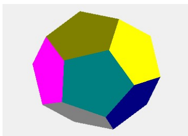
Markov chains
And their limit distributions

This is an article from my homepage : www.olewitthansen.dk



Ole Witt-Hansen 2021

Contents

Introduction

The theory of Markov chains is founded on the elementary part of probability. If you are not familiar with probability theory, you may for example be referred to the first four chapters of: **http://www.olewitthansen.dk/Mathematics/Probability_theory.pdf**.
An introduction to linear algebra is found in:
http://www.olewitthansen.dk/Mathematics/Eigenvalue_problems_in_linear_algebra.pdf

1. Definition and properties for a Markov chain

We consider a system, which may be in one of the states (1), (2), (3)…(n).
The state space S is the set of states. $S = \{s_1, s_2, ..., s_n\}$
X is a stochastic variable, that is, a function defined on U, such that the probability that the system is in the state j is $P(X = j) = p_j$.
The system changes its state stochastically with a constant finite time interval. The probabilities that the system is in a certain state is given by the a state probability vector $p = (p_1, p_2, ..., p_n)$, where $P(X = j) = p_j$. The initial probability state is denoted:

$$p^{(0)} = (p_1^{(0)}, p_2^{(0)}, ..., p_n^{(0)})$$

And the subsequent state probabilities are written as:

$$p^{(k)} = (p_1^{(k)}, p_2^{(k)}, ..., p_n^{(k)}).$$

It is important to distinguish between the state that a system occupy s_j, and the probabilities $P(X = s_i) = p_j$ that the system is in that state.
The system is at every instant is given by a state vector ($p_1, p_2, p_3, …, p_n$), where p_j is the probability that the system is in the state j.
Since the system must be in one of the states 1..n we must have :

$$p_1 + p_2 + p_3 + ... + p_n = 1$$

The conditional probability that the state i makes a transition into the state j is written p_{ij}, where

$$p_{ij} = P(X = i \mid X = j) \quad \text{or just} \quad P(X = i \rightarrow X = j) = p_{ij}$$

A Markov chain is a system, which has no history. The probability of a transition of the state (i) to the state (j) depends only on the state i and the state (j), but not on the states presiding the state (i). Written more formally:

$$P(X_{k+1} = j_{k+1} | P(X_k = j_k, X_{k-1} = j_{k-1}, X_{k-2} = j_{k-2}, ..., X_0 = j_0) = P(X_{k+1} = j_{k+1} | X_k = j_k)$$

So the conditional probability that the system reaches the state (j_{k+1}) depends only on the immediate presiding state, and not how the system reached that state.

An ordinary walk is not a Markov chain, since your position when you take the next step (in most circumstances) depends on some, if not all, the preceding steps you have taken.
A counterexample is the so called random walk, on which we shall deal with later.

The probabilities $P(X = i \rightarrow X = j) = p_{ij}$ form a transition matrix. If the system has *n* possible states, then a transition must result in one of these states, therefore the sum of transition possibilities in a row must add up to one.

$$\sum_{j=1}^{j=n} p_{ij} = 1 \quad \text{for} \;\; i = 1..n$$

The transition matrix, may be written

$$\begin{pmatrix} p_{11} & p_{12} & .. & p_{1,n-1} & p_{1,n} \\ p_{21} & p_{22} & .. & p_{2,n-1} & p_{2,n} \\ .. & \cdots & \cdots & \cdots & \cdots \\ p_{n-1,1} & p_{n-1,2} & \cdots & p_{n-1,n-1} & p_{n-1,n} \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n-1} & p_{n,n} \end{pmatrix}$$

Traditionally the probability distribution of a state *s* is written as a row vector. $(p_1, p_2,.., p_n)$ or $(q_1, q_2,.., q_n)$ or $(r_1, r_2,.., r_n)$
The transition to a new state is done by matrix multiplication with the transition matrix.
If $(r_1, r_2,.., r_n)$ is the probability distribution for the transformed state then

$$(q_1, q_2,..., q_n)\begin{pmatrix} p_{11} & p_{12} & .. & p_{1,n-1} & p_{1,n} \\ p_{21} & p_{22} & .. & p_{2,n-1} & p_{2,n} \\ .. & \cdots & \cdots & \cdots & \cdots \\ p_{n-1,1} & p_{n-1,2} & \cdots & p_{n-1,n-1} & p_{n-1,n} \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n-1} & p_{n,n} \end{pmatrix} = (r_1, r_2,...r_n)$$

Where we have: $r_j = q_1 p_{1j} + q_2 p_{2j} + ... + q_n p_{nj}$.
The probability $q_i p_{ij}$ is then the probability that the system is in state (*i*) times the probability of a transition to a state (*j*).
Written in a more compact form: $r_j = \sum_{i=1}^{n} q_i p_{ij} \quad j = 1..n$
Before we go on with the theory we shall present some examples.

