Probabilistic Reasoning Over Time

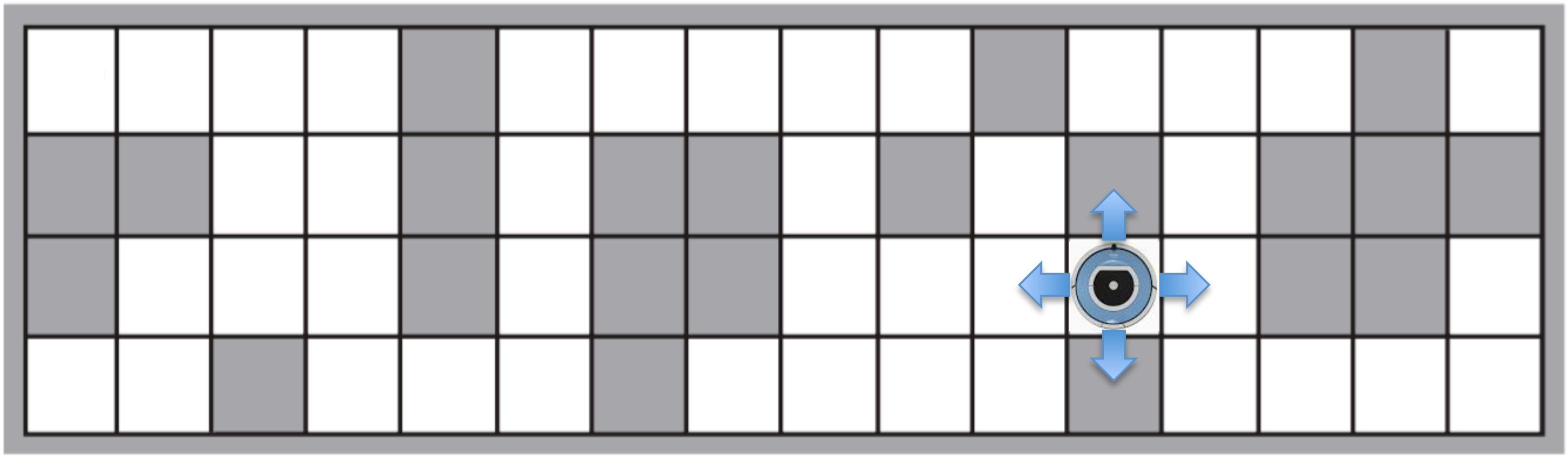
Artificial Intelligence
School of Computer Science
The University of Adelaide

Example: is it raining outside?

- You are the security guard permanently located at a secret underground installation.
- You cannot see the weather outside.
- Everyday, you see the director arriving with or without an umbrella.
- At day t+1, the director arrived with an umbrella. Is it raining outside?

Day	1	2	3	…	t	t+1
Observed umbrella?	✓	✓	✗	…	✗	✓

Example: robot localisation



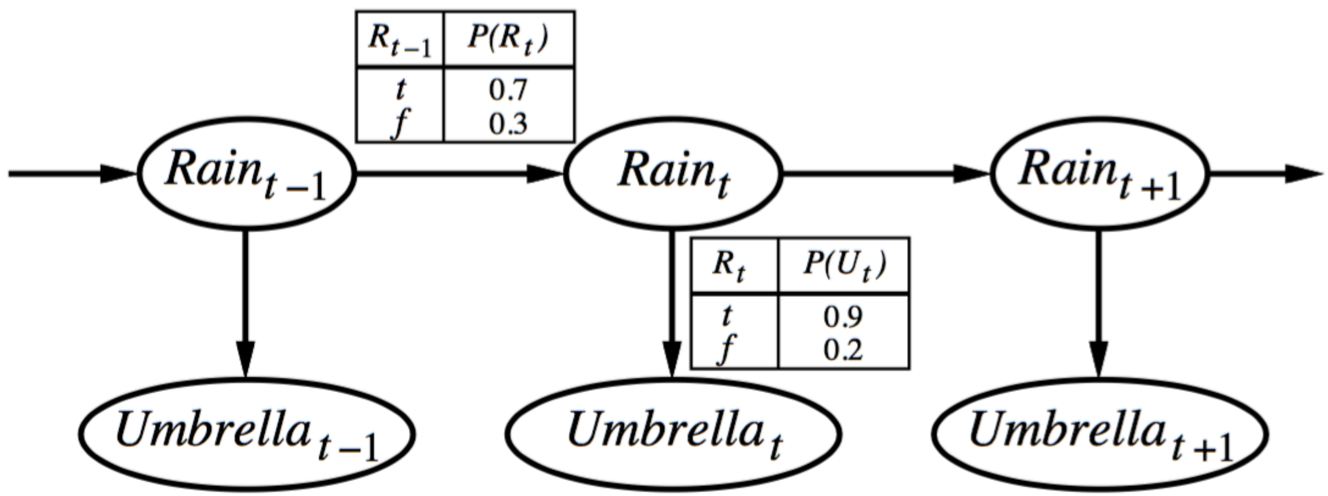
- A robot is moving in a 2D map with obstacles. Its sensor detects the directions are blocked by obstacles at the robot's location.
- A sequence of sensor readings is observed:

