

Chapter 1

Introduction

In the studies of natural systems, the main problem is usually to derive mathematical models for the response or responses in terms of input and/or risk variables as well as time. Almost all of these models are stochastic models because most of the risk variables are subject to random variations and most of the measurements of the responses and input variables are subject to random measurement errors. This is true for medical systems such as AIDS, cancer and genetics as well as for biological sciences, engineering sciences and social and economic systems. To study these systems, therefore, it is important to examine the probability laws governing the behavior of these systems. The purpose of this book is to present a systematic treatment on stochastic models which have been used in genetics, cancer, AIDS and some other medical systems.

To set up the basic background, in this chapter we will define some basic terminologies and give examples from genetic, cancer and AIDS to illustrate some basic concepts of stochastic processes.

1.1. Some Basic Concepts of Stochastic Processes and Examples

Definition 1.1. A *stochastic process* $\{X(t), t \in T\}$ is a family of random variables indexed by the parameter t in T .

In biomedical sciences as well as in many other areas, the parameter t is usually related to time and the set T is referred to as the parameter space. The sample space S_t of $X(t)$ is referred to as the state space and the elements of S_t the states. The space S_t may be discrete in which case the number of elements of S_t is either finite or countable infinite, or continuous in which case the number of elements of S_t is uncountable infinitely many. Similarly, T may either be discrete or continuous. It follows that there are four types of stochastic processes:

- (a) A stochastic process with discrete state space and discrete time t .
- (b) A stochastic process with discrete state space and continuous time t .
- (c) A stochastic process with continuous state space and discrete time t .
- (d) A stochastic process with continuous state space and continuous time t .

For given n and for given $t_0 < t_1 < \dots < t_n$, the observed values $\{X(t_0), X(t_1), \dots, X(t_n)\}$ of $X(t)$ at $\{t_0, t_1, \dots, t_n\}$ is referred to as a sample path of the process. We will refer the stochastic process as a finite stochastic process if the state space S contains only a finite number of states. To simplify notations, in what follows, we will let $T = \{t \geq 0\}$ if T is continuous and let $T = \{0, 1, \dots\}$ if T is discrete, unless otherwise stated (This can be achieved by defining the starting time of the process as 0). Similarly, in what follows, we will let $\{S = [a, b], -\infty \leq a < b \leq \infty\}$ if S is continuous and let $S = \{0, 1, \dots\}$ if S is discrete, unless otherwise stated (This can be achieved by defining the i th element of S as $i - 1$ ($i = 1, \dots, \infty$)).

The above definition of stochastic process can also be extended to k -dimensional stochastic processes with $k \geq 1$ being a positive integer.

Definition 1.2. A k -dimensional stochastic process $\{\underline{X}(t), t \in T\}$ is a family of k -dimensional random vectors indexed by the parameter t in T .

In k -dimensional stochastic processes, the state space S is then a subset of the k -dimensional Euclidean space $E^{(k)}$. Also, some of the variables of $\underline{X}(t)$ may be discrete while other random variables of $\underline{X}(t)$ may be continuous. These are mixed-type random vectors. In this book we will not consider cases with mixed types of random variables in $\underline{X}(t)$, unless otherwise stated; thus we will only consider cases in which either all random variables in $\underline{X}(t)$ are discrete or all random variables in $\underline{X}(t)$ are continuous.

Example 1.1. The frequency of mating types under full-sib mating in natural populations. In animal breeding, the breeders are usually confronted with the problem of sib mating (brother-sister mating) leading to inbred lines. Sib mating also are very common in wild animal populations. Hence, it is often of interest to compute the frequency of different mating types in sib mating in domestic as well as in wild populations. In a large population of diploid individuals, if we focus on a single locus with two alleles A and a , then there are three genotypes $\{AA, Aa, aa\}$ and there are 6 different mating types (see Remark 1.1): $\{AA \times AA, aa \times aa, AA \times aa, AA \times Aa, aa \times Aa, Aa \times Aa\}$ which we denote by $\{1, \dots, 6\}$ respectively. (As a convention, the genotype on the left denotes the genotype of the mother whereas the genotype on the right denotes the genotype of the father). Let t denote generation and let $X(t)$ denote the frequency of mating types at time t under sib-mating. Then $\{X(t), t \in T = (0, 1, \dots)\}$ is a stochastic process with discrete time and with state space $S = \{1, \dots, 6\}$. This is an example of stochastic process with discrete time and discrete state space.

