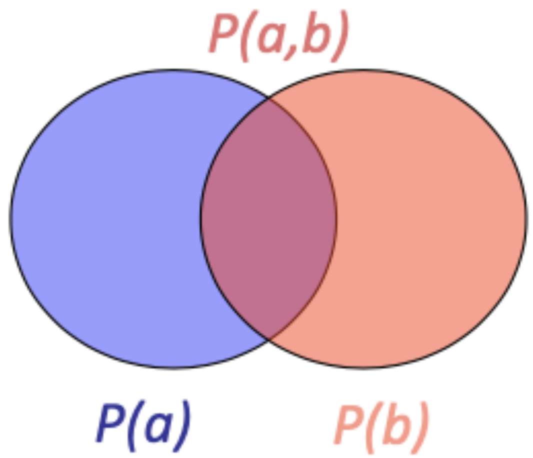
Recap: Probability

- Conditional probability
  - For any propositions a and b, the conditional probabilities is defined as

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)}$$

which holds whenever P(b) > 0



- Product rule

$$P(a \wedge b) = P(a \mid b)P(b).$$

Recap: Probability

- Marginalization rule
  - General **marginalization rule** for any sets of variables Y and Z:

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y}, \mathbf{Z} = \mathbf{z})$$

- Conditioning rule

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z})$$

Recap: Probability

- Bayes' Rule (Bayes' Law or Bayes' Theorem)
  - General case for multivalued variables can be written as follows:

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

  - General version conditionalized on some background evidence e:

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

Uncertainty

Uncertainty **Over Time**


週四
週五
週六
週日
週一
週二

Discrete-Time Models

The world is viewed as a series of snapshots or time slices.

- Each snapshot/time slice contains of a set of random variables
  - Some observable
  - Some unobservable

Example

- State variable X_t:
  - e.g., weather at time t

Probabilistic Reasoning

Probabilistic Reasoning over Time

Reasoning

Transition Model

- Given the previous state values $X_0, X_1, X_2, \ldots, X_{t-1}$,
  the transition model specifies the probability distribution over the latest state variables as

$$\mathbf{P}(X_t | X_{0:t-1})$$

  where $X_{0:t-1} = X_0, X_1, X_2, \ldots, X_{t-1}$
- Issue
  - $X_{0:t-1}$ is unbounded in size as t increases

Markov Assumption

- The current state depends on only a **finite fixed number** of previous state

Andrey Markov (1856-1922)

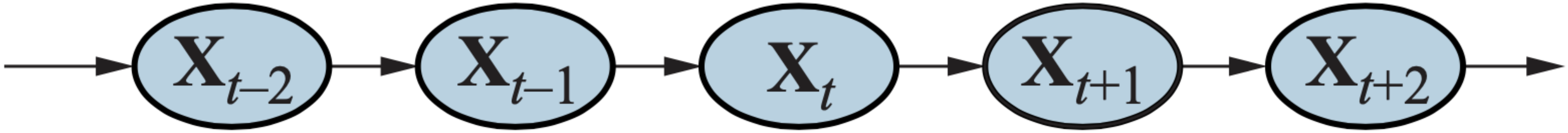
https://en.wikipedia.org/wiki/Andrey_Markov

Markov Chain (Markov Process)

- A sequence of random variables where the distribution of each variable follows the Markov assumption

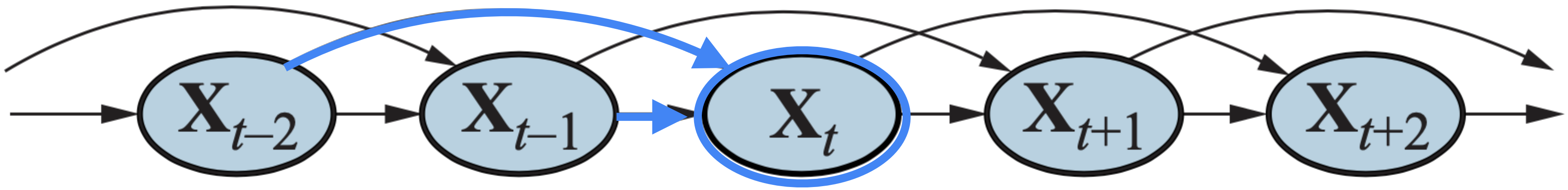
Example: Markov chain

- First-order Markov chain



$$\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{0:t-1}) = \mathbf{P}(\mathbf{X}_t|\mathbf{X}_{t-1}).$$

- Second-order Markov chain



$$\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{0:t-1}) = \mathbf{P}(\mathbf{X}_t|\mathbf{X}_{t-2}, \mathbf{X}_{t-1})$$

Different Distribution for Each Time Step?

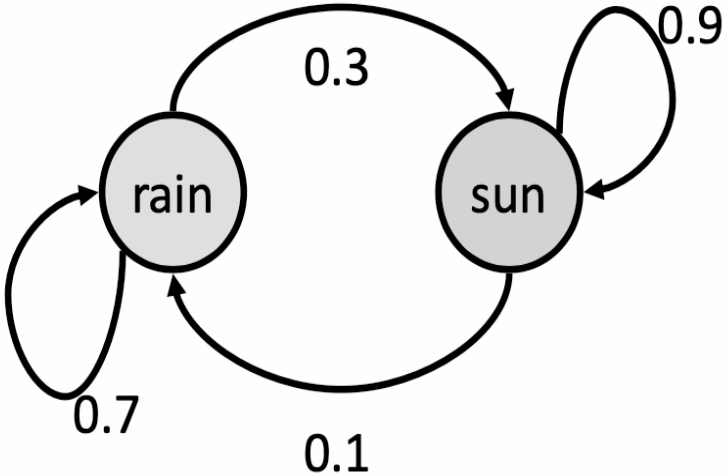
- **Time-homogeneous process (Stationarity assumption)**
  - A process of change that is governed by laws that do not themselves change over time

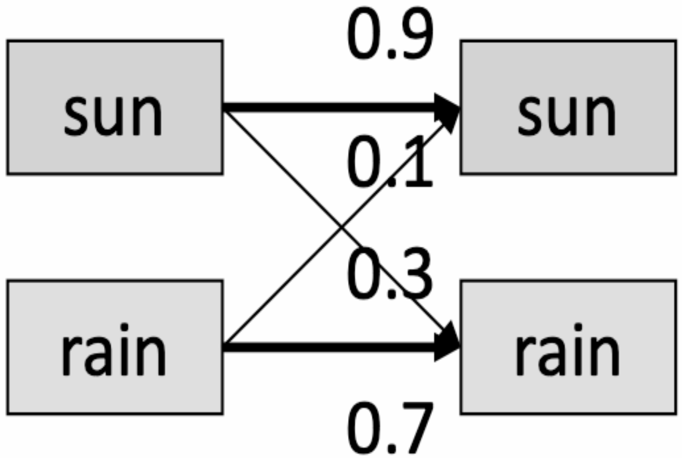
    i.e., transition probabilities are the same at all times

Example: Weather

- Let states: X = {rain, sun}
- Given initial distribution P(sun) = 1.0, and conditional probability table, CPT, $\mathbf{P}(X_t|X_{t-1})$, $P(X_2\text{=sun})$ = ?

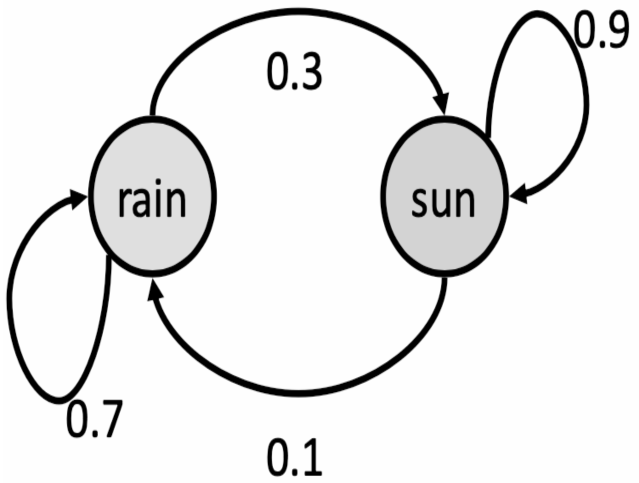
X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7





Example: Weather

- Let states: X = {rain, sun}
- Given initial distribution P(sun) = 1.0, and conditional probability table, CPT, $\mathbf{P}(X_t|X_{t-1})$, $P(X_2=sun)$ = ?