Time step	**1**	**2**	**3**	**…**	t	t**+1**
Blocked directions	N S W	N S	N	…	S E W	S E

- At time step t+1, where is the robot?

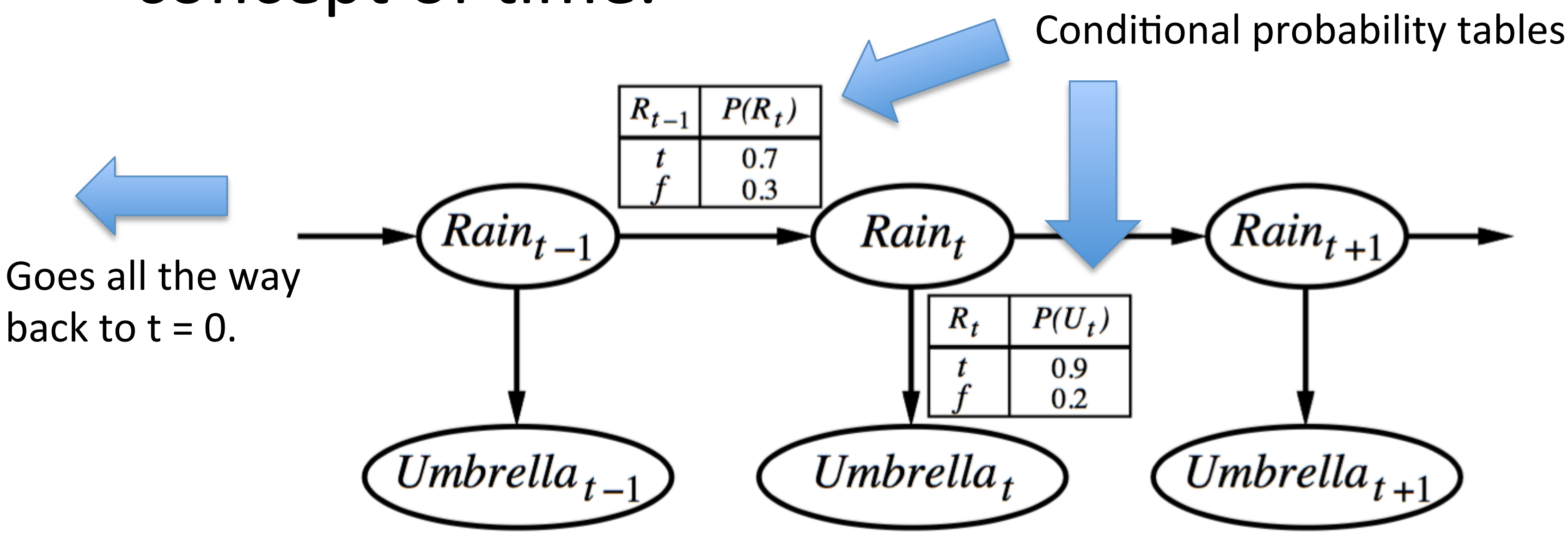
Example: is it raining outside?

