Recap: Q-Learning with state abstraction

- Using a feature representation, we can write a Q function (or value function) for any state using a few weights:

$$V(s) = w_1 f_1(s) + w_2 f_2(s) + \cdots + w_n f_n(s)$$

$$Q(s,a) = w_1 f_1(s,a) + w_2 f_2(s,a) + \cdots + w_n f_n(s,a)$$

- Advantage: our experience is summed up in a few powerful numbers
- Disadvantage: states may share features but actually be very different in value!

Function Approximation

$$Q(s,a) = w_1 f_1(s,a) + w_2 f_2(s,a) + \cdots + w_n f_n(s,a)$$

- Q-learning with linear Q-functions:

  transition = (s,a,r,s')

$$\text{difference} = \left[r + \gamma \max_{a'} Q(s',a')\right] - Q(s,a)$$

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left[\text{difference}\right]$$

Exact Q's

$$w_i \leftarrow w_i + \alpha\left[\text{difference}\right] f_i(s,a)$$

Approximate Q's

- Intuitive interpretation:
  - Adjust weights of active features
  - E.g. if something unexpectedly bad happens, disprefer all states with that state's features
- Formal justification: online least squares

Example: Q-Pacman

$Q(s,a)= 4.0f_{DOT}(s,a)-1.0f_{GST}(s,a)$

$f_{DOT}(s,\text{NORTH})=0.5$

$f_{GST}(s,\text{NORTH})=1.0$

$Q(s,a)=+1$

$R(s,a,s')=-500$

$difference=-501$

$w_{DOT} \leftarrow 4.0+\alpha[-501]0.5$

$w_{GST} \leftarrow -1.0+\alpha[-501]1.0$

$Q(s,a)= 3.0f_{DOT}(s,a)-3.0f_{GST}(s,a)$

s

a = North

r = -500

s'

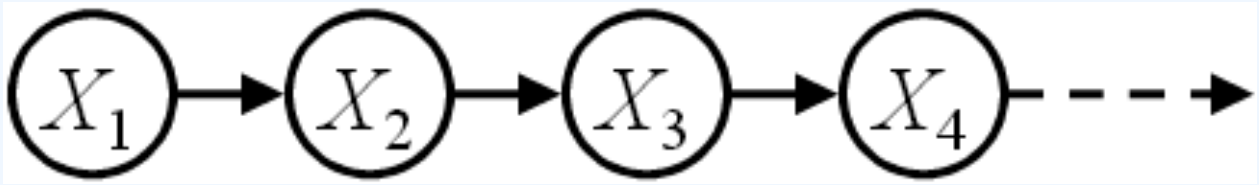
Today: Reasoning over Time

- Often, we want to reason about a sequence of observations
  - Speech recognition
  - Robot localization
  - User attention
  - Medical monitoring

- Need to introduce time into our models
- Basic approach: hidden Markov models (HMMs)
- More general: dynamic Bayes' nets

Markov Models

- A Markov model is a chain-structured BN
  - Conditional probabilities are the same (stationarity)
  - Value of X at a given time is called the state
  - As a BN:

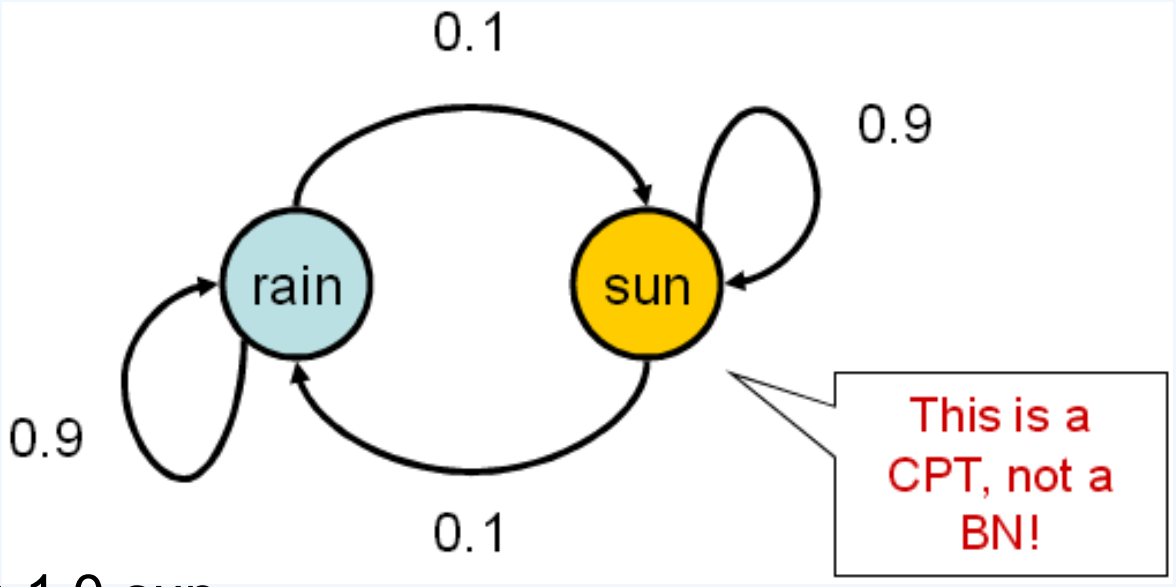


$p(X_1)$ $p(X|X_{-1})$

  - Parameters: called transition probabilities or dynamics, specify how the state evolves over time (also, initial probabilities)

Example: Markov Chain

- Weather:
  - States: X = {rain, sun}
  - Transitions:

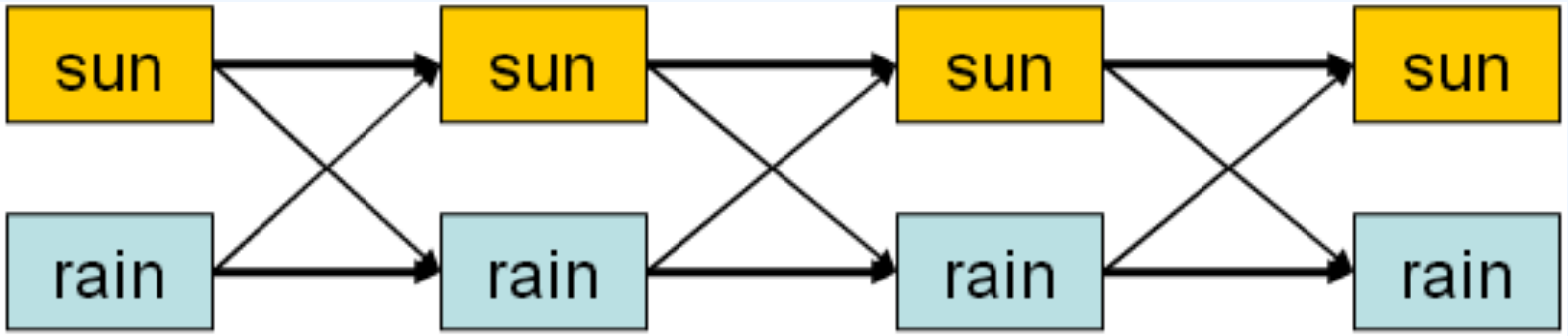


  - Initial distribution: 1.0 sun
  - What's the probability distribution after one step?

$$p(X_2=\text{sun})=p(X_2=\text{sun}|X_1=\text{sun})p(X_1=\text{sun})+ p(X_2=\text{sun}|X_1=\text{rain})p(X_1=\text{rain})$$
$$=0.9*1.0+1.0*0.0 \quad =0.9$$

Forward Algorithm

- Question: What's p(X) on some day t?



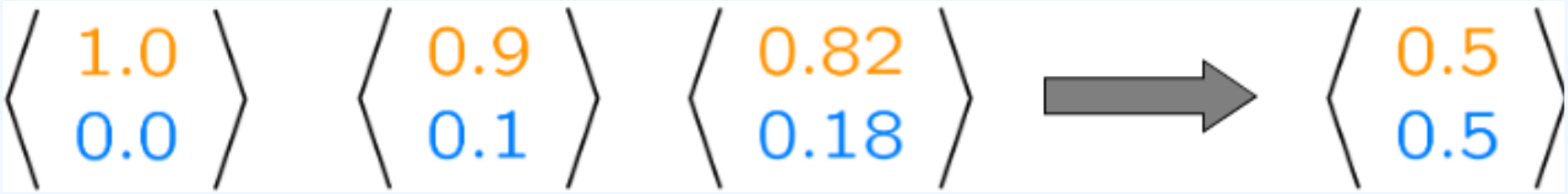
$$p(x_t) = \sum_{x_{t-1}} p(x_t \mid x_{t-1})\, p(x_{t-1})$$

