HMM

Inference

Notation

We will use the following notations:

$$\begin{aligned}
p(x_{t+1}|x_t) &= \mathbb{P}(X_{t+1} = x_{t+1}|X_t = x_t) \\
p(y_t|x_t) &= \mathbb{P}(Y_t = y_t|X_t = x_t) \\
p(x_t|y_{1:t}) &= \mathbb{P}(X_t = x_t|Y_1 = y_1, \ldots, Y_t = y_t) \\
p(y_{1:t}) &= \mathbb{P}(Y_1 = y_1, \ldots, Y_t = y_t)
\end{aligned}$$

Inference problems

- **Filtering:**

Given measurements up to time t, compute the distribution of Xt. $p(x_t|y_{1:t})$

- **Prediction:**

Given measurements up to time s, compute the distribution of Xt, s<t $p(x_t|y_{1:s}), \quad s < t$

- **Smoothing**

Given measurements up to time s, compute the distribution of Xt, s>t $p(x_t|y_{1:s}), \quad s > t$

- **Likelihood**

Find the likelihood of Y given the model $p(y_{1:T})$

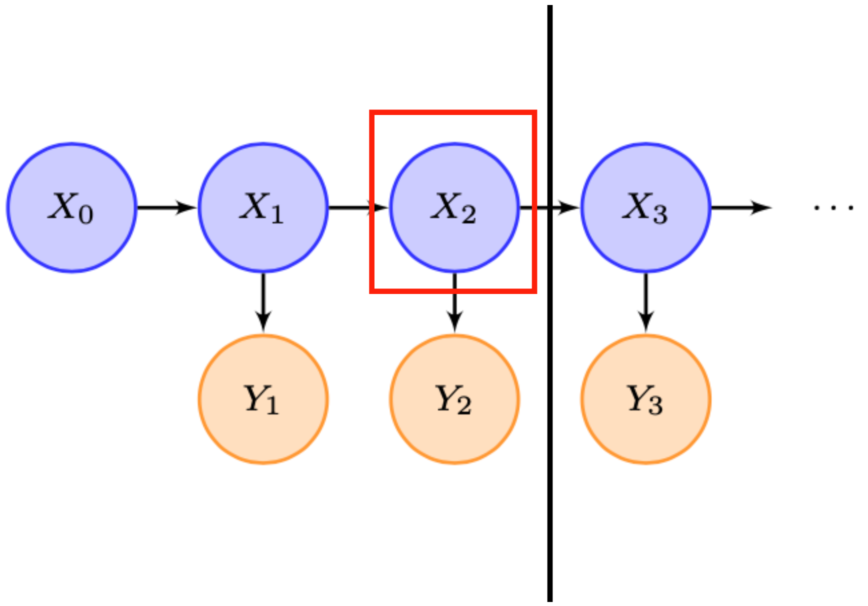
- **Decoding**

Find the most likely state history X given the observation history Y $\arg\max_{x_{0:T}} p(x_{0:T}|y_{1:T})$

