Bachelor Thesis

Bachelor of Science, LMU Munich, Department of Statistics

Discrete – Time Markov Chains

On the transition from countable to continuous state spaces

Lorenz Leopold Mihatsch

Munich, the 27th of July 2020

Supervisor: Dennis Mao, M.Sc.
Referent: Prof. Stefan Mittnik, PhD

Abstract

In this thesis we will discuss basic concepts of homogenous discrete time Markov chains. We will start by introducing stochastic processes as a series of random variables $\{X_n\}_{n\in\mathbb{N}_0}$ taking values in the same measurable space $(E, \mathcal{E})$, which are called *states*. The basic idea of Markov chains is that the probability of $\{X_n\}_{n\in\mathbb{N}_0}$ to adopt some state only depends on the previous state, but not on the ones before. This property is called *Markov property*. In case of discrete state spaces we describe the transition from one state to another by stochastic matrices, which we will then generalize to continous state spaces by introducing *Markov kernels*. Throughout this thesis we will analyze serval different examples of Markov chains in discrete as well as continous state spaces.

Preface

The content of this thesis is mainly based on three books, which are cited at the end of this work. As basic concepts of Markov chains are addressed here, it will be refrained from citing within the main text. The first chapter on *countable state space Markov chains* is mainly based on Brémaud (1999) while chapter two on *continuous state space Markov chains* is based on Cappé (2005). The introductory part of the second chapter is based on Breiman (1992). References for the introduction are formally cited within the text as well as at the end of the thesis.

I would like to thank my supervisor Dennis Mao, M.Sc. and referent Prof. Mittnik, PhD for giving me the oportunity to work on this topic and for their constant supervision and care.

Contents

1 Introduction

The genetic code across all known species consists of four different nucleobases, namely adenine [A], cytosine [C], thymine [T] and guanine [G]. For each gene, independently of the organism at scope, a unique sequence of these nucleobases is linked together to provide the information necessary for building proteins. When we arbitrarily pick a gene, we can easily calculate the frequency of each nucleobase from the genetic sequence. Taking the coding sequence of *CD274* (gene of the "Programmed Death Ligand 1") we obtain a sequence of 3634 nucleobases starting with `AGTTCTGCGCAGCTTCCCG...` and empirical (rel.) frequencies

A	C	T	G
0.284	0.201	0.308	0.206

.

If we assume the sequence of nucleobases to be a sequence of independent, identically distributed random variables $\{X_n\}_{n\in\mathbb{N}}$ we would expect the probability of a G following a C to be $P(X_i = \mathtt{C}, X_{i+1} = \mathtt{G}) = 0.201 \cdot 0.206 = 0.041$ for any i in $\{1, ..., n\}, n \in \mathbb{N}$. However empirically we obtain a frequency of 0.010, which diverges from the previous considerations by a factor of about 4. [4]

It was the very same argument Andrei Andrejewitsch Markov (1856–1922) used in 1913 to show the necessity of independence of random variables for the *law of large numbers*. As opposed to the above example Andrei Markov used the first 20.000 letters of Alexander Sergejewitsch Puschkin's novel *Eugen Onegin* to show that the frequency of two successive vocals differed considerably from the theoretical probability under the assumption of independence. With his work Andrei Markov contributed significantly to the theory of stochastic processes, where in a sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ the probability of observing some event is conditioned on the event before. [5] In honor of Andrei Markov, such stochastic processes were later called *Markov chains*[1] with the observed events called *states*.

In fact from the genetic sequence given, we can estimate the probabilities of the sequence transiting from one nucleobase (or state) to another.

	A	C	G	T
A	0.297	0.174	0.236	0.292
C	0.370	0.249	0.051	0.331
G	0.299	0.205	0.223	0.273
T	0.205	0.194	0.268	0.334

[1] Ususally one refers to *Markov chains* if the process is discrete in time, as opposed to *Markov processes* were time is assumed to be continuous.

We will later see that this table is indeed a transition matrix of a discrete time Markov chain in a countable finite state space. Markov chains of this type are the most simple form of Markov processes. They can further be extended to processes with countable infinite or even continuous state spaces and/or continuity in time.

Markov chains find their applications in various diciplines, ranging from natural sciences to social sciences, because many real life phenomenons can be described as such. Furthermore Markov chains also build the basis for a range of other model classes such as Hidden Markov Models, Variable-order Markov Models and Markov renewal processes.

This thesis will focus mainly on the extension of Markov chains in a countable state space (finite and infinite) to Markov chains in continuous state spaces. Therefore so called *Markov kernels* will be introduced. For some selected properties of discrete space Markov chains it will be shown how they translate to a continuous state space. In any case however, this thesis is restricted to Markov chains in discrete time.

2 Countable state space Markov chains

Definition 2.1 (Stochastic process)
Let $(\Omega, \mathcal{A}, P)$ *be a probability space and* $(E, \mathcal{E})$ *a measurable space. A* stochastic process *in discrete time is a sequence of random variables* $\{X_n\}_{n\in\mathbb{N}_0}$ *with* $X_n : \Omega \longrightarrow E$ *for all* $n \in \mathbb{N}_0$.

