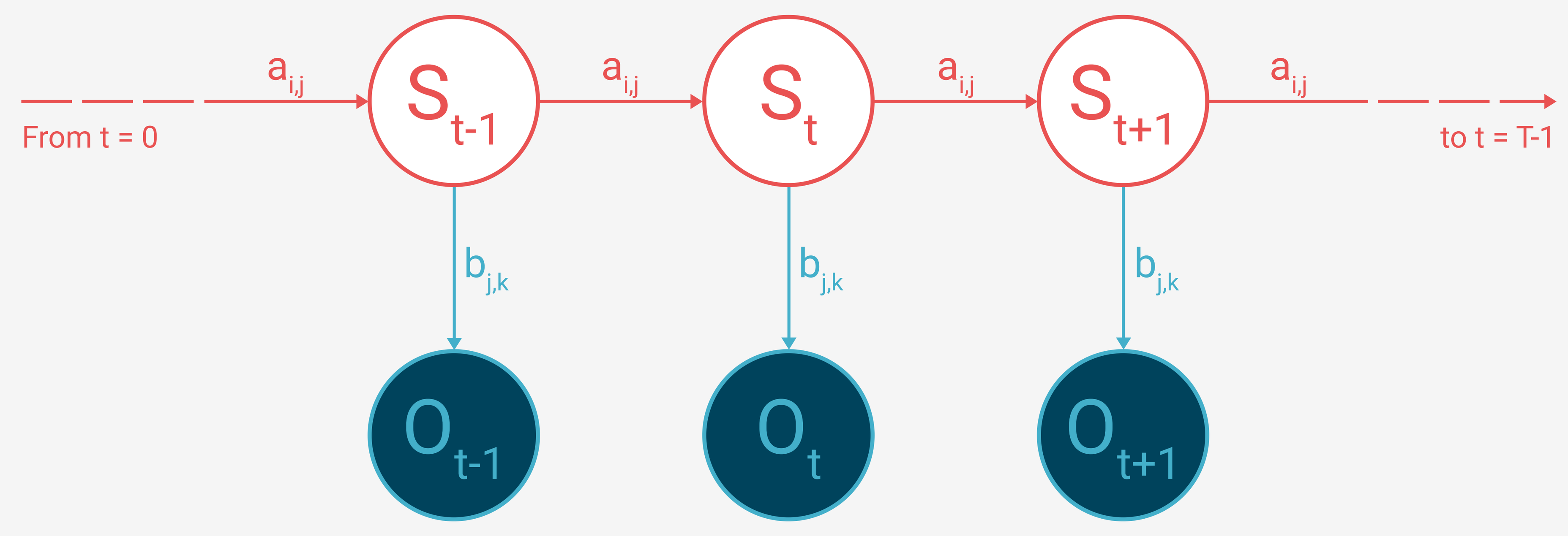
The Basic Theory

General illustration of state-/ observation probabilities



Intuitive meaning

If we are in state S_{t-1}, we use the *transition matrix* $a_{i,j}$ to find out what's the most probable state in the next time step *t*. If we know what state we are in, we can use the *observation matrix* $b_{j,k}$ to find out what's the most probable observation made in that same time step *t-1*. Another way to formulate the *transiton matrix* is:

$$a_{i,j} = A[\,i\,][\,j\,] = P(S_t = S_j \mid S_{t-1} = S_i)$$

This reads as *"the value positioned at the i^{th} row and at the j^{th} column in the transition matrix A is the probability that we are in state number j **given** that we were in state number i in the previous time step"*. We can formulate the *emission matrix* in the same way:

$$b_{j,k} = B[\,j\,][\,k\,] = P(O_t = O_k \mid S_t = S_j)$$

This reads as *"the value positioned at the j^{th} row and at the k^{th} column in the observation matrix B is the probability that we observe observation number k **given** that we are in state number j in the current time step"*.

Caveat When the red S_t is used, we imply the active state at time step t. Conversely, when the black S_n is used, we imply a specific state without taking time steps into account. E.g. if the set of states of possible states $\{S_0, S_1, S_2\}$ then S_5 could be either S_0, S_1 or S_2.