Filtering

Given measurements up to time t, compute the distribution of Xt

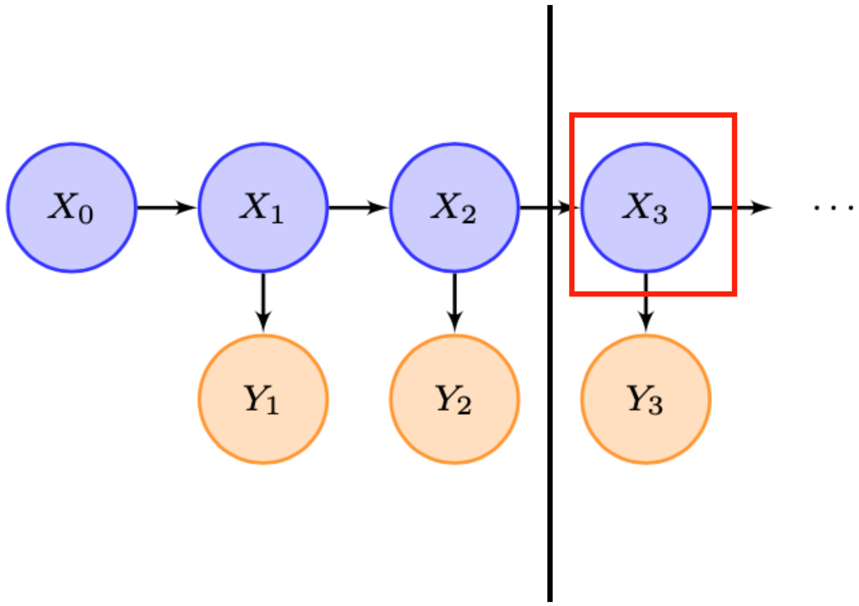
E.g. t=2



Prediction

Given measurements up to time s, compute the distribution of Xt, s<t

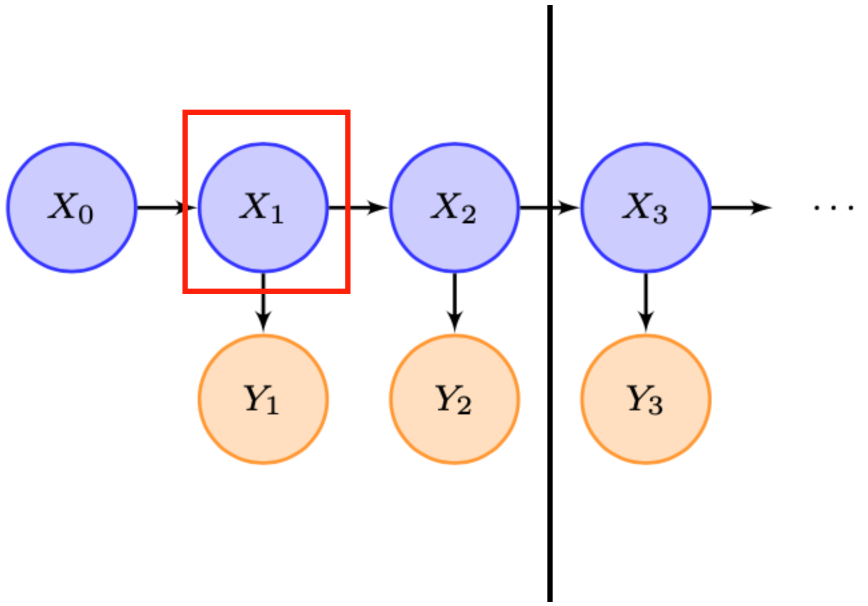
E.g. s=2
t=3



Smoothing

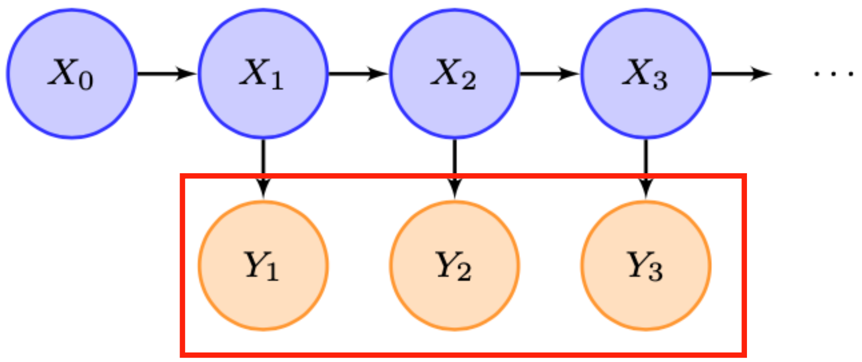
Given measurements up to time s, compute the distribution of Xt, s>t

