CSE P573: Artificial Intelligence
Spring 2014

Hidden Markov Models

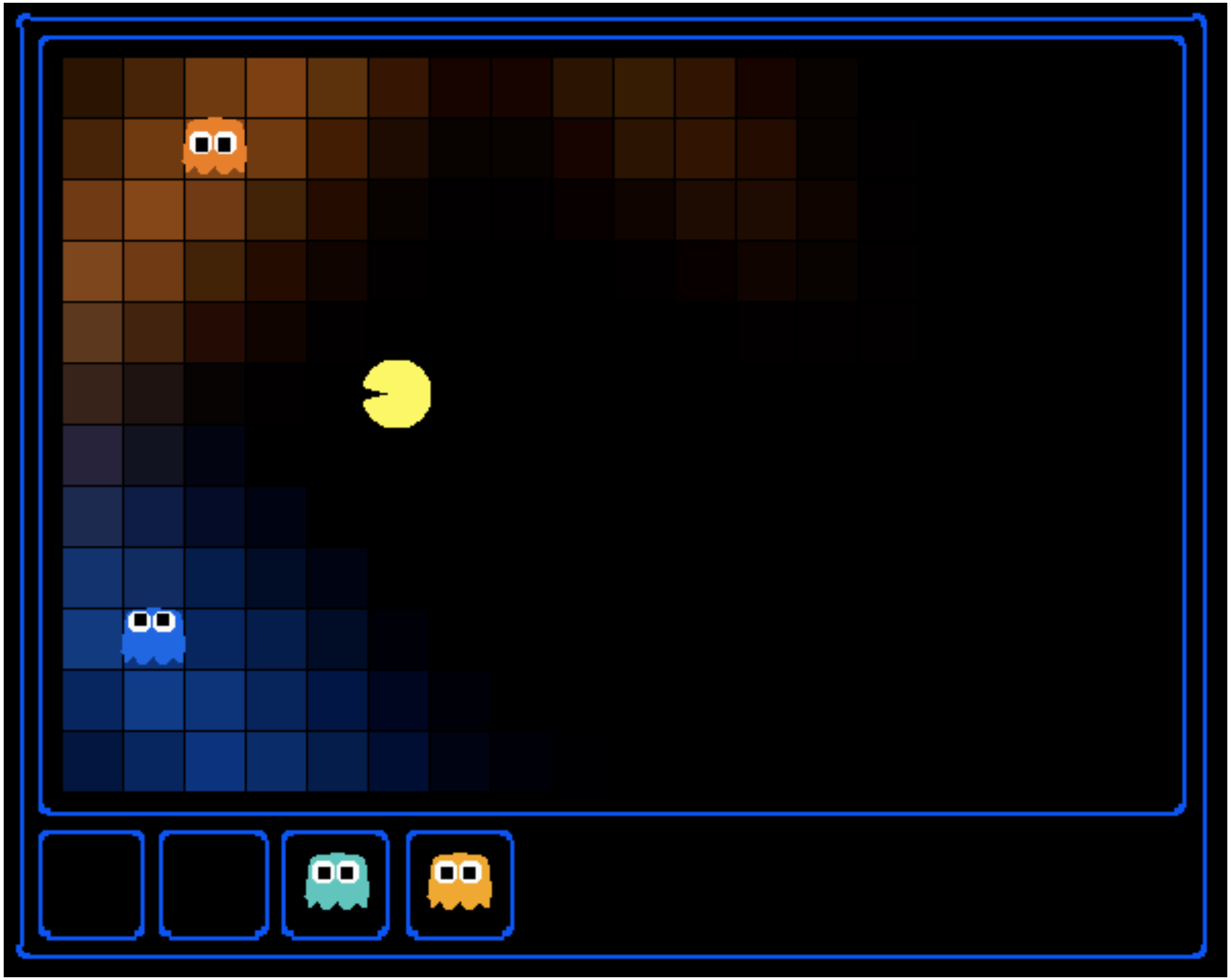
Ali Farhadi

Many slides adapted from Dan Weld, Pieter Abbeel, Dan Klein, Stuart Russell, Andrew Moore & Luke Zettlemoyer

Outline

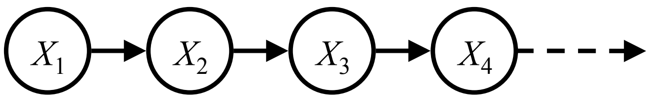
- Probabilistic sequence models (and inference)
  - Probability and Uncertainty – Preview
  - Markov Chains
  - Hidden Markov Models
  - Exact Inference
  - Particle Filters

