Intelligent Systems (AI-2)

Computer Science cpsc422, Lecture 14

Oct, 4, 2019

Slide credit: some slides adapted from Stuart Russell (Berkeley)

422 big picture

StarAI (statistical relational AI)

Hybrid: Det +Sto

Prob CFG
Prob Relational Models
Markov Logics

Deterministic | **Stochastic**

Query

Logics
First Order Logics
Ontologies

- Full Resolution
- SAT

Belief Nets

Approx. : Gibbs

Markov Chains and HMMs

Forward, Viterbi….
Approx. : Particle Filtering

Undirected Graphical Models
Markov Networks
Conditional Random Fields

Planning

Markov Decision Processes and Partially Observable MDP

- Value Iteration
- Approx. Inference

Reinforcement Learning

Applications of AI

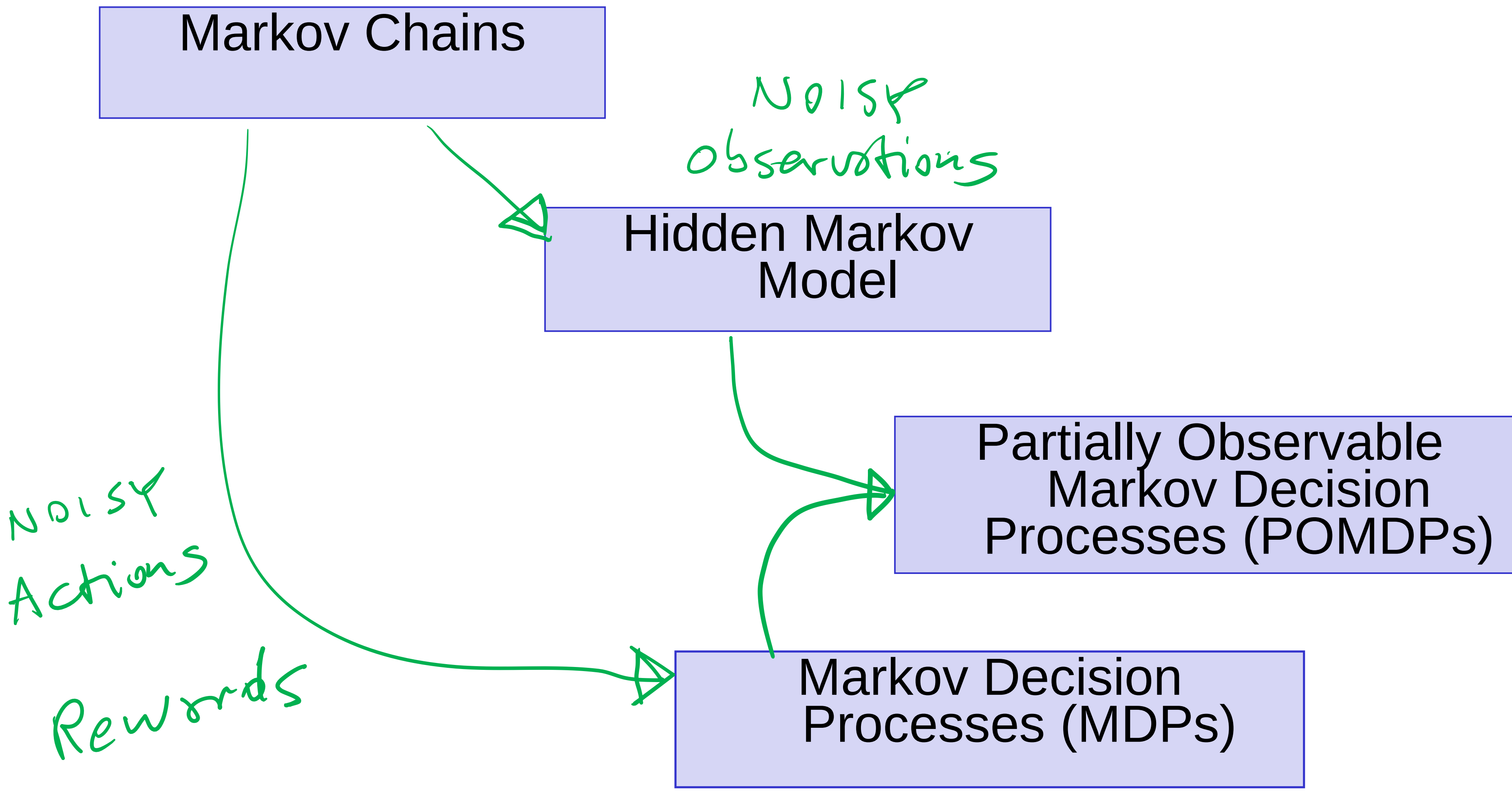
Representation

Reasoning Technique

Lecture Overview (Temporal Inference)