E.g. s=2
t=1



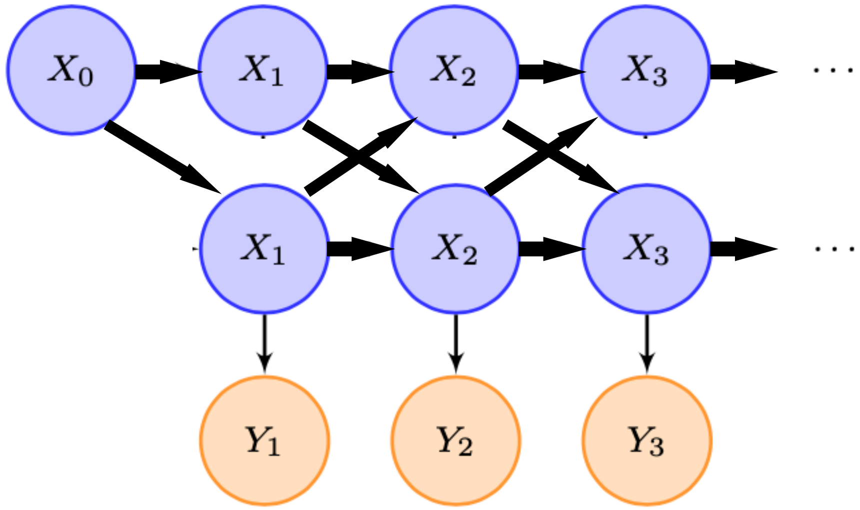
Likelihood

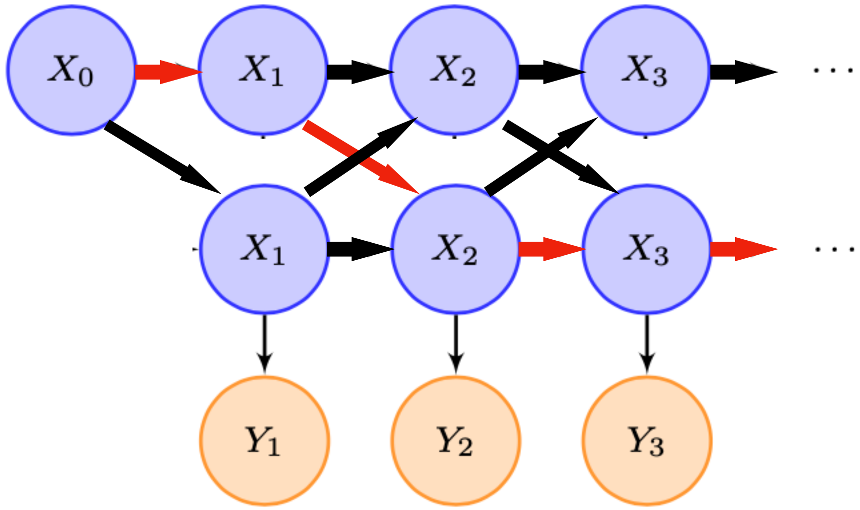
Find the likelihood of Y given the model



Decoding

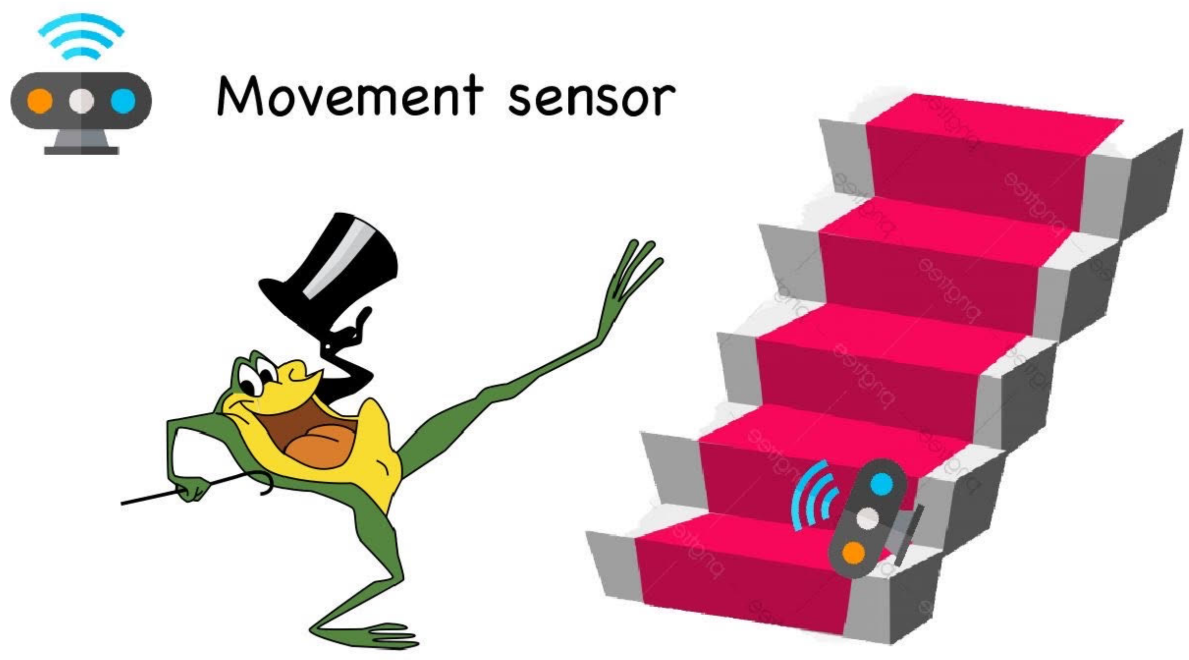
Find the most likely state history X given the observation history Y