$P(X_2=sun)$

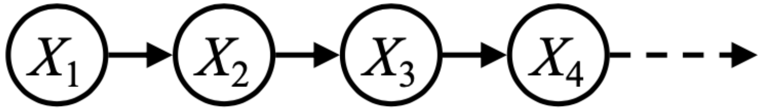
$= \sum_{x1} P(X_2=sun, x_1)$

$= \sum_{x1} P(X_2=sun|x_1)P(x_1)$

$= P(X_2=sun|X_1=sun)P(X_1=sun) + P(X_2=sun|X_1=rain)P(X_1=rain)$

$= 0.9 \cdot 1.0 + 0.3 \cdot 0.0$

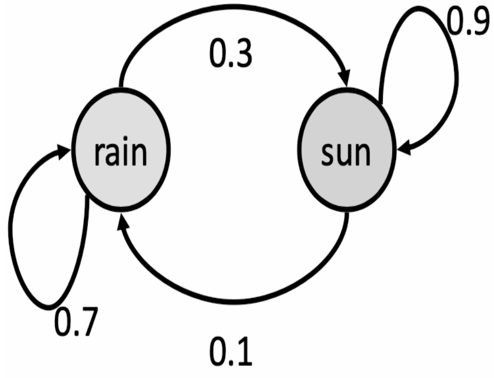
$= 0.9$



Example: Weather

- Given $\mathbf{P}(X_1)$ and CPT $\mathbf{P}(X_t|X_{t-1})$, what's $\mathbf{P}(X_t)$ on some day t?

Example: Weather



- Given $\mathbf{P}(X_1)$ and CPT $\mathbf{P}(X_t|X_{t-1})$, what's $\mathbf{P}(X_t)$ on some day t?
- Let initial observation $P(X_1\text{=sun}) = 1$

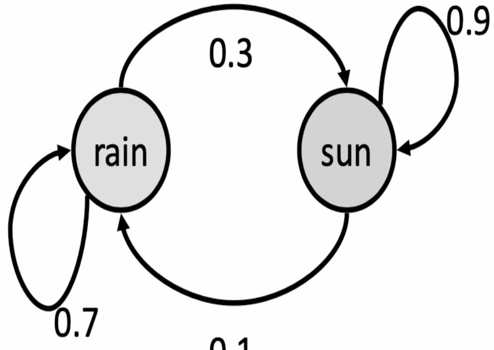
$$\begin{array}{c}\left\langle \begin{array}{c}1.0\\0.0\end{array}\right\rangle\\ \mathbf{P}(X_1)\end{array}$$

Example: Weather

- Given **P**(X_1) and CPT **P**($X_t|X_{t-1}$), what's **P**(X_t) on some day t?
- Let initial observation P(X_1=sun) = 1

$$\begin{pmatrix} 1.0 \\ 0.0 \end{pmatrix} \quad \begin{pmatrix} 0.9 \\ 0.1 \end{pmatrix}$$

$\mathbf{P}(X_1)$ $\quad$ $\mathbf{P}(X_2)$



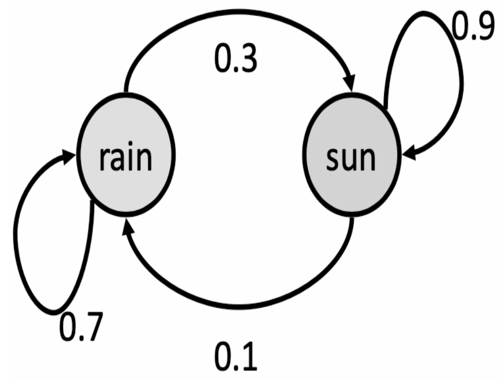
$$[1.0\ 0.0] \begin{array}{c|cc} & \text{sun} & \text{rain} \\ \hline \text{sun} & 0.9 & 0.1 \\ \text{rain} & 0.3 & 0.7 \end{array} = [0.9\ 0.1]$$

to

from [ ]

X_{t-1}	X_t	$P(X_t\|X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Example: Weather

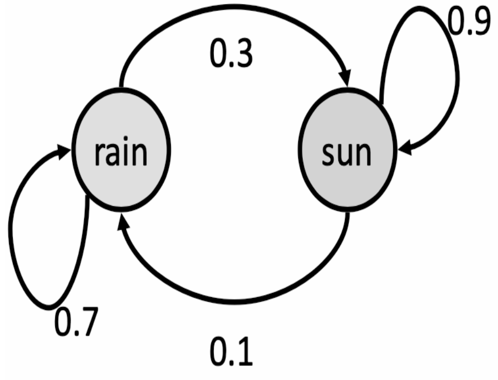


- Given $\mathbf{P}(X_1)$ and CPT $\mathbf{P}(X_t|X_{t-1})$, what's $\mathbf{P}(X_t)$ on some day t?
- Let initial observation $P(X_1\text{=sun}) = 1$

$$[0.9\ 0.1]\begin{bmatrix}0.9 & 0.1\\ 0.3 & 0.7\end{bmatrix} = [0.84\ 0.16]$$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix}1.0\\0.0\end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix}0.9\\0.1\end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_3)}{\left\langle \begin{matrix}0.84\\0.16\end{matrix} \right\rangle}$$

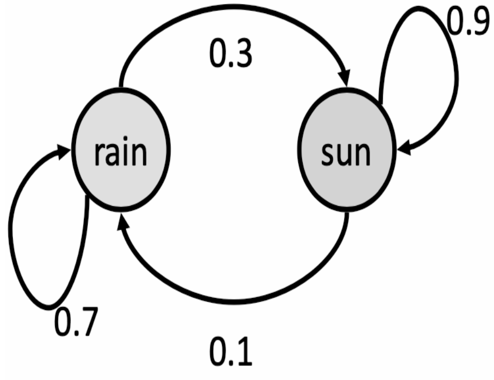
Example: Weather



- Given $\mathbf{P}(X_1)$ and CPT $\mathbf{P}(X_t|X_{t-1})$, what's $\mathbf{P}(X_t)$ on some day t?
- Let initial observation $P(X_1=\text{sun}) = 1$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_3)}{\left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_4)}{\left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle} \quad \Rightarrow \quad \underset{\mathbf{P}(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

Example: Weather



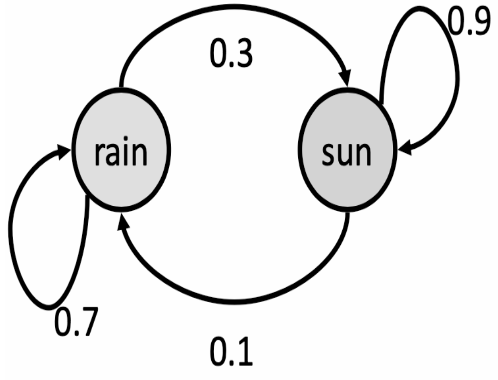
- Given $\mathbf{P}(X_1)$ and CPT $\mathbf{P}(X_t|X_{t-1})$, what's $\mathbf{P}(X_t)$ on some day t?
- Let initial observation $P(X_1\text{=sun}) = 1$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_3)}{\left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_4)}{\left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle} \longrightarrow \underset{\mathbf{P}(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

- Let initial observation $P(X_1\text{=rain}) = 1$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle}$$

Example: Weather



- Given $\mathbf{P}(X_1)$ and CPT $\mathbf{P}(X_t|X_{t-1})$, what's $\mathbf{P}(X_t)$ on some day t?
- Let initial observation $P(X_1\text{=sun}) = 1$

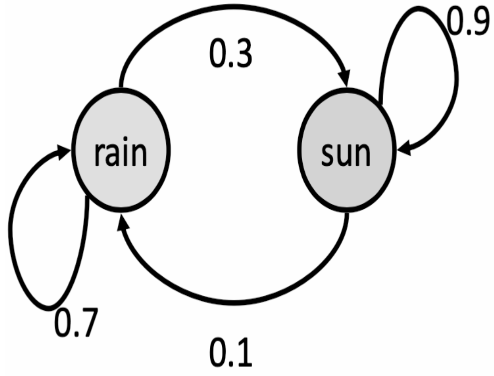
$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_3)}{\left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_4)}{\left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle} \Longrightarrow \underset{\mathbf{P}(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