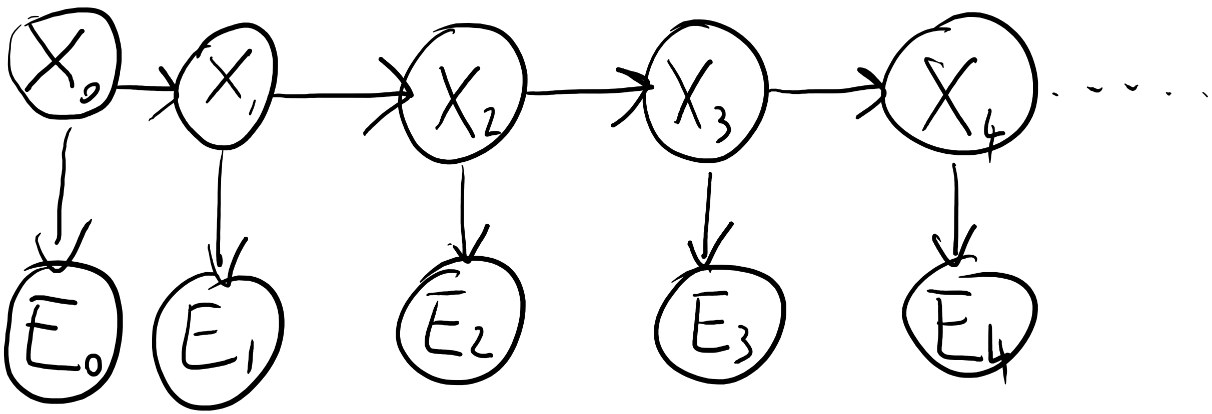
- **Filtering** (posterior distribution over the current state given evidence to date)
  - **From intuitive explanation to formal derivation**
  - **Example**
- **Prediction (**posterior distribution over a future state given evidence to date)
- **(start) Smoothing** (posterior distribution over a *past* state given all evidence to date)

Markov Models



Hidden Markov Model

- A Hidden Markov Model (HMM) starts with a Markov chain, and adds a noisy observation/evidence about the state at each time step:



- |domain(X)| = k
- |domain(E)| = h

- $P(X_0)$ specifies initial conditions K
- $P(X_{t+1}|X_t)$ specifies the dynamics K×K
- $P(E_t|S_t)$ specifies the sensor model K×h { K prob. dist. over O }

Simple Example

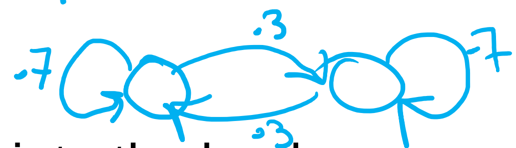
(We'll use this as a running example)

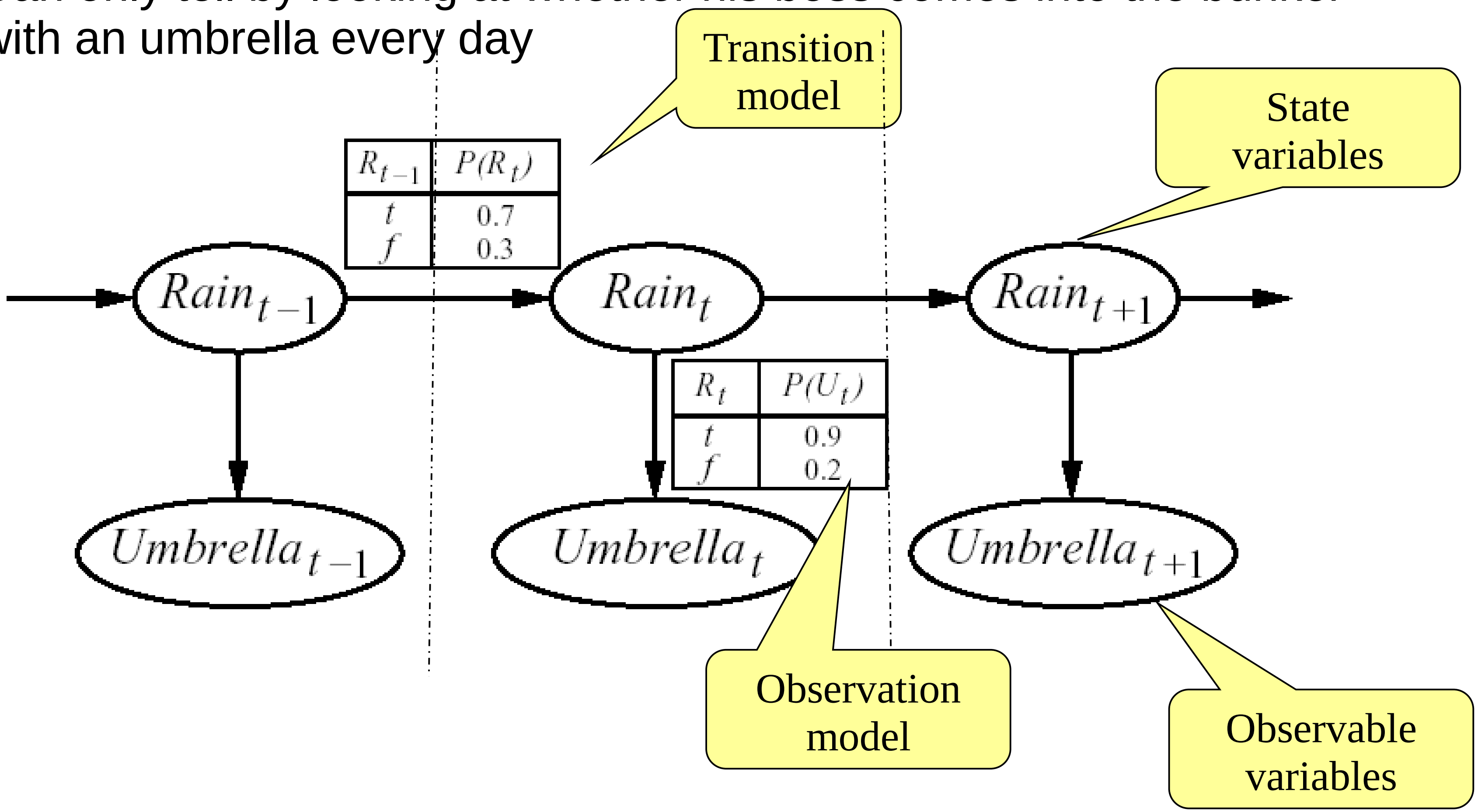
➢ Guard stuck in a high-security bunker

➢ Would like to know if it is raining outside

➢ Can only tell by looking at whether his boss comes into the bunker with an umbrella every day

