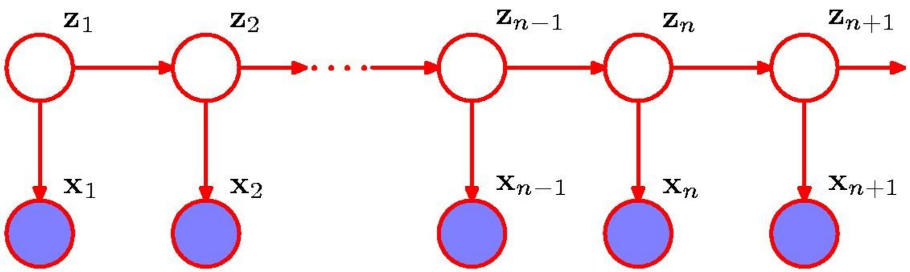
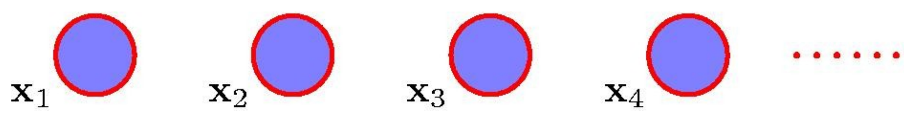
# Hidden Markov Models

## Terminology and Basic Algorithms

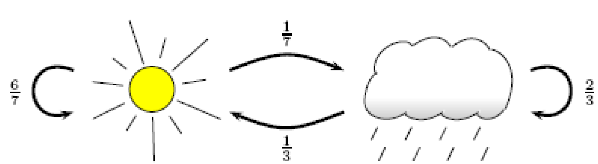


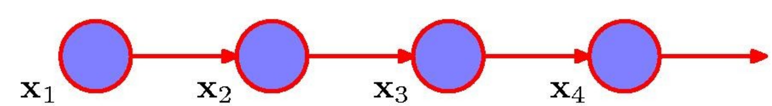
# Motivation

We make predictions based on models of observed data (machine learning). A simple model is that observations are assumed to be independent and identically distributed (iid) ...




but this assumption is not always the best, fx (1) measurements of weather patterns, (2) daily values of stocks, (3) acoustic features in successive time frames used for speech recognition, (4) the composition of texts, (5) the composition of DNA, or ...

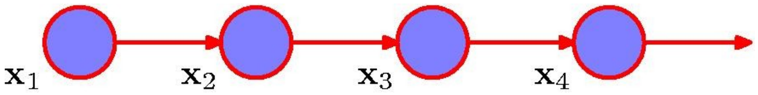


# Markov Models

If the *n*'th observation in a chain of observations is influenced only by the *n*-1'th observation, i.e.

$$p(\mathbf{x}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}) = p(\mathbf{x}_n|\mathbf{x}_{n-1})$$

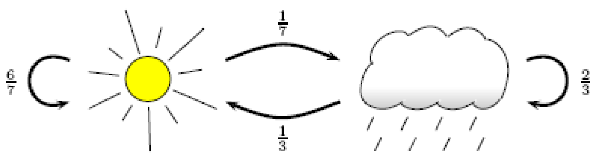



then the chain of observations is a 1st-order Markov chain, and the joint-probability of a sequence of *N* observations is

$$p(\mathbf{x}_1,\ldots,\mathbf{x}_N) = \prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}) = p(\mathbf{x}_1)\prod_{n=2}^{N} p(\mathbf{x}_n|\mathbf{x}_{n-1})$$

If the distributions $p(\mathbf{x}_n \mid \mathbf{x}_{n-1})$ are the same for all $n$, then the chain of observations is an *homogeneous* 1st-order Markov chain ...

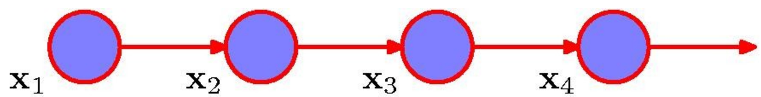
The model, i.e. $p(\mathbf{x}_n \mid \mathbf{x}_{n-1})$:




A sequence of observations:

If th y
by the $n$-1'th observation, i.e.

$$p(\mathbf{x}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}) = p(\mathbf{x}_n|\mathbf{x}_{n-1})$$

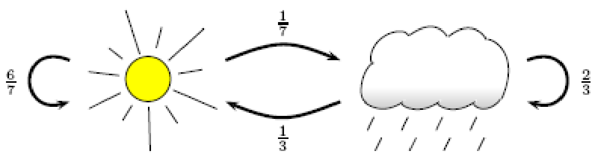



then the chain of observations is a 1st-order Markov chain, and the joint-probability of a sequence of $N$ observations is

$$p(\mathbf{x}_1,\ldots,\mathbf{x}_N) = \prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}) = p(\mathbf{x}_1)\prod_{n=2}^{N} p(\mathbf{x}_n|\mathbf{x}_{n-1})$$

If the distributions $p(\mathbf{x}_n \mid \mathbf{x}_{n-1})$ are the same for all $n$, then the chain of observations is an *homogeneous* 1st-order Markov chain ...

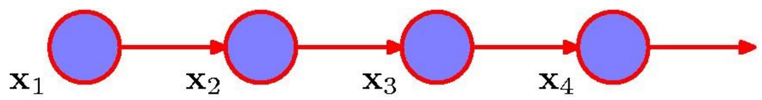
The model, i.e. $p(\mathbf{x}_n \mid \mathbf{x}_{n-1})$:



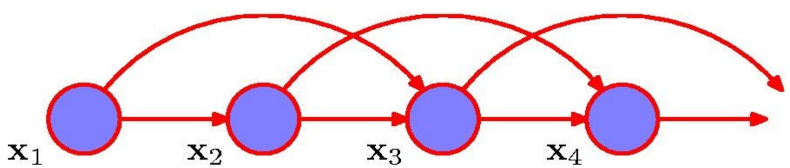

A sequence of observations:

If th … y
by the *n*-1'th observation, i.e.

$$p(\mathbf{x}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}) = p(\mathbf{x}_n|\mathbf{x}_{n-1})$$




## Extension – A higher order Markov chain




$$p(\mathbf{x}_1,\ldots,\mathbf{x}_N) = p(\mathbf{x}_1)p(\mathbf{x}_2|\mathbf{x}_1)\prod_{n=3}^{N} p(\mathbf{x}_n|\mathbf{x}_{n-1},\mathbf{x}_{n-2})$$

observations is an *homogeneous* 1st-order Markov chain ...

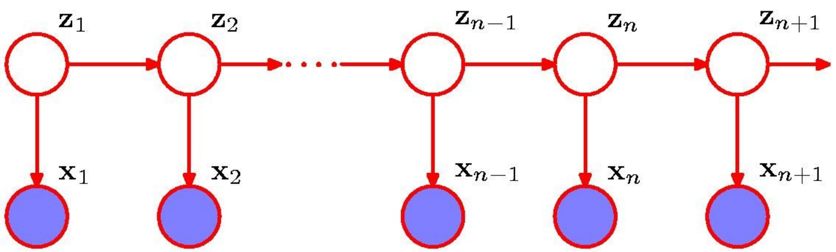
# Hidden Markov Models

What if the *n*'th observation in a chain of observations is influenced by a corresponding latent (i.e. hidden) variable?

Latent values

H H L L H

Observations

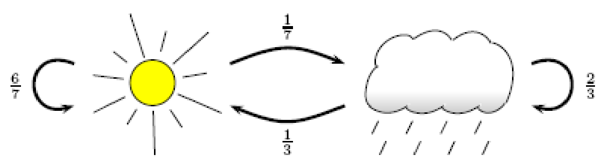



If the latent variables are discrete and form a Markov chain, then it is a hidden Markov model (HMM)

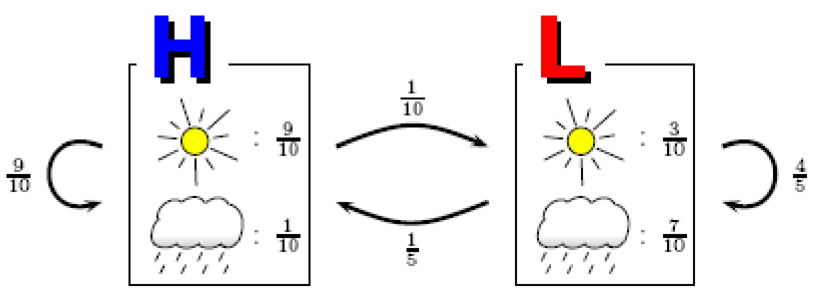
# Hidden Markov Models

What if the *n*'th observation in a chain of observations is influenced by a corresponding latent (i.e. hidden) variable?

