Lecture notes for *Principles of Artificial Intelligence* (COMS 4720/5720)
Yan-Bin Jia, Iowa State University

Inference in Temporal Models

Outline

I. Prediction

II. Smoothing

III. Most Likely Explanation

* Figures/images are from the textbook site or the instructor unless the source is cited specifically.

I. Prediction

Prediction is essentially filtering without the addition of new evidence.

Only prediction and no update at every time step.

For $k = 0,1,\ldots$

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

I. Prediction

Prediction is essentially filtering without the addition of new evidence.

Only prediction and no update at every time step.

For $k = 0,1,\ldots$

$$\boldsymbol{P}(\boldsymbol{X}_{t+k+1} \mid \boldsymbol{e}_{1:t}) = \sum_{\boldsymbol{x}_{t+k}} \underbrace{\boldsymbol{P}(\boldsymbol{X}_{t+k+1} \mid \boldsymbol{x}_{t+k})}_{\text{transition model}} \underbrace{P(\boldsymbol{x}_{t+k} \mid \boldsymbol{e}_{1:t})}_{\text{recursion}}$$

$$\boldsymbol{P}(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t}) = \alpha \underbrace{\boldsymbol{P}(\boldsymbol{e}_t \mid \boldsymbol{X}_t)}_{\text{sensor model}} \sum_{\boldsymbol{x}_{t-1}} \boldsymbol{P}(\boldsymbol{X}_t \mid \boldsymbol{x}_{t-1}) P(\boldsymbol{x}_{t-1} \mid \boldsymbol{e}_{1:t-1}) \quad \text{(filtering)}$$

I. Prediction

Prediction is essentially filtering without the addition of new evidence.

Only prediction and no update at every time step.

For $k = 0,1,\ldots$

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

$$\mathbf{P}(\mathbf{X}_t \mid \mathbf{e}_{1:t}) = \alpha \underbrace{\mathbf{P}(\mathbf{e}_t \mid \mathbf{X}_t)}_{\text{sensor model}} \sum_{\mathbf{x}_{t-1}} \mathbf{P}(\mathbf{X}_t \mid \mathbf{x}_{t-1}) P(\mathbf{x}_{t-1} \mid \mathbf{e}_{1:t-1}) \quad \text{(filtering)}$$

$$\mathbf{P}(\mathbf{X}_t \mid \mathbf{e}_{1:t}) \to \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \to \cdots \to \mathbf{P}(\mathbf{X}_{t+k} \mid \mathbf{e}_{1:t}) \to \mathbf{P}(\mathbf{X}_{t+k+1} \mid \mathbf{e}_{1:t})$$

I. Prediction

Prediction is essentially filtering without the addition of new evidence.

Only prediction and no update at every time step.

For $k = 0, 1, \ldots$

$$\boldsymbol{P}(\boldsymbol{X}_{t+k+1} \mid \boldsymbol{e}_{1:t}) = \sum_{\boldsymbol{x}_{t+k}} \underbrace{\boldsymbol{P}(\boldsymbol{X}_{t+k+1} \mid \boldsymbol{x}_{t+k})}_{\text{transition model}} \underbrace{P(\boldsymbol{x}_{t+k} \mid \boldsymbol{e}_{1:t})}_{\text{recursion}}$$

$$\boldsymbol{P}(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t}) = \alpha \underbrace{\boldsymbol{P}(\boldsymbol{e}_t \mid \boldsymbol{X}_t)}_{\text{sensor model}} \sum_{\boldsymbol{x}_{t-1}} \boldsymbol{P}(\boldsymbol{X}_t \mid \boldsymbol{x}_{t-1}) P(\boldsymbol{x}_{t-1} \mid \boldsymbol{e}_{1:t-1}) \quad \text{(filtering)}$$

$$\boldsymbol{P}(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t}) \to \boldsymbol{P}(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t}) \to \cdots \to \boldsymbol{P}(\boldsymbol{X}_{t+k} \mid \boldsymbol{e}_{1:t}) \to \boldsymbol{P}(\boldsymbol{X}_{t+k+1} \mid \boldsymbol{e}_{1:t})$$

- As k increases, the distribution will converge to the stationary distribution of the Markov process defined by the transition model.

I. Prediction

Prediction is essentially filtering without the addition of new evidence.

Only prediction and no update at every time step.

For $k = 0,1,\ldots$

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

$$\mathbf{P}(\mathbf{X}_t \mid \mathbf{e}_{1:t}) = \alpha \underbrace{\mathbf{P}(\mathbf{e}_t \mid \mathbf{X}_t)}_{\text{sensor model}} \sum_{\mathbf{x}_{t-1}} \mathbf{P}(\mathbf{X}_t \mid \mathbf{x}_{t-1}) P(\mathbf{x}_{t-1} \mid \mathbf{e}_{1:t-1}) \quad \text{(filtering)}$$

$$\mathbf{P}(\mathbf{X}_t \mid \mathbf{e}_{1:t}) \to \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \to \cdots \to \mathbf{P}(\mathbf{X}_{t+k} \mid \mathbf{e}_{1:t}) \to \mathbf{P}(\mathbf{X}_{t+k+1} \mid \mathbf{e}_{1:t})$$

- As k increases, the distribution will converge to the stationary distribution of the Markov process defined by the transition model.
- The value of k at which convergence happens is called the *mixing time*, which has been well studied.

I. Prediction

Prediction is essentially filtering without the addition of new evidence.

Only prediction and no update at every time step.

For $k = 0, 1, \ldots$

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

$$\mathbf{P}(\mathbf{X}_t \mid \mathbf{e}_{1:t}) = \alpha \underbrace{\mathbf{P}(\mathbf{e}_t \mid \mathbf{X}_t)}_{\text{sensor model}} \sum_{\mathbf{x}_{t-1}} \mathbf{P}(\mathbf{X}_t \mid \mathbf{x}_{t-1}) P(\mathbf{x}_{t-1} \mid \mathbf{e}_{1:t-1}) \quad \text{(filtering)}$$

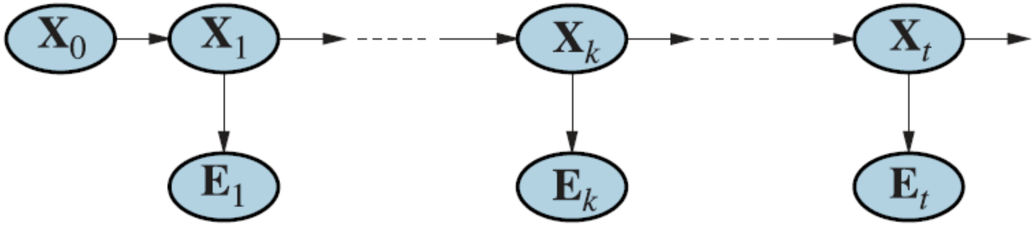
$$\mathbf{P}(\mathbf{X}_t \mid \mathbf{e}_{1:t}) \rightarrow \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \rightarrow \cdots \rightarrow \mathbf{P}(\mathbf{X}_{t+k} \mid \mathbf{e}_{1:t}) \rightarrow \mathbf{P}(\mathbf{X}_{t+k+1} \mid \mathbf{e}_{1:t})$$

- As k increases, the distribution will converge to the stationary distribution of the Markov process defined by the transition model.
- The value of k at which convergence happens is called the *mixing time*, which has been well studied.
- The more uncertainty, the shorter will be the mixing time.

II. Smoothing

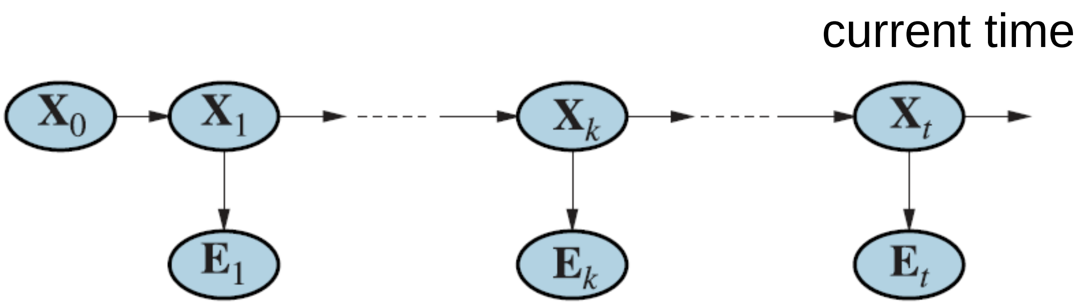
Compute $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t})$ for some $0 \leq k < t$.

current time



II. Smoothing

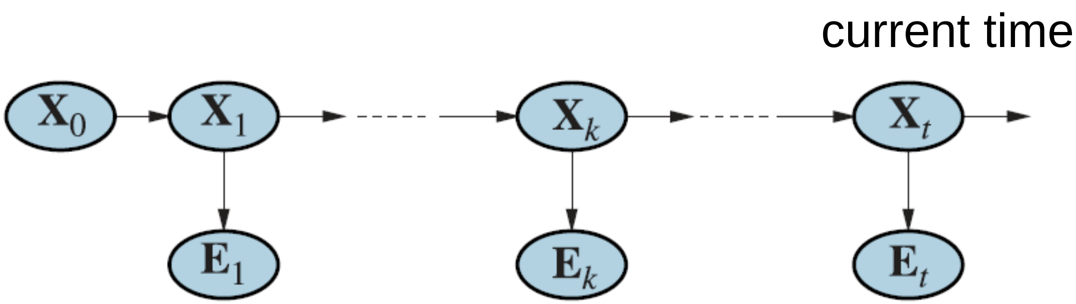
Compute $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t})$ for some $0 \le k < t$.



- Split the evidence into two parts: up to k and from $k+1$ to t.

II. Smoothing

Compute $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t})$ for some $0 \le k < t$.

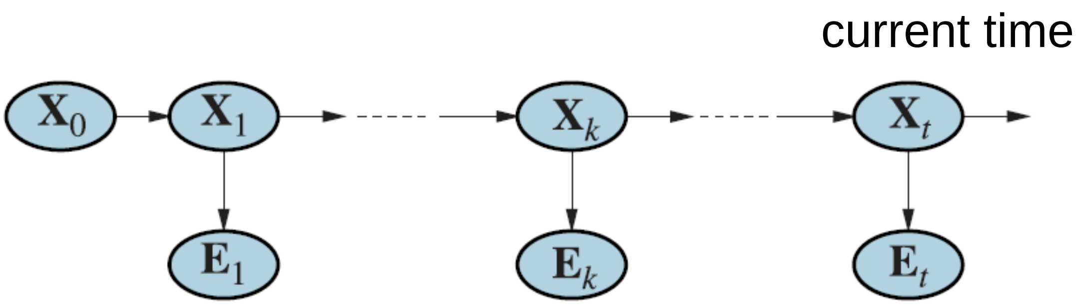


- Split the evidence into two parts: up to k and from $k+1$ to t.

