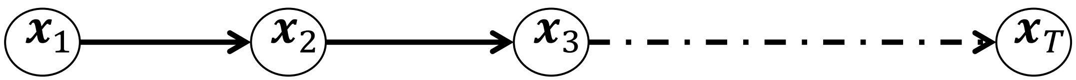
Hidden Markov Models (HMM)

Devise algorithms for computing the posteriors on a Markov Chain with hidden variables (HMM).

Markov Chains with Discrete Random Variables



Let's assume we have discrete random variables (e.g., taking 3 discrete values $\boldsymbol{x}_t = \{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\}$)

Markov Property: $p(\boldsymbol{x}_t|\boldsymbol{x}_1, .., \boldsymbol{x}_{t-1}) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$

$$\text{e.g. } p(\boldsymbol{x}_t = \begin{bmatrix}1\\0\\0\end{bmatrix} | \boldsymbol{x}_{t-1} = \begin{bmatrix}0\\1\\0\end{bmatrix})$$

Stationary, Homogeneous or Time-Invariant if the distribution $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ does not depend on t

Markov Chains with Discrete Random Variables

$$p(\boldsymbol{x}_1, .., \boldsymbol{x}_T) = p(\boldsymbol{x}_1) \prod_{t=2}^{\mathrm{T}} p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1})$$

What do we need in order to describe the whole procedure?

(1) A probability for the first frame/timestamp etc $p(\boldsymbol{x}_1)$. In order to define the probability we need to define the vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_{\mathrm{K}})$

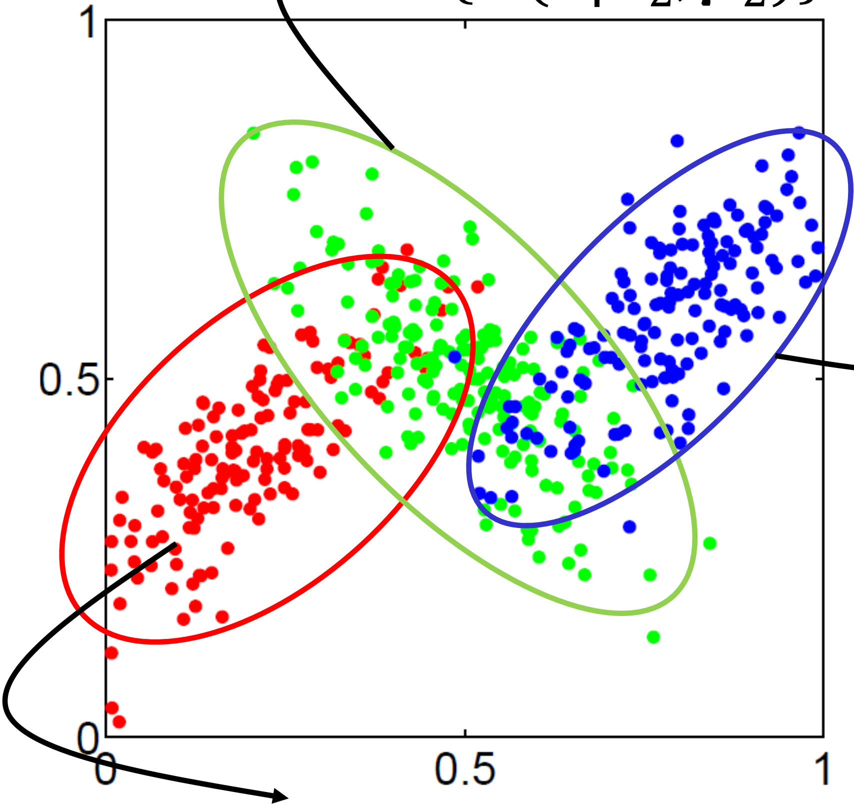
$$p(\boldsymbol{x}_1 | \boldsymbol{\pi}) = \prod_{c=1}^{K} \pi_c^{x_{1c}}$$

(2) A transition probability $p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1})$. In order to define it we need a KxK transition matrix $\boldsymbol{A} = [a_{ij}]$

$$p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{A}) = \prod_{j=1}^{K} \prod_{k=1}^{K} a_{jk}^{x_{t-1j} x_{tk}}$$