- A commonly used temporal model for this kind of problem:



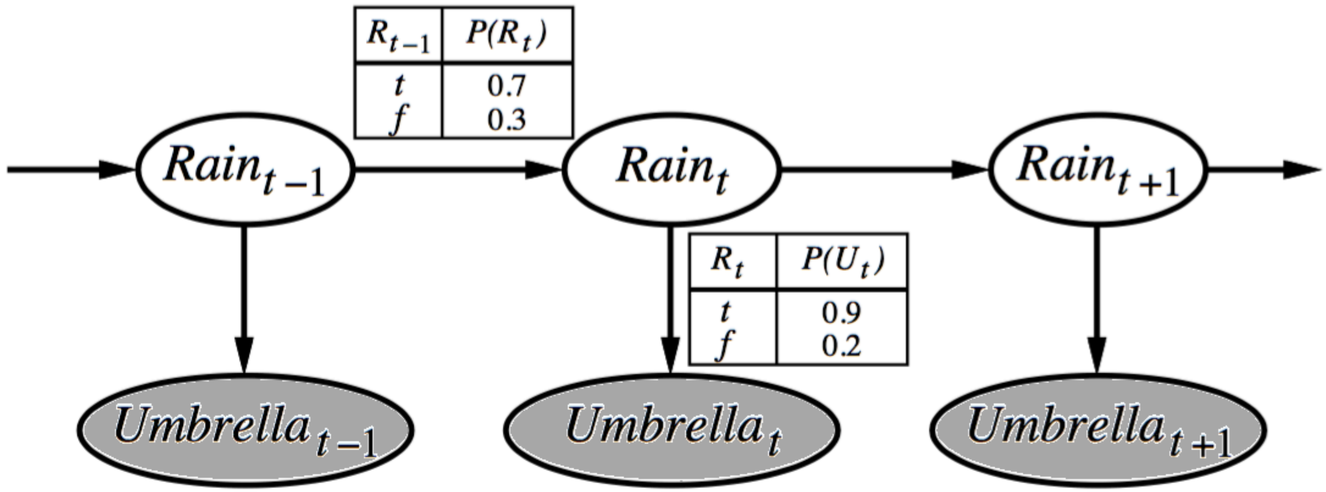
Example: is it raining outside?

- This is just a Bayesian Network with the concept of time.



- Variables = { R_0, R_1, …, R_{t+1}, U_1, …, U_{t+1} }.

Example: is it raining outside?



- You have observed evidences

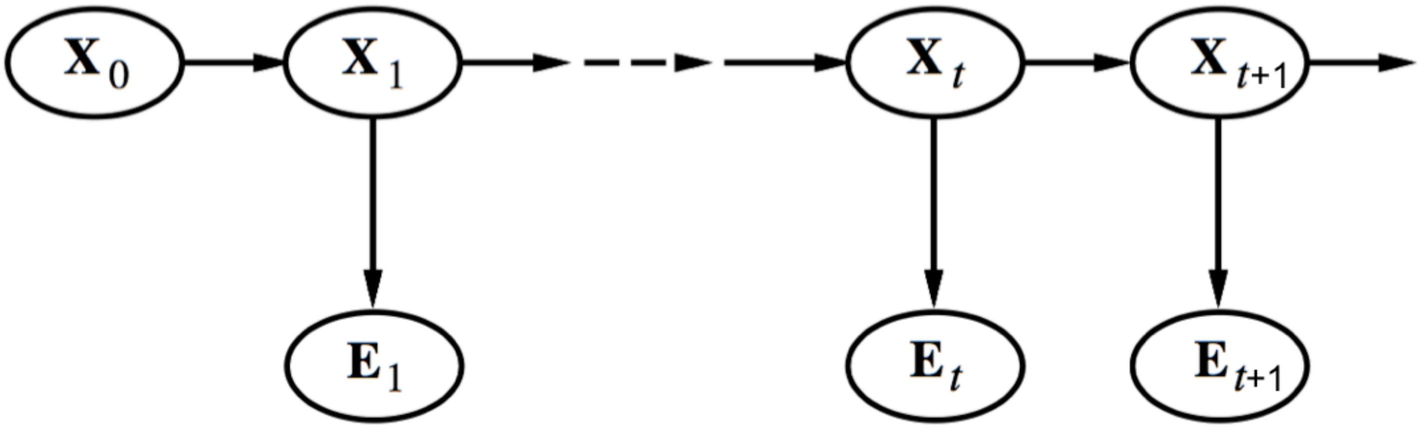
$$\{u_1, \ldots, u_{t+1}\} = \{\text{true,true,false,...,false,true}\}.$$