2. Conditions for a Markov-chain to be regular

In a regular Markov-chain, the sum of the probabilities of the states in the state vector should remain equal to one after each iteration.
A Markov chain is called regular if the sum of probabilities of each subsequent state vector is equal to one.

The random walk has two states (forward, backward). For the random walk, it is easy to see that it is regular. The transition matrix becomes:

$$\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$$

And a probability state: (q_1, q_2) is transformed into:

$$(q_1, q_2)\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix} = (q_1 p + q_2(1-p), q_1(1-p) + q_2 p)$$

Taking the sum of the state probabilities:

$$(q_1 p + q_2(1-p) + q_1(1-p) + q_2 p) = q_1(p + (1-p)) + q_2((1-p)p + p) = q_1 + q_2 = 1 .$$

Doing the same calculation for a *n* x *n* matrix is a bit more circumstantial But we shall show the same result for a 3x3 Markov matrix.

$$(q_1, q_2, q_3)\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} = (q_1 p_{11} + q_2 p_{21} + q_3 p_{31} \text{ , } q_1 p_{12} + q_2 p_{22} + q_3 p_{32}\text{, } \ q_1 p_{13} + q_2 p_{23} + q_3 p_{33})$$

And taking the sum:

$$(q_1 p_{11} + q_2 p_{21} + q_3 p_{31} + q_1 p_{12} + q_2 p_{22} + q_3 p_{32} + \ q_1 p_{13} + q_2 p_{23} + q_3 p_{33})$$

Taking q_1, $q_{2,,} q_3$ as a common factor of the three terms, we find;

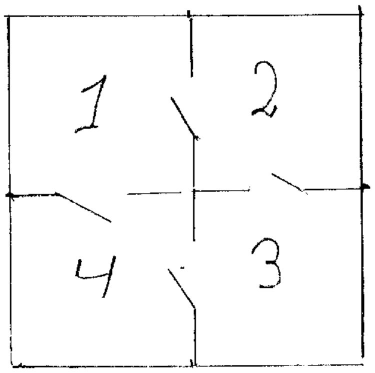
$$q_1(p_{11} + p_{12} + p_{13}) + q_2(p_{21} + p_{22} + p_{23}) + q_3(p_{31} + p_{32} + p_{33}) = q_1 + q_2 + q_3$$

We can then see that the sum of probabilities in the transformed state of a Markov matrice is equal to the sum of probabilities in the original state, if and only if the sum of probabilities in all three rows of the transition matrix equals one.
We have then shown that, when the sum of probabilities in each row equals one, then it is secured that the sum of the state vector probabilities is conserved.

It is a simple, (but rarely useful fact) that all Markov matrices have the eigenvector (1, 1, 1) from the right with the eigenvalue one.
We shall restrict ourselves by showing that it is true for 3 x3 matrices.

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} p_{11} + p_{12} + p_{13} \\ p_{21} + p_{22} + p_{33} \\ p_{31} + p_{32} + p_{33} \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

3. Examples: The cat and the mouse in the maze



The figure to the left shows a “maze” having 4 compartments. In each compartments, there are two exits to the neighbouring compartments.
Let us assume that a cat initially is in (1). We also assume that the probability of an exit to any the two neighbour compartments is $\frac{1}{2}$. We also assume that in every step the compartment is shifted
Thus, we have for example $P(1 \rightarrow 2) = \frac{1}{2}$ *and* $P(1 \rightarrow 3) = 0$.
The 4x4 transition matrix *T* therefore becomes.

$$T = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

Let us assume that the cat initially is in compartment (1).
The initial state probabilities are then: $p^{(0)} = (1,0,0,0)$
To find the state probabilities after one step, we multiply from the left with the transition matrix *T*.

1. step. $$(1,0,0,0)\begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix} = (0,\tfrac{1}{2},0,\tfrac{1}{2})$$

2. step. $$(0,\tfrac{1}{2},0,\tfrac{1}{2}) * T = (\tfrac{1}{4}+\tfrac{1}{4},0,\tfrac{1}{4}+\tfrac{1}{4},0) = (\tfrac{1}{2},0,\tfrac{1}{2},0)$$

3.step. $$(\tfrac{1}{2},0,\tfrac{1}{2},0) \bullet T = (0,\tfrac{1}{2},0,\tfrac{1}{2}).$$

We notice that the transition probabilities between the states are cyclic with period 2.
Now we shall do the same, but from the point of view of the mouse, which initially is in the compartment (3). The Initial state for the mouse is (0, 0, 1, 0).