$$p(x_1) = \text{known}$$

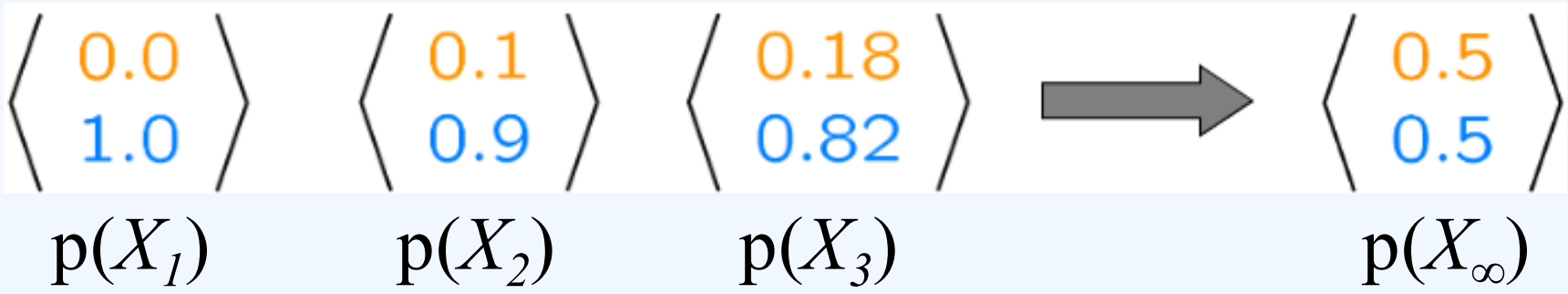
Forward simulation

Example

- From initial observation of sun



- From initial observation of rain

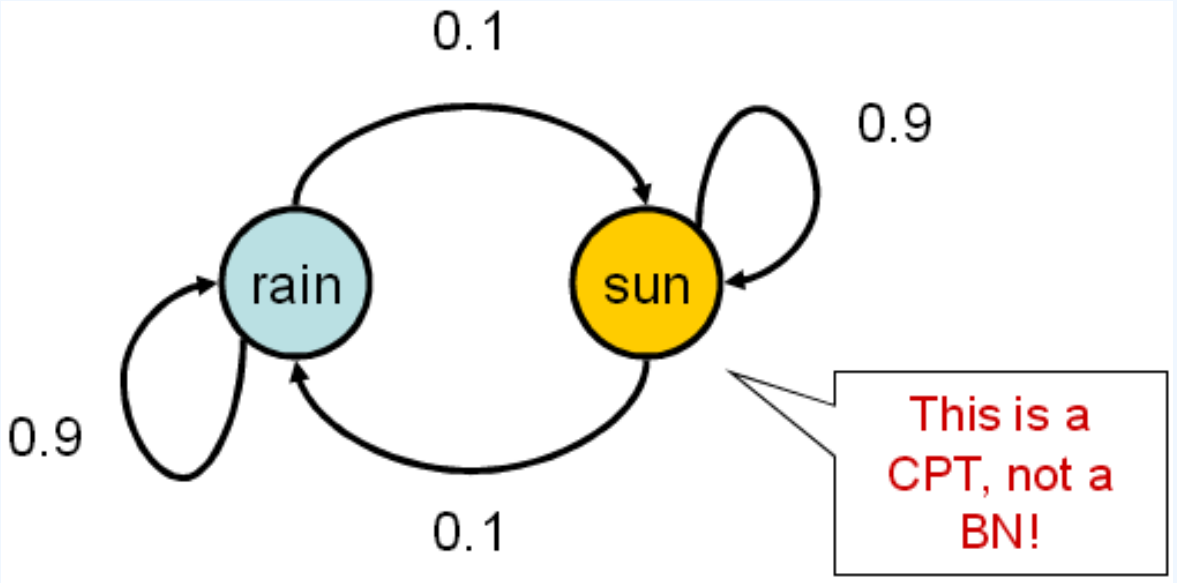


Stationary Distributions

- If we simulate the chain long enough:
  - What happens?
  - Uncertainty accumulates
  - Eventually, we have no idea what the state is!

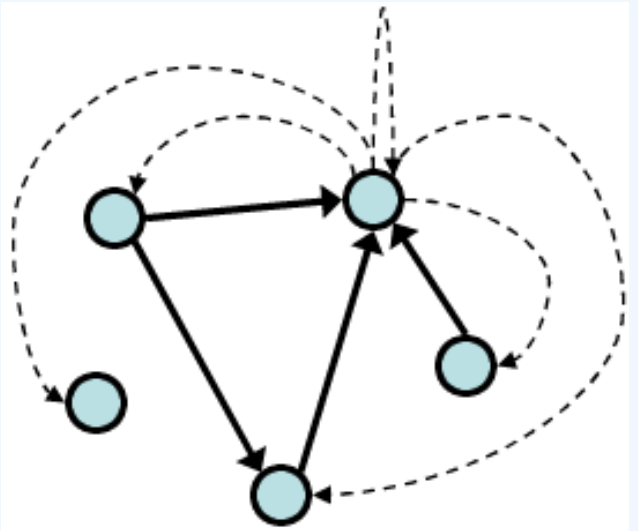
- Stationary distributions:
  - For most chains, the distribution we end up in is independent of the initial distribution
  - Called the stationary distribution of the chain
  - Usually, can only predict a short time out

Computing the stationary distribution



- $p(X=\text{sun})=p(X=\text{sun}|X_{-1}=\text{sun})p(X=\text{sun})+$
  $p(X=\text{sun}|X_{-1}=\text{rain})p(X=\text{rain})$
- $p(X=\text{rain})=p(X=\text{rain}|X_{-1}=\text{sun})p(X=\text{sun})+$
  $p(X=\text{rain}|X_{-1}=\text{rain})p(X=\text{rain})$

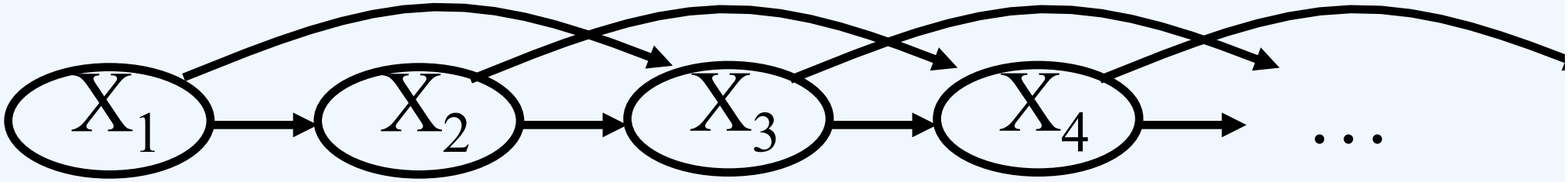
Web Link Analysis



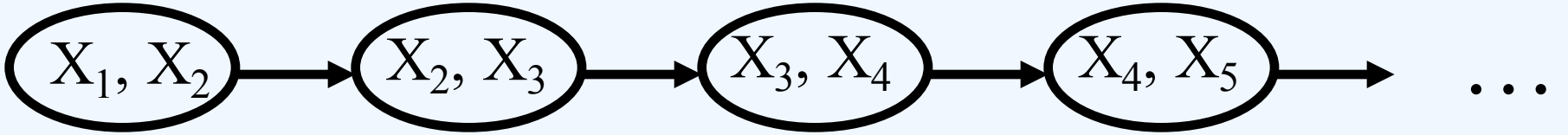
- PageRank over a web graph
  - Each web page is a state
  - Initial distribution: uniform over pages
  - Transitions:
    - With prob. c, uniform jump to a random page (dotted lines, not all shown)
    - With prob. 1-c, follow a random outlink (solid lines)
- Stationary distribution
  - Will spend more time on highly reachable pages
  - Somewhat robust to link spam

Restrictiveness of Markov models

- Are past and future really independent given current state?
- E.g., suppose that when it rains, it rains for at most 2 days



- Second-order Markov process
- Workaround: change meaning of "state" to events of last 2 days

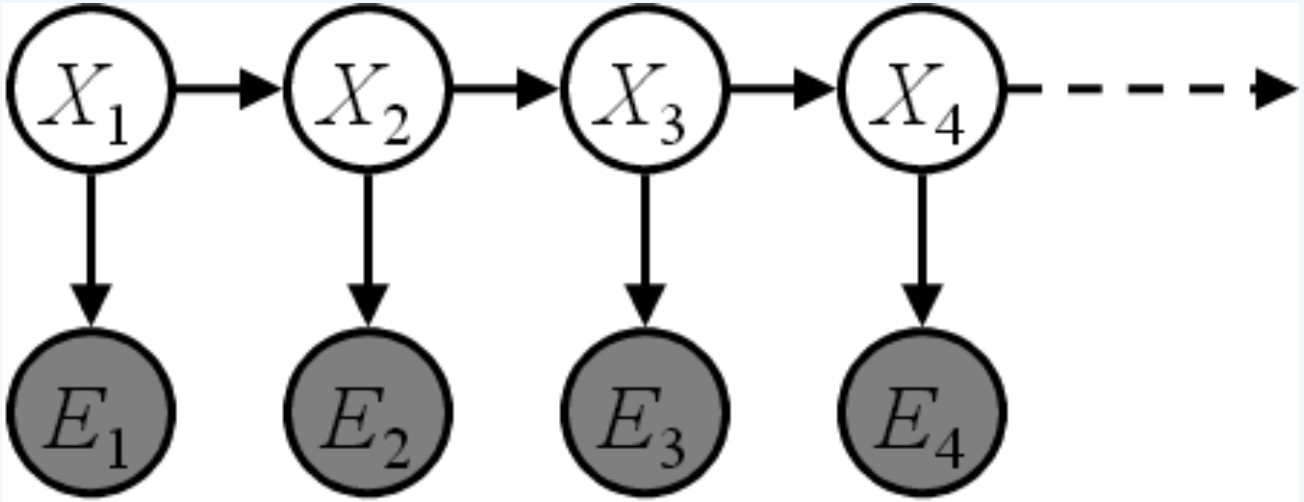


- Another approach: add more information to the state
- E.g., the full state of the world would include whether the sky is full of water
  - Additional information may not be observable
  - Blowup of number of states…