Going Hunting



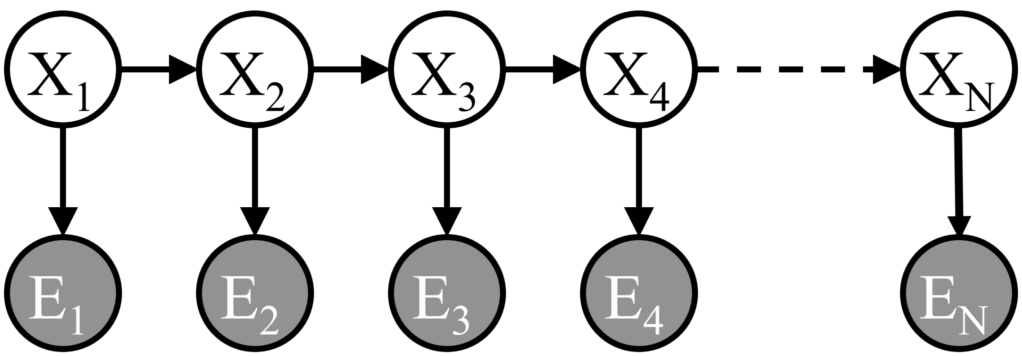
Hidden Markov Models

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

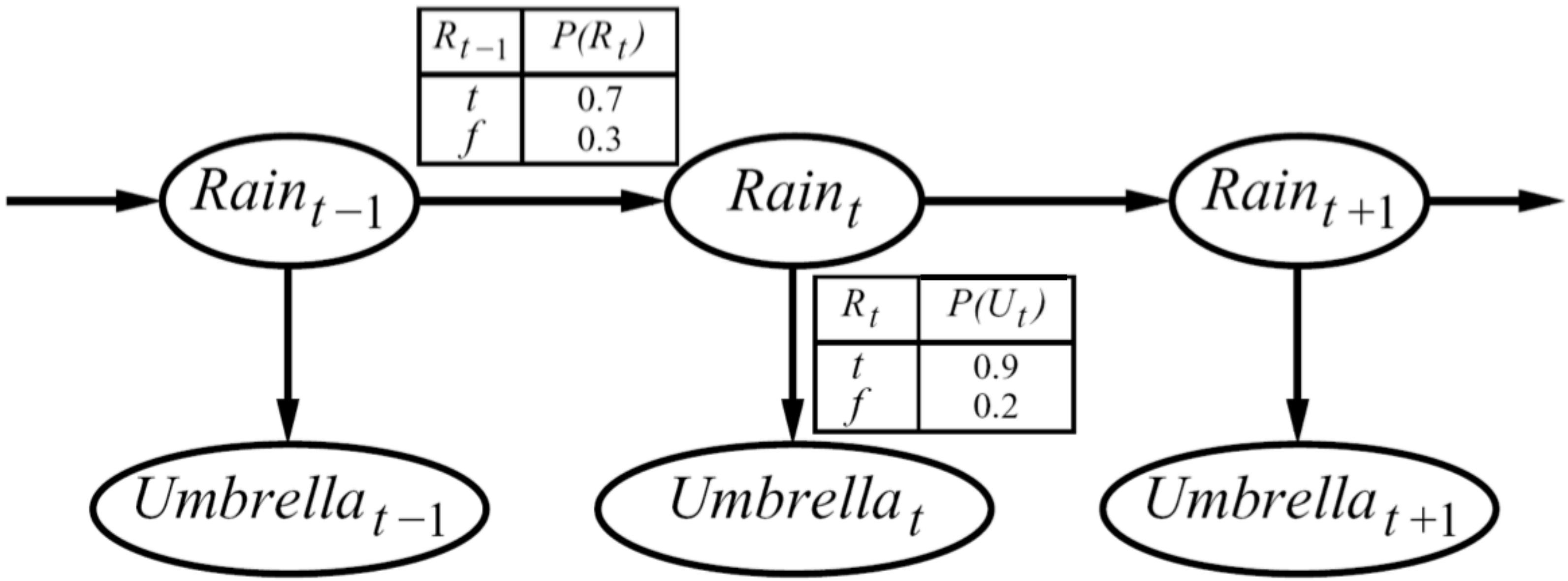


$P(X_1)$ $\qquad$ $P(X|X_{-1})$

- Hidden Markov models (HMMs)
  - Underlying Markov chain over states S
  - You observe outputs (effects) at each time step



Example: Weather HMM



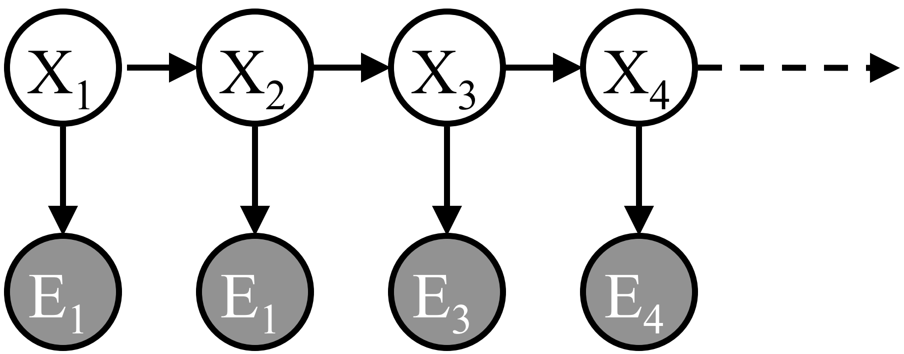
- An HMM is defined by:
  - Initial distribution: $P(X_1)$
  - Transitions: $P(X_t|X_{t-1})$
  - Emissions: $P(E|X)$

Ghostbusters HMM

- $P(X_1)$ = uniform
- $P(X'|X)$ = usually move clockwise, but sometimes move in a random direction or stay in place
- $P(E|X)$ = same sensor model as before: red means close, green means far away.

