2 DISCRETE-TIME MARKOV CHAINS

2.1 FUNDAMENTAL DEFINITIONS AND PROPERTIES

From now on we will consider processes with a countable or finite state space
\[ S \subseteq \{0, 1, 2, \ldots \}. \]

**Definition 1.** A discrete-time discrete-state stochastic process

\[ \{X_n : n \geq 0\} \]

with a state space \( S \subseteq \{0, 1, 2, \ldots \} \) is called a discrete-time Markov chain (or only a Markov chain) if and only if it has a Markov property

\[ P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i) \quad (2.1) \]

for all \( n \geq 0, i, j, i_0, i_1, \ldots, i_{n-1} \in S \) (the probability of the next state given the current state and the entire past depends only on the current state).

The Markovian property (2.1) is a constraint on the memory of the process: knowing the immediate past means the earlier outcomes are no longer relevant.

Alternatively: the future is conditionally independent of the past given the present.

Or: given the evolution of the process \( \{X_n : n \geq 0\} \) up to any ”current” time \( n \), the probabilistic description of its behaviour at time \( n+1 \) (and, by induction, the probabilistic description of all its subsequent behaviour) depends only on the current state \( \{X_n = i\} \), and not on the previous history of the process.

**Definition 2.** We say that a Markov chain \( \{X_n : n \geq 0\} \) is time-homogeneous, if the conditional probabilities (2.1) are independent of \( n \) for all \( i, j \in S \), that is, for all \( i, j \in S \) there exist \( p_{ij} \) such that

\[ P(X_{n+1} = j | X_n = i) = p_{ij} \quad \text{for all} \quad n \geq 0. \quad (2.2) \]

The numbers \( p_{ij} \) are called one-step transition probabilities and the matrix

\[ P = (p_{ij})_{i,j \in S} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} & \cdots \\ p_{21} & p_{22} & \cdots & p_{2m} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ p_{m1} & p_{m2} & \cdots & p_{mm} & \cdots \end{pmatrix} \quad (2.3) \]
is referred to as the **one-step transition matrix** of the Markov chain.

Thus, for a **time-homogeneous Markov chain** \( \{X_n : n \geq 0\} \) we have: given its evolution up to any "current" time \( n \), the probabilistic description of its behaviour at time \( n+1 \) (and, by extension, the probabilistic description of all its subsequent behaviour) depends only on the current state \( \{X_n : n \geq 0\} \), and not on the previous history of the process nor on the time \( n \) itself.

Throughout this course all Markov chains are assumed to be time-homogeneous.

Note that necessarily
\[
p_{ij} \geq 0 \quad \text{for all } i, j \in S, \quad \sum_{j \in S} p_{ij} = 1 \quad \text{for all } i \in S,
\]
i.e. \( P \) is a **stochastic matrix**.

The number \( p_{ij} \) is interpreted as the probability that the state of the process changes (has a "transition") from state \( i \) to state \( j \) in one time step.

**Proposition 1.** The sample paths of a Markov chain are completely characterized by the one-step transition probabilities, \( p_{ij} \), \( i, j \in S \), and initial-state probabilities, \( p_i \), \( i \in S \). Given values for these probabilities, the joint probability of any finite sample path can be decomposed into a product of an initial-state probability and the appropriate sequence of transition probabilities.

Let \( S = \{1, 2, \ldots, m\} \). The rows of a one-step transition matrix
\[
P = \begin{pmatrix}
p_{11} & p_{12} & \cdots & p_{1m} \\
p_{21} & p_{22} & \cdots & p_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
p_{m1} & p_{m2} & \cdots & p_{mm}
\end{pmatrix}
\] (2.4)

represent the last state of the process, while the columns represent the next state of the process. Each row of \( p \) corresponds to a conditional mass function and therefore sums to 1. For instance, row \( i \) corresponds to
\[
P(N = j | L = i) \quad \text{for } j = 1, 2, \ldots, m.
\]

For the purposes of simulation the matrix can be rewritten so that each row corresponds to a conditional cumulative distribution function:
\[
\Sigma P = \begin{pmatrix}
p_{11} & p_{11} + p_{12} & p_{11} + p_{12} + p_{13} & \cdots & 1 \\
p_{21} & p_{21} + p_{22} & p_{21} + p_{22} + p_{23} & \cdots & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
p_{m1} & p_{m1} + p_{m2} & p_{m1} + p_{m2} + p_{m3} & \cdots & 1
\end{pmatrix}
\] (2.5)

Row \( i \) of \( \Sigma P \) corresponds to
\[
P(N \leq j | L = i) \quad \text{for } j = 1, 2, \ldots, m.
\]
The initial-state probabilities can be organized into an initial-state vector 
\[ p = (p_1, p_2, \ldots, p_m) \]  
with \[ p_i = P(X_0 = i) \] for all \( i \in S \). (2.6)

This vector representing the probability mass function of \( X_0 \) is called the initial distribution of the given Markov chain.