Markov Model

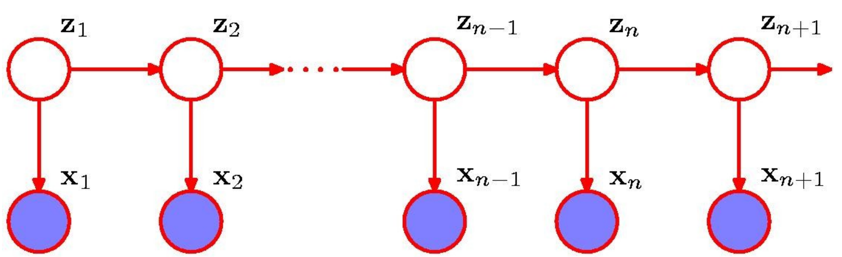



Hidden Markov Model




Latent values


Observations




If the latent variables are discrete and form a Markov chain, then it is a hidden Markov model (HMM)

## Computational problems

- Determine the likelihood of the sequence of observations
- Predict the next observation in the sequence of observations
- Find the most likely underlying explanation of the sequence of observation

# kov Models

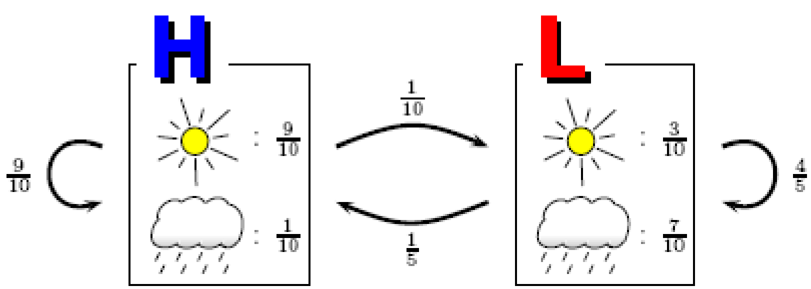
ain of observations is influenced
en) variable?

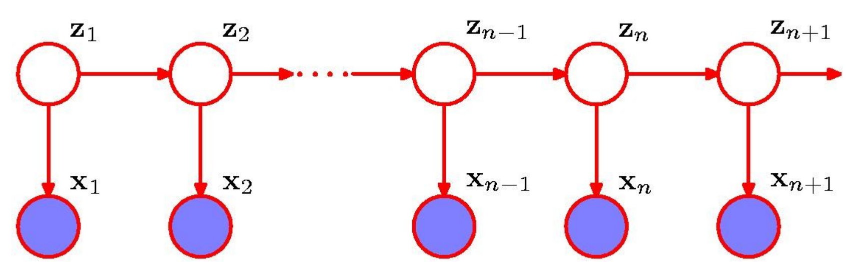
Latent values

H H L L H

Observations

Hidden Markov Model




If the latent variables are discrete and form a Markov chain, then it is a hidden Markov model (HMM)

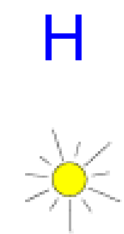
# Hidden Markov Models

What if the $n$'th observation in a chain of observations is influenced by a cor

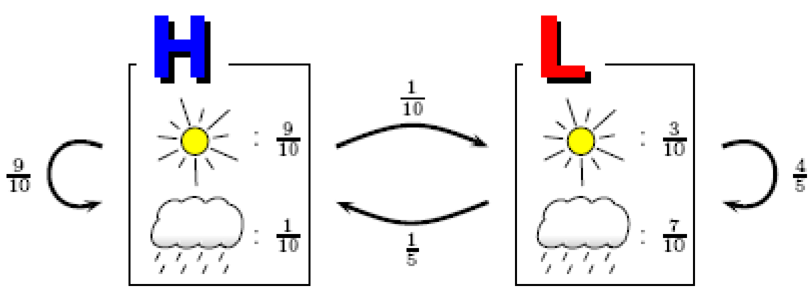
### The predictive distribution

$$p(\mathbf{x}_{n+1} \mid \mathbf{x}_1, \ldots, \mathbf{x}_n)$$

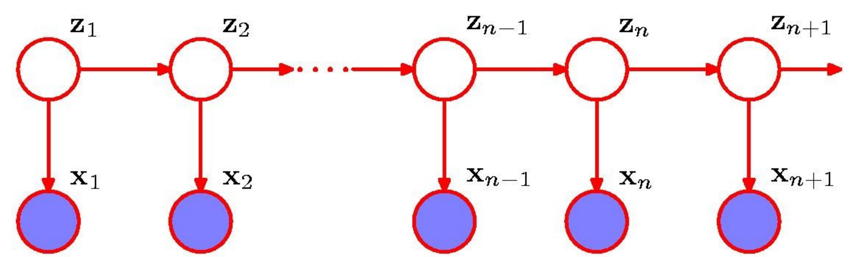
for observation $\mathbf{x}_{n+1}$ can be shown to depend on all previous observations, i.e. the sequence of observations is not a Markov chain of any order ...




Hidden Markov Model




Observations




If the latent variables are discrete and form a Markov chain, then it is a hidden Markov model (HMM)

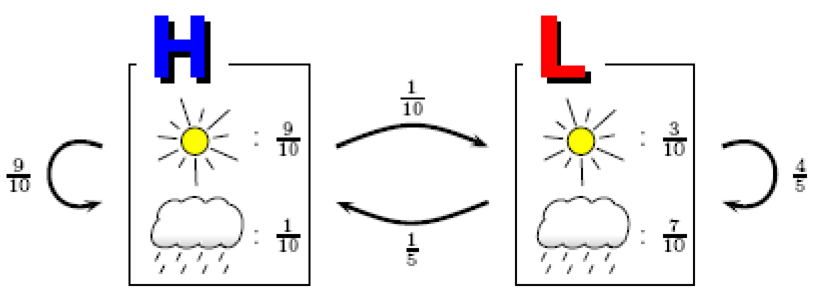
# Hidden Markov Models

What if the $n$'th observation in a chain of observations is influenced

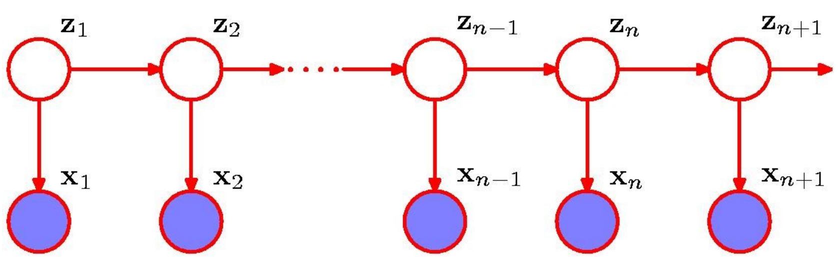
**The joint distribution**

$$p(\mathbf{x}_1,\ldots,\mathbf{x}_N,\mathbf{z}_1,\ldots,\mathbf{z}_N) = p(\mathbf{z}_1)\left[\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1})\right]\prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{z}_n)$$

Hidden Markov Model




Latent values




If the latent variables are discrete and form a Markov chain, then it is a hidden Markov model (HMM)

# Hidden Markov Models

What if … hain of o…

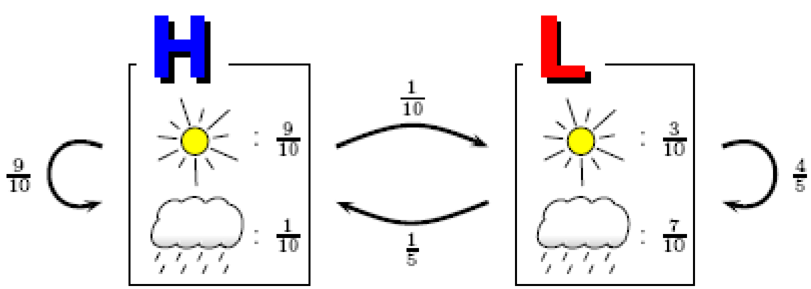
distribution

Transition probabilities

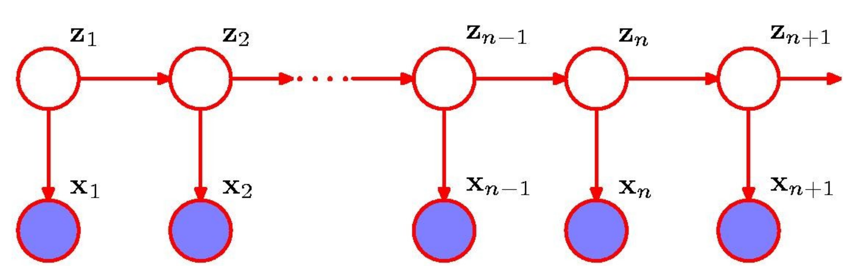
Emission probabilities

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{z}_1, \ldots, \mathbf{z}_N) = p(\mathbf{z}_1) \left[ \prod_{n=2}^{N} p(\mathbf{z}_n | \mathbf{z}_{n-1}) \right] \prod_{n=1}^{N} p(\mathbf{x}_n | \mathbf{z}_n)$$

Hidden Markov Model




Latent values




If the latent variables are discrete and form a Markov chain, then it is a hidden Markov model (HMM)

# Transition probabilities

**Notation:** In Bishop, the latent variables $\mathbf{z}_n$ are discrete variables, e.g. if $\mathbf{z}_n = (0,0,1)$ then the model in step $n$ is in state $k=3$ ...