Definitions

- **Background info** The abbreviation **HMM** stands for **H**idden **M**arkov **M**odel. The word hidden is used because the states are hidden. We make observations that will help us calculate the probability of being in a certain state at a certain time step. There are other types of Markov Models as well, but these are not covered on this page.
- **Cardinality** When using the word cardinality we refer to the number of elements in a certain set. For example the cardinality of the set $\mathbf{S}_1 = \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ is three because the set contains three unique elements. To be precise, the set $\mathbf{S}_2 = \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{Y}, \mathbf{Z}, \mathbf{X}\}$ also have the cardinality three, because the set also contains three unique elements. In a mathematical context we write $|\,\mathbf{S}_1\,| = 3$ to denote that the cardinality of set $\mathbf{S}_1$ is three.
- **Time steps** When we work with HMM, time is discrete. That is, we work with discrete time steps where each timestep is an integer. The first time step is either declared as t_0 or t_1, depending on which source your reading. On this page you have the opportunity to decide if the content should be 0-indexed or 1-indexed (the button in the top-right corner). The last time step is declared as t_{T-1} if 0-indexing is used and t_T if using 1-indexing.
- **States** As mentioned above, states in HMM are hidden. That is, we can't for sure know what state is present in the current time step. Depending on which source you're reading, states are called different things. Some of them, if not all, are: *Hidden States*, *States* and *Emitters*. On this page, we will use the notation *States*. There are many different ways to mathematically notate a certain state. On this site, we declare a *state* at a certain time step *t* to be denoted S_t.
- **Observations** As with states, observations have different names depending on which source you're getting your information from. Some of them, if not all, are: *Outputs*, *Emissions*, *Observations* and *Visible States*. On this site, we use the word *observation*, and we denote an observation at time step *t* as O_t.
- **Set of time steps** The set of all time steps are defined as $\{t_0, t_1, \dots, t_{T-1}\}$. Meaning that the cardinality of the set of time steps $|\,\{t_0, t_1, \dots, t_{T-1}\}\,| = \mathbf{T}$.
- **Set of states** The set of all states are defined as $\mathbf{S} = \{S_0, S_1, \dots, S_{N-1}\}$. Meaning that the cardinality of the set of states $|\,\{S_0, S_1, \dots, S_{N-1}\}\,| = \mathbf{N}$.
- **Set of possible observations** The set of possible observations are defined as $\mathbf{O} = \{O_0, O_1, \dots, O_{K-1}\}$. Meaning that the cardinality of the set of states $|\,\{O_0, O_1, \dots, O_{K-1}\}\,| = \mathbf{K}$.
- **State sequence** The state sequence is defined as $\mathbf{S} = \{S_0, S_1, \dots, S_{T-1}\} = S_{0:T-1}$. In other words, **S** is the set of all the observations being made in one specific sequence. Note that a specific observation S_n can occur several times in one observation sequence.
- **Observation sequence** The observation sequence is defined as $\mathbf{O} = \{O_0, O_1, \dots, O_{T-1}\}$. In other words, **O** is the set of all the observations being made in one specific sequence. Note that a specific observation O_k can occur several times in one observation sequence.
- **Row-stochastic** By using the word row-stochastic we mean that every row in the current matrix sums up to one. Similarly, the elements in a stochastic vector sums up to one. Let's define the following two matrices:

$$x_{i,j} = \begin{bmatrix} 0.30 & 0.35 & 0.53 \\ 0.43 & 0.12 & 0.31 \\ 0.12 & 0.48 & 0.37 \end{bmatrix} \qquad y_{i,j} = \begin{bmatrix} 0.30 & 0.30 & 0.40 \\ 0.25 & 0.25 & 0.50 \\ 0.10 & 0.10 & 0.80 \end{bmatrix}$$

Are these two matrices row-stochastic? Let's find out!

$$x_{0,0} + x_{0,1} + x_{0,2} = 0.30 + 0.35 + 0.53 = 1.18 \qquad y_{0,0} + y_{0,1} + y_{0,2} = 0.30 + 0.30 + 0.40 = 1.0$$
$$x_{1,0} + x_{1,1} + x_{1,2} = 0.43 + 0.12 + 0.31 = 0.86 \qquad y_{1,0} + y_{1,1} + y_{1,2} = 0.25 + 0.25 + 0.50 = 1.0$$
$$x_{2,0} + x_{2,1} + x_{2,2} = 0.12 + 0.48 + 0.37 = 0.97 \qquad y_{2,0} + y_{2,1} + y_{2,2} = 0.10 + 0.10 + 0.80 = 1.0$$

As we can see, the matrix $y_{i,j}$ is row-stochastic while $x_{i,j}$ is not.