The measurable space $(E, \mathcal{E})$ is also referred to as the *state space* and the index $n \in \mathbb{N}_0$ is interpreted as steps in time. $\{X_n\}_{n\in\mathbb{N}_0}$ therefore describes how some stochastic process evolves in time.[2] We will now narrow stochastic processes down to processes with a very specific dependency structure.

Definition 2.2 (Markov chain) *Let* $\{X_n\}_{n\in\mathbb{N}_0}$ *be a stochastic process on* $(\Omega, \mathcal{A}, P)$ *with values in a discrete state space* $(E, \mathcal{E})$. $\{X_n\}_{n\in\mathbb{N}_0}$ *is called* Markov chain, *if for all* $n \in \mathbb{N}_0$ *and all states* $i, i_0, ..., i_{n-1} \in E$ *and* $\{j\} \in \mathcal{E}$ *with* $P(X_0 = i_0, ..., X_{n-1} = i_{n-1}, X_n = i) > 0$

$$P(X_{n+1} = \{j\}|X_0 = i_0, ..., X_{n-1} = i_{n-1}, X_n = i) = P(X_{n+1} = \{j\}|X_n = i). \tag{1}$$

(1) *is called* Markov property.

The *Markov property* states that the probability of observing some state $\{j\} \in \mathcal{E}$ at some timepoint $n + 1 \in \mathbb{N}_0$, only depends on the state $i \in E$ at timepoint $n \in \mathbb{N}_0$, but not on the states before. More precisely one calls stochastic processes that satisfy (1) Markov chains of *1*st*-order*. If the probability depends on n previous time points, the Mrakov chain is of n^{th}*-order*.

If (1) is independet of $n \in \mathbb{N}_0$ the Markov chain is called *homogenous*. This means, that at any given timepoint $n \in \mathbb{N}$ the conditional probabilities are identical.

It seems a bit finical at this point to write $P(X_{n+1} = \{j\}|X_n = i)$ instead of $P(X_{n+1} = j|X_n = i)$. We use the more sophisticated notation to specify (firstly) that a probability measure $P : \mathcal{A} \longrightarrow [0, 1]$ is always defined on a $\sigma-$algebra and (secondly) that the conditional probabilities in terms of

[2]More general the sequence of random variables $\{X_n\}_{n\in I}$ could be on any index set $I \subseteq \mathbb{N}_0$ or even $I \subseteq \mathbb{R}_+ = [0, \infty)$, when a time continuous process is at scope.

Markov chains are conditioned on sets of single elements. This will be especially important for the introduction to *Markov kernels*, which are important for nondenumerable state spaces. This will be discussed later. For discrete state spaces we set $\mathcal{E} = \mathcal{P}(E)$.

Lemma 2.3 *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $B \in \mathcal{A}$ with $P(B) > 0$. Then*

$$P(\cdot|B) : \mathcal{A} \longrightarrow [0,1], \quad A \mapsto P(A|B),$$

is a probability measure for all $B \in \mathcal{A}$.

Proof: (P0) We show, that $P(A|B) \in [0,1]$ for all $B \in \mathcal{A}$. It is obvious that $P(A|B) \geq 0$. Due to the monotony of P

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \leq \frac{P(B)}{P(B)} = 1.$$

(P1)

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = 1$$

(P2) Let $A_1, A_2, ... \in \mathcal{A}$ be pairwise disjoint. Then

$$P(\bigcup_{i=1}^{\infty} A_i|B) = \frac{P(\bigcup_{i=1}^{\infty}(A_i|B))}{P(B)} = \sum_{i=1}^{\infty} \frac{P(A_i|A)}{P(B)} = \sum_{i=1}^{\infty} P(A_i|B)$$

□

Definition 2.4 (Stochastic matrix)
A matrix $\Pi \in [0,1]^{E \times \mathcal{E}}$ is called stochastic, *if*

$$\sum_{\{j\} \in \mathcal{E}} \Pi(i, \{j\}) = 1 \quad \textit{for all} \quad i \in E.$$

This means, that the row sum is equal to 1.

Definition 2.5 (Transition Matrix)
A stochastic matrix is called transition matrix *for a homogenous discrete state space Markov chain* $\{X_n\}_{n\in\mathbb{N}_0}$*, if*

$$\Pi(i,\{j\}) = P(X_{n+1} = \{j\}|X_n = i) \tag{2}$$

for all $i \in E, \{j\} \in \mathcal{E}$ *and* $n \in \mathbb{N}_0$ *with* $P(X_n = i) > 0$. $\Pi(i,\{j\})$ *are called* transition probabilities.

The prerequisite $P(X_n = i) > 0$ in (2) is neccessary for the transition matrix to be a stochastic matrix. Only then Lemma 2.3 garanties

$$\sum_{\{j\}\in\mathcal{E}} \Pi(i,\{j\}) = \sum_{\{j\}\in\mathcal{E}} P(X_{n+1} = \{j\}|X_n = i) = P(\bigcup_{\{j\}\in\mathcal{E}} \{X_{n+1} = j\}|X_n = i) = 1.$$

It is therefore advisable to choose the state space accordingly. Otherwise the transition matrix does not have a row sum equal to 1 and it is no stochastic matrix.