**Transition probabilities:** If the latent variables are discrete with $K$ states, the conditional distribution $p(\mathbf{z}_n \mid \mathbf{z}_{n-1})$ is a $K$ x $K$ table $\mathbf{A}$, and the marginal distribution $p(\mathbf{z}_1)$ describing the initial state is a $K$ vector $\boldsymbol{\pi}$ ...

The probability of going from state $j$ to state $k$ is:

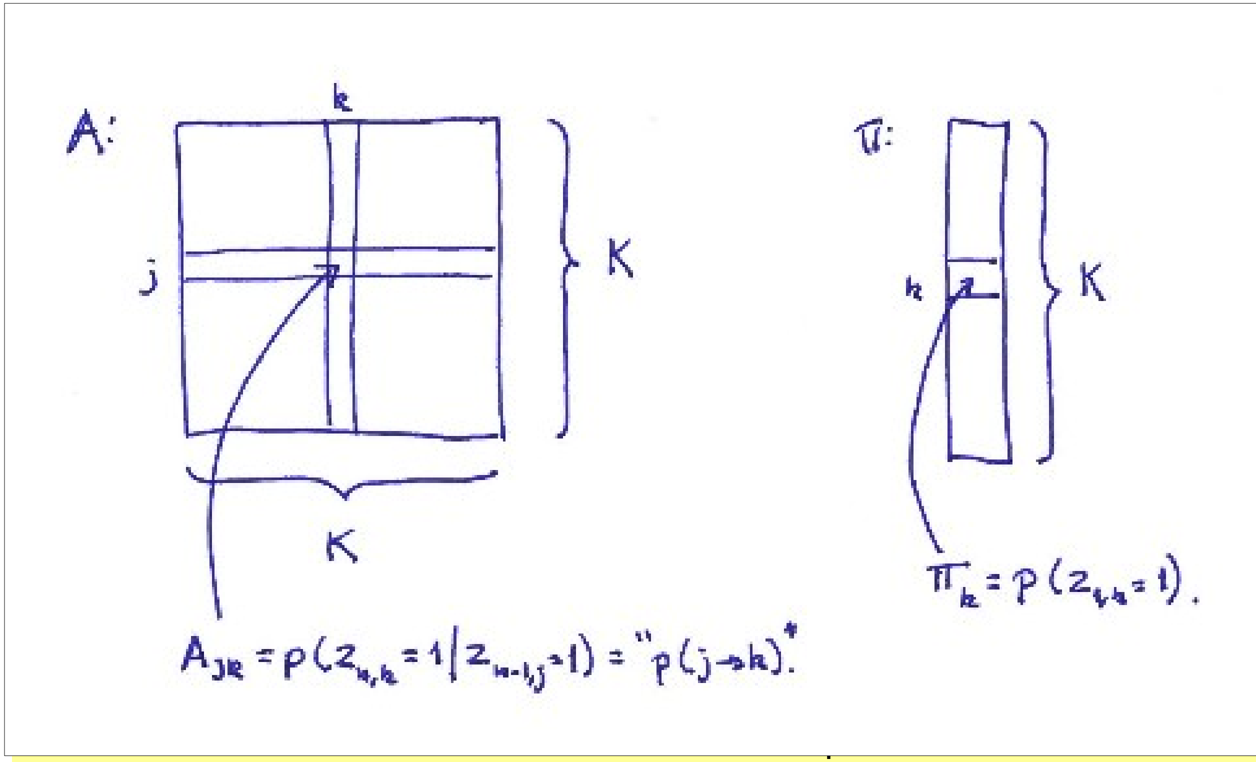
$$A_{jk} \equiv p(z_{nk} = 1 | z_{n-1,j} = 1)$$

$$\sum_k A_{jk} = 1$$

The probability of state $k$ being the initial state is:

$$\pi_k \equiv p(z_{1k} = 1)$$

$$\sum_k \pi_k = 1$$




# bilities

are discrete variables,
s in state $k$=3 ...

bles are discrete with $K$
is a $K$ x $K$ table **A**, and
initial state is a $K$
vector **π** ...

The probability of going from state $j$ to state $k$ is:

$$A_{jk} \equiv p(z_{nk} = 1 | z_{n-1,j} = 1)$$

$$\sum_k A_{jk} = 1$$

The probability of state $k$ being the initial state is:

$$\pi_k \equiv p(z_{1k} = 1)$$

$$\sum_k \pi_k = 1$$

**Notat** ... ables,
e.g. if

**Trans** ... th *K*
states ... and
the m
vecto

## The transition probabilities:

$$p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}) = \prod_{k=1}^{K} \prod_{j=1}^{K} A_{jk}^{z_{n-1,j} z_{nk}}$$

$$p(\mathbf{z}_1|\pi) = \prod_{k=1}^{K} \pi_k^{z_{1k}}$$

The probability of going from state *j* to state *k* is:

$$A_{jk} \equiv p(z_{nk} = 1|z_{n-1,j} = 1)$$

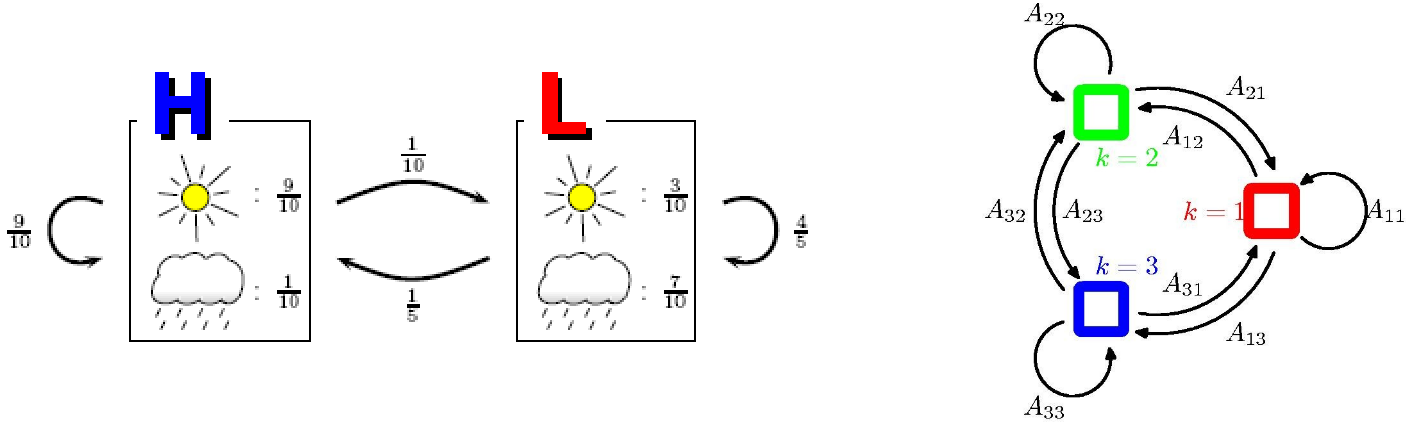
$$\sum_k A_{jk} = 1$$

The probability of state *k* being the initial state is:

$$\pi_k \equiv p(z_{1k} = 1)$$

$$\sum_k \pi_k = 1$$

## State transition diagram




**Notati** ... iables,
e.g. if **z** ...

**Transition probabilities:** If the latent variables are discrete with *K* states, the conditional distribution $p(\mathbf{z}_n \mid \mathbf{z}_{n-1})$ is a *K* x *K* table **A**, and the marginal distribution $p(\mathbf{z}_1)$ describing the initial state is a *K* vector $\boldsymbol{\pi}$ ...

