Chapter 15

Probabilistic Reasoning Over Time

Motivating Examples

- Car diagnosis (static problem)
  - There exists uncertainty
  - We don't care about time
  - Whatever is broken remains broken during diagnosis

- Diabetes management (dynamic problem)
  - It's a dynamic problem with uncertainty
  - Variable values change over time
    - Blood sugar level, stomach contents, etc
    - Measured blood sugar, food eaten, insulin doses, etc
  - We must model time to estimate present states and predict future states of a patient

Motivating Examples

- Statistical modeling
  - Economy, population, weather, etc
- Robotics
  - Tracking the location and velocity of a robot
- Computer vision
  - Recognize human actions

The world changes; we need to track and predict it.

Representation and Notation

- Variable representation over time
  - Basic idea: copy state and evidence variables from each time step
  - $\mathbf{X}_t$ = set of unobservable state variables at time t
    e.g., $BloodSugar_t$, $StomachContents_t$, etc.
  - $\mathbf{E}_t$ = set of observable evidence variables at time t
    e.g., $MeasuredBloodSugar_t$, $PulseRate_t$, $FoodEaten_t$
- Discrete time representation
  - The world is viewed as a series of snapshots or time slices: $\mathbf{X}_{a:b} = \mathbf{X}_a, \mathbf{X}_{a+1}, \ldots, \mathbf{X}_{b-1}, \mathbf{X}_b$
  - Step size depends on problem

Markov Process (Markov Chains)

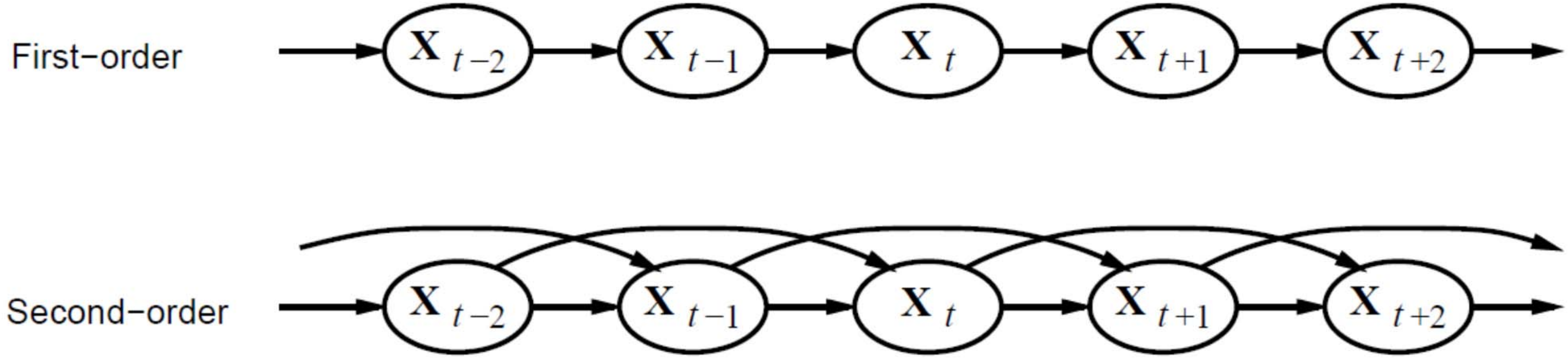
- Construct a Bayes net from variables over time
  - Transition model: how world evolves: $\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{0:t-1})$
  - Sensor model: how the evidence variables get their values: $\mathbf{P}(\mathbf{E}_t|\mathbf{X}_{0:t}, \mathbf{E}_{0:t-1})$
- Issue 1: $\mathbf{X}_{0:t-1}$ is unbounded in size as t increases

Markov Process (continued)

- Markov assumption: solution of issue 1
  - $\mathbf{X}_t$ depends on **bounded** subset of $\mathbf{X}_{0:t-1}$

First-order Markov process: $\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{0:t-1}) = \mathbf{P}(\mathbf{X}_t|\mathbf{X}_{t-1})$

Second-order Markov process: $\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{0:t-1}) = \mathbf{P}(\mathbf{X}_t|\mathbf{X}_{t-2}, \mathbf{X}_{t-1})$