Hidden Markov Models

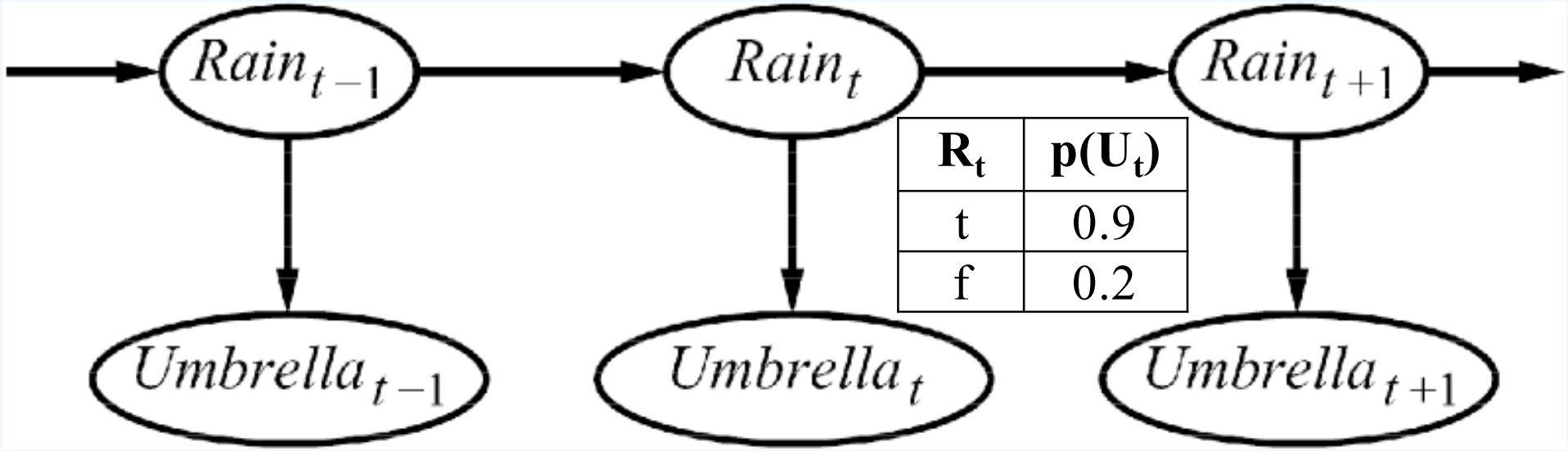
- Markov chains not so useful for most agents
  - Eventually you don't know anything anymore
  - Need observations to update your beliefs

- Hidden Markov models (HMMs)
  - Underlying Markov chain over state X
  - You observe outputs (effects) at each time step
  - As a Bayes' net:



Example

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

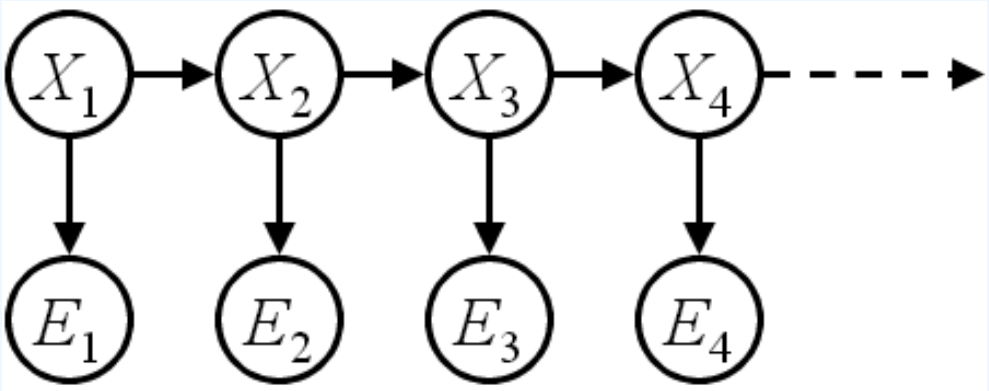


R_t	$p(U_t)$
t	0.9
f	0.2

- An HMM is defined by:
  - Initial distribution: $p(X_1)$
  - Transitions: $p(X|X_{-1})$
  - Emissions: $p(E|X)$

Conditional Independence

- HMMs have two important independence properties:
  - Markov hidden process, future depends on past via the present
  - Current observation independent of all else given current state



- Quiz: does this mean that observations are independent?
  - [No, correlated by the hidden state]

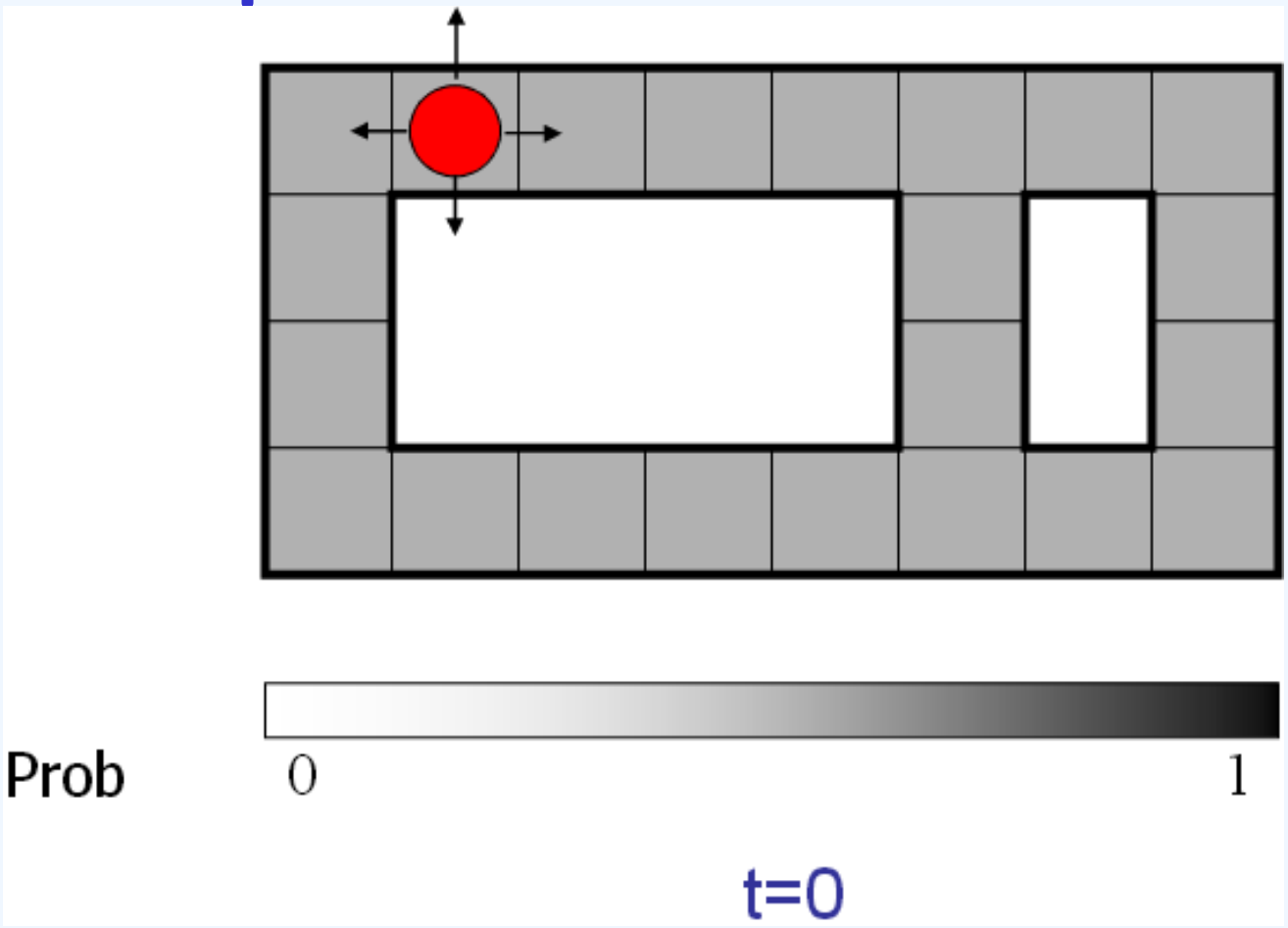
Real HMM Examples

- Speech recognition HMMs:
  - Observations are acoustic signals (continuous values)
  - States are specific positions in specific words (so, tens of thousands)
- Robot tracking:
  - Observations are range readings (continuous)
  - States are positions on a map (continuous)