Latent Variables

$\{N(\boldsymbol{x}|\boldsymbol{\Sigma}_2, \boldsymbol{\mu}_2)\}$



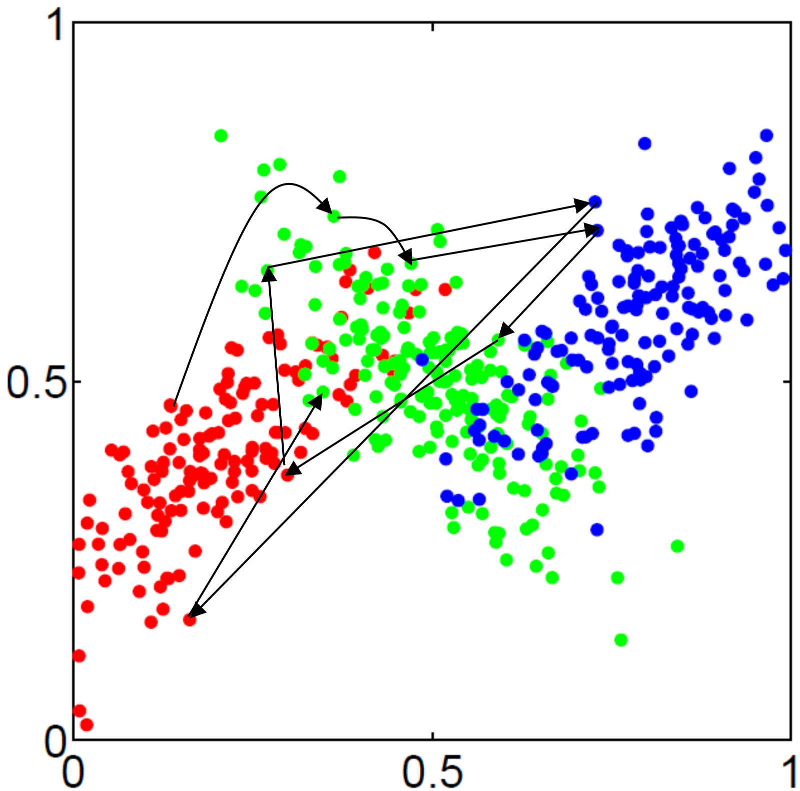
$\{\mathrm{N}(\boldsymbol{x}|\boldsymbol{\Sigma}_3, \boldsymbol{\mu}_3)\}$

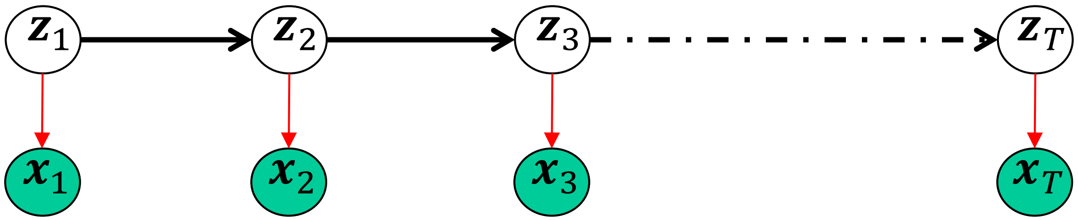
$$\boldsymbol{z}_t = \begin{bmatrix} z_{t1} \\ z_{t2} \\ z_{t3} \end{bmatrix} \in \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

$\{\mathrm{N}(\boldsymbol{x}|\boldsymbol{\Sigma}_1, \boldsymbol{\mu}_1)\}$

$$p(\boldsymbol{X}, \boldsymbol{Z}|\theta)$$
$$= p\left(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T, \boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_T|\theta\right)$$
$$= \prod_{t=1}^{T} p(\boldsymbol{x}_t|\boldsymbol{z}_t, \theta_x) \prod_{t=1}^{T} p(\boldsymbol{z}_t|\theta_z)$$

Latent Variables in a Markov Chain





$$p(\mathbf{z}_1, .., \mathbf{z}_T) = p(\mathbf{z}_1) \prod_{t=2}^{\mathrm{T}} p(\mathbf{z}_t | \mathbf{z}_{t-1})$$

$$p(\boldsymbol{X}, \boldsymbol{Z} | \theta)$$
$$= p\left(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T, \mathbf{z}_{1,} \mathbf{z}_2, \cdots, \mathbf{z}_T | \theta\right)$$
$$= \prod_{t=1}^{T} p(\boldsymbol{x}_t | \mathbf{z}_t, \theta_x)\, p(\mathbf{z}_1) \prod_{t=2}^{\mathrm{T}} p(\mathbf{z}_t | \mathbf{z}_{t-1})$$

Hidden Markov Models

A casino has two dice (our latent variable is probably the dice ☺)

$$\boldsymbol{z} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

One is fair and one is not

Each die has 6 sides.