1. step. $(0,0,1,0) \bullet T = (0,\frac{1}{2},0,\frac{1}{2})$
2. step: $((0,\frac{1}{2},0,\frac{1}{2}) \bullet T = (\frac{1}{2},0,\frac{1}{2},0)$
3. step: $(\frac{1}{2},0,\frac{1}{2},0) \bullet T = (0,\frac{1}{2},0,\frac{1}{2})$

Not surprisingly the sequence of states for the mouse is also cyclic with period 2. We further notice that in the states (1) and (3), the cat and the mouse have the same probability distribution. Therefore the probability that the cat gets the mouse is equal to: $(0,\frac{1}{2},0,\frac{1}{2}) \cdot (0,\frac{1}{2},0,\frac{1}{2}) = \frac{1}{2}$.

4. Further properties of Markov chains

We shall denote states probabilities with the letters *p, q, r*, having one or two indices.
The initial probability distribution of states is formally written: $p^{(0)} = (p_1{}^{(0)}, p_2{}^{(0)}, p_3{}^{(0)}, ..., p_n{}^{(0)})$,
and after one transition we have the probability distribution $p^{(1)} = (p_1{}^{(1)}, p_2{}^{(1)}, p_3{}^{(1)}, ..., p_n{}^{(1)})$.
The two distributions are related by the transition matrix:

$$(p_1{}^{(0)}, p_2{}^{(0)}, p_3{}^{(0)}, ..., p_n{}^{(0)}) \begin{pmatrix} p_{11} & p_{12} & \cdot\cdot & p_{1,n-1} & p_{1,n} \\ p_{21} & p_{22} & \cdot\cdot & p_{2,n-1} & p_{2,n} \\ \cdot\cdot & \cdots & \cdots & \cdots & \cdots \\ p_{n-1,1} & p_{n-1,2} & \cdots & p_{n-1,n-1} & p_{n-1,n} \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n-1} & p_{n,n} \end{pmatrix} = (p_1{}^{(1)}, p_2{}^{(1)}, p_3{}^{(1)}, ..., p_n{}^{(1)})$$

Or written in matrix formalism: $\underline{\underline{p}}^{(0)} \underline{\underline{T}} = \underline{\underline{p}}^{(1)}$

One step further we have: $\underline{\underline{p}}^{(1)} \underline{\underline{T}} = \underline{\underline{p}}^{(2)} = \underline{\underline{p}}^{(0)} \underline{\underline{T}}\,\underline{\underline{T}} = \underline{\underline{p}}^{(2)} = \underline{\underline{p}}^{(0)} \underline{\underline{T}}^2 = \underline{\underline{p}}^{(2)}$

It should then be obvious that:

$$\underline{\underline{p}}^{(0)} \underline{\underline{T}}^n = \underline{\underline{p}}^{(n)},$$

This may also be verified by induction. Assuming that the equation is valid for *n*, we shall prove that it is also valid for *n* + 1.

$$\underline{\underline{p}}^{(0)} \underline{\underline{T}}^n = \underline{\underline{p}}^{(n)} \;\wedge\; \underline{\underline{p}}^{(n)} \underline{\underline{T}} = \underline{\underline{p}}^{(n+1)} \Rightarrow \quad \underline{\underline{p}}^{(0)} \underline{\underline{T}}^n \underline{\underline{T}} = \underline{\underline{p}}^{(n+1)} \quad \Rightarrow \quad \underline{\underline{p}}^{(0)} \underline{\underline{T}}^{n+1} = \underline{\underline{p}}^{(n+1)}$$

3.1 Periodic, reducible and irreducible Markov chains.

A Markov chain is denoted <u>irreducible</u>, if any state can be reached from any other state. Otherwise it is called irreducible. It is important to be certain that the Markov chain in question is irreducible. If it is not, it may be decomposed into two or more irreducible Markov chains.
The irreducibility can often be decided from the transition matrix or rather from the physical phenomenon the Markov represents. If we look at the transition matrix

$$\begin{pmatrix} 0.3 & 0.7 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}$$

And a state vector (q_1, q_2, q_3, q_4) and we evaluate:

$$(q_1, q_2, q_3, q_4)\begin{pmatrix} 0.3 & 0.7 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix} = (0.3q_1 + 0.2q_2\ , 0.7q_1 + 0.8q_2\ , 0.4q_3 + 0.5q_4\ , 0.6q_3 + 0.5q_4)$$

We can see, that there are only transitions to state (1) and (2) from these states, and likewise for the states (3) and (4). The Markov chain is clearly reducible.

Above we saw in the Cat and mouse an example of a periodic Markov chain, since we recover the same state in only two iterations.
In general a Markov chain is said to **periodic**, if it reaches the same state (the initial state) in a finite number of iterations. If q is a state vector and T is the transition matrix, then the periodicity can be expressed as: There exists a positive number M, such that:

$$\underline{\underline{q}}\,\underline{\underline{T}}^M = \underline{\underline{q}}$$

This may also be formulated as $\underline{\underline{T}}^M$ has the eigenvalue 1 from the left.

