Chapter 2
Reliability Evaluation Techniques

Rong-Tsorng Wang

2.1 Introduction

From commercial to life-critical applications, the proliferation of computing systems in everyday life has substantially increased our dependence on them. Failures in air traffic control systems, nuclear reactors, or hospital patient monitoring systems can bring catastrophic consequences. In order to enhance the dependability of computing systems, an effective evaluation of their reliability is desired. This chapter presents methods for evaluating system reliability, and indicates that stochastic modeling has provided an effective and unified framework for analyzing various aspects of reliability. The content of this chapter is devoted to combinatorial methods, Markov models, and software models, and equally important is techniques based on the theory of stochastic processes.

Section 2.1.1 reviews some basic notation and terminology of reliability. In Sect. 2.2.1, we discuss combinatorial methods used in reliability analysis, including basic system structures, reliability block diagram, and fault tree analysis. When applying redundancy to improve system reliability, it is commonly assumed that components work independently and their intrinsic failure time distributions have no interdependence. In practical situations, however, the reliability of the system could be significantly affected by components (or subsystems) with dependency, which often occurs in three ways among others: components subjected to the same stresses, components sharing the same load, or components failing with common causes [11]. In each case, the random variables of interest tend to act similarly – degrade the performance and promote the system failure. For example, parallel systems such as the pair of engines of an aeroplane have components whose lifetimes are dependent. Thus, it is more realistic to assume certain forms of positive dependence among components.

R.-T. Wang (✉)
Department of Statistics, Tunghai University, Taichung, Taiwan
e-mail: rtwang@thu.edu.tw

J. Mathew et al. (eds.), *Energy-Efficient Fault-Tolerant Systems*,
DOI 10.1007/978-1-4614-4193-9_2, © Springer Science+Business Media New York 2014

In system reliability analysis, dependence concepts of multivariate models for nonrepairable system have been studied by many authors, e.g., Barlow and Proschan [11], Joe [43], and Balakrishnan and Lai [10]. For bivariate cases, for example, the emphasis is either on the relationship between the joint reliability function $P(T_1 > t_1, T_2 > t_2)$ and the product of marginal reliability functions $P(T_1 > t_1)P(T_2 > t_2)$, or on the conditional reliability functions $P(T_2 > t_2|T_1 > t_1)$. Balakrishnan and Lai [10] investigate various notions of positive dependence between two random variables. In particular, they study dependence concepts that are stronger (or weaker) than *positive quadrant dependence* (PQD), which appears as a straightforward concept of positive dependence and is given by $P(T_1 \geq t_1, T_2 \geq t_2) \geq P(T_1 \geq t_1) \cdot P(T_2 \geq t_2)$.

One technique for characterizing a system having dependent component failure rates is to apply the multi-version of exponential distributions. The most cited bivariate exponential distribution is the one proposed by Marshall and Olkin [67], known as the BVE distribution. The BVE distribution enable us to model common-cause failures, and it is obtained by considering a model in which two components fail separately or simultaneously upon receiving a shock that is governed by a homogeneous Poisson process. Freund [29] proposed a bivariate extension of the exponential distribution by allowing the failure rate of the surviving component to be affected after the failure of another component. Freund's model is one of the first to study bivariate distributions from reliability considerations, and it can be used to model load-sharing systems. In Sect. 2.2.2, we extend the BVE and Freund's model to n-components redundant systems for applying redundant system with common-cause and load-sharing failures.

Fault tree analysis has been used in analyzing safety systems in nuclear power plants, aerospace, and defense. However, dependencies of various types that do occur in practice are not easily captured by fault tree models. In contrast, Markov model is powerful in that it can solve system with dynamic and dependent behaviors, although Markov model has the significant disadvantage that its size grows exponentially as the size of the system increases. From the counting process point of view, the birth process which counts the number of failed components provides a simple representation for the failure process of a system. This approach, however, does not help us to analyze the condition when component lifetimes are dependent. Thus a more general model is required.

In Sect. 2.3, we discuss Markov chains, Poisson processes and nonhomogeneous Poisson processes, and birth-death processes. Continuous-time Markov chains can be used to model both nonrepairable and repairable system. By definition a repairable system is a system which, after failing to perform one or more of its functions satisfactorily, can be restored to fully satisfactory performance by a method other than replacement of the entire system [8]. Commonly used models for a repairable system are renewal processes (perfect repair) and nonhomogeneous Poisson processes (minimal repair) [54].

In Sect. 2.4, we discuss the software reliability models. We aim to extend common assumptions used in statistical software reliability, classify existing software reliability models, and propose mathematical models for describing software

reliability. We begin with the assumptions deduced from Jelinski and Moranda model [42]. Then, we show relations between the transition failure rates of the failure process and the conditional failure rates of times between failures, and specify several common types of software reliability models. A self-exciting and mixture model is then proposed as a generalization of perfect debugging models.

Section 2.5 discusses three examples. In Sect. 2.5.1, we consider a load-sharing system with common-cause failures. Specifically, we formulate reliability properties by introducing a continuous-time Markov chain to characterize systems in which component lifetimes are dependent in two ways: common-cause failures and increased failure rates for surviving components. In Sect. 2.5.2 we discuss a model for a shared-load k-out-of-n:G repairable system studied by Shao and Lamberson [97]. Specifically, a continuous-time Markov chain is used to create a set of differential equations. Section 2.5.3 reviews a bivariate birth-death model for a N-version programming fault-tolerant system proposed by Wang [115].

Effective and optimal techniques may be addressed according to system structures (e.g., hardware or software or integrated systems, binary-state or multi-state systems, parallel or k-out-of-n systems, individual or network systems), failure mechanisms (e.g., independent or common-cause failures), and repair patterns (e.g., repairable or non-repairable, perfect or minimal or general repair) [11, 22, 48, 64, 69, 71, 76, 78, 84, 85, 92, 100, 111]. Dependence concepts of Markov dependence, stationarity, and Martingales have been applied for system failure characterizations. Recent researches indicate that the modern theory of stochastic processes allows for developing a general reliability model that incorporates time dynamics and different information levels. More advanced issues can be found in for examples, Singpurwalla [99], Peña [82], and Aalen et al. [1].

2.1.1 Definitions of Reliability

The lifetime of a component is usually unknown and is characterized by a non-negative random variable (r.v.), say T, representing the time to failure of a system since its inception. Usually T is measured from time $t = 0$. By definition, *reliability* is the probability of a device performing its purpose adequately for the period of time intended under the operating conditions encountered [11]. (Failures can be defined in terms of degradation reaching a certain level, see, e.g., Meeker and Escobar [69].) Let $F(t) = P(T \leq t)$ be the cumulative distribution function (c.d.f.) of T such that $F(0-) = 0$ and $F(+\infty) = 1$. The *reliability function* (or survival function) of T is defined by

$$R(t) := \bar{F}(t) = P(T > t), \quad \forall t \geq 0. \tag{2.1}$$

For example, the reliability required for aircraft control systems has been specified as $R(t) > 1 - 10^{-9}$ for $t = 1\,\text{h}$. The corresponding conditional reliability function (or conditional survival function) of T of age x is given by

$$R(t|x) := \frac{\bar{F}(x+t)}{\bar{F}(x)}, \quad \forall t, x \geq 0,$$

provided that $\bar{F}(x) > 0$. The expected time of T, also known as the *mean time to failure* (MTTF), is given by

$$E[T] = \int_0^\infty \bar{F}(t)dt = \int_0^\infty R(t)dt. \tag{2.2}$$

In addition, the second moment of T is given by

$$E[T^2] = 2\int_0^\infty t \cdot R(t)dt.$$

The *mean residual life* of T of age x is given by

$$E[T - x|T > x] = \frac{\int_x^\infty \bar{F}(t)dt}{\bar{F}(x)} = \int_0^\infty \frac{\bar{F}(t+x)dt}{\bar{F}(x)} = \int_0^\infty R(t|x)dt.$$

Equation (2.1) may or may not continuous in t. Suppose that Eq. (2.1) is everywhere differentiable, then the probability density function (p.d.f.) $f(t)$ of T exists for all $t \geq 0$, and

$$f(t) \cdot \Delta t \approx \Pr(t \leq T < t + \Delta t),$$

where Δt is a small increment of t. The *failure rate* (or conditional failure rate) of the r.v. T at time t is defined as

$$h(t) := \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} \Pr(t < T \leq t + \Delta t | T > t), \qquad t \geq 0. \tag{2.3}$$

$h(t)$ gives the probability that a component of age t will almost immediately fail. Since, for those t where $h(t)$ is continuous,

$$h(t) = \frac{f(t)}{\bar{F}(t)}, \quad \forall t \geq 0. \tag{2.4}$$

$h(t)$ is also known as the hazard rate, the force of mortality, and the intensity rate. The functions $f(t)$, $F(t)$, $\bar{F}(t)$, and $h(t)$ are mathematically equivalent, since

$$\int_0^t h(x)dx = -\ln \bar{F}(t),$$

$$\bar{F}(t) = \exp\left\{-\int_0^t h(x)dx\right\}.$$

The function $\int_0^t h(x)dx$ is called the *cumulative hazard rate* function. Many non-normal univariate distributions have been used in reliability, such as exponential, Gamma, Weibull and so forth. In particular, these distributions are applied to three phases of the so-called bathtub curve: the infant mortality phase, the normal operating phase, and the wear-out phase. The most commonly used distribution is the exponential distribution.