$$\boldsymbol{x} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

(1) Fair die $p\left(x_j = 1 | z_1 = 1\right) = \frac{1}{6}$ for all $j = \{1, .., 6\}$

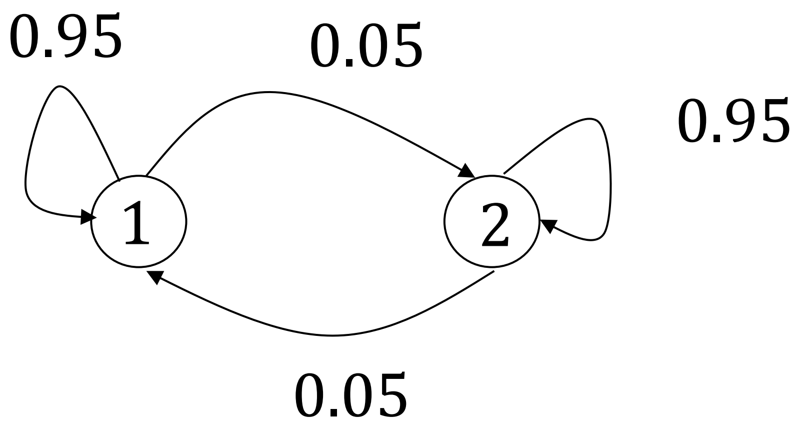
(2) Loaded die $p\left(x_j = 1 | z_2 = 1\right) = \frac{1}{10}$ for $j = \{1, .., 5\}$ and $p(x_6 = 6 | z_2 = 1) = \frac{1}{2}$

Emission probability

Hidden Markov Models

A casino player switches back & forth between fair and loaded die once every 20 turns

$$(\text{i.e., } p(z_{t\,1} = 1 | z_{t-1\,2} = 1) = p(z_{t\,2} = 1 | z_{t-1\,1} = 1) = 0.05)$$



Hidden Markov Models

Given a string of observations and the above model:

6641532161621152346532143566342616552342323151424641566663246

(1) We want to find for a timestamp t the probabilities of each die given the observations **that far**.

This process is called Filtering: $p(\boldsymbol{z}_t | \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t)$

(2) We want to find for a timestamp t the probabilities of each die given **the whole string**.

This process is called Smoothing: $p(\boldsymbol{z}_t | \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)$

Hidden Markov Models

(3) Given the observation string find the string of hidden variables that maximize the posterior.

This process is called Decoding (Viterbi).

$$\arg\max_{\boldsymbol{y}_1 \dots \boldsymbol{y}_t} p(\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_t | \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$$

664153216162115234653214356634261655234232315142464156663246

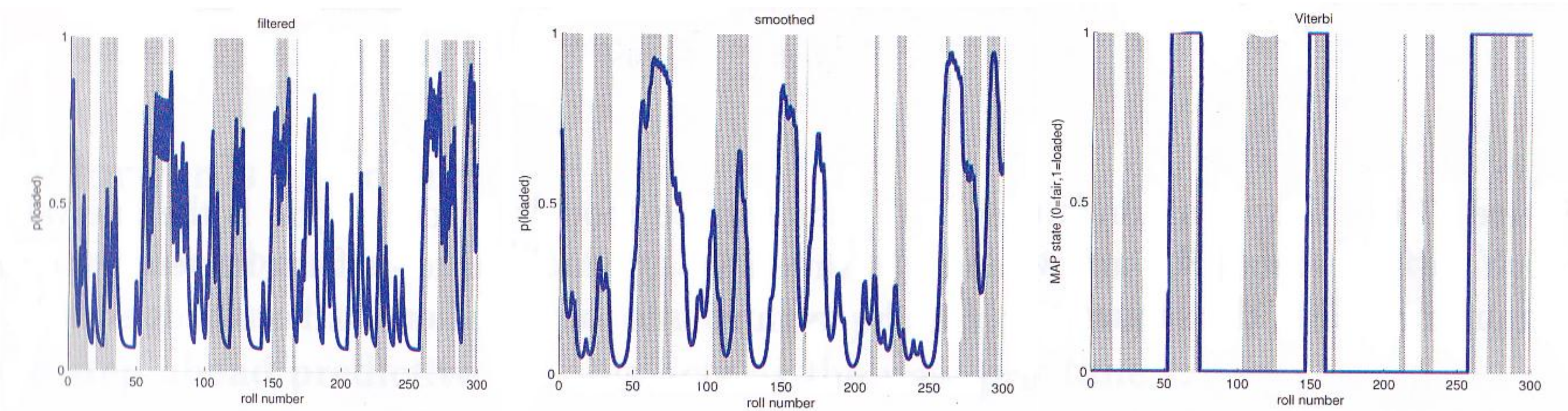
222222222222222111111222222222222221111111111111111112222222

(4) Find the probability of the model.