$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}, \boldsymbol{e}_{k+1:t})$$

II. Smoothing

Compute $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t})$ for some $0 \le k < t$.



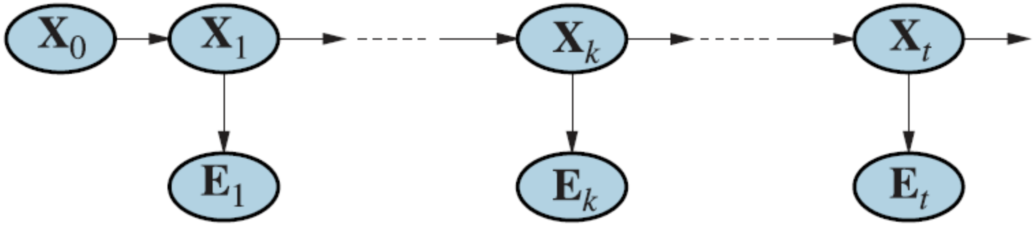
- Split the evidence into two parts: up to k and from $k+1$ to t.

$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}, \boldsymbol{e}_{k+1:t})$$

$$= \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k, \boldsymbol{e}_{1:k}) \quad \text{(Bayes' rule, given } \boldsymbol{e}_{1:k}\text{)}$$

II. Smoothing

Compute $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t})$ for some $0 \le k < t$.



- Split the evidence into two parts: up to k and from $k+1$ to t.

$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}, \boldsymbol{e}_{k+1:t})$$

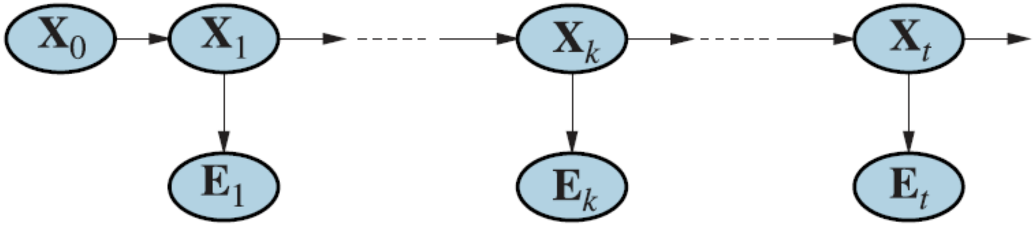
$$= \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k, \boldsymbol{e}_{1:k}) \quad \text{(Bayes' rule, given } \boldsymbol{e}_{1:k}\text{)}$$

$$= \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k) \quad \text{(conditional independence)}$$

II. Smoothing

Compute $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t})$ for some $0 \leq k < t$.

current time



- Split the evidence into two parts: up to k and from $k + 1$ to t.

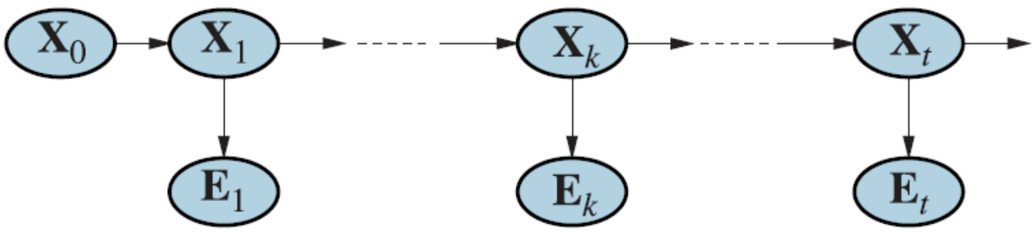
$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}, \boldsymbol{e}_{k+1:t})$$

$$= \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k, \boldsymbol{e}_{1:k}) \quad \text{(Bayes' rule, given } \boldsymbol{e}_{1:k}\text{)}$$

$$= \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \underbrace{\boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)}_{\text{“backward” message}} \quad \text{(conditional independence)}$$

II. Smoothing

Compute $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t})$ for some $0 \le k < t$.



- Split the evidence into two parts: up to k and from $k+1$ to t.

$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}, \boldsymbol{e}_{k+1:t})$$

$$= \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k})\boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k, \boldsymbol{e}_{1:k}) \quad \text{(Bayes' rule, given } \boldsymbol{e}_{1:k}\text{)}$$

$$= \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k})\underbrace{\boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)}_{\text{"backward" message}} \quad \text{(conditional independence)}$$

- The backward message can also be computed recursively.

$$\boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k) = \sum_{\boldsymbol{x}_{k+1}} \underbrace{P(\boldsymbol{e}_{k+1} \mid \boldsymbol{x}_{k+1})}_{\text{sensor model}} \underbrace{P(\boldsymbol{e}_{k+2:t} \mid \boldsymbol{x}_{k+1})}_{\text{recursion}} \underbrace{\boldsymbol{P}(\boldsymbol{x}_{k+1} \mid \boldsymbol{X}_k)}_{\text{transition model}}$$

Computation

Equation for smoothing ($0 \leq k < t$):

$$\mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:t}) = \alpha \underbrace{\mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:k})}_{\text{Old estimate}} \underbrace{\mathbf{P}(\mathbf{e}_{k+1:t} \mid \mathbf{X}_k)}_{\text{Rescaling factors}}$$

Computation

Equation for smoothing ($0 \leq k < t$):

$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \alpha \underbrace{\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k})}_{\text{Old estimate}} \underbrace{\boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)}_{\text{Rescaling factors}}$$

- Forward computation (filtering)

$$\boldsymbol{P}(\boldsymbol{X}_0) \to \boldsymbol{P}(\boldsymbol{X}_1 \mid \boldsymbol{e}_{1:1}) \to \cdots \to \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k})$$

where, for $0 \leq i \leq k-1$,

$$\boldsymbol{P}(\boldsymbol{X}_{i+1} \mid \boldsymbol{e}_{1:i+1}) = \alpha \boldsymbol{P}(\boldsymbol{e}_{i+1} \mid \boldsymbol{X}_{i+1}) \sum_{\boldsymbol{x}_i} \boldsymbol{P}(\boldsymbol{X}_{i+1} \mid \boldsymbol{x}_i) \boldsymbol{P}(\boldsymbol{x}_i \mid \boldsymbol{e}_{1:i})$$

Computation

Equation for smoothing ($0 \le k < t$):

$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \alpha \underbrace{\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k})}_{\text{Old estimate}} \underbrace{\boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)}_{\text{Rescaling factors}}$$

- Forward computation (filtering)

$$\boldsymbol{P}(\boldsymbol{X}_0) \to \boldsymbol{P}(\boldsymbol{X}_1 \mid \boldsymbol{e}_{1:1}) \to \cdots \to \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k})$$

where, for $0 \le i \le k-1$,

$$\boldsymbol{P}(\boldsymbol{X}_{i+1} \mid \boldsymbol{e}_{1:i+1}) = \alpha \boldsymbol{P}(\boldsymbol{e}_{i+1} \mid \boldsymbol{X}_{i+1}) \sum_{\boldsymbol{x}_i} \boldsymbol{P}(\boldsymbol{X}_{i+1} \mid \boldsymbol{x}_i) \boldsymbol{P}(\boldsymbol{x}_i \mid \boldsymbol{e}_{1:i})$$

- Backward computation (rescaling)

$$\underbrace{\boldsymbol{P}(\boldsymbol{e}_{t+1:t} \mid \boldsymbol{X}_t)} \to \boldsymbol{P}(\boldsymbol{e}_{t:t} \mid \boldsymbol{X}_{t-1}) \to \cdots \to \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$$

$= \boldsymbol{P}(\ \mid \boldsymbol{X}_t) = \mathbf{1}$ (vector of 1s since $\boldsymbol{e}_{t+1:t}$ is an empty sequence)

where, for $k \le j < t$,

$$\boldsymbol{P}(\boldsymbol{e}_{j+1:t} \mid \boldsymbol{X}_j) = \sum_{\boldsymbol{x}_{j+1}} P(\boldsymbol{e}_{j+1} \mid \boldsymbol{x}_{j+1})\, P(\boldsymbol{e}_{j+2:t} \mid \boldsymbol{x}_{j+1}) \boldsymbol{P}(\boldsymbol{x}_{j+1} \mid \boldsymbol{X}_j)$$

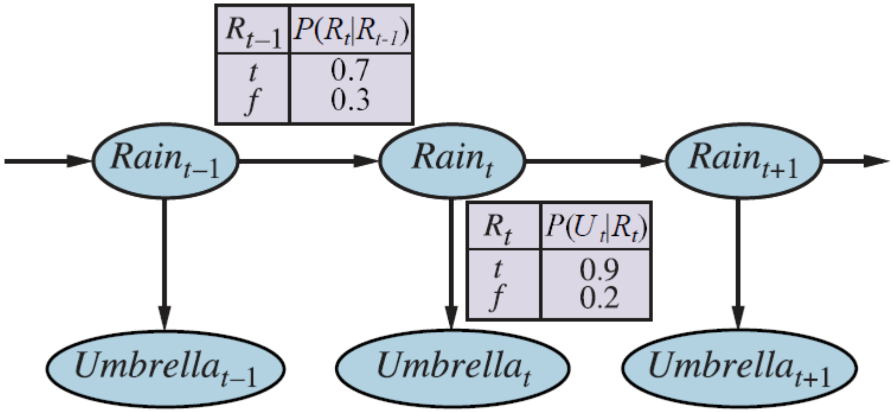
Smoothing in the Umbrella World

Compute $\boldsymbol{P}(R_1 \mid u_1, u_2)$ as follows:

probability of rain on day 1, given that umbrellas were observed on days 1 and 2.