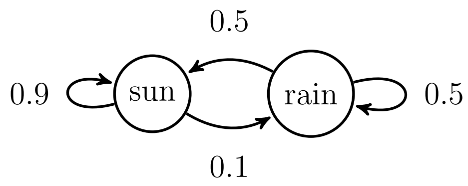
Having now given a short introduction into discrete state space Markov chains, we will now have a look at some examples.

Example 2.7 (Oversimplified model of the weather)

Let us examine the weather of some sunny place on a daily basis. To do so we confine to a *countable finite state space* $(E,\mathcal{E}) = (\{s,r\},\mathcal{P}(\{s,r\}))$, where $s \mathrel{\hat{=}}$ Sun and $r \mathrel{\hat{=}}$ Rain. We assume that for the weather of tomorrow only today's weather is relevant, but not the weather of the days before. Further we assume the transition probabilities from one weather state to another to be independent of the time and known with the following transition matrix

$$\Pi(i,\{j\}) = \begin{array}{c c} & \begin{array}{cc} \{s\} & \{r\} \end{array} \\ \begin{array}{c} s \\ r \end{array} & \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix} \end{array}, \quad i \in E \text{ and } \{j\} \in \mathcal{E}.$$

In this example the probability of enjoying sun another day following a sunny day is 90%. One can easily confirm that this transition matrix is indeed a stocastic matric, as the row sum is euqal to 1 for both rows. A common way to visualize the transitions between different states is by drawing *transition graphs* (Appendix A.1 and A.2). For our weather model we recieve the following transition graph:

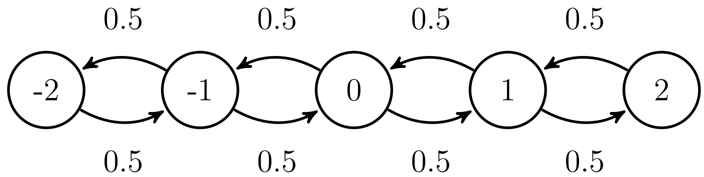


Example 2.8 (Random walk on $\mathbb{Z}$)

Let us assume we toss a fair coin, where we depict *head* $= 1$ and *tail* $= -1$. We model the outcome of a series of independent, identically distributed random variables $\{\xi_i\}, i \in \mathbb{N}$ which take values in $(\{1,-1\}, \mathcal{P}(\{1,-1\}))$ with $P(\xi_i = \{1\}) = \frac{1}{2} = P(\xi_i = \{-1\})$. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be another series of random variables, that takes values in $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$ and is given by

$$X_0 = 0, \quad X_{n+1} = X_n + \xi_{n+1} = \sum_{i=1}^{n+1} \xi_i.$$

One can illustrate the process as followed:



X_n gives us the position of some particle at timepoint $n \in \mathbb{N}_0$. This particle starts in 0. At any given timepoint $n \in \mathbb{N}_0$ it tosses a coin. When *head* shows, it jumps to the right, when *tail* shows, it jumps to the left. It therefore "walks" on $\mathbb{Z}$. We can define the transition probabilities as

$$\Pi(i,\{j\}) = \begin{cases} \frac{1}{2}, & \textit{if } \{j\} \in \{\{i+1\},\{i-1\}\} \\ 0, & \textit{else.} \end{cases}$$

Of note: $\Pi(i,\{j\})$ is the probability of transiting from state i to j. It does not give the probability for $\{X_n\}_{n\in\mathbb{N}_0}$ to be in state i or j.
The following proposition will help us show, that *Example 2.8* is indeed a (homogenous) discrete state space Markov Chain.

Proposition 2.9
Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a stochastic process with a countable state space $(E,\mathcal{E})$ and let $\Pi \in [0,1]^{E\times\mathcal{E}}$ be a stochastic matrix. If

$$P(X_{n+1} = \{j\} | X_0 = i_0, ..., X_{n-1} = i_{n-1}, X_n = i) = \Pi(i,\{j\}) \tag{3}$$

for all $n \in \mathbb{N}_0$*,* $i, i_0, ..., i_{n-1} \in E$ *and* $j \in \mathcal{E}$ *with* $P(X_0 = i_0, ..., X_{n-1} = i_{n-1}) > 0$*. Then* $\{X_n\}_{n\in\mathbb{N}_0}$ *is a homogenous Markov Chain with transition matrix* Π*.*

Proof: We start by showing the following assertion: For all $n \in \mathbb{N}_0$ and $i, i_0, ..., i_{n-1} \in E$ with $P(X_0 = i_0, ..., X_{n-1} = i_{n-1}, X_n = i) > 0$ the following applies:

$$P(X_0 = i_0, ..., X_{n-1} = i_{n-1}, X_n = i) = P(X_0 = i_0) \prod_{k=0}^{n-1} \Pi(i_k, \{i_{k+1}\}) \tag{4}$$