A Markov chain is said to be **non-periodic**, if it is not periodic, but nevertheless that any state can be reached from any other state in a finite number of iterations.
Otherwise the Markov chain is said to be **periodic.**

4. Limit distributions and stationary distributions for non-periodic and irreducible Markov chains.

A stationary distribution is a distribution which is unaltered after being brought one step forward by the transition matrix. More formally, if q is the state distribution and T is the transition matrix, we must have:

$$\underline{\underline{q}}\,\underline{\underline{T}} = \underline{\underline{q}}$$

This also means that q is an eigenvector to T with eigenvalue 1.
However, as we have already proven, a Markov transition matrix has the eigenvalue 1 from right, due to the Markov property that the sums of the elements in each row equals 1.

The eigenvalues λ of a matrix $\underline{\underline{A}}$, are among the solutions to the equation: $det(\underline{\underline{A}} - \lambda \underline{\underline{E}})=0$,
However, we know from linear algebra that the determinant of a matrix is the same as for the transposed matrix, that is, if the equation $det(\underline{\underline{A}} - \lambda \underline{\underline{E}}) = 0$, has a solution $\lambda = 1$, for an eigenvector to the right, then the determinant of the *transposed* matrix must also have the solution $\lambda = 1$ corresponding to an eigenvector from the left.

This means, however, that any Markov transition matrix has an eigenvector (from the left) with eigenvalue 1.

We shall illustrate this with a simple example using a general 2x2 Markov transition matrix:

$$\underline{\underline{T}} = \begin{pmatrix} p & 1-p \\ q & 1-q \end{pmatrix}$$

$$det(\underline{T} - \lambda\underline{\underline{E}}) = \begin{pmatrix} p-\lambda & 1-p \\ q & 1-q-\lambda \end{pmatrix} = (p-\lambda)(1-q-\lambda) - q(1-p) = \lambda^2 - (p+1-q)\lambda + p - q$$

It is easy to verify that this equation has the root 1. This corresponds to an eigenvector (α,α)
In the Markov transition matrix, we multiply however with a row vector from the left, that is, the transposed matrix. It still has the eigenvalue 1, but the eigenvector may not be the same.
Let us pursue this using the simple example above. So we search for an eigenvector from the left, having eigenvalue 1.

$$(x,y)\begin{pmatrix} p & 1-p \\ q & 1-q \end{pmatrix} = (x,y) \Leftrightarrow (xp+yq, x(1-p)+y(1-q)) = (x,y)$$

This results in two equations: $\begin{matrix} x(p-1)+yq=0 \\ x(1-p)-yq=0 \end{matrix}$ Each of these two equations results in

$\frac{y}{x} = \frac{1-p}{q}$, so we have for example with (p,q) = (0.8, 0.4) we have an eigenvector (0.2, 0.4)

It is obvious, that the state vector itself, $s = (s_1, s_2, ..., s_n)$, can have no a limiting distribution, since it is a stochastic variable on *S*. For any non-periodic Markov chain, the *probability* of returning to a given state s_i is, however, non zero, since any state can be reached from any state in a finite number of steps.
Furthermore the probability distribution will approach the probability distribution of a stationary state. This does not mean that the Markov chain will converge to the stationary state itself, but only it will approach (in a broad sense) the probability distribution of a stationary state.
This implies that there exists a stationary state for all irreducible, non-periodic Markov chains.
The formal proof of these theorems are non trivial and requires elements of number theory, so we shall start out with an illustrating (mathematical) example.

4.1 Example: Choice of a brand of goods

Let us assume that customers buy the same commodity, which is delivered in three different brands *A*, *B* and *C*. We assume the following transition table:
If a consumer one week have used the brand *A, B* or *C* then the probabilities of changing to another brand are the following:

$$P(A \to A) = 0.85, P(A \to B) = 0.10 \;\; and \;\; P(A \to C) = 0.05 .$$
$$P(B \to A) = 0.15, P(B \to B) = 0.80 \;\; and \;\; P(C \to C) = 0.05$$
$$P(C \to A) = 0.20, P(C \to B) = 0.10 \;\; and \;\; P(C \to C) = 0.65$$

Let us assume that the customers buy the commodity with an interval of one week, then we may describe the “system” as a Markov chain, having the transition matrix.

$$\begin{matrix} A & B & C \end{matrix}$$
$$\begin{pmatrix} 0.85 & 0.10 & 0.05 \\ 0.15 & 0.80 & 0.05 \\ 0.20 & 0.15 & 0.65 \end{pmatrix}$$

We can see that the customers (1) (2) and (3) favour the brand *A, B, C*, respectively, and we are interested in whether, this line of habit change from the Markov probability point of view.
For example will the product *C* eventually be deselected. Will the state probabilities eventually become a stationary state.
The latter is, by the way, one of the main results of a certain class the theory of Markov chains.
Below is shown a computer simulation of the development of the state probabilities..
We assume that the demand for the three brands initially have the following probabilities:

A	*B*	*C*
0.200	0.300	0.500

```
Probability table

 0.850   0.100   0.050
 0.150   0.800   0.050
 0.150   0.800   0.050

initial state:

 0.200   0.300   0.500
Sum of probabilities in initial state  1.000

Probabilities for new state:   1
 0.290   0.660   0.050

Probabilities for new state:   2
 0.353   0.597   0.050

Probabilities for new state:   3
 0.397   0.553   0.050

Probabilities for new state:   4
 0.428   0.522   0.050

Probabilities for new state:   5
 0.450   0.500   0.050

Probabilities for new state:   6
 0.465   0.485   0.050

We skip probabilities for the states 7-14
---------------------------------------
Probabilities for new state:  15
 0.499   0.451   0.050

Probabilities for new state:  16
 0.499   0.451   0.050
```

Counting from the tenth iteration it appears more than plausible that the state:
(0.499, 0,451, 0.050) is in fact an stationary state.

If there are zeroes in the transition matrix and in the initial state the approaching to a stationary state, becomes slower, as is illustrated in the example below

```
The transition matrix

(  0.10  0.00  0.90 )
(  0.90  0.00  0.10 )
(  0.00  0.90  0.10 )

initial vector

(   1.00  0.00  0.00 )
```

```
state no   1
(   0.10  0.00  0.90 )

state no   2
(   0.01  0.81  0.18 )

state no   3
(   0.73  0.16  0.11 )

We skip states 4 to 15

state no  16
(   0.33  0.28  0.40 )

state no  17
(   0.28  0.36  0.36 )

state no  18
(   0.35  0.32  0.33 )

state no  19
(   0.33  0.29  0.38 )

state no  20
(   0.30  0.34  0.36 )
```

Although we are near, it is still not a totally convincing stationary state, but it is reached after about 30 iterations

4.2 Consequences of non-periodicity.

One reason for the usefulness of non-periodicity is the following theorem:

Theorem 4.1 Suppose we have a non-periodic Markov chain $(X_0, X_1, ...)$ with state space $S = \{s_1, s_2, ..., s_n)$ and transition matrix *P*. Then there exists a $N \in Z_+$, such that

$$(P^n)_{i,j} > 0$$

For all $i \in \{1,...,k\}$ and for all $n \geq N$.
To prove **theorem 4.1** we need the following lemma from number theory, which we shall not prove.
Lemma 4.1 Let $A = \{a_1, a_2, ...\}$ be a set of positive integers, subject to the following two conditions:
(i): The greatest common divisor: $\gcd\{a_1, a_2, ...\} = 1$
(ii): *A* is closed under addition: if $a \in A \ \wedge \ a' \in A \ \Rightarrow \ a + a' \in A$

Then there exists a positive integer *N*, such that for all *n* >*N*: $n \in A$.

The theorem is a consequence of a theorem of partition of positive integers, and it is hard to imagine a counter example.

Proof of Theorem 4.1 For $s_i \in S$, let $A_i = \{n \geq 1 \mid (P^n)_{i,i} > 0\}$ In other words A_i is the set of possible return times to the state s_i starting at s_i .We assumed that the Markov chain was irreducible and non-periodic, so that gcd(A_i) = 1. Furthermore A_i is closed under addition for the following reason: If *a, a'* $\in A_i$ then $P(X_a = s_i \mid X_0 = s_i) > 0$ and $P(X_{a+a'} = s_i \mid X_a = s_i) > 0$.

This implies that:

$$P(X_{a+a'} = s_i \mid X_0 = s_i) \geq P(X_a = s_i \wedge X_{a+a'} = s_i \mid X_0 = s_i) =$$
$$P(X_a = s_i \mid X_0 = s_i) P(X_{a+a'} = s_i \mid X_0 = s_i) > 0$$

So that $a + a' \in A_i$.
We therefore have that *A* satisfies (i) and (ii) from lemma 4.1 which implies for the state s_i the existence of an integer N_i such that $(P^n)_{i,I} > 0$ for all $n \geq N_i$. Theorem 4.1 then follows, if we choose $N = \max\{N_1, N_2, ..., N_k\}$
By combining non-periodicity and irreducibility, we get the following important result, which will be used later to prove the so called *Markov chain convergence theorem.*

Corollary 4.1 Let $(X_0, X_1, ...)$ be a non-periodicity and irreducible Markov chain with state space $S = \{s_1, s_2, ..., s_n)$ and transition matrix *P*. Then there exists an integer *N*, such that
$(P^n)_{i,j} > 0 \text{ for all } i, j \in \{1,..,k\} \text{ and all } n \geq M$
Proof. By the assumed non-periodicity and theorem and theorem 4.1 , there exists an integer *N* such that $(P^n)_{i,i} > 0 \text{ for all } i \in \{1,..,k\} \text{ and all } n \geq N$. Fix two states $s_i, s_j \in S$. By the assumed irreducibility, we can find some $n_{i,j}$, such that $(P^{n_{i,j}})_{i,i} > 0$. Let $M_{i,j} = N + n_{i,j}$ then for $m \geq M_{i,j}$, we have:

$$P(X_m = s_j \mid X_0 = s_i) \geq P(X_{m-n_{i,j}} = s_i \wedge X_m = s_j \mid X_0 = s_i)$$
$$= P(X_{m-n_{i,j}} = s_i \mid X_0 = s_i) P(X_m = s_j \mid X_{m-n_{i,j}} = s_i)$$
$$> 0$$