For the purpose of simulation the vector can be rewritten so that it corresponds to the cumulative distribution function:

\[ \Sigma p = (p_1, p_1 + p_2, p_1 + p_2 + p_3, \ldots, 1). \]

**Definition 3.** A state \( i \) for which \( p_{ii} = 1 \) is called an absorbing state.

**Transition Diagrams**

A Markov chain transition matrix can be represented graphically as a transition-probability diagram where each node represents a state of the system and is numbered accordingly, directed arc connects state \( i \) to state \( j \) if a one-step transition from \( i \) to \( j \) is possible. The one-step transition probability \( p_{ij} \) is written next to the arc. Notice that a transition from a state to itself is represented by a loop.

**Example 1.** One of the simplest discrete-time Markov chains is one with two states, \( S = \{1, 2\} \). A transition matrix is in this case

\[ P = \begin{pmatrix} \alpha & 1 - \alpha \\ 1 - \beta & \beta \end{pmatrix} \]

\[ 1 - \alpha \]

\[ \alpha \]

\[ 1 - \beta \]

\[ \beta \]

**Fig. 2.1:** Transition-probability diagram for the model from Example 1

**Example 2.** As an example of a two-state Markov chain we can consider a simple weather forecasting model in which we classify the day’s weather as either "sunny" of "rainy". On the basis of previous data we have determined that if it is sunny today there is an 80% chance that it will be sunny tomorrow regardless of the past weather, whereas if it is rainy today there is a 30% chance that it will be rainy tomorrow, regardless of the past. Let \( X_n \) be the weather on day \( n \). We shall label "sunny" as state 1 and "rainy" as state 2. Then \( \{X_n : n \geq 0\} \) is a two-state Markov chain with the transition probability matrix
$\begin{align*}
P &= \begin{pmatrix}
0.8 & 0.2 \\
0.7 & 0.3
\end{pmatrix}
\end{align*}$

\textbf{Fig. 2.2: Transition-probability diagram for the model from Example 2}

\textbf{Example 3.} Consider

$$
P = \begin{pmatrix}
0 & 0.95 & 0.01 & 0.04 \\
0 & 0.27 & 0.63 & 0.10 \\
0 & 0.36 & 0.40 & 0.24 \\
0 & 0 & 0 & 1
\end{pmatrix}
$$

\textbf{Fig. 2.3: Transition-probability diagram for the model from Example 3}
2.2 MARGINAL DISTRIBUTIONS

To predict the state of the system at time \( n, n \geq 0 \), we need to compute the marginal distribution of \( X_n \).

Let

\[
p^{(n)}_j = P(X_n = j) = \sum_{i \in S} P(X_n = j | X_0 = i) P(X_0 = i)
\]

(2.7)

where \((p_i)\) is the initial distribution (2.6). Thus if we can compute the conditional probabilities

\[
p^{(n)}_{ij} = P(X_n = j | X_0 = i) \quad i, j \in S,
\]

(2.8)

we can compute the marginal distribution of \( X_n \).

The quantities \( p^{(n)}_{ij} \) are called the \textit{n-step probabilities}.

\textbf{Example 4. Beer Brand}

Consider two brands of beer (\textit{Pilsner} and \textit{Starobrno}). Let the following transition matrix describe the probability that you next purchase either brand, given your current purchase. (There is some brand loyalty: The probability of purchasing the same brand again is higher).

\[
P = \begin{pmatrix}
P_{\text{Pilsner}} & 0.95 & 0.05 \\
P_{\text{Starobrno}} & 0.10 & 0.90
\end{pmatrix}
\]

Suppose you currently purchase brand 1, \textit{Starobrno}. What is the probability that you purchase \textit{Pilsner} in two periods from now? In order to answer this question we have to derive \( p^{(2)}_{21} \):

\[
p^{(2)}_{21} = \sum_{i=1}^{2} p_{2i} p_{i1} = 0.10 \times 0.95 + 0.90 \times 0.10 = 0.1850,
\]

that is, the chance of switching from brand 2 (\textit{Starobrno}) to brand 1 (\textit{Pilsner}) in two periods is 18.5%.

