MARKOV MODELS

Nicole Bäuerle
Department of Optimization and Operations Research, University of Ulm, Germany.

Keywords: discrete-time Markov chains, continuous-time Markov chains, recurrence, transience, limit distribution, stationary distribution, Chapman-Kolmogorov equation, solidarity property, embedded Markov chain, intensity matrix.

Contents

Summary

We consider Markov chains in discrete and continuous time with a countable or finite state space. Markov chains are stochastic processes, which are characterized by the property that the evolution in the future depends only on the current state, not on the history. They constitute mathematical models for real-world processes in various fields like e.g. biology, economics, engineering, informational sciences, manufacturing, physics and telecommunication. The aim is to investigate the long run behavior of these processes. In a first step, we classify the system states according to their return behavior. We distinguish between transient, null recurrent and positive recurrent states. The latter ones are characterized by a finite expected return time. In a second step, we investigate the existence and computation of stationary distributions, which is a prerequisite for the existence of limit distributions. It will turn out that an irreducible Markov chain has a stationary distribution if and only if it is positive recurrent. The stationary distribution can be computed by solving a simple system of linear equations. Finally, we show that limit distributions exist, if and only if stationary distributions exist and if the Markov chain in discrete time is also aperiodic. In this case the stationary distribution is unique and is equal to the limit distribution.

1. Introduction

A sequence of independent random variables is seldom a useful stochastic model for real-world applications. In reality, we frequently encounter dependencies. For example, the size of a population in one year depends on its current size. The price of a certain stock in one week depends on its current value. The number of connected calls in a telephone network in 5 minutes depends on the current number. Thus, there is the need for a mathematical model that allows dependencies on the one hand and should not be too complicated to analyze on the other hand. Markov models are best suited to balance this trade-off. In a Markov model, the stochastic evolution is allowed to depend on the current state of the system. This is sufficient in many cases. Moreover, there is a well-developed mathematical theory about the behavior of such systems, in particular as far as the long run behavior is concerned.

One of the oldest Markov models is the queueing systems model (see *Queuing Systems*). More recent models are found in the area of finance and investment (see *Investment Models*). Typical applications can be found for example in biology, economics, engineering, informational sciences, manufacturing, physics and telecommunication.

In what follows we will distinguish Markov models which evolve in discrete and in continuous time. The first section is devoted to discrete-time processes, which are called Markov chains; the second section deals with continuous-time Markov chains. Throughout we assume that the state space is finite or countable.

2. Discrete-time Markov Chains

We will denote by S the state space of the Markov chain. S is assumed to be finite or countable and it will often be a subset of the natural numbers $\mathbb{N}$ or of the integers $\mathbb{Z}$. Let (X_n) be a sequence of random variables, taking values in S. Our interpretation is that X_n gives the random state of our system at time n.

2.1 Definition and first Properties

The sequence (X_n) is called a *discrete-time Markov chain*, if it satisfies the following property:

for all states $i, j, i_k \in S$

$$\begin{aligned}&\mathbb{P}(X_{n+1} = j \mid X_0 = i_0, ..., X_{n-1} = i_{n-1}, X_n = i)\\&= \mathbb{P}(X_{n+1} = j \mid X_n = i) = p_{ij}\end{aligned} \tag{1}$$

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

Thus, the distribution of the state tomorrow depends only on the current state and not on the whole history. This is the so-called *Markov property*. p_{ij} is the *transition probability* from state i to state j. Throughout this section we will assume that the transition probabilities do not depend on the time, the transition takes place. Sometimes a Markov chain with this property is called *homogeneous*. The matrix $P = (p_{ij})$ is called *transition matrix*. The transition probabilities certainly satisfy for all $i, j \in S$

$$p_{ij} \geq 0, \qquad \sum_{j \in S} p_{ij} = 1. \tag{2}$$

For example, every row of P is a distribution on the state space. A matrix P with this property is called *stochastic matrix*. We will assume that the Markov chain has an initial distribution (p_i). That means, $\mathbb{P}(X_0 = i) = p_i$ for $i \in S$. An easy observation is that the finite dimensional distribution of the Markov chain (X_n) is completely determined by the transition matrix P and the initial distribution p. More precisely, we have that (X_n) is a Markov chain with transition matrix P and initial distribution p if and only if for $i_k \in S$

$$\mathbb{P}(X_0 = i_0, ..., X_n = i_n) = p_{i_0} p_{i_0 i_1} ... p_{i_{n-1} i_n}. \tag{3}$$

Thus, to describe the system, it is sufficient to know the initial distribution p and the transition matrix P.

A Markov chain can easily be constructed in the following way: suppose (Y_n) is a sequence of independent and identically distributed random variables with values in a set Z. g is a mapping $g : S \times Z \to S$. If we define the sequence (X_n) recursively by

$$X_n := g(X_{n-1}, Y_n), \ n = 1, 2, ... \tag{4}$$

then (X_n) is a Markov chain.