- You want to calculate the probability

$$P(R_{t+1} | u_1, \ldots, u_{t+1})$$

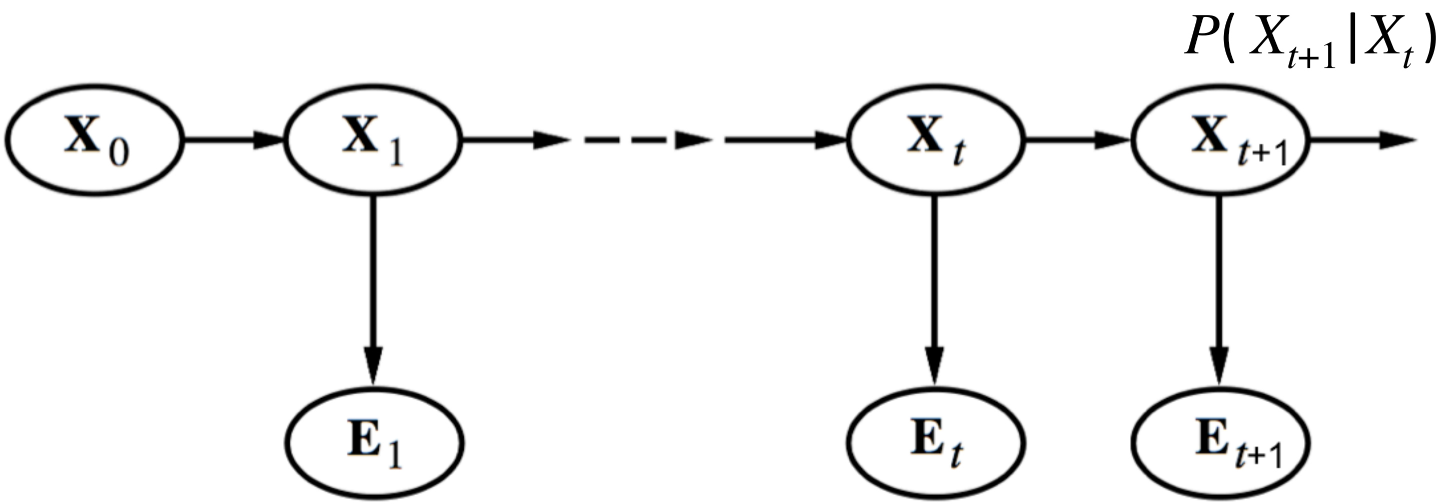
  for R_{t+1} = true and R_{t+1} = false.
- This is a special kind of probabilistic inference called filtering.

The general case



- State variables { X_0, X_1, …, X_{t+1} }.
- Evidence variables { E_1, …, E_{t+1} }.
- By convention, we assume X_t starts at t=0 while E_t starts at t=1.

