_{t+1} \mid \mathbf{x}_t, \mathbf{e}_{1:t}) P(\mathbf{x}_t \mid \mathbf{e}_{1:t})$$

$$= \alpha P(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} P(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{e}_{1:t})$$

In other words, take the old (time t) distribution, and update it with the new evidence at time t+1

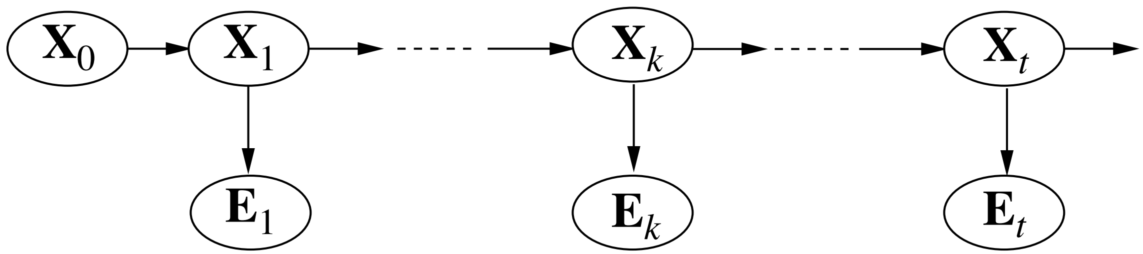
Filtering Example



Question: Prediction

- Given the equations for filtering, can you figure out how to do prediction?

Smoothing



Divide evidence $e_{1:t}$ into $e_{1:k}$, $e_{k+1:t}$

$$P(\mathbf{X}_k \mid \mathbf{e}_{1:t}) = P(\mathbf{X}_k \mid \mathbf{e}_{k+1:t}, \mathbf{e}_{1:k})$$

$$= \alpha P(\mathbf{e}_{k+1:t} \mid \mathbf{X}_k, \mathbf{e}_{1:k}) P(\mathbf{X}_k \mid \mathbf{e}_{1:k})$$

$$= \alpha P(\mathbf{X}_k \mid \mathbf{e}_{1:k}) P(\mathbf{e}_{k+1:t} \mid \mathbf{X}_k)$$

$$= \alpha \, \mathrm{f}_{1:k} \mathrm{b}_{k+1:t}$$

This is called the forward-backward algorithm

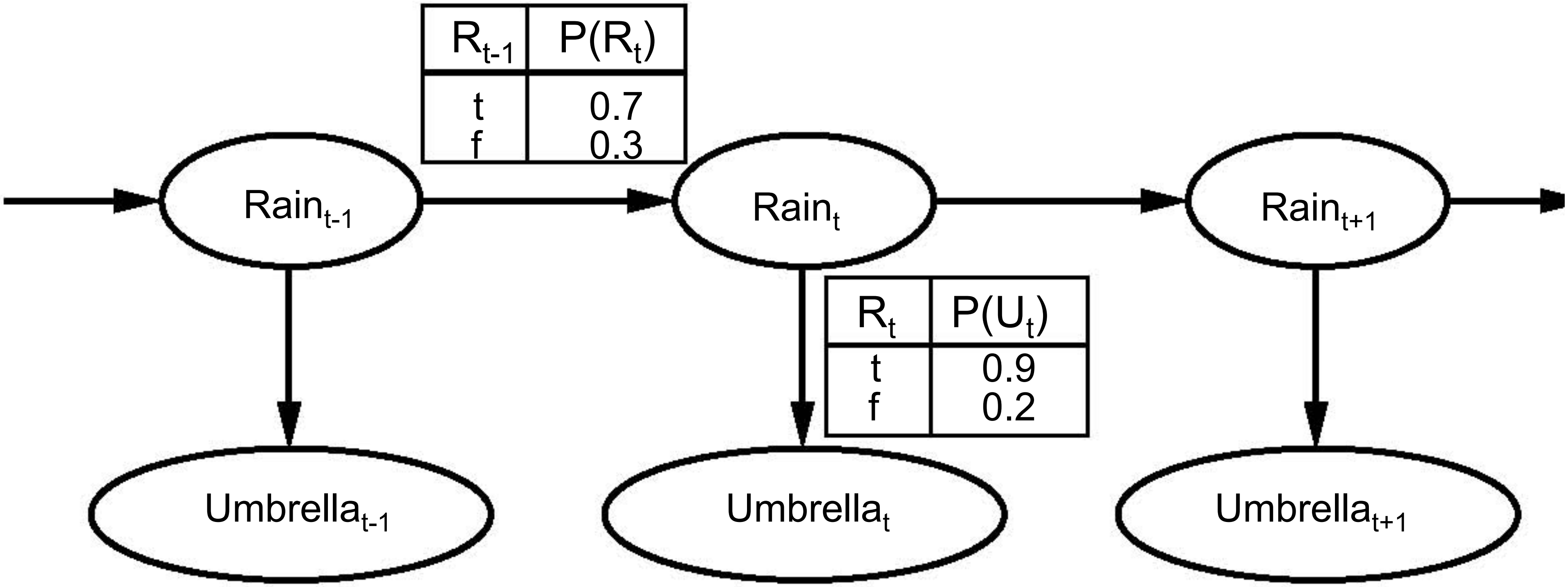
Backward message computed by a backwards recursion

