Introducing the basics of Markov Chains

Alexander Ritz

30. Juli 2020

1 Preface

The following introduction is meant to be a very basic overview of the most relevant theoretical concepts utilised when considering MCMC methods, i.e. Markov chains are introduced as a special case of stochastic processes and vital properties of theirs are discussed. This text ignores the intricacies of Markov chains with continuous state spaces, since the discrete case is sufficient for a basic understanding of Markov chains and for the sake of succinctness. Interestingly the text could still claim to provide a full introduction, since the vast body of literature concerned with Markov chains offers differing views on what constitutes such a chain. Often processes with continuous state spaces are excluded from the class of Markov chains anyways, with processes in continuous time being routinely excluded as well [1]. Again other works do not restrict the eligible processes at all. Suggestions, corrections and questions regarding this script can be addressed at alexander.ritz@stud.uni-goettingen.de

---

[1]Giving rise to the notion of Markov processes in the case of continuous time.

2 Stochastic processes

If one were to observe some arbitrary random quantity over the course of time, it would only seem natural to model this by constructing a sequence of random variables, stretched out over a set of time „points". Let T be this non-empty time set of real numbers and $(\Omega, \mathscr{A}, P)$ a probability space. Then a *stochastic process* is a family $(Z_t)_{t \in T}$ of random variables sharing the probability space $(\Omega, \mathscr{A}, P)$. In the case of Z_t assuming values in $\mathbb{R}^d$, we speak of a stochastic process of dimension d. For a given $\omega \in \Omega$ the function

$$Z_\bullet(\omega) : T \to \mathbb{R}^d \text{ with } t \mapsto Z_t(\omega)$$

is called path of the stochastic process. Intuitively, there are three types of index sets T to distinguish:

- $T = \{t_0, t_1, \dots, t_N\}$ with $N \in \mathbb{N}$ and $t_0 < t_1 < \cdots < t_N$. In this case, one is looking at a process in discrete time with finite time horizon. [2]

- $T = \{t_i \mid i \in \mathbb{N}_0\}$ with $t_i \in \mathbb{R}$ and $t_i < t_{i+1}$ for all $i \in \mathbb{N}$, as well as $\lim_{i \to \infty} t_i = \infty$. In this case, one is looking at a process in discrete time with infinite time horizon. [3]

- For T being an interval of $\mathbb{R}$ (most reasonably $T = [0, T_n]$ with $T_n \leq +\infty$), one is looking at a process in continuous time.

The properties of stochastic processes are often described analogously to the characterisation of random variables, i.e. expected value, variance, covariance, distribution etc.. In contrast to the „simple "case of random variables, processes pass on their time dependence to these quantities. Without involving ourselves too much with the details, the following list offers a useful overview:

- For $t_1, \dots t_n \in T$, $F_{t_1,\dots,t_n}$ denotes the *joint distribution function of the random variables* $Z_{t_1,\dots,t_n}$. Therefore it holds that

$$F_{t_1,\dots,t_n}(x_1, \dots, x_n) := P(\{Z_{t_1} \leq x_1, \dots, Z_{t_n} \leq x_n\})$$

The set $\{F_{Z_{t_1},\dots,Z_{t_k}} \mid t_i \in T, k \in \mathbb{N})\}$ is called *finite-dimensional distribution* of the process $(Z_t)_{t \in T}$.

---

[2]Due to the bijectivity of $l \mapsto t_l$ we can assume $T = \{0, 1, \dots, N\}$ w.l.o.g.

[3]Due to the bijectivity of $l \mapsto t_l$ we can assume $T = \mathbb{N}_0$ w.l.o.g.

- The expected value function of $(Z_t)_{t \in T}$ is given by

$$t \mapsto \mathbb{E}\left[Z_t\right]$$

, with the straightforward interpretation as a time dependant expected value.

