Chapter 7

Markov Chains

7.1 Introduction

7.1.1 Contents

This chapter gives a brief outline of the elementary properties of Markov chains. Main attention is given to ergodic Markov chains. Ergodic Markov chains are useful, e.g., in modelling the segmentation properties of DNA sequences, as will be shown in Chapter 11.

7.1.2 Definitions and Notations

Consider an alphabet $S = \{s_1, s_2, \ldots, s_J\}$ and a sequence of random variables $X_0, X_1, \ldots, X_n, \ldots$, assuming values in S. The symbols s_j in the alphabet are called *states* and S is also called the state space. We can encode the states by giving the state s_j the label j and take for convenience of typing $S = \{1, 2, \ldots, J\}$.

Definition 7.1.1 *A sequence of random variables* $\{X_n\}_{n=0}^{\infty}$ *is called a* **Markov chain**, **(MC)**, *if for all* $n \geq 1$ *and* $j_0, j_1, \ldots, j_n \in S$,

$$P(X_n = j_n | X_0 = j_0, \ldots, X_{n-1} = j_{n-1}) = P(X_n = j_n | X_{n-1} = j_{n-1}). \quad (1.1)$$

■

The significance of an MC lies in that if $X_n = j_n$ is a future event, then the conditional probability of this event given the past history $X_0 = j_0, X_1 =$

$j_1, \ldots, X_{n-1} = j_{n-1}$ depends only upon the immediate past $X_{n-1} = j_{n-1}$ and not upon the remote past $X_0 = j_0, X_1 = j_1, \ldots, X_{n-2} = j_{n-2}$.

The condition in (1.1) is known as the *Markov property*. The Markov property yields a simple and mathematically tractable alternative to the multinomial process from Chapter 3 as a model family for $\{X_n\}_{n \geq 0}$.

Let $\{X_n\}_{n=0}^{\infty}$ be Markov chain. If $X_n = j$, we say that the *the chain is in state j at time n* or that *the chain visits the state at time n*. The conditional probabilities

$$p_{i|j} = P\left(X_n = j | X_{n-1} = i\right), \qquad n \geq 1, \qquad i, j \in S \tag{1.2}$$

are assumed to be independent of n and are called *(stationary) one step transition probabilities*. A Markov chain with stationary transition probabilities is called *homogeneous*. If the conditional probability is not defined, we put $p_{i|j} = 0$.

The numbers $p_{i|j}$ are for many good purposes displayed in matrix form

$$P = \left(p_{i|j}\right)_{i=1,j=1}^{J,J} \tag{1.3}$$

or

$$P = \begin{pmatrix} p_{1|1} & p_{1|2} & \cdots & p_{1|J} \\ p_{2|1} & p_{2|2} & \cdots & p_{2|J} \\ \vdots & \vdots & \ddots & \vdots \\ p_{J|1} & p_{J|2} & \cdots & p_{J|J} \end{pmatrix}. \tag{1.4}$$

Thus P is an $J \times J$ matrix to be called a *transition matrix*. The ith row of P is the conditional probability distribution of X_n given that $X_{n-1} = i$. Therefore the following properties hold true:

$$p_{i|j} \geq 0, \qquad \sum_{j=1}^{J} p_{i|j} = 1. \tag{1.5}$$

A matrix with this property is called a *stochastic matrix*.

7.1.3 Examples

Example 7.1.1 [A binary Markov chain] A binary Markov chain is a sequential mechanism producing zeros (0) and ones (1). It can, e.g., be used

modelling a 'regulatory region' (of DNA) or 'not a regulatory region'. A regulatory region is a part of chromosomal DNA that serve as signals for turning the gene expression off or on. If at some stage 0 is produced, then at the next stage 1 will be produced with probability p and 0 will be produced with probability $1 - p$. If a 1 is produced, then at a next stage 0 will be produced with probability q and 1 will be produced with probability $1 - q$. This corresponds to the transition matrix

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}. \tag{1.6}$$

Figure 7.1 is a graphical representation of the state space and the transition matrix of this chain. Figure 7.2 shows (a segment of) the *trellis* for the sample paths of the Markov chains in this example. Each path from left to right through the trellis following at each step one of the arrows in the diagram is one possible sample path of the chain. For $n = 0, 1, 2$ there are eight ($= 2^3$) possible sample paths.

