

# Artificial Intelligence Foundation – JC3001

## Lecture 22: Uncertainty over Time I

**Prof. Aladdin Ayesh** (aladdin.ayesh@abdn.ac.uk)

**Dr. Binod Bhattarai** (binod.bhattarai@abdn.ac.uk)

**Dr. Gideon Ogunniye**, (g.ogunniye@abdn.ac.uk)

September 2025

Material adapted from:  
Russell and Norvig (AIMA Book): Chapter 14 (14.1–14.3)

- Part 1: Introduction
  - ① Introduction to AI ✓
  - ② Agents ✓
- Part 2: Problem-solving
  - ① Search 1: Uninformed Search ✓
  - ② Search 2: Heuristic Search ✓
  - ③ Search 3: Local Search ✓
  - ④ Search 4: Adversarial Search ✓
- Part 3: Reasoning and Uncertainty
  - ① Reasoning 1: Constraint Satisfaction ✓
  - ② Reasoning 2: Logic and Inference ✓
  - ③ Probabilistic Reasoning 1: BNs ✓
  - ④ **Probabilistic Reasoning 2: HMMs**
- Part 4: Planning
  - ① Planning 1: Intro and Formalism
  - ② Planning 2: Algos and Heuristics
  - ③ Planning 3: Hierarchical Planning
  - ④ Planning 4: Stochastic Planning
- Part 5: Learning
  - ① Learning 1: Intro to ML
  - ② Learning 2: Regression
  - ③ Learning 3: Neural Networks
  - ④ Learning 4: Reinforcement Learning
- Part 6: Conclusion
  - ① Ethical Issues in AI
  - ② Conclusions and Discussion

- Time and Uncertainty
- Inference in Temporal Models
- Hidden Markov Models



# Outline

## 1 Time and Uncertainty

► Time and Uncertainty

► Inference in Temporal Models

- discrete-time models: we view the world as a series of snapshots or time slices
- the time interval  $\Delta$  between slices, we assume to be the same for every interval
- $X_t$ : denotes the set of state variables at time  $t$ , which we assume to be unobservable
- $E_t$ : denotes the set of observable evidence variables: observation at time  $t$  is  $E_t = \mathbf{e}_t$

- discrete-time models: we view the world as a series of snapshots or time slices
- the time interval  $\Delta$  between slices, we assume to be the same for every interval
- $X_t$ : denotes the set of state variables at time  $t$ , which we assume to be unobservable
- $E_t$ : denotes the set of observable evidence variables: observation at time  $t$  is  $E_t = \mathbf{e}_t$

### Transition and sensor models

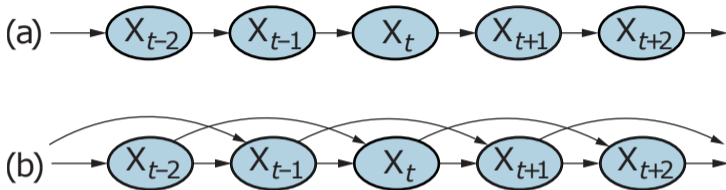
- The transition model specifies the probability distribution over the latest state variables, given the previous values:  $P(X_t \mid X_{0:t-1})$ .
- Problem: the set  $X_{0:t-1}$  is unbounded in size as  $t$  increases.

- discrete-time models: we view the world as a series of snapshots or time slices
- the time interval  $\Delta$  between slices, we assume to be the same for every interval
- $X_t$ : denotes the set of state variables at time  $t$ , which we assume to be unobservable
- $E_t$ : denotes the set of observable evidence variables: observation at time  $t$  is  $E_t = \mathbf{e}_t$

### Transition and sensor models

- The transition model specifies the probability distribution over the latest state variables, given the previous values:  $P(X_t \mid X_{0:t-1})$ .
- Problem: the set  $X_{0:t-1}$  is unbounded in size as  $t$  increases.
- Solution: Markov assumption  
the current state depends on only a finite fixed number of previous states
- $P(E_t \mid X_t)$  is our sensor model, sensor Markov assumption:

$$P(E_t \mid X_{0:t}, E_{1:t-1}) = P(E_t \mid X_t)$$



- (a) Bayesian network structure corresponding to a first-order Markov process with state defined by the variables  $X_t$