By full induction we show for all $n \in \mathbb{N}_0$:
$n = 0$: clear.
$n \to n + 1$: We set $A_n := \{X_0 = i_0, ..., X_n = i\}$. Due to (3)

$$\begin{aligned} P(X_0 = i_0, ..., X_n = i, X_{n+1} = i_{n+1}) &= P(A_n)P(X_{n+1} = i_{n+1}|A_n) \\ &= P(A_n)\Pi(i_n, \{i_{n+1}\}). \end{aligned}$$

(4) follows therefore from the induction requirement.
To now proove the Markov property (1), it is sufficient so assume $n \in \mathbb{N}_0$, $i, i_0, ..., i_n \in E$, $\{i_{n+1}\} \in \mathcal{E}$ and $P(X_0 = i_0, ..., X_n = i) > 0$ and proove that

$$P(X_{n+1} = \{i_{n+1}\}|X_n = i_n) = \Pi(i_n, \{i_{n+1}\}).$$

For that let $I = \{(i'_0, ..., i'_{n-1}) \in E : P(X_0 = i'_0, ..., X_{n-1} = i'_{n-1}, X_n = i_n) > 0\}$. From (4) it follows:

$$\begin{aligned} P(X_n = i_n) &= \sum_{(i'_0,...,i'_{n-1})\in I} P(X_0 = i'_0, ..., X_{n-1} = i'_{n-1}, X_n = i_n) \\ &= \sum_{(i'_0,...,i'_{n-1})\in I} P(X_0 = i'_0) \prod_{k=0}^{n-2} \Pi(i'_k, \{i'_{k+1}\})\Pi(i'_{n-1}, \{i_n\}). \end{aligned}$$

Analogously it follows that

$$\begin{aligned} &P(X_{n+1} = \{i_{n+1}\}, X_n = i_n) \\ &= \sum_{(i'_0,...,i'_{n-1})\in I} P(X_0 = i'_0, ..., X_{n-1} = i'_{n-1}, X_n = i_n, X_{n+1} = i_{n+1}) \\ &= \sum_{(i'_0,...,i'_{n-1})\in I} P(X_0 = i'_0) \prod_{k=0}^{n-2} \Pi(i'_k, \{i'_{k+1}\})\Pi(i'_{n-1}, \{i_n\})\Pi(i_n, \{i_{n+1}\}) \\ &= P(X_n = i_n)\Pi(i_n, \{i_{n+1}\}). \end{aligned}$$

The assertion follows therefore:

$$P(X_{n+1} = \{i_{n+1}\}|X_n = i_n) = \frac{P(X_{n+1} = \{i_{n+1}\}, X_n = i_n)}{P(X_n = i_n)} = \Pi(i_n, \{i_{n+1}\}).$$

$\square$

Proposition 2.10 (Homogenous Markov chain driven by white noise)
Let $Z_n, n \in \mathbb{N}$ be a random variable on $(\Omega, \mathcal{F}, P)$ with values in $(E', \mathcal{E}')$. Let further be $(E, \mathcal{E})$ be a countable measurable space, $f : E \times E' \longrightarrow E$ a measurable function and $X_0 : \Omega \longrightarrow E$ a random variable, which is independet of $Z_n, n \in \mathbb{N}$. We set

$$X_{n+1} := f(X_n, Z_{n+1}) \quad \textit{for all } n \in \mathbb{N}_0.$$

Then $\{X_n\}_{n\in\mathbb{N}_0}$ is a homogenous Markov chain with a countable state space and transition matrix

$$\Pi(i, \{j\}) = P(f(i, Z_1) = \{j\}) \quad \textit{for all } i, j \in E.$$

Proof: As defined recursively

$$\begin{aligned}
&X_1 = f(X_0, Z_1),\\
&X_2 = f(x_1, Z_2) = f(f(X_0, Z_1), Z_2),\\
&X_3 = f(X_2, Z_3) = f(f(f(X_0, Z_1), Z_2), Z_3), \ \textit{etc.}
\end{aligned}$$

In general

$$X_n = g_n(X_0, Z_1, ..., Z_n)$$

for all $n \in \mathbb{N}$ and a measurable function g_n.
Let $n \in \mathbb{N}, i_0, ..., i_{n-1}, i \in E$ and $\{j\} \in \mathcal{E}$ with

$$P(X_0 = i_0, ..., X_{n-1} = i_{n-1}, X_n = i) > 0.$$

We set $A := \{X_0 = i_0, ..., X_{n-1} = i_{n-1}, X_n = i\}$. It applies

$$\begin{aligned}
P(X_{n+1} = j|X_n = i, ..., X_0 = i_0) &= P(f(X_n, Z_{n+1}) = \{j\}|\{X_n = i\} \cap A)\\
&= P(f(i, Z_{n+1}) = \{j\}|\{X_n = i\} \cap A)\\
&= P(f(i, Z_{n+1}) = \{j\})\\
&= P(f(i, Z_1) = \{j\}) =: \Pi(i, \{j\})
\end{aligned}$$

Remarks: (firstly) $\{X_n = i\} \cap A$ *only depends on* $X_0, X_1, ..., X_n$ *and therefore on* $X_0, Z_1, ...Z_n$. *It is independent of* Z_{n+1}. *(secondly)* Z_{n+1} *is identically distributed as* Z_1.

