

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

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

(iv) For every  $x \in S$  and every  $t \geq 0$ ,

$$E\{|X(t+dt) - X(t)|^r | X(t) = x\} = o(dt) \text{ if } r \geq 3.$$

Any stochastic process  $\{X(t), t \in T\}$  which has continuous parameter space  $T$  and continuous state space  $S$  and which satisfies condition (i) is called a continuous stochastic process. Notice that condition (i) in Definition 1.7 implies that the probability of any significant change of state is very small in a small time interval (Convergence to 0 faster than the time interval  $dt$ ). That is, with probability one the process  $X(t)$  will not have jumps as time increases. Condition (iii) implies, however, that for any time interval  $dt$ , no matter how small, with positive probability changes do occur. Thus, condition (iii) guarantees that the process is a dynamic process unless it has been absorbed into some absorbing states.

Let  $f(x, y; s, t)$  be the conditional probability density function (pdf) of  $X(t)$  at  $y$  given  $X(s) = x$ . Then, condition (i) can be expressed alternatively as: For any  $\epsilon > 0$ ,

$$\int_{|y-x| \geq \epsilon} f(x, y; t, t+dt) dy = o(dt).$$

To be more precise, we notice that conditions (i)–(iv) are also equivalent to conditions (i), (ii)', (iii)' and (iv), where conditions (ii)' and (iii)' are given by the following:

(ii)'. For any  $\epsilon > 0$ ,

$$\int_{|y-x| \leq \epsilon} (y-x) f(x, y; t, t+dt) dy = m(x, t) dt + o(dt).$$

(iii)'. For any  $\epsilon > 0$ ,

$$\int_{|y-x| \leq \epsilon} (y-x)^2 f(x, y; t, t+dt) dy = v(x, t) dt + o(dt).$$

Condition (iii)' follows easily from the observation that if  $|y-x| > \epsilon \geq 1$ , then  $|y-x|^3 \geq (y-x)^2$  so that

$$\int_{|y-x| \geq \epsilon} (y-x)^2 f(x, y; t, t+dt) dy \leq \int_{|y-x| \geq \epsilon} |y-x|^3 f(x, y; t, t+dt) dy = o(dt);$$

on the other hand, if  $1 \geq |y - x| > \epsilon$ , then  $(y - x)^2 \leq 1$  so that

$$\int_{|y-x| \geq \epsilon} (y - x)^2 f(x, y; t, t + dt) dy \leq \int_{|y-x| \geq \epsilon} f(x, y; t, t + dt) dy = o(dt).$$

Condition (ii)' follows readily from the Schwarz inequality given by:

$$\int_{|y-x| \geq \epsilon} |y - x| f(x, y; t, t + dt) dy \leq \left\{ \int_{|y-x| \geq \epsilon} (y - x)^2 f(x, y; t, t + dt) dy \right\}^{1/2} \times \left\{ \int_{|y-x| \geq \epsilon} f(x, y; t, t + dt) dy \right\}^{1/2}.$$

Using the conditional pdf  $f(x, y; s, t)$ , one may also define the diffusion process as homogeneous iff  $f(x, y; s, t) = f(x, y; 0, t - s) = f(x, y; t - s)$ : That is,  $f(x, y; s, t)$  depends on the times  $s$  and  $t$  only through the difference  $t - s$ . Notice that, in order for the diffusion process to be homogeneous, a precondition is that  $m(x, t) = m(x)$  and  $v(x, t) = v(x)$  must be independent of time  $t$ . In Chaps. 6 and 7, we will provide some general theories of diffusion processes and illustrates its applications in detail.

**Example 1.14. Diffusion approximation of population growth models.** Let  $\{X(t), t \geq 0\}$  denote the number of bacteria at time  $t$  with  $M$  being the maximum population size. Then, under some general conditions, it is shown in Chap. 6 that to the order of  $O(M^{-2})$ ,  $Y(t) = X(t)/M$  follows a diffusion process with state space  $S = [0, 1]$ . For the stochastic logistic growth process, this was proved in [26, 27] by using alternative methods.

**Example 1.15. Diffusion approximation of the Galton–Watson branching processes.** Let  $\{X(t), t \in T = (0, 1, \dots, \infty)\}$  be a Galton–Watson branching process with progeny distribution  $\{p_j, j = 0, 1, \dots, \infty\}$ . Assume that the mean and the variance of the progeny distribution are given respectively by  $1 + \frac{1}{N}\alpha + O(N^{-2})$  and  $\sigma^2$ , where  $N$  is very large. Then it is shown in Example 6.6 that to the order of  $O(N^{-2})$ ,  $Y(t) = X(t)/N$  is a diffusion process with state space  $S = [0, \infty)$  and with coefficients  $\{m(x, t) = x\alpha, v(x, t) = x\sigma^2\}$ .

**Example 1.16. Diffusion approximation of the Wright model in population genetics.** In Example 1.11, we have considered the Wright model in population genetics. In this model,  $\{X(t), t \in T\}$  is the number of  $A$  allele in a large diploid population of size  $N$ , where  $T = \{0, 1, \dots, \infty\}$ . It is shown in

Example 1.11 that this is a Markov chain with state space  $S = \{0, 1, \dots, 2N\}$  and with one-step transition probabilities given by:

$$\Pr\{X_1(t+1) = j | X(t) = i\} = \binom{2N}{j} p_{t+1}^j q_{t+1}^{2N-j},$$

where  $p_{t+1}$  is the frequency of  $A$  allele at generation  $t+1$  and  $q_{t+1} = 1 - p_{t+1}$ .