An important role in the analysis of Markov chains play the matrix powers of P. We define $P^0 = I$, where I is the identity matrix and denote the elements of the matrix P^n by $p_{ij}^{(n)}$. It is important to realize that the matrix P^n is again a stochastic matrix and $p_{ij}^{(n)}$ exactly gives the probability of getting from state i to state j in n steps, i.e.

$$\mathbb{P}(X_{m+n} = j \mid X_m = i) = p_{ij}^{(n)} \quad \text{for all} \quad m \geq 0 \tag{5}$$

and if p is the initial distribution

$$\mathbb{P}(X_n = j) = \sum_{k \in S} p_i p_{ij}^{(n)}. \tag{6}$$

The probabilities $p_{ij}^{(n)}$ are therefore called *n-step transition probabilities*. It follows from $P^{n+m} = P^n \cdot P^m$ in particular that

$$p_{ij}^{n+m} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)}, \ \text{ for } i, j \in S. \tag{7}$$

This is the so-called *Chapman-Kolmogorov equation*. It states that the probability of getting from state i to state j in $n + m$ steps equals the sum over the probabilities of getting from state i in n steps to a state k and then from k in m steps to state j, where the sum is taken over all possible intermediate states k.

2.2 Examples

Typical examples of Markov chains are the following.

Example 2.1 (Random Walk)

Suppose that (Y_n) is a sequence of independent and identically distributed random variables with distribution $\mathbb{P}(Y_n = 1) = p$, $\mathbb{P}(Y_n = -1) = 1 - p$, and $p \in (0, 1)$. We assume that $X_0 = 0$ and for $n = 1, 2, ...$

$$X_n := \sum_{k=1}^{n} Y_k = X_{n-1} + Y_n. \tag{8}$$

According to our construction recipe, (X_n) is obviously a Markov chain on the state space $S = \mathbb{Z}$ with $g(i, y) = i + y$. The transition probabilities are given by $p_{ii+1} = p$ and $p_{ii-1} = 1 - p$.

Example 2.2 (Gambler's Ruin)

Two players with initial fortune i and $M - i$ start playing a sequence of coin tossing games. The probability that player 1 wins on a toss is $p \in (0, 1)$ and in this case player 2 has to pay player 1 one unit. $q := 1 - p$ is the probability to win for the second player, in which case player 2 is paid by player 1. If the process enters state 0, then player 1 is ruined. If the process enters state M, then player 2 is ruined. Let us denote by X_n the fortune of player 1 after n coin tosses. It is easy to see that X_n is a Markov chain on the state space $S = \{0, 1, ..., M\}$ with transition probabilities $p_{00} = p_{MM} = 1$ and for $i \neq 0$ and $i \neq M$, $p_{ii+1} = p$ and $p_{ii-1} = q$.

Example 2.3 (Inventory Model)

Let (X_n) be the inventory level of a certain item at time n. We take $S = \mathbb{N}$ and assume that the total demand in period $(n - 1, n]$ is given by a random variable D_n with values in $\mathbb{N}$. The sequence (D_n) should consist of independent, identically distributed random variables. A commonly used restocking policy consists of two values $0 < s < S$. If at time n the inventory is below s, we fill up the stock level to S. In all other cases no replenishment is undertaken. Thus, we obtain for $n = 1, 2, ...$, with given initial inventory $X_0 \in \mathbb{N}$

$$X_n = \begin{cases} (S - D_n)^+, & \text{if } X_{n-1} < s \\ (X_{n-1} - D_n)^+, & \text{if } X_{n-1} \geq s \end{cases} \tag{9}$$

where $x^+ = \max(x, 0)$ denotes the positive part of x. (X_n) is a Markov chain and the transition probabilities are given by

$$p_{ij} = \begin{cases} \mathbb{P}((S - D_n)^+ = j), & \text{if } i < s \\ \mathbb{P}((X_{n-1} - D_n)^+ = j), & \text{if } i \geq s \end{cases} \tag{10}$$

Example 2.4 (Discrete Queueing Model)

Suppose we have a queue with a single server. Arriving jobs are served according to first-come first-served. During the service of the n-th job, a random number of A_n jobs arrive. Again we assume that the sequence (A_n) consists of independent and identically distributed random variables with distribution $\mathbb{P}(A_n = i) = q_i \geq 0$, $i = 0,1, ...,$ and $X_0 \in \mathbb{N}_0$. Hence the state space is given by $S = \mathbb{N}_0$ and

$$X_n = (X_{n-1} - 1)^+ + A_n. \tag{11}$$

(X_n) is a Markov chain and the transition matrix P has the form

$$P = \begin{pmatrix} q_0 & q_1 & q_2 \cdots \\ q_0 & q_1 & q_2 \cdots \\ 0 & q_0 & q_1 \cdots \\ 0 & 0 & q_1 \cdots \\ \vdots & & \vdots \;\; \ddots \end{pmatrix} \tag{12}$$

2.3 Classification of States and Solidarity Properties

A first step in the analysis of Markov chains is to determine the possible paths through the state space. As we will see, not every arbitrary movement is possible. A reasonable question is: is it possible to get from state i to state j and then back to state i again? Two states with this property are said to communicate. More precisely, states i and j *communicate*, if there exist natural numbers n, m such that $p_{ij}^{(n)}\ p_{ji}^{(m)} > 0$. Moreover, we are now able to divide the state space into disjoint subsets of states that communicate with each other. It is easy to see that the relation ~ defined by $i \sim j$ if and only if $i = j$ or i and j communicate, is an equivalence relation. Thus, the relation induces a dissection of the state space in different *classes*. A Markov chain is called *irreducible* if the state space S consists of only one class, i.e. all states communicate with each other.

