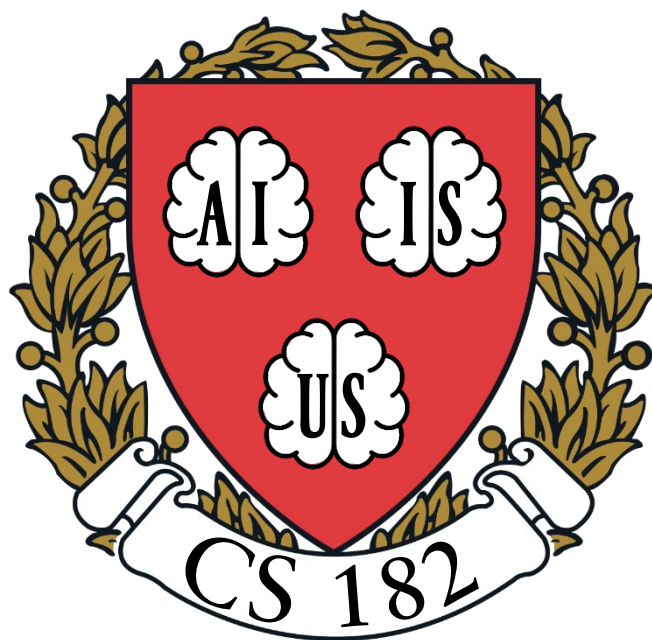




Fall 2021 | Lecture 13

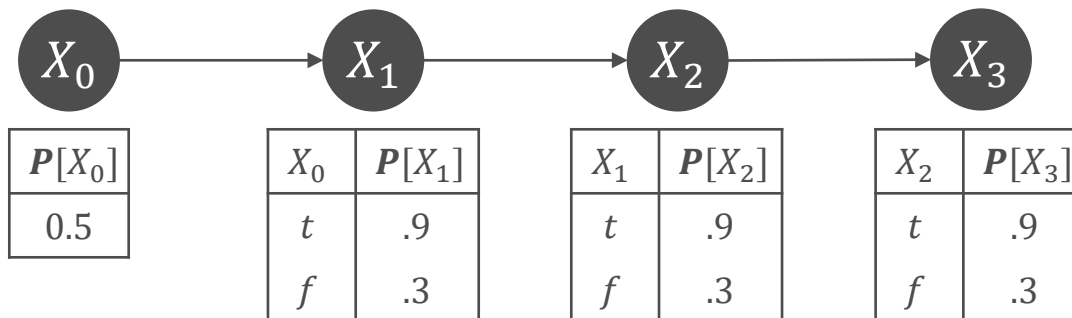
Hidden Markov Models

Ariel Procaccia | Harvard University

MARKOV CHAINS



Suppose the weather evolves according to the following Bayes net (with $t = \text{sun}$):



ASSUMPTIONS

- We will think of infinite processes described by random variables $X_0, X_1, \dots$
- We use $\mathbf{X}_{0:t}$ to denote $X_0, \dots, X_t$
- Our simple Bayes net is assumed to satisfy:

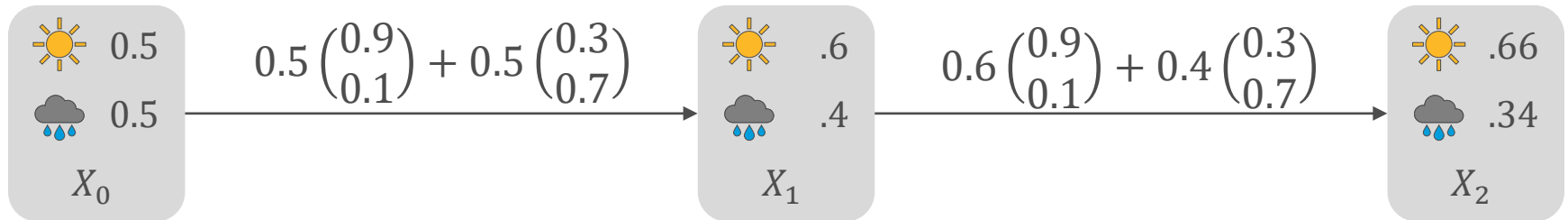
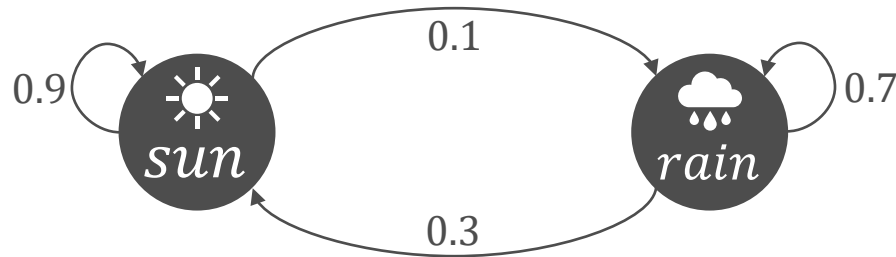
- **Markov assumption:**

$$\mathbf{P}[X_t \mid \mathbf{X}_{0:t-1}] = \mathbf{P}[X_t \mid X_{t-1}]$$

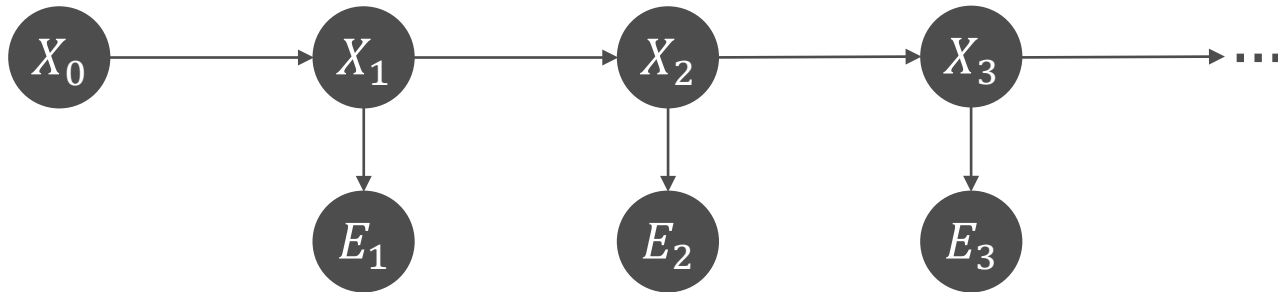
- **Stationarity assumption:** for all t, t' ,