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

$P(X_1)$



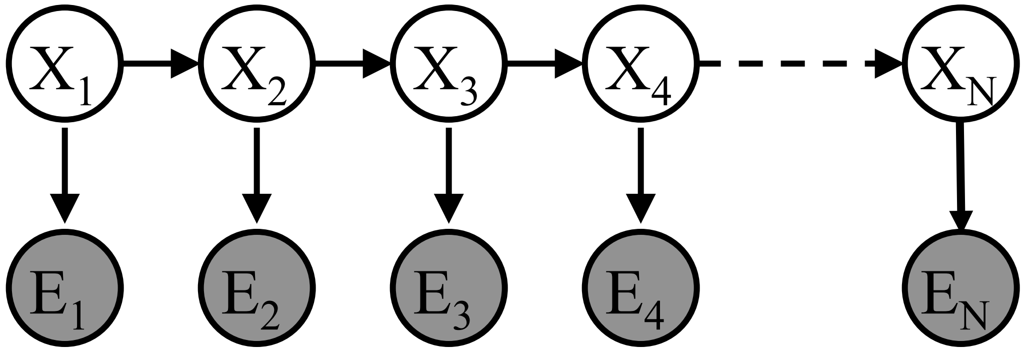
1/6	1/6	1/2
0	1/6	0
0	0	0

$P(X'|X=\langle 1,2\rangle)$

P(E|X)

P(red | 3)	P(orange | 3)	P(yellow | 3)	P(green | 3)
0.05	0.15	0.5	0.3

Hidden Markov Models



- Defines a joint probability distribution:

$$P(X_1, E_1, X_2, E_2, X_3, E_3) = P(X_1)P(E_1|X_1)P(X_2|X_1)P(E_2|X_2)P(X_3|X_2)P(E_3|X_3)$$

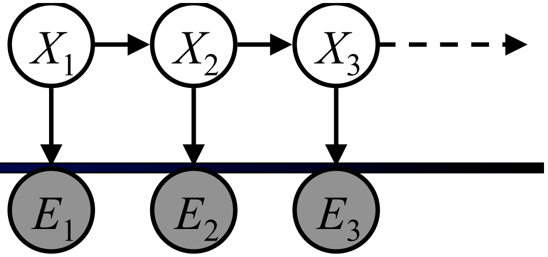
$$P(X_1, \ldots, X_n, E_1, \ldots, E_n) =$$

$$P(X_{1:n}, E_{1:n}) =$$

$$P(X_1)P(E_1|X_1)\prod_{t=2}^{N} P(X_t|X_{t-1})P(E_t|X_t)$$

- Questions to be resolved:
  - Does this indeed define a joint distribution?
  - Can every joint distribution be factored this way, or are we making some assumptions about the joint distribution by using this factorization?

Chain Rule and HMMs



- From the chain rule, *every* joint distribution over $X_1, E_1, X_2, E_2, X_3, E_3$ can be written as:

$$\begin{aligned} P(X_1, E_1, X_2, E_2, X_3, E_3) = & P(X_1)P(E_1|X_1)P(X_2|X_1, E_1)P(E_2|X_1, E_1, X_2) \\ & P(X_3|X_1, E_1, X_2, E_2)P(E_3|X_1, E_1, X_2, E_2, X_3) \end{aligned}$$

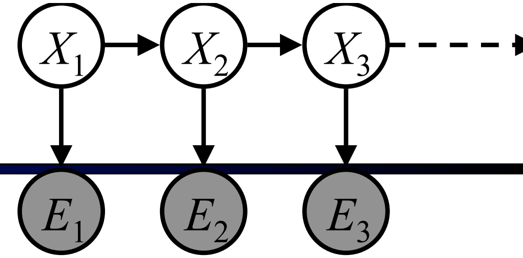
- *Assuming* that

$$X_2 \perp\!\!\!\perp E_1 \mid X_1, \quad E_2 \perp\!\!\!\perp X_1, E_1 \mid X_2, \quad X_3 \perp\!\!\!\perp X_1, E_1, E_2 \mid X_2, \quad E_3 \perp\!\!\!\perp X_1, E_1, X_2, E_2 \mid X_3$$

gives us the expression posited on the previous slide:

$$P(X_1, E_1, X_2, E_2, X_3, E_3) = P(X_1)P(E_1|X_1)P(X_2|X_1)P(E_2|X_2)P(X_3|X_2)P(E_3|X_3)$$

Chain Rule and HMMs



- From the chain rule, *every* joint distribution over $X_1, E_1, \ldots, X_T, E_T$ can be written as:

$$P(X_1, E_1, \ldots, X_T, E_T) = P(X_1)P(E_1|X_1)\prod_{t=2}^{T} P(X_t|X_1, E_1, \ldots, X_{t-1}, E_{t-1})P(E_t|X_1, E_1, \ldots, X_{t-1}, E_{t-1}, X_t)$$

- *Assuming* that for all *t*:
  - State independent of all past states and all past evidence given the previous state, i.e.:

$$X_t \perp\!\!\!\perp X_1, E_1, \ldots, X_{t-2}, E_{t-2}, E_{t-1} \mid X_{t-1}$$