HMM Parameters

- **Initial state distribution** The initial state distribution, which is sometimes called initial state probabilities, is a vector containing probabilities of being in a certain state at the initial time step t_0. The initial state distribution is denoted π and the vector is defined as $[\,\pi_0 \;\; \pi_1 \;\; \cdots \;\; \pi_{N-1}\,]$, where **N** is the number of states. The initial state distribution vector is always row-stochastic. Let's make an example:

$$\pi(S_i) = \begin{bmatrix} 0.42 & 0.12 & 0.46 \end{bmatrix}$$

The initial state distribution given above should be interpreted in the following way: the probability of being in state S_0 at time t_0 is 0.42, the probability of being in state S_1 at time t_0 is 0.12 and the probability of being in state S_2 at time t_0 is 0.46. We can also see that the vector is stochastic because 0.42 + 0.12 + 0.46 = 1.0. If we want, we can also write the probabilities with the following notation:

$$\pi(S_0) = 0.42$$
$$\pi(S_1) = 0.12$$
$$\pi(S_2) = 0.46$$

The initial state distribution π is stochastic.

- **Transition matrix** The transition matrix A, or $a_{i,j}$, is used to store probabilities when traveling between different states. E.g. if A[0][2] = 0.37, the probability of traveling from S_0 to S_2 between any given time steps is 0.37. As seen above we can express the transition matrix as:

$$a_{i,j} = A[\,i\,][\,j\,] = P(S_t = S_j \mid S_{t-1} = S_i)$$

This reads as *"the value positioned at the i^{th} row and at the j^{th} column in the transition matrix A is the probability that we are in state number j **given** that we were in state number i in the previous time step"*.

The transition matrix A is always a square matrix.

Let's look at an example:

$$a_{i,j} = \begin{bmatrix} 0.22 & 0.41 & 0.37 \\ 0.43 & 0.12 & 0.45 \\ 0.14 & 0.48 & 0.38 \end{bmatrix}$$

By just looking at the matrix we can see that there are three number of possible states. The following figure is a clarification of the same transition matrix:

State at time step t

State at time step t-1

A	S_0	S_1	S_2
S_0	0.22	0.41	0.37
S_1	0.43	0.12	0.45
S_2	0.14	0.48	0.38

$p(S_t = S_2 \mid S_{t-1} = S_1)$

Förslag: Vid hover kommer ovanstående information upp för varje ruta i matrisen!

The transition matrix A is row-stochastic.

- **Observation matrix** The observation matrix B, or $b_{j,k}$, is used to store probabilities of making specific observations given that the hidden model is in a particular state. E.g. if B[2][0] = 0.05, the probability of observing O_0 given that we are in state S_2 is 0.05. As seen above we can express the observation matrix as:

$$b_{j,k} = B[\,j\,][\,k\,] = P(O_t = O_k \mid S_t = S_j)$$

This reads as *"the value positioned at the j^{th} row and at the k^{th} column in the observation matrix B is the probability that we observe observation number k **given** that we are in state number j in the current time step"*.

The observation matrix B is **not** necessarily a square matrix.

Let's look at an example:

$$b_{j,k} = \begin{bmatrix} 0.13 & 0.24 & 0.16 & 0.18 & 0.29 \\ 0.15 & 0.55 & 0.04 & 0.11 & 0.15 \\ 0.05 & 0.14 & 0.43 & 0.22 & 0.16 \end{bmatrix}$$

By looking at the observation matrix we can see that the model consists of three different states and that we can make five types of observations. The following figure is a clarification of the same observation matrix:

B	O_0	O_1	O_2	O_3	O_4
S_0	0.13	0.24	0.16	0.18	0.29
S_1	0.15	0.55	0.04	0.11	0.15
S_2	0.05	0.14	0.43	0.22	0.38

$p(O_t = O_4 \mid S_t = S_1)$

The observation matrix B is row-stochastic.

- **Lambda** The combination of a specific initial state distribution vector, a specific transition matrix and a specific observation matrix is what together makes the Hidden Markov Model. Because we in different scenarios would like to distinguish different models, we can combine A, B and π in a variabel λ. In this way, we can discuss models without confusing ourselves. We define lambda as the set $\lambda = \{A, B, \pi\}$.

The combination of the initial state distribution vector, the transition matrix and the observation matrix is what together makes the HMM.