Sensor Markov assumption: $\mathbf{P}(\mathbf{E}_t|\mathbf{X}_{0:t}, \mathbf{E}_{0:t-1}) = \mathbf{P}(\mathbf{E}_t|\mathbf{X}_t)$

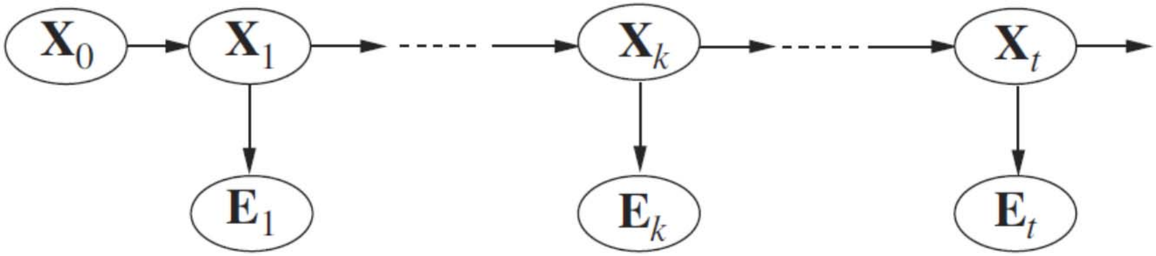
Markov Process (continued)

- Issue 2: Specify a different distribution for each time step?

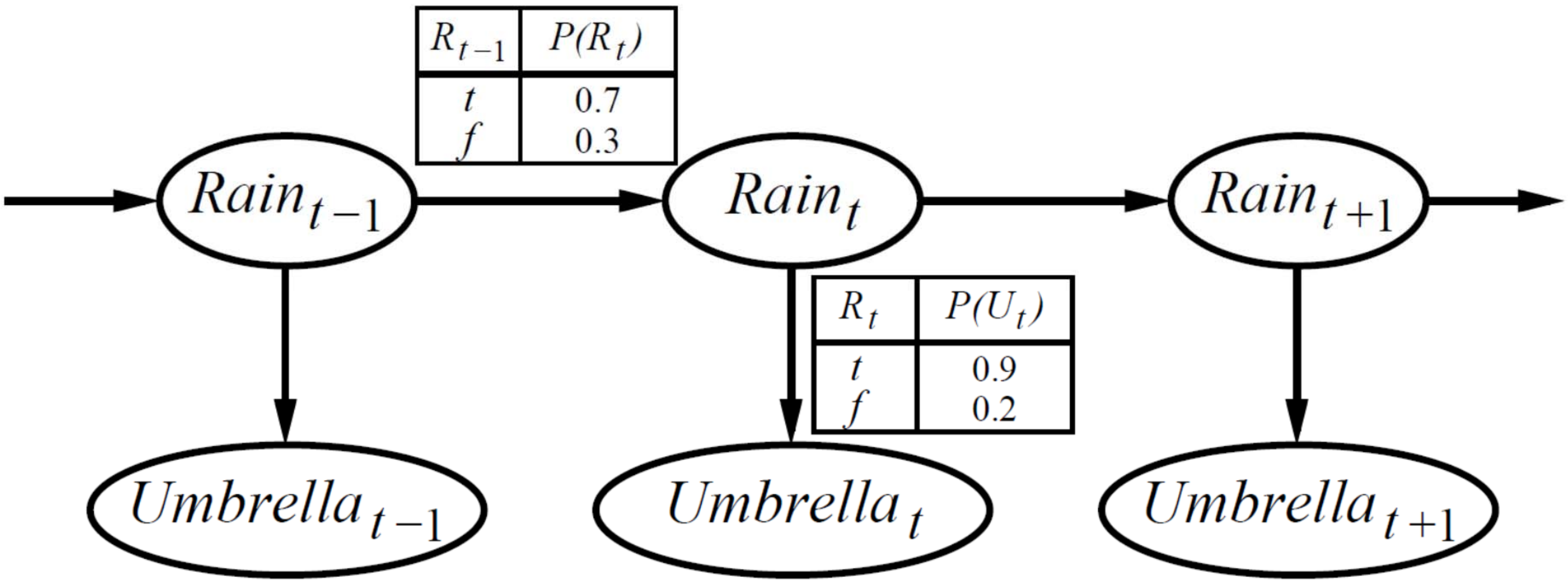
- Stationary process: solution of issue 2
  - Transition model $\mathbf{P}(\mathbf{X}_t|\mathbf{X}_{t-1})$ and sensor model $\mathbf{P}(\mathbf{E}_t|\mathbf{X}_t)$ are fixed for all t