The probability of going from state *j* to state *k* is:

$$A_{jk} \equiv p(z_{nk} = 1 | z_{n-1,j} = 1)$$

$$\sum_k A_{jk} = 1$$

The probability of state *k* being the initial state is:

$$\pi_k \equiv p(z_{1k} = 1)$$

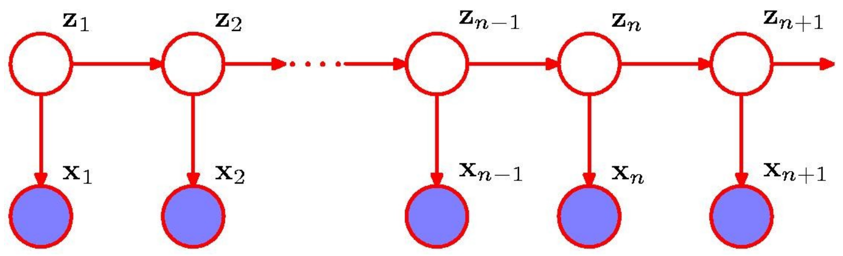
$$\sum_k \pi_k = 1$$

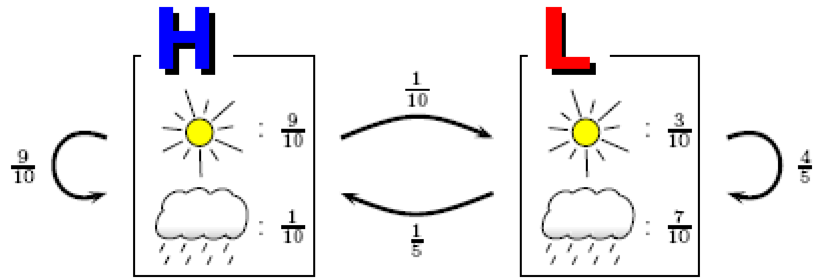
# Emission probabilities

**Emission probabilities:** The conditional distributions of the observed variables $p(\mathbf{x}_n \mid \mathbf{z}_n)$ from a specific state

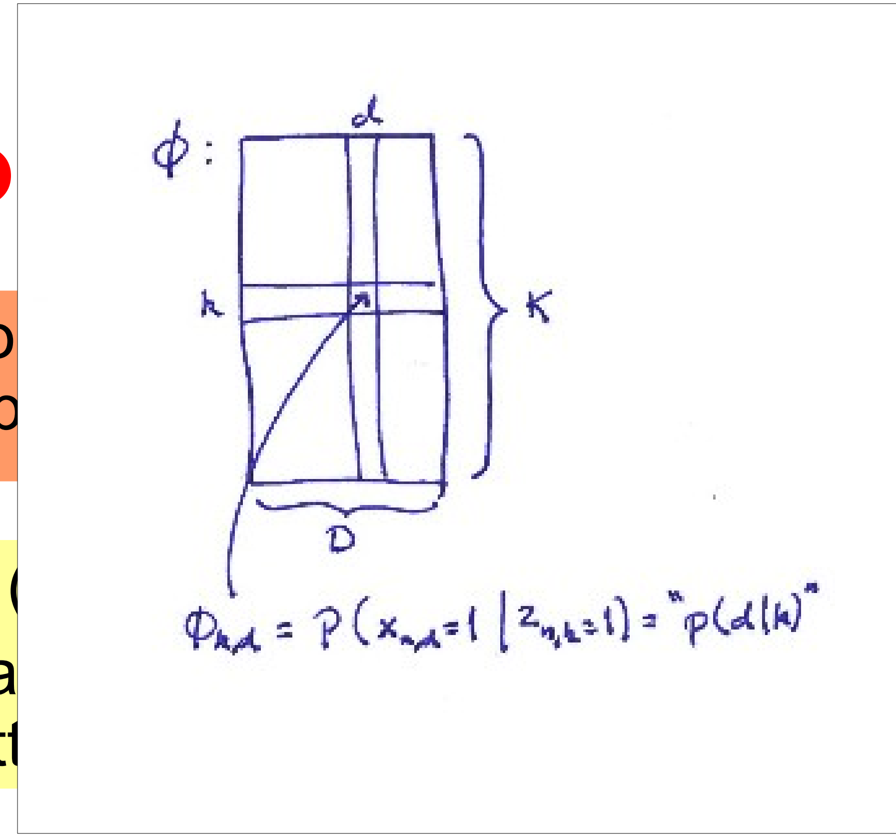
If the observed values $\mathbf{x}_n$ are discrete (e.g. *D* symbols), the emission probabilities $\boldsymbol{\Phi}$ is a *K*x*D* table of probabilities which for each of the *K* states specifies the probability of emitting each observable ...

$$p(\mathbf{x}_n|\mathbf{z}_n, \phi) = \prod_{k=1}^{K} p(\mathbf{x}_n|\phi_k)^{z_{nk}}$$



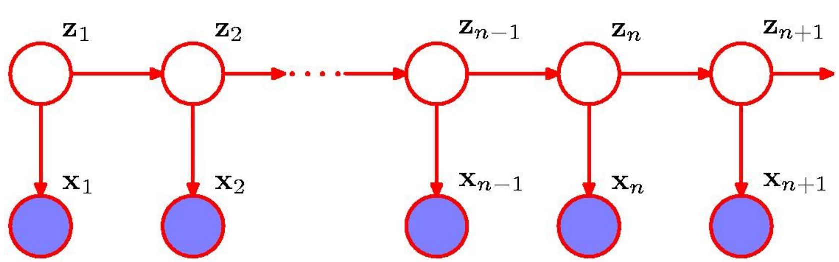
# Emission pro

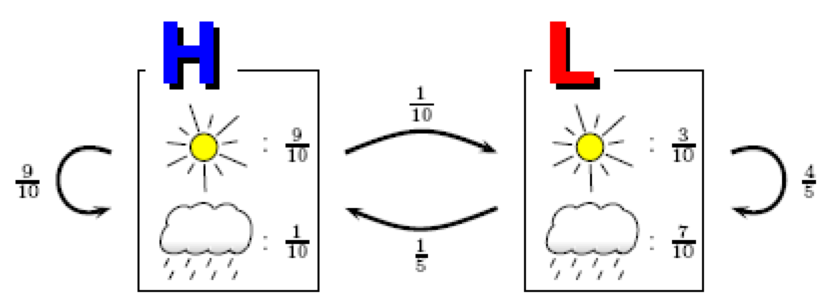



**Emission probabilities:** The conditio
observed variables $p(\mathbf{x}_n \mid \mathbf{z}_n)$ from a sp

If the observed values $\mathbf{x}_n$ are discrete (
probabilities ***Φ*** is a *K*x*D* table of proba
states specifies the probability of emitt

$$p(\mathbf{x}_n|\mathbf{z}_n,\phi) = \prod_{k=1}^{K} p(\mathbf{x}_n|\phi_k)^{z_{nk}}$$



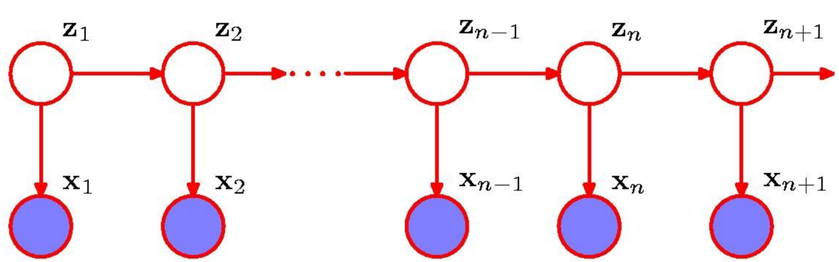
# Emission probabilities

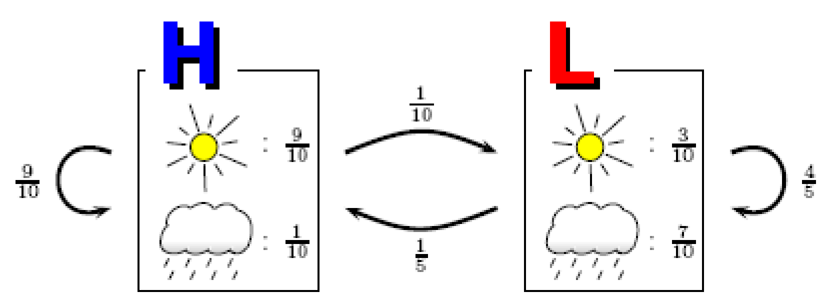
**Emission probabilities:** The conditional distributions of the observed variables $p(\mathbf{x}_n \mid \mathbf{z}_n)$ from a specific state

If the observed values $\mathbf{x}_n$ are di
probabilities $\boldsymbol{\Phi}$ is a $K$x$D$ table c
states specifies the probability