This process is called Evaluation

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T)$$

Hidden Markov Models



Filtering Smoothing Decoding

Taken from **Machine Learning: A Probabilistic Perspective by K. Murphy**

Hidden Markov Models

(5) Prediction

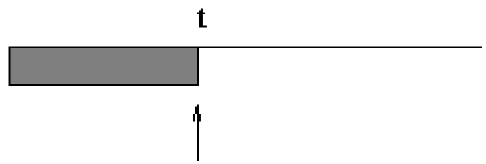
$$p(\boldsymbol{z}_{t+\delta}|\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t) \qquad p(\boldsymbol{x}_{t+\delta}|\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$$

(6) Parameter estimation (Baum-Welch algorithm)

$$\boldsymbol{A}, \boldsymbol{\pi}, \theta$$

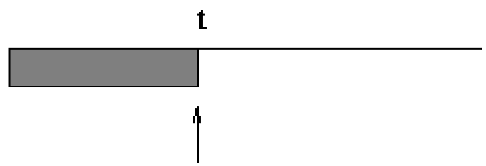
Hidden Markov Models

filtering



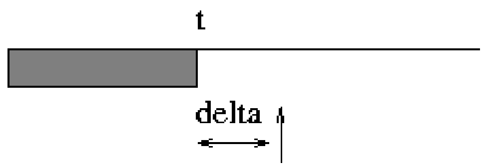
$$p(\boldsymbol{z}_t|\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$$

Viterbi



$$\arg\max_{\boldsymbol{y}_1 \dots \boldsymbol{y}_t} p(\boldsymbol{y}_1, \dots, \boldsymbol{y}_t|\boldsymbol{x}_1, \dots, \boldsymbol{x}_t)$$

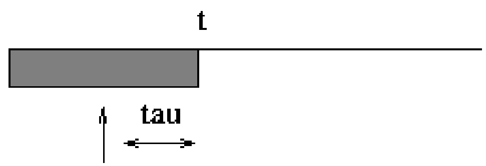
prediction



$$p(\boldsymbol{z}_{t+\delta}|\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$$

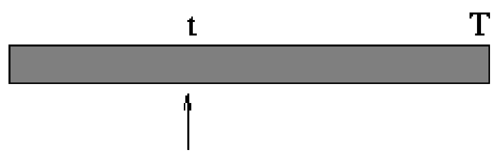
$$p(\boldsymbol{x}_{t+\delta}|\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$$

fixed–lag smoothing



$$p(\boldsymbol{z}_{t-\tau}|\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$$

fixed interval smoothing (offline)



$$p(\boldsymbol{z}_t|\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T)$$

Hidden Markov Models

$$p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_T \,|\boldsymbol{z}_t) = p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_t \,|\boldsymbol{z}_t)$$
$$p(\boldsymbol{x}_{t+1}, \dots, \boldsymbol{x}_T \,|\boldsymbol{z}_t)$$

$$p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1} \,|\boldsymbol{x}_t, \boldsymbol{z}_t) = p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1} \,|\boldsymbol{z}_t)$$

$$p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1} \,|\boldsymbol{z}_{t-1}, \boldsymbol{z}_t) = p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1} \,|\boldsymbol{z}_{t-1})$$

$$p(\boldsymbol{x}_{t+1}, \dots, \boldsymbol{x}_T \,|\boldsymbol{z}_t, \boldsymbol{z}_{t+1}) = p(\boldsymbol{x}_{t+1}, \dots, \boldsymbol{x}_T \,|\boldsymbol{z}_{t+1})$$

$$p(\boldsymbol{x}_{t+2}, \dots, \boldsymbol{x}_T \,|\boldsymbol{z}_{t+1}, \boldsymbol{x}_{t+1}) = p(\boldsymbol{x}_{t+2}, \dots, \boldsymbol{x}_T \,|\boldsymbol{z}_{t+1})$$

$$p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_T \,|\boldsymbol{z}_{t-1}, \boldsymbol{z}_t) = p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1} \,|\boldsymbol{z}_{t-1}) p(\boldsymbol{x}_t|\boldsymbol{z}_t)$$
$$p(\boldsymbol{x}_{t+1}, \dots, \boldsymbol{x}_T \,|\boldsymbol{z}_t)$$

$$p(\boldsymbol{z}_{T+1}|\boldsymbol{z}_T, \boldsymbol{x}_1, \dots, \boldsymbol{x}_T) = p(\boldsymbol{z}_{T+1}|\boldsymbol{z}_T)$$

