

(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

for every  $t_1 < \dots < t_n$ ,  $X(t_2) - X(t_1), X(t_3) - X(t_2), \dots, X(t_n) - X(t_{n-1})$  are independently distributed of one another. This latter condition has been referred to as “independent increment”; see Exercise 1.1.

**Definition 1.4.** A Markov process  $\{X(t), t \in T\}$  with state space  $S$  is called a *Markov chain* iff  $S$  is discrete. (With no loss of generosity, one may assume  $S = \{0, 1, \dots, \infty\}$ .) A *Markov chain*  $\{X(t), t \in T\}$  is called a finite Markov chain iff the state space  $S$  contains only a finite number of states.

By this definition, Examples 1.1 and 1.3 are finite Markov chains whereas Examples 1.2 and 1.4 are Markov chains with infinite state space. Examples 1.1–1.3 are Markov chains with discrete time whereas Example 1.4 is a Markov chain with continuous time. General theories and its applications of Markov chains with discrete times will be discussed in detail in Chaps. 2 and 3 whereas general theories and its applications of Markov chains with continuous times will be discussed in detail in Chaps. 4 and 5. Notice that these general theories are characterized by the transition probabilities

$$p_{ij}(s, t) = \Pr\{X(t) = j | X(s) = i\}, \quad i \in S, j \in S.$$

For Markov chains with discrete times, the  $p_{ij}(s, t)$ 's are further characterized and derived by the one step transition probabilities

$$p_{ij}(t) = p_{ij}(t, t + 1) = \Pr\{X(t + 1) = j | X(t) = i\}, \quad i \in S, j \in S.$$

The analog of the one-step transition probabilities in Markov chains with continuous times are

$$p_{ij}(t, t + \Delta t) = \Pr\{X(t + \Delta t) = j | X(t) = i\} = \alpha_{ij}(t)\Delta t + o(\Delta t), \quad i \in S, j \in S,$$

where  $o(\Delta t)$  is defined by  $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$ . In the literature, the  $\alpha_{ij}(t)$  have been referred to as the transition rates or infinitesimal parameters. Thus, for Markov chains with continuous time, the processes are characterized by the infinitesimal parameters or transition rates.

**Definition 1.5.** A Markov chain  $\{X(t), t \in T\}$  with state space  $S = \{0, 1, \dots, \infty\}$  is a *homogeneous Markov chain* iff  $p_{ij}(s, t) = p_{ij}(t - s)$  for all  $i \in S, j \in S$ .

From Definition 1.5, notice that homogeneous Markov chains depend on the time parameters only through the difference of times. It follows that if the chain is homogeneous, then  $p_{ij}(t) = p_{ij}(s, s + t) = P\{X(s + t) = j | X(s) = i\}$

$= P\{X(t) = j | X(0) = i\}$  for all  $s \geq 0$ . Hence, for Markov chains with discrete time, the 1-step transition probabilities are given by  $p_{ij}(1) = P\{X(n+1) = j | X(n) = i\} = p_{ij}$  which are independent of  $n$ .

In natural systems, homogeneous Markov chains are very common although there are also nonhomogeneous Markov chains. For instance, Examples 1.1–1.4 given above are homogeneous Markov chains. In Example 1.2, however, if the progeny distribution of  $X(n)$  depends on  $n$ , then the chain is not homogeneous although it is Markov. For ease of illustration, in what follows, we will assume that the chain  $X(t)$  is homogeneous although many of the results hold also for some nonhomogeneous Markov chains, unless otherwise stated.

**Remark 1.2.** Homogeneous Markov chains are not stationary chains. In fact, as shown in Example 5.4, stationary distributions may not exist in some homogeneous Markov chains. On the other hand, the *homogeneity* condition is a pre-condition for defining stationary distributions.