Example 2.1. If the failure rate is given by $h(t) = \beta^2 t/(1 + \beta t)$, where $\beta > 0$, $t > 0$. Then

$$\bar{F}(t) = \exp\left\{-\int_0^t h(x)dx\right\} = \exp\left(-\int_0^t \left[\beta - \frac{\beta}{1+\beta x}\right] dx\right) = (1+\beta t)e^{-\beta t}.$$

Observed that T has a gamma distribution. The MTTF is

$$E[T] = \int_0^\infty \bar{F}(t)dt = \int_0^\infty (1+\beta t)e^{-\beta t} = \frac{2}{\beta}.$$

Example 2.2. If T has a Weibull distribution with *shape* parameter $\beta > 0$ and *scale* parameter $\eta > 0$, whose c.d.f. is given by

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^\beta\right], \quad t > 0.$$

The failure rate is

$$h(t) = \frac{d}{dt}\left[-\ln \bar{F}(t)\right] = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}.$$

The hazard rate is increasing for $\beta > 1$, decreasing for $0 < \beta < 1$, and constant for $\beta = 1$.

2.2 Combinatorial Methods

System failure is modeled in terms of the failures of the components of the system. Both the system and its components are often allowed to take only two possible states: a working state and a failed state. The linking of component failures to system failures can be understood in several ways. Among these are the *reliability block diagram* (success-oriented) and the *fault tree analysis* (failure-oriented). In some cases, it is possible to convert the fault tree to a reliability block diagram, and vice versa. In Sect. 2.2.1, we review basic system structures, reliability block diagram, and fault tree analysis.

Parallel redundancy is a common method to increase system reliability and MTTF. Usually components in a reliability structure are dependent as they may share the same load or may be failed with common-cause failures. Bivariate and multivariate lifetime distributions play important roles in modeling these dependencies. Many bivariate and multivariate exponential distributions have been proposed [10], however, only a certain class of multivariate distributions is applicable for reliability systems. The BVE distribution of Marshall and Olkin [67] is suited for modeling common-cause failures. Freund's model [29] can be applied to the situation that the failed component increases the stress on the surviving component and consequently increases the other component's tendency of failure. In Sect. 2.2.2, we extend the BVE and Freund's distributions to n-components redundant systems, and discuss the reliability functions of series, parallel, and k-out-of-n:F structures.

2.2.1 Independent Components

2.2.1.1 Three Basic Structures

Consider a system with n independent components, and the system reliability $R(t)$ can be determined by components' reliability $R_i(t)$. We write $R(t)$ a function of $R_1(t), \ldots, R_n(t)$ as

$$R(t) = \psi(R_1(t), \ldots, R_n(t)).$$

The function ψ is decided by the structure of the system. We often use *reliability block diagram* or *network diagram* to describe the system reliability structure. Let $T_1, \ldots, T_n$ be the lifetimes of components. Assume that $T_1, \ldots, T_n$ are independent, there are three common structures:

$$\text{(1) series: } R(t) = \prod_{i=1}^{n} R_i(t), \tag{2.5}$$

$$\text{(2) parallel: } R(t) = 1 - \prod_{i=1}^{n} (1 - R_i(t)), \tag{2.6}$$

$$\text{(3A) } k\text{-out-of-}n\text{:F: } R(t) = \sum_{i=0}^{k-1} \binom{n}{i} [R_1(t)]^{n-i} [1 - R_1(t)]^i, \quad \text{if } R_i(t) = R_1(t), \tag{2.7}$$

$$\text{(3B) } k\text{-out-of-}n\text{:G: } R(t) = \sum_{i=k}^{n} \binom{n}{i} [R_1(t)]^{i} [1 - R_1(t)]^{n-i}, \quad \text{if } R_i(t) = R_1(t). \tag{2.8}$$

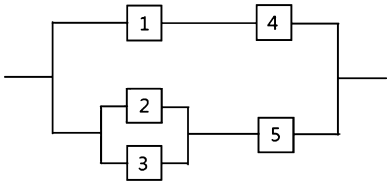


Fig. 2.1 A reliability block diagram of a complex system

A series system functions if and only if each component functions, and a parallel system functions if and only if at least one component functions. A system is said to be a k-out-of-n:F system if it fails if and only if at least k of the n components fail. A dual concept called k-out-of-n:G system is defined as that it is good if and only if at least k of n components are good. A system is a k-out-of-n:F system if and only if it is a $(n-k+1)$-out-of-n:G system. Likewise, a system is a k-out-of-n:G system if and only if it is a $(n-k+1)$-out-of-n:F system. Apparently,

$$\text{a system is 'series'} \iff \text{it is a 1-out-of-}n\text{:F system (or } n\text{-out-of-}n\text{:G)},$$

$$\text{a system is 'parallel'} \iff \text{it is an } n\text{-out-of-}n\text{:F system (or 1-out-of-}n\text{:G)}.$$

The reliability of a k-out-of-n:F system in (2.7) with independently and identically distributed (i.i.d.) components is equal to the probability that the number of failing components is less than or equal to $k-1$. As a k-out-of-n:F system is equivalent to a $(n-k+1)$-out-of-n:G system, (2.7) is equivalent to

$$\sum_{j=n-k+1}^{n} \binom{n}{j} [R_1(t)]^j [1-R_1(t)]^{n-j}.$$

If we denote $R_G(n,k)$ the reliability of a k-out-of-n:G system and $R_F(n,k)$ the reliability of a k-out-of-n:F system, we have

$$R_G(n,k) = R_F(n, n-k+1).$$

2.2.1.2 Reliability Block Diagram

A reliability block diagram is a success-oriented diagram describing a specified function of a system. Each rectangular block in the diagram represents a function of a component. If the function is available, we have connection through the block, and if the function is failed, there is no connection through the block.

Consider the reliability block diagram in Fig. 2.1. The system is a parallel structure of two independent modules: the structure comprising components 1 and 4,

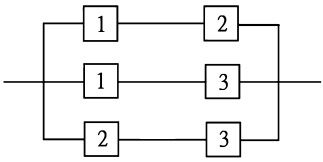


Fig. 2.2 A reliability block diagram for a 2-out-of-3:G system

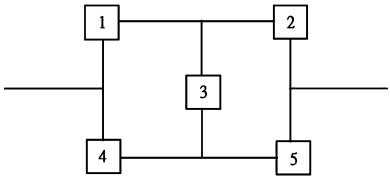


Fig. 2.3 A bridge structure

and the structure comprising components 2, 3, and 5. From (2.5) and (2.6), the reliability of the former structure equals R_1R_4, whereas the reliability of the latter equals $[1-(1-R_2)(1-R_3)]R_5$. Thus the system reliability is given by

$$R_{system} = 1-(1-R_1R_4)\{1-[1-(1-R_2)(1-R_3)]R_5\}.$$

A reliability block diagram of 2-out-of-3:G system is illustrated in Fig. 2.2.

2.2.1.3 A Bridge Structure

The reliability block diagram of a bridge configuration is given in Fig. 2.3. Each block in the diagram denotes a component. For independent components, the system reliability of the bridge structure can be obtained by using *minimal path* or *minimal cut sets* [11] and is expressed by

$$\begin{aligned} R_{system} = {} & R_1R_2 + R_4R_5 + R_1R_3R_5 + R_2R_3R_4 \\ & -R_1R_2R_3R_4 - R_1R_2R_3R_5 - R_1R_2R_4R_5 - R_1R_3R_4R_5 \\ & -R_2R_3R_4R_5 + 2R_1R_2R_3R_4R_5. \end{aligned}$$

The minimal path sets of the bridge structure are $\{1,2\}$, $\{4,5\}$, $\{1,3,5\}$, and $\{2,3,4\}$. The minimal path sets may represent the bridge as a parallel-series structure. The minimal cut sets are $\{1,4\}, \{2,5\}, \{1,3,5\}$, and $\{2,3,4\}$. The minimal cut sets may represent the bridge as a series-parallel structure. Alternative method to obtain the system reliability of a bridge structure is the *pivot-decomposition method* (or factoring method) [123]. The minimal path sets and minimal cut sets for the bridge structure are given in Fig. 2.4.