$u_1 \equiv (U_1 = \textit{true})$

$u_2 \equiv (U_2 = \textit{true})$



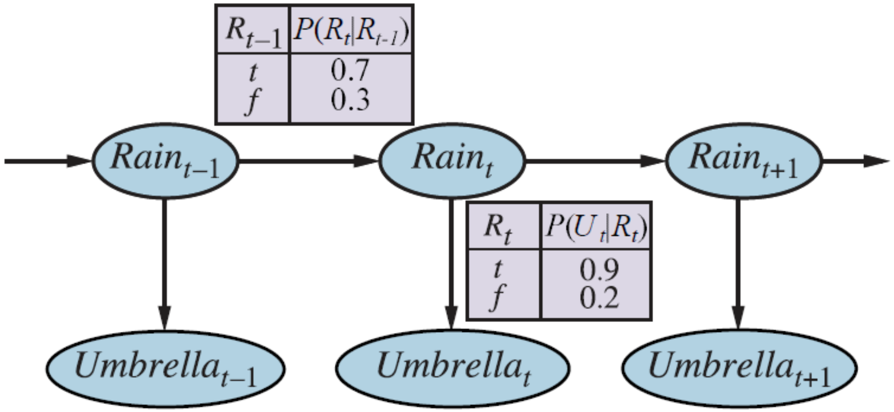
Smoothing in the Umbrella World

Compute $\boldsymbol{P}(R_1 \mid u_1, u_2)$ as follows:

probability of rain on day 1, given that umbrellas were observed on days 1 and 2.

$u_1 \equiv (U_1 = \textit{true})$

$u_2 \equiv (U_2 = \textit{true})$



$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$$

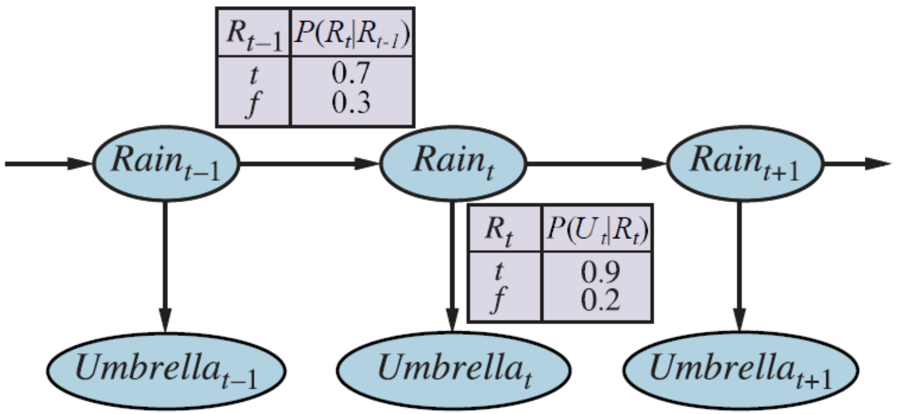
Smoothing in the Umbrella World

Compute $\boldsymbol{P}(R_1 \mid u_1, u_2)$ as follows:

probability of rain on day 1, given that umbrellas were observed on days 1 and 2.

$u_1 \equiv (U_1 = \mathit{true})$

$u_2 \equiv (U_2 = \mathit{true})$



$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$$

$t = 2$

$k = 1$

$$\boldsymbol{P}(R_1 \mid u_1, u_2) = \alpha \boldsymbol{P}(R_1 \mid u_1) \boldsymbol{P}(u_2 \mid R_1)$$

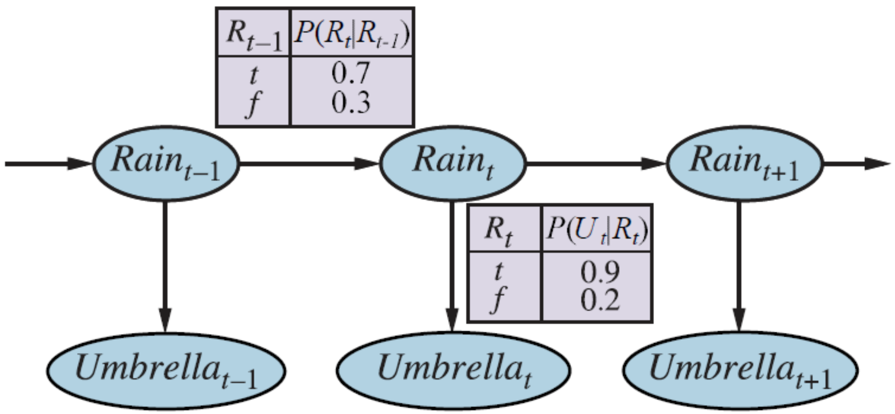
Smoothing in the Umbrella World

Compute $\boldsymbol{P}(R_1 \mid u_1, u_2)$ as follows:

probability of rain on day 1, given that umbrellas were observed on days 1 and 2.

$u_1 \equiv (U_1 = \textit{true})$

$u_2 \equiv (U_2 = \textit{true})$



$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$$

$t = 2$

$k = 1$

$$\boldsymbol{P}(R_1 \mid u_1, u_2) = \alpha \boldsymbol{P}(R_1 \mid u_1) \boldsymbol{P}(u_2 \mid R_1)$$

$$= \alpha \boldsymbol{P}(u_1 \mid R_1) \boldsymbol{P}(R_1) \approx \langle 0.818, 0.182 \rangle$$

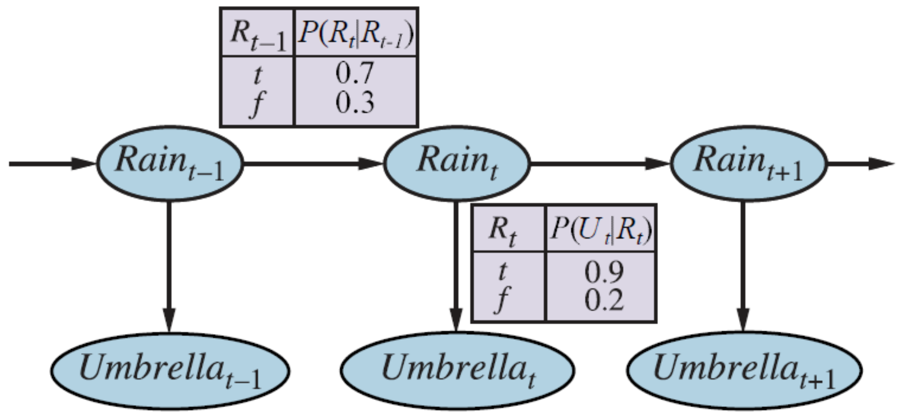
Smoothing in the Umbrella World

Compute $\boldsymbol{P}(R_1 \mid u_1, u_2)$ as follows:

probability of rain on day 1, given that umbrellas were observed on days 1 and 2.

$u_1 \equiv (U_1 = \textit{true})$

$u_2 \equiv (U_2 = \textit{true})$



$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$$

$\Downarrow$ $t = 2$, $k = 1$

$$\boldsymbol{P}(R_1 \mid u_1, u_2) = \alpha \boldsymbol{P}(R_1 \mid u_1) \boldsymbol{P}(u_2 \mid R_1)$$

$$= \alpha \boldsymbol{P}(u_1 \mid R_1) \boldsymbol{P}(R_1) \approx \langle 0.818, 0.182 \rangle$$

$$\boldsymbol{P}(u_2 \mid R_1) = \sum_{r_2' \in \{r_2, \neg r_2\}} P(u_2 \mid r_2') P(\mid r_2') \boldsymbol{P}(r_2' \mid R_1)$$

$$= (0.9 \cdot 1 \cdot \langle 0.7, 0.3 \rangle) + (0.2 \cdot 1 \cdot \langle 0.3, 0.7 \rangle)$$

$$= \langle 0.69, 0.41 \rangle$$

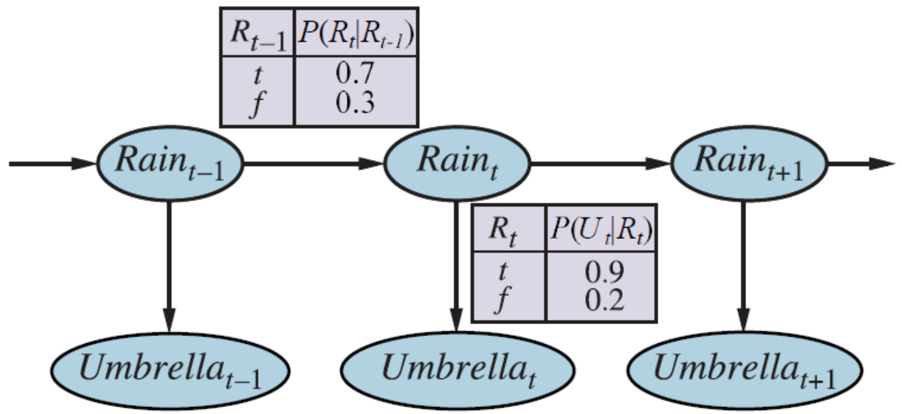
Smoothing in the Umbrella World

Compute $\boldsymbol{P}(R_1 \mid u_1, u_2)$ as follows:

probability of rain on day 1, given that umbrellas were observed on days 1 and 2.

$u_1 \equiv (U_1 = \textit{true})$

$u_2 \equiv (U_2 = \textit{true})$



$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$$

$t = 2$

$k = 1$

$$\boldsymbol{P}(R_1 \mid u_1, u_2) = \alpha \boldsymbol{P}(R_1 \mid u_1) \boldsymbol{P}(u_2 \mid R_1)$$

$$= \alpha \boldsymbol{P}(u_1 \mid R_1) \boldsymbol{P}(R_1) \approx \langle 0.818, 0.182 \rangle$$

$$\boldsymbol{P}(u_2 \mid R_1) = \sum_{r'_2 \in \{r_2, \neg r_2\}} P(u_2 \mid r'_2) P(\mid r'_2) \boldsymbol{P}(r'_2 \mid R_1)$$

$$= (0.9 \cdot 1 \cdot \langle 0.7, 0.3 \rangle) + (0.2 \cdot 1 \cdot \langle 0.3, 0.7 \rangle)$$

$$= \langle 0.69, 0.41 \rangle$$

$$\approx \alpha \langle 0.818, 0.182 \rangle \times \langle 0.69, 0.41 \rangle$$

$$\approx \langle 0.883, 0.117 \rangle$$

Intuition Behind Smoothing

$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$$

Intuition Behind Smoothing

$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$$

- The old estimate $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k})$, which consists of the probabilities for individual values of the state at time k, was obtained based on the evidences from time 1 through k.

Intuition Behind Smoothing

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

- The old estimate $\mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:k})$, which consists of the probabilities for individual values of the state at time k, was obtained based on the evidences from time 1 through k.

- Now we have gathered additional evidences (from time $k+1$ to t), so we hope to polish the old state estimate.