$$\mathbf{P}[X_t \mid X_{t-1}] = \mathbf{P}[X_{t'} \mid X_{t'-1}]$$

PREDICTING THE WEATHER



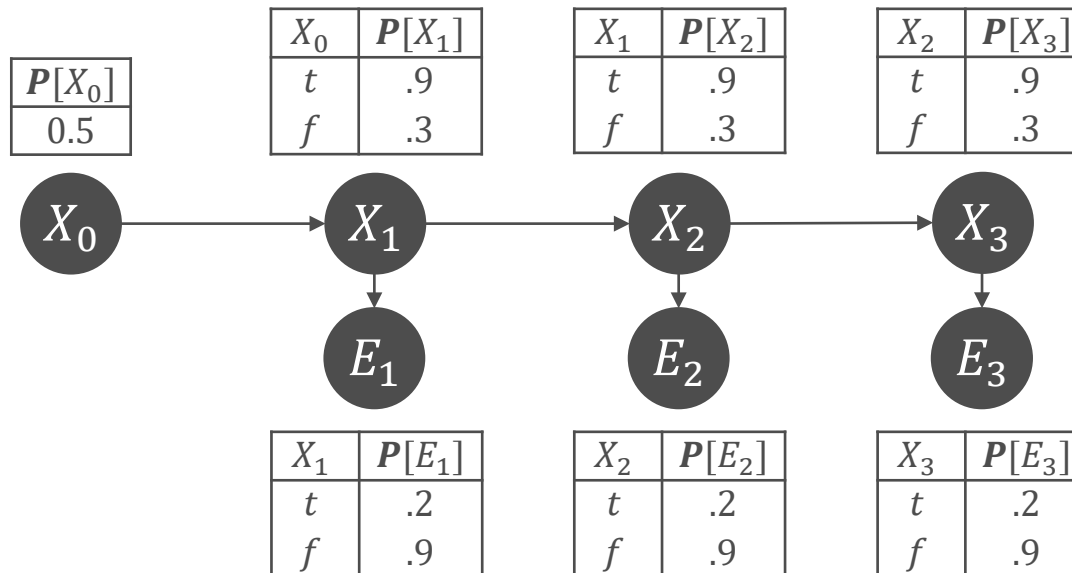
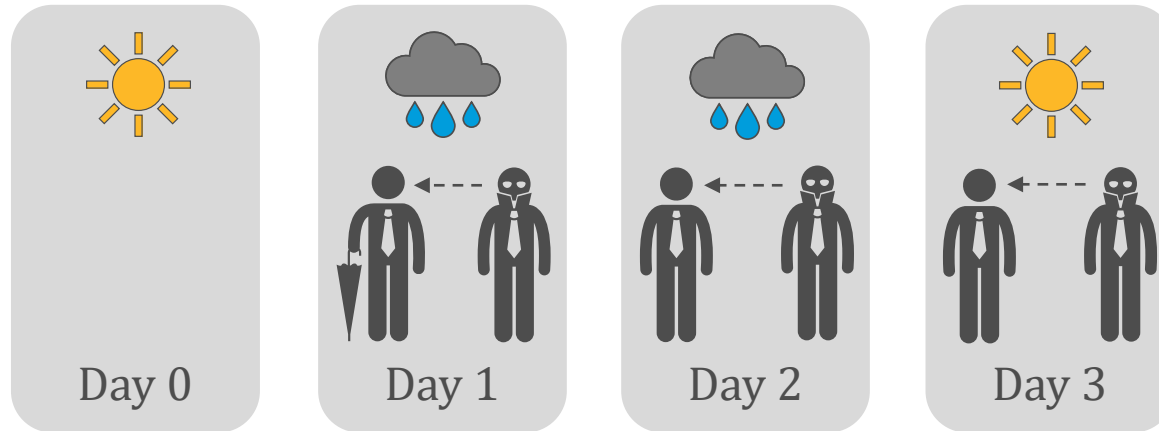
HIDDEN MARKOV MODELS