Filtering and smoothing

Filtering: $p(\boldsymbol{z}_t|\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t)$ Smoothing: $p(\boldsymbol{z}_t|\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)$

$$p(\boldsymbol{z}_t|\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T) = \frac{p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T|\boldsymbol{z}_t)p(\boldsymbol{z}_t)}{p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)}$$

$$= \frac{p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t|\boldsymbol{z}_t)p(\boldsymbol{x}_{t+1}, \ldots, \boldsymbol{x}_T|\boldsymbol{z}_t)p(\boldsymbol{z}_t)}{p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)}$$

$$= \frac{p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t, \boldsymbol{z}_t)p(\boldsymbol{x}_{t+1}, \ldots, \boldsymbol{x}_{\boldsymbol{T}}|\boldsymbol{z}_t)}{p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)}$$

$$= \frac{\alpha(\boldsymbol{z}_t)\beta(\boldsymbol{z}_t)}{p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)}$$

Forward Probabilities

$$\alpha(\boldsymbol{z}_t) = p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_t, \boldsymbol{z}_t)$$

$$= p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_t | \boldsymbol{z}_t) p(\boldsymbol{z}_t)$$

$$= p(\boldsymbol{x}_t | \boldsymbol{z}_t) p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1} | \boldsymbol{z}_t) p(\boldsymbol{z}_t)$$

$$= p(\boldsymbol{x}_t | \boldsymbol{z}_t) p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1}, \boldsymbol{z}_t)$$

$$= p(\boldsymbol{x}_t | \boldsymbol{z}_t) \sum_{\boldsymbol{z}_{t-1}} p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1}, \boldsymbol{z}_t, \boldsymbol{z}_{t-1})$$

$$= p(\boldsymbol{x}_t | \boldsymbol{z}_t) \sum_{\boldsymbol{z}_{t-1}} p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1}, \boldsymbol{z}_t | \boldsymbol{z}_{t-1}) p(\boldsymbol{z}_{t-1})$$

$$= p(\boldsymbol{x}_t | \boldsymbol{z}_t) \sum_{\boldsymbol{z}_{t-1}} p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1} | \boldsymbol{z}_{t-1}, \boldsymbol{z}_t) p(\boldsymbol{z}_t | \boldsymbol{z}_{t-1}) p(\boldsymbol{z}_{t-1})$$

Forward Probabilities

$$= p(\boldsymbol{x}_t|\boldsymbol{z}_t)\sum_{\boldsymbol{z}_{t-1}} p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{t-1}\,|\boldsymbol{z}_{t-1})p(\boldsymbol{z}_{t-1})p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1})$$

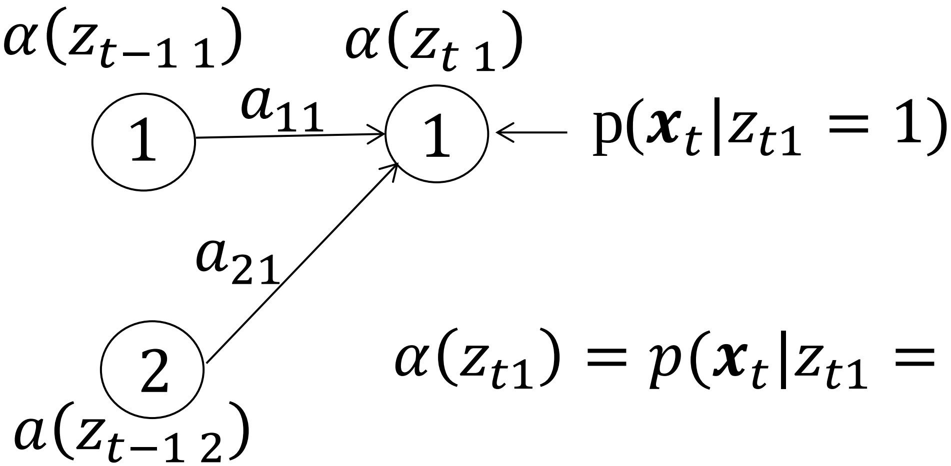
$$= p(\boldsymbol{x}_t|\boldsymbol{z}_t)\sum_{\boldsymbol{z}_{t-1}} p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{t-1}\,,\boldsymbol{z}_{t-1})p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1})$$

$$= p(\boldsymbol{x}_t|\boldsymbol{z}_t)\sum_{\boldsymbol{z}_{t-1}} \alpha(\boldsymbol{z}_{t-1})p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1})$$