The general case

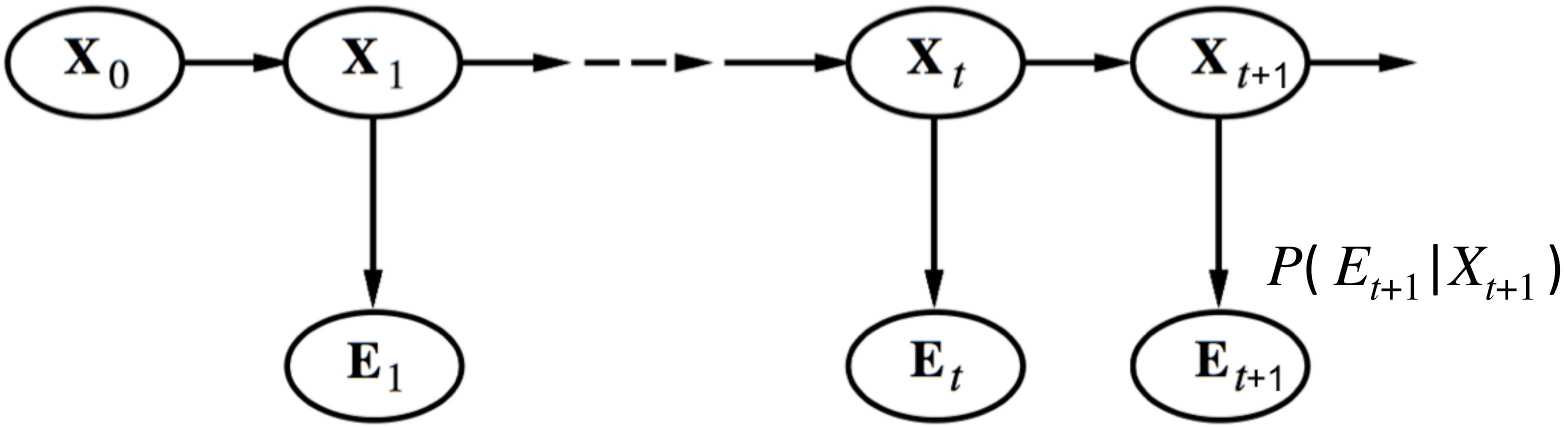


- State transition model

$$P(X_{t+1} | X_0, \ldots, X_t) = P(X_{t+1} | X_t)$$

- First order Markov assumption: the present state depends only on the immediate previous state.

The general case



- Observation model

$$P(E_{t+1} | X_{0:t+1}, E_{0:t}) = P(E_{t+1} | X_{t+1})$$

- Sensor Markov assumption: the probability of observing E_t depends only on the state X_t.

*Note: $X_{0:t} = X_0, X_1, \ldots, X_t$

Filtering

- We have observed $\mathbf{e}_1$, ..., $\mathbf{e}_{t+1}$ = $\mathbf{e}_{1:t+1}$. We wish to calculate

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

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \quad \text{(dividing up the evidence)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}, \mathbf{e}_{1:t})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(using Bayes' rule)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(by the sensor Markov assumption).}
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t, \mathbf{e}_{1:t}) P(\mathbf{x}_t \mid \mathbf{e}_{1:t}) \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{e}_{1:t}) \quad \text{(Markov assumption).}
\end{aligned}$$

Filtering

- We have observed $\mathbf{e}_1$, ..., $\mathbf{e}_{t+1}$ = $\mathbf{e}_{1:t+1}$. We wish to calculate

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

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \quad \text{(dividing up the evidence)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}, \mathbf{e}_{1:t})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(using Bayes' rule)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(by the sensor Markov assumption).}
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t, \mathbf{e}_{1:t}) P(\mathbf{x}_t \mid \mathbf{e}_{1:t}) \\
&= \alpha\, \underbrace{\mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1})}_{\text{Observation model}} \sum_{\mathbf{x}_t} \underbrace{\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t)}_{\text{Transition model}} P(\mathbf{x}_t \mid \mathbf{e}_{1:t}) \quad \text{(Markov assumption).}
\end{aligned}$$

Filtering

- We have observed $\mathbf{e}_1, \ldots, \mathbf{e}_{t+1} = \mathbf{e}_{1:t+1}$. We wish to calculate

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

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \quad \text{(dividing up the evidence)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}, \mathbf{e}_{1:t})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(using Bayes' rule)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(by the sensor Markov assumption).}
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t, \mathbf{e}_{1:t}) P(\mathbf{x}_t \mid \mathbf{e}_{1:t}) \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{e}_{1:t}) \quad \text{(Markov assumption).}
\end{aligned}$$

Has the same form, but at one time step before!
This process is called recursive estimation.

Filtering

- We have observed $\mathbf{e}_1, \ldots, \mathbf{e}_{t+1} = \mathbf{e}_{1:t+1}$. We wish to calculate

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

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \quad \text{(dividing up the evidence)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}, \mathbf{e}_{1:t})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(using Bayes' rule)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(by the sensor Markov assumption).}
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t, \mathbf{e}_{1:t}) P(\mathbf{x}_t \mid \mathbf{e}_{1:t}) \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \underbrace{\sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{e}_{1:t})}_{\text{Calculating this is called prediction.}} \quad \text{(Markov assumption).}
\end{aligned}$$

Filtering

- We have observed $\mathbf{e}_1, \ldots, \mathbf{e}_{t+1} = \mathbf{e}_{1:t+1}$. We wish to calculate

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

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \,|\, \mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1} \,|\, \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \quad \text{(dividing up the evidence)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \,|\, \mathbf{X}_{t+1}, \mathbf{e}_{1:t})\, \mathbf{P}(\mathbf{X}_{t+1} \,|\, \mathbf{e}_{1:t}) \quad \text{(using Bayes' rule)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \,|\, \mathbf{X}_{t+1})\, \mathbf{P}(\mathbf{X}_{t+1} \,|\, \mathbf{e}_{1:t}) \quad \text{(by the sensor Markov assumption).}
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \,|\, \mathbf{e}_{1:t+1}) &= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \,|\, \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \,|\, \mathbf{x}_t, \mathbf{e}_{1:t}) P(\mathbf{x}_t \,|\, \mathbf{e}_{1:t}) \\
&= \underbrace{\alpha\, \mathbf{P}(\mathbf{e}_{t+1} \,|\, \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \,|\, \mathbf{x}_t) P(\mathbf{x}_t \,|\, \mathbf{e}_{1:t})} \quad \text{(Markov assumption).}
\end{aligned}$$

Combining the prediction with the new evidence is called update.