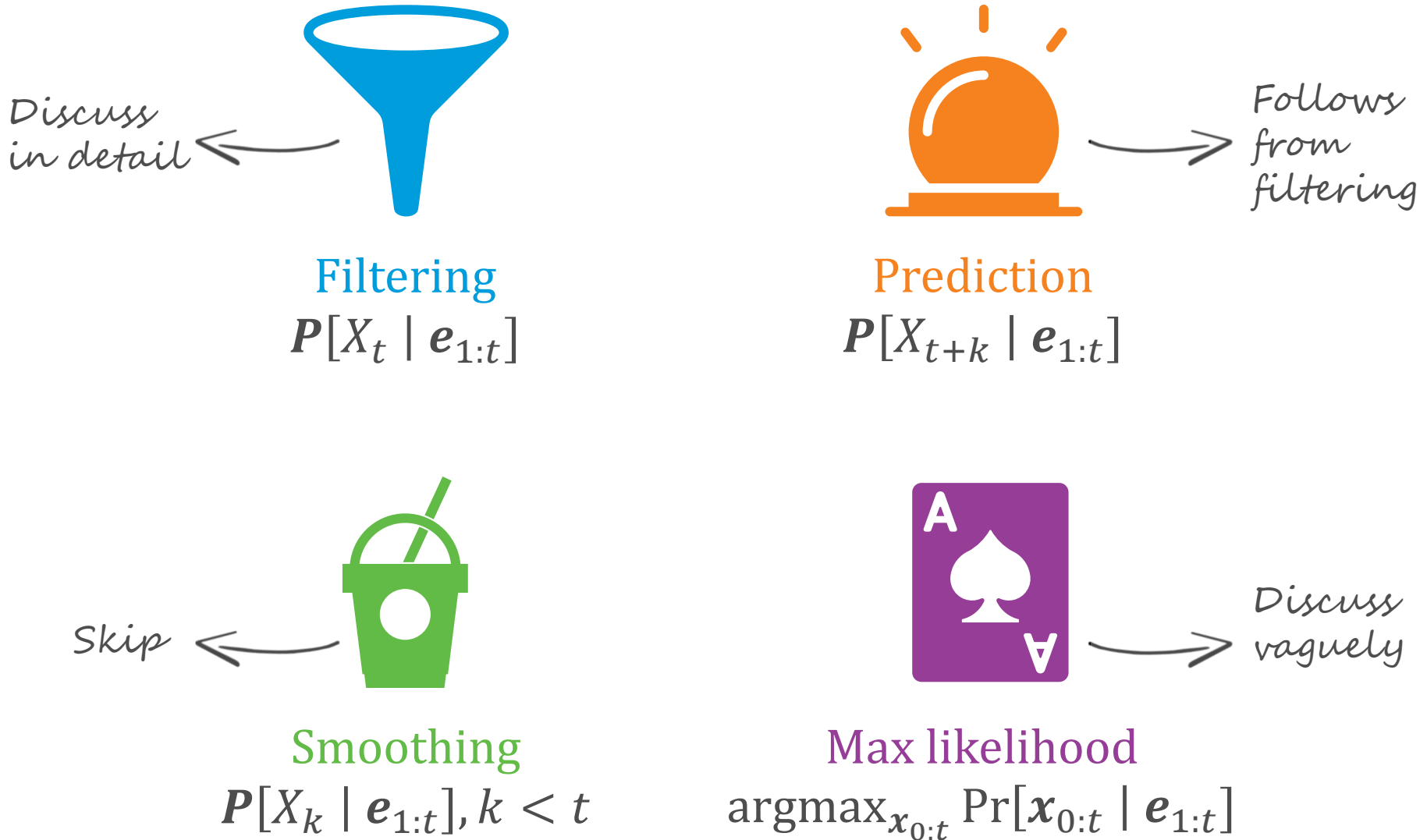
- Sometimes we can't directly observe the state of the world, but rather only observe evidence
- We are given $P[E_t \mid \mathbf{X}_{0:t}, \mathbf{E}_{0:t-1}]$
- A hidden Markov model satisfies the same assumptions as before, plus the **Markov sensor assumption**:

$$P[E_t \mid \mathbf{X}_{0:t}, \mathbf{E}_{0:t-1}] = P[E_t \mid X_t]$$

HIDDEN MARKOV MODELS: EXAMPLE



INFERENCE



FILTERING

- We want an iterative algorithm that would compute $\mathbf{P}[X_{t+1} \mid \mathbf{e}_{1:t+1}]$ given e_{t+1} and our previous calculation of $\mathbf{P}[X_t \mid \mathbf{e}_{1:t}]$
- Given $\mathbf{e}_{1:t}$ we will first compute $\mathbf{P}[X_1 \mid e_1]$, use that estimate to compute $\mathbf{P}[X_2 \mid \mathbf{e}_{1:2}]$ based on e_2 , and so on

FILTERING

$$P[X_{t+1} \mid \mathbf{e}_{1:t+1}] = P[X_{t+1} \mid \mathbf{e}_{1:t}, e_{t+1}]$$

Bayes' Rule

$$\propto P[e_{t+1} \mid X_{t+1}, \mathbf{e}_{1:t}] \cdot P[X_{t+1} \mid \mathbf{e}_{1:t}]$$