- Let initial observation $P(X_1\text{=rain}) = 1$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix} 0.3 \\ 0.7 \end{matrix} \right\rangle}$$

$$[0.0\ 1.0] \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix} = [0.3\ 0.7]$$

Example: Weather



- Given $\mathbf{P}(X_1)$ and CPT $\mathbf{P}(X_t|X_{t-1})$, what's $\mathbf{P}(X_t)$ on some day t?
- Let initial observation $P(X_1{=}sun) = 1$

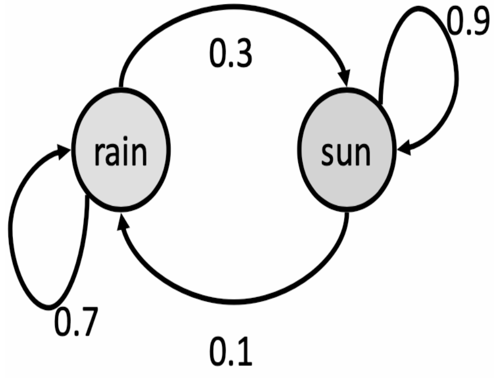
$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_3)}{\left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_4)}{\left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle} \longrightarrow \underset{\mathbf{P}(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

- Let initial observation $P(X_1{=}rain) = 1$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix} 0.3 \\ 0.7 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_3)}{\left\langle \begin{matrix} 0.48 \\ 0.52 \end{matrix} \right\rangle}$$

$$[0.3\ 0.7] \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix} = [0.48\ 0.52]$$

Example: Weather



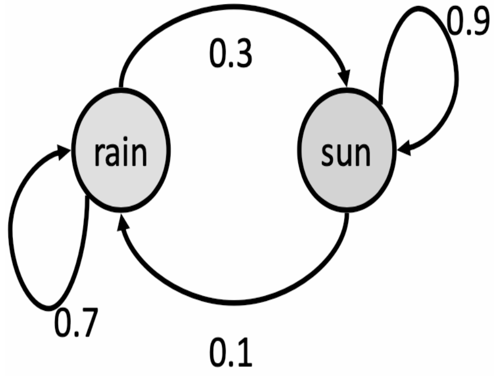
- Given $\mathbf{P}(X_1)$ and CPT $\mathbf{P}(X_t|X_{t-1})$, what's $\mathbf{P}(X_t)$ on some day t?
- Let initial observation $P(X_1\text{=sun}) = 1$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_3)}{\left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_4)}{\left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle} \longrightarrow \underset{\mathbf{P}(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

- Let initial observation $P(X_1\text{=rain}) = 1$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix} 0.3 \\ 0.7 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_3)}{\left\langle \begin{matrix} 0.48 \\ 0.52 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_4)}{\left\langle \begin{matrix} 0.588 \\ 0.412 \end{matrix} \right\rangle} \longrightarrow \underset{\mathbf{P}(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

Example: Weather



- Given $\mathbf{P}(X_1)$ and CPT $\mathbf{P}(X_t|X_{t-1})$, what's $\mathbf{P}(X_t)$ on some day t?
- Let initial observation $P(X_1=\text{sun}) = 1$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_3)}{\left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_4)}{\left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle} \quad \Rightarrow \quad \underset{\mathbf{P}(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

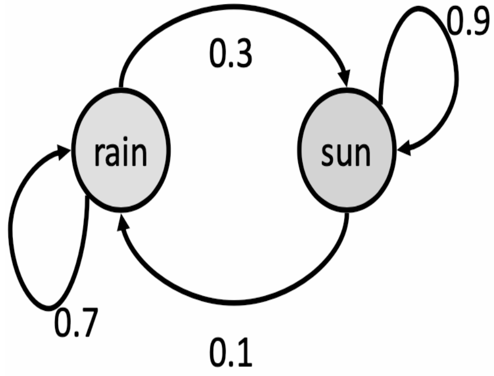
- Let initial observation $P(X_1=\text{rain}) = 1$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix} 0.3 \\ 0.7 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_3)}{\left\langle \begin{matrix} 0.48 \\ 0.52 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_4)}{\left\langle \begin{matrix} 0.588 \\ 0.412 \end{matrix} \right\rangle} \quad \Rightarrow \quad \underset{\mathbf{P}(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

- Let initial distribution $\mathbf{P}(X_1) = \langle p, 1-p \rangle$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} p \\ 1-p \end{matrix} \right\rangle}$$

Example: Weather



- Given $\mathbf{P}(X_1)$ and CPT $\mathbf{P}(X_t|X_{t-1})$, what's $\mathbf{P}(X_t)$ on some day t?
- Let initial observation $P(X_1\text{=sun}) = 1$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_3)}{\left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_4)}{\left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle} \longrightarrow \underset{\mathbf{P}(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

- Let initial observation $P(X_1\text{=rain}) = 1$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_2)}{\left\langle \begin{matrix} 0.3 \\ 0.7 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_3)}{\left\langle \begin{matrix} 0.48 \\ 0.52 \end{matrix} \right\rangle} \quad \underset{\mathbf{P}(X_4)}{\left\langle \begin{matrix} 0.588 \\ 0.412 \end{matrix} \right\rangle} \longrightarrow \underset{\mathbf{P}(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

- Let initial distribution $\mathbf{P}(X_1) = \langle p, 1-p \rangle$

$$\underset{\mathbf{P}(X_1)}{\left\langle \begin{matrix} p \\ 1-p \end{matrix} \right\rangle} \quad \cdots \quad \longrightarrow \underset{\mathbf{P}(X_\infty)}{\left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle}$$

Stationary Distributions

- For most Markov chains:
  - Influence of the initial distribution gets less and less over time
  - The distribution we end up in is independent of the initial distribution
- The distribution we end up with is called
  the **stationary distribution** P_∞ of the Markov chain, and it satisfies

$$P_\infty(X) = P_{\infty+1}(X) = \sum_x P(X|x)P_\infty(x)$$

$$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \dashrightarrow$$

Example

X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

- Given **P**(X_1) and CPT **P**($X_t|X_{t-1}$), what's **P**(X_∞) at time t = ∞?

$$P_\infty(X) = P_{\infty+1}(X) = \sum_x P(X|x)P_\infty(x)$$

$$\begin{cases} P_\infty(sun) = P(sun|sun)P_\infty(sun) + P(sun|rain)P_\infty(rain) \\ P_\infty(rain) = P(rain|sun)P_\infty(sun) + P(rain|rain)P_\infty(rain) \end{cases}$$

$$\begin{cases} P_\infty(sun) = 0.9P_\infty(sun) + 0.3P_\infty(rain) \\ P_\infty(rain) = 0.1P_\infty(sun) + 0.7P_\infty(rain) \end{cases}$$

$$P_\infty(sun) = 3P_\infty(rain)$$

$$P_\infty(rain) = 1/3P_\infty(sun)$$

$$P_\infty(sun) + P_\infty(rain) = 1$$

$$\Rightarrow \begin{cases} P_\infty(sun) = 3/4 \\ P_\infty(rain) = 1/4 \end{cases}$$

Real World

(unknown)

Hidden State	Observation
Weather	Umbrella
Words Spoken	Audio Waveforms
User Engagement	Website or App Analytics
Robot’s Position	Robot’s Sensor Data

Sensor Model (Observation Model)

- The evidence variables could depend on previous evidence variables as well as the current state variables

Sensor Markov Assumption

- The evidence variable depends only the corresponding state

states before evidence

$$\mathbf{P}(\mathbf{E}_t | \mathbf{X}_{0:t}, \mathbf{E}_{1:t-1}) = \mathbf{P}(\mathbf{E}_t | \mathbf{X}_t)$$