		R_t	
		T	F
R_{t-1}	T	.7	.3
	F	.3	.7

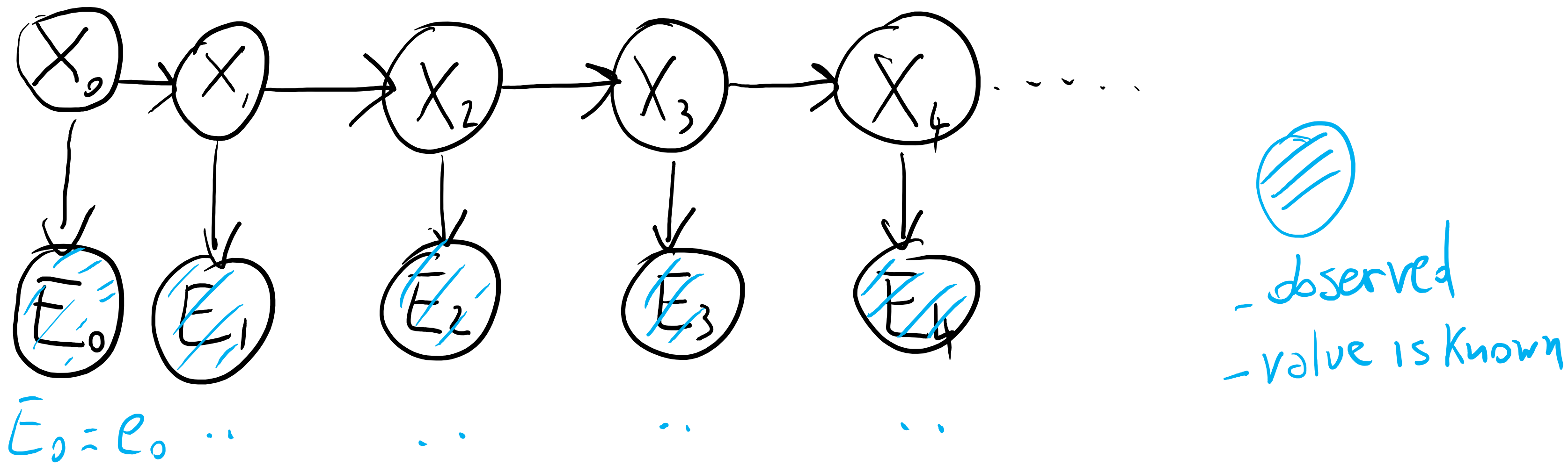




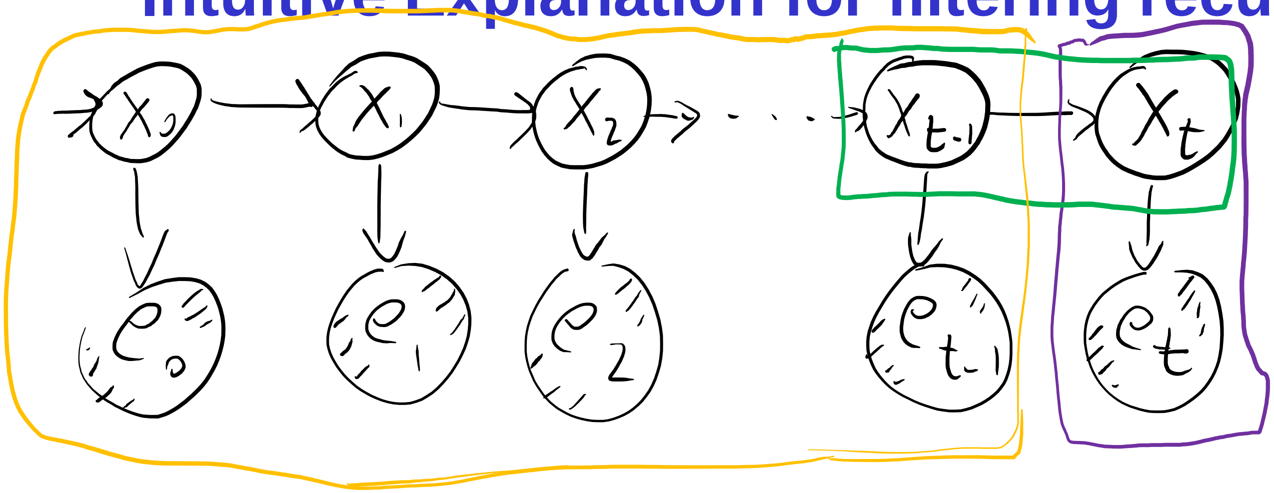
Useful inference in HMMs

- **In general (Filtering):** compute the posterior distribution over the current state given all evidence to date

$$P(X_t \mid e_{0:t})$$



Intuitive Explanation for filtering recursive formula



sequence of evidences $e_0:e_t$

$$P(X_t \mid e_{0:t}) = \alpha\, P(e_t \mid X_t) * \sum_{x_{t-1}} P(X_t \mid x_{t-1}) * P(x_{t-1} \mid e_0:e_{t-1})$$