Conditional independence

$$= P[e_{t+1} \mid X_{t+1}] \cdot P[X_{t+1} \mid \mathbf{e}_{1:t}]$$

Condition on x_t

$$= P[e_{t+1} \mid X_{t+1}] \cdot \sum_{x_t} \Pr[x_t \mid \mathbf{e}_{1:t}] \cdot P[X_{t+1} \mid x_t, \mathbf{e}_{1:t}]$$

Conditional independence

$$= P[e_{t+1} \mid X_{t+1}] \cdot \sum_{x_t} \Pr[x_t \mid \mathbf{e}_{1:t}] \cdot P[X_{t+1} \mid x_t]$$

Given

Previously computed

Given

FILTERING: EXAMPLE

$$P[X_{t+1} | \mathbf{e}_{1:t+1}] \propto P[e_{t+1} | X_{t+1}] \cdot \sum_{x_t} \Pr[x_t | \mathbf{e}_{1:t}] \cdot P[X_{t+1} | x_t]$$

In our weather example:

$P[X_0]$
0.5

X_0	$P[X_1]$
t	.9
f	.3

X_1	$P[E_1]$
t	.2
f	.9

Suppose the piece of evidence on Day 1 is that the director is carrying an umbrella ($e_1 = t$) then

$$P[X_1 = t | E_1 = t] \propto 0.2 \cdot (0.5 \cdot 0.9 + 0.5 \cdot 0.3) = 0.12$$

$$P[X_1 = f | E_1 = t] \propto 0.9 \cdot (0.5 \cdot 0.1 + 0.5 \cdot 0.7) = 0.36$$

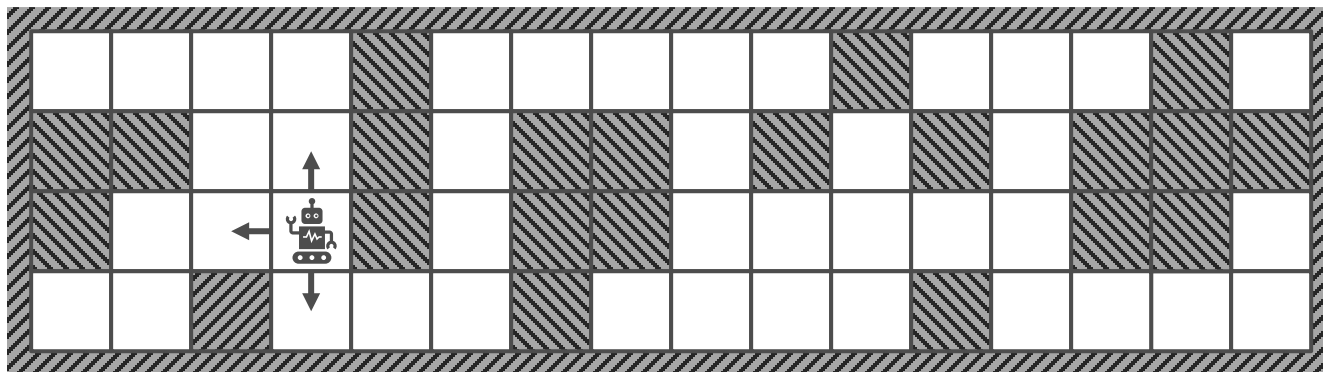
By normalizing we get a prediction of (0.25,0.75)

FILTERING: RUNNING TIME

$$P[X_{t+1} | \mathbf{e}_{1:t+1}] \propto P[e_{t+1} | X_{t+1}] \cdot \sum_{x_t} P[x_t | \mathbf{e}_{1:t}] \cdot P[X_{t+1} | x_t]$$