- The relationships between the random variables of the stochastic process $(Z_t)_{t \in T}$ is described by the variance or autocovariance function given by

$$t \mapsto \operatorname{Var}(Z_t) \text{ and } t, h \mapsto \operatorname{Cov}(Z_t, Z_{t-h})$$

3 Markov chains

A Markov chain is simpy a stochastic process $(Z_t)_{t \in T}$ for which the eponymous *Markov*[4] *property* holds. In order to avoid the concept of *filtrations*[5], we define the Markov property only for processes with finite or countable state spaces Ω. In order to make our continued efforts even more bearable, we will restrict ourselves to discrete time as well. Given these restrictions, the Markov property is then given by:

$$P(Z_{t+1} = x \mid Z_t = c_t, Z_{t-1} = c_{t-1}, \ldots, Z_0 = c_0) = P(Z_{t+1} = x \mid Z_t = c_t)$$

This property is often also described as *memorylessness*. In simplifying terms, one could say that a Markov chains future$(t+1)$ value only depends upon its past$(t-i)$ values through the present(t) value. A trivial example of a process fulfilling this property would be the repeated drawing with replacement of a ball from a bag containing a countable number of colored balls, each time denoting the drawn color as the current realisation. If one were to omit the replacement, the process would cease to be a Markov chain however, since the probabilities of different outcomes of future draws then depend on not only the last draw, but every single one that came before.

Notationally, the probability $P(Z_{t+1} = j \mid Z_t = i)$ is often denoted by p_{ij}, since it describes the probability of the chain transitioning from state i at time t to

---

[4]Named after the Russian mathematician Andrei Andrejewitsch Markow, whose surname is usually transcribed as Markov, in German texts formerly also as Markoff.
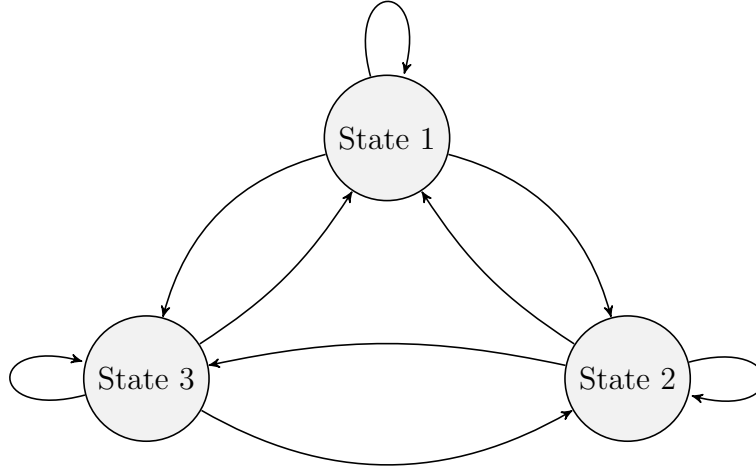
[5]The appendix gives a short definition of filtrations.

Abbildung 1: Example of a Markov chain depicted as a directed graph.

state j in $t+1$. Furthermore, this notation provides an elegant way of defining a Markov chain through utilisation of an $N \times N$ *transition matrix*[6]. A transition matrix $\mathbf{P}$ contains the probabilities of transitioning between states at time t. The depicted chain in figure 1 possesses the transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

. The probabilities have to satisfy two obvious properties:

$$p_{ij} \geq 0 \quad \text{and} \quad \forall i \in \{1, 2, \ldots, N\} : \quad \sum_j p_{ij} = 1$$

With N denoting the number of states. Statisticians with strong mathematical instincts[7] might now be curious about the distribution described by such a Markov chain specified by its transition matrix. This, of course, depends on what initial distribution of states is chosen, a starting point for the chain so to speak. This initial distribution π_0 (a row vector of N probabilites such that $\pi_i \geq 0$ and $\sum_i \pi_i = 1$) then fully defines all further distributions, through the *evolution equations*. With π_t denoting the marginal distribution of Z_t, with

---

[6]Also often called *stochastic matrix*.
[7]i.e. competent ones.