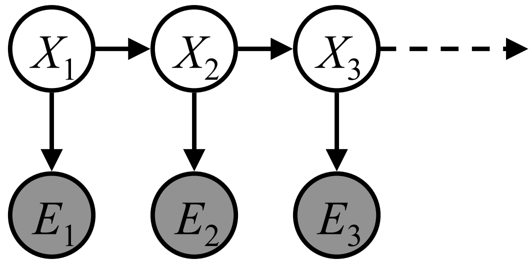
  - Evidence is independent of all past states and all past evidence given the current state, i.e.:

$$E_t \perp\!\!\!\perp X_1, E_1, \ldots, X_{t-2}, E_{t-2}, X_{t-1}, E_{t-1} \mid X_t$$

gives us the expression posited on the earlier slide:

$$P(X_1, E_1, \ldots, X_T, E_T) = P(X_1)P(E_1|X_1)\prod_{t=2}^{T} P(X_t|X_{t-1})P(E_t|X_t)$$

Implied Conditional Independencies



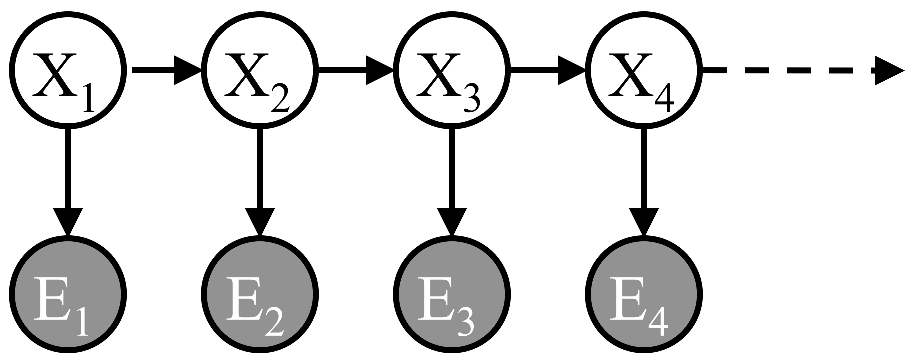
- Many implied conditional independencies, e.g.,

$$E_1 \perp\!\!\!\perp X_2, E_2, X_3, E_3 \mid X_1$$

- To prove them
  - Approach 1: follow similar (algebraic) approach to what we did in the Markov models lecture
  - Approach 2: directly from the graph structure (3 lectures from now)
    - Intuition: If path between U and V goes through W, then $U \perp\!\!\!\perp V \mid W$ [Some fineprint later]

Conditional Independence

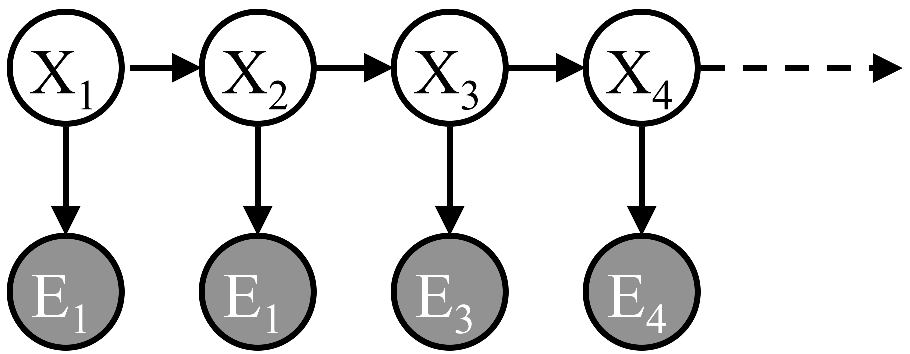
- HMMs have two important independence properties:
  - Markov hidden process, future depends on past via the present
  - Current observation independent of all else given current state



- Quiz: Are observations E1, E2 independent?
  - [No, correlated by the hidden state]

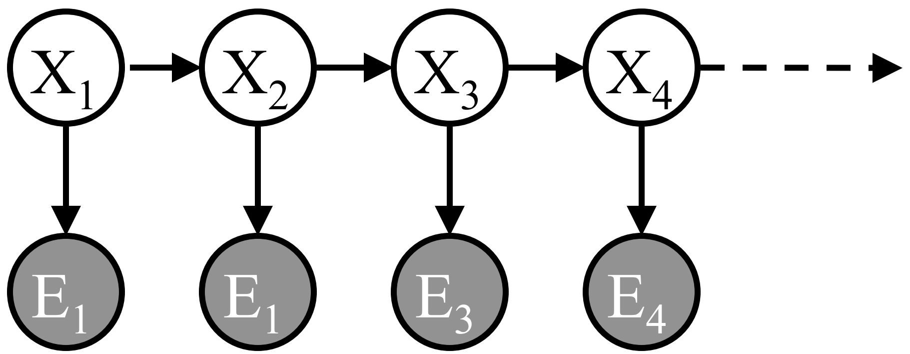
Real HMM Examples

- Speech recognition HMMs:
  - Observations are acoustic signals (continuous valued)
  - States are specific positions in specific words (so, tens of thousands)