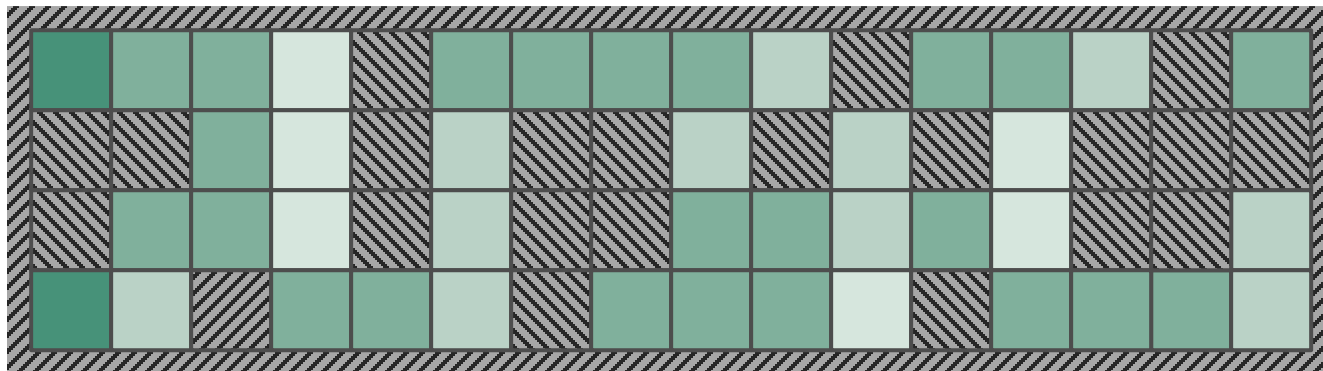
- **Poll 1:** What is the running time of computing $P[X_t | \mathbf{e}_{1:t}]$ as a function of t ?
 - $\Theta(1)$
 - $\Theta(t)$ ✓
 - $\Theta(t^2)$
 - $\Theta(t^3)$

ROBOT LOCALIZATION AS HMM

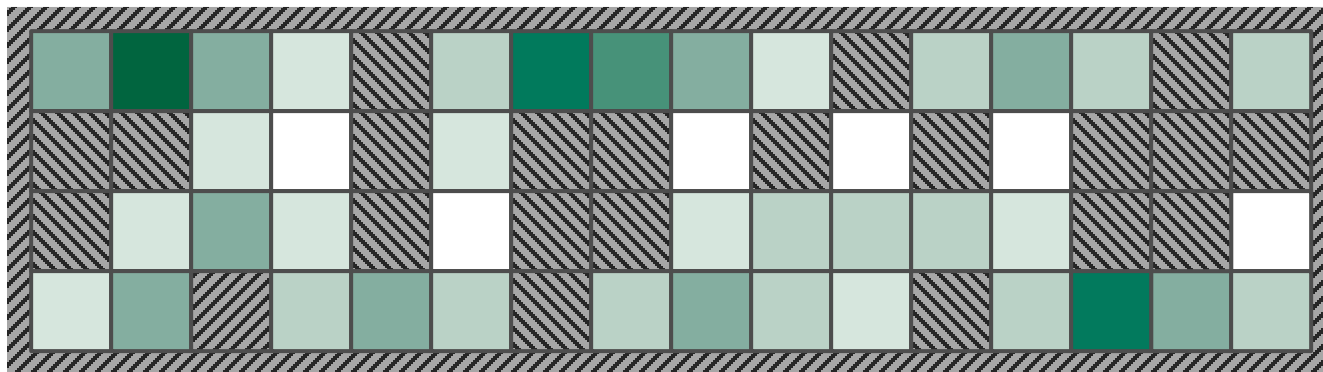


- State X_t is position of robot on grid
- In each step the robot moves in a random unblocked direction
- The robot starts in a random cell
- E_t consists of a string of four bits that indicate obstacles in N, E, S, W
- The sensor error rate for each bit is 0.2