$z_{nk}$ = 1 iff the $n$'th latent variable in the sequence is in state $k$, otherwise it is 0, i.e. the product just “picks” the emission probabilities corresponding to state $k$ ...

$$p(\mathbf{x}_n|\mathbf{z}_n, \phi) = \prod_{k=1}^{K} p(\mathbf{x}_n|\phi_k)^{z_{nk}}$$



# HMM joint probability distribution

$$p(\mathbf{X},\mathbf{Z}|\mathbf{\Theta}) = p(\mathbf{z}_1|\pi)\left[\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1},\mathbf{A})\right]\prod_{n-1}^{N} p(\mathbf{x}_n|\mathbf{z}_n,\phi)$$

Observables:

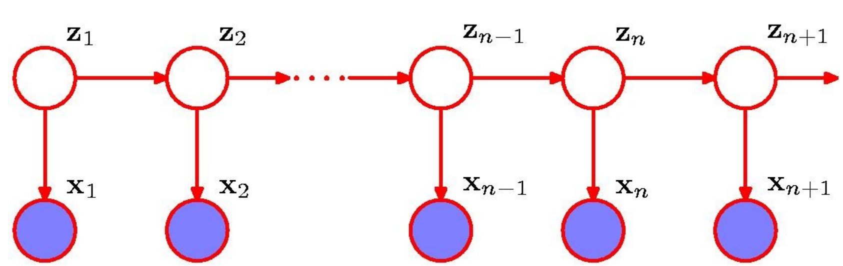
$$\mathbf{X} = \{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$$

Latent states:

$$\mathbf{Z} = \{\mathbf{z}_1,\ldots,\mathbf{z}_N\}$$

Model parameters:

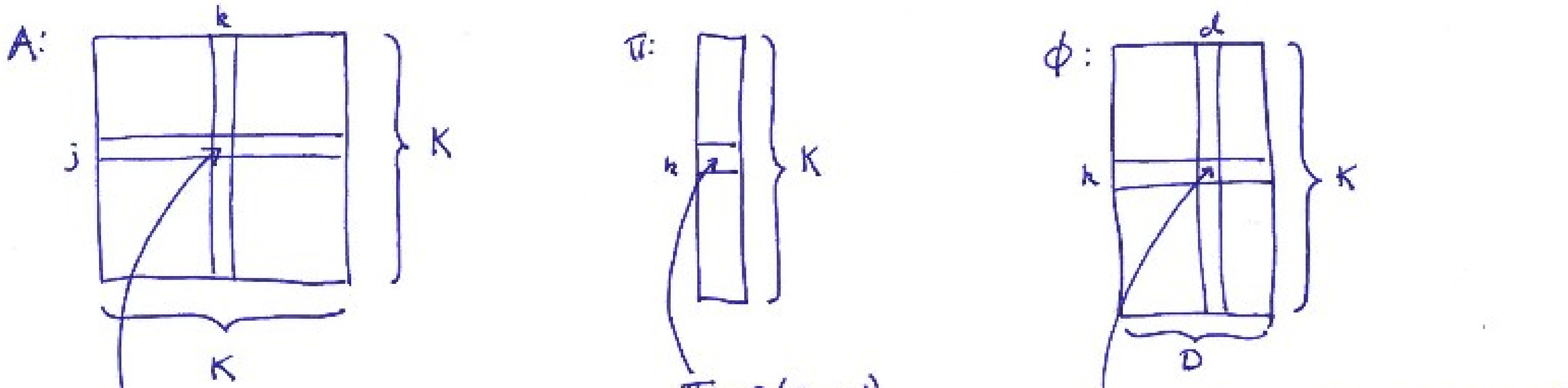
$$\mathbf{\Theta} = \{\pi,\mathbf{A},\phi\}$$




If A and ***Φ*** are the same for all *n* then the HMM is *homogeneous*

# HMM joint probability distribution

$$p(\mathbf{X}, \mathbf{Z}|\mathbf{\Theta}) = p(\mathbf{z}_1|\pi) \left[ \prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}) \right] \prod_{n-1}^{N} p(\mathbf{x}_n|\mathbf{z}_n, \phi)$$




$A_{jk} = p(z_{n,k}=1|z_{n-1,j}=1) =$ "$p(j \to k)$"

$\pi_k = p(z_{1,k}=1)$.

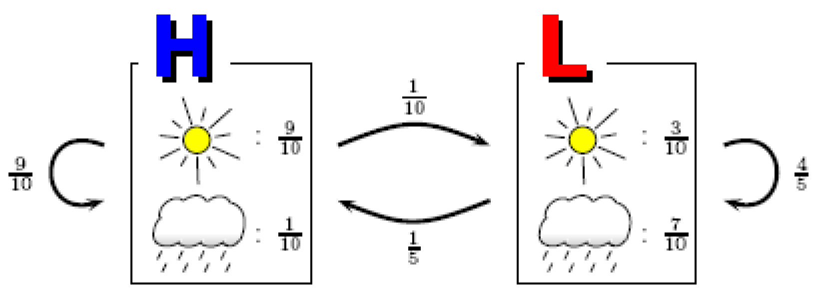
$\phi_{k,d} = P(x_{n,d}=1|z_{n,k}=1) =$ "$p(d|k)$"

If A and ***Φ*** are the same for all *n* then the HMM is *homogeneous*

# HMMs as a generative model

A HMM ***generates a sequence of observables*** by moving from latent state to latent state according to the transition probabilities and ***emitting an observable*** (from a discrete set of observables, i.e. a finite alphabet) from each latent state visited ***according to the emission probabilities*** of the state ...

Model *M*:




A run follows a sequence of states:

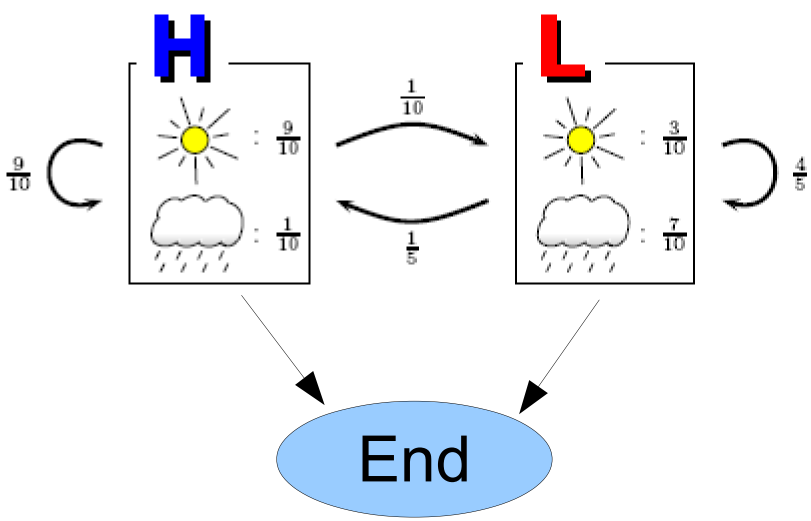
H H L L H

And emits a sequence of symbols:

# HMMs as a generative model

A HMM ***generates a sequence of observables*** by moving from latent state to latent state according to the transition probabilities and ***emitting an observable*** (from a discrete set of observables, i.e. a finite alphabet) from each latent state visited ***according to the emission probabilities*** of the state ...

Model *M*:




A run follows a sequence of states:

H H L L H

And emits a sequence of symbols:

A special **End**-state can be added to generate finite output

# Using HMMs

- Determine the likelihood of the sequence of observations
- Predict the next observation in the sequence of observations
- Find the most likely underlying explanation of the sequence of observation

# Using HMMs

- Determine the likelihood of the sequence of observations
- Predict the next observation in the sequence of observations
- Find the most likely underlying explanation of the sequence of observation

$$p(\mathbf{X}|\mathbf{\Theta}) = \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\mathbf{\Theta})$$

# Using HMMs

- Determine the likelihood of the sequence of observations
- Predict the next observation in the sequence of observations
- Find the most likely underlying explanation of the sequence of observation

$$p(\mathbf{X}|\mathbf{\Theta}) = \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\mathbf{\Theta})$$

The sum has $K^N$ terms, but it can be computed in $O(K^2N)$ time ...