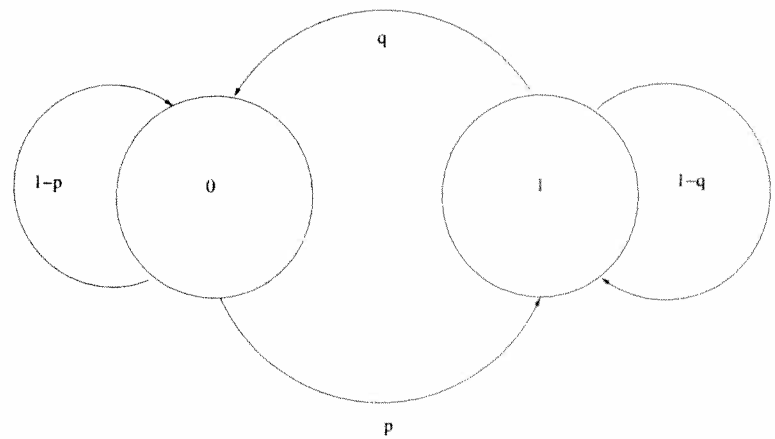

Figure 7.1: A state space graph

7. The random variables $\{Z_n\}_{n=-1}^{\infty}$ are independent and identically distributed with $Z_n \in \mathrm{Be}\left(\frac{1}{2}\right)$ for all n. We set

$$X_n = [Z_n, Z_{n-1}], \qquad n = 0, 1, 2, \ldots,$$

so that $X_0 = [Z_0, Z_{-1}]$, $X_1 = [Z_1, Z_0]$ and so on.

(a) Show that $\{X_n\}_{n=0}^{\infty}$ is a Markov chain.

(b) Find the invariant distribution of the chain $\{X_n\}_{n=0}^{\infty}$.

7.5.2 Stationary Markov Processes

8. A stochastic process $\{Z_n\}_{n=0}^{\infty}$ with values in $\mathcal{Z} = \{z_1, \ldots, z_r\}$ is called **(strictly) stationary** is the finite dimensional distributions are invariant to shift, or

$$P(Z_{n_1} = z_{l_1}, \ldots, Z_{n_k} = z_{l_k}) = P(Z_{n_1+h} = z_{l_1}, \ldots, Z_{n_k+h} = z_{l_k}) \quad (5.7)$$

for any sequence $z_{l_1}, \ldots, z_{l_k}$ of symbols in $\mathcal{Z}$ and any $n_1, \ldots, n_k$ and for any integer $h > 0$. Let

$$\pi = (\pi_1, \ldots, \pi_J)$$

be the invariant distribution of a time homogeneous Markov chain $\{X_n\}_{n=0}^{\infty}$ with $\pi_j > 0$ for all j. Verify that if $\pi_j = P(X_0 = j)$ for all j then $\{X_n\}_{n=0}^{\infty}$ is stationary in the sense of (5.7).

7.5.3 Entropy Rate of a Stationary Markov Chain

The *entropy rate* of a stochastic process$\{X_n\}_{0=1}^{\infty}$ with a finite alphabet is defined as

$$H_{\infty}(X) := \lim_{n \longrightarrow \infty} \frac{H(X_0, X_1 \cdots, X_n)}{n+1}. \quad (5.8)$$

9. Prove that this limit exists for stationary random processes and is given by

$$H_{\infty}(X) = \lim_{n \to \infty} H(X_n | X_{n-1}, \ldots, X_0),$$

where $H(X_n | X_n, \ldots, X_0)$ is defined as the entropy of the conditional distribution or

$$H(X_n | X_{n-1}, \ldots, X_0) = \sum_{x_n} \cdots \sum_{x_{n-1}} p_{X_{n-1},\ldots,X_0}(x_{n-1}, \ldots, x_0)\, h,$$

where

$$h = -\sum_{x_n} p_{X_n|X_{n-1},\ldots,X_0}(x_n | x_{n-1}, \ldots, x_0) \log p_{X_n|X_{n-1},\ldots,X_0}(x_n | x_{n-1}, \ldots, x_0).$$

10. Let $\{X_n\}_{0=1}^{\infty}$ be a stationary Markov chain with the transition matrix $P = (p_{i|j})_{i=1,j=1}^{J,J}$ and the invariant distribution π. Show that the entropy rate $\mathcal{H}$ of $\{X_n\}_{0=1}^{\infty}$ is equal to

$$H_{\infty}(X) = -\sum_{i=1}^{J} \pi_i \sum_{j=1}^{J} p_{i|j} \cdot \log p_{i|j}.$$

11. Let $\{X_n\}_{0=1}^{\infty}$ be a time invariant Markov chain with the transition matrix the transition matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ 0 & 1 \end{pmatrix}. \quad (5.9)$$

Determine the entropy rate of this Markov chain.