Recursive formula and initial condition $\alpha(\boldsymbol{z}_1)$

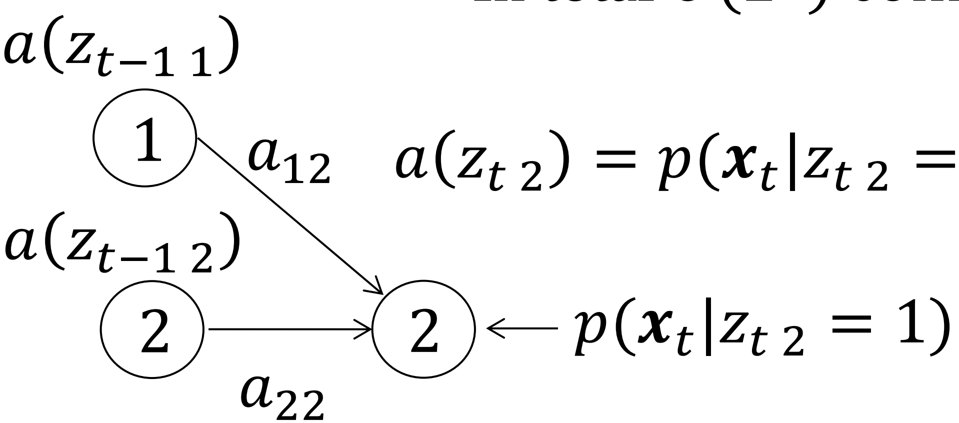
$$\alpha(\boldsymbol{z}_1) = p(\boldsymbol{x}_1,\boldsymbol{z}_1) = p(\boldsymbol{z}_1)p(\boldsymbol{x}_1|\boldsymbol{z}_1) = \prod_{k=1}^{2}\{\pi_k p(\boldsymbol{x}_1|z_{1k})\}^{z_{1k}}$$

Forward Probabilities



$$\alpha(z_{t1}) = p(\boldsymbol{x}_t|z_{t1} = 1)(a_{11}a(z_{t-1\,1}) + a_{21}a(z_{t-1\,2}))$$

In total $O(2^2)$ computation (in general $O(K^2)$)



$$a(z_{t\,2}) = p(\boldsymbol{x}_t|z_{t\,2} = 1)(a_{12}a(z_{t-1\,1}) + a_{22}a(z_{t-1\,2}))$$

Filtering

$$p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_t) = \sum_{\boldsymbol{z}_t} p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_t, \boldsymbol{z}_t) = \sum_{\boldsymbol{z}_t} \alpha(\boldsymbol{z}_t) = \alpha(z_{t1}) + \alpha(z_{t2})!!$$

Filtering $p(\boldsymbol{z}_t | \boldsymbol{x}_1, \dots, \boldsymbol{x}_t) = \dfrac{p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_t, \boldsymbol{z}_t)}{p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_t)} = \dfrac{\alpha(\boldsymbol{z}_t)}{\sum_{\boldsymbol{z}_t} \alpha(\boldsymbol{z}_t)} = \tilde{a}(\boldsymbol{z}_t)$

Evaluation: How can we compute $p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_T)$?

$$p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_T) = \sum_{\boldsymbol{z}_T} p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_T, \boldsymbol{z}_T) = \sum_{\boldsymbol{z}_T} \alpha(\boldsymbol{z}_T)$$

Backward probabilities

$$\beta(\mathbf{z}_t) = p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T \,|\mathbf{z}_t)$$

$$= \sum_{\mathbf{z}_{t+1}} p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T \,, \mathbf{z}_{t+1}|\mathbf{z}_t)$$