# The forward-backward algorithm

$\alpha(\mathbf{z}_n)$ is the joint probability of observing $\mathbf{x}_1,\ldots,\mathbf{x}_n$ and being in state $\mathbf{z}_n$

$$\alpha(\mathbf{z}_n) \equiv p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n)$$

$\beta(\mathbf{z}_n)$ is the conditional probability of future observation $\mathbf{x}_{n+1},\ldots,\mathbf{x}_N$ assuming being in state $\mathbf{z}_n$

$$\beta(\mathbf{z}_n) \equiv p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{z}_n)$$

# The forward-backward algorithm

$\alpha(\mathbf{z}_n)$ is the joint probability of observing $\mathbf{x}_1,\ldots,\mathbf{x}_n$ and being in state $\mathbf{z}_n$

$$\alpha(\mathbf{z}_n) \equiv p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n)$$

$\beta(\mathbf{z}_n)$ is the conditional probability of future observation $\mathbf{x}_{n+1},\ldots,\mathbf{x}_N$ assuming being in state $\mathbf{z}_n$

$$\beta(\mathbf{z}_n) \equiv p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{z}_n)$$

Using $\alpha(\mathbf{z}_n)$ and $\beta(\mathbf{z}_n)$ we get the likelihood of the observations as:

$$p(\mathbf{X}) = \sum_{\mathbf{z}_n} \alpha(\mathbf{z}_n)\beta(\mathbf{z}_n) \qquad p(\mathbf{X}) = \sum_{\mathbf{z}_N} \alpha(\mathbf{z}_N)$$

# The forward algorithm

$\alpha(\mathbf{z}_n)$ is the joint probability of observing $\mathbf{x}_1,\ldots,\mathbf{x}_n$ and being in state $\mathbf{z}_n$

$$\alpha(\mathbf{z}_n) \equiv p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n)$$

# The α-recursion

$$
\begin{aligned}
\alpha(\mathbf{z}_n) &= p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n) \\
&= p(\mathbf{x}_1, \ldots, \mathbf{x}_n | \mathbf{z}_n) p(\mathbf{z}_n) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1} | \mathbf{z}_n) p(\mathbf{z}_n) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{z}_n) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{z}_{n-1}, \mathbf{z}_n) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{z}_n | \mathbf{z}_{n-1}) p(\mathbf{z}_{n-1}) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1} | \mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1}) p(\mathbf{z}_{n-1}) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1}) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} \alpha(\mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1})
\end{aligned}
$$

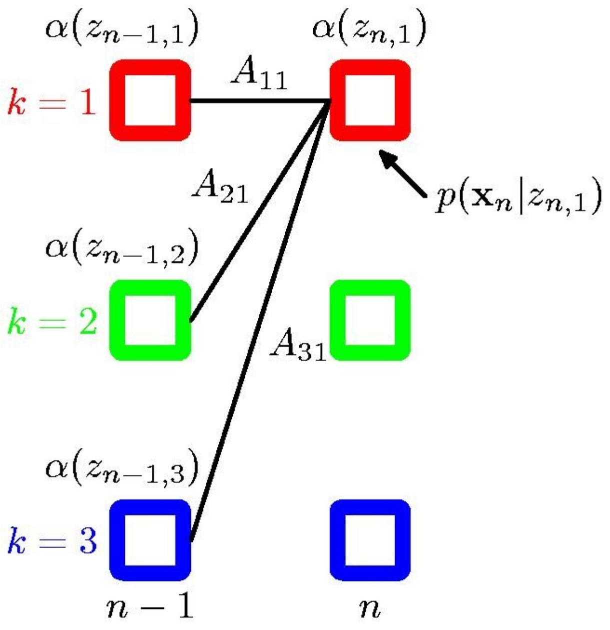
# The α-recursion

$$
\begin{aligned}
\alpha(\mathbf{z}_n) &= p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n) \\
&= p(\mathbf{x}_1, \ldots, \mathbf{x}_n | \mathbf{z}_n) p(\mathbf{z}_n) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1} | \mathbf{z}_n) p(\mathbf{z}_n) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{z}_n) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{z}_{n-1}, \mathbf{z}_n) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{z}_n | \mathbf{z}_{n-1}) p(\mathbf{z}_{n-1}) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1} | \mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1}) p(\mathbf{z}_{n-1}) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1}) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} \alpha(\mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1})
\end{aligned}
$$

# The forward algorithm

$\alpha(\mathbf{z}_n)$ is the joint probability of observing $\mathbf{x}_1,\ldots,\mathbf{x}_n$ and being in state $\mathbf{z}_n$

$$\alpha(\mathbf{z}_n) \equiv p(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_n)$$



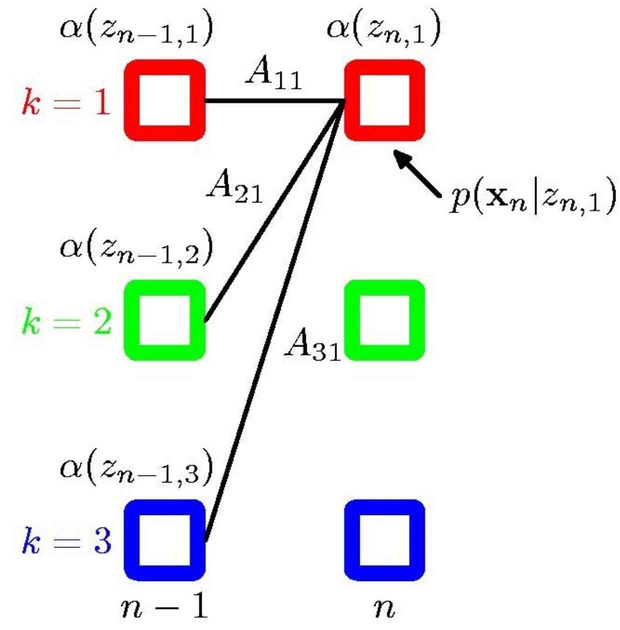

**Recursion:**

$$\alpha(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n)\sum_{\mathbf{z}_{n-1}}\alpha(\mathbf{z}_{n-1})p(\mathbf{z}_n|\mathbf{z}_{n-1})$$

# The forward algorithm

$\alpha(\mathbf{z}_n)$ is the joint probability of observing $\mathbf{x}_1,\ldots,\mathbf{x}_n$ and being in state $\mathbf{z}_n$

$$\alpha(\mathbf{z}_n) \equiv p(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_n)$$




## Recursion:

$$\alpha(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n)\sum_{\mathbf{z}_{n-1}}\alpha(\mathbf{z}_{n-1})p(\mathbf{z}_n|\mathbf{z}_{n-1})$$

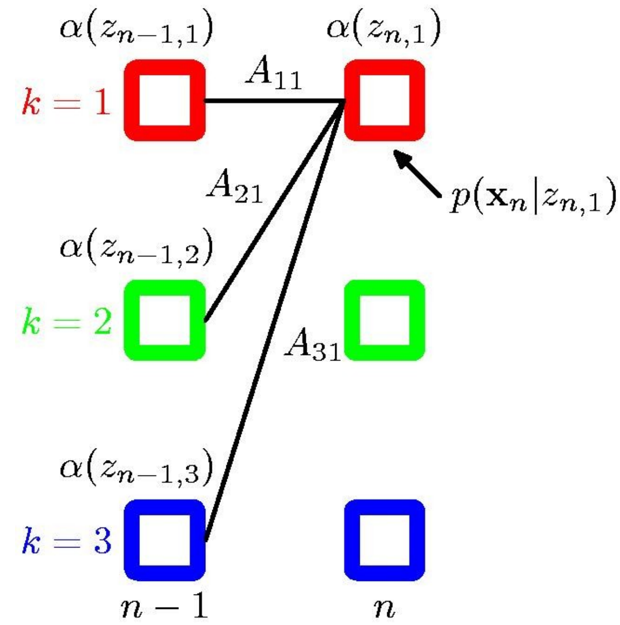
## Basis:

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

# The forward algorithm

$\alpha(\mathbf{z}_n)$ is the joint probability of observing $\mathbf{x}_1,\ldots,\mathbf{x}_n$ and being in state $\mathbf{z}_n$

$$\alpha(\mathbf{z}_n) \equiv p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n)$$




**Recursion:**

$$\alpha(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} \alpha(\mathbf{z}_{n-1}) p(\mathbf{z}_n|\mathbf{z}_{n-1})$$

**Basis:**

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

Takes time $O(K^2N)$ and space $O(KN)$ using memorization

# The backward algorithm

$\beta(\mathbf{z}_n)$ is the conditional probability of future observation $\mathbf{x}_{n+1},\ldots,\mathbf{x}_N$ assuming being in state $\mathbf{z}_n$

$$\beta(\mathbf{z}_n) \equiv p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{z}_n)$$

# The $\beta$-recursion

$$\begin{aligned}
\beta(\mathbf{z}_n) &= p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N, \mathbf{z}_{n+1} | \mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{z}_n, \mathbf{z}_{n+1}) p(\mathbf{z}_{n+1} | \mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{z}_{n+1}) p(\mathbf{z}_{n+1} | \mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+2}, \ldots, \mathbf{x}_N | \mathbf{z}_{n+1}) p(\mathbf{x}_{n+1} | \mathbf{z}_{n+1}) p(\mathbf{z}_{n+1} | \mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} \beta(\mathbf{z}_{n+1}) p(\mathbf{x}_{n+1} | \mathbf{z}_{n+1}) p(\mathbf{z}_{n+1} | \mathbf{z}_n)
\end{aligned}$$