Intuition Behind Smoothing

$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k})\boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$$

- The old estimate $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k})$, which consists of the probabilities for individual values of the state at time k, was obtained based on the evidences from time 1 through k.

- Now we have gathered additional evidences (from time $k+1$ to t), so we hope to polish the old state estimate.

- Polishing is done by rescaling the probabilities in the distribution $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k})$ to get a new distribution $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t})$.

Intuition Behind Smoothing

$$\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \alpha \boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$$

- The old estimate $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k})$, which consists of the probabilities for individual values of the state at time k, was obtained based on the evidences from time 1 through k.

- Now we have gathered additional evidences (from time $k+1$ to t), so we hope to polish the old state estimate.

- Polishing is done by rescaling the probabilities in the distribution $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:k})$ to get a new distribution $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t})$.

- $\boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{X}_k)$ gives the scaling factors for all possible values based on how likely each would result in a sequence of states from time $k+1$ to t to produce the sequence of evidences $\boldsymbol{e}_{k+1:t}$.

III. Most Likely Explanation

Umbrella sequence on the guard's first five days: [*true*, *true*, *false*, *true*, *true*]

Question What weather sequence is most likely?

III. Most Likely Explanation

Umbrella sequence on the guard's first five days: [*true*, *true*, *false*, *true*, *true*]

Question What weather sequence is most likely?

- ♣ 2^5 possible weather sequences to pick from.

III. Most Likely Explanation

Umbrella sequence on the guard's first five days: [*true*, *true*, *false*, *true*, *true*]

Question What weather sequence is most likely?

♣ 2^5 possible weather sequences to pick from.

♠ Use smoothing to find posterior distribution for weather at each time step, and select at the step the weather that scores the highest.

- Posterior distribution by smoothing are distributed over single time steps.
- Joint probabilities over all the time steps must be considered to find the most likely sequence.

III. Most Likely Explanation

Umbrella sequence on the guard's first five days: [*true*, *true*, *false*, *true*, *true*]

Question What weather sequence is most likely?

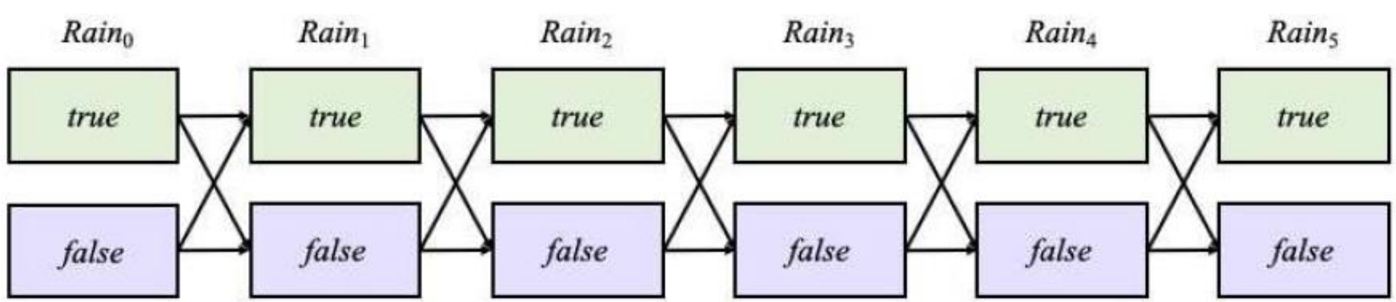
- ♣ 2^5 possible weather sequences to pick from.
- ♠ Use smoothing to find posterior distribution for weather at each time step, and select at the step the weather that scores the highest.
  - Posterior distribution by smoothing are distributed over single time steps.
  - Joint probabilities over all time steps must be considered to find the most likely sequence.

III. Most Likely Explanation

Umbrella sequence on the guard's first five days: [*true*, *true*, *false*, *true*, *true*]

Question What weather sequence is most likely?

- ♣ 2^5 possible weather sequences to pick from.
- ♠ Use smoothing to find … ster… distribution for weather at each time step, and select at the step … ther that scores the highest.
  - Posterior distribution … othing are distributed over single time steps.
  - Joint probabilities o… … me steps must be considered to find the most likely sequence.
- ♦ View each sequence as a *path* through a graph whose nodes are the possible states at each time step.

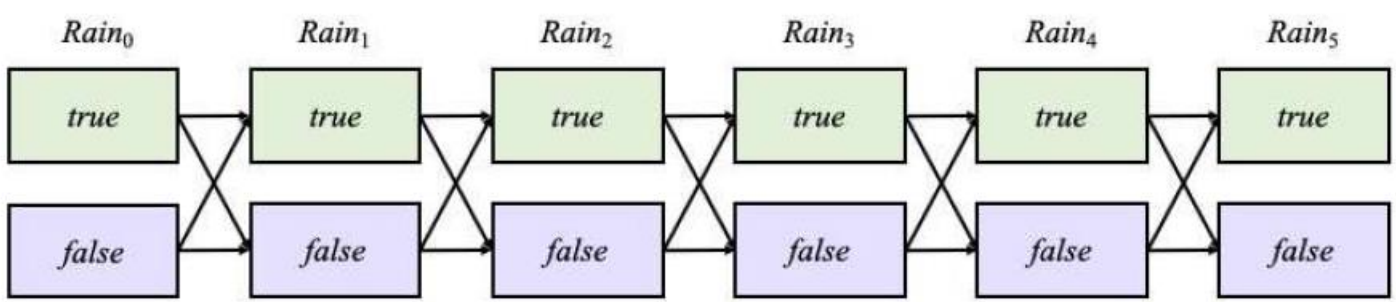


III. Most Likely Explanation

Umbrella sequence on the guard's first five days: [*true*, *true*, *false*, *true*, *true*]

Question What weather sequence is most likely?

- 2^5 possible weather sequences to pick from.
- Use smoothing to find [illegible] distribution for weather at each time step, and select at the step [illegible]ther that scores the highest.
  - Posterior distribution [illegible]othing are distributed over single time steps.
  - Joint probabilities o[illegible]me steps must be considered to find the most likely sequence.
- View each sequence as a *path* through a graph whose nodes are the possible states at each time step.



- Find the most likely path.

Computing the Most Likely Path

Likelihood of a path is the product of

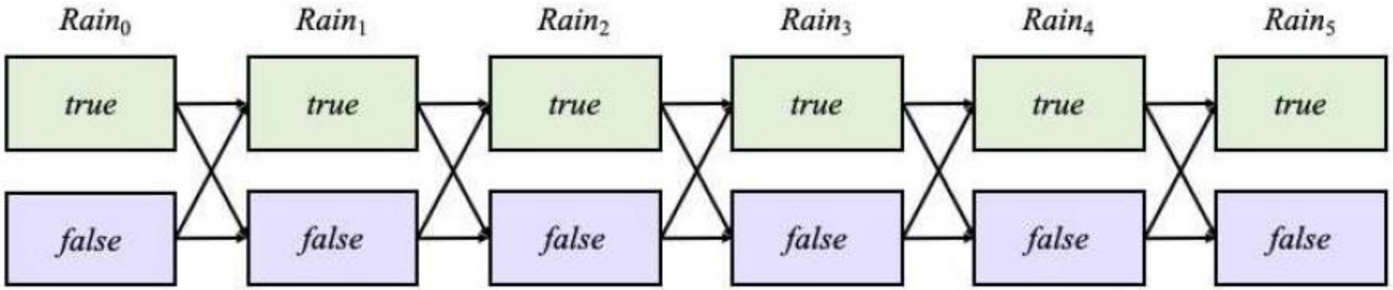
- the transition probabilities along the path, and
- the probabilities of the given observations at each state.

Computing the Most Likely Path

Likelihood of a path is the product of

- the transition probabilities along the path, and
- the probabilities of the given observations at each state.

Consider paths that reach the state $Rain_5 = true$

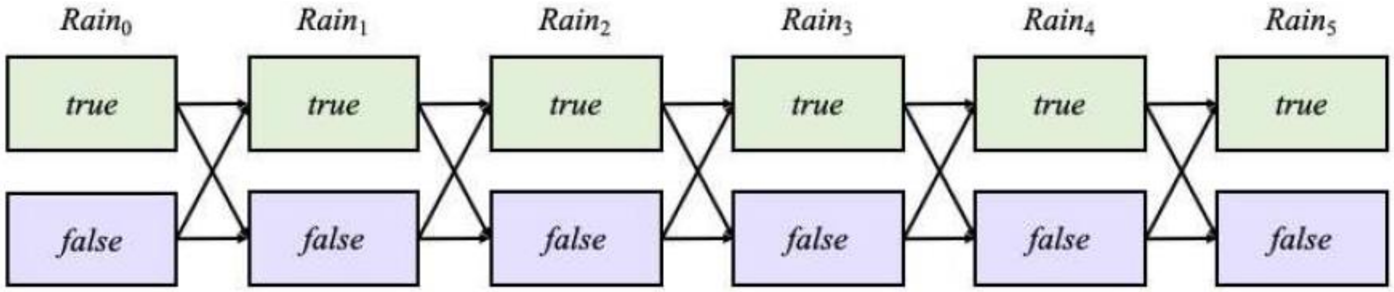


Computing the Most Likely Path

Likelihood of a path is the product of

- the transition probabilities along the path, and
- the probabilities of the given observations at each state.

Consider paths that reach the state $Rain_5 = true$



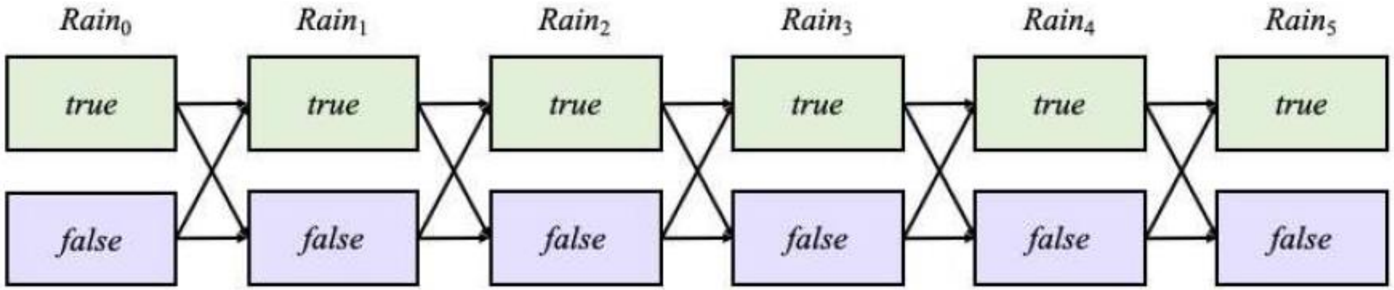
- The most likely path to $Rain_5 = true$ consists of the most likely path to some state at time 4 (optimal substructure) followed by a transition.