Remark 1.1. In most of the plants, animals and human beings, the chromosomes are grouped into a fixed number of pairs of homologous chromosomes, one from the mother and the other from the father. This type of individuals has been referred to as diploid. For example, in human being, there are 23 pairs of chromosomes and hence human being are diploid individuals. Biologists have also shown that all characters are controlled by genes which are segments of DNA in the chromosomes. This segment has been referred to as locus and different genes in the same locus are referred to as alleles.

Example 1.2. Survival of mutant genes in natural population-branching processes. In human beings, many of the inherited disease are caused by mutation of certain genes [1, 2]. Suppose that at a certain time, a mutant gene is introduced into the population. Suppose further that each mutant gene produces j mutant genes with probability p_j ($j = 0, 1, \dots, \infty$) in the next generation independently of other genes. Let $X(t)$ be the number of mutant genes at generation t . Then $X(t)$ is a stochastic process with discrete time and with state space $S = \{0, 1, \dots, \infty\}$. As we shall see, this type of processes belongs to a class of stochastic processes referred to as Galton–Watson branching processes [3]. This is an example of stochastic process with discrete time and discrete state space.

Example 1.3. The change of frequency of genes in natural populations. In studying the theory of evolution, it is of interest to find the probability law in natural populations governing the changes of the frequencies of types or genes. Thus, in a large population of diploid individuals, for a single locus with two alleles A and a , one would need to find the probability law for the number of A allele over time. Let N be the population size and let t denote generation. Then the number $X(t)$ of A allele at time t is a stochastic process with discrete time and with state space $S = \{0, 1, \dots, 2N\}$. (Note that in a diploid population, each individual has two alleles for each locus; hence the total number of alleles for each locus in the population is $2N$.) This is an example of stochastic process with discrete time and discrete state space.

Let $Y(t) = \frac{X(t)}{2N}$. Since the population size N is usually very large and since the evolution process is an extremely slow process taking place over millions and millions of years, as we shall see in Chap. 6, $Y(t)$ can be closely approximated by a stochastic process with continuous time and continuous state space $S = [0, 1]$.

Example 1.4. The number of drug-resistant cancer tumor cells. In treating cancer by chemotherapy, a major difficulty is the development of drug-resistant cancer tumor cells. Thus, questions of the possible efficiency and optimal timing of cancer chemotherapy can be studied by mathematical models for the development of drug-resistant cancer tumor cells. Let $X_1(t)$ be the number of sensitive cancer tumor cells at time t and $X_2(t)$ the number of resistant cancer tumor cells at time t . Let 0 be the time starting treatment. Then $\{[X_1(t), X_2(t)], t > 0\}$ is a two-dimensional stochastic process with parameter space $T = \{t > 0\}$ and state space $S = \{(i, j), i, j = 0, 1, \dots\}$. Stochastic process of this type has been studied in [4]. This is an example of two-dimensional stochastic process with discrete state space and continuous parameter space.

Example 1.5. The multi-stage model of carcinogenesis. Cancer tumors develop from normal stem cells by going through a finite number of genetic changes or mutations with intermediate cells subjecting to stochastic proliferation (birth) and differentiation (death). That is, cancer tumors develop from normal stem cells by a multistage model with intermediate cells subjecting to stochastic birth and death. Assume that there are k ($k \geq 2$) intermediate stages. For $t \geq 0$ with 0 being the time of birth of the individual, let $X_0(t)$ denote the number of normal stem cells at time t , $X_i(t)$ ($i = 1, \dots, k$)

the number of the i th stage intermediate cells at time t and $T(t)$ the number of malignant cancer tumors at time t . Then $\{[X_i(t), i = 0, 1, \dots, k, T(t)], t \geq 0\}$ is a $(k+2)$ -dimensional stochastic process with parameter space $T = \{t > 0\}$ and with state space $S = \{(i_0, i_1, \dots, i_k, i_{k+1}), i_r = 0, 1, \dots, ; r = 0, 1, \dots, k+1\}$. In general, $\{X_i(t), i = 0, 1, \dots, k, T(t)\}$ involves both stochastic birth-death processes for cell proliferation and differentiation of normal stem cells, intermediate cells and cancer tumors and Poisson processes for generating intermediate cells through genetic changes or mutations. This is an example of multi-dimensional stochastic processes with discrete state space and continuous time. These processes have been discussed in detail in [5, 6].