$\pi_t(i) = P(Z_t = i)$, the mentioned equation holds

$$\pi_t = \pi_0 \mathbf{P}^t$$

. With $\mathbf{P}^t$ offering the transition probabilites like so

$$(\mathbf{P}^t)_{ij} = P(Z_t = j \mid Z_0 = i)$$

. A central concept in the application of Markov chains is the *stationary distribution*. A row vector π of familiar form is called stationary distribution of the relevant Markov chain if

$$\pi = \pi \mathbf{P}$$

, honoring its name in an obvious manner. The term *invariant measure* is often used synonymously. This condition can be rewritten as

$$\forall j \in \{1, 2, \ldots, N\}: \quad \pi_j = \sum_i \pi_i p_{ij}$$

In large parts of the literature these conditions are known as the *global balance equations*. The most tempting remaining questions now involve the existence and uniqueness of such a stationary distribution for a given Markov chain. In order to answer these, the following text will define several important properties a Markov chain might fulfill and how some of these relate to their stationary distributions. At first we will consider only chains with finite state spaces.

- A state i is called *recurrent*, if $P(T_i < \infty \mid Z_0 = i) = 1$, with $T_i = \inf(t \geq 1 : Z_t = i)$ denoting the first time state i is realised. inf being the *infimum*, i.e. the greatest lower bound. This simply describes the process almost surely returning to state i after initially starting there. If a state is not recurrent, it is *transient*. Transcience therefore describes the case of positive probability for the event of never returning to a given state.

- A state i is said to have period k if the return to this state can only occur in multiples of k time steps, which can be written as

$$k = gcd\{n > 0 : P(Z_n = i \mid Z_o = i) > 0\}$$

, as long as this set is not empty. *gcd* denoting the *greatest common divisor*. In the case of $k = 1$ the state is called aperiodic.

- A state j is called *accessible* from state i, tentatively written as $i \to j$ if $\exists t \geq 0 : \mathbf{P}_{ij}^t > 0$. States i and j are furthermore said to *communicate* if $i \to j$ and $j \to i$ hold. And building on these properties a Markov chain is called *irreducible* if all of its states communicate.

- A state is called *ergodic* if it is recurrent and aperiodic, and the whole chain is called ergodic if all of its states are.

When considering an irreducible finite Markov chain, all its states are recurrent and ergodicity as well as periodicity are global properties, in the sense that they are shared among all states. Wielding this new knowledge delivers the useful insight that for any ergodic, finite Markov chain one may construct its unique stationary distribution like so:

$$\lim_{t \to \infty} P(Z_t = j \mid Z_0 = i) = \lim_{t \to \infty} \mathbf{P}_{ij}^t =: \pi_j^\infty$$

Only minimal changes to this conclusion arise in the case of a countably infinite state space. The property of recurrence is then extended in a natural manner:

- A recurrent state i is called *positive recurrent*, if the time to return to the state possesses a finite mean, i.e. $\mathbb{E}(T_i \mid Z_o = i) < \infty$. Otherwise it is considered *null recurrent*.

Ergodicity of a state is then in turn defined as reliant on the positive reccurrence of this state, and aperiodicity as before. In an irreducible Markov chain the properties of positive and null recurrence and periodicity as well as ergodicity are now global. Lastly we can conclude that in this case an irreducible chain possesses a stationary distribution if and only if all of its states are positive recurrent. This distribution is indeed also unique.

A   Filtrations

A family $(\mathscr{I}_t)_{t \in T}$ of σ-algebras in Ω is called *filtration* in σ-algebra $\mathscr{A}$ if

$$\forall t \in T : \quad \mathscr{I}_t \subset \mathscr{A}$$

and
$$\forall t_1, t_2 \in T \text{ with } t_1 \leq t_2 : \mathscr{I}_{t_1} \subset \mathscr{I}_{t_2}$$

. The quadruple $(\Omega, \mathscr{A}, P, (\mathscr{I}_t)_{t \in T})$ is called filtered probability space.