X_t generated evidence e_t

whatever x_{t-1} was, X_t was reached from there

and evidence $e_0:e_{t-1}$ must have been generated before getting to x_{t-1}

Filtering

- Idea: recursive approach
  - Compute filtering up to time t-1, and then include the evidence for time t (***recursive estimation***)

- $P(X_t \mid e_{0:t}) = P(X_t \mid e_{0:t-1}, e_t)$ dividing up the evidence

$= \alpha\, P(e_t \mid X_t, e_{0:t-1})\, P(X_t \mid e_{0:t-1})$ **WHY?**

$= \alpha\, P(e_t \mid X_t)\, P(X_t \mid e_{0:t-1})$ **WHY?**

A. Bayes Rule

B. Cond. Independence

C. Product Rule

Inclusion of new evidence: **this is available from..**

One step prediction of current state given evidence up to *t-1*

- So we only need to compute $P(X_t \mid e_{0:t-1})$

$$P(X,Y,Z) = P(X \mid Y,Z)\,P(Y,Z)$$

$$P(X,Y,Z) = P(Y \mid X,Z)\,P(X,Z)$$

$$= P(Y \mid Z)\,P(Z)$$

$$P(Y \mid X,Z) = \frac{P(X \mid Y,Z)\,P(Y,Z)}{P(X,Z)}$$

$$= P(X \mid Z)\,P(Z)$$

$$= \frac{P(X \mid Y,Z)\,P(Y \mid Z)}{P(X \mid Z)}$$

$$= \alpha\,P(X \mid Y,Z)\,P(Y \mid Z)$$

$$\sum_y P(x\,y \mid z)$$

$$\Downarrow$$

$$\sum_y P(x \mid y\,z)\,P(y \mid z)$$

"moving" the conditioning

$$\boxed{P(AB|C)} = \frac{P(ABC)}{P(C)} * \frac{P(BC)}{P(BC)} =$$

$$= \frac{P(ABC)}{P(BC)} * \frac{P(BC)}{P(C)} =$$

$$= \boxed{P(A|BC) * P(B|C)}$$

Filtering

Prove it?

- Compute $\boldsymbol{P(X_t \mid e_{0:t-1})}$

$$P(X_t \mid e_{0:t-1}) = \sum_{x_{t-1}} P(X_t, x_{t-1} \mid e_{0:t-1}) = \sum_{x_{t-1}} P(X_t \mid x_{t-1}, e_{0:t-1})\, P(x_{t-1} \mid e_{0:t-1}) =$$

$$= \sum_{x_{t-1}} P(X_t \mid x_{t-1})\, P(x_{t-1} \mid e_{0:t-1})$$ because of..

Transition model!

Filtering at time *t-1*

- Putting it all together, we have the desired recursive formulation

$$P(X_t \mid e_{0:t}) = \alpha\, P(e_t \mid X_t) \sum_{x_{t-1}} P(X_t \mid x_{t-1})\, P(x_{t-1} \mid e_{0:t-1})$$

Inclusion of new evidence (sensor model)

Propagation to time *t*

Filtering at time *t-1*

- $P(X_{t-1} \mid e_{0:t-1})$ can be seen as a message $f_{0:t-1}$ that is propagated forward along the sequence, modified by each transition and updated by each observation

Filtering

- Thus, the recursive definition of filtering at time *t* in terms of filtering at time *t-1* can be expressed as a FORWARD procedure
  - $\boldsymbol{f}_{0:t} = \alpha \, FORWARD\,(\boldsymbol{f}_{0:t-1}, \boldsymbol{e}_t)$
- which implements the update described in

$$\boldsymbol{P}(\boldsymbol{X}_t \mid \boldsymbol{e}_{0:t}) = \alpha\, \boldsymbol{P}(\boldsymbol{e}_t \mid \boldsymbol{X}_t) \sum_{x_{t-1}} \boldsymbol{P}(\boldsymbol{X}_t \mid \boldsymbol{x}_{t-1})\, P(\boldsymbol{x}_{t-1} \mid \boldsymbol{e}_{0:t-1})$$

Filtering at time *t-1*

Inclusion of new evidence (sensor model)

Propagation to time *t*

Analysis of Filtering

- Because of the recursive definition in terms for the forward message, when all variables are discrete the time for each update is constant (i.e. independent of t)

- The constant depends of course on the size of the state space

Rain Example

- Suppose our security guard came with a prior belief of 0.5 that it rained on day 0, just before the observation sequence started.