**Example 1.8. The full-sib mating model for one locus with two alleles in natural populations.** In Example 1.1, we have considered a large diploid population under full-sib mating. In this example, we have focused on one locus with two alleles  $A : a$  and let  $X(t)$  denote the mating types at time  $t$ . Then, the state space consists of the six mating types  $AA \times AA, aa \times aa, AA \times aa, AA \times Aa, aa \times Aa, Aa \times Aa$  which are denoted by  $(1, \dots, 6)$  respectively. Thus,  $\{X(t), t \in T = (0, 1, 2, \dots)\}$  is a finite homogeneous Markov chain with state space  $S = \{1, \dots, 6\}$ . For this Markov chain, the matrix of the one-step transition probabilities is given by:

$$\begin{array}{l}
 \begin{array}{cccccc}
 & AA \times AA & aa \times aa & AA \times aa & AA \times Aa & aa \times Aa & Aa \times Aa \\
 AA \times AA & \left( \begin{array}{c} 1 \\ 0 \\ 0 \\ \frac{1}{4} \\ 0 \\ \frac{1}{16} \end{array} \right. & \begin{array}{c} 0 \\ 1 \\ 0 \\ 0 \\ \frac{1}{4} \\ \frac{1}{16} \end{array} & \begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{8} \end{array} & \begin{array}{c} 0 \\ 0 \\ 0 \\ \frac{1}{2} \\ 0 \\ \frac{1}{4} \end{array} & \begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{2} \\ \frac{1}{4} \end{array} & \begin{array}{c} 0 \\ 0 \\ 1 \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{array} \\
 aa \times aa & & & & & & \\
 AA \times aa & & & & & & \\
 AA \times Aa & & & & & & \\
 aa \times Aa & & & & & & \\
 Aa \times Aa & & & & & & 
 \end{array}
 \end{array}$$

The above matrix of one-step transition probabilities are derived by noting that matings occur only between brother and sister within the family. For

example, the mating type  $AA \times Aa$  gives only progenies  $AA$  and  $Aa$  with relative frequencies  $\{\frac{1}{2} AA, \frac{1}{2} Aa\}$ ; hence the frequencies of mating types in the next generation is  $\{\frac{1}{4} AA \times AA, \frac{1}{4} Aa \times Aa, \frac{1}{2} AA \times Aa\}$ . As another example, notice that the mating type  $Aa \times Aa$  gives progenies  $\{AA, aa, Aa\}$  with relative frequencies  $\{\frac{1}{4} AA, \frac{1}{2} Aa, \frac{1}{4} aa\}$ ; hence the frequencies of mating types in the next generation is  $\{\frac{1}{16} AA \times AA, \frac{1}{16} aa \times aa, \frac{1}{8} AA \times aa, \frac{1}{4} AA \times Aa, \frac{1}{4} aa \times Aa, \frac{1}{4} Aa \times Aa\}$ .

**Example 1.9. The simple Galton–Watson branching processes.** In genetics, in biological problems as well as in many other stochastic systems, an important class of Markov processes is the branching process; see [3]. This includes the simple Galton–Watson process which is a homogeneous Markov chain with discrete time. This latter process has been used to examine the stochastic behavior of mutant genes in populations; in particular, the survival of mutant genes as time progresses.

**Definition 1.6.** A Markov chain  $\{X(t), t \in T = (0, 1, 2, \dots)\}$ , with state space  $S = \{0, 1, 2, \dots\}$  is called a *simple branching process* (or *Galton–Watson process*) with progeny distribution  $\{p_k, k = 0, 1, 2, \dots (p_k \geq 0, \sum_{k=0}^{\infty} p_k = 1)\}$  iff  $P\{X(0) = 1\} = 1$  and the one-step transition probabilities  $p_{ij} = \Pr\{X(n+1) = j | X(n) = i\}$  are given by:

- (i) If  $i = 0, j \geq 0$ , then  $p_{ij} = \delta_{ij}$ , where,  $\delta_{ij} = 1$  if  $i = j$  and  $\delta_{ij} = 0$  if  $i \neq j$ .
- (ii) If  $i > 0$ , then

$$p_{ij} = \Pr\{Z_1 + Z_2 + \dots + Z_i = j\},$$

where  $Z_1, Z_2, \dots$  are independently and identically distributed with probability density function (pdf) given by  $p_j, j = 0, 1, 2, \dots$ .