# The β-recursion

$$
\begin{aligned}
\beta(\mathbf{z}_n) &= p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N, \mathbf{z}_{n+1} | \mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{z}_n, \mathbf{z}_{n+1}) p(\mathbf{z}_{n+1} | \mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{z}_{n+1}) p(\mathbf{z}_{n+1} | \mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+2}, \ldots, \mathbf{x}_N | \mathbf{z}_{n+1}) p(\mathbf{x}_{n+1} | \mathbf{z}_{n+1}) p(\mathbf{z}_{n+1} | \mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} \beta(\mathbf{z}_{n+1}) p(\mathbf{x}_{n+1} | \mathbf{z}_{n+1}) p(\mathbf{z}_{n+1} | \mathbf{z}_n)
\end{aligned}
$$

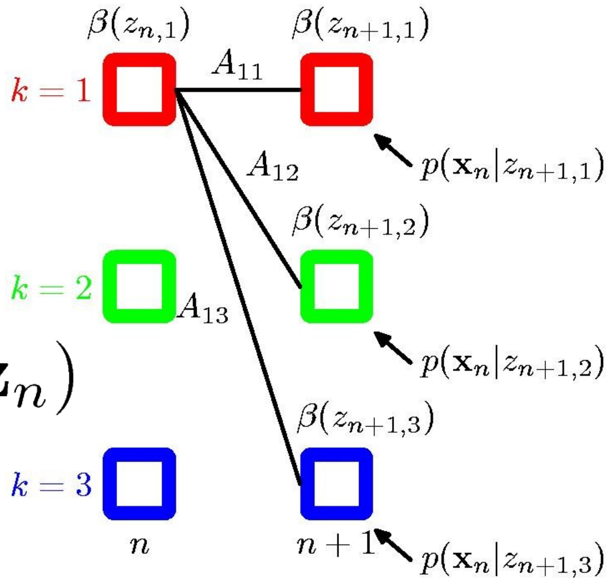
# The backward algorithm

$\beta(\mathbf{z}_n)$ is the conditional probability of future observation $\mathbf{x}_{n+1},\ldots,\mathbf{x}_N$ assuming being in state $\mathbf{z}_n$

$$\beta(\mathbf{z}_n) \equiv p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)$$

**Recursion:**

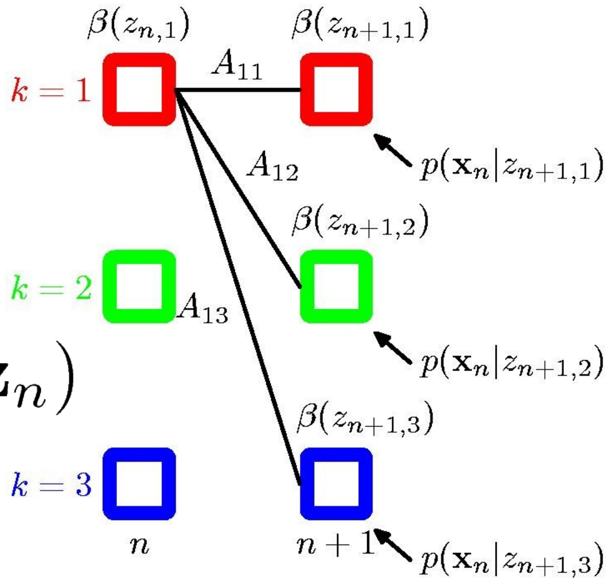
$$\beta(\mathbf{z}_n) = \sum_{\mathbf{z}_{n+1}} \beta(\mathbf{z}_{n+1}) p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1}) p(\mathbf{z}_{n+1}|\mathbf{z}_n)$$



# The backward algorithm

$\beta(\mathbf{z}_n)$ is the conditional probability of future observation $\mathbf{x}_{n+1},\ldots,\mathbf{x}_N$ assuming being in state $\mathbf{z}_n$

$$\beta(\mathbf{z}_n) \equiv p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)$$




**Recursion:**

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

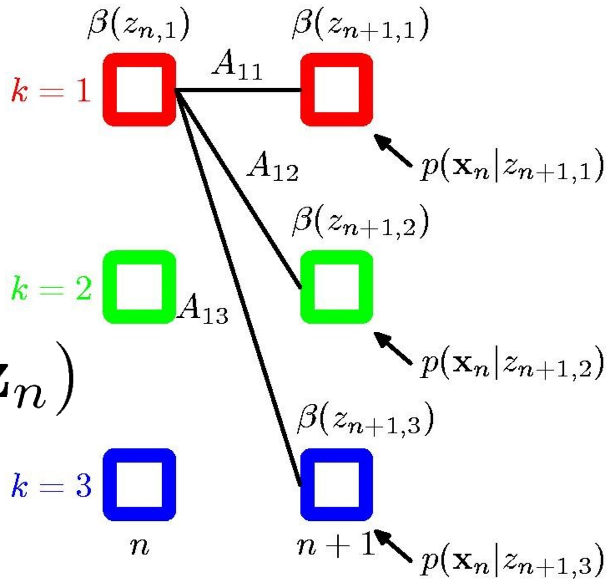
**Basis:**

$$\beta(\mathbf{z}_N) = 1$$

# The backward algorithm

$\beta(\mathbf{z}_n)$ is the conditional probability of future observation $\mathbf{x}_{n+1},\ldots,\mathbf{x}_N$ assuming being in state $\mathbf{z}_n$

$$\beta(\mathbf{z}_n) \equiv p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)$$




**Recursion:**

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

**Basis:**

$$\beta(\mathbf{z}_N) = 1$$

Takes time $O(K^2N)$ and space $O(KN)$ using memorization

# Using HMMs

- Determine the likelihood of the sequence of observations
- Predict the next observation in the sequence of observations
- Find the most likely underlying explanation of the sequence of observation

$$p(\mathbf{x}_{N+1}|\mathbf{X})$$

# Predicting the next observation

$$
\begin{aligned}
p(\mathbf{x}_{N+1}|\mathbf{X}) &= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}, \mathbf{z}_{N+1}|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) p(\mathbf{z}_{N+1}|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}, \mathbf{z}_N|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}|\mathbf{z}_N) p(\mathbf{z}_N|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}|\mathbf{z}_N) \frac{p(\mathbf{z}_N, \mathbf{X})}{p(\mathbf{X})} \\
&= \frac{1}{p(\mathbf{X})} \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}|\mathbf{z}_N) \alpha(\mathbf{z}_N)
\end{aligned}
$$

# Predicting the next observation

$$
\begin{aligned}
p(\mathbf{x}_{N+1}|\mathbf{X}) &= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}, \mathbf{z}_{N+1}|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) p(\mathbf{z}_{N+1}|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}, \mathbf{z}_N|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}|\mathbf{z}_N) p(\mathbf{z}_N|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}|\mathbf{z}_N) \frac{p(\mathbf{z}_N, \mathbf{X})}{p(\mathbf{X})} \\
&= \frac{1}{p(\mathbf{X})} \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}|\mathbf{z}_N) \alpha(\mathbf{z}_N)
\end{aligned}
$$

# Predicting the next observation

$$
\begin{aligned}
p(\mathbf{x}_{N+1}|\mathbf{X}) &= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}, \mathbf{z}_{N+1}|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) p(\mathbf{z}_{N+1}|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}, \mathbf{z}_N|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}|\mathbf{z}_N) p(\mathbf{z}_N|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}|\mathbf{z}_N) \frac{p(\mathbf{z}_N, \mathbf{X})}{p(\mathbf{X})} \\
&= \frac{1}{p(\mathbf{X})} \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}|\mathbf{z}_N) \alpha(\mathbf{z}_N)
\end{aligned}
$$

$$
p(\mathbf{X}) = \sum_{\mathbf{z}_N} \alpha(\mathbf{z}_N)
$$

# Using HMMs

- Determine the likelihood of the sequence of observations
- Predict the next observation in the sequence of observations
- Find the most likely underlying explanation of the sequence of observation

$$\mathbf{Z}^* = \arg\max_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z} | \mathbf{\Theta})$$

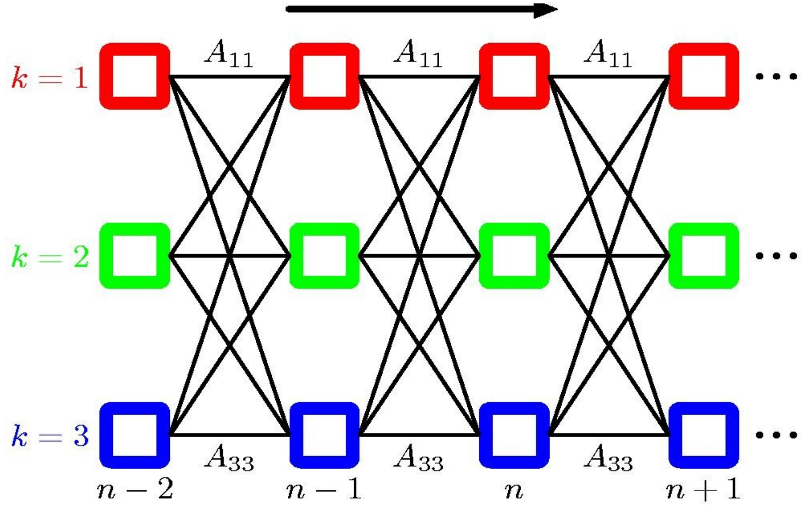
**The Viterbi algorithm:** Finds the most probable sequence of states generating the observations ...