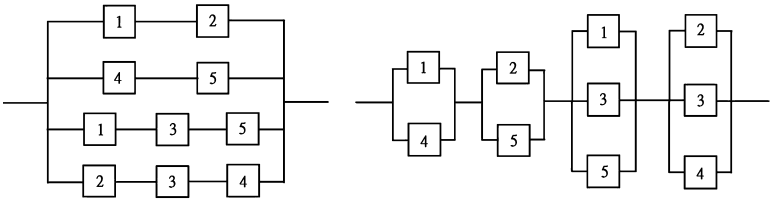


Fig. 2.4 `Minimal path and minimal cut representations for bridge`

2.2.1.4 A Non-i.i.d. *k*-out-of-*n*:F System

The reliability of a non-i.i.d. k-out-of-n:F system is given by

$$R(t) = \sum_{i=0}^{k-1} \sum_{j=1}^{\binom{n}{i}} \prod_{m=1}^{n} R_m^{\delta_{m,j}} (1 - R_m)^{\bar{\delta}_{m,j}}, \quad k = 1, \ldots, n,$$

where $\delta_{m,j}$ and $\bar{\delta}_{m,j}$ are complementary indicator functions for which

$$\sum_{m=1}^{n} \delta_{m,j} = n - k, \qquad \sum_{m=1}^{n} \bar{\delta}_{m,j} = k,$$

and

$$\delta_{m,j} = \begin{cases} 1 & \text{if the } m\text{th component functions} \\ 0 & \text{if the } m\text{th component fails.} \end{cases}$$

Myers [77] has studied extensively about k-out-of-n:G systems of this type.

Example 2.3. The reliability of a non-i.i.d. 3-out-of-4:F system is given by

$$\begin{aligned} R &= \sum_{i=0}^{2} \sum_{j=1}^{\binom{4}{i}} \prod_{m=1}^{4} R_m^{\delta_{m,j}} (1 - R_m)^{\bar{\delta}_{m,j}} \\ &= R_1R_2R_3R_4 + R_1R_2R_3(1 - R_4) + R_1R_2R_4(1 - R_3) + R_1R_3R_4(1 - R_2) \\ &\quad + R_2R_3R_4(1 - R_1) + R_1R_2(1 - R_3)(1 - R_4) + R_1R_3(1 - R_2)(1 - R_4) \\ &\quad + R_1R_4(1 - R_2)(1 - R_3) + R_2R_3(1 - R_1)(1 - R_4) \\ &\quad + R_2R_4(1 - R_1)(1 - R_3) + R_3R_4(1 - R_1)(1 - R_2). \end{aligned}$$

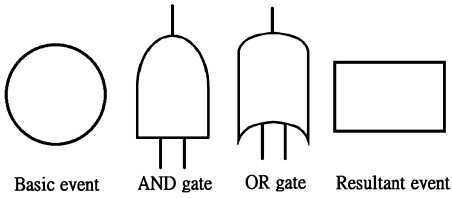


Fig. 2.5 Common fault tree symbols

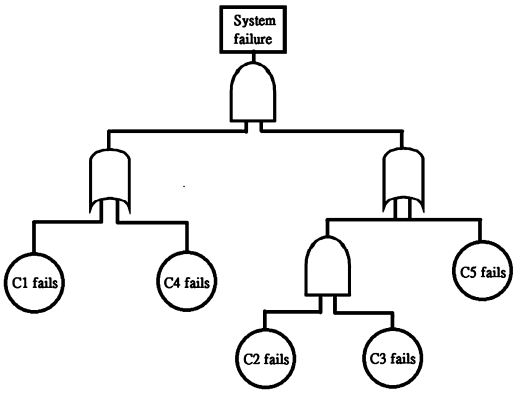


Fig. 2.6 A fault tree diagram

2.2.1.5 Fault Tree Analysis

Fault tree analysis is a widely used method in the industrial sector to perform reliability analysis of systems during their design and development, particularly in system safety analysis. A *fault tree diagram* is the underlying top-down graphical model in fault tree analysis. The starting point of a fault tree is a system failure, known as a *top event*. Fault events that can cause the occurrence of the top event are generated and connected by logic operators such as 'AND' and 'OR'. The status of output/top event can be derived by the status of input events and the connections of the logical gates. The logical meaning of AND(OR) gate is the same as that of the intersection(union) of events. That is, the AND gate provides a True output (fault) if all its inputs are True (fault) and the OR gate provides a True output (fault) if one or more of its inputs are True (fault). In other words, an AND gate in a fault tree is logically equivalent to a parallel reliability block diagram, and an OR gate corresponds to a series reliability block diagram. The fault tree is developed by repeatedly asking which combinations of the component failures will result in a system failure. Fault events are generated successively until the fault events need not be developed any further. These lowest level causes are called *basic events*.

Four commonly used symbols in the construction of fault trees are shown in Fig. 2.5. A fault tree diagram corresponding to Fig. 2.1 is shown in Fig. 2.6.

2.2.1.6 Order Statistics

A r.v. T with d.f. F is said to be exponential distributed with parameter λ, denoted by $T \sim \text{Exp}(\lambda)$ or $F \sim \text{Exp}(\lambda)$, if the reliability function is given by $R(t) = e^{-\lambda t}$ for $t > 0, \lambda > 0$.

Suppose that $T_1, \ldots, T_n$ are i.i.d. r.v.'s with d.f. F. We index the failure times chronologically by $0 \leq T_{1:n} \leq \cdots \leq T_{n:n}$, where $T_{k:n}$ denotes the occurrence time of the kth failure, for $k = 1, \ldots, n$. That is, $T_{1:n}, \ldots, T_{n:n}$ are the corresponding *order statistics* of $T_1, \ldots, T_n$. We also denote the *spacings* by, for $k = 1, \ldots, n$, $D_k = T_{k:n} - T_{k-1:n}$, where $T_{0:n} = 0$. That is, D_k denotes the time between the $(k-1)$th and the kth failure. The reliability of a system can then be studied through the order statistics and spacings as an *general order statistics* model [90]. In particular, if $F \sim \text{Exp}(\lambda)$ the model is called an i.i.d. *exponential order statistics* [70]. Moreover the kth order statistic $T_{k:n}$ can be considered as the time of failure of a k-out-of-n:F system [11]. In the case that $T_1, \ldots, T_n$ are i.i.d. r.v.'s with d.f. F, the reliability function of a k-out-of-n:F system at time t is given by

$$P(T_{k:n} > t) = \sum_{i=0}^{k-1} \binom{n}{i} [F(t)]^i [1 - F(t)]^{n-i}. \tag{2.9}$$

Example 2.4. Let $T_1, \ldots, T_n$ be i.i.d. r.v.'s with d.f. F. If $F \sim Exp(\lambda)$, then the spacings D_k are independent exponential distributions with parameters $(n-k+1)\lambda$, for $k = 1, \ldots, n$ [11]. Since $T_{k:n} = \sum_{i=1}^{k} D_i$, the p.d.f. of $T_{k:n}$ is given by

$$f_{T_{k:n}}(t) = \beta_1 \ldots \beta_k \sum_{i=1}^{k} \left[\Psi_{i,k} e^{-\beta_i t} \right], \tag{2.10}$$

where

$$\begin{aligned} \beta_i &= (n - i + 1)\lambda, \\ \Psi_{i,k}^{-1} &= (\beta_1 - \beta_i) \ldots (\beta_{i-1} - \beta_i)(\beta_{i+1} - \beta_i) \ldots (\beta_k - \beta_i). \end{aligned}$$

The distributions of $T_{k:n}$ are Erlang distributions. The reliability functions of $T_{k:n}$ are more conveniently obtained from Eq. (2.9) than from the p.d.f. (2.10). Namely, for $k = 1, \ldots, n$,

$$P(T_{k:n} > t) = \sum_{i=0}^{k-1} \binom{n}{i} (1 - e^{-\lambda t})^i (e^{-\lambda t})^{n-i}.$$

In particular, the reliability functions of series and parallel systems are, respectively,

$$P(T_{1:n} > t) = e^{-n\lambda t} \quad \text{and} \quad P(T_{n:n} > t) = 1 - (1 - e^{-\lambda t})^n.$$

2.2.1.7 Linear Combinations of Reliability of Series Systems

Consider a parallel system of n components. Let T and T_i be the lifetimes of the parallel system and the component i, respectively, and consider the events $A = \{T > t\}$ and $A_i = \{T_i > t\}$. Then $A = \bigcup_{i=1}^{n} A_i$. Apparently, the sum of the component reliabilities gives an upper bound of the parallel system reliability

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

To evaluate $P(A)$, apart from (2.6), we consider the inclusion-exclusion identity:

$$P(A) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right). \quad (2.11)$$

Here

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(T_{i_1} > t, T_{i_2} > t, \ldots, T_{i_k} > t) = P(U > t),$$

where U is the lifetime of the series system of components $i_1, i_2, \ldots, i_k$. Note that the series system reliability has a lower bound

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1),$$

known as *Bonferroni's inequality*.