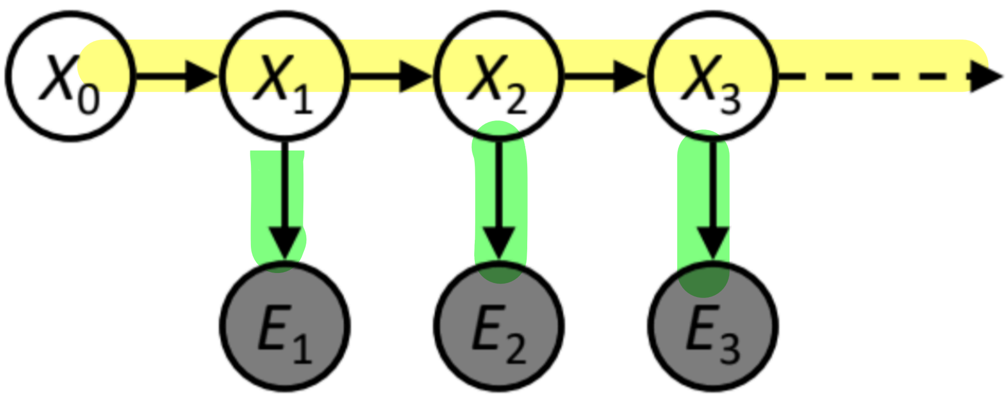
states after evidence

  - X_t: state variable at time t
  - E_t: observable evidence variable at time t

$x_0, x_1, \ldots, x_n$ are hidden (unknown)

Hidden Markov Model (HMM)

- A Markov model for a system with hidden states that generate some observed event
  - Markov assumption
  - Sensor Markov assumption
- Example

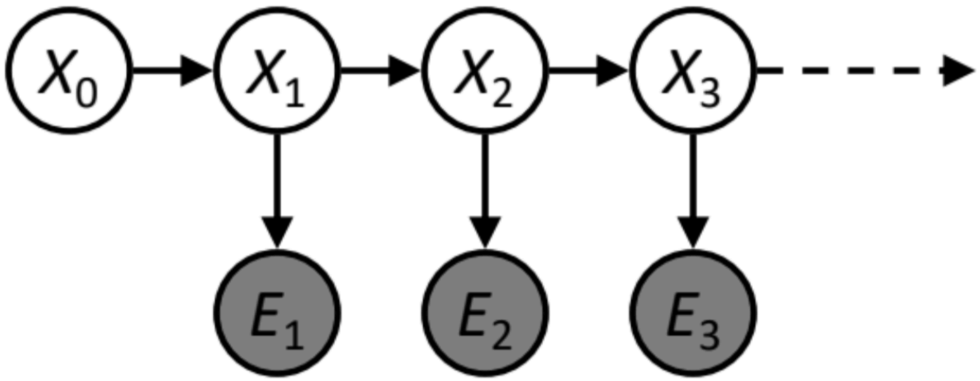


Probability Model: HMM

- Joint distribution for hidden Markov model:

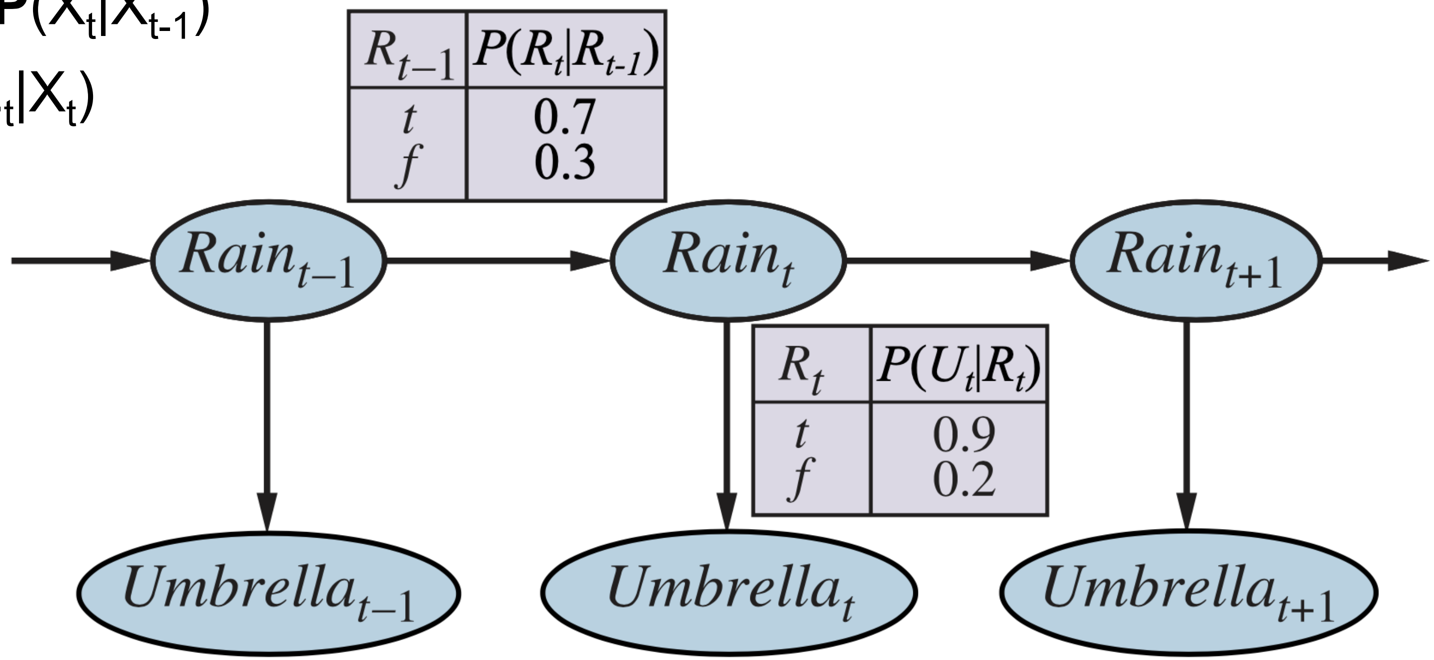
$$\mathbf{P}(\mathbf{X}_{0:t}, \mathbf{E}_{1:t}) = \mathbf{P}(\mathbf{X}_0) \prod_{i=1}^{t} \mathbf{P}(\mathbf{X}_i \mid \mathbf{X}_{i-1}) \mathbf{P}(\mathbf{E}_i \mid \mathbf{X}_i)$$

Initial State Model Transition Model Sensor Model



Example: HMM

- HMM is defined by
  - Initial distribution/initial state model $\mathbf{P}(X_1)$
  - Transition model $\mathbf{P}(X_t|X_{t-1})$
  - Sensor model $\mathbf{P}(E_t|X_t)$



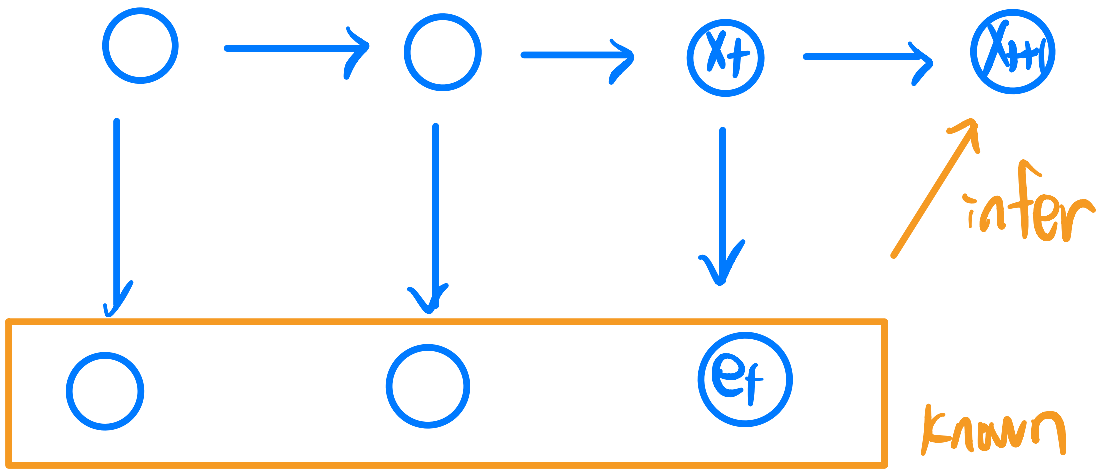
Inference Tasks in Temporal Models

- Filtering (State estimation)
  - Given observations from start until now, calculate distribution for **current** state
- Prediction
  - Given observations from start until now, calculate distribution for a **future** state
- Smoothing
  - Given observations from start until now, calculate distribution for **past** state
- Most likely explanation
  - Given observations from start until now, calculate distribution for most likely **sequence** of states (that have generated those observations)