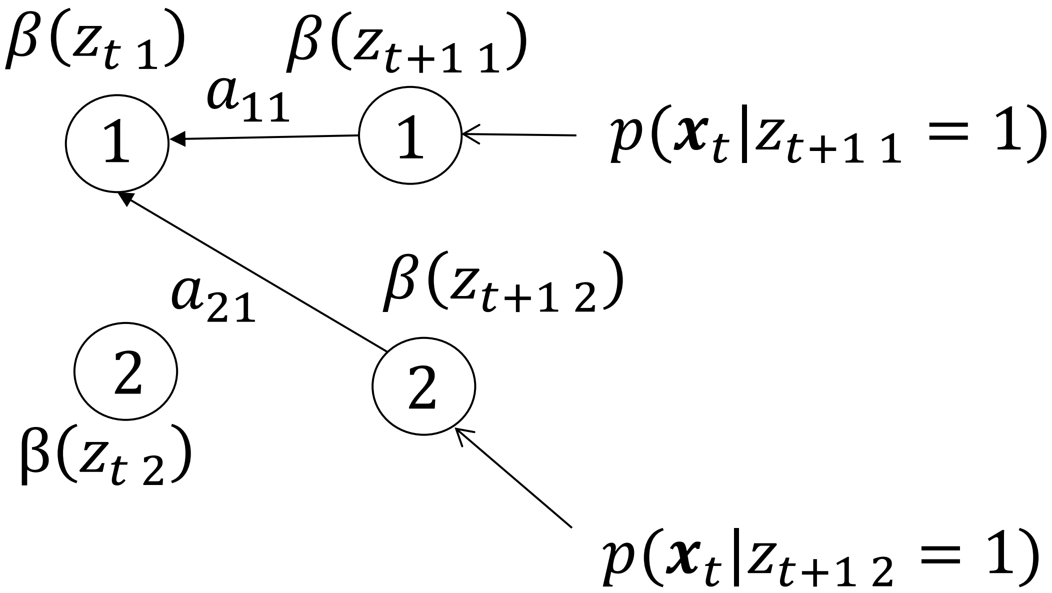
$$= \sum_{\mathbf{z}_{t+1}} p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T \,|\mathbf{z}_t, \mathbf{z}_{t+1}) p(\mathbf{z}_{t+1}|\mathbf{z}_t)$$

$$= \sum_{\mathbf{z}_{t+1}} p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T \,|\mathbf{z}_{t+1}) p(\mathbf{z}_{t+1}|\mathbf{z}_t)$$

$$= \sum_{\mathbf{z}_{t+1}} p(\mathbf{x}_{t+2}, \dots, \mathbf{x}_T \,|\mathbf{z}_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{z}_{t+1}) p(\mathbf{z}_{t+1}|\mathbf{z}_t)$$

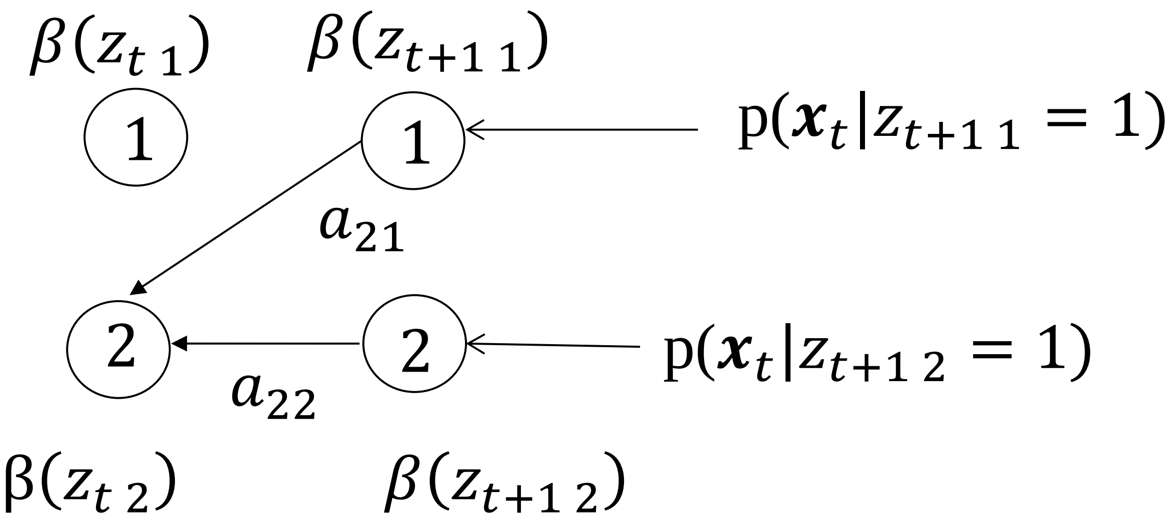
$$= \sum_{\mathbf{z}_{t+1}} \beta(\mathbf{z}_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{z}_{t+1}) p(\mathbf{z}_{t+1}|\mathbf{z}_t)$$