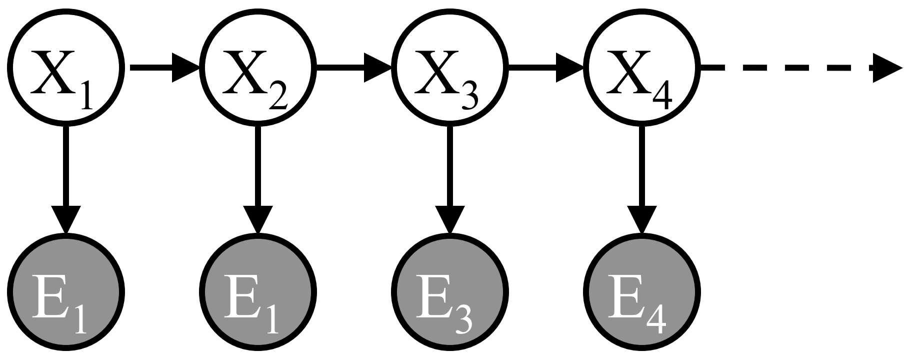
Real HMM Examples

- Machine translation HMMs:
  - Observations are words (tens of thousands)
  - States are translation options



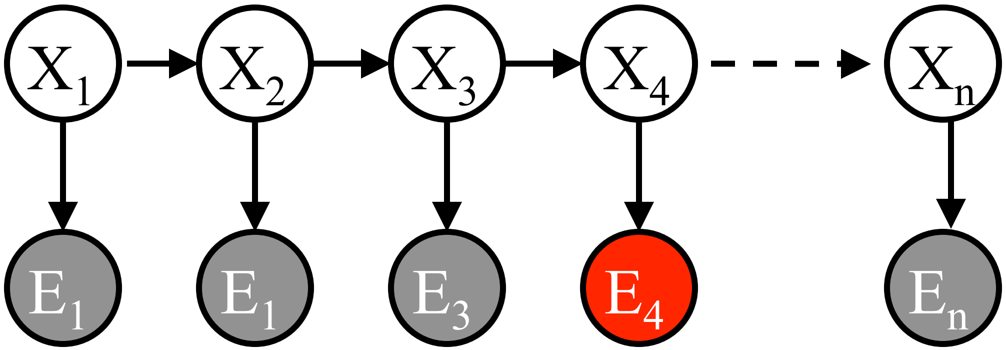
Real HMM Examples

- Robot tracking:
  - Observations are range readings (continuous)
  - States are positions on a map (continuous)



HMM Computations

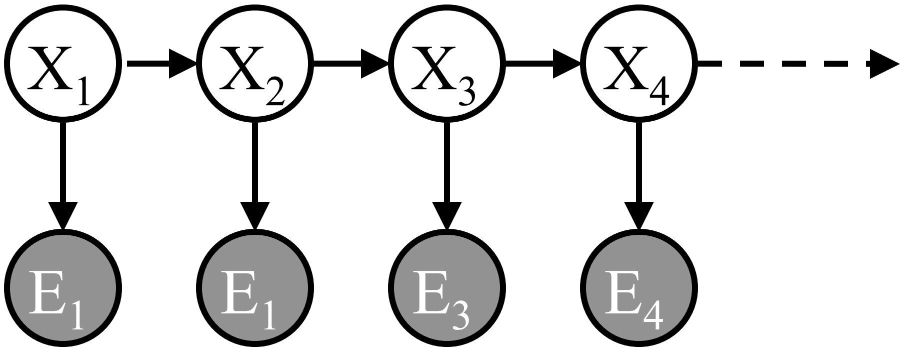
- Given
  - joint $P(X_{1:n}, E_{1:n})$
  - evidence $E_{1:n} = e_{1:n}$



- Inference problems include:
  - Filtering, find $P(X_t | e_{1:t})$ for current t
  - Smoothing, find $P(X_t | e_{1:n})$ for past t

HMM Computations

- Given
  - joint $P(X_{1:n}, E_{1:n})$
  - evidence $E_{1:n} = e_{1:n}$



- Inference problems include:
  - Filtering, find $P(X_t|e_{1:t})$ for current t
  - Smoothing, find $P(X_t|e_{1:n})$ for past t
  - Most probable explanation, find

