MARKOV PROCESSES

K. Grill
Institut für Statistik und Wahrscheinlichkeitstheorie, TU Wien, Austria

Keywords: Markov process, Markov chain, Markov property, stopping times, strong Markov property, transition matrix, infinitesimal generator, recurrence, transience, transition operator, infinitesimal operator, Feller process, random walk, birth and death process, Galton-Watson process, diffusion process

Contents

Summary

A Markov process is characterized by the property that, given the value of the process at a specified time, the past and the future of the process are conditionally independent. The simplest examples of Markov processes are the ones for which either the set of possible values of the parameters or the set of possible states (or both) is at most countable, the so-called Markov chains. Their discussion leads to a number of basic concepts, which are then extended to the general case. In particular, this leads to the definition of the transition function and the transition operators of a Markov process, and to the infinitesimal operator. An important set of questions is studied next, namely how path properties of the process can be deduced from properties of the transition function or the infinitesimal operator. Finally, the results obtained are applied to a few examples.

1. Introduction

1.1. The Markov Property and the Transition Function

Markov process can be viewed as a generalization of the notion of independent random variables. They are characterized by the Markov property which states that the conditional distributions of the past and the future, given the present, are independent. In order to give a formal definition, let T be a subset of the real line (either a finite or infinite interval or the intersection of one of those with the integers; in the sequel, it will be assumed that the left endpoint of this interval is 0); T is called the parameter space of the process. Furthermore, let $(\xi(t), t \in T)$ be a stochastic process (i.e., a collection of random variables indexed by T) taking values in some subset X of a Euclidean space; usually, this space- the state space – will be the real line. One can consider more general state spaces, but this generates some additional (mostly technical) difficulties, and will be avoided in this exposition. The stochastic process $\xi(.)$ is called a Markov process if for all

$$s_1 < s_2 < \ldots < s_n < t < t_1 < \ldots < t_k \tag{1}$$

and

$$A_1, \ldots A_n, B_1, \ldots, B_k \in \mathcal{B} \tag{2}$$

it holds with probability one that

$$\mathbf{P}(\xi(s_i) \in A_i, i \le n, \xi(t_j) \in B_j, j \le k \mid \xi(t)) = \tag{3}$$

$$\mathbf{P}(\xi(s_i) \in A_i, i \le n \mid \xi(t)) \mathbf{P}(\xi(t_j) \in B_j, j \le k \mid \xi(t)) .$$

There is a number of equivalent definitions in the literature. All of them are just clever ways of saying that given the value of the process at time t, an event defined in terms of the values of the process before t and one that is defined in terms of the values of the process after t are conditionally independent. One that is very important because it is the simplest to verify is the following: for $s_1 < s_2 < \ldots < s_n < t$

$$\mathbf{P}(\xi(t) \in B \mid \xi(s_1), \ldots, \xi(s_n)) = \mathbf{P}(\xi(t) \in B \mid \xi(s_n)) . \tag{4}$$

If the parameter space is a subset of the integers, the Markov process is traditionally called a Markov chain A case of special interest is given by the discrete Markov chains; for these the state space, too, is at most countable, which make their analytic treatment fairly simple.

Another important special case is that of a Markov process with a continuous parameter (the parameter space is a whole interval) and a discrete (i.e., at most countable) state space. These processes are often called continuous parameter (or continuous time) Markov chains (or, putting emphasis on the discreteness of the state space, "discrete

Markov chains in continuous time”). The most important notion in the theory of Markov processes is the transition function of the process: a function

$$P(s,x,t,B) s,t \in T, s \le t, x \in X, B \in \mathcal{B} \tag{5}$$

is called the transition function of the Markov process $\xi(.)$ if the following conditions hold true:

1. For s,x,t fixed, the function $P(s,x,t,.)$ is a probability measure on $\mathcal{B}$.

2. For fixed s,t,B the functions $P(s,.,t,B)$ is measurable.

3. It holds that $P(s,x,s,B)=\delta_x(B)$, the measure that assigns mass 1 to the point x.

4. For any $s<t, B\in\mathcal{B}$, the following equation holds with probability one:

$$\mathbf{P}(\xi(t)\in B \mid \xi(s)) = P(s,\xi(s),t,B) . \tag{6}$$

The question of the existence of a transition function is somewhat hard, and for general Markov processes, a transition function need not exist. In the case considered here – the process taking values in some Euclidean space – there is always a transition function.