As Π is a stochastic matrix, the assertion follows from Proposition 2.9.

□

This last proposition will now help us prove, that the random walk on $\mathbb{Z}$ is indeed a Markov chain.

Proof: (Example 2.8 continued)
Let $\{\xi_n\}, n \in \mathbb{N}$ be a series of independent fair coin tosses, with

$$P(\xi_n = -1) = \frac{1}{2} = P(\xi_n = 1).$$

We set $Z_n := \xi_n$ for $n \in \mathbb{N}$, $f(i,j) = i + j$ and for $X_0 := 0$ we obtain

$$X_{n+1} := f(X_n, Z_{n+1}) = X_n + Z_{n+1} = X_n + \xi_{n+1}.$$

By induction we get

$$X_n = \sum_{i=1}^{n} \xi_i,$$

which means, that $\{X_n\}_{n \in \mathbb{N}_0}$ is the random walk on $\mathbb{Z}$. Following Proposition 2.10, it is indeed a homogenous Markov chain with transition matrix

$$\begin{aligned}\Pi(i, \{j\}) &= P(f(i, Z_1) = \{j\}) = P(i + Z_1 = \{j\}) \\ &= P(i + \xi_1 = \{j\}) = P(\xi_1 = \{j - i\}) \\ &= \begin{cases} \frac{1}{2}, & if\ \{j\} \in \{\{i+1\}, \{i-1\}\}, \\ 0, & else. \end{cases}\end{aligned}$$

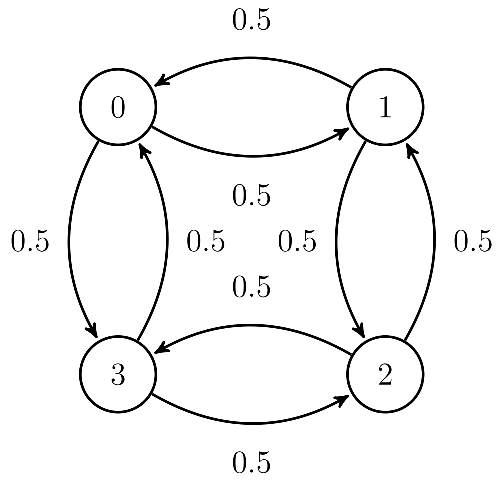
□

Example 2.11 (Random walks on the cyclic group $\mathbb{Z}/4\mathbb{Z}$)
We will now extend the random walk on $\mathbb{Z}$ to a random walk on $\mathbb{Z}/4\mathbb{Z} = \{[0], \dots, [4]\}$ For this purpose we map our random variable $\{X_n\}_{n \in \mathbb{N}_0}$ from the previous example to the residue class modulo 4. This means, that two integers that are maped to the same residue class, have the same residue if divided by 4.

The transition matrix of this process is given by

$$\Pi(i, \{j\}) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} \{0\} & \{1\} & \{2\} & \{3\} \end{array} \\ \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix} \end{array}$$

One can illustrate the process as followed:



We refrain from proving that this is indeed a Markov chain at this point. However it can be observed that for $X_0 = 0$ the process adopts to states $\{0, 2\}$ at even time points and $\{1, 3\}$ at uneven time points.
It follows

$$\Pi^2 = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}, \quad \Pi^3 = \Pi.$$

In general it applies for all $n \in \mathbb{N}_0$:

$$\Pi^{2n} = \Pi^2, \quad \Pi^{2n+1} = \Pi.$$

This means, that the random walk on $\mathbb{Z}/4\mathbb{Z}$ is an example for a *non homogenous* Markov chain.

3 Continuous state space Markov chains

Up to now, we have studied Markov chains in countable finite or infite state spaces $(E, \mathcal{E})$, where the model is described mainly by a stochastic matrix - a non negative matrix $\Pi(i, \{j\})$ where the row sum $\sum_{\{j\}\in\mathcal{E}} \Pi(i, \{j\}) = 1$ for all $i \in E$.
Evidently we can think of Π as family of discrete distributions $\Pi(i, \cdot)$, one for each $i \in E$, where $\Pi(i, \cdot)$ gives us the distribution of X_{n+1} given $X_n = i, n \in \mathbb{N}_0$.

At the end of the past section we have proven that a a random walk on $\mathbb{Z}$ can be modelled by a Markov chain. We have conceived this random walk on $\mathbb{Z}$ a particle jumping from one integer to an adjacent one, meaning $X_{n+1} - X_n \in \{-1, 1\}$. If we stick to the concept of a moving particle, a much more intuitive way would be, if we allow it to move any distance $X_{n+1} - X_n \in \mathbb{R}$ between two timepoints of observation.

Let $(E, \mathcal{E})$ now be a continuous state space, with $E \subset \mathbb{R}$ and $\mathcal{E} = \sigma(\mathcal{O})$ be a Borel σ-algebra on E, with $\mathcal{O}$ a favourable set of open sets on E. Heuristically speaking the family of discrete distributions $\Pi(i, \cdot)$ is now replaced by a family of densities $Q(x, \cdot)$, one for each $x \in E$.