Filtering / Monitoring

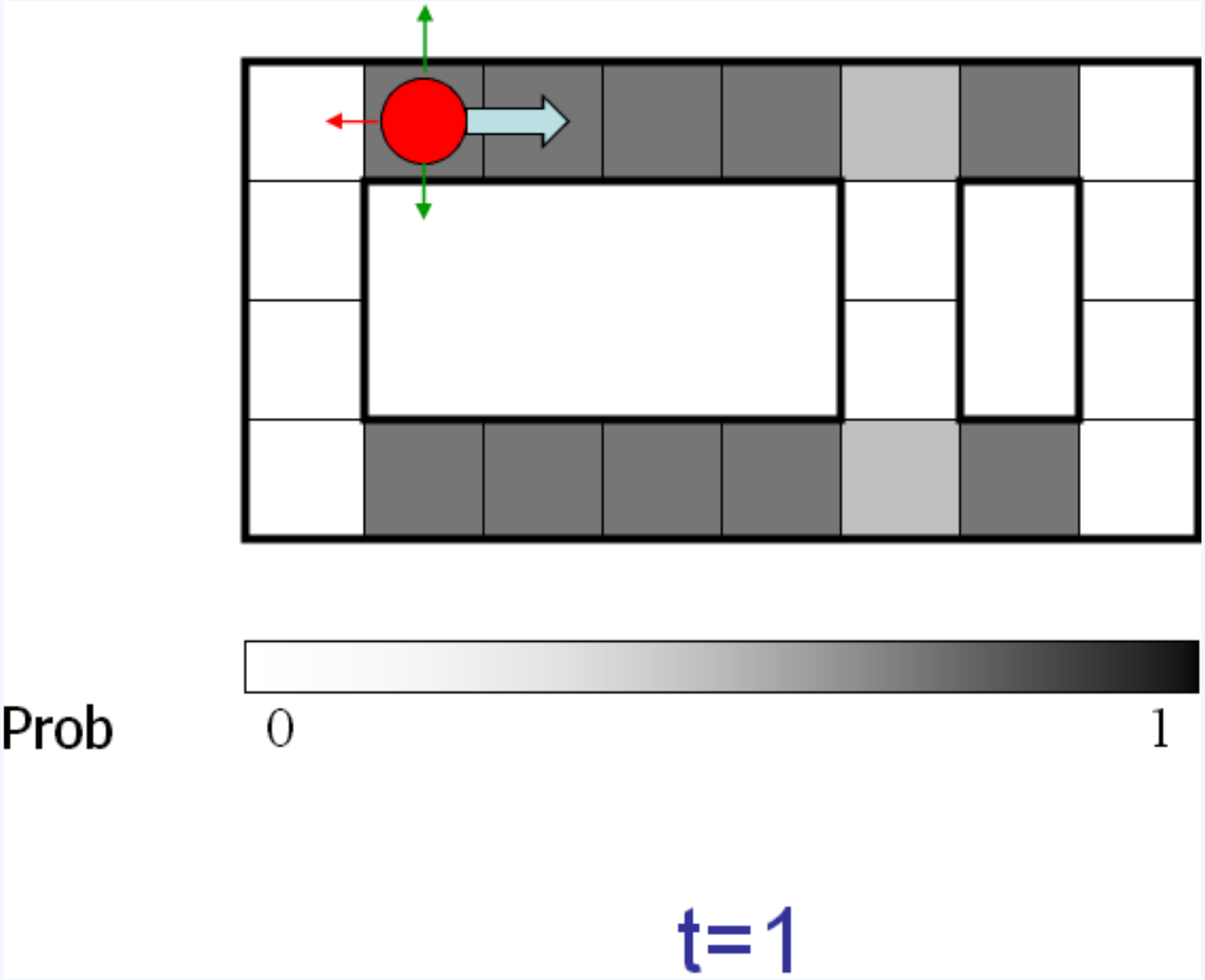
- Filtering, or monitoring, is the task of tracking the distribution B(X) (the belief state) over time
- We start with B(X) in an initial setting, usually uniform
- As time passes, or we get observations, we update B(X)
- The Kalman filter was invented in the 60's and first implemented as a method of trajectory estimation for the Apollo program

Example: Robot Localization

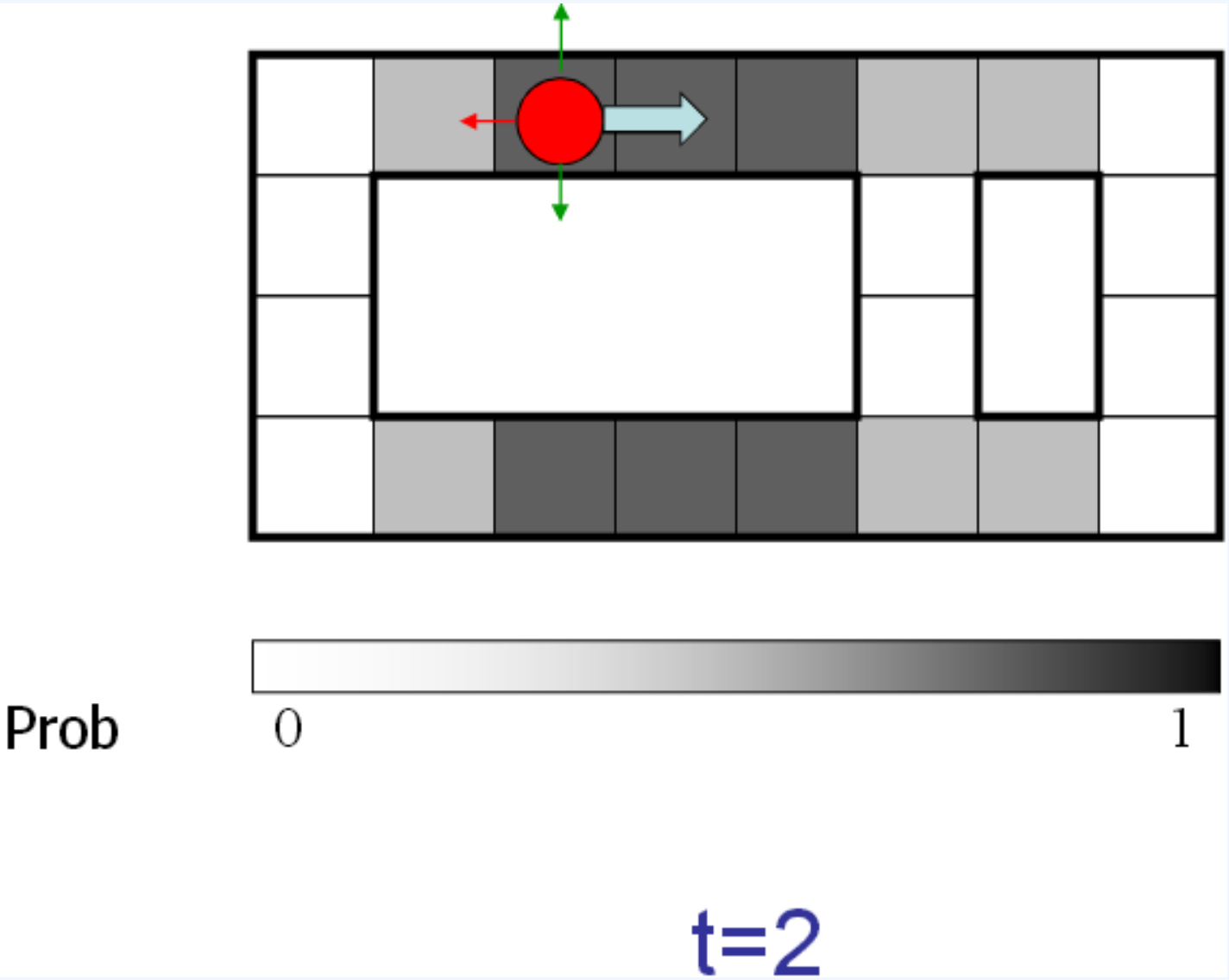


May not execute action with small prob.