Equation (2.11) gives that the reliability of a parallel system can be expressed as a linear combination of the reliability of all possible series systems. Taking integration in (2.11) and applying (2.2), we have that the MTTF of parallel systems can be found as a linear combination of the MTTF of all possible series systems. Rade [89] gives the following theorem.

Theorem 2.1. *Let μ be the MTTF of a parallel system of n components and let $\mu_{i_1 i_2 \ldots i_k}$ be the MTTF of the series system of components $i_1, i_2, \ldots, i_k$, assuming the joint distribution of the components' lifetime is the same as in the parallel system. Then*

$$\mu = \sum_{i=1}^{n} \mu_i - \sum_{i<j} \mu_{ij} + \sum_{i<j<k} \mu_{ijk} - \cdots + (-1)^{n+1} \mu_{12\ldots n}. \quad (2.12)$$

If all series systems of k components have the same MTTF $\mu_{(k)}$, then

$$\mu = \binom{n}{1}\mu_{(1)} - \binom{n}{2}\mu_{(2)} + \binom{n}{3}\mu_{(3)} - \cdots + (-1)^{n+1}\binom{n}{n}\mu_{(n)}. \tag{2.13}$$

Rade [89] also generalizes Theorem 2.1 to a series-parallel system and a k-out-of-n:G system.

Example 2.5. Let the lifetimes $T_1, T_2, \ldots, T_n$ of the components be independent and exponentially distributed with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then (2.12) gives

$$\mu = \sum_{i=1}^{n} \frac{1}{\lambda_i} - \sum_{i<j} \frac{1}{\lambda_i + \lambda_j} + \cdots + (-1)^{n+1} \frac{1}{\lambda_1 + \cdots + \lambda_n}.$$

In this case, the lifetime of the series system of components with intensities $\lambda_1, \lambda_2, \ldots, \lambda_k$ has an exponential distribution with parameter $\lambda_1 + \cdots + \lambda_k$ and mean $1/(\sum_{j=1}^{k} \lambda_j)$. If, furthermore, all λ_i are equal to λ, the formula (2.13) gives

$$\mu = \frac{1}{\lambda} \sum_{j=1}^{n} \frac{(-1)^{j+1}\binom{n}{j}}{j} = \frac{1}{\lambda} \sum_{j=1}^{n} \frac{1}{j}.$$

A direct calculation from (2.6) yields

$$\mu = \int_0^\infty \left[1 - (1 - e^{-\lambda t})^n\right] dt = \frac{1}{\lambda} \sum_{j=1}^{n} \frac{1}{j}.$$

For large n, μ is approximately $\frac{1}{\lambda}(\ln n + \gamma)$, where $\gamma = 0.577215664\ldots$ is Euler's constant. If the system is a k-out-of-n:G system, the mean time to failure of the system is

$$\int_0^\infty \left[\sum_{j=k}^{n} \binom{n}{j} e^{-j\lambda t} (1 - e^{-\lambda t})^{n-j} \right] dt = \frac{1}{\lambda} \sum_{j=k}^{n} \frac{1}{j}.$$

2.2.2 Dependent Components

2.2.2.1 The Model of Marshall and Olkin

Marshall and Olkin [67] introduced a bivariate exponential distribution (BVE) by considering a reliability model in which two components fail separately or simultaneously upon receiving a shock that is governed by a homogeneous Poisson process. They derived the BVE in several ways: the bivariate lack of memory property, shock models, a random sum model, and a minima model. The BVE distribution is the

most important bivariate exponential distribution. It contains many properties which inspire intensive investigations of the bivariate exponential distributions.

We say that (T_1, T_2) has the BVE distribution if the joint reliability function is given by, for $t_1 \geq 0$ and $t_2 \geq 0$,

$$\bar{F}(t_1, t_2) = e^{-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} \max(t_1, t_2)},$$

where $\lambda_1, \lambda_2, \lambda_{12}$ are non-negative parameters such that $\lambda_1 + \lambda_{12} > 0, \lambda_2 + \lambda_{12} > 0$. T_1 and T_2 are independent when $\lambda_{12} = 0$. For $0 < t_1 < t_2$ and $0 < t_2 < t_1$ the density can be obtained by $\partial^2 \bar{F}(t_1, t_2)/\partial t_1 \partial t_2$, and then for $t > 0$

$$P(t < T_1 < T_2) = \frac{\lambda_1}{\lambda} e^{-\lambda t},$$

$$P(t < T_2 < T_1) = \frac{\lambda_2}{\lambda} e^{-\lambda t},$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$. It follows that $P(T_1 = T_2 > t) = \frac{\lambda_{12}}{\lambda} e^{-\lambda t}$. So (T_1, T_2) has density f on the region $\{(t_1, t_2) : t_1 > 0, t_2 > 0, t_1 \neq t_2\}$ with respect to two-dimensional Lebesgue measure, and a density g on the line $\{(t, t) : t > 0\}$ in the sense that $P(T_1 = T_2 \leq t) = \int_0^t g(u) du$ with respect to one-dimensional Lebesgue measure. Density functions f and g are given by

$$f(t_1, t_2) = \begin{cases} \lambda_1(\lambda_2 + \lambda_{12}) e^{-\lambda_1 t_1 - (\lambda_2 + \lambda_{12}) t_2}, & 0 < t_1 < t_2 \\ \lambda_2(\lambda_1 + \lambda_{12}) e^{-(\lambda_1 + \lambda_{12}) t_1 - \lambda_2 t_2}, & 0 < t_2 < t_1, \end{cases} \tag{2.14}$$

and

$$g(t) = \lambda_{12} e^{-\lambda t}. \tag{2.15}$$

In addition, $P(T_1 = T_2) = \int_0^\infty g(t) dt = \frac{\lambda_{12}}{\lambda} \geq 0$.

The Laplace transform $\psi(s, t) = E(e^{-sT_1 - tT_2})$ is given by, from (2.14) and (2.15),

$$\psi(s, t) = \frac{(\lambda_1 + \lambda_{12})(\lambda_2 + \lambda_{12})(\lambda + s + t) + st\lambda_{12}}{(s + t + \lambda)(s + \lambda_1 + \lambda_{12})(t + \lambda_2 + \lambda_{12})}.$$

It follows that the marginal expectations, variances, and correlation coefficient can be obtained. In particular, the correlation coefficient $\rho(T_1, T_2) = \frac{\lambda_{12}}{\lambda} \geq 0$.

For the n-dimensional multivariate exponential distribution (MVE), the joint reliability function is given by

$$\bar{F}(t_1, \ldots, t_n) = \exp\left\{ - \left[\sum_{i=1}^n \lambda_i t_i + \sum_{i<j}^n \lambda_{ij} \max(t_i, t_j) \right.\right.$$
$$\left.\left. + \sum_{i<j<k}^n \lambda_{ijk} \max(t_i, t_j, t_k) + \cdots + \lambda_{1\ldots n} \max(t_1, \ldots, t_n) \right] \right\}, \tag{2.16}$$

where $t_i \geq 0$ and the $2^n - 1$ different λ-symbols are non-negative constants. By setting $t_i = 0$, we have a $(n-1)$-dimensional MVE distribution. Hence the marginal distributions of all order are MVE. For example,

$$\bar{F}(t_1, t_2, 0) = \exp\left\{-\left[\sum_{i=1}^{2}(\lambda_i + \lambda_{i3})t_i + (\lambda_{12} + \lambda_{123})\max(t_1, t_2)\right]\right\}$$

is a BVE.