Example: is it raining outside?

- On day 0, we have no observations, only the security guard's prior beliefs; let's assume that consists of $\mathbf{P}(R_0) = \langle 0.5, 0.5 \rangle$.
- On day 1, the umbrella appears, so $U_1 = true$. The prediction from $t = 0$ to $t = 1$ is

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

Then the update step simply multiplies by the probability of the evidence for $t = 1$ and normalizes, as shown in Equation (15.4):

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

Example: is it raining outside?

- On day 2, the umbrella appears, so $U_2 = true$. The prediction from $t = 1$ to $t = 2$ is

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

and updating it with the evidence for $t = 2$ gives

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

Can keep on going as new observations are made.

Hidden Markov Model (HMM)

- A HMM is obtained if X_t and E_t for all t are single discrete random variables.

  e.g., "is it raining outside?" is a HMM.

- In a HMM, the transition model can be encoded in an SxS matrix

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}[1,1] & \mathbf{T}[1,2] & \cdots & \mathbf{T}[1,S] \\ \mathbf{T}[2,1] & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ \mathbf{T}[S,1] & \cdots & \cdots & \mathbf{T}[S,S] \end{bmatrix}$$

where S in the number of possible values of X_t, and

$$\mathbf{T}[i,j] = P(X_t = j \mid X_{t\text{-}1} = i)$$

Hidden Markov Model (HMM)

- Given evidence e_t at time step t, the observation model can be encoded in an SxS diagonal matrix

$$\mathbf{O}_t = \begin{bmatrix} \mathbf{O}_t[1,1] & 0 & \cdots & 0 \\ 0 & \mathbf{O}_t[2,2] & \cdots & \vdots \\ \vdots & \cdots & \mathbf{O}_t[3,3] & 0 \\ 0 & \cdots & 0 & \mathbf{O}_t[S,S] \end{bmatrix}$$

where

$$\mathbf{O}_t[\mathrm{i},\mathrm{i}] = P(e_t | X_t = i)$$

Hidden Markov Model (HMM)

- Recursive estimation in HMM can be computed as

$$\mathbf{f}_{t+1} = \alpha \mathbf{O}_{t+1} \mathbf{T}^T \mathbf{f}_t$$

where

$\mathbf{f}_{t+1}$ is column vector form of $P(\mathrm{X}_{t+1} \mid \mathrm{e}_{1:t+1})$

$\mathbf{f}_t$ is column vector form of $P(\mathrm{X}_t \mid \mathrm{e}_{1:t})$

Example: is it raining outside?

- From transition model:

$$\mathbf{T} = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$$

- At t=0, rain or no rain is equally likely, so

$$\mathbf{f}_0 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

- At t=1, umbrella is observed, so

$$\mathbf{O}_1 = \begin{bmatrix} 0.9 & 0 \\ 0 & 0.2 \end{bmatrix}$$

- Filtering result at t=1 is

$$\mathbf{f}_1 = \alpha \begin{bmatrix} 0.9 & 0 \\ 0 & 0.2 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}^T \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \alpha \begin{bmatrix} 0.45 \\ 0.1 \end{bmatrix} \approx \begin{bmatrix} 0.818 \\ 0.182 \end{bmatrix}$$

Example: is it raining outside?