Computing the Most Likely Path

Likelihood of a path is the product of

- the transition probabilities along the path, and
- the probabilities of the given observations at each state.

Consider paths that reach the state $Rain_5 = true$



- The most likely path to $Rain_5 = true$ consists of the most likely path to some state at time 4 (optimal substructure) followed by a transition.
- The state at time 4 that gets chosen maximizes the likelihood of the path to $Rain_5 = true$.

Recurrence

Define the message

$$\boldsymbol{m}_{1:k} \equiv \max_{\boldsymbol{x}_{1:k-1}} \boldsymbol{P}(\boldsymbol{x}_{1:k-1}, \boldsymbol{X}_k, \boldsymbol{e}_{1:k})$$

Recurrence

Define the message

$$\boldsymbol{m}_{1:k} \equiv \max_{\boldsymbol{x}_{1:k-1}} \boldsymbol{P}(\boldsymbol{x}_{1:k-1}, \boldsymbol{X}_k, \boldsymbol{e}_{1:k})$$

We derive a recurrence for the term as follows:

$$\boldsymbol{m}_{1:k+1} = \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k+1})$$

Recurrence

Define the message

$$\boldsymbol{m}_{1:k} \equiv \max_{\boldsymbol{x}_{1:k-1}} \boldsymbol{P}(\boldsymbol{x}_{1:k-1}, \boldsymbol{X}_k, \boldsymbol{e}_{1:k})$$

We derive a recurrence for the term as follows:

$$\boldsymbol{m}_{1:k+1} = \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k+1})$$

$$= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}, e_{k+1})$$

Recurrence

Define the message

$$\boldsymbol{m}_{1:k} \equiv \max_{\boldsymbol{x}_{1:k-1}} \boldsymbol{P}(\boldsymbol{x}_{1:k-1}, \boldsymbol{X}_k, \boldsymbol{e}_{1:k})$$

We derive a recurrence for the term as follows:

$$\begin{aligned}
\boldsymbol{m}_{1:k+1} &= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k+1}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}, e_{k+1}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(e_{k+1} \mid \boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k})
\end{aligned}$$

Recurrence

Define the message

$$\boldsymbol{m}_{1:k} \equiv \max_{\boldsymbol{x}_{1:k-1}} \boldsymbol{P}(\boldsymbol{x}_{1:k-1}, \boldsymbol{X}_k, \boldsymbol{e}_{1:k})$$

We derive a recurrence for the term as follows:

$$\begin{aligned}
\boldsymbol{m}_{1:k+1} &= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k+1}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}, e_{k+1}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(e_{k+1} \mid \boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(e_{k+1} \mid \boldsymbol{X}_{k+1}) \boldsymbol{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{x}_k) P(\boldsymbol{x}_{1:k}, \boldsymbol{e}_{1:k})
\end{aligned}$$

Recurrence

Define the message

$$\boldsymbol{m}_{1:k} \equiv \max_{\boldsymbol{x}_{1:k-1}} \boldsymbol{P}(\boldsymbol{x}_{1:k-1}, \boldsymbol{X}_k, \boldsymbol{e}_{1:k})$$

We derive a recurrence for the term as follows:

$$\begin{aligned}
\boldsymbol{m}_{1:k+1} &= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k+1}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}, e_{k+1}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(e_{k+1} \mid \boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(e_{k+1} \mid \boldsymbol{X}_{k+1}) \boldsymbol{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{x}_k) P(\boldsymbol{x}_{1:k}, \boldsymbol{e}_{1:k}) \\
&= \boldsymbol{P}(e_{k+1} \mid \boldsymbol{X}_{k+1}) \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{x}_k) P(\boldsymbol{x}_{1:k}, \boldsymbol{e}_{1:k})
\end{aligned}$$

Recurrence

Define the message

$$\boldsymbol{m}_{1:k} \equiv \max_{\boldsymbol{x}_{1:k-1}} \boldsymbol{P}(\boldsymbol{x}_{1:k-1}, \boldsymbol{X}_k, \boldsymbol{e}_{1:k})$$

We derive a recurrence for the term as follows:

$$\begin{aligned}
\boldsymbol{m}_{1:k+1} &= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k+1}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}, e_{k+1}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(e_{k+1} \mid \boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(e_{k+1} \mid \boldsymbol{X}_{k+1}) \boldsymbol{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{x}_k) P(\boldsymbol{x}_{1:k}, \boldsymbol{e}_{1:k}) \\
&= \boldsymbol{P}(e_{k+1} \mid \boldsymbol{X}_{k+1}) \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{x}_k) P(\boldsymbol{x}_{1:k}, \boldsymbol{e}_{1:k}) \\
&= \boldsymbol{P}(e_{k+1} \mid \boldsymbol{X}_{k+1}) \max_{\boldsymbol{x}_k} \boldsymbol{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{x}_k) \max_{\boldsymbol{x}_{1:k-1}} P(\boldsymbol{x}_{1:k-1}, \boldsymbol{x}_k, \boldsymbol{e}_{1:k})
\end{aligned}$$

Recurrence

Define the message

$$\boldsymbol{m}_{1:k} \equiv \max_{\boldsymbol{x}_{1:k-1}} \boldsymbol{P}(\boldsymbol{x}_{1:k-1}, \boldsymbol{X}_k, \boldsymbol{e}_{1:k})$$

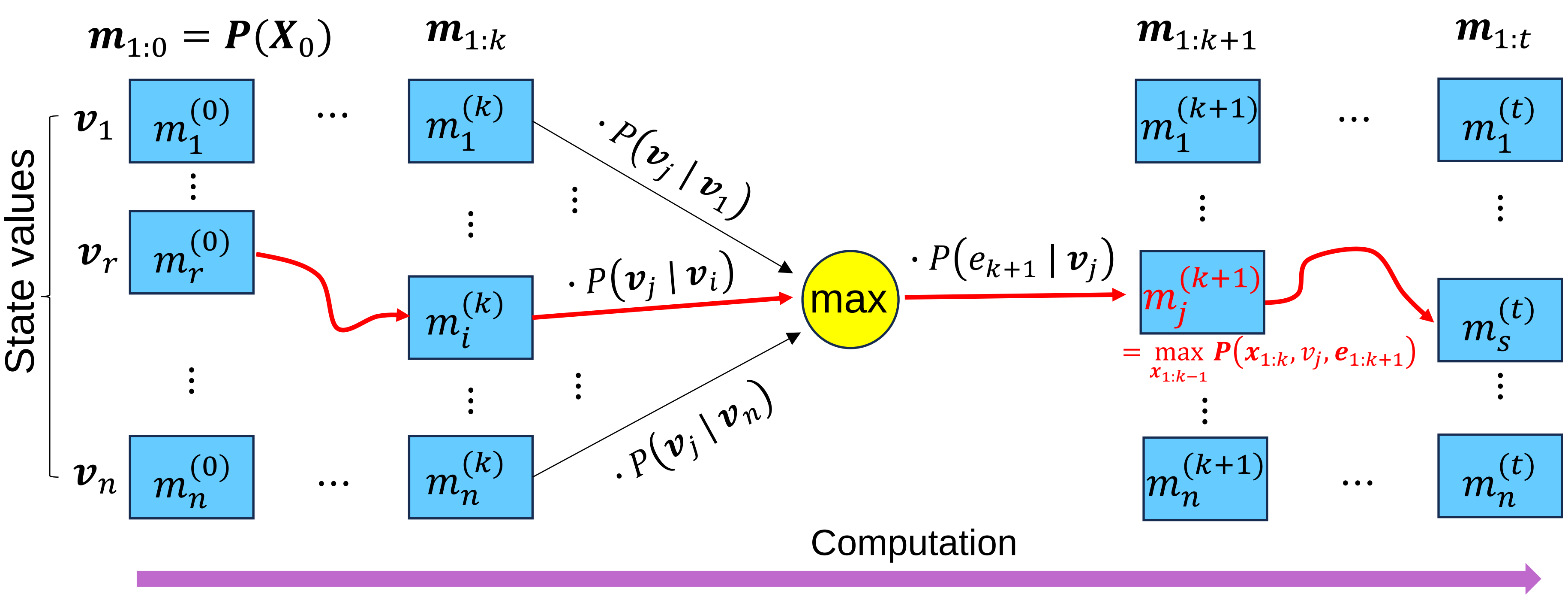
We derive a recurrence for the term as follows:

$$\begin{aligned}
\boldsymbol{m}_{1:k+1} &= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k+1}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}, e_{k+1}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(e_{k+1} \mid \boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}) \boldsymbol{P}(\boldsymbol{x}_{1:k}, \boldsymbol{X}_{k+1}, \boldsymbol{e}_{1:k}) \\
&= \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(e_{k+1} \mid \boldsymbol{X}_{k+1}) \boldsymbol{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{x}_k) P(\boldsymbol{x}_{1:k}, \boldsymbol{e}_{1:k}) \\
&= \boldsymbol{P}(e_{k+1} \mid \boldsymbol{X}_{k+1}) \max_{\boldsymbol{x}_{1:k}} \boldsymbol{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{x}_k) P(\boldsymbol{x}_{1:k}, \boldsymbol{e}_{1:k}) \\
&= \boldsymbol{P}(e_{k+1} \mid \boldsymbol{X}_{k+1}) \max_{\boldsymbol{x}_k} \boldsymbol{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{x}_k) \underbrace{\max_{\boldsymbol{x}_{1:k-1}} P(\boldsymbol{x}_{1:k-1}, \boldsymbol{x}_k, \boldsymbol{e}_{1:k})}_{\text{An element of } \boldsymbol{m}_{1:k} \text{ corresponding to } \boldsymbol{x}_k}
\end{aligned}$$

Viterbi Algorithm

$$\boldsymbol{m}_{1:k} \equiv \max_{\boldsymbol{x}_{1:k-1}} \boldsymbol{P}(\boldsymbol{x}_{1:k-1}, \boldsymbol{X}_k, \boldsymbol{e}_{1:k}) = \left(m_1^{(k)}, m_2^{(k)}, \dots, m_n^{(k)}\right)^T$$ (definition)