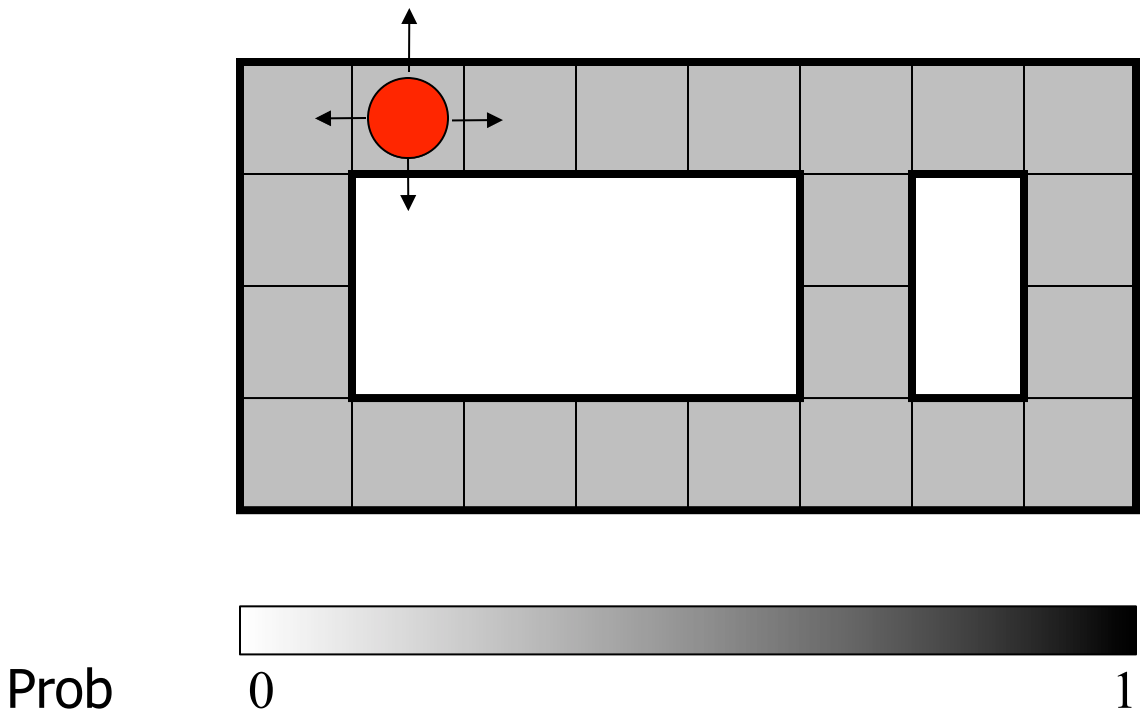
$$x^*_{1:n} = \mathrm{argmax}_{x_{1:n}} P(x_{1:n}|e_{1:n})$$

Filtering / Monitoring

- Filtering, or monitoring, is the task of tracking the distribution $B(X)=P(X_t|e_{1:t})$ (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

