Chapter 1
Stochastic Processes

Contents

1 Stochastic Processes 1
- 1.1 Introduction 2
  - 1.1.1 Characterization of a stochastic process 3
- 1.2 Independent processes 3
- 1.3 Random walks 4
  - 1.3.1 Properties of random walks 6
  - 1.3.2 Stopping times 6
- 1.4 Markov Chains 8
  - 1.4.1 Transition graph 9
  - 1.4.2 State probabilities 10
  - 1.4.3 Stationary distribution 10
  - 1.4.4 Computing the stationary distribution 11
- 1.5 Stationary Processes 13
  - 1.5.1 Characterization of stochastic processes 13
  - 1.5.2 Stationarity 15
  - 1.5.3 Ergodicity 18
  - 1.5.4 Linear systems with stochastic inputs 21
  - 1.5.5 Gaussian processes 23
  - 1.5.6 Power Spectral Density 23
- References 26

1.4 Markov Chains

One important property of the simple random walk is that the effect of the past on the future is summarized only by the current state, rather than the whole history. In other words, the distribution of X_{n+1} depends only on the value of X_n and not on the whole set of values of $X_0, X_1, \ldots, X_n$. A stochastic process with such property is called a **Markov chain**.

Definition 5 (Markov property) Let X_n be a discrete-time SP that, for each n, takes value in some discrete set $\mathscr{S}$, which we call the **state space**. We say that the SP has the Markov property if

$$P\{X_{n+1} = i | X_n, X_{n-1}, \ldots, X_0\} = P\{X_{n+1} = i | X_n\} \tag{1.15}$$

for all n in the time-domain of the process and for all $i \in \mathscr{S}$.

Definition 6 (Markov chain)
Let X_n be a discrete-time SP. We say that the SP is a Markov chain (MC) if it satisfies the Markov property.

We will say that an MC is **finite** if $\mathscr{S}$ is a finite set. In this case, we usually take the elements of $\mathscr{S}$ as consecutive integers. For instance, for a state space with m elements, we will take $\mathscr{S} = \{0, 1, \ldots, m-1\}$ or $\mathscr{S} = \{1, 2, \ldots, m\}$ under our convenience.

Note that the SRW is an MC, but it is not finite.

Any finite MC can be described in terms of the transition probabilities

$$p_{ij} = P\{X_{n+1} = j | X_n = i\} \tag{1.16}$$

All the elements of a MC model can be encoded in a transition probability matrix

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0,m-1} \\ p_{10} & p_{11} & \cdots & p_{1,m-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m-1,0} & p_{m-1,1} & \cdots & p_{m-1,m-1} \end{pmatrix} \tag{1.17}$$

Using (1.16), it is easy to see that

$$\sum_{j \in \mathscr{S}} p_{ij} = 1 \tag{1.18}$$

thus, all rows in $\mathbf{P}$ sum up to one. For this reason, we say that $\mathbf{P}$ is a right-stochastic matrix. In matrix form, this property can be expressed as

$$\mathbf{P}\mathbb{1}_m = \mathbb{1}_m \tag{1.19}$$

where $\mathbb{1}_m$ is an all-ones vector with dimension m.

Note that, in general, the transition probabilities p_{ij} (and, thus, the transition matrix) may change with time. However, the case of constant probabilities is particularly interesting.

Definition 7 (Homogeneous Markov Chain) A markov chain X_n is homogeneous if the transition probabilities do not depend on time, that is.

$$P\{X_{n+1} = j | X_n = i\} = P\{X_{m+1} = j | X_m = i\}, \qquad i \in \mathscr{S}, \quad j \in \mathscr{S} \tag{1.20}$$

for any m and n in the time domain of the process.

Example 5 (Two-state machine) A machine can be either working or broken on a given day. If it is working, it will break down in the next day with probability 0.01, and will continue working with probability 0.99. If it breaks down on a given day, it will be repaired and be working in the next day with probability 0.8, and will continue to be broken down with probability 0.2. We can model this machine by a Markov chain with two states: working, and broken down. The transition probability matrix is given by

$$\begin{pmatrix} 0.99 & 0.8 \\ 0.01 & 0.2 \end{pmatrix}$$

Example 6 (SRW) A simple random walk is an example of an MC. However, there is no transition probability matrix associated with the simple random walk since the sample space is infinite.

1.4.1 Transition graph

The transition probability matrix of a given homogeneous MC can be represented by a directed graph, whose nodes are the states, and and edges from node i to node j represents a transition from state i to state j with nonzero probability. The weight of the edge is the transition probability.

Example 7 The transition graph of an MC given by state space $\mathscr{S} = \{0,1,2\}$ and transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 0.4 & 0.2 & 0.4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.21}$$

is shown in Figure 1.3.

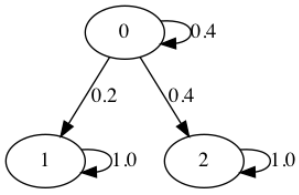

Fig. 1.3 The transition graph for the markov chain given by (1.21)

1.4.2 State probabilities

The transition probability matrix is useful to compute any probabilities about the state of the process at any given time. For instance, the n-th step transition probabilities, defined as $r_{ij}(n) = P\{X_n = j | X_0 = i\}$ can be computed recursively, by applying the total probability theorem, as

$$r_{ij}(n) = \sum_{k \in \mathscr{S}} P\{X_n = j | X_{n-1} = k, X_0 = i\} P\{X_{n-1} = k | X_0 = i\} = \sum_{k \in \mathscr{S}} r_{ik}(n-1) p_{kj}, \qquad n > 1 \tag{1.22}$$

Note that $r_{ij}(1) = p_{ij}$.