Before giving the formal definition of *transition kernels* we will derive its concept starting with conditional probabilities as a motivation. Lemma 2.3 extends directly to conditional probabilities of random variables that adopt to a countable number of values/states. Instead of $P(A|B) = \frac{P(A\cap B)}{P(B)}, A, B \in \mathcal{F}$ we obtain $P(A|X = x) = \frac{P(A,X=x)}{P(X=x)}$ for a discrete random variable X. In case $P(X = x) = 0$ we arbitrarily define $P(A|X = x) = 0$. However this last restriction can be circumvent by choosing $(E, \mathcal{E})$ well.

Suppose now a random variable X on $(\Omega, \mathcal{A}, P)$ that takes nondenumerable many values. Let further be $B \in \mathcal{E}$ a Borel σ-algebra such that $P(X \in B) > 0$. The conditional probability of $A \in \mathcal{A}$ given $X \in B$ is given by

$$P(A|X \in B) = \frac{P(A, X \in B)}{P(X \in B)}.$$

In sepcific, we are interested in the conditional probability of $A \in \mathcal{A}$ given $X(\omega) = x, \omega \in \Omega$ a single element. In case of continous random variables $P(X(\omega) = x) = 0$ for all x, which is unfavourbale. Therefore we are looking for a way to generalize the definition of conditional probabilities in Lemma

2.3.

If we let $(x - h, x + h) = B \in \mathcal{E}$ an open interval, we can try to take the limit, given by

$$P(A|X = x) = \lim_{h \searrow 0} \frac{P(A, X \in (x - h, x + h))}{P(X \in (x - h, x + h))}, \tag{5}$$

we run into the trouble, that the existence of $x = x_0$ is not garanteed, if $P(X = x_0) = 0$. If we however percieve (5) as a function of x it looks as if we try to take the derivative of one measure with respect to another. Let us therefore define two measures on $\mathcal{E}$. For $B \in \mathcal{E}$, let

$$\tilde{Q}(B) = P(A, X \in B), \tag{6}$$

$$\tilde{P}(B) = P(X \in B). \tag{7}$$

Just like in Lemma 2.3 $\tilde{Q}$ is absolutely continuous with respect to $\tilde{P}$ ($0 \leq \tilde{Q}(B) \leq \tilde{P}(B)$). The Radon-Nikodym-Theorem (Appendix A.3) ensures, that a $\mathcal{E}$-measurable function $\varphi(x)$ exists such that

$$\tilde{Q}(B) = \int_B \varphi(x) \tilde{P}(dx), \quad \textit{for all } B \in \mathcal{E}.$$

It needs to be noted that φ is not uniquely defined. However, with this restriction we can now generalize the conditional probability $P(A|x = x)$ as follows:

Definition 3.1
The conditional probability $P(A|x = x)$ is defined as any $\mathcal{E}$-measurable function satisfying

$$P(A, X \in B) = \int_B P(A|X = x) \tilde{P}(dx),$$

for all $B \in \mathcal{E}$.

Let us now have a closer look how the transition from one state to another is defined in terms of continuous state spaces.

Definition 3.2 (Transition Kernel)
Let $(E, \mathcal{E})$ and $(E', \mathcal{E}')$ be two measurable spaces. An unnormalized transition kernel *from $(E, \mathcal{E})$ to $(E', \mathcal{E}')$ is a function $Q : E \times \mathcal{E}' \longrightarrow [0, \infty)$ that satisfies*

1. *for all $x \in E, Q(x, \cdot)$ is a positive measure on $(E', \mathcal{E}')$*
2. *for all $A \in \mathcal{E}'$, the function $x \mapsto Q(x, A)$ is measurable.*

Remarks: (firstly) If $Q(x, E') = 1$ for all $x \in E$, then Q is called a normalized transition kernel *or simply* kernel. *In this case $Q(x, \cdot)$ is a probability measure. (secondly) If $(E, \mathcal{E}) = (E', \mathcal{E}')$ and $Q(x, E') = 1$, then we refer to Q as a* Markov kernel. *From now on, we will only consider Markov kernels.*

The first requirement of Definition 3.2 makes sure that we get a probability measure for every $x \in E$. Which means that whatever state the Markov chain adopts to, we can specify the probability for any state at the next timepoint.
The second requirement can be viewed in analogy to definition 3.1. It ensures that the conditional probability of $A \in \mathcal{E}'$ exist for single events $x \in E$.

Definition 3.3 (Density of a transition kernel)
Let Q be a Markov kernel and $q : E \times E' \longrightarrow [0, \infty)$ a non negative function, measurabel with respect to the product σ-field $\mathcal{E} \otimes \mathcal{E}'$. Q adopts a density with respect to the positive measure P on E' such that

$$Q(x, A) = \int_{A \in \mathcal{E}'} q(x, y) P(dy), \quad A \in \mathcal{E}'.$$

The function q is then referred to as a transition density.