Denote by  $x = i/2N$ ,  $m(x, t) = (2N)(x - p_{t+1})$ , and

$$v(x, t) = (2N)(x - p_{t+1})^2 + p_{t+1}(1 - p_{t+1}).$$

If  $m(x, t)$  and  $v(x, t)$  are bounded functions of  $x$  and  $t$  for all  $i \geq 0$  and for all  $t \geq 0$ , then, it is shown in Theorem 6.6, that to order of  $O(N^{-2})$ ,  $\{Y(t) = X(t)/(2N), t \geq 0\}$  is a diffusion process with state space  $S = [0, 1]$  and with diffusion coefficients  $\{m(x, t), v(x, t)\}$ .

**Example 1.17. Diffusion approximation of the SIR model in infectious disease.** Consider the SIR model in infectious diseases. Let  $S(t)$  and  $I(t)$  denote the number of  $S$  people and  $I$  people at time  $t$ . Let  $c(t)\Delta t$  be the average number of partners of each  $S$  person during  $[t, t + \Delta t)$ . Let the transition rates of  $I \rightarrow R$  be  $\gamma(t)$  and  $q(t)$  the per partner transmission probability of the disease given contacts between a  $S$  person and an  $I$  person during  $[t, t + \Delta t)$ . Let the death rate and the immigration and recruitment rate of  $I$  people be  $\mu_I(t)$  and  $\nu_I(t)$  respectively. Suppose that the following conditions hold:

- (1) There is no contact between  $S$  people and  $R$  people.
- (2) The population size changes very little over time so that  $S(t) + I(t) = N(t) \sim N$  is approximately independent of time  $t$ .
- (3) There is only one sexual activity level and the mixing pattern is random mixing.

Then it is shown in Example 6.8 that to the order of  $O(N^{-2})$ ,  $\{Y(t) = \frac{I(t)}{N}, t \geq 0\}$  is a diffusion process with state space  $S = [0, 1]$  and with coefficients  $\{m(x, t) = \alpha(t)x(1-x) + \mu_I(t) - x[\gamma(t) + \nu_I(t)], v(x, t) = \alpha(t)x(1-x) + \mu_I(t) + x[\gamma(t) + \nu_I(t)]\}$ , where  $\alpha(t) = c(t)q(t)$ .

**Example 1.18. Diffusion approximation of initiated cancer cells in carcinogenesis.** Consider the MVK two-stage model of carcinogenesis as described in Example 1.13. Let  $N_0$  denote the number of normal stem cells at time 0 and denote by  $X(t) = \frac{1}{N_0}I(t)$ . For large  $N_0$ , it is shown in Example 6.7

that to the order of  $O(N_0^{-2})$ ,  $\{X(t), t \geq 0\}$  is a diffusion process with state space  $\Omega = [0, \infty)$  and with coefficients

$$\{m(x, t) = \alpha_N(t) + x\gamma(t), v(x, t) = \frac{1}{N_0}x\omega(t)\},$$

where  $\gamma(t) = b_I(t) - d_I(t)$ ,  $\omega(t) = b_I(t) + d_I(t)$  and  $\alpha_N(t)$  is the mutation rate from normal stem cells to initiated cells.

#### 1.4. State Space Models and Hidden Markov Models

To validate the stochastic models and to estimate unknown parameters in the model, one usually generates observed data from the system. Based on these data sets, statisticians have constructed statistical models to make inferences about the unknown parameters and to validate the model. To combine information from both the mechanism and the data, the state space model then combines the stochastic model and the statistical model into one model. Thus, the state space model has two sub-models:

- (1) The stochastic system model which is the stochastic model of the system, and
- (2) the observation model which is the statistical model based on some observed data from the system.

**Definition 1.8.** Let  $X(t)$  be a stochastic process with parameter space  $T$  and with state space  $S$ . Let  $\{Y(t_j) = Y_j, j = 1, \dots, n\}$  be the observed values on  $X(t)$  at the time points  $t_1 \leq t_2 \leq \dots \leq t_{n-1} \leq t_n$ . Suppose that  $Y_j = f[X(t), t \leq t_j] + e_j$  for some function  $f()$  of  $X(t), t \leq t_j$ , where  $e_j$  is the random measurement error for measuring  $Y_j$ . Then the combination  $\{X(t), t \in T; Y_j, j = 1, \dots, n\}$  is called a *state space model of the stochastic system* with stochastic system model given by the stochastic process  $\{X(t), t \in T\}$  and with the observation model given by the statistical model  $Y_j = f[X(t), t \leq t_j] + e_j$  for the system. In other word, a state space model of a stochastic system is the stochastic model of the system plus some statistical model based on some observed data from the system.

From this definition, it appears that if some data are available on the system, then one may always construct a state space model for the system. For this state space model, the stochastic process of the system is the stochastic