$$P(\mathbf{e}_{k+1:t} \mid X_k) = \sum_{x_{k+1}} P(\mathbf{e}_{k+1:t} \mid X_k, x_{k+1}) P(x_{k+1} \mid X_k)$$

$$= \sum_{x_{k+1}} P(\mathbf{e}_{k+1:t} \mid x_{k+1}) P(x_{k+1} \mid X_k)$$

$$= \sum_{x_{k+1}} P(e_{k+1} \mid x_{k+1}) P(\mathbf{e}_{k+2:t} \mid x_{k+1}) P(x_{k+1} \mid X_k)$$

Smoothing Example



R_{t-1}	$P(R_t)$
t	0.7
f	0.3

R_t	$P(U_t)$
t	0.9
f	0.2

Smoothing Example

$P(u_t \mid r_t)$	$P(u_t \mid \sim r_t)$
0.9	0.2

$P(r_0)$	$P(r_1 \mid u_1)$	$P(r_2 \mid u_1,u_2)$	$P(r_1 \mid u_1,u_2)$
0.5	0.818	0.883	

$P(r_{t+1} \mid r_t)$	$P(r_{t+1} \mid \sim r_t)$
0.7	0.3

$$P(r_1 \mid u_1, u_2) = \alpha P(r_1 \mid u_1) P(u_2 \mid r_1)$$

$$P(u_2 \mid r_1) = \sum_{r_2} P(u_2 \mid r_2) P(u_3 \mid r_2) P(r_2 \mid r_1)$$

$$= P\left(u_2 \mid r_2\right) P\left(r_2 \mid r_1\right) + P\left(u_2 \mid \neg r_2\right) P\left(\neg r_2 \mid r_1\right)$$

$$P(\neg r_1 \mid u_1, u_2) = \alpha P(\neg r_1 \mid u_1) P(u_2 \mid \neg r_1)$$

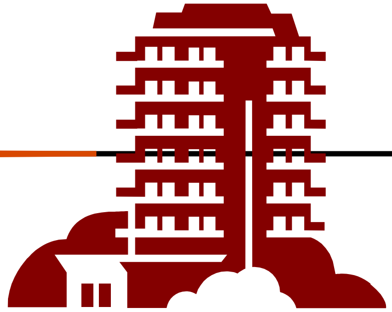
$$P(u_2 \mid \neg r_1) = \sum_{r_2} P(u_2 \mid r_2) P(u_3 \mid r_2) P(r_2 \mid \neg r_1)$$

$$= P\left(u_2 \mid r_2\right) P\left(r_2 \mid \neg r_1\right) + P\left(u_2 \mid \neg r_2\right) P\left(\neg r_2 \mid \neg r_1\right)$$

Hidden Markov Models

- simplest dynamic Bayesian Network
- state is represented by a single "megavariable"
- evidence is represented by a single evidence variable
- applications in speech recognition, handwriting recognition, gesture recognition, musical score following, and bioinformatics