We next evaluate the reliability functions for a k-out-of-n:F system whose components lifetimes are characterized by the MVE. Recall in Sect. 2.2.1.2 that the time of failure of a k-out-of-n:F system may be represented in terms of the kth order statistic, i.e., $R_{k:n}(t) = P(T_{k:n} > t)$. The system reliability functions for 1-out-of-2:F and 2-out-of-2:F are $R_{1:2}(t) = e^{-\lambda t}$ and $R_{2:2}(t) = e^{-(\lambda_1+\lambda_{12})t} + e^{-(\lambda_2+\lambda_{12})t} - e^{-\lambda t}$, respectively, where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$. The MVE distribution gives a formidably large task to get the system reliability functions, as there are $2^n - 1$ different $\lambda's$ in Eq. (2.16). For simplicity, we consider the case of identical component lifetime distributions. Let

$$\begin{aligned}
\lambda_i &= \beta_1, \quad i = 1, \ldots, n, \\
\lambda_{ij} &= \beta_2, \quad i, j = 1, \ldots, n,\ i \neq j, \\
&\vdots \\
\lambda_{1\ldots n} &= \beta_n,
\end{aligned}$$

then the joint reliability function (2.16) becomes

$$\bar{F}(t_1, \ldots, t_n) = \exp\left\{-\left[\beta_1 \sum_{i=1}^{n} t_i + \beta_2 \sum_{i<j}^{n} \max(t_i, t_j) + \beta_3 \sum_{i<j<k}^{n} \max(t_i, t_j, t_k) + \cdots + \beta_n \max(t_1, \ldots, t_n)\right]\right\}. \quad (2.17)$$

The system reliability functions for 1-out-of-n:F and n-out-of-n:F are, respectively,

$$R_{1:n}(t) = \bar{F}(t, \ldots, t) = \exp\left\{-\left[\binom{n}{1}\beta_1 + \binom{n}{2}\beta_2 + \cdots + \binom{n}{n}\beta_n\right]t\right\},$$

and

$$\begin{aligned}
R_{n:n}(t) &= P\left(\bigcup_{k=1}^{n}\{T_k > t\}\right) \\
&= \sum_{k=1}^{n} P(T_k > t) - \sum_{1\leq i_1\leq i_2\leq n} P(T_{i_1} > t, T_{i_2} > t) + \cdots + (-1)^{n-1} P\left(\bigcap_{k=1}^{n}\{T_k > t\}\right).
\end{aligned}$$

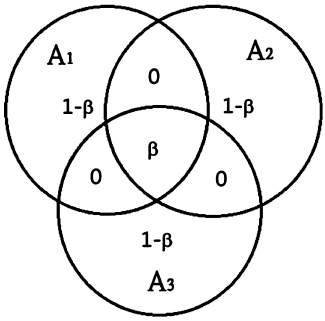


Fig. 2.7 Failure events for a system with three identical components when using the beta-factor model

From (2.17), for $k = 1, \ldots, n$, we have

$$
\begin{aligned}
&P(T_1 > t, \ldots, T_k > t) \\
&\quad = P(T_1 > t, \ldots, T_k > t, T_{k+1} = 0, \ldots, T_n = 0) \\
&\quad = \exp\left\{-\left[\beta_1 k t + \beta_2 \left(\binom{n}{2} - \binom{n-k}{2}\right) t + \beta_3 \left(\binom{n}{3} - \binom{n-k}{3}\right) t + \cdots + \beta_n t\right]\right\} \\
&\quad = \exp\left\{-\sum_{i=0}^{n-1} \sum_{j=1}^{k} \binom{n-j}{i} \beta_{i+1} t\right\},
\end{aligned}
$$

where we use the equality

$$
\sum_{j=1}^{k} \binom{n-j}{r} = \binom{n}{r+1} - \binom{n-k}{r+1}, \quad r = 1, \ldots, n-1,
$$

where $\binom{x}{y} = 0$ if $x < y$. Hence

$$
R_{n:n}(t) = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \cdot \exp\left\{-\sum_{i=0}^{n-1} \sum_{j=1}^{k} \binom{n-j}{i} \beta_{i+1} t\right\}.
$$

Reliability modeling of common cause failures was introduced in the nuclear power industry more than 30 years ago [108]. According to NEA [79] a common cause failure is defined as "a dependent failure in which two or more component fault states exist simultaneously, or within a short time interval, and are a direct result of a shared cause." Common cause failures are especially important for redundant components, and they may be classified in two main types: (i) multiple failures that occur at the same time due to a common cause, and (ii) multiple failures that occur due to a common cause, but not necessarily at the same time.

The *beta-factor model* is the most commonly used common cause failure model. Figure 2.7 gives failure events for a system with three identical components

when using the beta-factor model. Extensions of the beta-factor model have been suggested, for example, the binomial failure rate model, the alpha factor model, the multiple Greek letter model, and the multiple beta-factor model [38].

As for time-varying failure rates, there are considerable studies in the topic of common cause failures for the system reliability modeling and analysis, such as models apply to systems with exponential time-to-failure distributions [7, 17, 28]; components being affected by at most a single common-cause [105, 110]; a single common-cause affects all components of a system [5]. Zhang and Horigome [125] present a method for analyzing availability & reliability of repairable systems with common cause failures among components. Xing [117] studies the reliability analysis of computer network systems by allowing for multiple common cause that can affect different subsets of system components.

2.2.2.2 Freund's Model

Freund [29] proposed a bivariate extension of the exponential distribution by allowing the failure rate of the surviving component to be affected after the failure of another component. Freund's bivariate distribution is absolutely continuous and possesses the bivariate lack of memory property [12]. Suppose components T_1 and T_2 follow independently exponential distributions with parameters α_1 and α_2, respectively, until one of components fails. When component $T_1(T_2)$ has failed, the parameter of the surviving component $T_2(T_1)$ changes to $\alpha_2'(\alpha_1')$. T_1 and T_2 are not independent, since the failure of one component changes the parameter of the surviving one. Consequently, the joint density of (T_1, T_2) is

$$f(t_1,t_2)=\begin{cases}\alpha_1\alpha_2' e^{-(\alpha_1+\alpha_2-\alpha_2')t_1-\alpha_2' t_2}, & 0<t_1<t_2\\ \alpha_2\alpha_1' e^{-\alpha_1' t_1-(\alpha_1+\alpha_2-\alpha_1')t_2}, & 0<t_2<t_1.\end{cases} \tag{2.18}$$

For convenience, we call this distribution the FBVE.

The joint reliability function of (T_1, T_2) is, for $0 < t_1 < t_2$,

$$\begin{aligned}\bar{F}(t_1,t_2)&=\int_{t_2}^{\infty}\int_{t_1}^{t_2}\alpha_1\alpha_2' e^{-(\alpha_1+\alpha_2-\alpha_2')x-\alpha_2' y}dxdy+\bar{F}(t_2,t_2)\\ &=\frac{\alpha_1}{\alpha_1+\alpha_2-\alpha_2'}e^{-(\alpha_1+\alpha_2-\alpha_2')t_1-\alpha_2' t_2}+\frac{\alpha_2-\alpha_2'}{\alpha_1+\alpha_2-\alpha_2'}e^{-(\alpha_1+\alpha_2)t_2}.\end{aligned}$$

It is similar for $0 < t_2 < t_1$. Thus the joint reliability function of the FBVE is given by

$$\bar{F}(t_1,t_2)=\begin{cases}\kappa_1 e^{-(\alpha_1+\alpha_2-\alpha_2')t_1-\alpha_2' t_2}+(1-\kappa_1)e^{-(\alpha_1+\alpha_2)t_2}, & t_1<t_2\\ \kappa_2 e^{-(\alpha_1+\alpha_2-\alpha_1')t_2-\alpha_1' t_1}+(1-\kappa_2)e^{-(\alpha_1+\alpha_2)t_1}, & t_2<t_1,\end{cases} \tag{2.19}$$

where $\kappa_1=\frac{\alpha_1}{\alpha_1+\alpha_2-\alpha_2'}$ and $\kappa_2=\frac{\alpha_2}{\alpha_1+\alpha_2-\alpha_1'}$. T_1 and T_2 are independent when $\alpha_1'=\alpha_1$ and $\alpha_2'=\alpha_2$.