The same result can be obtained by using matrix multiplication:

\[
p^2 = \begin{pmatrix}
0.95 & 0.05 \\
0.10 & 0.90
\end{pmatrix} \begin{pmatrix}
0.95 & 0.05 \\
0.10 & 0.90
\end{pmatrix} = \begin{pmatrix}
0.9075 & 0.0925 \\
0.1850 & 0.8150
\end{pmatrix}
\]

Each element of this matrix, \( P^{(2)}_{ij} \), tells you the probability that you go from state \( i \) to state \( j \) after 2 periods. For example, the probability that someone who buys brand 1 in the current period will still buy brand 1 in two periods is 0.9075.

If the initial distribution is \( p = (0.30, 0.70) \), i.e., about 30% of customers buy \textit{Pilsner} and about 70% buy \textit{Starobrno}, what is the fraction of consumers that buys \textit{Starobrno} after 2 periods?

This distribution is given by

\[
p^{(n)} = p P^2 = (0.30, 0.70) \begin{pmatrix}
0.9075 & 0.0925 \\
0.1850 & 0.8150
\end{pmatrix} = (0.4018, 0.5983),
\]

that is, the fraction that buy \textit{Starobrno} shrinks to 59.8%.
Similarly, we can continue further and observe the system evolution over time:

\[
P^3 = \begin{pmatrix} 0.8714 & 0.1286 \\ 0.2573 & 0.7428 \end{pmatrix} \quad P^{10} = \begin{pmatrix} 0.7323 & 0.2677 \\ 0.5354 & 0.4646 \end{pmatrix}
\]

\[
P^{20} = \begin{pmatrix} 0.6796 & 0.3204 \\ 0.6408 & 0.3592 \end{pmatrix} \quad P^{100} = \begin{pmatrix} 0.6667 & 0.3333 \\ 0.6667 & 0.3333 \end{pmatrix}
\]

Note that there is one striking result, the fraction of people buying Pilsner vs. Starobroň is independent of the initial distribution.

\section*{Example 5. The Gambler’s Ruin}

You currently have $2 dollars and you can win a dollar with probability \( p \) or lose it with probability \( (1 - p) \). You stop playing if you reach $4 or if you go broke. The transition matrix is

\[
P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ (1 - p) & 0 & p & 0 & 0 \\ 0 & (1 - p) & 0 & p & 0 \\ 0 & 0 & (1 - p) & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}
\]

Since you start with $2, the initial distribution is \( p = (0, 0, 1, 0, 0) \).

If the game is fair, the probability \( p = 0.5 \) and the transition matrix is

\[
P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}
\]

In the long run the game will end and there are only two possible outcomes: winning enough to stop at $4 or go broke. And this time it matters whether we start at $2 or $3 whether we will end the game by winning enough to get to $4 or go broke.

\[
P^2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0.25 & 0 & 0.25 & 0 \\ 0.25 & 0 & 0.5 & 0 & 0.25 \\ 0 & 0.25 & 0 & 0.25 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad P^{100} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.75 & 0 & 0 & 0.25 & 0 \\ 0.50 & 0 & 0 & 0 & 0.50 \\ 0.25 & 0 & 0 & 0 & 0.75 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}
\]

Note that both the state of going broke and reaching $4 are absorbing states (compare Definition 3).

The $n$-step transition probabilities satisfy the following equations:

$$ p_{ij}^{(n)} = \sum_{r \in S} p_{ir}^{(k)} p_{rj}^{(n-k)} \quad \text{for all } i, j \in S, \quad (2.9) $$

where $k$ is a fixed integer such that $0 \leq k \leq n$.

Proof. Fix an integer $k \in \{0, 1, 2, \ldots, n\}$. It is

$$ p_{ij}^{(n)} = P(X_n = j | X_0 = i) = \sum_{r \in S} P(X_n = j, X_k = r | X_0 = i) $$

$$ = \sum_{r \in S} P(X_n = j | X_k = r, X_0 = i) P(X_k = r | X_0 = i) \quad \text{[from Markov property (2.1)]} $$

$$ = \sum_{r \in S} P(X_{n-k} = j | X_0 = r) P(X_k = r | X_0 = i) \quad \text{[from time-homogeneity (def. 2)]} $$

$$ = \sum_{r \in S} p_{rj}^{(n-k)} p_{ir}^{(k)} \quad \text{[by Equation (2.8)]} $$

which proves the theorem. \(\square\)

In Matrix notation: let

$$ P^{(n)} = \begin{pmatrix} p_{ij}^{(n)} \end{pmatrix} $$

be a matrix of the $n$-step transition probabilities (obviously, $P^{(1)} = P$) and

$$ p^{(n)} = (p_1^{(n)}, p_2^{(n)}, \ldots, p_m^{(n)}) $$

be a vector of the probability mass function of $X_n$ (obviously, $p^{(0)} = p$).