Computing Backward probabilities



$$\beta(z_{t\,1}) = \beta(z_{t+1\,1})a_{11}p(\boldsymbol{x}_t|z_{t+1\,1}=1)$$
$$+\beta(z_{t+1\,2})a_{12}p(\boldsymbol{x}_t|z_{t+1\,2}=1)$$

Computing Backward probabilities



$$\beta(z_{t\,2}) = \beta(z_{t+1\,1}) a_{21} p(\boldsymbol{x}_t|z_{t+1\,1}=1) + \beta(z_{t+1\,2}) a_{22} p(\boldsymbol{x}_t|z_{t+1\,2}=1)$$

$$p(\boldsymbol{z}_T|\boldsymbol{x}_1,\dots,\boldsymbol{x}_T) = \frac{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_T,\boldsymbol{z}_T)}{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_T)} = \frac{\alpha(\boldsymbol{z}_T) * 1}{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_T)} \Rightarrow \beta(\boldsymbol{z}_T) = 1$$

Prediction

$$p(\boldsymbol{z}_{t+2}|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_t) = \sum_{\boldsymbol{z}_{t+1}}\sum_{\boldsymbol{z}_t} p(\boldsymbol{z}_{t+2},\boldsymbol{z}_{t+1},\boldsymbol{z}_t|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_t)$$

$$= \sum_{\boldsymbol{z}_{t+1}}\sum_{\boldsymbol{z}_t} p(\boldsymbol{z}_{t+2},\boldsymbol{z}_{t+1}|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_t,\boldsymbol{z}_t)p(\boldsymbol{z}_t|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_t)$$

$$= \sum_{\boldsymbol{z}_{t+1}}\sum_{\boldsymbol{z}_t} p(\boldsymbol{z}_{t+2},\boldsymbol{z}_{t+1}|\boldsymbol{z}_t)\underbrace{p(\boldsymbol{z}_t|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_t)}_{\tilde{a}(\boldsymbol{z}_t)}$$

$$= \sum_{\boldsymbol{z}_{t+1}}\sum_{\boldsymbol{z}_t} p(\boldsymbol{z}_{t+2}|\boldsymbol{z}_{t+1})p(\boldsymbol{z}_{t+1}|\boldsymbol{z}_t)\tilde{a}(\boldsymbol{z}_t)$$

$$\Rightarrow \boldsymbol{A}^T\boldsymbol{A}^T\tilde{\boldsymbol{a}}$$

Prediction

and

$$p(\boldsymbol{x}_{t+2}|\boldsymbol{x}_1,\dots,\boldsymbol{x}_t) = \sum_{\boldsymbol{z}_{t+2}} p(\boldsymbol{x}_{t+2}|\boldsymbol{z}_{t+2})\, p(\boldsymbol{z}_{t+2}|\boldsymbol{x}_1,\dots,\boldsymbol{x}_t)$$

Smoothed transition

$$\xi(\boldsymbol{z}_{t-1}, \boldsymbol{z}_t) = p(\boldsymbol{z}_{t-1}, \boldsymbol{z}_t | \boldsymbol{x}_1, \dots, \boldsymbol{x}_T)$$

$$= \frac{p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_T | \boldsymbol{z}_{t-1}, \boldsymbol{z}_t) p(\boldsymbol{z}_{t-1}, \boldsymbol{z}_t)}{p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_T)}$$

$$= \frac{p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1} | \boldsymbol{z}_{t-1}) p(\boldsymbol{x}_t | \boldsymbol{z}_t) p(\boldsymbol{x}_{t+1}, \dots, \boldsymbol{x}_T | \boldsymbol{z}_t) p(\boldsymbol{z}_t | \boldsymbol{z}_{t-1}) p(\boldsymbol{z}_{t-1})}{p(\boldsymbol{x})}$$

$$= \frac{\alpha(\boldsymbol{z}_{t-1}) p(\boldsymbol{x}_t | \boldsymbol{z}_t) p(\boldsymbol{z}_t | \boldsymbol{z}_{t-1}) \beta(\boldsymbol{z}_t)}{p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_T)}$$

Parameter Estimation (Baum-Welch algorithm)

We need to define an EM algorithm

and we have all the necessary ingredients ☺

$$p(\boldsymbol{z}_t | \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{\boldsymbol{T}}) \qquad p(\boldsymbol{z}_{t-1}, \boldsymbol{z}_t | \boldsymbol{x}_1, \dots, \boldsymbol{x}_T)$$

Summary

We saw how to perform

(a) Filtering

(b) Smoothing

(c) Evaluation

(d) Prediction

Next we will see how to perform (e) EM and (f) Decoding