The Laplace transform $\psi(s,t)$ is given by

$$\psi(s,t) = E(e^{-sT_1 - tT_2}) = (\alpha_1 + \alpha_2 + s + t)^{-1}\left[\frac{\alpha_1'\alpha_2}{\alpha_1' + s} + \frac{\alpha_1\alpha_2'}{\alpha_2' + t}\right].$$

We remark that the moment generating function $E(e^{sT_1+tT_2})$ given by Freund [29] was wrong. It follows that the marginal expectations, variances, and correlation coefficient can be obtained. In particular, the correlation coefficient is

$$\rho(T_1, T_2) = \frac{\alpha_1'\alpha_2' - \alpha_1\alpha_2}{\sqrt{(\alpha_1'^2 + 2\alpha_1\alpha_2 + \alpha_2^2)(\alpha_2'^2 + 2\alpha_1\alpha_2 + \alpha_1^2)}}.$$

It can be shown that $-1/3 < \rho(T_1, T_2) < 1$.

The marginals are not exponentials; each is a mixture of exponentials

$$f_1(t_1) = (1-\kappa_2)(\alpha_1+\alpha_2)e^{-(\alpha_1+\alpha_2)t_1} + \kappa_2\alpha_1' e^{-\alpha_1' t_1},$$
$$f_2(t_2) = (1-\kappa_1)(\alpha_1+\alpha_2)e^{-(\alpha_1+\alpha_2)t_2} + \kappa_1\alpha_2' e^{-\alpha_2' t_2},$$

provided that $\alpha_1 + \alpha_2 - \alpha_1' \neq 0$ and $\alpha_1 + \alpha_2 - \alpha_2' \neq 0$. The marginal density $f_i(t_i)$ contains various parameter ranges $\alpha_1' < \alpha_1$, $\alpha_1' = \alpha_1$, $\alpha_1 < \alpha_1' < \alpha_1 + \alpha_2$ and $\alpha_1 + \alpha_2 < \alpha_1'$. If $\alpha_1 + \alpha_2 - \alpha_1' = 0$ and $\alpha_1 + \alpha_2 - \alpha_2' = 0$, then

$$f_1(t_1) = (\alpha_1 + \alpha_1'\alpha_2 t_1)e^{-\alpha_1' t_1},$$
$$f_2(t_2) = (\alpha_2 + \alpha_2'\alpha_1 t_2)e^{-\alpha_2' t_2}.$$

Thus $f_i(t_i)$ is a mixture of an exponential density $g_i(t_i)$ and a gamma density $g_i^*(t_i)$, i.e.,

$$f_1(t_1) = \frac{\alpha_1}{\alpha_1'} g_1(t_1) + \frac{\alpha_2}{\alpha_1'} g_1^*(t_1),$$

where $g_1(t_1) = \alpha_1' e^{-\alpha_1' t_1}$ and $g_1^*(t_1) = \alpha_1'^2 t_1 e^{-\alpha_1' t_1}$.

From (2.19), the reliability functions for 1-out-of-2:F and 2-out-of-2:F systems are, respectively, $R_{1:2}(t) = e^{-(\alpha_1+\alpha_2)t}$ and

$$R_{2:2}(t) = \begin{cases} \kappa_1 e^{-\alpha_2' t} + \kappa_2 e^{-\alpha_1' t} + (1-\kappa_1-\kappa_2)e^{-(\alpha_1+\alpha_2)t}, & \alpha_1+\alpha_2 \neq \alpha_1', \alpha_2' \\ \kappa_1 e^{-\alpha_2' t} + (1-\kappa_1+\alpha_2 t)e^{-(\alpha_1+\alpha_2)t}, & \alpha_1+\alpha_2 = \alpha_1' \neq \alpha_2' \\ \kappa_2 e^{-\alpha_1' t} + (1-\kappa_2+\alpha_1 t)e^{-(\alpha_1+\alpha_2)t}, & \alpha_1+\alpha_2 = \alpha_2' \neq \alpha_1' \\ [1+(\alpha_1+\alpha_2)t]\, e^{-(\alpha_1+\alpha_2)t}, & \alpha_1+\alpha_2 = \alpha_1' = \alpha_2', \end{cases}$$

where $\kappa_1 = \frac{\alpha_1}{\alpha_1+\alpha_2-\alpha_2'}$ and $\kappa_2 = \frac{\alpha_2}{\alpha_1+\alpha_2-\alpha_1'}$.

Example 2.6. When $\alpha_1 = \alpha_2 = \alpha$ and $\alpha_1' = \alpha_2' = 2\alpha$, the system reliability $R_{2:2}(t)$ is given by

$$R_{2:2}(t) = (1 + 2\alpha t)e^{-2\alpha t}.$$

From (2.4), the system's model failure rate function is

$$\lambda(t) = \frac{4\alpha^2 t}{1 + 2\alpha t}.$$

This system is related to *cascading failures* models, see Singpurwalla [100].

For a n-component system of this type, the joint density can be formulated in which the parameters of the surviving components change as components fail. Let the n components fail at times $T_{1:n} \le T_{2:n} \le \cdots \le T_{n:n}$. For $j = 1, \ldots, n$ and $k = 1, \ldots, j-1$, $T_{j:n}$ has failure rate $\alpha_j^{(k)}$ after k of the components have failed, i.e., for $j = 1, \ldots, n$,

$$T_{j:n} : \alpha_j \overset{1 \text{ failed}}{\longrightarrow} \alpha_j^{(1)} \overset{2 \text{ failed}}{\longrightarrow} \alpha_j^{(2)} \overset{3 \text{ failed}}{\longrightarrow} \cdots \overset{(j-1) \text{ failed}}{\longrightarrow} \alpha_j^{(j-1)}.$$

The joint density of $T_1, \ldots, T_n$ is, similar to (2.18),

$$f(t_1, \ldots, t_n) = \prod_{k=1}^{n} \left[\alpha_k^{(k-1)} \prod_{j=k}^{n} \exp\left(-\alpha_j^{(k-1)} (t_{k:n} - t_{k-1:n}) \right) \right],$$

where $0 = t_{0:n} \le t_{1:n} \le \cdots \le t_{n:n}$. We call this distribution the FMVE.

The formulae become more complicated as the number of components becomes larger than two. However, if we assume these different components are the same versions with equal parameters, then the reliability functions have simpler forms. If k components failed, the conditional distribution of the surviving components will have exponential distributions with parameters α_{k+1}, for $k = 0, 1, \ldots, n-1$. That is, for $j = 1, \ldots, n$,

$$T_{j:n} : \alpha_1 \overset{1 \text{ failed}}{\longrightarrow} \alpha_2 \overset{2 \text{ failed}}{\longrightarrow} \alpha_3 \overset{3 \text{ failed}}{\longrightarrow} \cdots \overset{(j-1) \text{ failed}}{\longrightarrow} \alpha_j.$$

The joint density of $T_1, \ldots, T_n$ becomes

$$f(t_1, \ldots, t_n) \prod_{k=1}^{n} \alpha_k \exp\left[-\alpha_k (n - k + 1)(t_{k:n} - t_{k-1:n}) \right],$$

where $0 = t_{0:n} \leq t_{1:n} \leq \cdots \leq t_{n:n}$. The spacings $D_k = T_{k:n} - T_{k-1:n}$ are independent exponential distributions with mean $1/\theta_k$, where

$$\theta_k = (n-k+1)\alpha_k, \quad k = 1, \ldots, n.$$

The kth order statistics $T_{k:n} = \sum_{i=1}^{k} D_i$ can be considered as the time of failure of a k-out-of-n:F system. Thus the reliability function for the 1-out-of-n:F system is $R_{1:n}(t) = P(T_{1:n} > t) = e^{-\theta_1 t} = e^{-n\alpha_1 t}$. In order to obtain the reliability function for the n-out-of-n:F system, we need to obtain the p.d.f. of $T_{k:n}$. The Laplace transform of $T_{k:n}$ is

$$\psi(s) = E(e^{-sT_{k:n}}) = \prod_{i=1}^{k} \frac{\theta_i}{\theta_i + s}.$$

The p.d.f. of $T_{k:n}$ is obtained by expressing the Laplace transform in partial fraction form as

$$\sum_{i=1}^{k} p_i \frac{\theta_i}{\theta_i + s},$$

where

$$p_i = \prod_{j=1, j\neq i}^{n} \frac{\theta_j}{\theta_j - \theta_i}.$$

Thus the p.d.f. of $T_{k:n}$ is

$$f_{T_{k:n}}(t) = \sum_{i=1}^{k} p_i \theta_i \exp[-\theta_i t].$$

It follows that the reliability function for the n-out-of-n:F system is

$$R_{n:n}(t) = P(T_{n:n} > t) = \int_t^{\infty} f_{T_{n:n}}(v)dv = \sum_{i=1}^{n} p_i \exp[-\theta_i t],$$

that is,

$$R_{n:n}(t) = \sum_{i=1}^{n} \left[\prod_{j=1, j\neq i}^{n} \frac{(n-j+1)\alpha_j}{(n-j+1)\alpha_j - (n-i+1)\alpha_i} \right] \exp[-(n-i+1)\alpha_i t], \tag{2.20}$$

provided that $(n-j+1)\alpha_j - (n-i+1)\alpha_i \neq 0$, for $j \neq i$. The reliability function $R_{n:n}(t)$ is a mixture of exponentials, since $\sum_{i=1}^{n} p_i = 1$.