The first factor on the right side is positive because $m - n_{i,j} \geq N$ and the second factor is positive by the choice of $n_{i,j}$. Hence we have shown that $(P^m)_{i,j} > 0 \text{ for all } m > M_{i,j}$. The corollary now follows from $M = \max\{M_{1,1}, M_{1,2}, .., M_{1,k}, ..., M_{2k}, .., M_{k,k}\}$.

5. Stationary distributions

In this chapter, we shall deal with one of the central issues in Markov theory, which is the asymptotic behaviour for long term Markov chains. Do non-periodic Markov chains have a limiting *probability* distribution? The answer is confirmative, as we have argued for ealier.
This does, however, not mean that there exist a limiting state vector, since if $(X_0, X_1, ...,)$ is a any , nontrivial Markov chain, then it will keep fluctuating infinitely many times, as $n \to \infty$ and therefore a limiting state (in a mathematical sense) is excluded.
But we may hope that the *probability distribution* of X_n approaches a limit. This is equivalent to saying that the transition matrix *P* has an eigenvector with eigenvalue one from the left.
A matrix and its transposed matrix, have the same determinant, and therefore share de same eigenvalues (but not the same eigenvectors) The eigenvalues of a regular matrix *A* are determined by the equation.

$$det(\underline{A} - \lambda \underline{E})$$

The determinant is an *n*-degree polynomial in λ . According to the fundamental theorem of Algebra an *n*-degree polynomial has *n* (complex) roots. If *n* is odd, it has at least one real root. For a stationary state, however the eigenvalue is one.

We have seen above that any Markov transition matrix has the eigenvalue 1, with eigenvector (1,1,1,1,..,1). But then any Markov transition matrix must also have the same eigenvalue one from the left, but with a different eigenvector, as demonstrated in the example above.

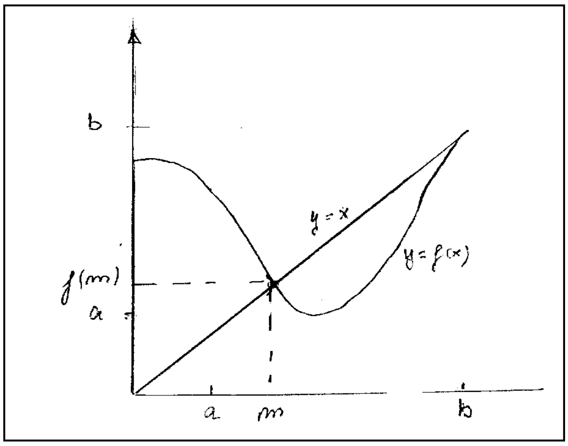
So from an algebraic point of view, every Markov transition matrix has a stationary state, that is, an eigenvector with eigenvalue 1.
That any state approaches a stationary state in the long run cannot be proven algebraically.
The main theorem of Markov chains is the following:

Theorem 1.4 (existence of stationary distributions)

For any irreducible and non-periodic Markov chain, there exists at least one stationary distribution, and that this limiting distribution is approached in a finite number of iteration steps.

What we still lack to demonstrate is approach to the stationary distribution.
However, the number-theoretic proof of this theorem is unbearable long and abstract, and only minor illustrative of the algebraic mechanism behind the formation of a stationary state.
Firstly we will remind you of the fixed point theorem that a real function, which maps an interval $[a, b]$ on itself has at least one fixed point m, so that

$$f(m) = m.$$



This theorem may be illustrated geometrically using th graph to the left, where we assume that the continuous function $y = f(x)$ maps the interval $[a, b]$ on itself, such that $f([a, b]) = [a, b]$.
To the left is shown a function $y = f(x)$ which fulfils these conditions together with the line $y = x$, between the two points (a, a) and (b, b). Since a continuous function obtain all values between $f(a)$ and $f(b)$ at least once, then the function must intersects the line $y = x$ at least once.
If $f(a) = a$ then a is a fixed point, else if $f(a) \neq a$, then the function must cross the line $y = x$ either to obtain values that are greater than a or less than a. The intersections with the line $y = x$ are fixed points, since $f(x) = x$.

The theorem that all functions which naps an interval to itself may also be proven indirectly from the theorem:
Any continuous function $y = f(x)$, defined in an interval $[a, b]$, where $f(a)f(b) < 0$ intersects the first axis at least once, that is where $f(x_0) = 0$.