Prediction

- Given observations from start until now, calculate distribution for a **future** state

$$\mathbf{P}(X_{t+1} \mid e_{1:t})$$



Prediction

$1 \sim t$ $\quad$ $t+1$

- Given observations from start until now, calculate distribution for a **future** state

$$\mathbf{P}(X_{t+1} \mid e_{1:t})$$

- Calculation:

$(x_t$: current state$)$

$$\mathbf{P}(X_{t+1} \mid e_{1:t}) = \frac{\mathbf{P}(X_{t+1}, e_{1:t})}{P(e_{1:t})} = \Sigma_{x_t} \frac{\mathbf{P}(X_{t+1}, x_t, e_{1:t})}{P(e_{1:t})} \underset{\text{Chain Rule}}{=} \Sigma_{x_t} \frac{\mathbf{P}(X_{t+1}, x_t, e_{1:t})}{P(x_t, e_{1:t})} \frac{P(x_t, e_{1:t})}{P(e_{1:t})}$$

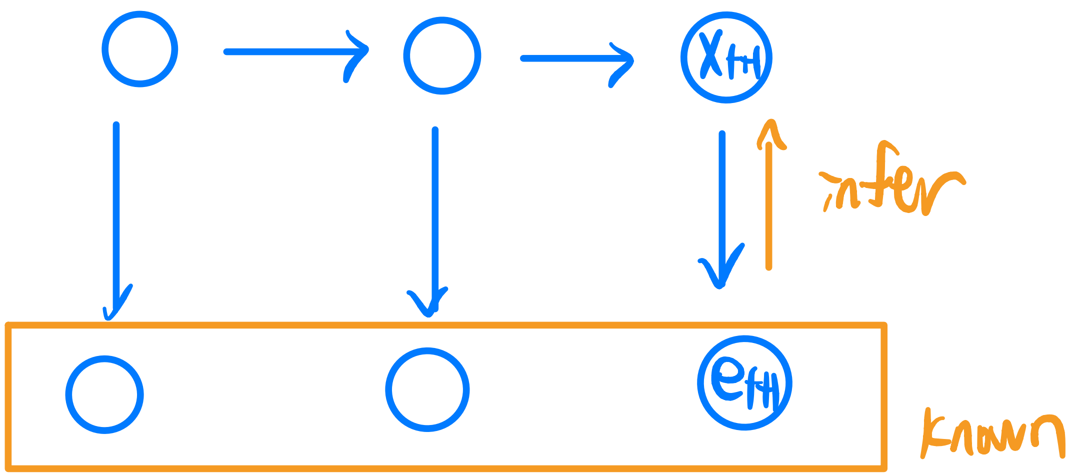
$$= \Sigma_{x_t} \mathbf{P}(X_{t+1} \mid x_t, e_{1:t}) P(x_t \mid e_{1:t})$$

$$= \sum_{\mathbf{x}_t} \underbrace{\mathbf{P}\left(\mathbf{X}_{t+1} \mid \mathbf{x}_t\right)}_{\text{transition model}} \underbrace{P\left(\mathbf{x}_t \mid \mathbf{e}_{1:t}\right)}_{\text{recursion}}$$ (By Markov assumption)

Filtering (State Estimation)

- Given observations from start until now, calculate distribution for **current** state

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



Filtering (State Estimation)

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

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

1 ~ t+1

- Given observations from start until now, calculate distribution for **current** state

t+1

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

- Calculation:

$$\mathbf{P}(\mathbf{X}_{t+1} | \mathbf{e}_{1:t+1}) = \mathbf{P}(\mathbf{X}_{t+1} | \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \quad \text{(dividing up the evidence)}$$

$$= \alpha \mathbf{P}(\mathbf{e}_{t+1} | \mathbf{X}_{t+1}, \mathbf{e}_{1:t}) \, \mathbf{P}(\mathbf{X}_{t+1} | \mathbf{e}_{1:t}) \quad \text{(using Bayes'rule, given } \mathbf{e}_{1:t}\text{)}$$

$$= \alpha \mathbf{P}(\mathbf{e}_{t+1} | \mathbf{X}_{t+1}) \mathbf{P}(\mathbf{X}_{t+1} | \mathbf{e}_{1:t}) \quad \text{(by the sensor Markov assumption)}$$

$$= \alpha \underbrace{\mathbf{P}(\mathbf{e}_{t+1} | \mathbf{X}_{t+1})}_{\text{sensor model}} \sum_{\mathbf{x}_t} \underbrace{\mathbf{P}(\mathbf{X}_{t+1} | \mathbf{x}_t)}_{\text{transition model}} \underbrace{P(\mathbf{x}_t | \mathbf{e}_{1:t})}_{\text{recursion}}$$

this state ↓ this observation

last state

last state ↓ this state

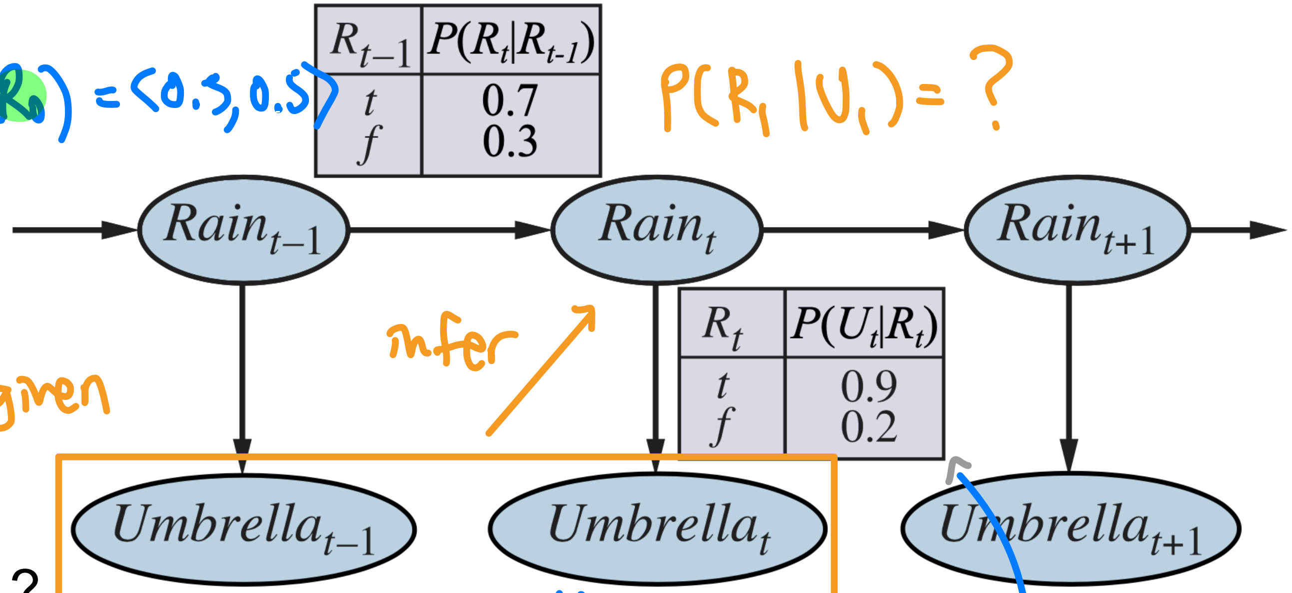
prob. of last state

transition model

Example: Filtering

$P(R_0) = \langle 0.5, 0.5 \rangle$

$P(R_1 | U_1) = ?$



infer

given

$U_1 = \text{true}$

sensor model

- Day 0: No observations

$$\mathbf{P}(R_0) = \langle 0.5, 0.5 \rangle$$

- Day 1: U_1 = true, $\mathbf{P}(R_1 \,|\, u_1)$ = ?

$$\begin{aligned}\mathbf{P}(R_1|u_1) &= \alpha \mathbf{P}(u_1|R_1)\mathbf{P}(R_1) = \alpha \langle 0.9, 0.2 \rangle \langle 0.5, 0.5 \rangle \\ &= \alpha \langle 0.45, 0.1 \rangle \approx \langle 0.818, 0.182 \rangle .\end{aligned}$$

$$\begin{aligned}\mathbf{P}(R_1) &= \sum_{r_0} \mathbf{P}(R_1|r_0)P(r_0) \\ &= \langle 0.7, 0.3 \rangle \times 0.5 + \langle 0.3, 0.7 \rangle \times 0.5 = \langle 0.5, 0.5 \rangle\end{aligned}$$