- Joint probability over all variables: chain rule

$$\mathbf{P}(\mathbf{X}_{0:t}, \mathbf{E}_{1:t}) = \mathbf{P}(\mathbf{X}_0) \prod_{i=1}^{t} \mathbf{P}(\mathbf{X}_i \mid \mathbf{X}_{i-1}) \, \mathbf{P}(\mathbf{E}_i \mid \mathbf{X}_i)$$



Example: Markov Process



- First-order Markov assumption is not exactly true in real world
- Problem can be addressed by:
  - Increase order of Markov process
  - Augment states, e.g., add *season*, *temperature*, etc.

Inference Tasks

Filtering: $\mathbf{P}(\mathbf{X}_t|\mathbf{e}_{1:t})$

belief state—input to the decision process of a rational agent

Prediction: $\mathbf{P}(\mathbf{X}_{t+k}|\mathbf{e}_{1:t})$ for $k > 0$

evaluation of possible action sequences;
like filtering without the evidence

Smoothing: $\mathbf{P}(\mathbf{X}_k|\mathbf{e}_{1:t})$ for $0 \leq k < t$

better estimate of past states, essential for learning

Most likely explanation: $\arg\max_{\mathbf{x}_{1:t}} P(\mathbf{x}_{1:t}|\mathbf{e}_{1:t})$

speech recognition, decoding with a noisy channel

Filtering

- Objective: design a ***recursive*** state estimation algorithm

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

- Two-step process
  - Prediction: the current state distribution is projected ***forward*** from t to $t+1$
  - Update: the distribution is updated using the new evidence

$$\begin{aligned}\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \\ &= \alpha \mathbf{P}(\mathbf{e}_{t+1}|\mathbf{X}_{t+1}, \mathbf{e}_{1:t}) \mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t}) \\ &= \alpha \mathbf{P}(\mathbf{e}_{t+1}|\mathbf{X}_{t+1}) \mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t})\end{aligned}$$

Update Prediction

  - Prediction by summing out $\mathbf{X}_t$

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

Sensor model Transition model

Filtering (continued)

- View as massage passing

$$\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t+1}) = \alpha \mathbf{P}(\mathbf{e}_{t+1}|\mathbf{X}_{t+1}) \Sigma_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1}|\mathbf{x}_t) P(\mathbf{x}_t|\mathbf{e}_{1:t})$$

  – Consider $\mathbf{P}(\mathbf{X}_t|\mathbf{e}_{1:t})$ as a message $\mathbf{f}_{1:t}$ which is
    - Prorogated forward along the sequence
    - Modified by each transition
    - Updated by each new observation

$\mathbf{f}_{1:t+1} = \text{FORWARD}(\mathbf{f}_{1:t}, \mathbf{e}_{t+1})$ where $\mathbf{f}_{1:t} = \mathbf{P}(\mathbf{X}_t|\mathbf{e}_{1:t})$

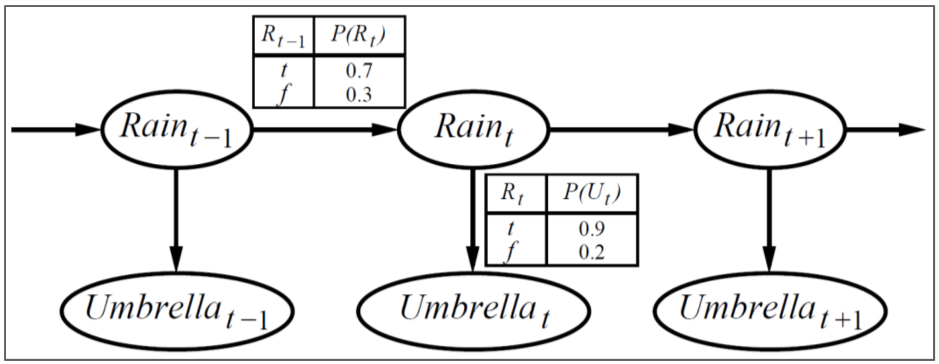
Time and space **constant** (independent of t)