Example 1.6. The AIDS epidemiology in homosexual populations.

Consider a large population of homosexual men such as the San Francisco homosexual population which is at risk for AIDS. Then there are three types of people regarding HIV epidemic in the population: The S (susceptible) people, the I (infective) people and the A (clinical AIDS cases) people. A S person does not carry the AIDS virus but can contract it through sexual contact with I people or AIDS cases or by sharing needles in IV drug use or through blood transfusion of contaminated blood. An I person carries the AIDS virus and can transmit the virus to S people through sexual contact or sharing contaminated needles with I people; there is a chance that he/she will develop AIDS symptoms to become an AIDS case. An AIDS case (An A person) is a person who has developed AIDS symptoms or who has $CD4^{(+)}$ T cell counts in the blood falling below $200/\text{mm}^3$ [7].

Let $S(t)$, $I(t)$ and $A(t)$ denote the numbers of susceptible people (S people), infected people (I people) and AIDS cases at time t respectively and write $\underline{X}(t) = \{S(t), I(t), A(t)\}'$, where \prime denotes transpose. Let $t_0 = 0$ be the time at which a few HIV were introduced into the population to start the AIDS epidemic. Then $\{\underline{X}(t), t \geq 0\}$ is a three-dimensional stochastic process with parameter space $T = \{t \geq 0\}$ and with state space $\Omega = \{(i, j, k), i, j, k \text{ being non-negative integers}\}$. This is an example of multi-dimensional stochastic process with discrete state space and continuous parameter space [8, Chap. 3].

Example 1.7. The HIV pathogenesis in HIV-infected individuals.

In a HIV-infected individual, let time 0 denote the time of HIV infection. Then, there are three types of $CD4^{(+)}$ T cells, the uninfected $CD4^{(+)}$ T cells

(denoted by T_1), the latently infected $CD4^{(+)}$ T cells (denoted by T_2) and the productively HIV-infected $CD4^{(+)}$ T cells (denoted by T_3 , also referred to as actively HIV-infected T cells). Let $T_i(t)$ ($i = 1, 2, 3$) denote the number of T_i ($i = 1, 2, 3$) cells at time t per mm^3 of blood and let $V(t)$ denote the number of free HIV at time t per mm^3 of blood. Denote by $\underline{X}(t) = \{T_i(t), i = 1, 2, 3, V(t)\}'$. Then $\{\underline{X}(t), t \geq 0\}$ is a four-dimensional stochastic process with parameter space $T = \{t \geq 0\}$ and with discrete state space $S = \{(i, j, k, l), i, j, k, l \text{ being non-negative integers}\}$; for more detail, see [8, Chaps. 7–8] and [9].

1.2. Markovian and Non-Markovian Processes, Markov Chains and Examples

In genetics, carcinogenesis, AIDS as well as in many other stochastic systems, many processes can be characterized by a *dependence condition* referred to as the *Markov condition*. These processes are classified as Markov processes.

Definition 1.3. Let $\{X(t), t \in T\}$ be a stochastic process with parameter space T and with state space S . Then $X(t)$ is called a *Markov process* iff (if and only if) for every n and for every $t_1 < \dots < t_n \leq t$ in T ,

$$\Pr\{X(t) \in A | X(t_1) = x_1, \dots, X(t_n) = x_n\} = P\{X(t) \in A | X(t_n) = x_n\},$$

for any event $A \subset S$. (1.1)

where $\Pr\{X(t) \in A | X(t_1) = x_1, \dots, X(t_n) = x_n\}$ is the conditional probability of $X(t) \in A$ given $\{X(t_1) = x_1, \dots, X(t_n) = x_n\}$ and $P\{X(t) \in A | X(t_n) = x_n\}$ the conditional probability of $X(t) \in A$ given $X(t_n) = x_n$.

The above definition is equivalent to stating that the probability distribution of $X(t)$ depends only on results in the most recent time and is independent of past history. From this definition, it is then seen that most of the processes in genetics and in evolution theory are Markov processes. Similarly, many process in carcinogenesis [5] and in AIDS epidemiology [8] are Markov processes. Thus, Examples 1.1–1.4 are Markov processes. However, there are also many processes in nature which are not Markov. An example from AIDS epidemiology is given in Example 1.10 whereas an example from cancer is given in Example 1.12 below. A sufficient condition for which $X(t)$ is Markov is that