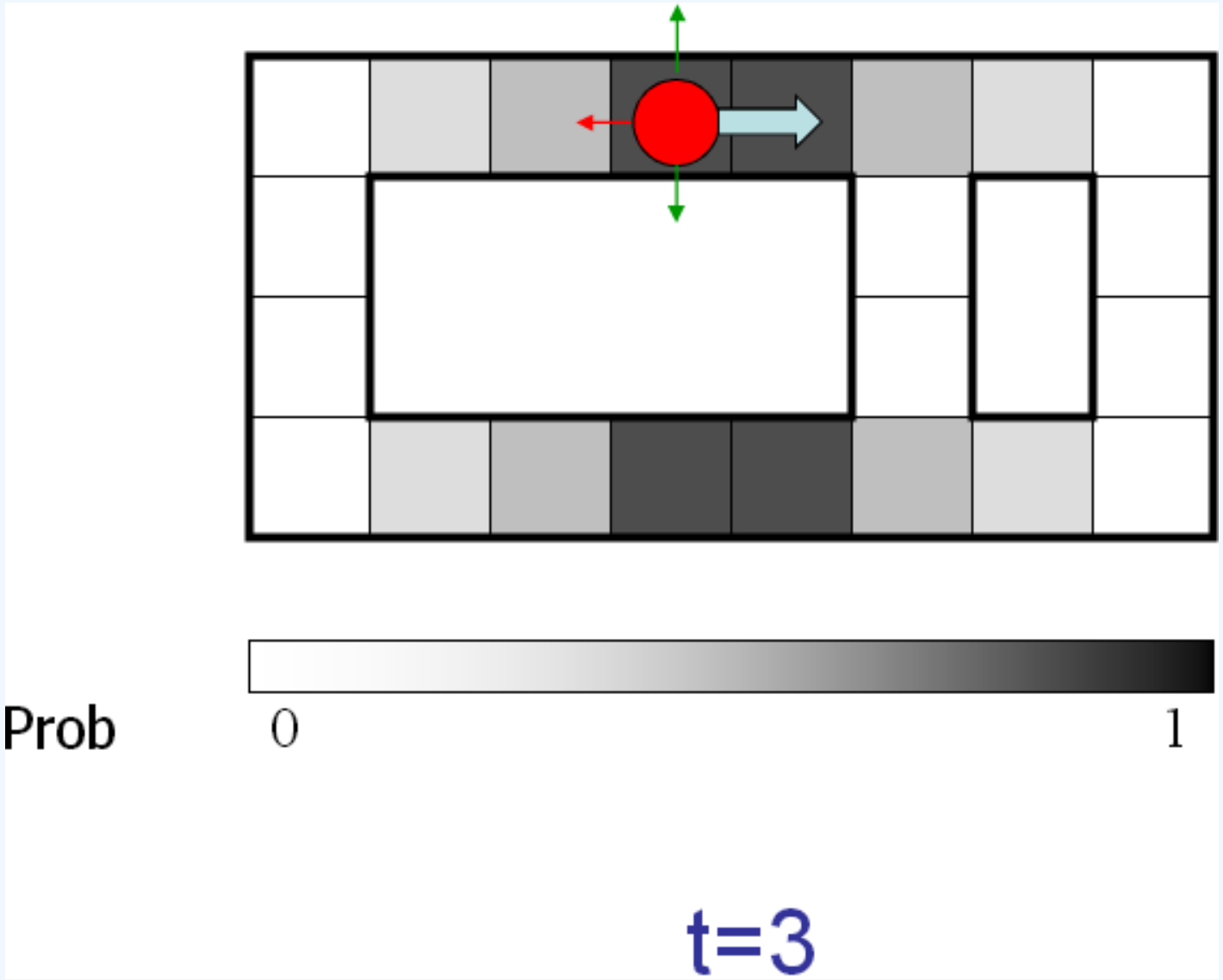
Example: Robot Localization



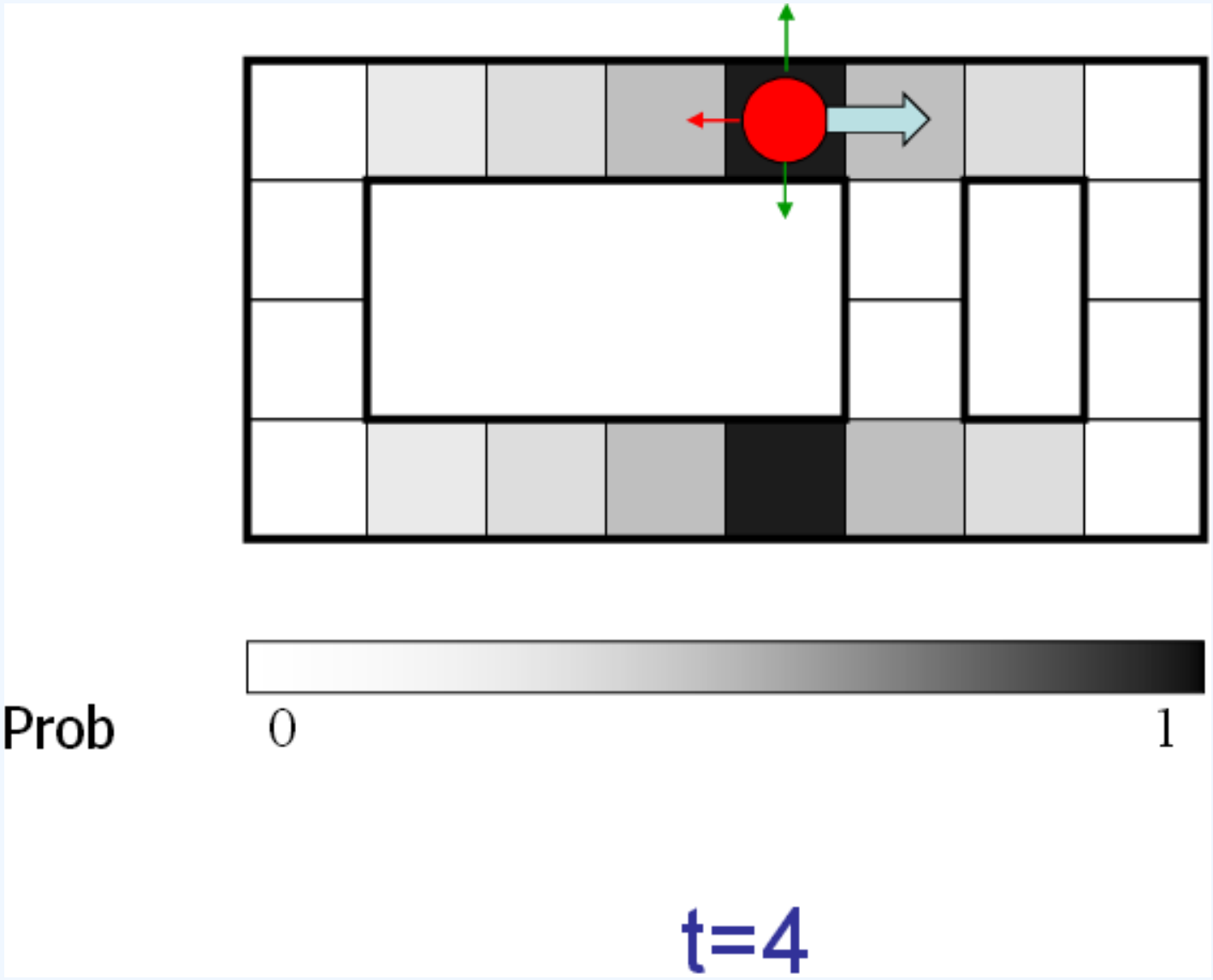
Example: Robot Localization



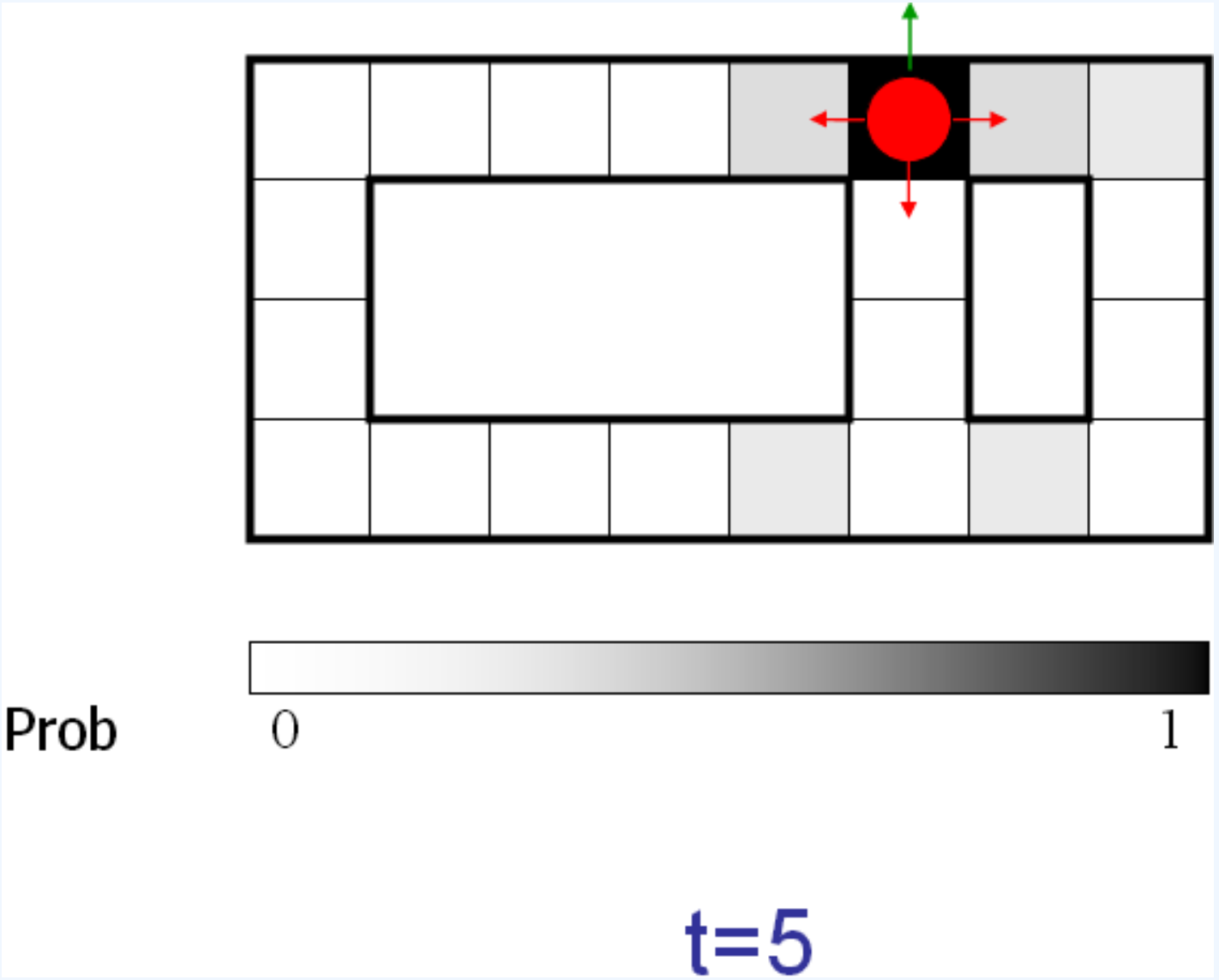
Example: Robot Localization



Example: Robot Localization