# The Viterbi algorithm

$\omega(\mathbf{z}_n)$ is the probability of the most likely sequence of states $\mathbf{z}_1,\ldots,\mathbf{z}_n$ generating the observations $\mathbf{x}_1,\ldots,\mathbf{x}_n$

$$\omega(\mathbf{z}_n) \equiv \max_{\mathbf{z}_1,\ldots,\mathbf{z}_{n-1}} p(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_1,\ldots,\mathbf{z}_n)$$

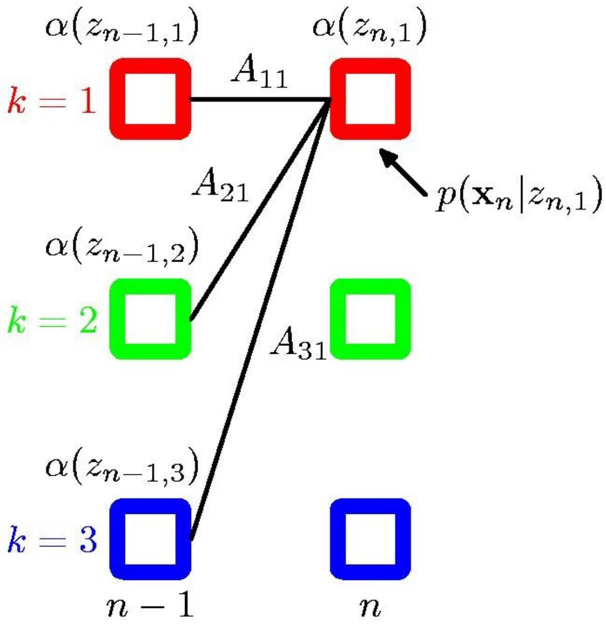
**Intuition:** Find the “longest path” from column 1 to column n, where “length” is its total probability, i.e. the probability of the transitions and emissions along the path ...



# The Viterbi algorithm

$\omega(\mathbf{z}_n)$ is the probability of the most likely sequence of states $\mathbf{z}_1,\ldots,\mathbf{z}_n$ generating the observations $\mathbf{x}_1,\ldots,\mathbf{x}_n$

$$\omega(\mathbf{z}_n) \equiv \max_{\mathbf{z}_1,\ldots,\mathbf{z}_{n-1}} p(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_1,\ldots,\mathbf{z}_n)$$



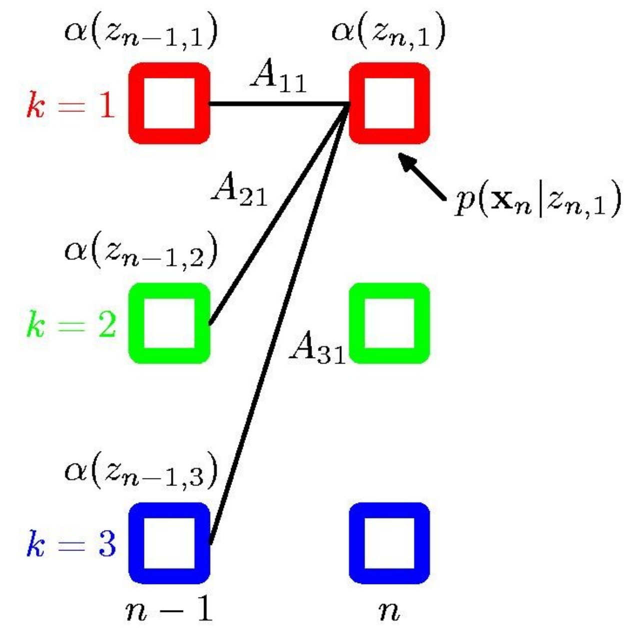

**Recursion:**

$$\omega(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n)\max_{\mathbf{z}_{n-1}}\omega(\mathbf{z}_{n-1})p(\mathbf{z}_n|\mathbf{z}_{n-1})$$

# The Viterbi algorithm

$\omega(\mathbf{z}_n)$ is the probability of the most likely sequence of states $\mathbf{z}_1,\ldots,\mathbf{z}_n$ generating the observations $\mathbf{x}_1,\ldots,\mathbf{x}_n$

$$\omega(\mathbf{z}_n) \equiv \max_{\mathbf{z}_1,\ldots,\mathbf{z}_{n-1}} p(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_1,\ldots,\mathbf{z}_n)$$




**Recursion:**

$$\omega(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n) \max_{\mathbf{z}_{n-1}} \omega(\mathbf{z}_{n-1}) p(\mathbf{z}_n|\mathbf{z}_{n-1})$$

**Basis:**

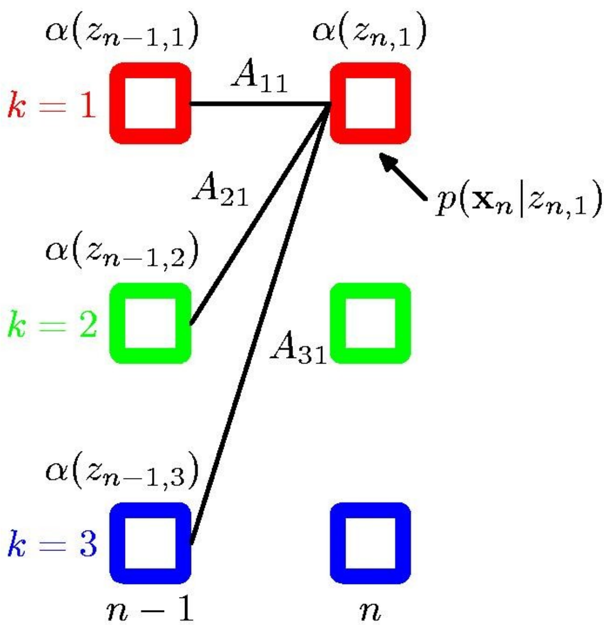
$$\omega(\mathbf{z}_1) = p(\mathbf{x}_1,\mathbf{z}_1) = p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1) = \prod_{k=1}^{K} \{\pi_k p(\mathbf{x}_1|\phi_k)\}^{z_{1k}}$$

Takes time O($K^2N$) and space O($KN$) using memorization

# The Viterbi algorithm

$\omega(\mathbf{z}_n)$ is the probability of the most likely sequence of states $\mathbf{z}_1,\ldots,\mathbf{z}_n$ generating the observations $\mathbf{x}_1,\ldots,\mathbf{x}_n$

$$\omega(\mathbf{z}_n) \equiv \max_{\mathbf{z}_1,\ldots,\mathbf{z}_{n-1}} p(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_1,\ldots,\mathbf{z}_n)$$




**Recursion:**

$$\omega(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n) \max_{\mathbf{z}_{n-1}} \omega(\mathbf{z}_{n-1}) p(\mathbf{z}_n|\mathbf{z}_{n-1})$$

**Basis:**

$$\omega(\ldots \prod_{k=1}^{K} \ldots$$

The path itself can be retrieve in time O($KN$) by backtracking

Takes time O($K^2N$) and space O($KN$) using memorization

# Summary

- Introduced hidden Markov models (**HMMs**)
- The **forward-backward algorithms** for determining the likelihood of a sequence of observations, and predicting the next observation in a sequence of observations.
- The **Viterbi-algorithm** for finding the most likely underlying explanation (sequence of latent states) of a sequence of observation
- **Next:** How to implement the basic algorihtms (forward, backward, and Viterbi) in a “numerically” sound manner.