From the above definition, we see that the Galton–Watson process is a homogeneous Markov chain. (Because  $p_0 = 1$  indicates that the mutant is certainly to be lost in the next generation while  $p_0 = 0$  is the situation that the mutant will never get lost, to avoid trivial cases we will assume  $0 < p_0, p_1 < 1$  in what follows).

To obtain  $p_{ij}$  for  $i > 0$ , let  $f(s)$  denote the *probability generating function* (pgf) of the progeny distribution  $\{p_j, j = 0, \dots, \infty\}$ ,  $f_n(s)$  the pgf of  $X(n)$  given  $X(0) = 1$  and  $g_i(s)$  the pgf of  $\{p_{ij}, i > 0, j = 0, \dots, \infty\}$ . Then, by

(ii) above,  $g_i(s) = [f(s)]^i$  and  $p_{ij}(n) = [p_{1j}(n)]^i$ . By definition of pgf, we have:

$$p_{ij} = \frac{1}{j!} \left\{ \frac{d^j}{ds^j} g_i(s) \right\}_{s=0};$$

$$p_{1j}(n) = \frac{1}{j!} \left\{ \frac{d^j}{ds^j} f_n(s) \right\}_{s=0}.$$

As an example, consider a diploid population with only one allele  $A$  at the  $A$  locus before the  $t_0$ th generation (With no loss of generality, one may assume that  $t_0 = 0$ ). Suppose that at the 0th generation, an  $A$  allele has mutated to  $a$  so that at the 0th generation there is an individual with genotype  $Aa$  in the population. Let  $X(t)$  be the number of mutant  $a$  at generation  $t$ . Assume that each mutant  $a$  reproduces itself independently of one another and that each mutant has probability  $p_j$  of giving  $j$  mutants in the next generation. Then, barring further mutations from  $A$  to  $a$  in the future,  $\{X(t), t \in T = \{0, 1, 2, \dots\}\}$  is a Galton–Watson process.

To specify  $p_j$ , let the fitness (i.e., average number of progenies per generation) of  $AA$  and  $Aa$  genotypes be given by  $\mu$  and  $\mu(1+v)$  ( $\mu > 0$ ) respectively. Let  $N$  be the population size. Then in the 0th generation, the frequency of the  $a$  allele is

$$\frac{(1+v)\mu}{(2N-1+1+v)\mu} = \frac{1}{2N}(1+v) + o((2N)^{-1}) = p + o((2N)^{-1})$$

for finite  $v$ , where  $p = \frac{1}{2N}(1+v)$ . When  $N$  is sufficiently large, and if the mating is random, then to order of  $o((2N)^{-1})$ , the probability that there are  $j$  “ $a$ ” mutants in the next generation is

$$p_j = \binom{2N}{j} p^j (1-p)^{2N-j}.$$

Since  $\lambda = 2Np = (1+v) + (2N)o((2N)^{-1}) \rightarrow (1+v)$  as  $N \rightarrow \infty$ , when  $N$  is sufficiently large,  $1+v$  is then the average number of progenies of the  $a$  allele and

$$p_j \sim e^{-(1+v)} \frac{(1+v)^j}{j!}, \quad j = 0, 1, 2, \dots$$

(Notice that the Poisson distribution is the limit of the binomial distribution if  $N \rightarrow \infty$  and if  $\lim_{N \rightarrow \infty} (2Np)$  is finite.)