Example 2.7. Suppose that a system consists of n identical components connected in parallel. Let us suppose that failures of certain components lead to increase in the functional load applied to the components still in operation. The failure rate of each component is independent of time but depends on the number of components that have not failed. At a given instant, $(n-k+1)$ components are in operation, then the failure rate of each of them is equal to α_k. The reliability function is given by (2.20). It can be shown that the mean lifetime of the system is [32]

$$\frac{1}{n\alpha_1} + \frac{1}{(n-1)\alpha_2} + \cdots + \frac{1}{\alpha_n}.$$

Load-sharing model is originated from the study of the failure properties of composite materials using the concept of fiber bundles. The model can be either static or time-dependent. Experiments usually favor the dynamic failure fiber bundle models [50]. The load-sharing rule is an important element of the load-sharing models. It governs how the loads on the working components change after some components in the system fail. There are equal, local, monotone, and non-monotone load-sharing rules, and the equal rule is the one most applied. FBVE is a simple load-sharing model and is applicable to almost all load-sharing rule.

Studies of k-out-of-n system related to FMVE are Scheuer [94], Lin et al. [53], and Amari and Misra [4]. Amari et al. [6] study tampered failure rate load-sharing k-out-of-n:G systems. Huang and Xu [41], Liu [62], and Lu [63] study reliability functions of load-sharing redundant systems with non-exponential or arbitrary failure distributions. Shao and Lamberson [97] models a load-sharing k-out-of-n:G system by a continuous-time Markov chain and solves its differential equations using inverse Laplace transforms. We return to this model in Sect. 2.5.2.

2.3 Markov Models

Markov models are commonly used to perform reliability analysis of engineering systems and fault-tolerant systems. They are also used to handle reliability and availability analysis of repairable systems. We first give notations and several properties of stochastic processes. Next, we explore Markov chains focusing on criteria of recurrent/transient state, and long-run probabilities. We then discuss basic properties of the homogeneous Poison process, which is one of the most important stochastic processes. The nonhomogeneous Poisson process (NHPP) is an time-dependent variation of the homogeneous Poisson process. NHPP models has been applied to various fields, such as repairable systems and software reliability modeling. The discussion is then going to the continuous-time Markov chain, including the birth, the death, and the birth-death processes. It is not an easy task to solve the state equations. A number of solution techniques exist, such as analytical solution [9], Laplace-Stieltjes transforms [92], numerical integration and computer-assisted evaluation [88].

2.3.1 Basic Properties of Stochastic Processes

A *stochastic process* is a mathematical model for the occurrence, at each moment after the initial time, of a random phenomenon. Mathematically, a stochastic process is a collection of random variables $\{X(t), t \in T\}$ defined on a common probability space indexed by a suitable index set T which describes the evolution of some systems. That is, for each $t \in T$, $X(t)$ is a random variable. Generally these random variables $X(t)$ are dependent. Stochastic processes are then characterized by different dependency relationships among $X(t)$. We often interpreted t as time and call $X(t)$ the *state* of the process at time t. If T is a countable set, we call $\{X(t) : t \in T\}$ a discrete-time stochastic process. If T is a continuum, we call $\{X(t) : t \in T\}$ a continuous-time stochastic process.

Often $T = [0, \infty)$ if the system evolves in continuous time. For example, $X(t)$ might be the number of people in a queue at time t, or the accumulated claims paid by an insurance company in $[0, t]$. Alternatively, we could have $T = \{0, 1, \dots\}$ if the system evolves in discrete time. Then $X(n)$ might represent the number of arrivals to a queue during the service interval of the nth customer.

Any *realization* of a stochastic process $\{X(t) : t \in T\}$ is called a *sample path*. The space of the possible values of each $X(t)$ is called the *state space* S. If S is a countable set, we call $\{X(t) : t \in T\}$ a discrete state stochastic process. If S is a continuum, we call $\{X(t) : t \in T\}$ a continuous state stochastic process. Several common properties of stochastic processes are given as follows.

Markov property: A Markov process is a stochastic process having the Markov property that, given the present state at time s and all past states, the future state at $t + s$ is independent of the past and depends only on the present state at s. Formally, if $\forall s, t \geq 0$ and $0 \leq u \leq s$,

$$P(X(t+s) = j | X(s) = i, X(u) = x(u), 0 \leq u < s) = P(X(t+s) = j | X(s) = i).$$

A Markov process having a countable state space is called a *Markov chain*. A discrete-time Markov process is simply called a Markov chain. Random walk is a Markov chain. The Poisson process and birth-death processes are *continuous-time Markov chains*. A Markov process for which all realization are continuous functions is called a *diffusion process*. Brownian motion is a diffusion process.

Homogeneous (or Stationary): A stochastic process is said to be homogeneous or stationary if the transition probability $p_{ij}(t) = P(X(t+s) = j | X(s) = i)$ is independent of s. Many continuous-time Markov chain we consider will have homogeneous property. Let τ_i be the amount of time that the process stays in state i. By the Markovian property, $P(\tau_i > s + t | \tau_i > s) = P(\tau_i > t)$ for all $s, t \geq 0$. Hence, the random variable τ_i is memoryless and must be exponentially distributed. That is, the amount of time a continuous-time Markov chain spends in each state, before jumping to the next state, is exponentially distributed.

Independent increments: A continuous-time stochastic process $\{X(t), t \in T\}$ is said to have *independent increments* if, for all $0 \leq t_0 < t_1 < t_2 < \cdots < t_n$, the

random variables $X(t_1) - X(t_0), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$ are jointly independent, where the difference $X(t_i) - X(t_{i-1})$ is called the *increment*. That is, for all n,

$$P(X(t_0, t_1) = k_1, \ldots, X(t_{n-1}, t_n) = k_n) = \prod_{i=1}^{n} P(X(t_{i-1}, t_i) = k_i),$$

where $X(t_2) - X(t_1) = X(t_1, t_2)$. If the index set contains a smallest index t_0, it is also assumed that $X(t_0), X(t_1) - X(t_0), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$ are jointly independent. *Brownian motions*, *Poisson processes* and *Lévy processes* have homogeneous independent increments.

Martingale processes: A stochastic process $\{X_t : t \in T\}$ is said to be a *martingale* process if, for $t_1 < \cdots < t_n < t_{n+1}$,

1. $E(|X_t|) < \infty, \quad \forall t$,
2. $E\left(X_{t_{n+1}} | X_{t_1}, X_{t_2}, \ldots, X_{t_n}\right) = X_{t_n}$,
3. For any $t \in T$ the random variable $X(t)$ is $\mathcal{F}_t$-measurable.

Taking expectation of (2) gives $E(X_{t_{n+1}}) = E(X_{t_n})$, and so $E(X_{t_n}) = E(X_{t_1})$ for all t's. For any $s \leq t, s, t \in T$, if $E(X(t)|\mathcal{F}_s) \geq X(s)$ P-a.s., then we have the definition of a submartingale; if $E(X(t)|\mathcal{F}_s) \leq X(s)$ P-a.s., then we have a supermartingale.

A martingale is a generalized version of a fair game. Martingale concepts and methodologies have provided a far-reaching apparatus vital to the analysis of all kinds of functionals of stochastic processes. In particular, martingale constructions serve decisively in the investigating of stochastic models of *diffusion type*.

2.3.2 Markov Chains

2.3.2.1 Transition Probabilities

A Markov chain with discrete-state space S is a collection of random variables X_n with the Markov property

$$\begin{aligned} P(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i) \\ = P(X_{n+1} = j | X_n = i) = p_{ij}, \end{aligned}$$

for all states $i_0, i_1, \ldots, i_{n-1}, i, j$ and all $n \geq 0$. p_{ij} is called a 1-step *transition probability*, representing the probability that a process in state i will next make a transition into state j. We write p_{ij} instead of $p_{ij}^{n,n+1}$ to represent the homogeneous transition probabilities.