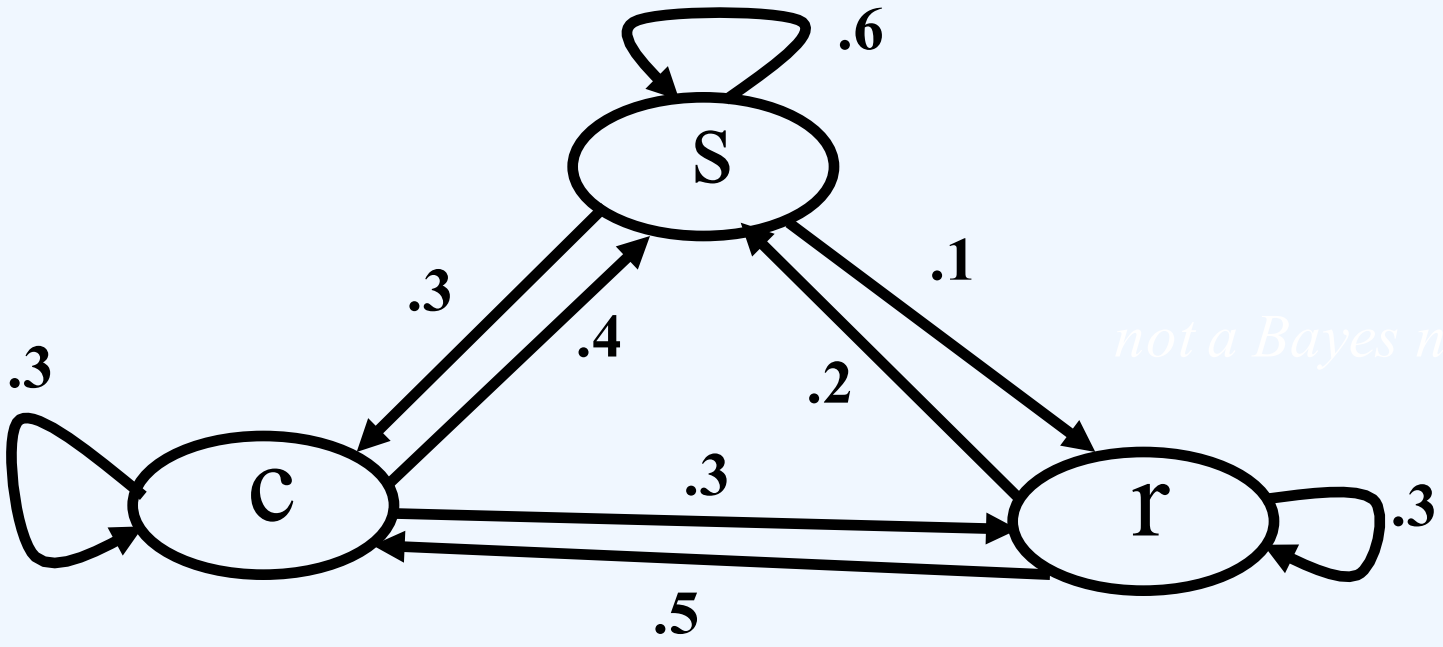


Example: Robot Localization



Another weather example

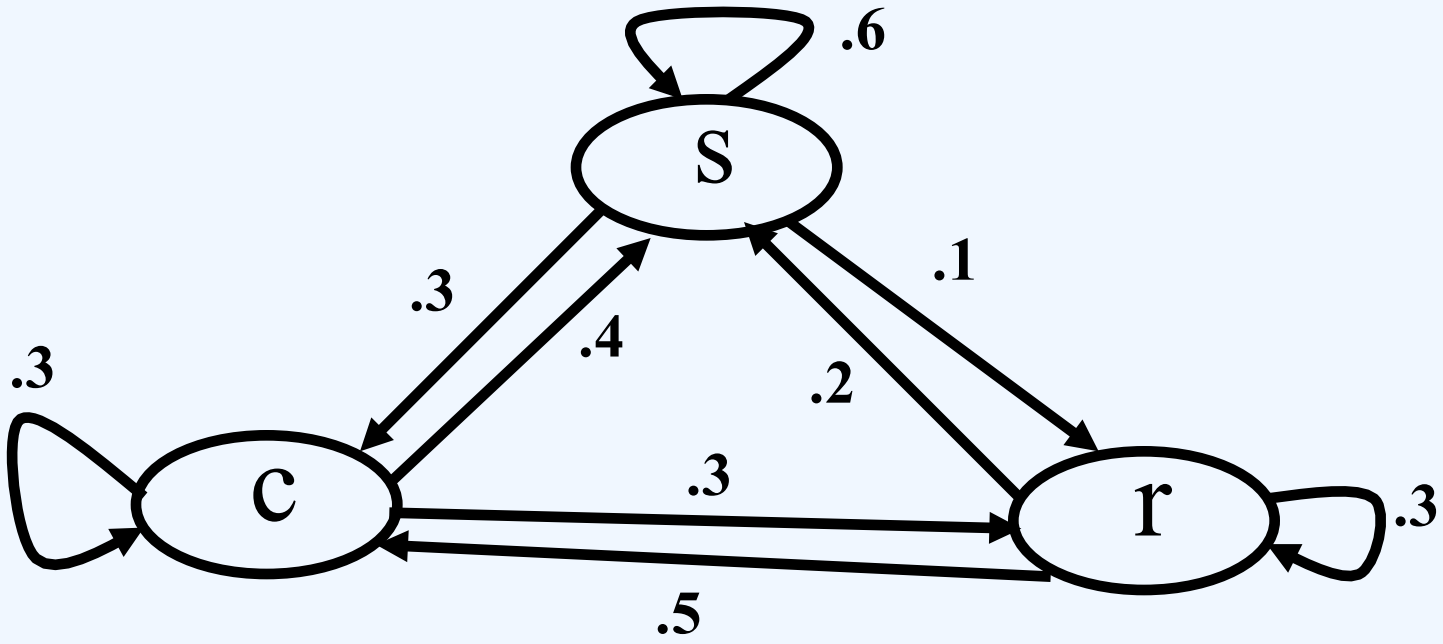
- X_t is one of {s, c, r} (sun, cloudy, rain)
- Transition probabilities:



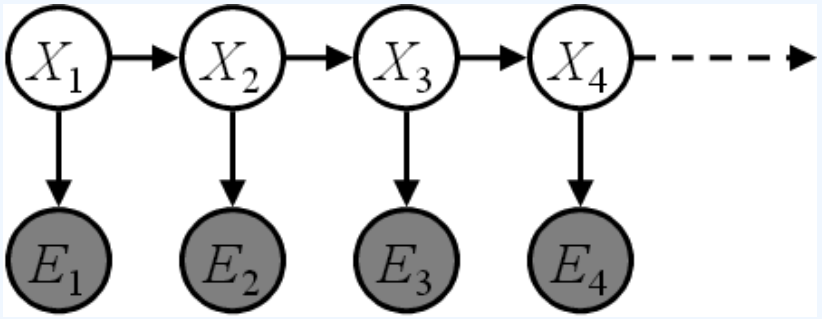
- Throughout, assume uniform distribution over X_1

Weather example extended to HMM

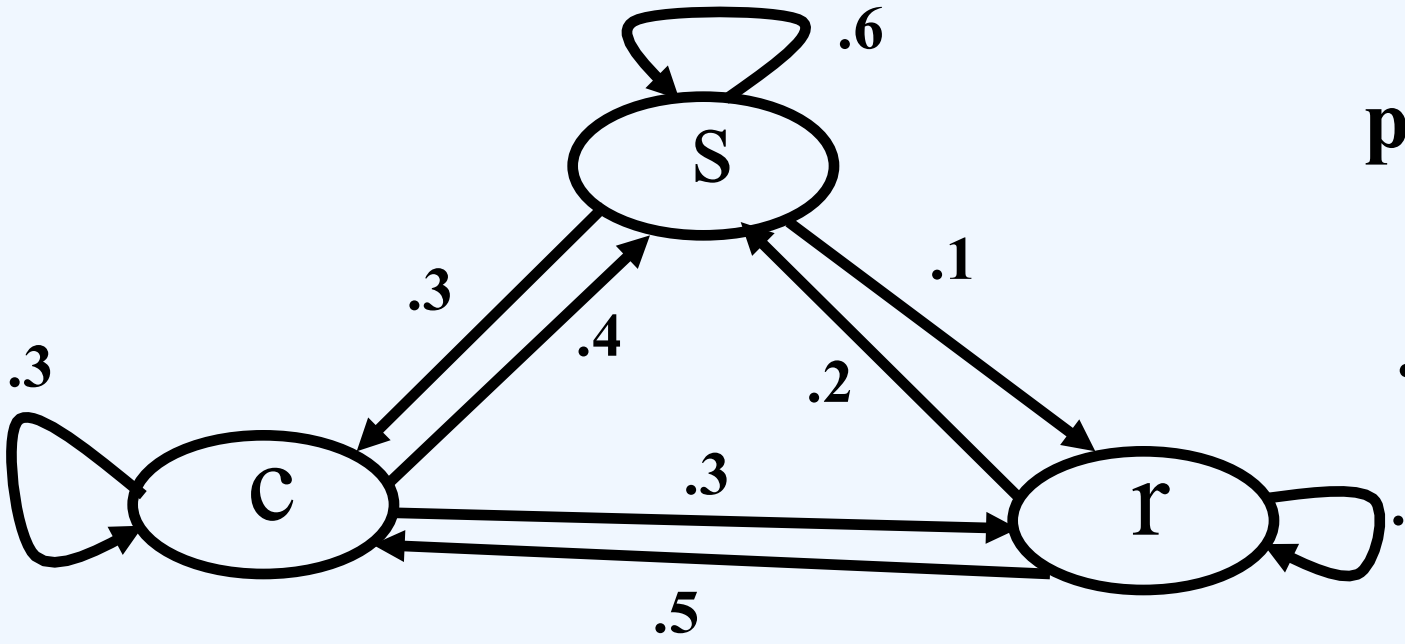
- Transition probabilities:



- Observation: roommate wet or dry
- p(w|s) = .1, p(w|c) = .3, p(w|r) = .8



HMM weather example: a question

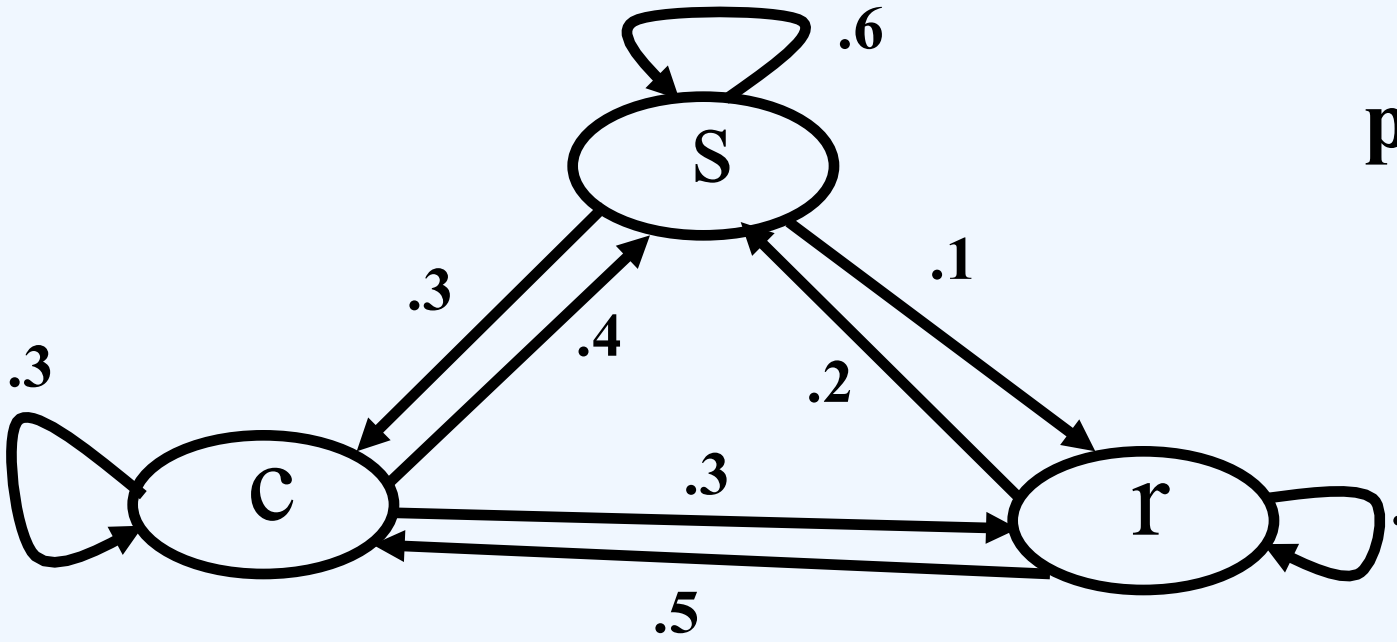


p(w|s) = .1

p(w|c) = .3 p(w|r) = .8

- You have been stuck in the dorm for three days (!)
- On those days, your roommate was dry, wet, wet, respectively
- What is the probability that it is now raining outside?
- $p(X_3 = r \mid E_1 = d, E_2 = w, E_3 = w)$
- By Bayes' rule, really want to know $p(X_3, E_1 = d, E_2 = w, E_3 = w)$