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

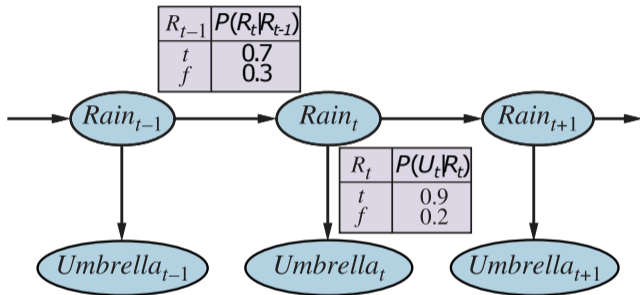
- (b) A second-order Markov process

$$P(X_t | X_{t-2}, X_{t-1})$$

- the prior probability distribution at time 0,  $P(X_0)$ .

$$P(X_{0:t}, E_{1:t}) = P(X_0) \prod_{i=1} P(X_i | X_{i-1}) P(E_i | X_i)$$

- Umbrella World: first-order Markov process—–the probability of rain is assumed to depend only on whether it rained the previous day
- The first-order Markov assumption says that the state variables contain all the information needed to characterize the probability distribution for the next time slice.
- Ways to improve the accuracy of the approximation
  - Increasing the order of the Markov process mode
  - Increasing the set of state variables



- Bayesian network structure and conditional distributions describing the umbrella world.
- The transition model is  $P(Rain_t | Rain_{t-1})$  and the sensor model is  $P(Umbrella_t | Rain_t)$ .



# Outline

## 2 Inference in Temporal Models

► Time and Uncertainty

► Inference in Temporal Models

- Formulate the basic inference tasks that must be solved:
  - **Filtering or state estimation** is the task of computing the **belief state**  $P(X_t \mid \mathbf{e}_{1:t})$
  - **Prediction:** This is the task of computing the posterior distribution over the future state, given all evidence to date.
  - **Smoothing:** This is the task of computing the posterior distribution over a past state, given all evidence up to the present
  - **Most likely explanation:** Given a sequence of observations, we might wish to find the sequence of states that is most likely to have generated those observations
- Besides inference tasks:
  - **Learning:** The transition and sensor models, if not yet known, can be learned from observations

We want to derive the belief state at time  $t + 1$ , given:

- new evidence  $\mathbf{e}_{t+1}$ , and
- the belief state  $P(X_t \mid \mathbf{e}_{1:t})$  so far

We want to derive the belief state at time  $t + 1$ , given:

- new evidence  $\mathbf{e}_{t+1}$ , and
- the belief state  $P(X_t \mid \mathbf{e}_{1:t})$  so far

$$P(X_{t+1} \mid \mathbf{e}_{1:t+1}) = \mathbf{f}(\mathbf{e}_{t+1}, P(X_t \mid \mathbf{e}_{1:t}))$$

# Inference in Temporal Models

Derivation of Recursive Formula

$$P(X_{t+1} \mid \mathbf{e}_{1:t+1}) = P(X_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1})$$

(dividing up the evidence)

$$\begin{aligned} P(X_{t+1} \mid \mathbf{e}_{1:t+1}) &= P(X_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) && \text{(dividing up the evidence)} \\ &= \alpha P(\mathbf{e}_{t+1} \mid X_{t+1}, \mathbf{e}_{1:t}) P(X_{t+1} \mid \mathbf{e}_{1:t}) && \text{(using Bayes' rule, given } \mathbf{e}_{1:t}) \end{aligned}$$

Recall: Bayes Rule

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

$$\begin{aligned} P(X_{t+1} \mid \mathbf{e}_{1:t+1}) &= P(X_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) && \text{(dividing up the evidence)} \\ &= \alpha P(\mathbf{e}_{t+1} \mid X_{t+1}, \mathbf{e}_{1:t}) P(X_{t+1} \mid \mathbf{e}_{1:t}) && \text{(using Bayes' rule, given } \mathbf{e}_{1:t}) \end{aligned}$$

Recall: Bayes Rule

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

more general form of Bayes Rule, conditionalised on background evidence

$$\begin{aligned}
 P(X_{t+1} \mid \mathbf{e}_{1:t+1}) &= P(X_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) && \text{(dividing up the evidence)} \\
 &= \alpha P(\mathbf{e}_{t+1} \mid X_{t+1}, \mathbf{e}_{1:t}) P(X_{t+1} \mid \mathbf{e}_{1:t}) && \text{(using Bayes' rule, given } \mathbf{e}_{1:t}) \\
 &= \alpha \underbrace{P(\mathbf{e}_{t+1} \mid X_{t+1})}_{\text{update}} \underbrace{P(X_{t+1} \mid \mathbf{e}_{1:t})}_{\text{prediction}} && \text{(by the sensor Markov Assumption)}
 \end{aligned}$$