Intuition

A dancing frog



Prior distribution

Where is the frog at the beginning?

1	2	3	4	5	6
1/6	1/6	1/6	1/6	1/6	1/6

Transition probabilities (Matrix A)

Where can the frog jump?

Level to

Level from

	1	2	3	4	5	6
1	0.4	0.6				
2	0.3	0.4	0.6			
3		0.3	0.4	0.6		
4			0.3	0.4	0.6	
5				0.3	0.4	0.6
6	0.6				0.3	0.4

Emission probabilities (Matrix B)

What is the probability to detect the frog?

Movement sensor

	Detection	No detection
1	0.9	0.1
2	0.5	0.5
3	0.1	0.9
4	0	1
5	0	1
6	0	1

Observations

After 14 times, this are the sensor's results:

1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	0	0	0	1	1	0	0	0	0	1	1	0	1

Using filtering algorithm

Inizialization

$$\alpha_1(i) = \pi_i \cdot b_i(O_1)$$

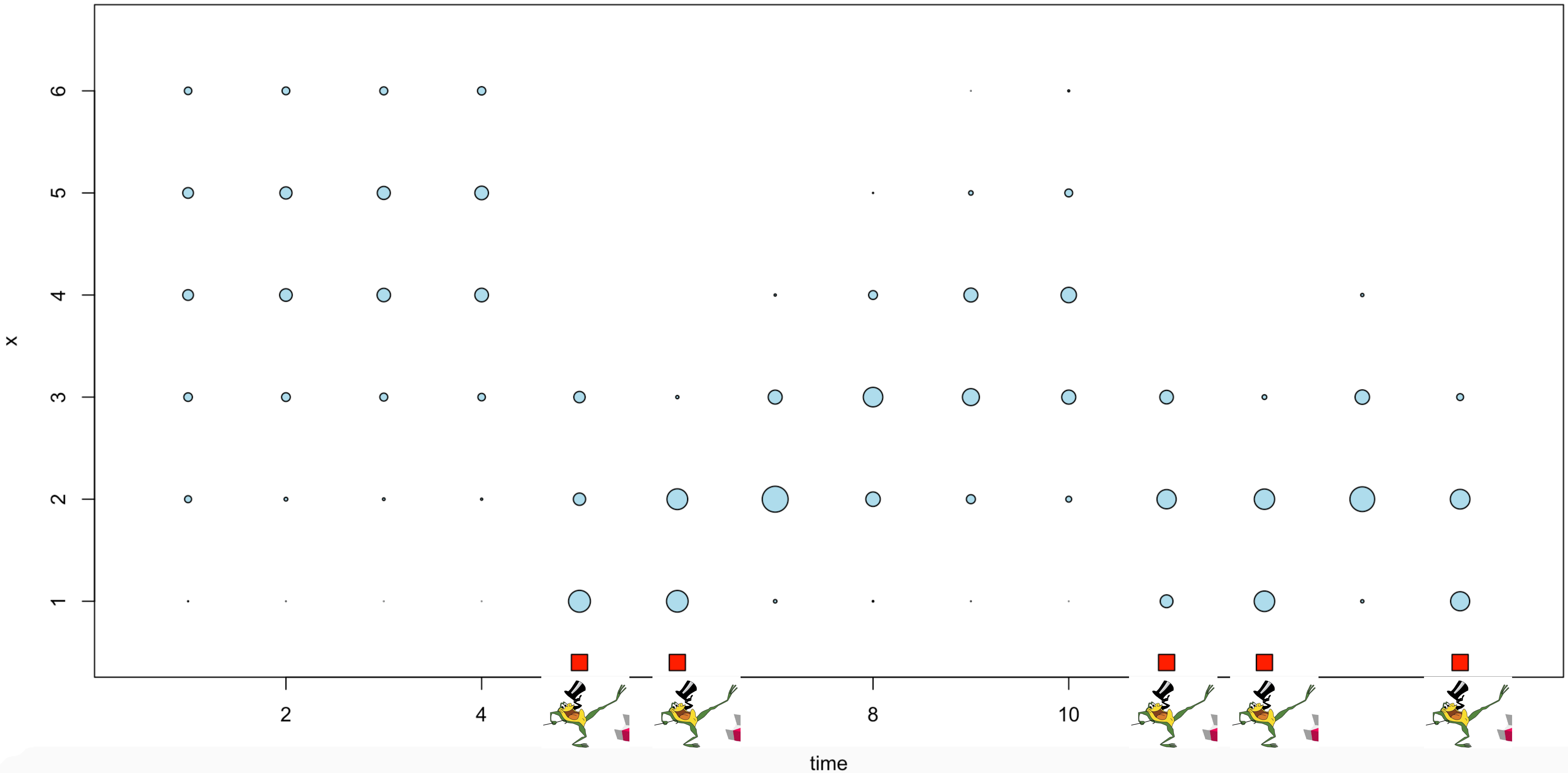
Recursion

$$\alpha_{t+1}(j) = \sum_{i=1}^{N} \alpha_t(i) \cdot a_{ij} \cdot b_j(O_{t+1})$$

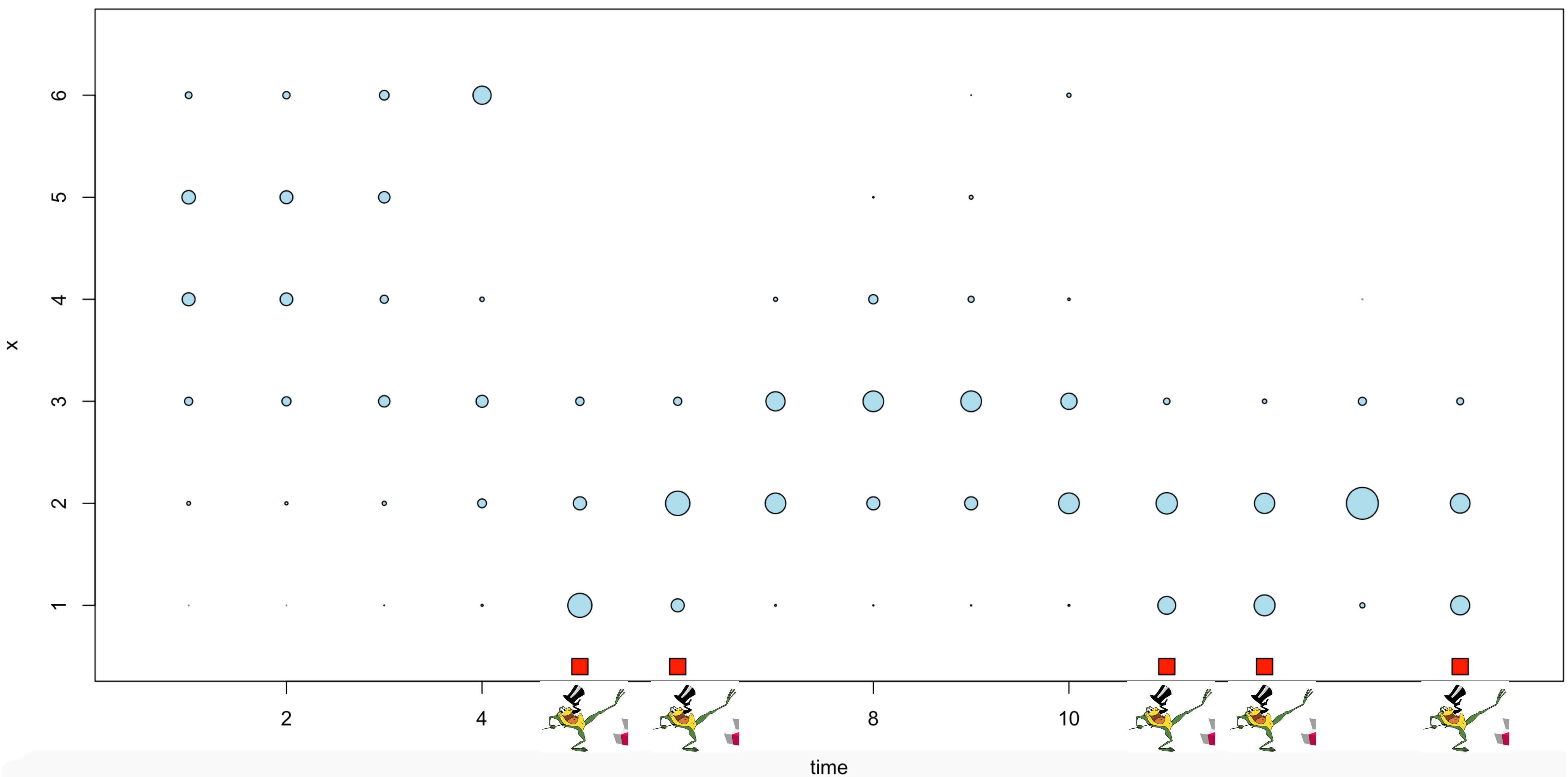
Termination

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

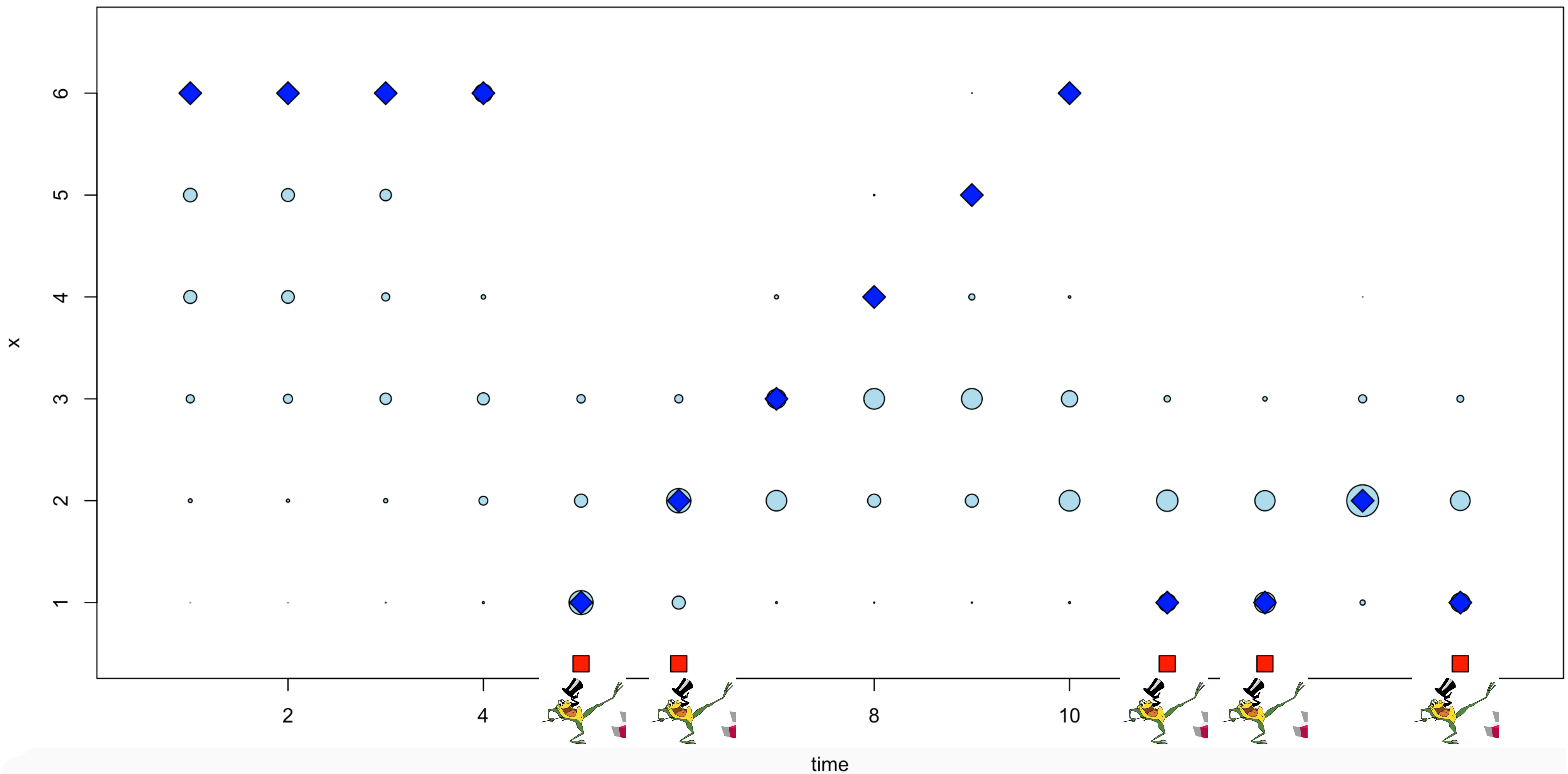
Distribution of X (filtering)