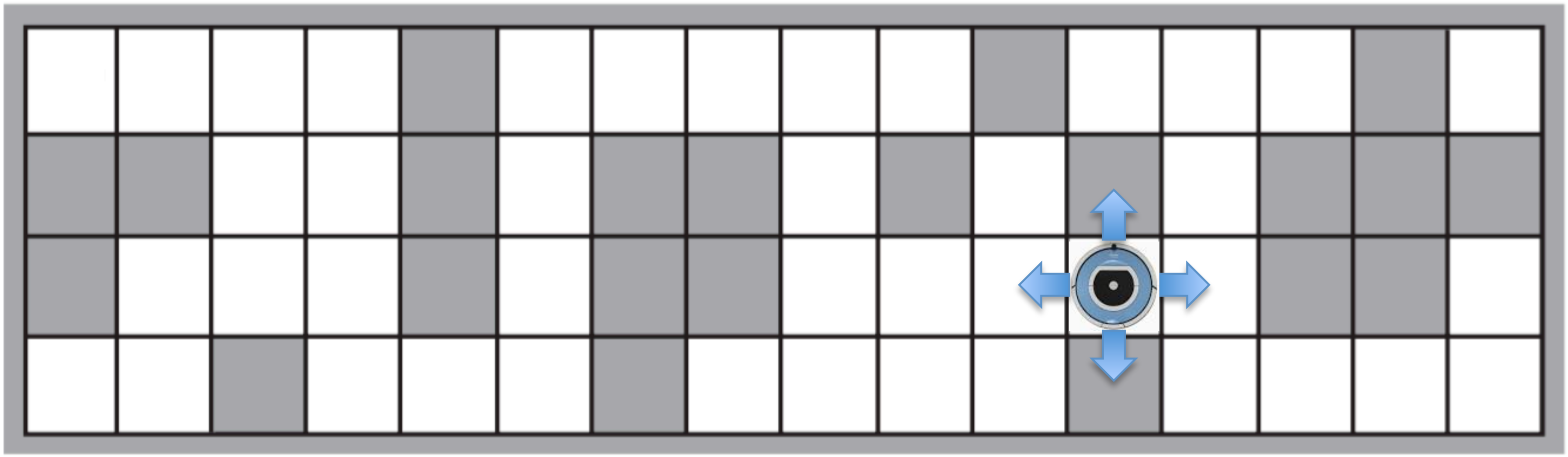
- At t=2, umbrella is observed, so

$$\mathbf{O}_2 = \begin{bmatrix} 0.9 & 0 \\ 0 & 0.2 \end{bmatrix}$$

- Filtering result at t=2 is

$$\mathbf{f}_2 = \alpha \begin{bmatrix} 0.9 & 0 \\ 0 & 0.2 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}^T \begin{bmatrix} 0.45 \\ 0.1 \end{bmatrix} = \alpha \begin{bmatrix} 0.3105 \\ 0.041 \end{bmatrix} \approx \begin{bmatrix} 0.883 \\ 0.117 \end{bmatrix}$$

Example: robot localisation



- State variable represents the robot location:

$$X_t \in \{1, 2, \ldots, S\}$$

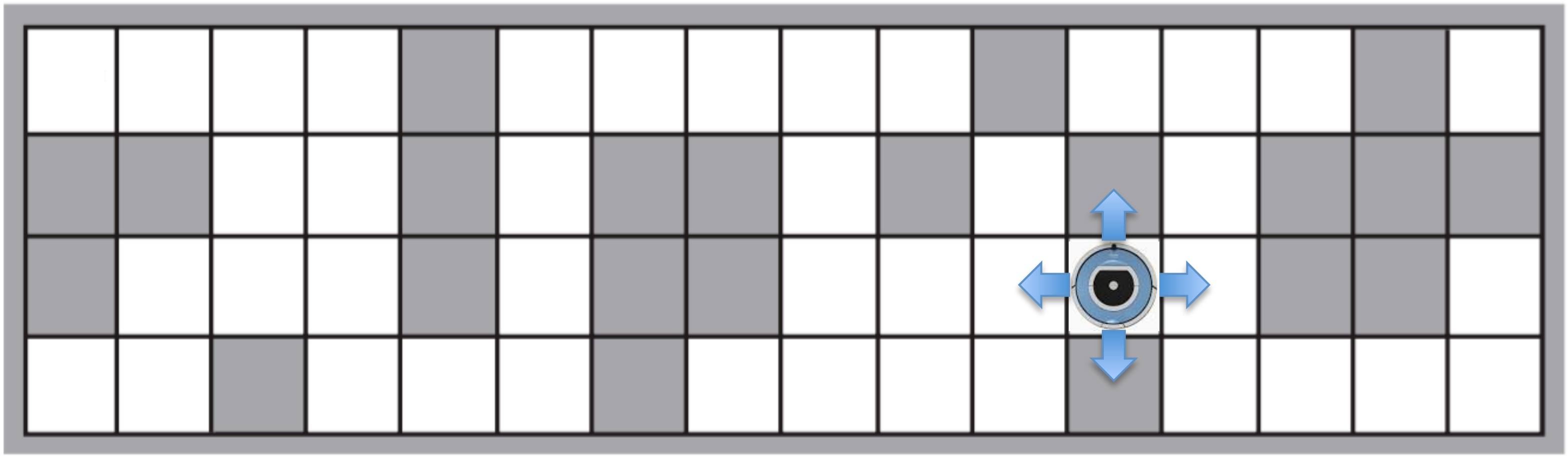
- Assuming random walk, the transition model is