Solving the question

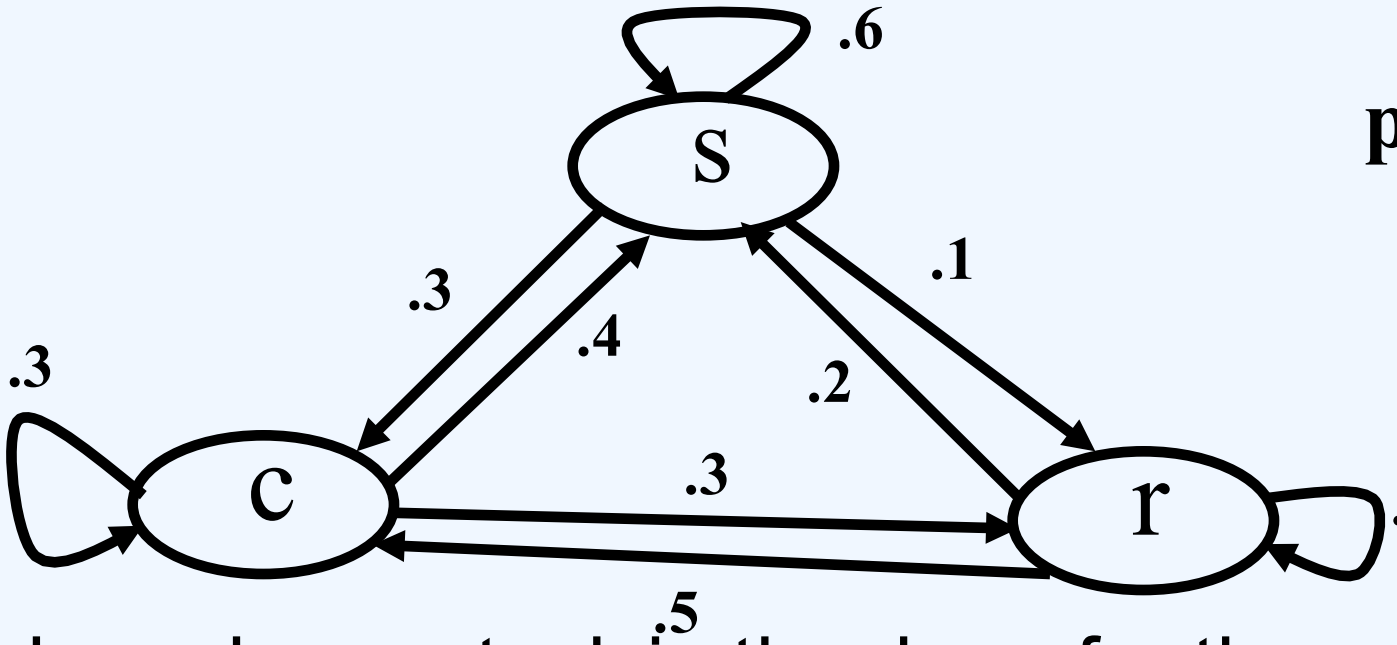


p(w|s) = .1

p(w|c) = .3 p(w|r) = .8

- Computationally efficient approach: first compute $p(X_1 = i, E_1 = d)$ for all states i
- General case: solve for $p(X_t, E_1 = e_1, \ldots, E_t = e_t)$ for t=1, then t=2, … This is called monitoring
- $p(X_t, E_1 = e_1, \ldots, E_t = e_t) = \Sigma_{X_{t-1}} p(X_{t-1} = x_{t-1}, E_1 = e_1, \ldots, E_{t-1} = e_{t-1}) P(X_t \mid X_{t-1} = x_{t-1}) P(E_t = e_t \mid X_t)$

Predicting further out

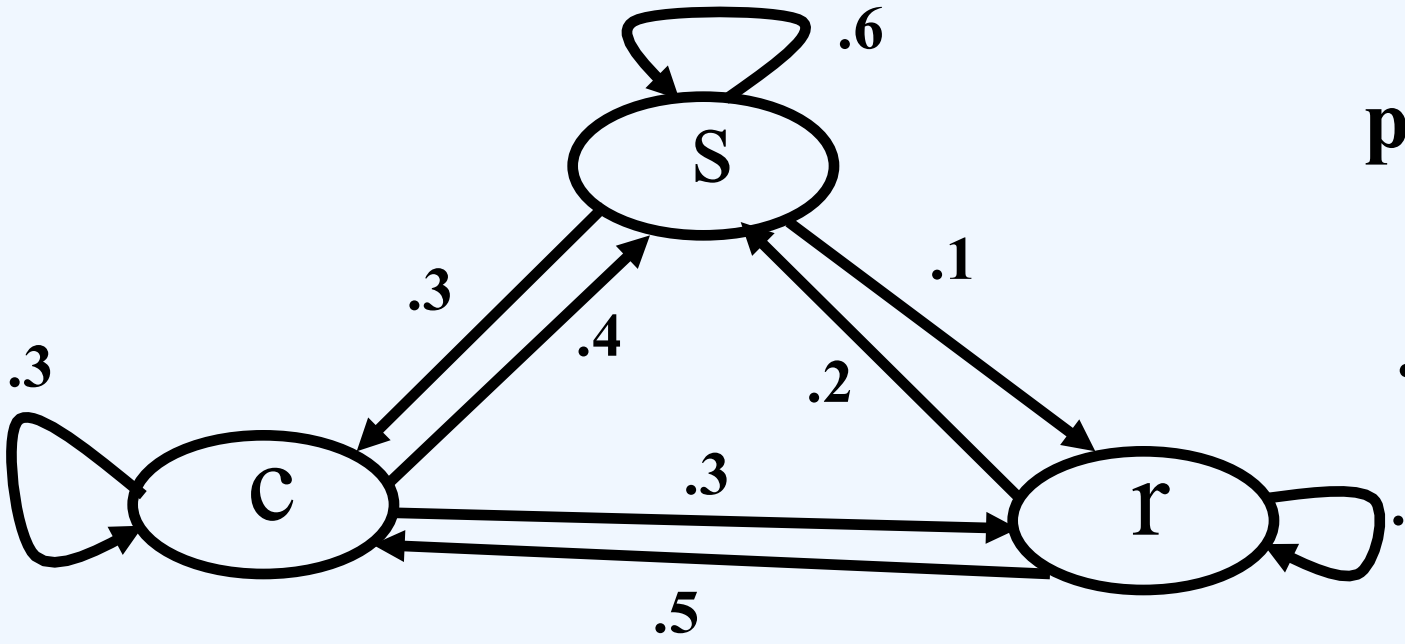


p(w|s) = .1

p(w|c) = .3 p(w|r) = .8

- You have been stuck in the dorm for three days
- On those days, your roommate was dry, wet, wet, respectively
- What is the probability that **two days from now** it will be raining outside?
- $p(X_5 = r \mid E_1 = d, E_2 = w, E_3 = w)$

Predicting further out, continued…

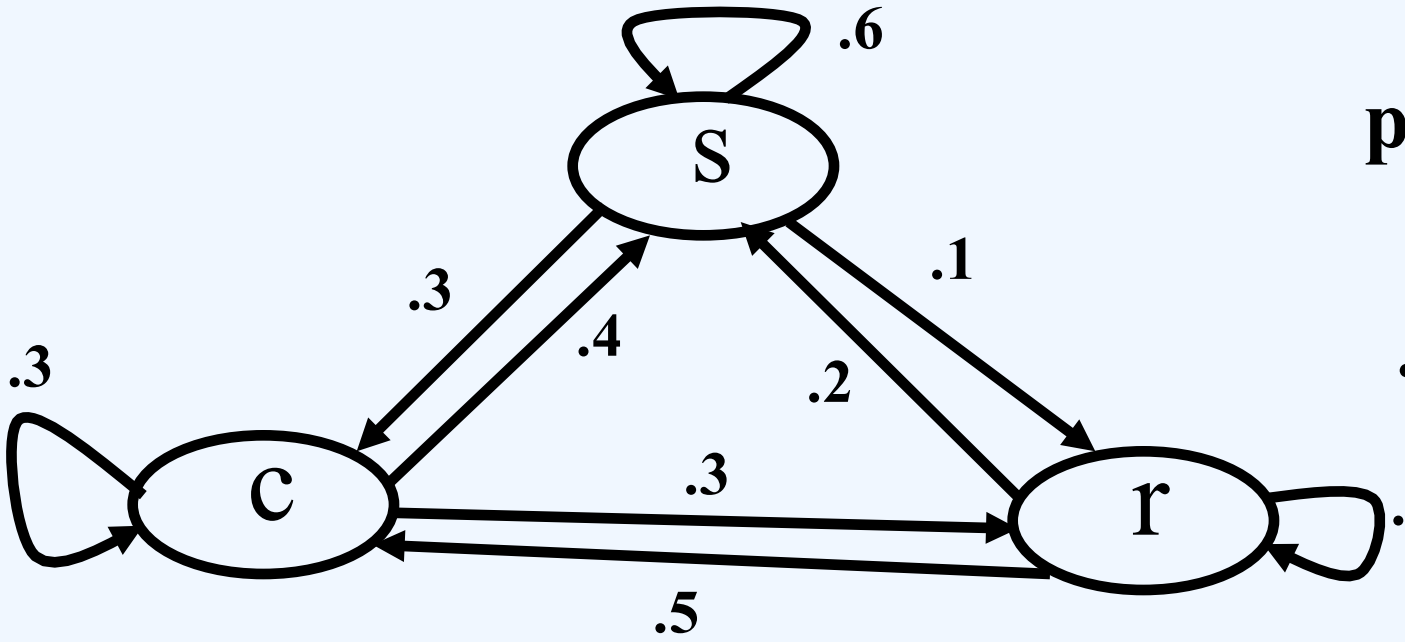


p(w|s) = .1

p(w|c) = .3 p(w|r) = .8

- Want to know: $p(X_5 = r \mid E_1 = d, E_2 = w, E_3 = w)$
- Already know how to get: $p(X_3 \mid E_1 = d, E_2 = w, E_3 = w)$
- $p(X_4 = r \mid E_1 = d, E_2 = w, E_3 = w) =$
  $\Sigma_{X_3} \; P(X_4 = r, X_3 = x_3 \mid E_1 = d, E_2 = w, E_3 = w)$
  $= \Sigma_{X_3} \; P(X_4 = r \mid X_3 = x_3) P(X_3 = x_3 \mid E_1 = d, E_2 = w, E_3 = w)$
- Etc. for X_5
- So: monitoring first, then straightforward Markov process updates

Integrating newer information



p(w|s) = .1

p(w|c) = .3 p(w|r) = .8

- You have been stuck in the dorm for **four** days (!)
- On those days, your roommate was dry, wet, wet, dry respectively
- What is the probability that **two days ago** it was raining outside? $p(X_2 = r \mid E_1 = d, E_2 = w, E_3 = w, E_4 = d)$
  - Smoothing or hindsight problem