Sensor Markov Assumption

$$P(E_t \mid X_{0:t}, E_{1:t-1}) = P(E_t \mid X_t)$$

$$\begin{aligned}
 P(X_{t+1} \mid \mathbf{e}_{1:t+1}) &= P(X_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) && \text{(dividing up the evidence)} \\
 &= \alpha P(\mathbf{e}_{t+1} \mid X_{t+1}, \mathbf{e}_{1:t}) P(X_{t+1} \mid \mathbf{e}_{1:t}) && \text{(using Bayes' rule, given } \mathbf{e}_{1:t}) \\
 &= \alpha \underbrace{P(\mathbf{e}_{t+1} \mid X_{t+1})}_{\text{update}} \underbrace{P(X_{t+1} \mid \mathbf{e}_{1:t})}_{\text{prediction}} && \text{(by the sensor Markov Assumption)}
 \end{aligned}$$

Plugging in  $P(X_{t+1} \mid \mathbf{e}_{1:t}) = \sum_{x_t} P(X_{t+1} \mid x_t, \mathbf{e}_{1:t}) P(x_t \mid \mathbf{e}_{1:t})$

$$\begin{aligned}
 P(X_{t+1} \mid \mathbf{e}_{1:t+1}) &= \alpha P(\mathbf{e}_{t+1} \mid X_{t+1}) \sum_{x_t} P(X_{t+1} \mid x_t, \mathbf{e}_{1:t}) P(x_t \mid \mathbf{e}_{1:t}) \\
 &= \alpha \underbrace{P(\mathbf{e}_{t+1} \mid X_{t+1})}_{\text{sensor model}} \sum_{x_t} \underbrace{P(X_{t+1} \mid x_t)}_{\text{transition model}} \underbrace{P(x_t \mid \mathbf{e}_{1:t})}_{\text{recursion}} && \text{(Markov assumption)}
 \end{aligned}$$

We can think of the filtered estimate  $P(X_t \mid \mathbf{e}_{1:t})$  as a “message”  $\mathbf{f}_{1:t}$ :

- Propagated forward along the sequence
- Modified by each transition
- Updated by each new observation

So that

$$\mathbf{f}_{1:t+1} = \text{Forward}(\mathbf{f}_{1:t}, \mathbf{e}_{t+1})$$

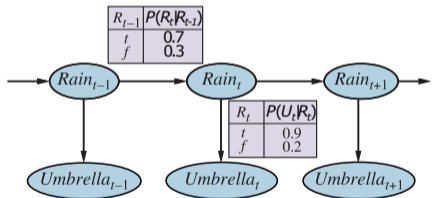
We bootstrap the process with  $\mathbf{f}_{1:0} = P(X_0)$

# Prediction and Filtering example

Umbrella scenario

$$P(X_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha P(\mathbf{e}_{t+1} \mid X_{t+1}) \sum_{x_t} P(X_{t+1} \mid x_t) P(x_t \mid \mathbf{e}_{1:t})$$

Let us compute  $P(R_2 \mid u_{1:2})$ ,



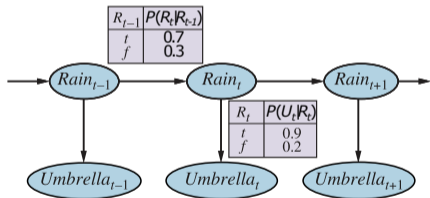
# Prediction and Filtering example

Umbrella scenario

$$P(X_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha P(\mathbf{e}_{t+1} \mid X_{t+1}) \sum_{x_t} P(X_{t+1} \mid x_t) P(x_t \mid \mathbf{e}_{1:t})$$

Let us compute  $P(R_2 \mid u_{1:2})$ ,  
assume  $P(R_0) = \langle 0.5, 0.5 \rangle$

Umbrella appears on day 1, so  $U_1 = \text{true}$

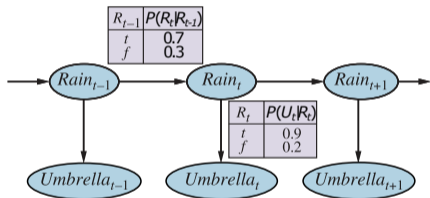


$$P(R_1 \mid u_1) = \alpha P(u_1 \mid R_1) P(R_1)$$

# Prediction and Filtering example

Umbrella scenario

$$P(X_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha P(\mathbf{e}_{t+1} \mid X_{t+1}) \sum_{x_t} P(X_{t+1} \mid x_t) P(x_t \mid \mathbf{e}_{1:t})$$