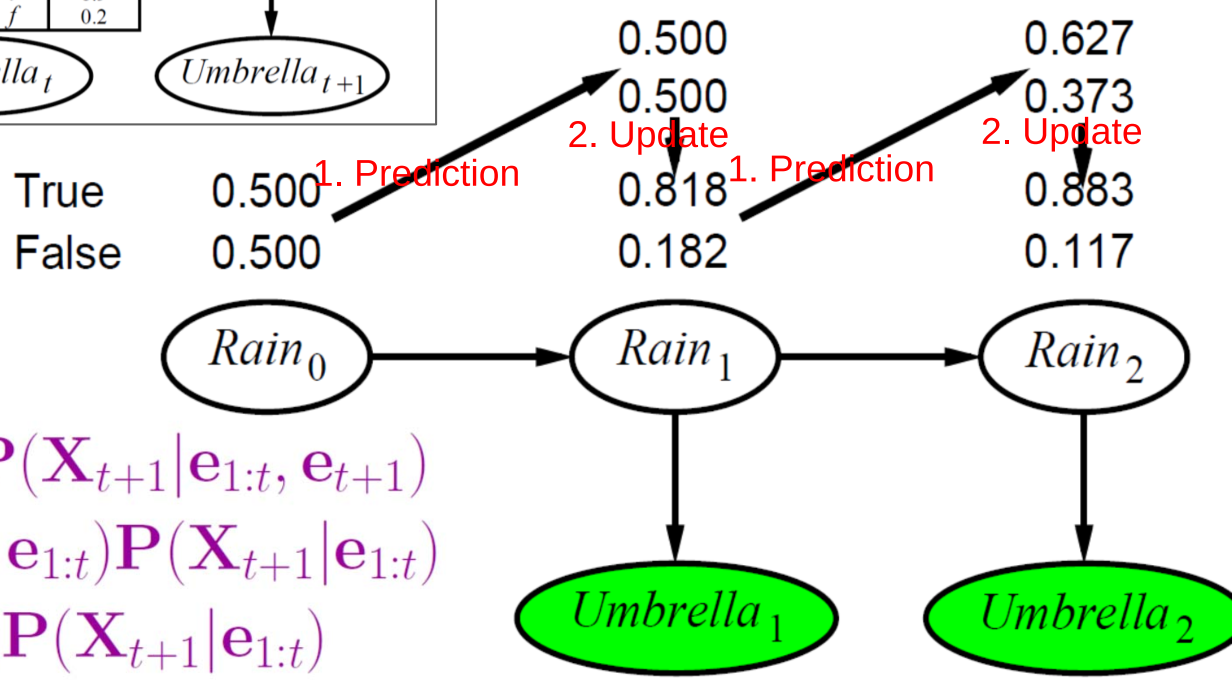
Example: Filtering

- View as massage passing

$$\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t+1}) = \alpha\mathbf{P}(\mathbf{e}_{t+1}|\mathbf{X}_{t+1})\Sigma_{\mathbf{x}_t}\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{x}_t)P(\mathbf{x}_t|\mathbf{e}_{1:t})$$

2. Update — 1. Prediction





$$\begin{aligned}\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \\ &= \alpha\mathbf{P}(\mathbf{e}_{t+1}|\mathbf{X}_{t+1}, \mathbf{e}_{1:t})\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t}) \\ &= \alpha\mathbf{P}(\mathbf{e}_{t+1}|\mathbf{X}_{t+1})\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t})\end{aligned}$$

Prediction

- Prediction can be viewed as filtering without the addition of new evidence
- Predication can be recursively computed by:

$$\mathbf{P}(\mathbf{X}_{t+k+1} \mid \mathbf{e}_{1:t}) = \sum_{\mathbf{x}_{t+k}} \mathbf{P}(\mathbf{X}_{t+k+1} \mid \mathbf{x}_{t+k}) P(\mathbf{x}_{t+k} \mid \mathbf{e}_{1:t})$$