Defining the n-step transition probability matrix $\mathbf{R}_n$ as the matrix with components $r_{ij}(n)$, the recurrent relation (1.22) can be written in matrix form as

$$\mathbf{R}_n = \mathbf{R}_{n-1} \mathbf{P} \tag{1.23}$$

and, noting that $\mathbf{R}_1 = \mathbf{P}$, we get

$$\mathbf{R}_n = \mathbf{P}^n \tag{1.24}$$

Using these relations we can compute the state probabilities at any given time, given the initial state or the initial state probabilities.

For instance, let $\mathbf{q}_n$ be the state probability vector at time n, that is

$$q_{n,i} = P\{X_n = i\}, \quad n \geq 0, \quad i \in \mathscr{S} \tag{1.25}$$

then we can express the state probability vector at time n as a function of the initial state probabilities using (1.24) as

$$\begin{aligned} q_{n,j} = P\{X_n = j\} &= \sum_{i \in \mathscr{S}} P\{X_n = j | X_0 = i\} P\{X_0 = i\} \\ &= \sum_{i \in \mathscr{S}} r_{ij}(n) P\{X_0 = i\} \end{aligned} \tag{1.26}$$

which, in vector form, can be computed as

$$\mathbf{q}_n = \mathbf{R}_n^\intercal \mathbf{q}_0 = (\mathbf{P}^\intercal)^n \mathbf{q}_0 \tag{1.27}$$

1.4.3 Stationary distribution

Definition 8 (Stationary distribution) A stationary distribution of an MC is a probability distribution over the state space $\mathscr{S}$ (where $P\{X_0 = j\} = \pi_j$) such that

$$\mathbf{P}^\intercal \boldsymbol{\pi} = \boldsymbol{\pi} \tag{1.28}$$

Example 8 Let $\mathscr{S} = \mathbb{Z}_m$ and $X_0 = 0$. Consider the MC X_n such that $X_{n+1} = (X_n + 1) \mod m$ with probability $\frac{1}{2}$ and $X_{n+1} = (X_n - 1) \mod m$ with probability $\frac{1}{2}$.

The stationary distribution of this MC is $\pi_i = \frac{1}{m}$.

Note that the vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_m)^\intercal$ is an eigenvector of $\mathbf{P}^\intercal$ with eigenvalue 1. Hence the following theorem can be deduced from the Perron-Frobenius theorem.

Theorem 1 *If $p_{ij} > 0$ for all $i, j \in \mathscr{S}$, then there exists a unique stationary distribution of the system.*

Moreover, if $\mathbf{r}_i(n)$ is the state conditional probability vector with components $r_{ij}(n)$, $j \in \mathscr{S}$ (i.e., the values in the i-th row of $\mathbf{R}_n$),

$$\lim_{n\to\infty} \mathbf{r}_i(n) = \boldsymbol{\pi}, \qquad \forall i, j \in \mathscr{S} \tag{1.29}$$

A corresponding theorem is not true if we consider infinite state spaces.

Finite Markov chains can be shown to have at least one stationary distribution. However, if at least one of the transition probabilities is zero, the behavior of the process may depend on the topology of the transition graph:

- The stationary distribution may be not unique.
- The limiting distributions $\lim_{n\to\infty} \mathbf{r}_i(n)$ may depend on i.

Example 9 The MC in example in 7 has at least two limiting distributions: $\boldsymbol{\pi}_A = (0,1,0)^\intercal$ and $\boldsymbol{\pi}_B = (0,0,1)^\intercal$. Also, any convex combination of them $\boldsymbol{\pi} = \alpha\boldsymbol{\pi}_A + (1-\alpha)\boldsymbol{\pi}_B$, with $0 \leq \alpha \leq 1$ is a stationary distribution.

The limiting distributions depend on the initial state: is is easy to verify that

$$\lim_{n\to\infty} \mathbf{r}_i(n) = (0,0,1) \tag{1.30}$$

$$\lim_{n\to\infty} \mathbf{r}_i(n) = (0,1,0) \tag{1.31}$$

$$\lim_{n\to\infty} \mathbf{r}_i(n) = \left(0, \frac{1}{3}, \frac{2}{3}\right) \tag{1.32}$$

1.4.4 Computing the stationary distribution

In general, the stationary distribution are all the solution of the set of equations (1.28) and the probabilistic constraint $\mathbb{1}^\intercal \mathbf{p} = 1$, that is

$$\begin{pmatrix} \mathbf{P}^\intercal - \mathbf{I} \\ \mathbb{1}^\intercal \end{pmatrix} \boldsymbol{\pi} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{1.33}$$

The system of linear equations given by (1.33) is consistent (it has at least one solution) but overdetermined, because the left matrix has $m+1$ rows but only m columns. In fact, using (1.19), it is easy to see that

$$\mathbb{1}_m^\intercal (\mathbf{P}^\intercal - \mathbf{I}) = \mathbf{0} \tag{1.34}$$

so that the rank of $\mathbf{P}^\intercal - \mathbf{I}$ is at most $m-1$. Thus, we can remove one of the rows in the matrix (any one of them) to solve the system.