Using the Poisson progeny distribution as above, we have that  $f(s) = e^{\lambda(s-1)}$  and  $g_i(s) = [f(s)]^i = e^{i\lambda(s-1)}$ ,  $i = 1, 2, \dots$ . Hence,

$$p_{0j} = \delta_{0j}, \quad j = 0, 1, \dots, \infty,$$

$$p_{ij} = e^{-i\lambda} \frac{(i\lambda)^j}{j!}, \quad i = 1, 2, \dots; \quad j = 0, 1, 2, \dots, \infty.$$

**Example 1.10. Nonhomogeneous Galton–Watson processes.** In the Galton–Watson processes, the progeny distribution may change as time progresses. This is true for new mutants which are usually selectively disadvantageous comparing with wild allele when they were first introduced into the population; however, environmental changes at latter times may make the mutants selectively more advantageous over the wild allele. In these cases, the branching processes become nonhomogeneous. To illustrate how to derive transition probabilities in these cases, assume that the progeny distributions of the mutant are given by  $\{p_j^{(i)}, i = 1, 2\}$  for  $n \leq t_1$  and  $t_1 < n$  respectively, where

$$p_j^{(i)} = e^{-\lambda_i} \frac{\lambda_i^j}{j!}, \quad i = 1, 2, \quad j = 0, 1, \dots, \infty,$$

and where  $\lambda_i = 1 + s_i$ ,  $i = 1, 2$ .

( $1 + s_i$  is the relative fitness of the mutant comparing with the wild allele over time with 1 for time  $n \leq t_1$  and 2 for time  $n > t_1$ .)

Let  $p_{ij}(n, n + 1) = \Pr\{X(n + 1) = j | X(n) = i\}$ . Then, for  $j = 0, 1, \dots, \infty$ ,

$$p_{0j}(n, n + 1) = \delta_{0j} \quad \text{for all } n \in T = (0, 1, \dots, \infty);$$

$$p_{ij}(n, n + 1) = e^{-i\lambda_1} \frac{(i\lambda_1)^j}{j!} \quad \text{for } n \leq t_1 \text{ and for all } i = 1, 2;$$

$$p_{ij}(n, n + 1) = e^{-i\lambda_2} \frac{(i\lambda_2)^j}{j!} \quad \text{for } n > t_1 \text{ and for all } i = 1, 2.$$

**Example 1.11. The Wright model in population genetics.** In Example 1.3, we have considered a large diploid population and have focused on one locus with two alleles, say  $A$  and  $a$ . Let the population size be  $N$  (In reality,  $N$  is the number of individuals who mature to produce progenies). Denote by  $\{X_1(t), X_2(t)\}$  the numbers of the genotypes  $AA$  and  $Aa$  at generation  $t$  respectively (The number of the genotype  $aa$  at generation  $t$  is

$X_3(t) = N - X_1(t) - X_2(t)$  as the population size is  $N$ ). Since the genotype  $AA$  contributes 2  $A$  alleles while the genotype  $Aa$  contributes only one  $A$  allele,  $X(t) = 2X_1(t) + X_2(t)$  is then the number of  $A$  allele at generation  $t$ . Let  $p(i, t)$  denote the frequency of  $A$  allele at generation  $t + 1$  given  $X(t) = i$ . Since the total number of alleles in the population is  $2N$  as each individual has two alleles, hence  $p(i, t)$  is a function of  $\frac{X(t)}{2N} = \frac{i}{2N}$  given  $X(t) = i$ . Under the assumption that the mating is random among individuals, the conditional probability that  $\{X_1(t+1) = m, X_2(t+1) = n\}$  given  $X(t) = i$  is then given by:

$$\begin{aligned} & \Pr\{X_1(t+1) = m, X_2(t+1) = n | X(t) = i\} \\ &= \frac{N!}{m!n!(N-n-m)!} [p(i, t)^2]^m \\ & \quad \times [2p(i, t)q(i, t)]^n [q(i, t)^2]^{N-m-n}, \end{aligned}$$

where  $q(i, t) = 1 - p(i, t)$ .