ROBOT LOCALIZATION AS HMM

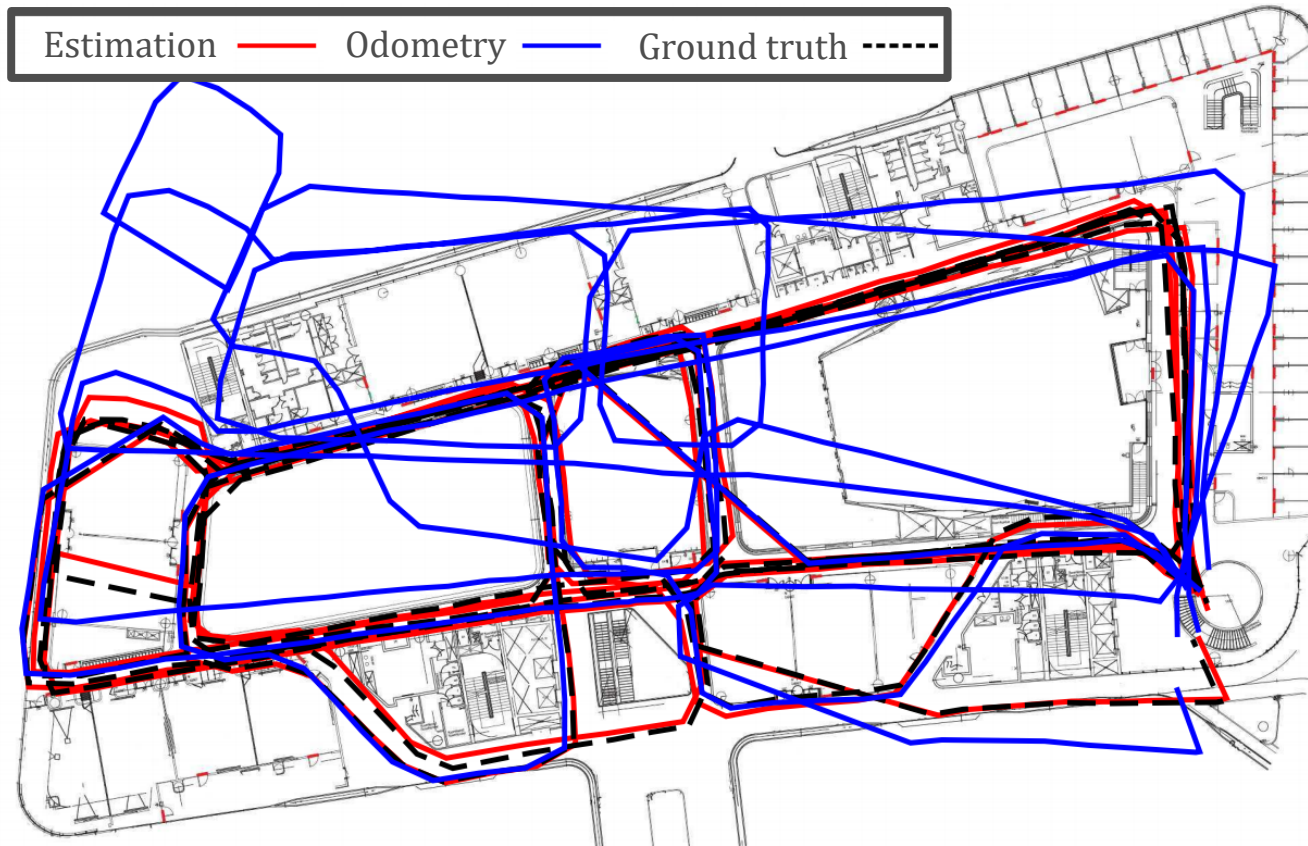


Posterior distribution over robot location after $E_1 = 1011$



Posterior distribution over robot location after $E_1 = 1011, E_2 = 1010$

LOCALIZATION IN THE REAL WORLD



[Liu et al., 2019]