- Without loss of generality, this can be modelled with a fictitious state R_0 with no associated observation and $P(R_0) = \langle 0.5, 0.5 \rangle$

- **Day 1**: umbrella appears (u_1). Thus

no previous evidence

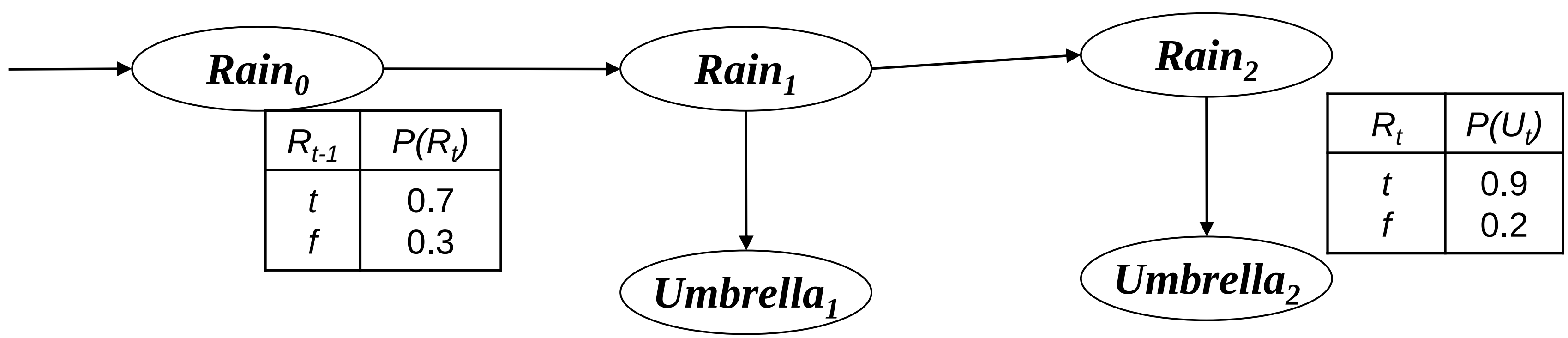
R_{t-1} \ R_t	T	F
T	.7	.3
F	.3	.7

$$\mathbf{P}(R_1 \mid \mathbf{e}_{0:t-1}) = \mathbf{P}(R_1) = \sum_{r_0} \mathbf{P}(R_1 \mid r_0)\, P(r_0)$$

$$= \langle \underset{T}{0.7}, \underset{F}{0.3} \rangle * \underset{T}{0.5} + \langle \underset{T}{0.3}, \underset{F}{0.7} \rangle * \underset{F}{0.5} = \langle 0.5, 0.5 \rangle$$

0.5
0.5

TRUE 0.5
FALSE 0.5



Rain Example

- Updating this with evidence from for $t = 1$ (umbrella appeared) gives

  $\mathbf{P}(R_1 | u_1) = \alpha\, \mathbf{P}(u_1 | R_1)\, \mathbf{P}(R_1) =$

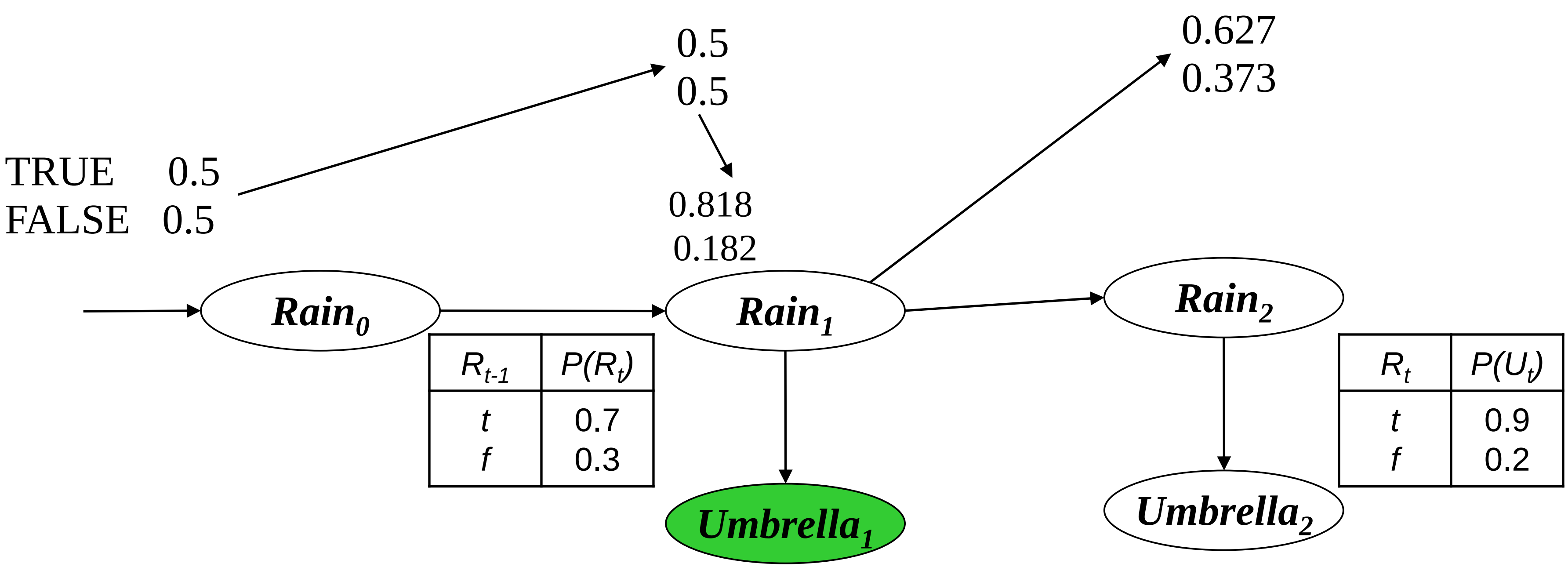
  $\alpha<0.9, 0.2><0.5,0.5> = \alpha<0.45, 0.1> \sim <0.818, 0.182>$