The probability generating function (pgf) of  $X(t+1) = 2X_1(t+1) + X_2(t+1)$  given  $X(t) = i$  is

$$\begin{aligned} \theta(s) &= \sum_{r=0}^{2N} s^r \Pr\{X(t+1) = r | X(t) = i\} \\ &= \sum_{m=0}^N \sum_{n=0}^{N-m} s^{2m+n} \frac{N!}{m!n!(N-n-m)!} [p(i, t)^2]^m [2p(i, t)q(i, t)]^n \\ & \quad \times [q(i, t)^2]^{N-m-n} \\ &= \{[sp(i, t)]^2 + 2sp(i, t)q(i, t) + [q(i, t)]^2\}^N \\ &= \{sp(i, t) + q(i, t)\}^{2N} \\ &= \sum_{j=0}^{2N} s^j \binom{2N}{j} [p(i, t)]^j [q(i, t)]^{2N-j}. \end{aligned}$$

It follows that the process  $\{X(t), t \in T\}$  is a Markov chain with discrete time  $T = (0, 1, \dots, \infty)$  and with state space  $S = \{0, 1, \dots, 2N\}$ . The one step

transition probability is:

$$\Pr\{X(t + 1) = j | X(t) = i\} = \binom{2N}{j} [p(i, t)]^j [q(i, t)]^{2N-j},$$

where  $q(i, t) = 1 - p(i, t)$ .

The above model has been referred to as the Wright model in population genetics [10]. Whether or not this chain is homogeneous depending on  $p(i, t)$ . The following cases have been widely considered in the literature of population genetics.

(i) If there are no mutation, no selection among the individuals and no immigration and migration, then  $p(i, t) = \frac{i}{2N}$  if  $X(t) = i$ . This case has been referred to as the *Random Genetic Drift* in population genetics; see Chaps. 6 and 7. In this case, the chain is homogeneous.

(ii) Suppose that there are mutations from  $A$  to  $a$  and from  $a$  to  $A$  in each generation but there are no selection, no immigration and no migration. Let the mutation rates per generation from  $A$  to  $a$  be  $u$  and from  $a$  to  $A$  be  $v$ . Then, given  $X(t) = i$ ,  $p(i, t) = \frac{i}{2N}(1 - u) + (1 - \frac{i}{2N})v$ . If both  $u$  and  $v$  are independent of time  $t$ , then the chain is homogeneous. However, because of the changing environment, it is expected that both  $u$  and  $v$  are functions of time. In this latter case, the chain is not homogeneous.

(iii) Suppose that there are no mutations, no immigration and no migration but there are selections among the individuals. Let the fitness (i.e., the expected number of progenies) of the three genotypes  $\{AA, Aa, aa\}$  be given by  $c(1 + s_1)$ ,  $c(1 + s_2)$  and  $c$  respectively ( $c > 0$ ); see Remark 1.3. Then, given  $X(t) = i$ , with  $x = \frac{i}{2N}$ ,  $p_{t+1} = p(i, t)$  is given by:

$$\begin{aligned} p(i, t) &= \frac{2x^2c(1 + s_1) + 2x(1 - x)c(1 + s_2)}{2x^2c(1 + s_1) + 2 \times 2x(1 - x)c(1 + s_2) + 2(1 - x)^2c} \\ &= \frac{x[1 + xs_1 + (1 - x)s_2]}{1 + x[xs_1 + 2(1 - x)s_2]}. \end{aligned}$$

Hence, if both  $s_1$  and  $s_2$  are independent of time  $t$ , then the chain is homogeneous; if any of  $s_1$  and  $s_2$  depend on time  $t$ , then the chain is not homogeneous.