To apply this to determine fixed points, we shall assume that $f(a) < f(b)$. The case $f(a) > f(b)$ can be treated, if we consider the function $y = -f(x)$

Put $g(x) = f(x) - k \iff f(x) = g(x) + k$, where k is a real number.
To fulfill the condition that $g(x)$ intersects the first axis, we must claim that
$g(a) < 0 \;\; og \;\; g(b) > 0$, and thus $f(a) - k < 0 \;\; and \;\; f(b) - k > 0$
Solving these inequalities with respect to k, we have:

$$k > f(a) \; and \; k < f(b) \quad \Leftrightarrow \quad f(a) < k < f(b).$$

According to the zero point theorem there exists at least one $m \in [a,b]$, such that, $g(m) = 0$.
If we put $f_1(x) = f(x) + (m-k) = g(x) + k + (m-k)$, we obtain that $f_1(x)$ has a fixed point

$$f_1(m) = m$$

The function $f_1(x)$ deviates from $f(x)$ only by an additive constant. However we are not able to show that it also applies for $f(x)$, without prior knowledge of *m*.

In the following we shall lean on a theorem which is named by Brouwer
The theorem states that if a function has a fixed point, then under rather general circumstances one may approach the fixed point by recursion, starting with an (almost) arbitrary x_0:

$$x_1 = f(x_0), \quad x_2 = f(x_1), \quad x_3 = f(x_2), \ldots$$

The claim is that this series will approach a fixed point *m* where: $m = f(m)$.

In the following we shall consequently work with a system having 3 states. This is no restriction on the generality, but rather a typographical convenience

A Markov chain consist of a state vector $s = (s_1, s_2, s_3)$ with the probabilities $q = (q_1, q_2, q_3)$.

The probabilities obey: $\sum_{i=1}^{3} q_i = 1$

And an transition matrix

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{3,2} & p_{3,3} \end{pmatrix}$$

which obeys : $\sum_{j=1}^{3} p_{i,j} = 1$ for $i = 1..3$

A transition from a state $q^{(0)} = (q_1^{(0)}, q_2^{(0)}, q_3^{(0)})$ to a state $q^{(1)} = (q_1^{(1)}, q_2^{(1)}, q_3^{(1)})$ goes as follows (common matrix multiplication)

$$\underline{\underline{q}}^{(k+1)} = \underline{\underline{q}}^{(k)} \underline{\underline{P}} \qquad \text{or written out} \qquad q_i^{(k+1)} = \sum_{j=1}^{3} q_j^{(k)} p_{i,j} \qquad i = 1..3.$$

Since the system must be in one of the states (s_1, s_2, s_3) then the sum of the probabilities must be equal to one at every stage

$$q_1^{(k)} + q_2^{(k)} + q_3^{(k)} = 1$$

But that means that $q_i^{(k+1)}$ is the *mean value* (denoted $<q_i^{(k+1)}>$) of the *i*'th column of $\underline{\underline{P}}$, with respect to the distribution $(q_1^{(k)}, q_2^{(k)}, q_3^{(k)})$ since $q_i^{(k+1)} = q_1^{(k)} p_{1,i} + q_2^{(k)} p_{2,i} + q_3^{(k)} p_{3,i} = < \underline{\underline{p_i}} >_k$

If we think of the transition $\underline{\underline{\mathbf{P}}}$ matrix as a linear function L of 3 variables (q_1, q_2, q_3) we thus have

$$(q_1^{(k+1)}, q_2^{(k+1)}, q_3^{(k+1)}) = L(q_1^{(k)}, q_2^{(k)}, q_3^{(k)}) \text{ where } (q_1^{(k+1)}, q_2^{(k+1)}, q_3^{(k+1)}) = (< \underline{\underline{p_1}} >_k, < \underline{\underline{p_2}} >_k, < \underline{\underline{p_3}} >_k)$$

L is thus a recursive function.

According to Brouwers theorem: That a continuous function $f(\mathbf{x})$, which maps an n-dimensional interval on itself, has at least one fixed point.
The fixed point can be reached by recursion, if the function is constrained by a Lifschitz condition, which may be formulated: There exists a constant c, such that

$$\| f(y) - f(x) \| < c \| y - x \| \ ,$$

for all x and y, and || || is an appropriate length measure in the space in question:

Since all Markov state vectors have a length less than or equal to one. They obay a Lifschitz condition. So we have argued (if not proven)

> *Any Markov chain with the probably transition matrix P has a stationary state* (*that is, a fixed point for the transition matrix), and beginning with an arbitrary state the state will approach to the stationary state by recursion.*

Below is shown some computer evaluation of subsequent states of a Markov chain. It is seen that The first example is a matrix where all elements are nonzero, and likewise for the initial vector. The stationary state is reached within just three iterations.