- Day 2: umbella appears (u_2). Thus

  $\mathbf{P}(R_2 | \mathbf{e}_{0:t-1}) = \mathbf{P}(R_2 | u_1) = \sum_{r_1} \mathbf{P}(R_2 | r_1)\, P(r_1 | u_1) =$

  $= <0.7, 0.3> * 0.818 + <0.3,0.7> * 0.182 \sim <0.627,0.373>$

R_{t-1} \ R_t	T	F
T	.7	.3
F	.3	.7



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

R_t	$P(U_t)$
t	0.9
f	0.2

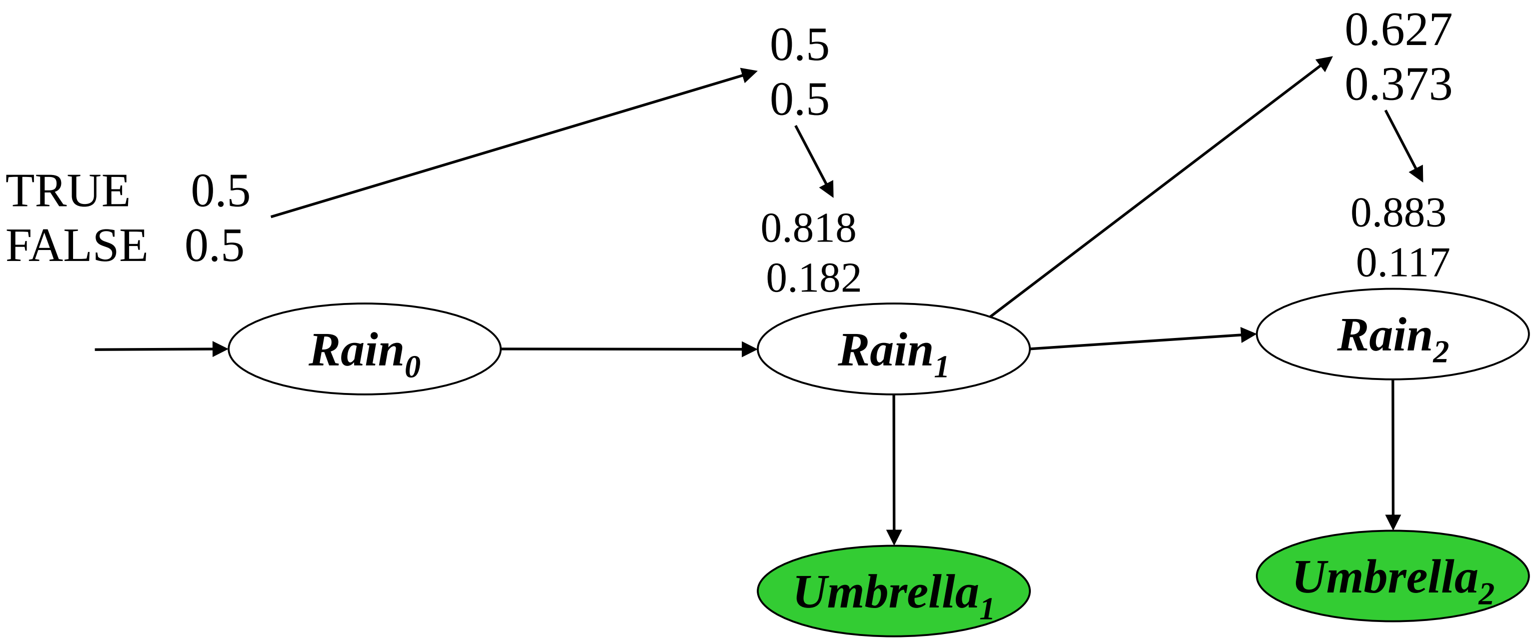
Rain Example

- Updating this with evidence from for $t = 2$ (umbrella appeared) gives

  $\mathbf{P}(R_2 | u_1, u_2) = \alpha\, \mathbf{P}(u_2 | R_2)\, \mathbf{P}(R_2 | u_1) =$

  $\alpha\langle 0.9, 0.2\rangle\langle 0.627, 0.373\rangle = \alpha\langle 0.565, 0.075\rangle \sim \langle 0.883, 0.117\rangle$

- Intuitively, the probability of rain increases, because the umbrella appears twice in a row



Practice exercise (home)

Compute filtering at t_3 if the 3rd observation/evidence is *no umbrella* (will put solution on inked slides)

$\langle 0.7, 0.3 \rangle * 0.883 + \langle 0.3, 0.7 \rangle * 0.117$

$\langle 0.618, 0.264 \rangle + \langle 0.035, 0.081 \rangle = \langle 0.653, 0.345 \rangle$

sensor model

$\alpha \langle 0.653, 0.345 \rangle * \langle 0.1, 0.8 \rangle$

$\alpha \langle 0.065, 0.276 \rangle$

normalize / divide by the sum.