Distribution of X (smoothing)



Decoding of X



Formal Filtering

We are interested in the conditional probability mass function $p(x_t|y_{1:t})$ of the state X_t given the data observed up to time t

$$p(x_t|y_{1:t}) = \frac{p(x_t, y_{1:t})}{\sum_{x'_t \in \mathcal{X}} p(x'_t, y_{1:t})}$$

We will derive a recursion for

$$\begin{aligned}\boxed{p(x_t, y_{1:t})} &= \sum_{x_{t-1} \in \mathcal{X}} p(x_t, x_{t-1}, y_t, y_{1:t-1}) \\ &= \sum_{x_{t-1} \in \mathcal{X}} p(y_t|x_t, x_{t-1}, y_{1:t-1}) p(x_t|x_{t-1}, y_{1:t-1}) p(x_{t-1}, y_{1:t-1}) \\ &= p(y_t|x_t) \sum_{x_{t-1} \in \mathcal{X}} p(x_t|x_{t-1}) p(x_{t-1}, y_{1:t-1})\end{aligned}$$

We can define the **alpha recursion** as: $\boxed{\alpha_t(x_t)} = \boxed{p(x_t, y_{1:t})}$ with: $\alpha_0(x_0) = p(x_0).$

Formal Filtering and likelihood

So, we compute a filtering using:

For t = 1, . . . , T , $x_t \in$ X:

$$\boxed{\alpha_t(x_t)} = p(y_t|x_t) \sum_{x_{t-1} \in \mathcal{X}} p(x_t|x_{t-1}) \boxed{\alpha_{t-1}(x_{t-1})}$$

The filtering pmf is obtained by normalizing $\alpha_t(x_t)$ as:

$$p(x_t|y_{1:t}) = \frac{p(x_t, y_{1:t})}{p(y_{1:t})} = \frac{\alpha_t(x_t)}{\sum_{x \in \mathcal{X}} \alpha_t(x)}.$$

The likelihood term p(y1:T) can be computed from the α-recursion

$$p(y_{1:T}) = \sum_{x \in \mathcal{X}} \alpha_T(x)$$

Considerations in Filtering

The computation cost of the whole forward recursion is $O(T|\mathcal{X}|^2)$

The proposed recursion may suffer from numerical underflow/overflow, as α_t may become very small or very large for large t.

To avoid this, we can normalize α_t, or propagate the filtering pmf $p(x_t|y_{1:t})$ instead of α_t, using the following two-step predict-update recursion

$$p(x_t|y_{1:t-1}) = \sum_{x_{t-1}\in\mathcal{X}} p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1}) \qquad \text{Predict}$$

$$p(x_t|y_{1:t}) = \frac{g_{x_t}(y_t)p(x_t|y_{1:t-1})}{\sum_{x'_t\in\mathcal{X}} g_{x'_t}(y_t)p(x'_t|y_{1:t-1})} \qquad \text{Update}$$