MAXIMUM LIKELIHOOD

- We want an iterative algorithm to compute

$$\operatorname{argmax}_{\mathbf{x}_{0:t}} \mathbf{P}[\mathbf{x}_{0:t}, X_{t+1} \mid \mathbf{e}_{1:t+1}]$$

given e_{t+1} and our previous computation of

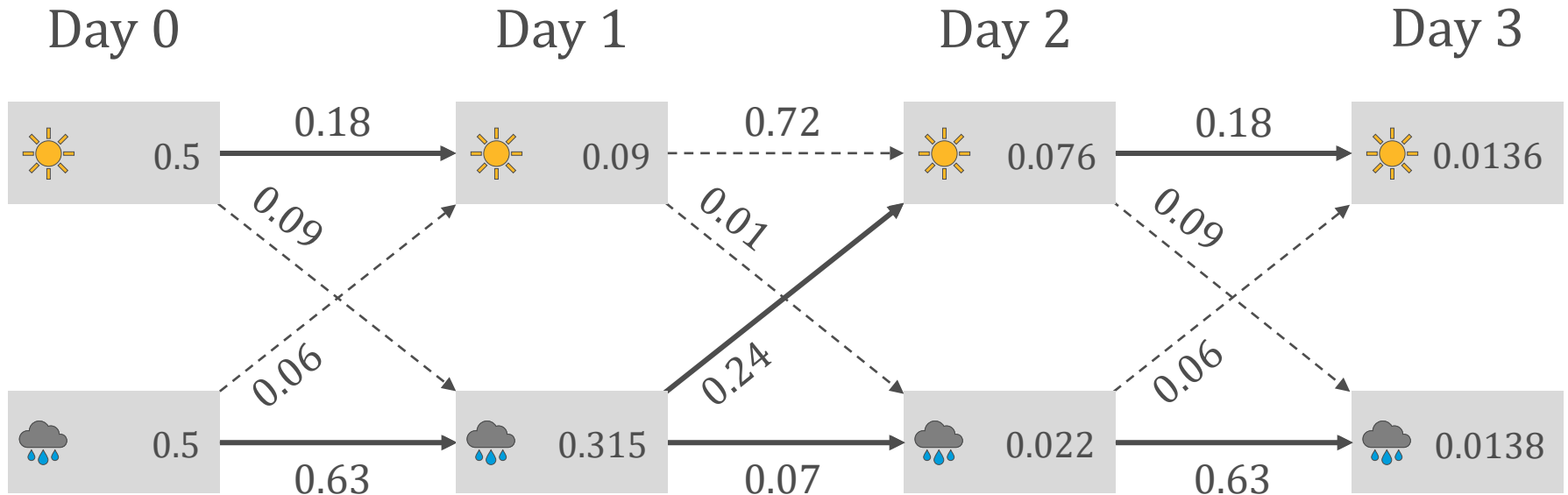
$$\operatorname{argmax}_{\mathbf{x}_{0:t-1}} \mathbf{P}[\mathbf{x}_{0:t-1}, X_t \mid \mathbf{e}_{1:t}]$$

- By a calculation similar to the one we did for filtering, we can show that

$$\begin{aligned} & \max_{\mathbf{x}_{0:t}} \mathbf{P}[\mathbf{x}_{0:t}, X_{t+1} \mid \mathbf{e}_{1:t+1}] \\ & \propto \mathbf{P}[e_{t+1} \mid X_{t+1}] \max_{x_t} \mathbf{P}[X_{t+1} \mid x_t] \max_{\mathbf{x}_{0:t-1}} \Pr[\mathbf{x}_{0:t} \mid \mathbf{e}_{1:t}] \end{aligned}$$

- Andrew Viterbi showed in 1967 how to compute this efficiently

VITERBI: EXAMPLE



Evidence:



The weight on each edge is
 $\Pr[e_{t+1} \mid x_{t+1}] \cdot \Pr[x_{t+1} \mid x_t]$

VITERBI: RUNNING TIME

- **Poll 2:** What is the running time of the Viterbi Algorithm as a function of t ?
 - $\Theta(1)$
 - $\Theta(t)$ ✓
 - $\Theta(t^2)$
 - $\Theta(t^3)$

APPLICATION: POS TAGGING

