Basics about Markov Chains

We consider processes on a finite set Ω. Each element $\omega \in \Omega$ is called a *state* and Ω is called *state space*. Let us denote the set of all probability distributions on the measurable space $\big(\Omega, \mathscr{P}(\Omega)\big)$ by $\mathscr{M}_1(\Omega)$, whereby the σ-algebra $\mathscr{P}(\Omega)$ represents the power set on Ω. In our framework, we examine discrete-time Markov chains on a finite state space Ω. Such a process moves along the elements of Ω such that the probability of the next state only depends of the current state. Hence, we define:

Definition 1.1. *A sequence of random variables* $(X_t)_{t\geq 0}$ *is called a* discrete-time Markov chain with finite state space Ω *if for all* $\omega, \omega' \in \Omega$*, all* $t \geq 1$*, and all events* $E_{t-1} := \bigcap_{i=0}^{t-1}\{X_i = \omega^{(i)}\}$ *that satisfy* $\mathbb{P}(E_{t-1} \cap \{X_t = \omega\}) > 0$*, we have*

$$\mathbb{P}(X_{t+1} = \omega' \mid E_{t-1} \cap \{X_t = \omega\}) = \mathbb{P}(X_{t+1} = \omega' \mid X_t = \omega)\,.$$

Since the probability of moving from ω to ω' does not depend on the past, i.e. on the sequence $\omega^{(0)}, \dots, \omega^{(t-1)}$, an $|\Omega| \times |\Omega|$ matrix P suffices to describe the transitions of the process:

Definition 1.2. *Let* $(X_t)_{t\geq 0}$ *be a discrete-time Markov chain with finite state space* Ω*. Then the matrix* $P \in \mathbb{R}^{|\Omega|\times|\Omega|}$ *with entries*

$$P(\omega, \omega') := \mathbb{P}(X_{t+1} = \omega' \mid X_t = \omega)$$

is called transition matrix *and its entries are called* transition probabilities*.*

Due to the simplicity of such a process, we can compute the distribution of the Markov chain at each time $t \geq 1$ by simple matrix multiplication if we know the initial distribution $\mu_0 \in \mathscr{M}_1(\Omega)$, as stated in the following theorem:

Theorem 1.3. *Let* $(X_t)_{t\geq 0}$ *be a discrete-time Markov chain with finite state space* Ω *and transition matrix* P*. Let* X_t *be distributed according to* $\mu_t \in \mathscr{M}_1(\Omega)$ *for all* $t \geq 0$*. Then*

$$\mu_t = \mu_0 P^t \quad \textit{for } t \geq 0\,.$$

For further contemplations we introduce two important properties of Markov chains, namely *irreducibility* and *aperiodicity*. Definitions are given below and some additional explanations will be given in the presentation.

Definition 1.4. *A transition matrix* P *of a Markov chain is called* irreducible *if for all* $\omega, \omega' \in \Omega$ *there exists* $t \in \mathbb{N}$ *such that* $P^t(\omega, \omega') > 0$*.*

Definition 1.5. *An irreducible Markov chain* $(X_t)_{t\geq 0}$ *with state space* Ω *is called* aperiodic *if* $\gcd\{t \in \mathbb{N} \mid P^t(\omega, \omega) > 0\} = 1$ *for all* $\omega \in \Omega$*. Otherwise, the chain will be called* periodic.

For our purpose, we also are interested in so-called *stationary distributions* that have an interesting property:

Definition 1.6. *Let $(X_t)_{t\geq 0}$ be a Markov chain with state space Ω and transition matrix P. Then $\pi \in \mathscr{M}_1(\Omega)$ is called* stationary distribution *for P if*

$$\pi = \pi P \,.$$

If we start a Markov chain in a stationary distribution π, then we have $\mu_t = \pi P^t = \pi$ for all $t \geq 0$ according to Theorem 1.3. Therefore, we call π stationary. The following proposition provides a criteria for a given distribution to be stationary:

Proposition 1.7. *Let $(X_t)_{t\geq 0}$ be an irreducible Markov chain with state space Ω and transition matrix P. If $\pi \in \mathscr{M}_1(\Omega)$ satisfies the* detailed balance equations

$$\pi(\omega)P(\omega,\omega') = \pi(\omega')P(\omega',\omega) \quad \textit{for all} \quad \omega,\omega' \in \Omega \,,$$

then π is stationary for P.

Definition 1.8. *A Markov chain $(X_t)_{t\geq 0}$ with state space Ω and transition matrix P that satisfies the detailed balance equations is called* reversible *with respect to $\pi \in \mathscr{M}_1(\Omega)$.*

In order to talk about convergence of measures, we will define a distance between two probability measures. This distance will be based on the *total variation distance*:

Definition 1.9. *Let $\mu, \nu \in \mathscr{M}_1(\Omega)$. Then the* total variation distance *between μ and ν is defined by*

$$\|\mu - \nu\|_{\mathrm{TV}} := \max_{A \subset \Omega} |\mu(A) - \nu(A)| \,.$$

The above definitions are primarily taken from the book “Markov Chains and Mixing Times” by Levin et al. (2009). For further explanations, see chapter 1 of this book.