Let us now have s closer look at the transition matrix in case of homogenous discrete state space Markov chains. Alternatively we can define the transition matrix from Definition 2.5 as a family of functions

$$\Pi : E \times \mathcal{E}' \longrightarrow [0, 1], \quad \Pi(i, B) \mapsto \sum_{j \in B} \Pi(i, \{j\}), \ \forall B \in \mathcal{E}', \tag{8}$$

where $\Pi(i, \{j\}) = P(X_{n+1} = \{j\}|X_n = i)$ as before. We claim that (8) defines a Markov kernel.

Proof: In oder to show that (8) is a Markov kernel, we need to show that (8) satisfies the two requirements from Definition 3.2.

1. follows directly from Lemma 2.3. It shows that for every $i \in E$ we get a probability measure $\Pi(i, \cdot)$.

2. As we consider the case of a discrete state space, that σ-algebra $\mathcal{E}$ is equivalent to $\mathcal{P}(E)$. Therefore any function $i \mapsto \Pi(i, B)$, $\forall B \in \mathcal{E}'$ is measurable with respect to the powerset $\mathcal{P}(E)$.

□

Remarks: This last requirement means, that for every $i \in E$ we assign a row of the transition matrix Π.

Before we can define a homogenous Markov chain in a continuous state space, we need to be able to grasp the history of a stochastic process. For this, we need to introduce filtrations.

Definition 3.4 (Filtration)
Let $(\Omega, \mathcal{F}, P)$ be a probability space. For every $n \in \mathbb{N}$ $\mathcal{A}_n$ is a sub-σ-algebra of $\mathcal{F}$ and the family of σ-algebra is denoted by $\mathbb{A} = \{A_n\}_{n \in \mathbb{N}_0}$.
$\mathbb{A}$ is called a filtration on $(\Omega, \mathcal{F}, P)$ if it is ordered, such that for any $n, m \in \mathbb{N}$ with $m \leq n : \mathcal{A}_m \subseteq \mathcal{A}_n$.

If $\mathbb{A} = \{\mathcal{A}_n\}_{n \in \mathbb{N}_0}$ is a filtration we call $(\Omega, \mathcal{F}, \{\mathcal{A}_n\}_{n \in \mathbb{N}_0}, P)$ a filtered probability space.

Remarks: (firstly) One says a stochastic process $\{X_n\}_{n \in \mathbb{N}_0}$ adapts to $\mathbb{A}$, if X_n is $\mathcal{A}_n$-measurable for all $n \geq 0$. (secondly) The *natural filtration* of a process $\{X_n\}_{n \in \mathbb{N}_0}$ is the smalles filtration $\{X_n\}_{n \in \mathbb{N}_0}$ adapts to and is denoted by $\mathbb{A}^X = \{A_n\}_{n \in \mathbb{N}_0}^X$.

In order to give a better understanding about filtrations, we can think back to the previous Example 2.8, the random walk on $\mathbb{Z}$.

Example 3.5

We intend to model a random walk on $\mathbb{Z}$. Let $\{X_n\}_{n\in\mathbb{N}_0}$ therefore take values in the measurable space $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$. Assuming we are interessted in modelling the probability of adopting state $z \in \mathbb{Z}$ at time point $n \in \mathbb{N}_0$. A possible filtration would be

$$\mathcal{A}_n := \sigma(\mathcal{P}(\{-z, \dots, z\})).$$

Definition 3.6 (Markov chain in a continuous state space)
Let $(\Omega, \mathcal{F}, \{\mathcal{A}_n\}_{n\in\mathbb{N}_0}, P)$ be a filtered probability space and let Q be a Markov kernel on a measurable space $(E, \mathcal{E})$. A stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ with $X_0 : \Omega \longrightarrow E$ is said to be a Markov chain *under P, with respect to the filtration $\mathbb{A} = \{\mathcal{A}_n\}_{n\in\mathbb{N}_0}$ and transition kernel Q, if it is adapted to $\mathbb{A}$ for all $n \in \mathbb{N}_0$ and $A \in \mathcal{E}$, with*

$$P(X_{n+1} \in A|\mathcal{A}_n) = Q(X_n, A). \tag{9}$$

X_0 is called the initial distribution of the chain.

As already mentioned, $\mathbb{A} = \{\mathcal{A}_n\}_{n\in\mathbb{N}_0}$ represents the history of the (homogenous) stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$. One can see from (9) that the transition to the next time point only relies on X_n. In analogy to Definition 2.2 the transition of a Markov chain in a continuous state space only depends on the present and not on the past.

Example 3.7 (Random walk with normal increment)
Consider the following continuous state space Markov chain. We set $X_0 = 0$ and

$$X_{n+1} = X_n + \epsilon_n,$$

for $n \geq 1$, where $\{\epsilon_n\}_{n\in\mathbb{N}}$ are independent and identically normal distributed random variables with mean 0 and variance σ^2. The Markov kernel of this chain is given by

$$Q(x_n, A) = \int_A \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(y - x_n)^2}{2\sigma^2}\right\} dy.$$

In the recursive definition of the chain, the increment from one step to the next is given by a random number drawn from a normal distribution $N(0, \sigma^2)$.