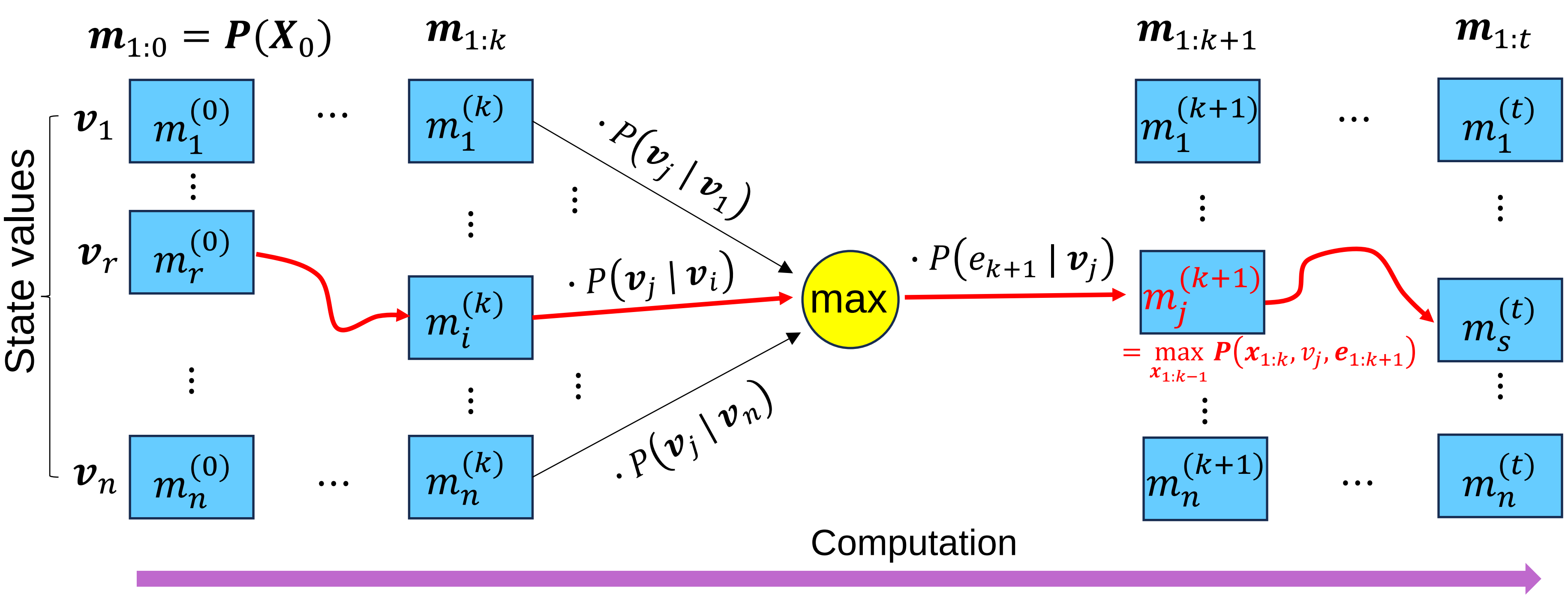
$$\boldsymbol{m}_{1:k+1} = \boldsymbol{P}(e_{k+1} \mid \boldsymbol{X}_{k+1}) \max_{\boldsymbol{x}_k} \boldsymbol{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{x}_k) \max_{\boldsymbol{x}_{1:k-1}} P(\boldsymbol{x}_{1:k-1}, \boldsymbol{x}_k, \boldsymbol{e}_{1:k})$$ (recurrence)



Viterbi Algorithm

$$\boldsymbol{m}_{1:k} \equiv \max_{\boldsymbol{x}_{1:k-1}} \boldsymbol{P}(\boldsymbol{x}_{1:k-1}, \boldsymbol{X}_k, \boldsymbol{e}_{1:k}) = \left(m_1^{(k)}, m_2^{(k)}, \dots, m_n^{(k)}\right)^T$$ (definition)

$$\boldsymbol{m}_{1:k+1} = \boldsymbol{P}(e_{k+1} \mid \boldsymbol{X}_{k+1}) \max_{\boldsymbol{x}_k} \boldsymbol{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{x}_k) \max_{\boldsymbol{x}_{1:k-1}} P(\boldsymbol{x}_{1:k-1}, \boldsymbol{x}_k, \boldsymbol{e}_{1:k})$$ (recurrence)



- To identify the actual sequence, record, for each state, the best predecessor state.

Execution on the Umbrella World

R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2

Problem Determine the most likely weather sequence from the umbrella sequence [*true*, *true*, *false*, *true*, *true*].

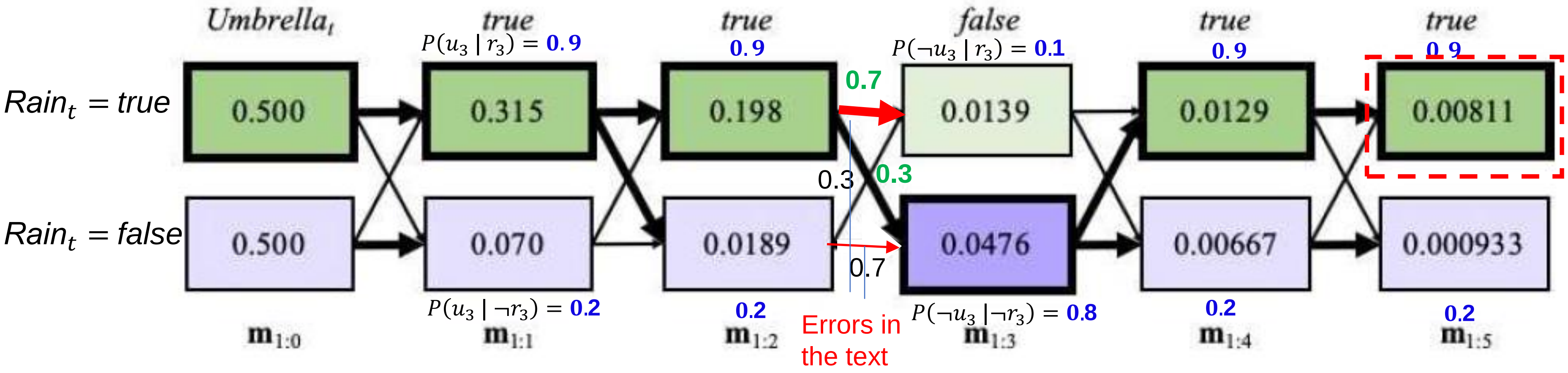
i.e., $\{u_1, u_2, \neg u_3, u_4, u_5\}$

Execution on the Umbrella World

R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2

Problem Determine the most likely weather sequence from the umbrella sequence [*true*, *true*, *false*, *true*, *true*]. i.e., $\{u_1, u_2, \neg u_3, u_4, u_5\}$



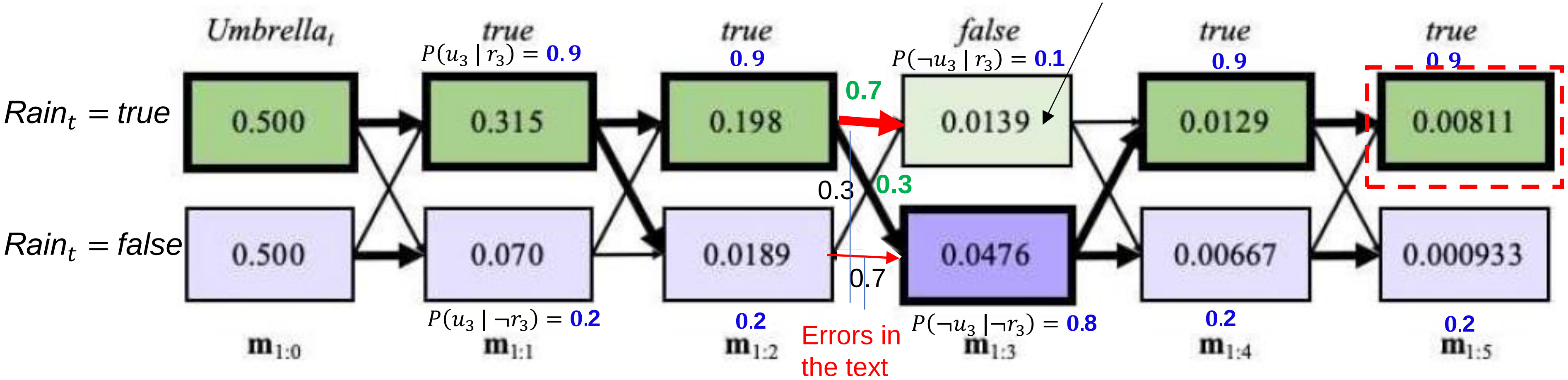
Execution on the Umbrella World

R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2

Problem Determine the most likely weather sequence from the umbrella sequence [*true*, *true*, *false*, *true*, *true*]. i.e., $\{u_1, u_2, \neg u_3, u_4, u_5\}$

$$P(\neg u_3 \mid r_3) \cdot \max\left(\left(P(r_3 \mid r_2), P(r_3 \mid \neg r_2)\right) \cdot (0.198, 0.0189)\right)$$
$$= \mathbf{0.1} \cdot \max(\mathbf{0.7} \cdot 0.198, \mathbf{0.3} \cdot 0.0189) \approx 0.0139$$