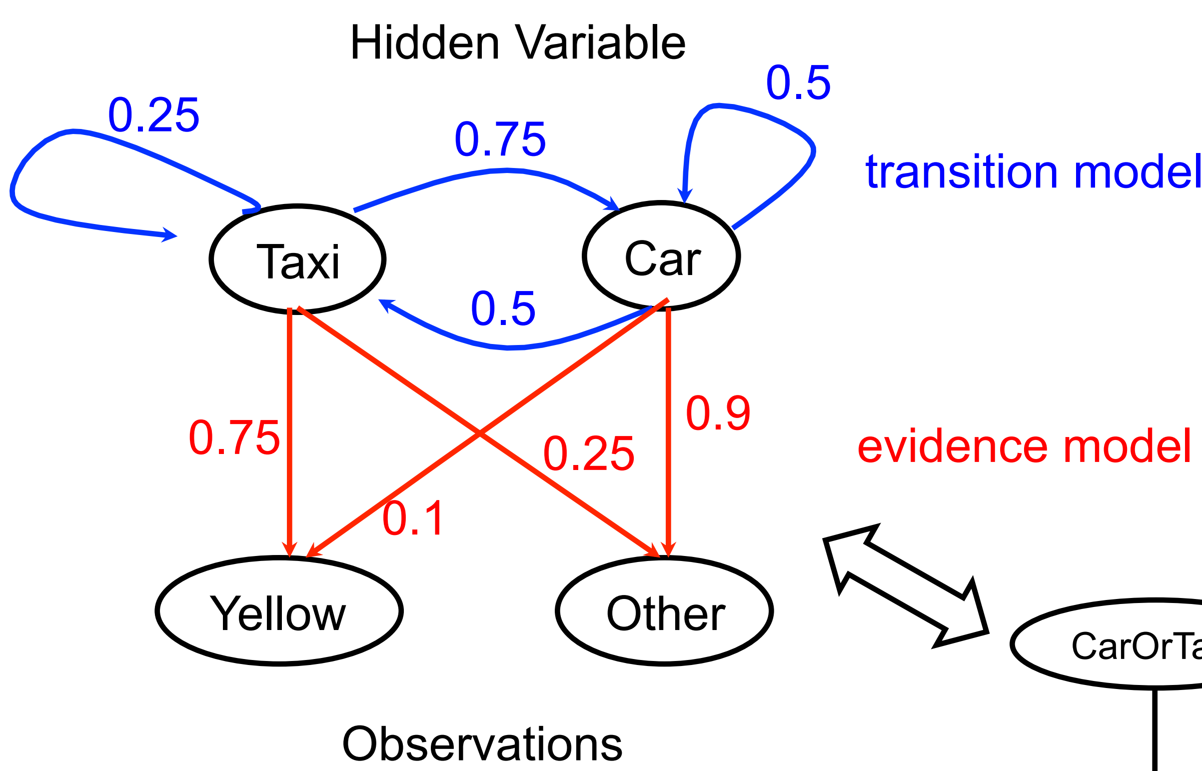
Your turn

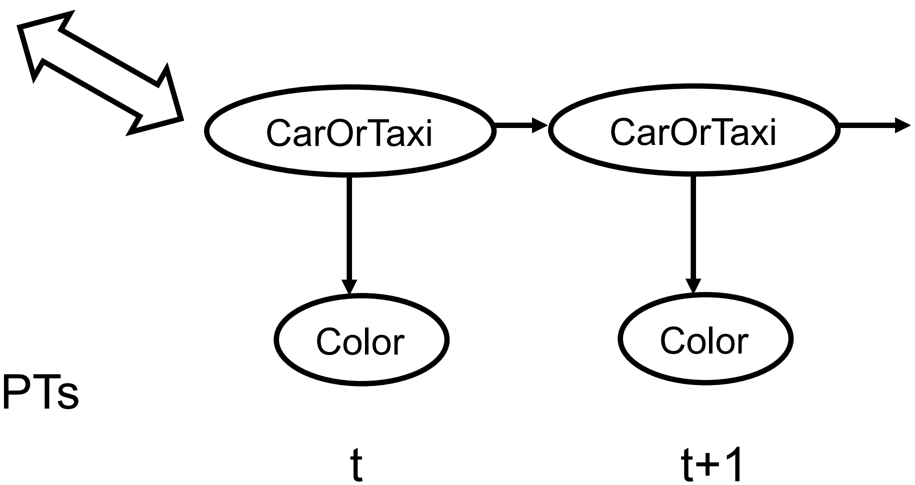


- Problem: Identify the taxi

Your turn

- Problem: Identify the taxi





Translate the diagram above into two CPTs

Your Turn

- You just saw a red car, and now you see a yellow car. What is the probability that the car you see is a taxi? (Assume taxis and cars are equally likely when you start looking)
- What is the probability that the next car will be a taxi?

Filtering Equation

$$P(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1})$$

$$= \alpha P(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} P(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{e}_{1:t})$$

Prediction Equation

$$P(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) = \sum_{\mathbf{x}_t} P(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{e}_{1:t})$$

Your Turn

- You decide not to go down, and you observe that the next car is also yellow. Use this new information to update your belief that the last car you saw was a taxi

Smoothing Equation

$$P(\mathbf{X}_k \mid \mathbf{e}_{1:t}) = \alpha P(\mathbf{X}_k \mid \mathbf{e}_{1:k}) P(\mathbf{e}_{k+1:t} \mid \mathbf{X}_k)$$

$$P(\mathbf{e}_{k+1:t} \mid X_k) = \sum_{x_{k+1}} P(e_{k+1} \mid x_{k+1}) P(\mathbf{e}_{k+2:t} \mid x_{k+1}) P(x_{k+1} \mid X_k)$$

Viterbi - Most Likely Explanation