0.19 0.81

Lecture Overview

- **Filtering** (posterior distribution over the current state given evidence to date)
  - **From intuitive explanation to formal derivation**
  - **Example**
- **Prediction (**posterior distribution over a future state given evidence to date)
- **(start) Smoothing** (posterior distribution over a *past* state given all evidence to date)

Prediction $P(X_{t+k+1} \mid e_{0:t})$

- Can be seen as filtering without addition of new evidence
- In fact, filtering already contains a one-step prediction

$$\mathbf{P}(\mathbf{X}_t \mid \mathbf{e}_{0:t}) = \alpha\, \mathbf{P}(\mathbf{e}_t \mid \mathbf{X}_t) \sum_{x_{t-1}} \mathbf{P}(\mathbf{X}_t \mid \mathbf{x}_{t-1})\, P(\mathbf{x}_{t-1} \mid \mathbf{e}_{0:t-1})$$

Inclusion of new evidence (sensor model)

Propagation to time t

Filtering at time $t-1$

- We need to show how to recursively predict the state at time $t+k+1$ from a prediction for state $t+k$

$$\mathbf{P}(\mathbf{X}_{t+k+1} \mid \mathbf{e}_{0:t}) = \sum_{x_{t+k}} \mathbf{P}(\mathbf{X}_{t+k+1}, \mathbf{x}_{t+k} \mid \mathbf{e}_{0:t}) = \sum_{x_{t+k}} \mathbf{P}(\mathbf{X}_{t+k+1} \mid \mathbf{x}_{t+k}, \mathbf{e}_{0:t})\, P(\mathbf{x}_{t+k} \mid \mathbf{e}_{0:t}) =$$

$$= \sum_{x_{t+k}} \mathbf{P}(\mathbf{X}_{t+k+1} \mid \mathbf{x}_{t+k})\, P(\mathbf{x}_{t+k} \mid \mathbf{e}_{0:t})$$

Prediction for state $t+k$

Transition model

- Let‘s continue with the rain example and compute the probability of *Rain* on day four after having seen the umbrella in day one and two: $\mathbf{P}(R_4 \mid u_1, u_2)$

Rain Example

- Prediction from day 2 to day 3

$$\mathbf{P}(\mathbf{X}_3 \mid \mathbf{e}_{1:2}) = \sum_{x_2} \mathbf{P}(\mathbf{X}_3 \mid \mathbf{x}_2)\, P(\mathbf{x}_2 \mid \mathbf{e}_{1:2}) = \sum_{r_2} \mathbf{P}(R_3 \mid r_2)\, P(r_2 \mid u_1 u_2) =$$

$$= \langle 0.7, 0.3\rangle * 0.883 + \langle 0.3, 0.7\rangle * 0.117 = \langle 0.618, 0.265\rangle + \langle 0.035, 0.082\rangle$$

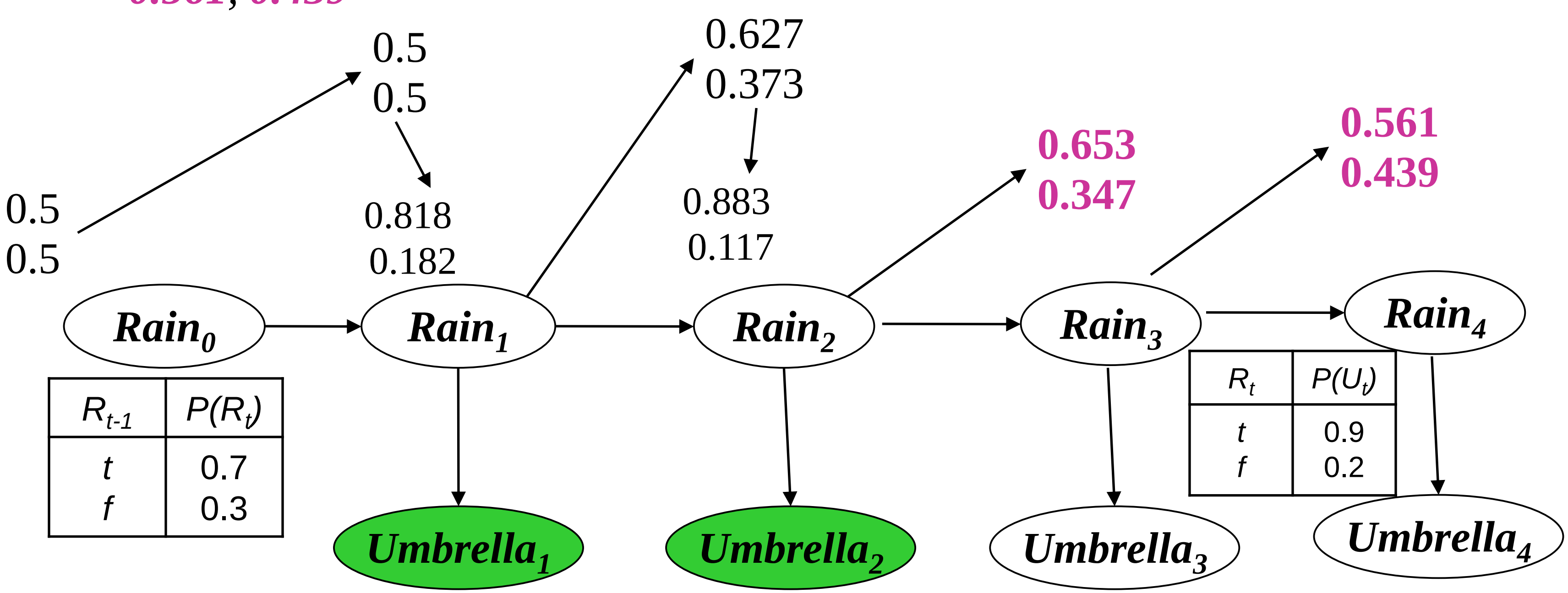
$$= \langle 0.653, 0.347\rangle$$

- Prediction from day 3 to day 4

$$\mathbf{P}(\mathbf{X}_4 \mid \mathbf{e}_{1:2}) = \sum_{x_3} \mathbf{P}(\mathbf{X}_4 \mid \mathbf{x}_3)\, P(\mathbf{x}_3 \mid \mathbf{e}_{1:2}) = \sum_{r_3} \mathbf{P}(R_4 \mid r_3)\, P(r_3 \mid u_1 u_2) =$$