Example from
Michael Pfeiffer

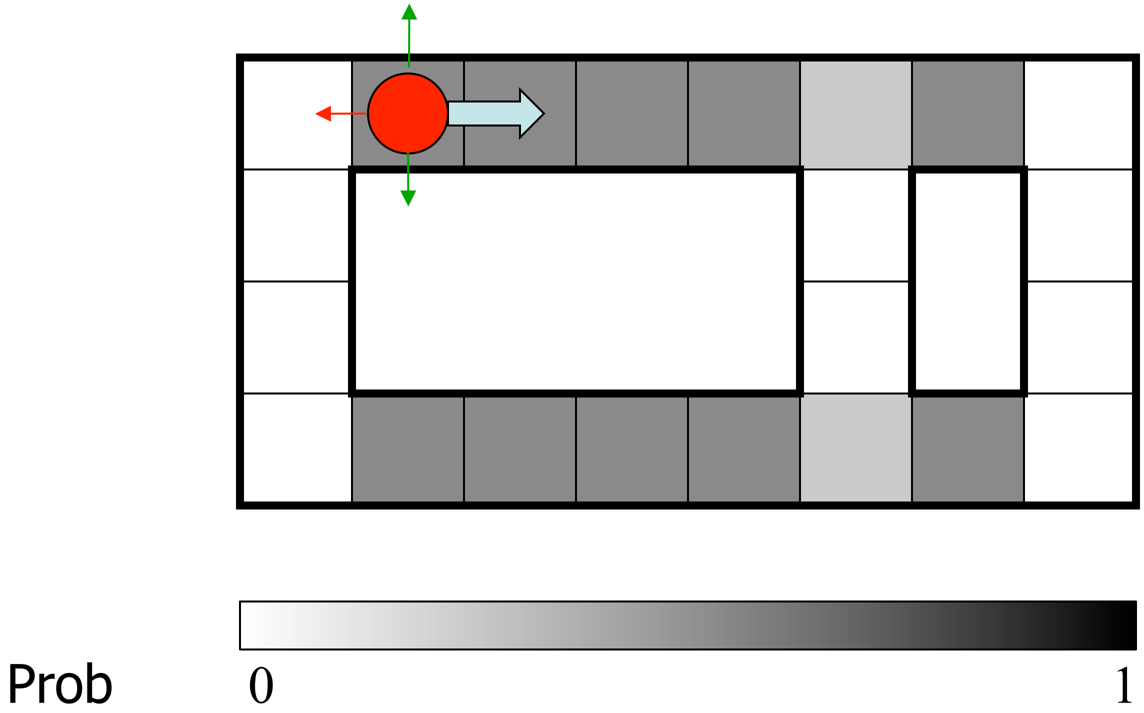


t=0

Sensor model: never more than 1 mistake

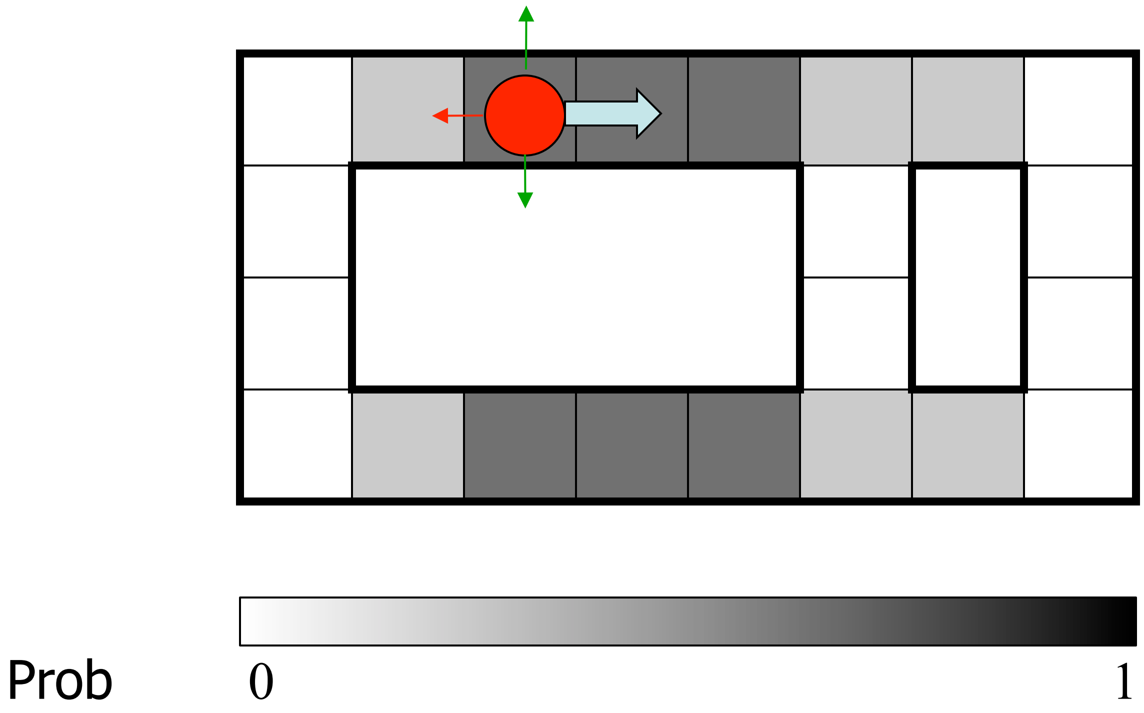
Motion model: may not execute action with small prob.