(vi) Suppose that there are no immigration and no migration but there are selections among the individuals and there are mutations from  $A$  to  $a$  and from  $a$  to  $A$ . Let the mutation rates per generation be given in (ii) and let the fitness of the three genotypes be given in (iii). Then, given  $X(t) = i$ , with

$x = \frac{i}{2N}$ ,  $p_{t+1} = p(i, t)$  is given by:

$$\begin{aligned} p(i, t) &= (1 - u) \frac{x^2(1 + s_1) + x(1 - x)(1 + s_2)}{x^2(1 + s_1) + 2x(1 - x)(1 + s_2) + (1 - x)^2} \\ &\quad + v \frac{x(1 - x)(1 + s_2) + (1 - x)^2}{x^2(1 + s_1) + 2x(1 - x)(1 + s_2) + (1 - x)^2} \\ &= (1 - u) \frac{x[1 + xs_1 + (1 - x)s_2]}{1 + x[xs_1 + 2(1 - x)s_2]} \\ &\quad + v \frac{(1 - x)[1 + xs_2]}{1 + x[xs_1 + 2(1 - x)s_2]}. \end{aligned}$$

(v) Suppose that there are immigration and migration but there are no mutations and no selections among the individuals. To model this, we follow Wright [11] to assume that the population exchanges the  $A$  alleles with those from outside populations at the rate of  $m$  per generation. Let  $x_I$  denote the frequency of  $A$  allele among the immigrants. Then, given  $X(t) = i$ , with  $x = \frac{i}{2N}$ ,  $p_{t+1} = p(i, t)$  is given by:

$$p(i, t) = x + m(x_I - x).$$

The chain is homogeneous if  $m$  and/or  $x_I$  are independent of time  $t$ ; otherwise, the chain is not homogeneous.

**Remark 1.3.** Because the frequency of the alleles are the major focus in population genetics, one may with no loss of generality assume relative fitness for the genotypes. This is equivalent to delete the constant  $c$  from the fitness of the genotypes.

**Example 1.12. The staged model of the AIDS epidemiology.** In the AIDS epidemic, for clinical management and for taking into account the effects of infection duration, the infective stage ( $I$  stage) is usually divided into substage  $I_1, \dots, I_k$  with stochastic transitions between these substage [12–17]. For example, based on the total number of CD4<sup>(+)</sup> T cell counts per mm<sup>3</sup> of blood, Satten and Longini [16–17] have classified the  $I$  stage into 6 substage given by:  $I_1$ , CD4 counts  $\geq 900/\text{mm}^3$ ;  $I_2$ ,  $900/\text{mm}^3 > \text{CD4 counts} \geq 700/\text{mm}^3$ ;  $I_3$ ,  $700/\text{mm}^3 > \text{CD4 counts} \geq 500/\text{mm}^3$ ;  $I_4$ ,  $500/\text{mm}^3 > \text{CD4 counts} \geq 350/\text{mm}^3$ ;  $I_5$ ,  $350/\text{mm}^3 > \text{CD4 counts} \geq 200/\text{mm}^3$ ;  $I_6$ ,  $200/\text{mm}^3 > \text{CD4 counts}$ . (Because of the 1993 AIDS definition by CDC [7], we will merge

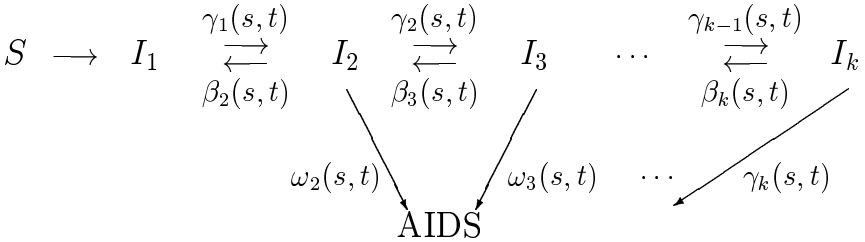


Fig. 1.1. A general model of the HIV epidemic with reverse transition.

the  $I_6$  stage with the AIDS stage (A stage).) Let  $S$  denote the susceptible stage. Then the model and the transition is expressed schematically in Fig. 1.1.

Let  $S(t)$  and  $I_i(t)$  ( $i = 1, \dots, k$ ) denote the numbers of  $S$  people and  $I_i$  ( $i = 1, \dots, k$ ) people at time  $t$  respectively and  $A(t)$  the number of AIDS cases at time  $t$ . Let  $t = 0$  be the time to start the epidemic. Then we have a  $(k + 2)$ -dimensional stochastic process  $\underline{U}(t) = \{S(t), I_i(t), i = 1, \dots, k, A(t)\}$  with parameter space  $T = \{t \geq 0\}$  and with sample space  $\Omega$  which is a subset of an  $(k + 2)$ -dimensional Euclidean space with non-negative integers as components.