$$= \langle 0.7, 0.3\rangle * \mathbf{0.653} + \langle 0.3, 0.7\rangle * \mathbf{0.347} = \langle \mathbf{0.457}, \mathbf{0.196}\rangle + \langle \mathbf{0.104}, \mathbf{0.243}\rangle$$

$$= \langle \mathbf{0.561}, \mathbf{0.439}\rangle$$



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

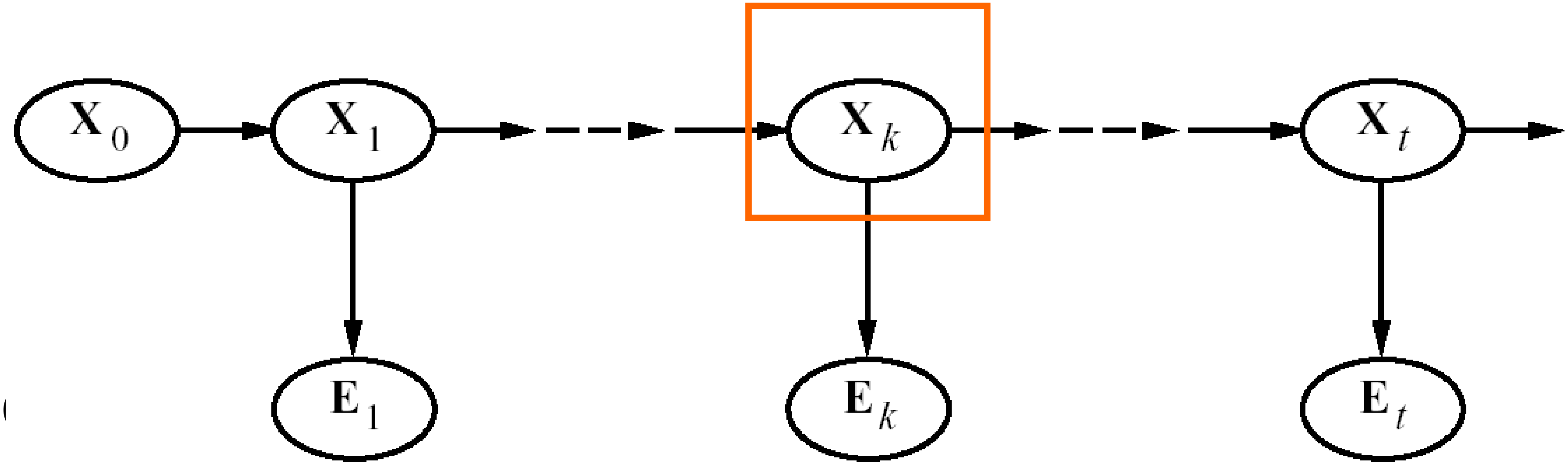
R_t	$P(U_t)$
t	0.9
f	0.2

Lecture Overview

- **Filtering** (posterior distribution over the current state given evidence to date)
  - **From intuitive explanation to formal derivation**
  - **Example**
- **Prediction** (posterior distribution over a future state given evidence to date)
- **(start) Smoothing** (posterior distribution over a *past* state given all evidence to date)

Smoothing

- ***Smoothing***: Compute the posterior distribution over a *past* state given all evidence to date
  - $P(\mathbf{X}_k \mid \mathbf{e}_{0:t})$ for $1 \leq k < t$



Smoothing

- $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{0:t}) = \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{0:k}, \boldsymbol{e}_{k+1:t})$ dividing up the evidence

$= \alpha\, \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{0:k})\, \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k, \boldsymbol{e}_{0:k})$ using…

$= \alpha\, \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{0:k})\, \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$ using…

forward message from filtering up to state k, $\boldsymbol{f}_{0:k}$

backward message, $\boldsymbol{b}_{k+1:t}$ computed by a recursive process that runs backwards from *t*

Smoothing

- $\boldsymbol{P(X_k \mid e_{0:t}) = P(X_k \mid e_{0:k}, e_{k+1:t})}$ dividing up the evidence

  $= \alpha\, \boldsymbol{P(X_k \mid e_{0:k})\, P(e_{k+1:t} \mid X_k, e_{0:k})}$ using Bayes Rule

  $= \alpha\, \boldsymbol{P(X_k \mid e_{0:k})\, P(e_{k+1:t} \mid X_k)}$ By Markov assumption on evidence

forward message from filtering up to state k, $\boldsymbol{f}_{0:k}$

backward message, $\boldsymbol{b}_{k+1:t}$ computed by a recursive process that runs backwards from *t*

Learning Goals for today's class

- **You can:**

- Describe Filtering and derive it by manipulating probabilities
- Describe Prediction and derive it by manipulating probabilities
- Describe Smoothing and derive it by manipulating probabilities

TODO for Mon

- **Keep Reading Textbook Chp 8.5**
- **Keep working on assignment-2 (due on Fri, Oct 18)**