Example: Robot Localization



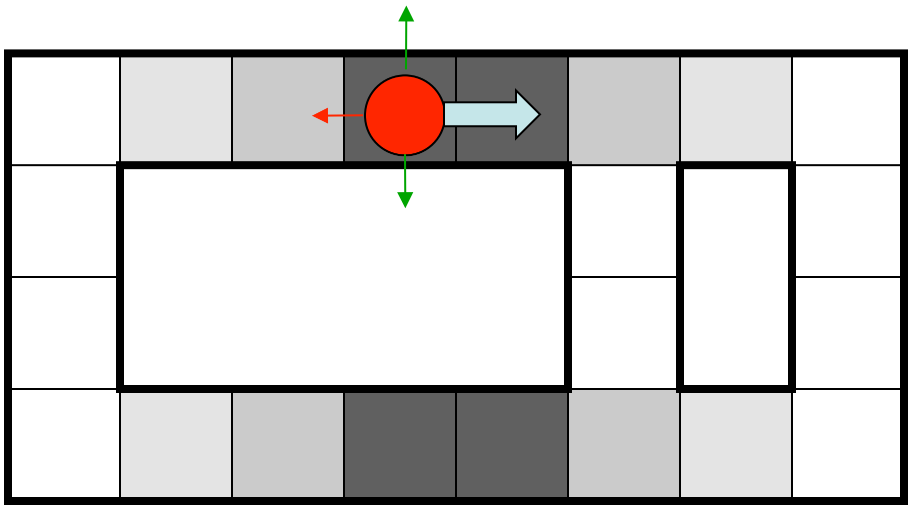
t=1

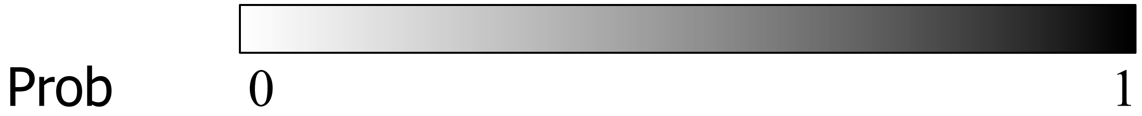
Example: Robot Localization



t=2

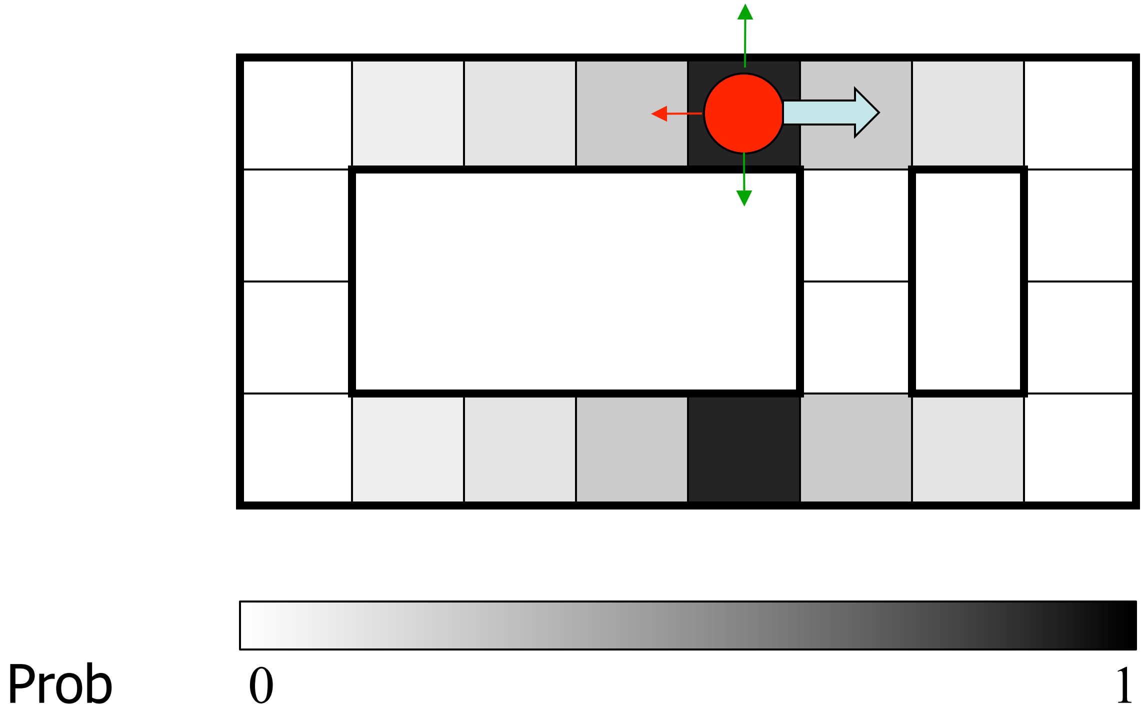
Example: Robot Localization





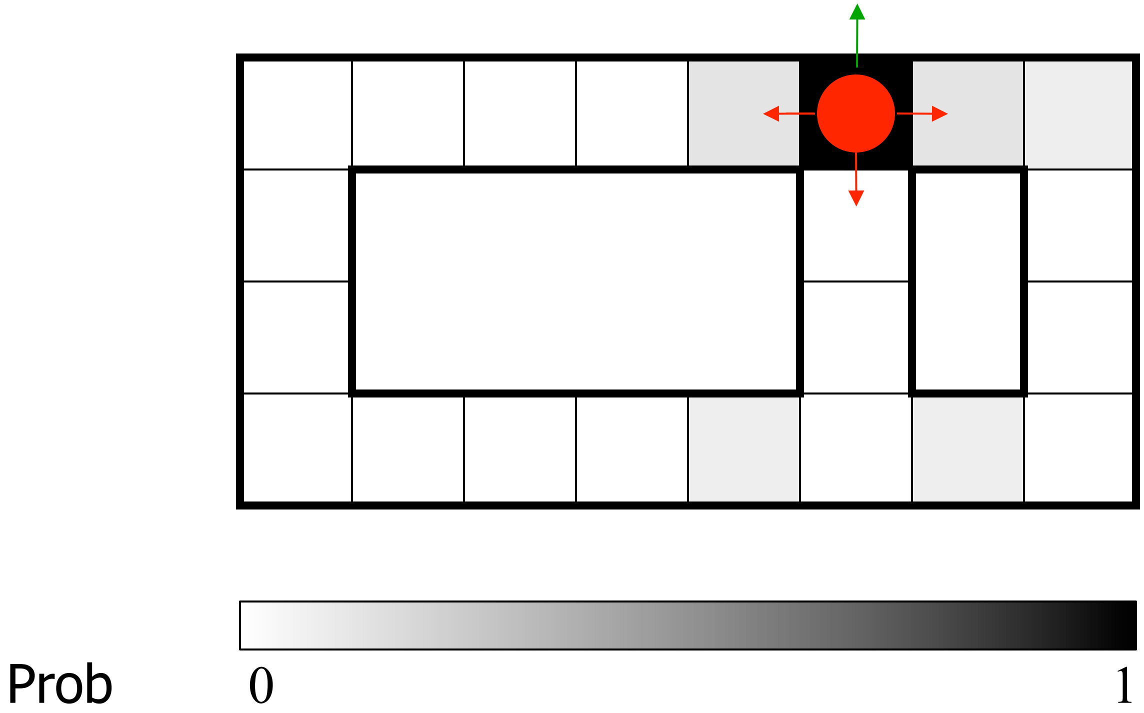
t=3

Example: Robot Localization



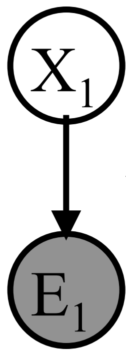
t=4

Example: Robot Localization



t=5

Inference : Simple Cases



$P(X_1)$ $\quad P(X_t|X_{t-1})$

$P(E|X)$

$$P(X_1|e_1)$$

$$P(x_1|e_1) = P(x_1, e_1)/P(e_1)$$

$$\propto_{X_1} P(x_1, e_1)$$

$$= P(x_1)P(e_1|x_1)$$