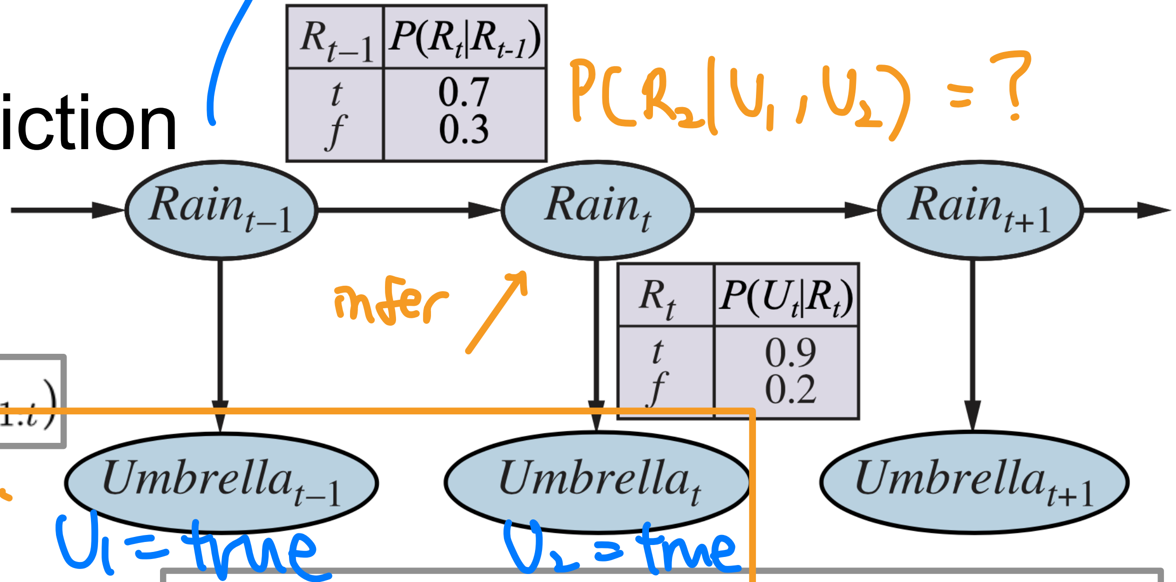
Example: Filtering and Prediction

$P(R_1 \mid U_1) = \langle 0.818, 0.182 \rangle$

$P(R_2 \mid U_1, U_2) = ?$

- Day 2: U_2 = true, $\mathbf{P}(R_2 \mid u_1, u_2)$ = ?

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



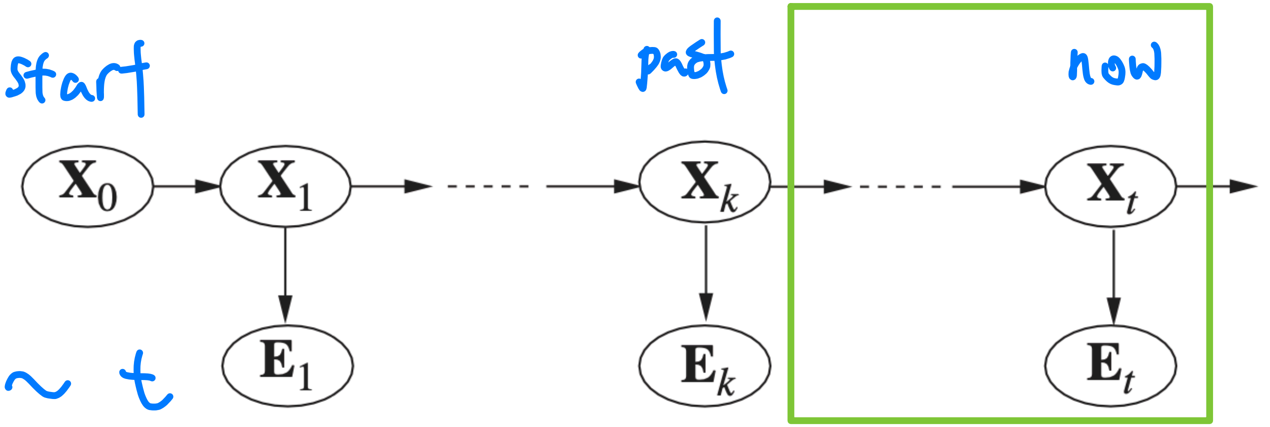
$$\begin{aligned}
&\mathbf{P}(R_2|u_1, u_2) \\
&= \alpha \mathbf{P}(u_2|R_2)\mathbf{P}(R_2|u_1) \\
&= \alpha \langle 0.9, 0.2 \rangle \langle 0.627, 0.373 \rangle \\
&= \alpha \langle 0.565, 0.075 \rangle \approx \langle 0.883, 0.117 \rangle .
\end{aligned}$$

given

$$\boxed{\mathbf{P}(X_{t+1} \mid e_{1:t}) = \sum_{\mathbf{x}_t} \underbrace{\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{x}_t)}_{\text{transition model}} \underbrace{P(\mathbf{x}_t|\mathbf{e}_{1:t})}_{\text{recursion}}}$$

$$\begin{aligned}
\mathbf{P}(R_2|u_1) &= \sum_{r_1} \mathbf{P}(R_2|r_1)P(r_1|u_1) \\
&= \langle 0.7, 0.3 \rangle \times 0.818 + \langle 0.3, 0.7 \rangle \times 0.182 \approx \langle 0.627, 0.373 \rangle
\end{aligned}$$

Smoothing



1 ~ t

- Given observations from start until now, calculate distribution for **past** state k

Gives us a more accurate result

before past state

$$\mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:t}) \text{ for } 0 \leq k < t.$$

after past state

- Calculation:

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

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:t}) &= \mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:k}, \mathbf{e}_{k+1:t}) \\
&= \alpha\, \mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:k})\mathbf{P}(\mathbf{e}_{k+1:t} \mid \mathbf{X}_k, \mathbf{e}_{1:k}) \quad \text{(using Bayes' rule)} \\
&= \alpha\, \mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:k})\mathbf{P}(\mathbf{e}_{k+1:t} \mid \mathbf{X}_k) \quad \text{(using conditional independence)} \\
&= \alpha \left[\mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:k})\right] \left[\sum_{\mathbf{x}_{k+1}} \underbrace{P(\mathbf{e}_{k+1} \mid \mathbf{x}_{k+1})}_{\text{sensor model}} \underbrace{P(\mathbf{e}_{k+2:t} \mid \mathbf{x}_{k+1})}_{\text{recursion}} \underbrace{\mathbf{P}(\mathbf{x}_{k+1} \mid \mathbf{X}_k)}_{\text{transition model}}\right]
\end{aligned}$$

forward pass (filtering up to k)

the state immediately after the past state

backward pass

backward pass

$$\mathbf{P}(\mathbf{e}_{k+1:t}|\mathbf{X}_k) = \frac{\mathbf{P}(e_{k+1:t}, X_k)}{\mathbf{P}(X_k)} = \Sigma_{x_{k+1}} \frac{\mathbf{P}(e_{k+1:t}, X_k, x_{k+1})}{\mathbf{P}(X_k)}$$

$$= \Sigma_{x_{k+1}} \frac{\mathbf{P}(e_{k+1:t}, X_k, x_{k+1})}{P(x_{k+1}, X_k)} \frac{P(x_{k+1}, X_k)}{P(X_k)}$$