Theorem 3.

$$ P^{(n)} = P^n, \quad (2.10) $$

where $P^{(n)}$ is the $n$-th power of $P$.

Proof. The Chapman-Kolmogorov equations (2.9) can be written in a matrix form:

$$ P^{(n)} = P^{(n-k)} P^{(k)} \quad (2.11) $$

Denote

$$ P(X_0 = j | X_0 = i) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise}. \end{cases} $$

Thus

$$ P^{(0)} = E, \quad P^{(1)} = P. $$
The theorem is therefore true for \( n = 0 \) and \( n = 1 \).

Suppose it is true for \( n = k \); we want to prove that it is true for \( n = k + 1 \), too. From Chapman-Kolmogorov equations (2.9) and using the induction hypothesis,

\[
P^{(k+1)} = P^{(k)} P^{(1)} = P^k P = P^{k+1}.
\]

The theorem follows by induction. \( \square \)

**Corollary 1.**

\[
p^{(n)} = p P^n .
\]  
(2.12)

**Proof.** \( p^{(n)} = p^{(0)} P^{(n)} = a P^n \). \( \square \)

**Example 6.** Let \( \{X_n : n \geq 0\} \) be a Markov chain with state space \( S = \{1, 2, 3, 4\} \) and the transition probability matrix

\[
P = \begin{pmatrix}
0.1 & 0.2 & 0.3 & 0.4 \\
0.2 & 0.2 & 0.3 & 0.3 \\
0.5 & 0 & 0.5 & 0 \\
0.6 & 0.2 & 0.1 & 0.1
\end{pmatrix}.
\]

Let the initial distribution be

\[
p = (0.25, 0.25, 0.25, 0.25).
\]

Find the following probabilities:

a) \( P(X_3 = 4, X_2 = 1, X_1 = 3, X_0 = 1) \)

b) \( P(X_3 = 4, X_2 = 1, X_1 = 3) \)

c) Compute \( p^{(4)} \).

**Solution**

a) \( P(X_3 = 4, X_2 = 1, X_1 = 3, X_0 = 1) = P(X_0 = 1)p_{13}P_{31}p_{14} = \)

\[
= 0.25 \times 0.3 \times 0.5 \times 0.4 = 0.015.
\]

b) \( P(X_3 = 4, X_2 = 1, X_1 = 3) = \)

\[
= \sum_{i=1}^{4} P(X_3 = 4, X_2 = 1, X_1 = 3|X_0 = i)P(X_0 = i) = \sum_{i=1}^{4} p_{i} p_{i3} p_{31} p_{14} =
\]

\[
= (0.5 \times 0.4) \sum_{i=1}^{4} p_{i} p_{i3} = 0.20 \times 0.25 \times (0.3 + 0.3 + 0.5 + 0.1) = 0.06
\]

c) According to Corollary 2 we have

\[
p^{(4)} = p P^4 = (0.25, 0.25, 0.25, 0.25) \times \begin{pmatrix}
0.3616 & 0.1344 & 0.3192 & 0.1848 \\
0.3519 & 0.1348 & 0.3222 & 0.1911 \\
0.3330 & 0.1320 & 0.3340 & 0.2010 \\
0.3177 & 0.1404 & 0.3258 & 0.2161
\end{pmatrix} =
\]

\[
= (0.34105, 0.1354, 0.3253, 0.19825)
\]
2.2. MARGINAL DISTRIBUTIONS

2.2.1 Computation of Matrix Powers

Methods of Direct Multiplication

Apparently the most straightforward way how to compute the matrix power that is necessary for finding the marginal distribution of $X_n$ is a direct matrix multiplication. When the matrixes are small or sparse, it is not difficult, but for large matrices that are not sparse, the direct algorithms requires too many arithmetic steps ($O(nm^3)$ or $O(\log nm^3)$) and require $3m^2$ memory space.

Numerically inferior to these methods is the following one.