The transition function can be interpreted in the following way: $P(s,x,t,B)$ is the probability that the Markov process will be in the set B at time t, if it starts from x at time s. If this idea is followed a little further, one arrives at the notion of a family of Markov processes, or for short, a “Markov family”: this is just the collection of all Markov processes with a given set of transition functions, and started at different times and states. More precisely, suppose that for each $s\in T$ and $x \in X$, there is a probability measure $\mathbf{P}_{s,x}$ on the σ-algebra generated by the set of random variables $(\xi(t), t\in T\cap[s,\infty))$. The collection of probability measures $(\mathbf{P}_{s,x})$ is then called a Markov family with transition function $P(.,.,.,.)$ if the following three conditions are fulfilled:

1. The process $(\xi(t), t\in T\cap[s,\infty))$ is a Markov process with respect to the probability measure $(\mathbf{P}_{s,x})$.
2. This process has $P(.,.,.,.)$ as its transition function.
3. It starts in x at time s:

$$\mathbf{P}_{s,x}(\xi(s)=x)=1. \tag{7}$$

Now, there is the question of whether a given function $P(.,.,.,.)$ is the transition function of a Markov family. First, observe that the finite-dimensional distributions of the process can be calculated in terms of the transition function. In fact, for $s<t_1<\ldots<t_n$, it holds that

$$\mathbf{P}_{s,x}(\xi(t_1)\in A_1,...,\xi(t_n)\in A_n)= \tag{8}$$

$$\int_{A_{n-1}}\cdots\int_{A_1} P(s,x,t_1,dx_1)P(t_1,x_1,t_2,dx_2)\ldots$$

$$P(t_{n-2},x_{n-2},t_{n-1},dx_{n-1})P(t_{n-1},x_{n-1},t_n,A_n)\,.$$

Now, by Kolmogorov's extension theorem, one can construct a process with these finite-dimensional distributions if and only if they are consistent, i.e., if one adds the conditions

$$\xi(s_1)\in\mathbf{R},...,\xi(s_k)\in\mathbf{R} \tag{9}$$

in the probability on the left-hand side of (8), then the integral on the right-hand side must not change. This condition is equivalent to the Chapman-Kolmogorov equation

$$P(s,x,t,A)=\int P(s,x,u,dy)P(u,y,t,A)(s<u<t)\,. \tag{10}$$

If one considers only a single Markov process instead of a Markov family, the above argument gets just a little more involved; first, one needs to supply the distribution P_0 of $\xi(0)$. This enters formula (8) in the following way:

$$\mathbf{P}(\xi(t_1)\in A_1,...,\xi(t_n)\in A_n)= \tag{11}$$

$$\int_{A_{n-1}}\cdots\int_{A_1}\int_{\mathbf{R}} P_0(dx_0)P(0,x_0,t_1,dx_1)P(t_1,x_1,t_2,dx_2)\ldots P(t_{n-1},x_{n-1},t_n,A_n)\,.$$

The Chapman-Kolmogorov equations are still important – it is clear that they are sufficient for the existence of a Markov process with the given transition function. A necessary and sufficient condition is obtained by demanding that equation (10) is satisfied for almost all x (with respect to the distribution of $\xi(s)$). Of particular importance are those Markov processes whose transition function is only a function of the difference $t-s$. In other words, one has

$$P(s,x,t,A)=P(s+h,x,t+h,A)=P(t-s,x,A)\,, \tag{12}$$

which means that the dynamics of the process does not change if it shifted in time. Such a Markov process is called a homogenous Markov process, or a Markov process with stationary transition probabilities. This assumption is not necessarily a restriction; it is easily seen that from any Markov process one can obtain a homogeneous one by adding the parameter as an additional state variable. In Particular, if

$$\xi(t)=(\xi_1(t),\ldots,\xi_d(t)) \tag{13}$$

is a *d*-dimensional Markov process, then

$$\eta(t) = (\eta_1(t), \ldots \eta_{d+1}(t)) = (\xi_1(t), \ldots, \xi_d(t), t) \tag{14}$$

is a homogeneous Markov process with transition function

$$P(y_{d+1}, (y_1, \ldots, y_d), y_{d+1} + t - s, \{(x_1, \ldots, x_d) : (x_1, \ldots, x_d, y_{d+1} + t - s \in A\}) . \tag{15}$$

2. Discrete Markov Chains

Obviously, this depends on s and t only via $t-s$.

In the sequel, only homogeneous Markov processes will be considered.

The simplest examples of Markov processes are the Markov chains, both in discrete and continuous time. These will be studied next.