$$P(X_{t+1} = j \mid X_t = i) = \begin{cases} 1/N(i) & \text{if } j \in \text{NEIGHBOURS(i)} \\ 0 & \text{otherwise} \end{cases}$$

where $\text{NEIGHBOURS}(i)$ = set of neighbours of cell i.

$N(i)$ = number of neighbours of cell i.

Example: robot localisation



- The sensor's error rate is ϵ and error occurs independently in the four directions. This gives the observation model

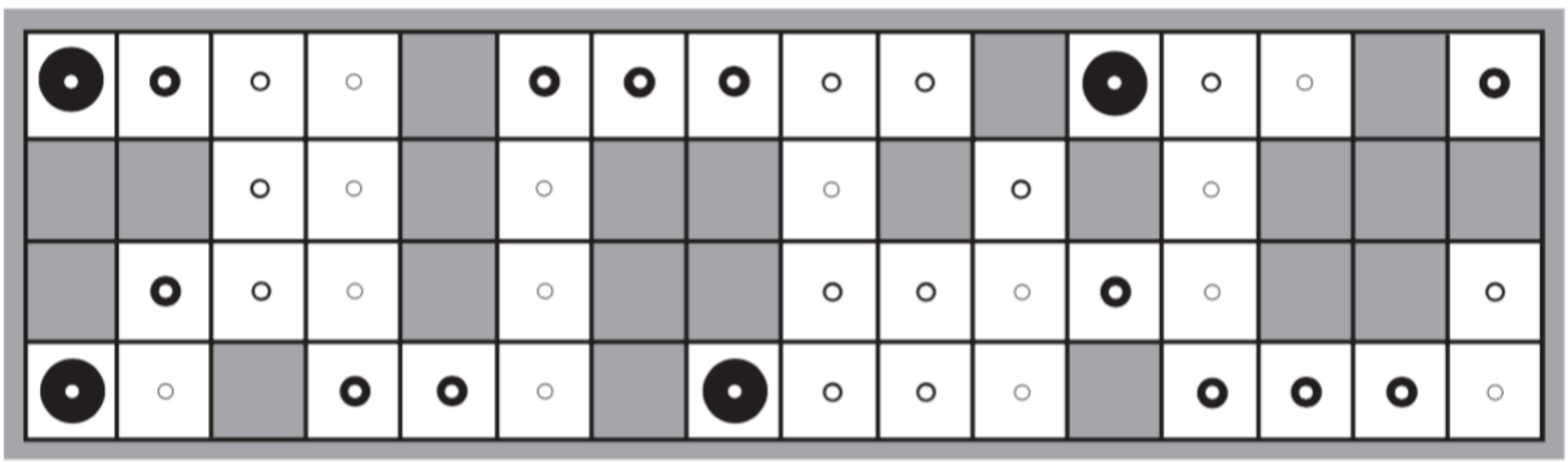
$$P(E_t = e_t \mid X_t = i) = (1-\epsilon)^{4-d_{it}}\epsilon^{d_{it}}$$

where d_{it} is the number of directions that are wrong given location $X_t = i$ and sensor reading $E_t = e_t$.

- Example: at the robot's position in the map above, the probability of observing $e_t = NSE$ is $(1-\epsilon)^3\epsilon^1$.

Example: robot localisation

- Assume the robot is equally likely to be at any square at $t = 0$, i.e., $\mathbf{f}_0$ is uniform.
- After observing $E_1 = NSW$, $\mathbf{f}_1$ becomes



Example: robot localisation

- After observing $E_2 = NS$, $\mathbf{f}_2$ becomes