Method of Generating Functions

Suppose we are interested in studying a sequence $\{p_k : k \geq 0\}$. One useful method is to first compute its generating function $P(z)$, defined by

$$P(z) = \sum_{k=0}^{\infty} p_k z^k. \quad (2.13)$$

In many cases, $P(z)$ can be easily computed. If it is a rational function of $z$, it can be inverted to obtain expressions for $p_k$.

Generating functions can also be easily defined for sequences of matrices. In our case, we need to study the generating function of $\{P^n, n \geq 0\}$ defined as follows:

$$P(z) = \sum_{n=0}^{\infty} z^n P^n, \quad (2.14)$$

where $z$ is a complex number. The above series converges absolutely whenever $|z| < 1$. Now,

$$P(z) = E + \sum_{n=0}^{\infty} z^n P^n = E + zP(z)P. \quad (2.15)$$

Hence,

$$P(z) = (E - zP)^{-1}. \quad (2.16)$$

In this way we have reached the method of generating function:

Step 1 Compute

$$A(z) = (a_{ij}(z)) = (E - zP)^{-1}.$$  

Step 2 The function $a_{ij}(z)$ is a rational function of $z$. Expand it as a power series in $z$,

$$a_{ij}(z) = \sum_{n=0}^{\infty} a_{ij}^{(n)} z^n.$$  

Step 3 $P^n = (a_{ij}^{(n)})$ is the required power of $P$.

More on generating functions can be found for example in the Internet address:

2.3 FIRST PASSAGE TIMES

2.3.1 First Passage Time and Its Probability Distribution

Definition 4. Let \( \{X_n, \ n \geq 0\} \) be a discrete-time Markov Chain with state space \( S = \{0, 1, 2, \ldots\} \), transition probability matrix \( P \) and initial distribution \( p \).

Denote
\[
T = \min\{n \geq 0 : X_n = 0\}.
\]
(2.17)

The random variable \( T \) is called a first passage time (into state 0).

Probability distribution of \( T \)

Let
\[
\alpha_i = P(T = n|X_0 = i)
\]
(2.18)
denote the probability mass function of \( T \),

\[
u_i(n) = P(T \leq n|X_0 = i).
\]
(2.19)

The following theorem provides a recursive method of computing \( u_i(n) \).

Theorem 4.

\[
u_i(n) = p_{i0} + \sum_{j=1}^{\infty} p_{ij}u_j(n-1) \quad \text{for all } i \geq 1, \ n \geq 1,
\]
(2.20)

with
\[
u_i(0) = 0 \quad \text{for all } i \geq 1.
\]

2.3.2 Absorption Probabilities

Now we will study methods of computing the so-called probability of visiting state 0:

\[
u = P(X_n = 0 \text{ for some } n \geq 0) = P(T < \infty),
\]
(2.21)

where \( T \) is the first passage time defined by Equation (2.17).

In many applications the state 0 is an absorbing state – once \( X_n \) visits 0 it can not leave it. The quantity \( \nu \) is therefore called the absorption probability.

Let
\[
u = P(T < \infty|X_0 = i), \quad i \geq 1.
\]
(2.22)

Then
\[
u_i = \lim_{n \to \infty} u_i(n),
\]
(2.23)

where \( u_i(n) \) is defined by Equation (2.20).
Theorem 5. The \( \{u_i, \ i \geq 1\} \) is the smallest non-negative solution to

\[
 u_i = p_{i0} + \sum_{j=1}^{\infty} p_{ij} u_j .
\]  

This theorem is not directly useful since it does not tell us how to identify the smallest solution. For a finite-state discrete-time Markov chain we can proceed as follows:

Suppose state 0 is not accessible from a given state \( i \). Then if the Markov chain starts in state \( i \), time to reach 0 is \( \infty \). Hence \( u_i = 0 \) for such a state. Let \( A \) be the set of all such states with \( u_i = 0 \). Then Equation (2.24) becomes

\[
 u_i = p_{i0} + \sum_{\substack{j \neq 1 \ j \notin A}}^{\infty} p_{ij} u_j .
\]

These equations have a unique solution, it must therefore be the required solution.
2.4 FURTHER READING

Discrete-Time Markov Chains
In addition to the recommended literature, more on discrete-time Markov chains can be found in following Internet addresses:

- http://en.wikipedia.org/wiki/Markov_chain
- http://screwdriver.bu.edu/cn760-lectures/19n/19.htm

More Examples and Applications