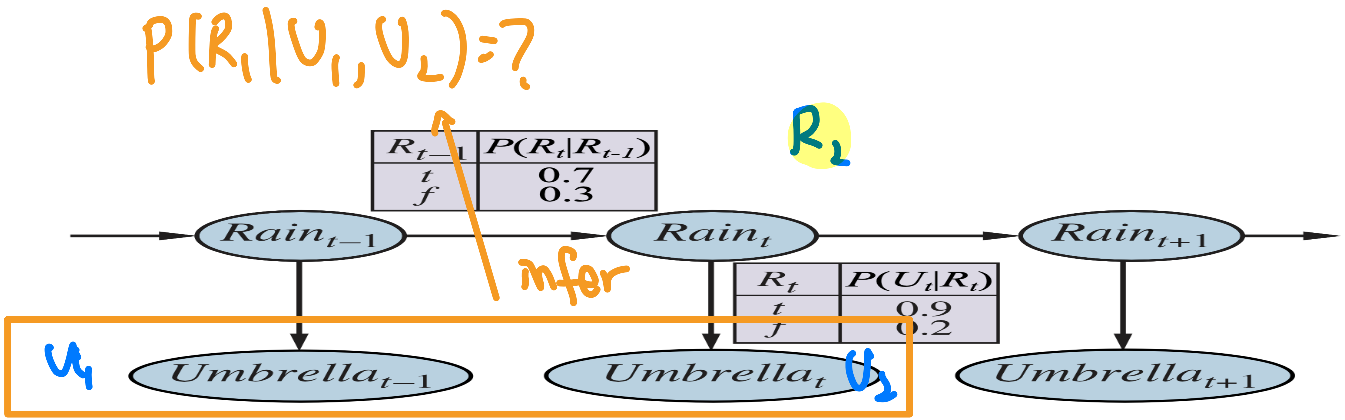
$$= \sum_{\mathbf{x}_{k+1}} \mathbf{P}(\mathbf{e}_{k+1:t}|\mathbf{X}_k, \mathbf{x}_{k+1})\mathbf{P}(\mathbf{x}_{k+1}|\mathbf{X}_k) \quad (\text{conditioning on } \mathbf{X}_{k+1})$$

$$= \sum_{\mathbf{x}_{k+1}} P(\mathbf{e}_{k+1:t}|\mathbf{x}_{k+1})\mathbf{P}(\mathbf{x}_{k+1}|\mathbf{X}_k) \quad (\text{by conditional independence})$$

$$= \sum_{\mathbf{x}_{k+1}} P(\mathbf{e}_{k+1}, \mathbf{e}_{k+2:t}|\mathbf{x}_{k+1})\mathbf{P}(\mathbf{x}_{k+1}|\mathbf{X}_k) \qquad \frac{\mathbf{P}(e_{k+1}, e_{k+2:t}, x_{k+1})}{P(e_{k+2:t}, x_{k+1})} \frac{P(e_{k+2:t}, x_{k+1})}{P(x_{k+1})}$$

$$= \sum_{\mathbf{x}_{k+1}} \underbrace{P(\mathbf{e}_{k+1}|\mathbf{x}_{k+1})}_{\text{sensor model}} \underbrace{P(\mathbf{e}_{k+2:t}|\mathbf{x}_{k+1})}_{\text{recursion}} \underbrace{\mathbf{P}(\mathbf{x}_{k+1}|\mathbf{X}_k)}_{\text{transition model}},$$

Example: Smoothing



- Given the umbrella observations on days 1 and 2, for time k = 1,

  $\mathbf{P}(R_1|u_1,u_2)$ = ?

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

$$= \alpha\, \mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:k}) \sum_{\mathbf{x}_{k+1}} \underbrace{P(\mathbf{e}_{k+1}|\mathbf{x}_{k+1})}\underbrace{P(\mathbf{e}_{k+2:t}|\mathbf{x}_{k+1})}\underbrace{\mathbf{P}(\mathbf{x}_{k+1}|\mathbf{X}_k)}$$

The probability of observing an empty sequence is 1

- $\mathbf{P}(R_1|u_1,u_2) = \alpha \mathbf{P}(R_1|u_1)\mathbf{P}(u_2|R_1)$

$$= \alpha\langle 0.818, 0.182\rangle \sum_{r_2} P(u_2|r_2)P(\ |r_2)\mathbf{P}(r_2|R_1)$$

r_2 : true $\qquad\qquad$ r_2 : false

$$= \alpha\langle 0.818, 0.182\rangle \left[(0.9 \times 1 \times \langle 0.7, 0.3\rangle) + (0.2 \times 1 \times \langle 0.3, 0.7\rangle)\right]$$

$$= \alpha\langle 0.818, 0.182\rangle \times \langle 0.69, 0.41\rangle \approx \langle 0.883, 0.117\rangle.$$

Most Likely Explanation

- Given observations from start until now, calculate distribution for most likely **sequence** of states (that have generated those observations)

$$\arg\max_{x_{i:t}} P(x_{1:t} \mid e_{1:t})$$

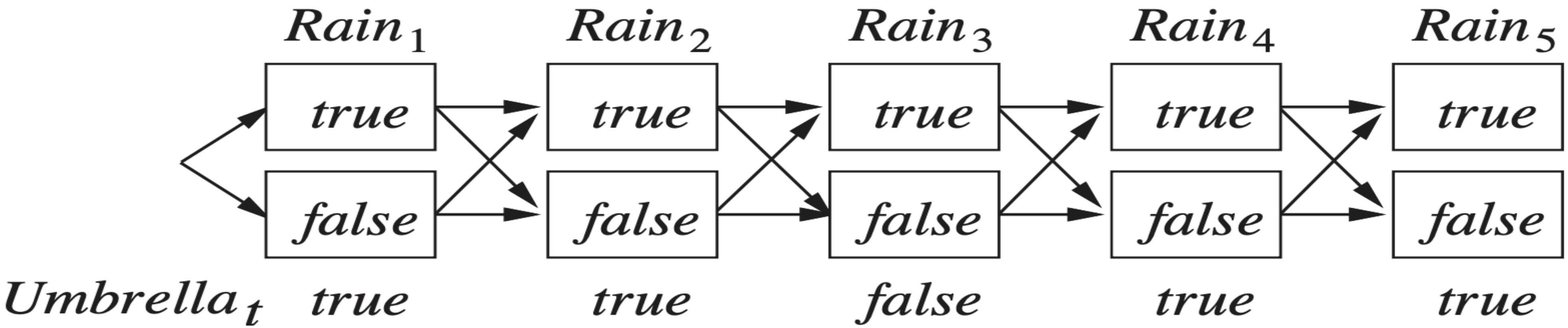
- Calculation:
  - $\arg\max_{x_{i:t}} P(x_{1:t} \mid e_{1:t}) \overset{\text{Bayes Rule}}{=} \arg\max_{x_{i:t}} \alpha P(x_{1:t}, e_{1:t})$

$$= \arg\max_{x_{i:t}} P(x_{1:t}, e_{1:t})$$

$$= \arg\max_{x_{i:t}} \underbrace{P(x_0)}_{\text{initial state}} \prod_{x_{1:t}} \underbrace{P(x_t \mid x_{t-1})}_{\text{transition model}} \underbrace{P(e_t \mid x_t)}_{\text{sensor model}}$$

Example: Most Likely Explanation

- Graph of states and transitions over time
  - Given umbrella sequence: [true,true,false,true,true], find the most probable path



  - Each arc represents some transition $x_{t-1} \rightarrow x_t$
  - Each arc has weight $P(x_t|x_{t-1})P(e_t|x_t)$
  - The product of weights on a path is proportional to that state sequence's probability
  - Viterbi algorithm:
    - For each state at time t, keep track of the maximum probability of any path to it

Example: Most Likely Explanation

transition model

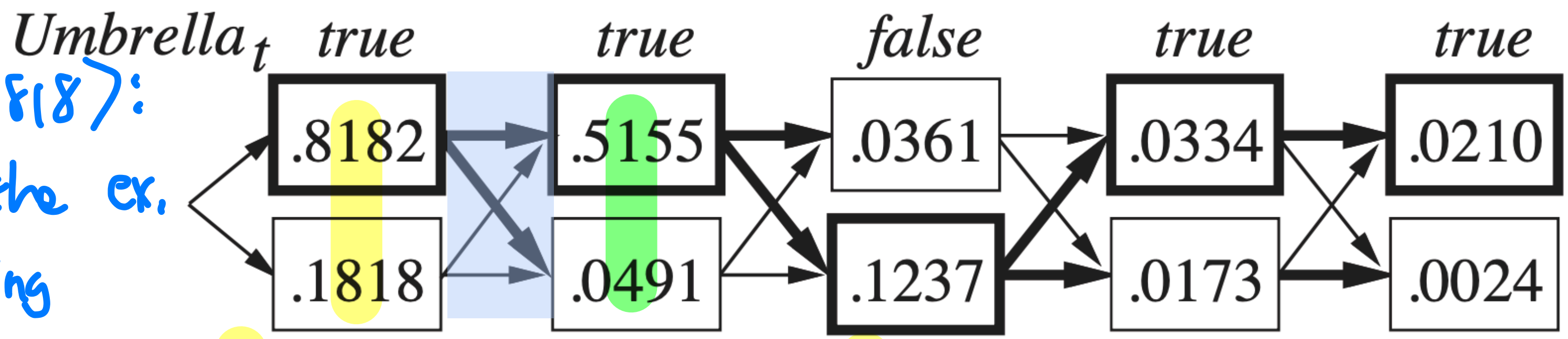
R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

sensor model

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2

- Graph of states and transitions over time
  - Given umbrella sequence: [true,true,false,true,true], find the most probable path
  - Each arc has weight $P(x_t|x_{t-1})P(e_t|x_t)$
  - The product of weights on a path is proportional to that state sequence's probability