As stated above, these are Markov processes for which both the parameter space T and the state space X are discrete; without loss of generality one can assume that they are equal to the set of natural numbers (or, for finite state chains, X may be a set of the form $\{1, \ldots, n\}$, in which case the infinite matrices below reduce to ordinary square matrices). Furthermore, as stated above, the Markov chains studied here will be assumed to be homogeneous. The first observation one makes is that, since a discrete distribution is determined by the probabilities of the singletons, it is sufficient to define the transition function for those; let the transition probabilities be defined as

$$p_{ij}(t) = P(t, i, \{j\}). \tag{16}$$

These can be written as a matrix

$$P(t) = (p_{ij}(t))_{(X \times X)}, \tag{17}$$

the so-called transition matrix. He Using this, the Chapman-Kolmogorov equation can be written in the simple form

$$P(s+t) = P(s)P(t). \tag{18}$$

Denoting $P(1)$ simply by P, one obtains from this

$$P(t) = P^t. \tag{19}$$

In addition, if $p(t)$ denotes the row vector with entries $p_i(t) = \mathbf{P}(\xi(t) = i)$, then

$$p(s+t) = p(s)P(t). \tag{20}$$

Thus, the distribution of a discrete Markov chain is completely specified by its one-step transition Matrix P.

2.1. Classification of the State of a Markov Chain

A state i is said to be a predecessor of another state j if there is a $t > 0$ such that $p_{ij}(t) > 0$, i.e., if it is possible that the process visits j some time in the future, if it starts at i. If j is also a predecessor of i, then i and j are said to communicate. This obviously constitutes an equivalence relation between the states of the Markov chain, and the set of states can be split up into the corresponding equivalence classes. Many of the properties that will be studied later are the same for all states in the same equivalence class. This type of property will be called a “class property”.

A case of particular interest is that of a Markov chain whose states all communicate, or in other words, for which there is only one equivalence class. Such a chain is called irreducible. With many important questions, the general case can be reduced to a study of irreducible Markov chains.

One instance f a class property is periodicity. The period d of a state i is the greatest common divisor of the set of all n such that $p_{ii}(n) > 0$. Now, if j and i communicate, there are numbers a and b such that $p_{ij}(a)$ and $p_{ji}(b)$ are positive. This implies that

$$p_{ii}(a+b) \geq p_{ij}(a)\, p_{ji}(b) > 0\,, \qquad (21)$$

and if $p_{jj}(c) > 0$, then also

$$p_{ii}(a+b+c) \geq p_{ij}(a)\, p_{jj}(c)\, p_{ji}(b) > 0\,. \qquad (22)$$

By the definition of the period, this implies that the period d of i is a divisor of both $a+b$ and $a+b+c$, hence also of c. This holds for any c with $p_{jj}(c) > 0$, so the period of i is a divisor of the period of j, and by reversing the roles of i and j, one finds that both periods are equal.

If the period d of a class is different from zero (it can only be zero if it contains only one state i with $p_{ii} = 0$), then it is readily seen that the class can be divided into d subsets $S_0, \ldots, S_{d-1}$ such that p_{ij} is zero except for the case when $i \in S_k$ and $j \in S_{k \oplus 1}$ for some $k \in \{0, \ldots, d-1\}$, where $\oplus$ denotes addition modulo d. If the period of a class equals one, the class is called aperiodic.

-
-
-

Bibliography

Billingsley, P. (1968) *Statistical Inference for Markov Processes*. 160pp. Chicago: University of Chicago Press [This studies questions of estimation and testing for Markov processes depending on a parameter]

Dynkin, E.B. (1965) *Markov Processes* I (364pp) ,II(274) . Berlin: Springer [One of the fundamental reference works]

Itô, K., and Mckean, H.P.jr. (1974) *Diffusion Processes and their sample paths*. Berlin: Springer [A detailed discussion of diffusion processes, i.e., Markov processes generated by a second-order differential operator]

Karlin, S., and Taylor, H.M. (1974) *A First Course in Stochastic Processes*. 557pp. New York: Academic Press [This is an introduction to the theory of stochastic processes and, in particular, Markov processes]

Revuz, D. (1975) *Markov Chains*. 336pp. Amsterdam: North-Holland [This discusses Markov chains in discrete time with a general state space]

Biographical Sketch

Karl Grill received the PhD degree in 1983 from TU Wien. He joined TU Wien in 1982 as Assistant Professor in the Department of Statistics and became an Associate Professor in 1988. During 1991-92 he was a visiting professor at the Department of Statistics, University of Arizona, USA. In 1994 he received a six month NSERC Foreign Researcher Award at Carleton University, Ottawa, Canada