<u>The transition matrix</u>

(0.10 0.40 0.50)
(0.40 0.30 0.30)
(0.20 0.30 0.50)

<u>The initial vector</u>

(0.20 0.30 0.50)

The subsequent probably states are

(0.24 0.34 0.47)
(0.25 0.34 0.46)
(0.26 0.34 0.46)

However, if you start out with a state vector (1,0,0) and a transition matrix with several zeroes, the convergence is substantially slower. In the example below, the stationary state is barely reached in 15 iterations. And it is still not the case for 30 iterations.

```
The transition matrix

( 0.10  0.00  0.90 )
( 0.90  0.00  0.10 )
( 0.00  0.90  0.10 )

initial state
( 1.00  0.00  0.00 )
Subsequent states
```

```
( 0.10  0.00  0.90 )
( 0.01  0.81  0.18 )
( 0.73  0.16  0.11 )
( 0.22  0.10  0.68 )
( 0.11  0.62  0.28 )
( 0.56  0.25  0.19 )
( 0.28  0.17  0.55 )
( 0.18  0.50  0.32 )
( 0.47  0.29  0.24 )
( 0.31  0.22  0.47 )
( 0.23  0.42  0.35 )
( 0.41  0.31  0.28 )
( 0.32  0.25  0.42 )
( 0.26  0.38  0.36 )
state no  15
( 0.37  0.32  0.31 )
( 0.33  0.28  0.40 )
( 0.28  0.36  0.36 )
( 0.35  0.32  0.33 )
( 0.33  0.29  0.38 )
state no  20
( 0.30  0.34  0.36 )
```

The equation for reaching the stationary state $q^{(n)}$, where the initial state is $q^{(0)}$, and where P is the transition matrix is:

$$\underline{q}^{(n)}\underline{\underline{P}} = \underline{q}^{(n)}$$

This may also be written:

$$\underline{q}^{(0)}\underline{\underline{P}}^{(n)} = \underline{q}^{(n)}, \text{ where } \underline{\underline{P}}^{(n)} = \underline{\underline{P}}\,\underline{\underline{P}}\,\underline{\underline{P}}\ldots\,\underline{\underline{P}} \text{ (} n \text{ times)}$$

Below we have shown the calculation of $\underline{\underline{P}}^{(\mathrm{n})}$ with itself for two probability distributions:
The first matrix has no zeroes, whether the last matrix has one zero in each line.

```
The transition matrix

  0.10  0.40  0.50
  0.40  0.30  0.30
  0.20  0.30  0.50

2 multiplications

  0.27  0.31  0.42
  0.22  0.34  0.44
  0.24  0.32  0.44

4 multiplications

  0.24  0.32  0.43
  0.24  0.32  0.44
  0.24  0.32  0.44

8 multiplications

  0.24  0.32  0.44
  0.24  0.32  0.44
  0.24  0.32  0.44

16 multiplications

  0.24  0.32  0.44
  0.24  0.32  0.44
  0.24  0.32  0.44
```

```
The transition matrix

  0.10  0.00  0.90
  0.90  0.00  0.10
  0.00  0.90  0.10

2 multiplications

  0.01  0.81  0.18
  0.09  0.09  0.82
  0.81  0.09  0.10

4 multiplications

  0.22  0.10  0.68
  0.67  0.15  0.17
  0.10  0.67  0.23

8 multiplications

  0.18  0.50  0.32
  0.27  0.21  0.53
  0.50  0.27  0.23

16 multiplications

  0.33  0.28  0.40
  0.36  0.32  0.32
  0.28  0.36  0.36

32 multiplications

  0.32  0.32  0.36
  0.32  0.32  0.36
  0.32  0.32  0.35

64 multiplications

  0.32  0.32  0.36
  0.32  0.32  0.36
  0.32  0.32  0.36
```

It is noticeable that the each column shares exactly the same number in the three rows. This is no coincidence however, but a consequence of the fact that the vectors are a probability distribution

If we look at a probability vector $p = (p_{1,} p_{2,} p_{3,}) \quad where \quad p_1 + p_2 + p_3 = 1$ and a matrix

$$P^{(n)} = \begin{pmatrix} a & b & c \\ a & b & c \\ a & b & c \end{pmatrix}$$

$$(p_{1,} p_{2,} p_{3,}) \begin{pmatrix} a & b & c \\ a & b & c \\ a & b & c \end{pmatrix} = (p_1 a + p_2 a + p_3 a, p_1 b + p_2 b + p_3 b, p_1 c + p_2 c + p_3 c) =$$
$$((p_1 + p_2 + p_3)a, (p_1 + p_2 + p_3)b, (p_1 + p_2 + p_3)c) = (a, b, c).$$

So $(p_{1,} p_{2,} p_{3,})$ is an eigenvector with eigenvalue one (a stationary state) only if $(a, b, c) = (p_{1,} p_{2,} p_{3,})$.
It is easy to convince oneself, that this is also a necessary condition.