Another important question is: starting in state i, at which time points is it possible to return to state i?

We define by

$$d_i := \gcd\{n \in \mathbb{N} \mid p_{ii}^{(n)} > 0\} \tag{13}$$

the *period* of state $i \in S$, where *gcd* is the usual greatest common divisor. We set *gcd* $\varnothing$:= 0. If $d_i = 0$ or 1, then state i is called *aperiodic*.

Hence, starting in state i, it is only possible to return to state i at time points nd_i, $n = 1, 2, \ldots$. Finally, the crucial question is: starting in state i do we return to a state i with probability 1 or less? Let us denote by $\tau_i := \inf\{n \in \mathbb{N} \mid X_n = i\}$ the first entrance time into state i. In order to simplify the notation, we write $\mathbb{P}_i(A) = \mathbb{P}(A \mid X_0 = i)$. A state i is called *recurrent*, if $\mathbb{P}_i(\tau_i < \infty) = 1$ and *transient*, if $\mathbb{P}_i(\tau_i < \infty) < 1$. An important fact is that whenever two states i and j communicate they are both either recurrent or transient and have the same period. Thus, speaking in terms of classes: within one class of states, say $C \subset S$, all states $i \in C$ are either recurrent or transient and all have the same period. These kind of properties are called *solidarity properties*. Therefore, we will also say that a class is recurrent or transient or the whole chain, provided it is irreducible. Next, it will be important to characterize recurrence and transience of states. Sometimes the following criteria are useful:

State i is recurrent if and only if $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$. Hence, state i is transient if and only if the sum is finite.

In particular when the state space S is finite, it holds that class $C \subset S$ is recurrent, if and only if $(p_{ij})_{i,j \in C}$ is a stochastic matrix. Hence, class C is transient if and only if this matrix is not stochastic.

Example 2.5 (Random Walk)

Let us look at the random walk of Example 2.1 with $p \in (0, 1)$. Obviously all states communicate, thus the Markov chain is irreducible. The period of the chain is two. We will use the preceding criterion to determine whether the Markov chain is recurrent or transient. It is sufficient to look at state 0. We get that

$$\begin{aligned} p_{00}^{(2n)} &= \mathbb{P}\left(n \text{ steps to the right, } n \text{ steps to the left}\right) \\ &= \binom{2n}{n} p^n (1-p)^n \end{aligned} \tag{14}$$

Using Stirling's formula to approximate $n!$, we derive that the Markov chain is recurrent for $p = \frac{1}{2}$ and transient for $p \neq \frac{1}{2}$.

-
-
-

Bibliography

Anderson W.J. (1991). *Continuous-Time Markov Chains* 335 pp. New York: Springer-Verlag. [Presents a comprehensive treatment of continuous-time Markov chains.]

Brémaud P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, 444 pp. New York: Springer-Verlag. [Includes an introduction to Monte Carlo simulation]

Davis M.H.A. (1993). *Markov Models and Optimization*, 295 pp. London: Chapman & Hall. [This book gives an introduction to piecewise deterministic Markov processes.]

Meyn S.P. and Tweedie R.L. (1993). *Markov Chains and Stochastic Stability*, 548 pp. London: Springer-Verlag. [This book provides the state of the art of Borel state Markov chains.]

Resnick S.I. (1992). *Adventures in Stochastic Processes*, 626 pp. Boston: Birkhäuser. [Excellent text book with many examples.]

Ross S.M. (1997). *An Introduction to Probability Models*, 669 pp. San Diego: Academic Press. [Elementary textbook on Markov models, including chapters on reliability and simulation.]

Stewart W.J. (ed.). (1991). *Numerical Solution of Markov Chains. Probability Pure and Applied: 8*, 704 pp. New York: Marcel Dekker. [Collection of 40 papers on numerical aspects of Markov chains.]

Biographical Sketch

Nicole Bäuerle was born in 1968. She received the M.S. degree in mathematics & economics in 1992 , the Ph.D. degree in mathematics in 1994 and the Habilitation in mathematics in 1999, all from the University of Ulm. Since 2001 she has been Associate Professor at the University of Ulm. Her research interests include the analysis and control of stochastic processes with applications in telecommunication, computer science, finance and insurance. Dr. Bäuerle received two awards in 1996 for her Ph.D. thesis: one from the German special interest group in stochastics and the other one from the Ulmer Universitätsgesellschaft. She is currently Editor of the journal *Stochastic Models*.