<0.8182, 0.1818>: Refer to the ex. in filtering



true : **0.8182*0.7*0.9=0.5155** > 0.1818*0.3*0.9=0.0491

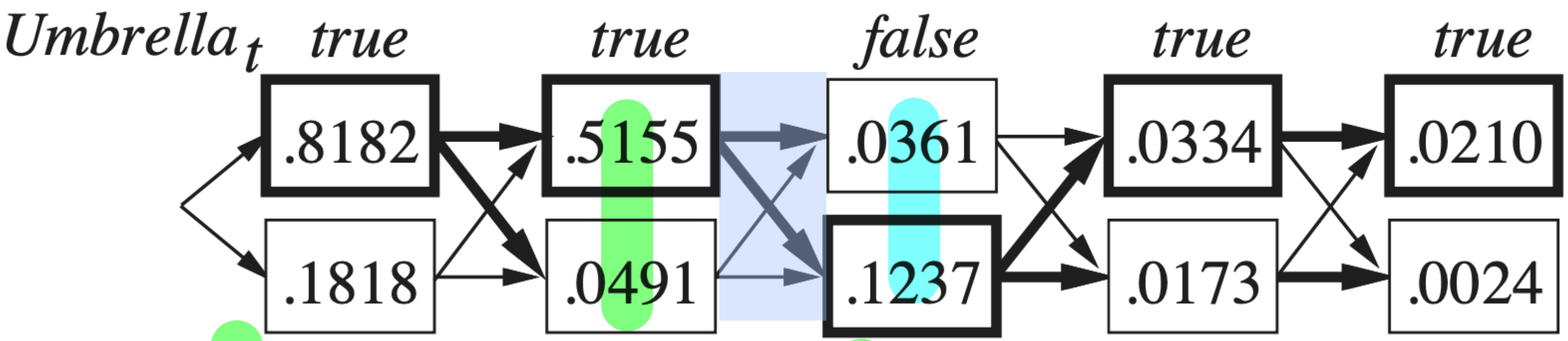
false : **0.8182*0.3*0.2=0.0491** > 0.1818*0.7*0.2=0.0255

Example: Most Likely Explanation

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

- Graph of states and transitions over time
  - Given umbrella sequence: [true,true,false,true,true], find the most probable path
  - Each arc has weight $P(x_t|x_{t-1})P(e_t|x_t)$
  - The product of weights on a path is proportional to that state sequence's probability



true: **0.5155*0.7*0.1=0.0361** > 0.0491*0.3*0.1=0.0015

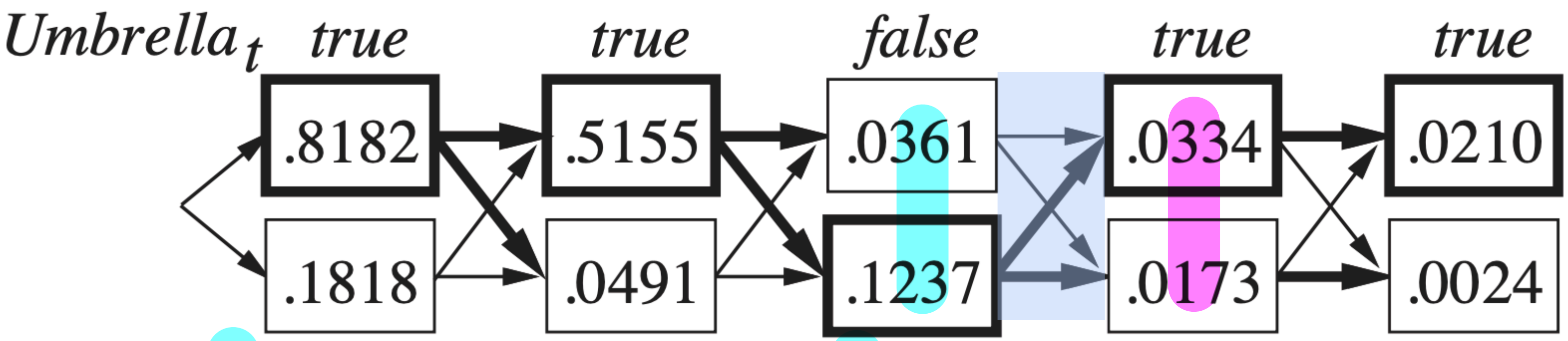
false: **0.5155*0.3*0.8=0.1237** > 0.0491*0.7*0.8=0.0275

Example: Most Likely Explanation

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

- Graph of states and transitions over time
  - Given umbrella sequence: [true,true,false,true,true], find the most probable path
  - Each arc has weight $P(x_t|x_{t-1})P(e_t|x_t)$
  - The product of weights on a path is proportional to that state sequence's probability



true: 0.0361*0.7*0.9=0.0227< **0.1237*0.3*0.9=0.0334**

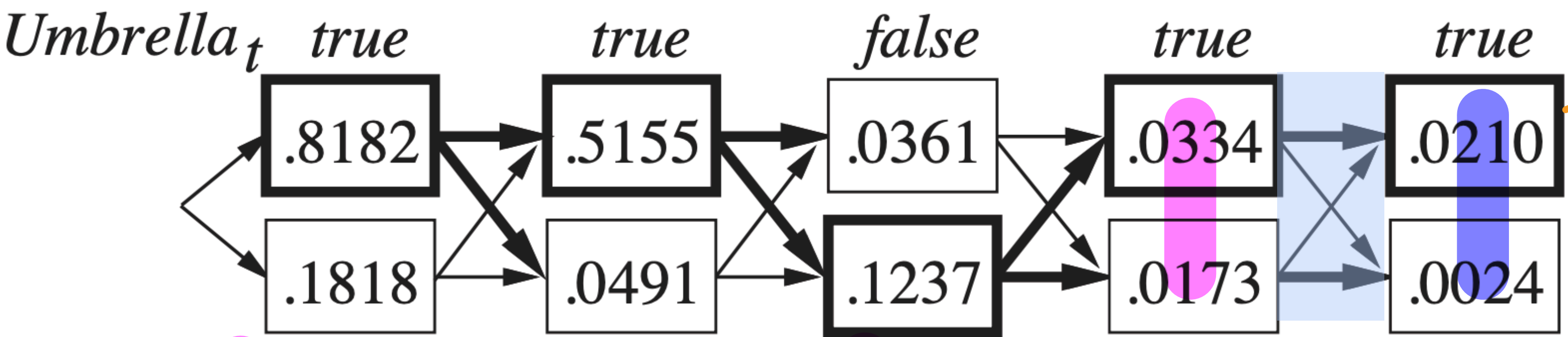
false: 0.0361*0.3*0.2=0.0022< **0.1237*0.7*0.2=0.0173**

Example: Most Likely Explanation

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

- Graph of states and transitions over time
  - Given umbrella sequence: [true,true,false,true,true], find the most probable path
  - Each arc has weight $P(x_t|x_{t-1})P(e_t|x_t)$
  - The product of weights on a path is proportional to that state sequence's probability



.021 > .0024

true: **0.0334*0.7*0.9=0.0210** > 0.0173*0.3*0.9=0.0047

false: 0.0334*0.3*0.2=0.0020 < **0.0173*0.7*0.2=0.0024**