Let  $p_S(t)dt$  be the probability of  $S \rightarrow I_1$  during  $[t, t + dt)$ . With  $I_0 = S$  and  $I_{k+1} = A$ , let the transition rates of  $I_i \rightarrow I_{i+1}$ ,  $I_i \rightarrow I_{i-1}$  and  $I_i \rightarrow A$  at time  $t$  be given by  $\gamma_i(s, t)$ ,  $\beta_i(s, t)$  and  $\omega_i(s, t)$ , respectively, for  $I_i$  ( $i = 1, \dots, k$ ) people who have arisen from  $I_{i-1}$  at time  $s$ . (Note that  $\beta_1(s, t) = 0$ ,  $\gamma_k(s, t) = \omega_k(s, t)$  from Fig. 1.1.) If  $\gamma_i(s, t) = \gamma_i(t)$ ,  $\beta_i(s, t) = \beta_i(t)$  and  $\omega_i(s, t) = \omega_i(t)$  are independent of  $s$ , then the process  $\underline{U}(t)$  is Markov. This process is in fact a Markov chain since the number of states is countable infinite. These are the processes considered by Longini and his associates [12]. On the other hand, if  $\gamma_i(s, t)$ ,  $\beta_i(s, t)$  and  $\omega_i(s, t)$  are dependent on  $s$ , then the process  $\underline{U}(t)$  is not Markov [13, 14, 18]. The non-Markovian processes arise because of treatment of HIV-infected individuals by anti-viral drugs such as AZT. These are the processes considered by Longini *et al.* [13, 14]; see also [18].

**Example 1.13. The MVK two-stage model of carcinogenesis.** The two-stage model of carcinogenesis which fits the cancer biological mechanism was first proposed by Knudson [19], Moolgavkar and Venzon [20] and Moolgavkar and Knudson [21] and has been referred to as the MVK two-stage model. This model assumes that a cancer tumor develops from a single normal stem

cell by clonal expansion and views carcinogenesis as the end result of two discrete, heritable and irreversible events in normal stem cells; each event occurs during cell division. The MVK two-stage model has assumed that the parameter values are constant independent of time and has made the assumption that with probability one each cancer tumor cell grows instantaneously into a malignant tumor. Furthermore, it is assumed that the proliferation and differentiation of normal stem cells follow deterministic growth.

Let  $N(t)$  denote the number of normal stem cells at time  $t$  and  $\{I(t), T(t)\}$  the numbers of initiated cells and cancer tumor cells. (Note that  $I(t)$  is actually the number of the first initiated tumor cells and  $T(t)$  the number of second initiated cells under the assumption that each second initiated cell grows instantaneously into a malignant tumor; see [5, Chap. 3].) Then  $N(t)$  is a deterministic function of time  $t$  and  $\{I(t), T(t)\}$  is a two-dimensional Markov chain with continuous time; see Tan [5, Chap. 3]. To find the transition rates (or incidence functions) of this process, let  $M_I(t)$  be the number of mutations from normal stem cells to  $I$  cells during  $[t, t + \Delta t)$  and denote by  $\lambda(t) = N(t)\alpha_N(t)$ , where  $\alpha_N(t)$  is the mutation rate per cell division from  $N$  to  $I$  at time  $t$ . Then, to order of  $O(N(0)^{-1})$ ,  $M_I(t)$  follows a Poisson distribution with mean  $\lambda(t)\Delta t$ ; see Exercise 1.13. Let  $b_I(t)$ ,  $d_I(t)$  and  $\alpha_I(t)$  be the birth rate, the death rate and the mutation rate of the  $I$  cells. Then, during  $[t, t + \Delta t)$ , the probabilities that an  $I$  cell will yield two  $I$  cells, 0  $I$  cells and 1  $I$  cell and 1  $T$  cell are given respectively by  $b_I(t)\Delta t + o(\Delta t)$ ,  $d_I(t)\Delta t + o(\Delta t)$  and  $\alpha_I(t)\Delta t + o(\Delta t)$ , respectively. It follows that as defined in Chap. 4,  $\{I(t), t \leq 0\}$  is a stochastic birth-death process with birth rate  $jb_I(t) + \lambda(t)$  and death rate  $jd_I(t)$ . This is a nonhomogeneous stochastic Feller–Arley birth-death process with immigration as defined in [22].