Let us compute  $P(R_2 \mid u_{1:2})$ ,  
assume  $P(R_0) = \langle 0.5, 0.5 \rangle$

Umbrella appears on day 1, so  $U_1 = true$

$$P(R_1 \mid u_1) = \alpha P(u_1 \mid R_1) P(R_1)$$

Prediction from  $t = 0$  to  $t = 1$

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

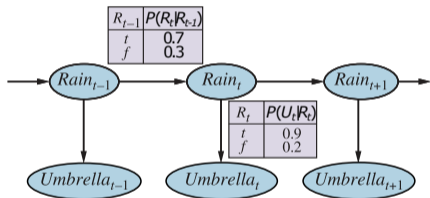
# Prediction and Filtering example

Umbrella scenario

$$P(X_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha P(\mathbf{e}_{t+1} \mid X_{t+1}) \sum_{x_t} P(X_{t+1} \mid x_t) P(x_t \mid \mathbf{e}_{1:t})$$

Let us compute  $P(R_2 \mid u_{1:2})$ ,  
assume  $P(R_0) = \langle 0.5, 0.5 \rangle$

Umbrella appears on day 1, so  $U_1 = \text{true}$



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

Prediction from  $t = 0$  to  $t = 1$

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

# Prediction and Filtering example

Umbrella scenario

$$P(X_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha P(\mathbf{e}_{t+1} \mid X_{t+1}) \sum_{x_t} P(X_{t+1} \mid x_t) P(x_t \mid \mathbf{e}_{1:t})$$

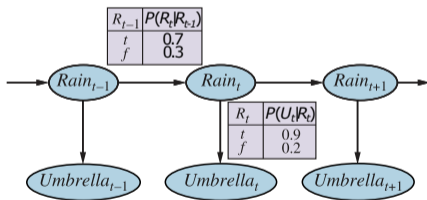
Let us compute  $P(R_2 \mid u_{1:2})$ ,

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

$P(R_1 \mid u_1) = \langle 0.818, 0.182 \rangle$

Umbrella appears on day 2, so  $U_2 = \text{true}$

$$P(R_2 \mid u_2, u_2) = \alpha P(u_2 \mid R_2) P(R_2 \mid u_1)$$

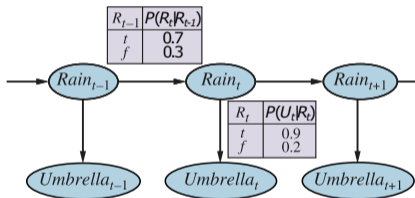


# Prediction and Filtering example

Umbrella scenario

$$P(X_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha P(\mathbf{e}_{t+1} \mid X_{t+1}) \sum_{x_t} P(X_{t+1} \mid x_t) P(x_t \mid \mathbf{e}_{1:t})$$

Let us compute  $P(R_2 \mid u_{1:2})$ ,  
assume  $P(R_0) = \langle 0.5, 0.5 \rangle$   
 $P(R_1 \mid u_1) = \langle 0.818, 0.182 \rangle$



Umbrella appears on day 2, so  $U_2 = true$

$$P(R_2 \mid u_2, u_2) = \alpha P(u_2 \mid R_2) P(R_2 \mid u_1)$$

Prediction from  $t = 1$  to  $t = 2$

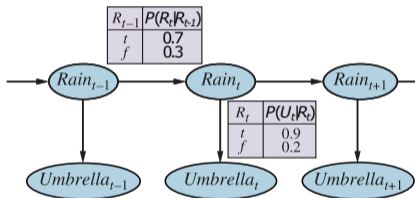
$$\begin{aligned} P(R_2 \mid u_1) &= \sum_{r_1} P(R_2 \mid r_1) P(r_1 \mid u_1) \\ &= \langle 0.7, 0.3 \rangle \times 0.818 + \langle 0.3, 0.7 \rangle \times 0.182 = \langle 0.627, 0.373 \rangle \end{aligned}$$

# Prediction and Filtering example

Umbrella scenario

$$P(X_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha P(\mathbf{e}_{t+1} \mid X_{t+1}) \sum_{x_t} P(X_{t+1} \mid x_t) P(x_t \mid \mathbf{e}_{1:t})$$

Let us compute  $P(R_2 \mid u_{1:2})$ ,  
assume  $P(R_0) = \langle 0.5, 0.5 \rangle$   
 $P(R_1 \mid u_1) = \langle 0.818, 0.182 \rangle$



Umbrella appears on day 2, so  $U_2 = true$

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

Prediction from  $t = 1$  to  $t = 2$

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

To continue in the next session.