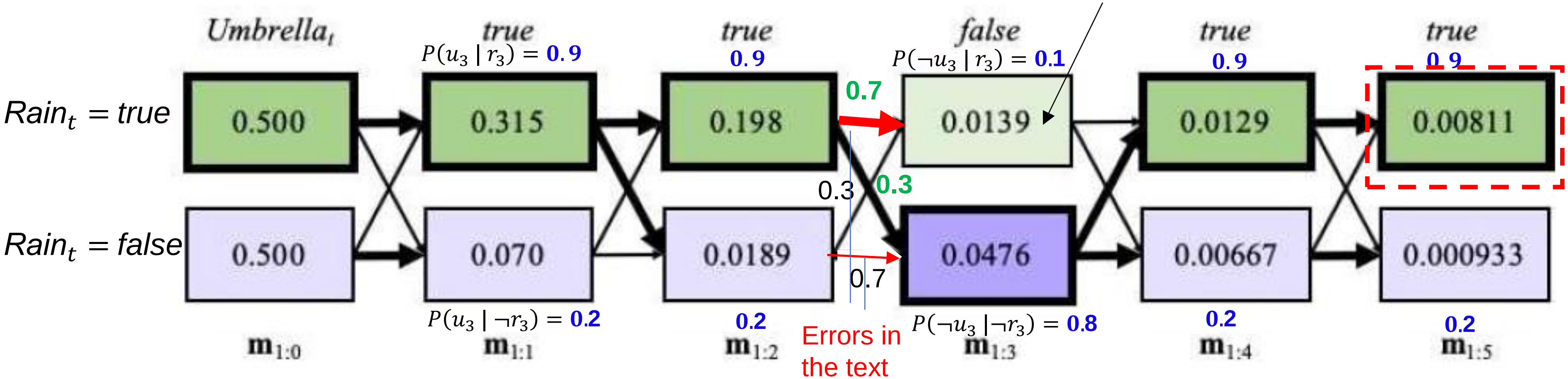
Execution on the Umbrella World

R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2

Problem Determine the most likely weather sequence from the umbrella sequence [*true*, *true*, *false*, *true*, *true*].
i.e., $\{u_1, u_2, \neg u_3, u_4, u_5\}$

$$P(\neg u_3 \mid r_3) \cdot \max\left(\left(P(r_3 \mid r_2), P(r_3 \mid \neg r_2)\right) \cdot (0.198, 0.0189)\right)$$
$$= \mathbf{0.1} \cdot \max(\mathbf{0.7} \cdot 0.198, \mathbf{0.3} \cdot 0.0189) \approx 0.0139$$



$$0.00811 \approx 0.5 \times (0.7 \times 0.9) \times (0.7 \times 0.9) \times ((1-0.7) \times (1-0.2)) \times (0.3 \times 0.9) \times (0.7 \times 0.9)$$

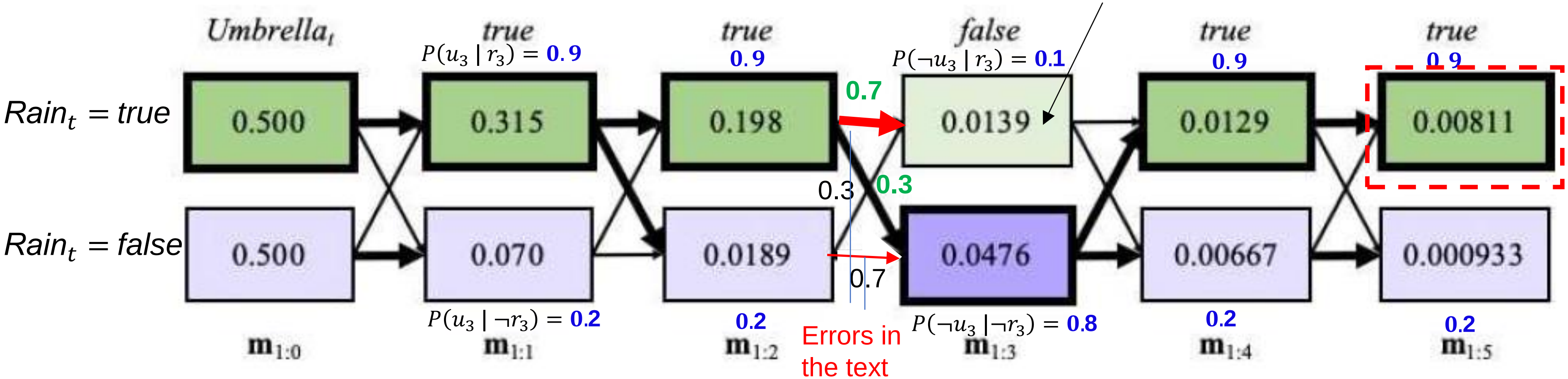
Execution on the Umbrella World

R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2

Problem Determine the most likely weather sequence from the umbrella sequence [*true*, *true*, *false*, *true*, *true*].
i.e., $\{u_1, u_2, \neg u_3, u_4, u_5\}$

$$P(\neg u_3 \mid r_3) \cdot \max\left(\left(P(r_3 \mid r_2), P(r_3 \mid \neg r_2)\right) \cdot (0.198, 0.0189)\right)$$
$$= \mathbf{0.1} \cdot \max(\mathbf{0.7} \cdot 0.198, \mathbf{0.3} \cdot 0.0189) \approx 0.0139$$



$$0.00811 \approx \underbrace{0.5 \times (0.7 \times 0.9)}_{0.315} \times (0.7 \times 0.9) \times ((1 - 0.7) \times (1 - 0.2)) \times (0.3 \times 0.9) \times (0.7 \times 0.9)$$

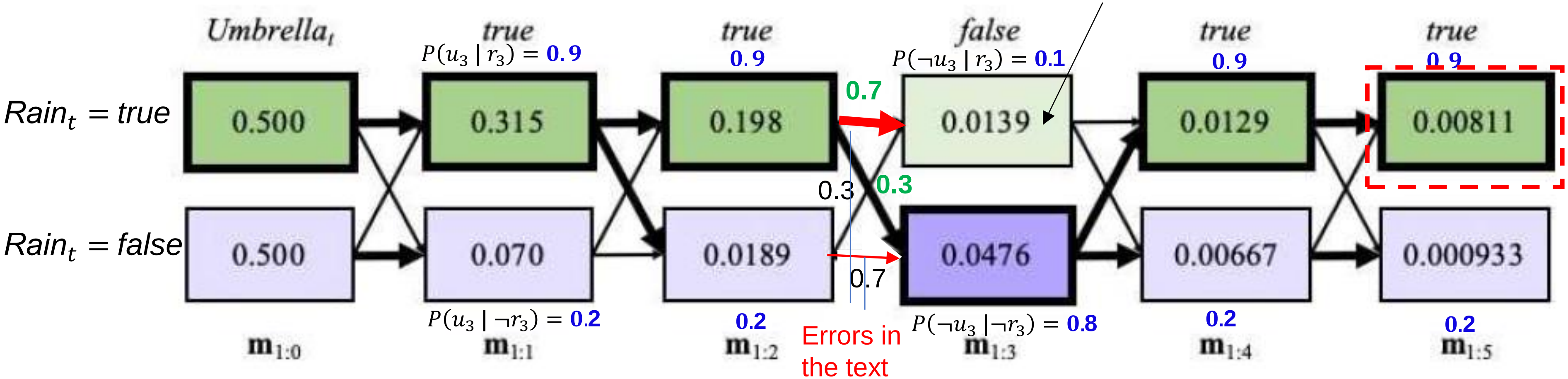
Execution on the Umbrella World

R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2

Problem Determine the most likely weather sequence from the umbrella sequence [*true*, *true*, *false*, *true*, *true*]. i.e., $\{u_1, u_2, \neg u_3, u_4, u_5\}$

$$P(\neg u_3 \mid r_3) \cdot \max\left(\left(P(r_3 \mid r_2), P(r_3 \mid \neg r_2)\right) \cdot (0.198, 0.0189)\right)$$
$$= \mathbf{0.1} \cdot \max(\mathbf{0.7} \cdot 0.198, \mathbf{0.3} \cdot 0.0189) \approx 0.0139$$



$$0.00811 \approx \underbrace{\underbrace{0.5 \times (0.7 \times 0.9)}_{0.315} \times (0.7 \times 0.9)}_{0.198} \times ((1-0.7) \times (1-0.2)) \times (0.3 \times 0.9) \times (0.7 \times 0.9)$$

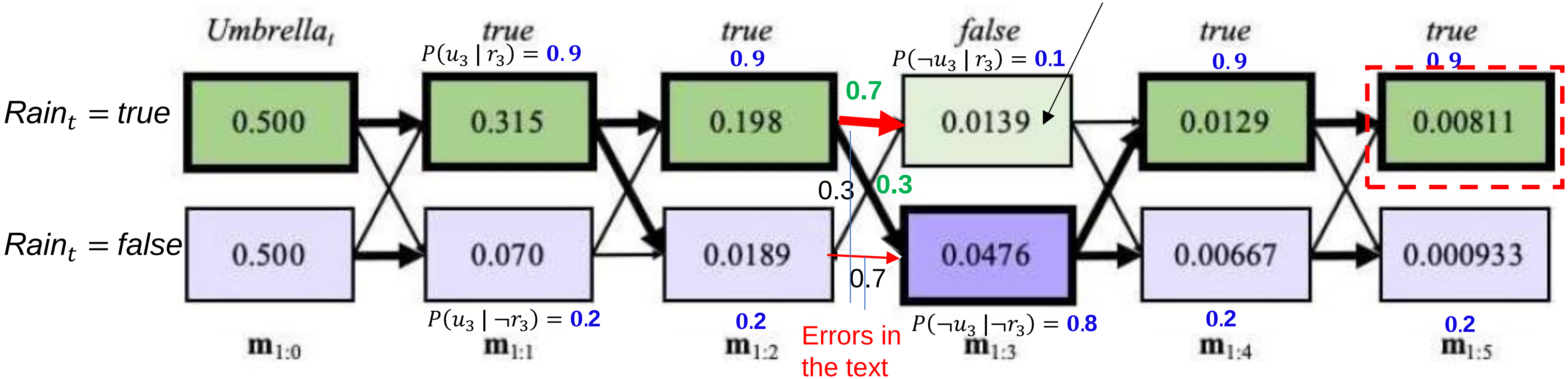
Execution on the Umbrella World

R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2

Problem Determine the most likely weather sequence from the umbrella sequence [*true*, *true*, *false*, *true*, *true*].
i.e., $\{u_1, u_2, \neg u_3, u_4, u_5\}$

$P(\neg u_3 \mid r_3) \cdot \max\left((P(r_3 \mid r_2), P(r_3 \mid \neg r_2)) \cdot (0.198, 0.0189)\right)$
$= \mathbf{0.1} \cdot \max(\mathbf{0.7} \cdot 0.198, \mathbf{0.3} \cdot 0.0189) \approx 0.0139$



$$0.00811 \approx \underbrace{\underbrace{\underbrace{0.5 \times (0.7 \times 0.9)}_{0.315} \times (0.7 \times 0.9)}_{0.198} \times ((1-0.7) \times (1-0.2))}_{0.0476} \times (0.3 \times 0.9) \times (0.7 \times 0.9)$$