Example 3.8
Let us now consider a stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ on a probability space $(\Omega, \mathcal{F}, P)$ that takes values in $([0,1], \sigma(\mathcal{O}_{[0,1]}))$, where $\sigma(\mathcal{O}_{[0,1]})$ denotes the Borel σ-algebra generated by open sets on the interval $[0,1]$. We set $X_0 = 1$ and define a Markov kernel as

$$Q(x_n, A) = \int_A \frac{\lambda(A)}{\lambda([0,x_n])} d\lambda(x) = \frac{\lambda(A)}{\lambda([0,x_n])} = \frac{\lambda(A)}{x_n},$$

for $A \subseteq [0, x_n]$ and λ the lebesgue measure.

In a first step we draw a random number from a continuous uniform distribution on the interval $[0,1]$. This number will then be the upper bound of the intervall from wich the number of the next step is drawn, again from a continuous uniform distribution on that new interval.

4 Outlook

Starting from stochastic processes, we have defined Markov chains in discrete and continuous state spaces. However, we have resticted ourselves to homogenous discrete time Markov chains of first order. There are several ways to further extend such models. In the following we will discuss three possible extensions:

1. From discrete time to continuous time: such processes $\{X_t\}_{t\geq 0}, t \in \mathbb{R}^+$ are defined by a (countable) state space $(E, \mathcal{E})$ and instead of a transition matrix Π a *transition rate matrix* Ψ. Each $\psi_{ij} \in \mathbb{R}_0^+$ is non-negative and gives the rate of transiting from state i to state j, while $\psi_{ii} \in \mathbb{R}$ are chosen such that the rowsums are equal to 0.

2. From first order to n$^{\text{th}}$ - order: in such processes the probability of observing some state is conditioned on $n \in \mathbb{N}$ previously observed states. The transition kernel then needs to be extended to a function $Q : E^n \times \mathcal{E}' \longrightarrow [0, \infty)$, that satisfies

   i. for all $(x_1, \ldots, x_n) \in E^n, Q((x_1, \ldots, x_n), \cdot)$ is a positive measure on $(E', \mathcal{E}')$

   ii. for all $A \in \mathcal{E}'$, the function $(x_1, \ldots, x_n) \mapsto Q((x_1, \ldots, x_n), A)$ is measurable.

3. Convergence of Markov chains: in Example 3.8 the interval $[0, x_n], n \in \mathbb{N}_0$ from which the upper bound of the next interval is drawn, becomes narrower with every step in time. One could hypothesize, that the interval converges to 0.

5 References

[1] Breiman L (1992), *Probability*, Society for Industrial and Applied Mathematics, Philadelphia.

[2] Brémaud P (1999), *Markov chains. Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer, New York.

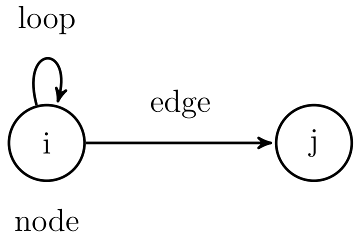
[3] Cappé O, Moulines E & Rydén T (2005), *Inference in Hidden Markov Models*, Springer, New York.

[4] CD274 - NC_000009.12 - GRCh38.p13 Primary Assembly (2004), Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/nuccore/568815589?report=genbank&from=5447492&to=5473576, (accessed July 20th, 2020).

[5] Strick HK (2016), *Der Mathematische Monatskalender: Andrei Markow (1856–1922)*, Spektrum der Wissenschaft, Leverkusen, https://www.spektrum.de/wissen/andrei-markow-1856-1922/1411967, (accessed July 20th, 2020).

6 Appendix

A.1 Definition (Transition Graph)
The graph $G = (\mathcal{N}, \mathcal{K})$ is determined by a countable set of *nodes* $\mathcal{N}$ and a set of directed *edges* $\mathcal{K} \subseteq \mathcal{N} \times \mathcal{N}$. For $k = (i, \{j\}) \in \mathcal{K}$ there is an edge pointing from i to j. If $k = (i, \{i\})$ then there is a *loop* in i.



A.2 Definition
If $\{X_n\}_{n\in\mathbb{N}_0}$ is a homogenous Markov chain with countable state space $(E, \mathcal{E})$ and transition matrix Π, its *transition graph* is defined by $G = (\mathcal{N}, \mathcal{K})$ with

$$\mathcal{N} = E \quad and \quad \mathcal{K} = \{(i, \{j\}) \in E \times \mathcal{E} : \Pi(i, \{j\}) > 0\}.$$

A.3 Proposition (Radon-Nikodym-Theorem)
Let $(\Omega, \mathcal{F})$ be a measurable space and μ, ν two measures on $\mathcal{F}$. If $\nu << \mu$, then there exists a non-negative $\mathcal{F}$-measurable function X determined up to equivalence, such that for any $A \in \mathcal{F}$,

$$\nu(A) = \int_A X d\mu.$$

This means that the Radon derivative of ν with respect to μ exists and is equal to X. That is

$$\frac{d\nu}{d\mu} = X.$$

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Munich, 27th of July 2020 Lorenz Leopold Mihatsch