7.5.4 High-order Markov Chains

12. Suppose that $\{Z_n\}_{n=0}^{\infty}$ is a second order Markov chain with the state space

$$\mathcal{M} = \{1, 2, \ldots, m\}$$

and that

$$p_{i,k|j} = P(Z_n = j | Z_{n-1} = i, Z_{n-2} = k), \qquad n \geq 1, i, j \in \mathcal{M} \quad (5.10)$$

are the time-homogeneous transition probabilities. The invariant distribution is a bivariate distribution

$$\mu_{j,k} = P(Z_{n-1} = j, Z_n = k).$$

Hence the probabilities $\mu_{j,k}$ satisfy

$$\mu_{j,k} = \sum^{m} \mu_{i,j} p_{i,k|j} \quad (5.11)$$

and

$$\sum_{j=1}^{m}\sum_{k=1}^{m}\mu_{j,k} = 1. \quad (5.12)$$

The process $\{Z_n\}_{n=0}^{\infty}$ can be transformed into a first order Markov chain by the definition

$$X_n = [Z_{n-1}, Z_n].$$

The effect of this is to enlarge the state space.

Take for example $\{Z_n\}_{n=0}^{\infty}$ assuming values in the binary state space $\mathcal{M} = \{1, 2\}$. The second order probability transition probabilities are fully specified by the four transition probabilities

$$p_{1,1|2} = P(Z_n = 2|Z_{n-1} = 1, Z_{n-2} = 1),$$

$$p_{2,2|1} = P(Z_n = 1|Z_{n-1} = 2, Z_{n-2} = 2),$$

$$p_{1,2|1} = P(Z_n = 1|Z_{n-1} = 2, Z_{n-2} = 1),$$

$$p_{2,1|2} = P(Z_n = 1|Z_{n-1} = 1, Z_{n-2} = 2).$$

The corresponding first order Markov chain X_n defined above then has the state space

$$\{(1,1), (1,2), 2,1), (2,2)\}$$

and the transition matrix

$$P = \begin{pmatrix} 1-p_{1,1|2} & p_{1,1|2} & 0 & 0 \\ 0 & 0 & p_{1,2|1} & 1-p_{1,2|1} \\ 1-p_{2,1|2} & p_{2,1|2} & 0 & 0 \\ 0 & 0 & p_{2,2|1} & 1-p_{2,2|1} \end{pmatrix}, \quad (5.13)$$

Find the invariant distribution for this chain.

13. **MTD**
A high order Markov property can be modelled by **mixture transition distribution** (MTD). We let $Q = \left(q_{i|j}\right)_{i=1,j=1}^{J,J}$ be a matrix with $q_{i|j} \geq 0$, $\sum_{j=1}^{J} q_{i|j} = 1$. Actually the original construction in (Raftery 1985) deals with *columnwise* stochastic matrices. A mixing distribution is given by $\lambda_i \geq 0, i = 1, \ldots, l, \sum_{i=1}^{l} \lambda_i = 1$. We consider a sequence of random variables $\{X_n\}_{n=0}^{\infty}$ assuming values in $\{1, \ldots, J\}$ such that for all n and $j_l, j_1, \ldots, j_n \in S$,

$$P(X_n = j_n|X_{n-l} = j_{n-l}, \ldots, X_{n-1} = j_{n-1}) = \sum_{i=1}^{l} \lambda_i q_{j_{n-i},j_n} \quad (5.14)$$

for a positive integer l. The process $\{X_n\}_{n=0}^{\infty}$ is called an MTD(l) process. The integer l is called the order of the process.

(a) How can the values of an MTD(l) process be generated?
(b) Introduce the vector process

$$[X_n, \ldots, X_{n-l+1}],$$

which assumes values in the alphabet J^l. Find its transition matrix, i.e., a matrix with the arrays

$$P(X_n = j_n, \ldots, X_{n-l+1} = j_{n-l+1}|X_{n-1} = j_{n-1}, \ldots, X_{n-l} = j_{n-l}).$$

(Raftery 1985).

7.6 References and Further Reading:

T.M. Cover and J.A. Thomas (1991): *Elements of Information Theory.* John Wiley and Sons, New York.

I. Csiszár (1963): Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publications of the Mathematical Institute of the Hungarian Academy of Science.* Ser. A, Fasc. 1-2, 8, pp. 85–108.

R.M. Gray (1988): *Probability, Random Processes and Ergodic Properties.*, Springer Verlag, New York, Berlin, etc..

L.L. Helms (1997): *Introduction to Probability Theory With Contemporary Applications.* W.H. Freeman and Company. New York.

O. Häggström (2001): *Finite Markov Chains and Algorithmic Applications.* `http://www.math.chalmers.se/~olleh/surveys.html`

J.G. Kemeny, J.L. Sell and A.W. Knapp (1976): *Denumerable Markov Chains.* Second Edition. Springer Verlag, New York, Berlin, etc..

A.I. Khuri (1993): *Advanced Calculus with Applications in Statistics.* John Wiley and Sons, New York.