Execution on the Umbrella World

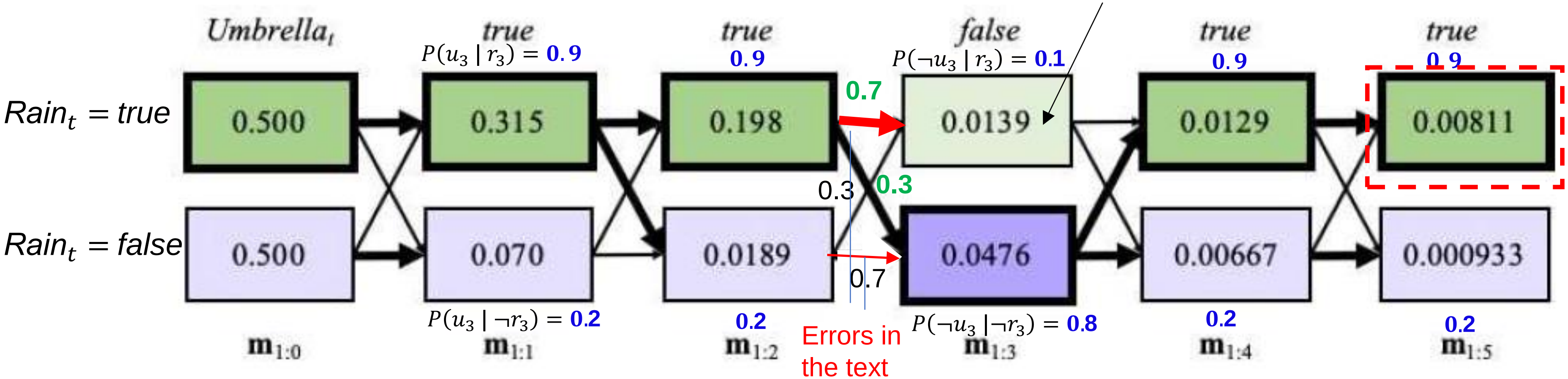
R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2

Problem Determine the most likely weather sequence from the umbrella sequence [*true*, *true*, *false*, *true*, *true*]. i.e., $\{u_1, u_2, \neg u_3, u_4, u_5\}$

$$P(\neg u_3 \mid r_3) \cdot \max\left(\left(P(r_3 \mid r_2), P(r_3 \mid \neg r_2)\right) \cdot (0.198, 0.0189)\right)$$

$$= \mathbf{0.1} \cdot \max(\mathbf{0.7} \cdot 0.198, \mathbf{0.3} \cdot 0.0189) \approx 0.0139$$



$$0.00811 \approx 0.5 \times (0.7 \times 0.9) \times (0.7 \times 0.9) \times ((1 - 0.7) \times (1 - 0.2)) \times (0.3 \times 0.9) \times (0.7 \times 0.9)$$

0.315

0.198

0.0476

0.0129

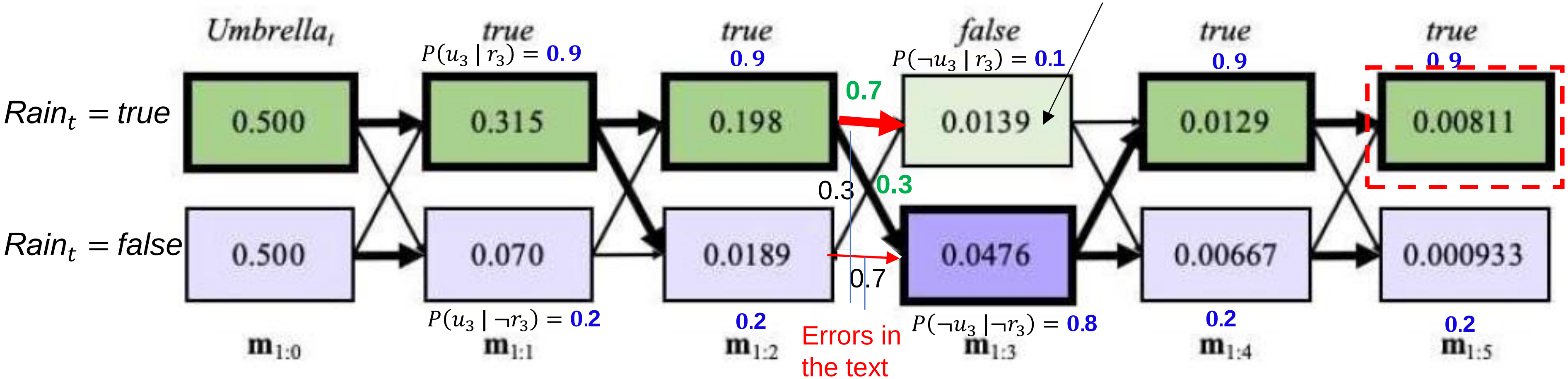
Execution on the Umbrella World

R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2

Problem Determine the most likely weather sequence from the umbrella sequence [*true*, *true*, *false*, *true*, *true*].
i.e., $\{u_1, u_2, \neg u_3, u_4, u_5\}$

$$P(\neg u_3 \mid r_3) \cdot \max\left(\left(P(r_3 \mid r_2), P(r_3 \mid \neg r_2)\right) \cdot (0.198, 0.0189)\right)$$
$$= \mathbf{0.1} \cdot \max(\mathbf{0.7} \cdot 0.198, \mathbf{0.3} \cdot 0.0189) \approx 0.0139$$



$$0.00811 \approx 0.5 \times (0.7 \times 0.9) \times (0.7 \times 0.9) \times ((1-0.7) \times (1-0.2)) \times (0.3 \times 0.9) \times (0.7 \times 0.9)$$

0.315

0.198

0.0476

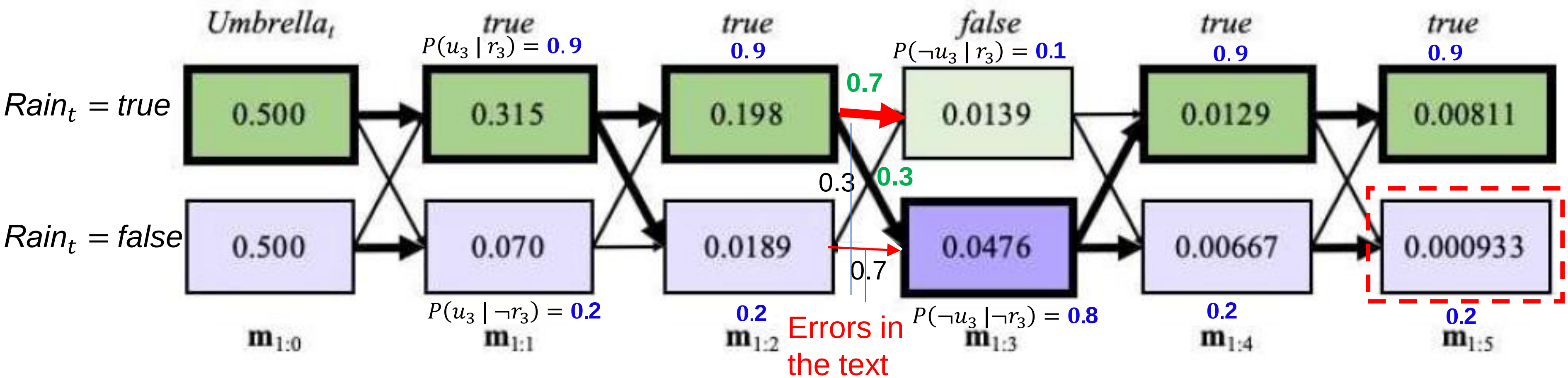
0.0129

0.00811 is the maximum joint probability of any rain scenario on the first four days, rain on day 5, and the umbrella sequence [*true*, *true*, *false*, *true*, *true*] on the first five days. It is achieved by the weather sequence [*true*, *true*, *false*, *true*, *true*].

Execution (cont'd)

R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2

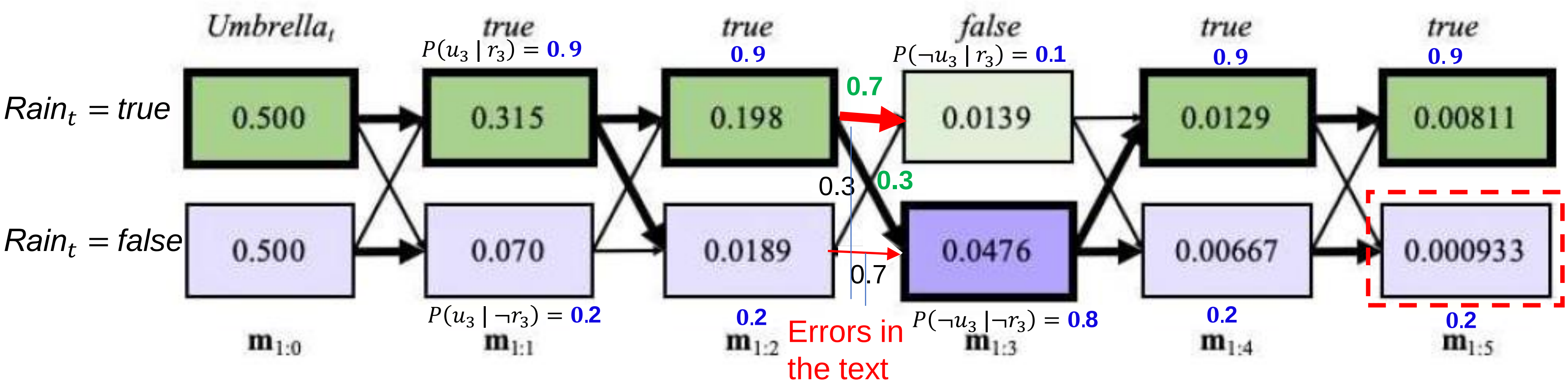


$$0.000933 \approx 0.5 \times (0.7 \times 0.9) \times (0.7 \times 0.9) \times (0.3 \times 0.8) \times (0.7 \times 0.2) \times (0.7 \times 0.2)$$

Execution (cont'd)

R_{t-1}	$P(R_t \mid R_{t-1})$
t	0.7
f	0.3

R_t	$P(U_t \mid R_t)$
t	0.9
f	0.2



$$0.000933 \approx 0.5 \times (0.7 \times 0.9) \times (0.7 \times 0.9) \times (0.3 \times 0.8) \times (0.7 \times 0.2) \times (0.7 \times 0.2)$$

If it does not rain on day 5 and the umbrella sequence on the first five days is [*true*, *true*, *false*, *true*, *true*], then the most likely weather sequence is [*true*, *true*, *false*, *false*, *false*] with the maximum join probability of 0.000933.