For the above process, notice that because the number of stem cells after birth is usually very large ( $10^6 \sim 10^8$ ), it is a good approximation to assume that  $N(t)$  is a deterministic function of  $t$ ; in fact it has been shown by Tan and Brown [23] through continuous multiple branching process that to order of  $O(N(0)^{-1})$ ,  $N(t)$  is indeed a deterministic function. However, it has been recognized that the assumption that with probability one each cancer tumor cell grows instantaneous into a malignant tumor does not hold in many real world situations [6, 24, 25]; in fact it has been shown by Yang and Chen [24] and Tan and Chen [6] that malignant tumor cells develop by clonal expansion from primary second initiated cells. It follows that conditional on the number of  $I(s)$

cells for all  $s \leq t$ ,  $T(t)$  follows a Poisson distribution with conditional mean given by  $\lambda_T(t) = \int_0^t I(x)\alpha_I(s)P_T(s, t)dx$ , where  $P_T(s, t)$  is the probability that a second initiated cell arising at time  $s$  will develop into a malignant cancer tumor by time  $t$ ; for proof of this, see Chap. 8. Since the distribution of  $T(t)$  depends on the  $I(s)$  for all  $s \leq t$ ,  $T(t)$  is not even a Markov process; for more detail, see Chap. 8.

### 1.3. Diffusion Processes and Examples

Let  $\{X(t), t \geq 0\}$  be a stochastic process with continuous parameter space  $T = \{t \geq 0\}$  and with continuous state space  $S = [a, b]$ . ( $a$  can be  $-\infty$  and  $b$  can be  $\infty$ .) Suppose that the increment  $dX(t) = X(t + dt) - X(t)$  changes continuously in probability when  $dt$  is very small so that the probability of any jump (say  $\epsilon > 0$ ) would be nil. Also, in many practical problems, it is reasonable to assume that if  $dt \cong 0$ , one may practically ignore higher order moments (i.e., with order  $\geq 3$ ) of  $dX(t)$ . This leads to a class of stochastic processes which involve only the first and second moments of  $dX(t)$ . If these processes are Markov processes, then they are classified as *Diffusion Processes*.

**Definition 1.7.** Let  $X(t)$  be a Markov stochastic process with parameter space  $T = \{t \geq 0\}$  and with state space  $S = [a, b]$ . Then  $X(t)$  is called a *diffusion process with coefficients*  $\{m(x, t), v(x, t)\}$  if and only if the following conditions hold:

- (i) For every  $\epsilon > 0$  given, for every  $x \in S$  and for every  $t \geq 0$ ,

$$P\{|X(t + dt) - X(t)| \geq \epsilon | X(t) = x\} = o(dt),$$

where  $o(dt)$  is defined by  $\lim_{dt \rightarrow 0} o(dt)/dt = 0$ .

- (ii) There exists a continuous function  $m(x, t)$  of  $x \in S$  and  $t \geq 0$  satisfying the condition:

$$E[X(t + dt) - X(t) | X(t) = x] = m(x, t)dt + o(dt).$$

- (iii) There exists a positive continuous function  $v(x, t)$  of  $x \in S$  and  $t \geq 0$  satisfying

$$E\{[X(t + dt) - X(t)]^2 | X(t) = x\} = v(x, t)dt + o(dt).$$