We often display the 1-step transition probabilities $p_{ij} = P(X_{n+1} = j | X_n = i)$ in a matrix $\mathbf{P} = [p_{ij}]$, called the *transition probability matrix* (or simply transition matrix), such that

$$\mathbf{P} = [p_{ij}] = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ \vdots & \vdots & \vdots & \cdots \\ p_{i0} & p_{i1} & p_{i2} & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{bmatrix},$$

where $p_{ij} \geq 0 \ \forall i, j \geq 0$, and $\sum_{j=0}^{\infty} p_{ij} = 1, i = 0, 1, \ldots$.

A Markov chain $\{X_n, n = 0, 1, 2, \ldots\}$ is completely defined by its (1-step) transition probability matrix and its initial distribution (i.e., the specification of a probability distribution on the state of the process at time 0):

$$\begin{aligned} & P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) \\ & \quad = P(X_n = i_n | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}) \\ & \qquad \cdot P(X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}) \\ & \quad = p_{i_{n-1}, i_n} \cdot p_{i_{n-2}, i_{n-1}} \cdot \cdots \cdot p_{i_1, i_2} \cdot p_{i_0, i_1} \cdot P(X_0 = i_0), \end{aligned}$$

where $P(X_0 = i_0)$ is the *initial probability*.

The analysis of a Markov chain concerns mainly the calculation of the probability of the possible realizations of the process. Central in these calculations are the n-step transition probability matrices $\mathbf{P}^{(n)} = [p_{ij}^{(n)}]$ where $p_{ij}^{(n)}$ is the n-step transition probability and $p_{ij}^{(n)} = P(X_{n+m} = j | X_m = i), n \geq 0, i, j \geq 0$. The interest is to calculate $p_{ij}^{(n)}$ in terms of the transition probability p_{ij}.

2.3.2.2 Chapman-Kolmogorov Equations

The Chapman-Kolmogorov equations calculate the $(n + m)$-step transition probability $p_{ij}^{(n+m)}$ by summing over all the intermediate state k at time n and moving to state j from state k at the remaining time m: $\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \cdot \mathbf{P}^{(m)}$. A consequence of the Chapman-Kolmogorov equations is that $\mathbf{P}^{(n)} = \mathbf{P}^n$, i.e., the n-step transition probability matrix is equal to the nth power of the 1-step transition matrix.

The distribution of position at any time, $P(X_n = j)$, is given by the initial distribution and the n-step transition probabilities: $(n \geq 1)$

$$P(X_n = j) = \sum_{i=0}^{\infty} P(X_0 = i) P(X_n = j | X_0 = i) = \sum_{i=0}^{\infty} P(X_0 = i) p_{ij}^{(n)}.$$

Example 2.8. Consider a system with three state components situated in a shock environment that generates shocks according to a Poisson process with intensity λ. The three states can be described as 0 or good, 1 or degraded, and 2 or failed. When a component is hit by a shock it will (independently of the other components) pass from state 0 to state 1 with probability $p_1 = 1 - q_1$ and from state 1 to state 2 with probability $p_2 = 1 - q_2$ ($p_1 \neq p_2$). Let P_k be the probability that a component in state 0 survives k shocks. Then

$$P_k = q_1^k + \sum_{j=0}^{k-1} q_1^j \, p_1 q_2^{k-1-j} = \frac{p_1 q_2^k - p_2 q_1^k}{q_2 - q_1}.$$

2.3.2.3 Stationary Distributions

Let $w = [w_0, w_1, w_2, \dots]$ and $\sum_{i=0}^{\infty} w_i = 1$. Then w is said to be a *stationary distribution* (or equilibrium vector, steady-state vector) for the Markov chain if $w\,\mathbf{P} = w$. If w is a stationary distribution, then $\forall n \geq 1$,

$$w\mathbf{P}^n = (w\mathbf{P})\mathbf{P}^{n-1} = w\mathbf{P}^{n-1} = \cdots = w\mathbf{P} = w.$$

The chain is sometimes said to be in *equilibrium*.

Example 2.9. Consider a two-state Markov chain with the state space $S = \{0, 1\}$ and the transition probability matrix $P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$, $0 < \alpha, \beta < 1$. The stationary distribution of the Markov chain is $\left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$.

2.3.2.4 Classification of States

- *Accessible*: If $p_{ij}^{(n)} > 0$ for some $n \geq 0$. That is, the state j is accessible from the state i if there is positive probability that in a finite number of steps a Markov chain moves from i to j.
- *Communicate*: If $\exists\, m, n \geq 0$ such that $p_{ij}^{(m)} > 0$ and $p_{ji}^{(n)} > 0$. That is, two states communicate if each is accessible from the other.
- *Irreducible*: "Communicate with" is an equivalence relation, which breaks state space S up into disjoint equivalence classes. Two states that communicate are in the same class. If there is just one class, the chain is irreducible (all states communicate with each other), otherwise it is *Reducible*. A *closed* class is one we cannot leave. A closed class containing exactly one state is called *absorbing*.
- *Transient/recurrent*: A state is transient if it is not certain we shall return to it sometime, otherwise it is recurrent. Formally,

$$P(X_n = i \text{ for some } n \geq 1 | X_0 = i) \begin{cases} = 1, & \text{state } i \text{ is recurrent} \\ < 1, & \text{state } i \text{ is transient.} \end{cases}$$

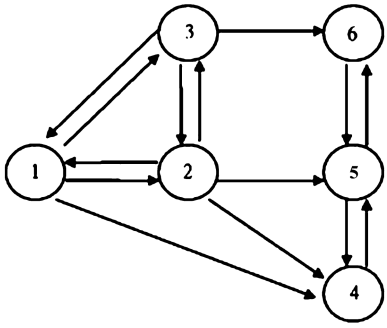


Fig. 2.8 A six-state system transitions

- *Periodicity*: Period $d(i)$ of a state i is the greatest common divisor of those time $n \geq 1$ taken to return to i, given we start there. That is, $d(i) = \gcd\{n : n \geq 1, p_{ii}^{(n)} > 0\}$. In particular, if $d(i) = 1$, then the state i is called *aperiodic*.

Example 2.10. If $S = \{1, 2, 3, 4, 5, 6\}$ and the transition probability matrix is given by

$$\mathbf{P} = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/6 & 1/6 & 0 \\ 1/3 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Then $\{1, 2, 3\}$: transient, aperiodic; $\{4, 5, 6\}$: irreducible closed set, recurrent, period 2 (Fig. 2.8).

2.3.2.5 First Hitting Probability

The states of a Markov chain are classified independently from two viewpoints: (i) recurrent and transient states, and (ii) periodic and aperiodic states. The vast majority of Markov chains we deal with are aperiodic, and criteria for a state to be recurrent or transient is closely related to the *first hitting probability*. The first hitting probability is defined as

$$f_{ij}^{(n)} = P(T_j = n | X_0 = i), \qquad n \geq 1,$$

representing the probability that, starting from state i, the first transition into state j occurs at time n. The random variable T_j represents the time (i.e., number of steps) for the first transition into state j, i.e., $T_j = \min\{n \geq 1 : X_n = j\}$. If $\forall\, n \geq 1$, $X_n \neq j$, then define $T_j = \infty$. Define $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$ to be the probability that the

chain ever visits j, starting from i. For $i \neq j$, $f_{ij} > 0$ if and only if j is accessible from i. Of course, $f_{ij} \leq 1$. Now we have a criterion for a state to be recurrent or transient.

Definition 2.1. State j is recurrent if and only if $f_{jj} = 1$, and transient if and only if $f_{jj} < 1$.

If state j is an absorbing state, then $P(T_j = 1|X_0 = j) = 1$. This implies that $f_{jj} = 1$. Thus an absorbing state is recurrent. If a Markov chain is in state j and j is recurrent, the probability that the process will return to that state is 1. Since the process is a Markov chain, this is equivalent to the process beginning once more from state j, and with probability 1, it will once again return to that state. Repeating this argument leads to the conclusion that state j will be entered *infinitely often*. Hence, a recurrent state has the property that the expected number of time periods that the process is in state j is infinite. If state j is transient, then the probability of reentering state j is f_{jj} (< 1) and the probability of not reentering state j is $1 - f_{jj}$. Then the expected number of time periods that the process is in state j is finite and given by $\frac{1}{1-f_{jj}}$.

2.3.2.6 A Criterion for Recurrent and Transient

Theorem 2.2. *An important formula that connect* $f_{ij}^{(n)}$ *and* $p_{ij}^{(n)}$ *is*

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}, \qquad \forall i, j \text{ and } 1 \leq n < \infty.$$

Theorem 2.3. *A criterion for identifying a state to be recurrent or transient in terms of the n-step transition probabilities* $p_{ij}^{(n)}$ *is: (1) State* j *is recurrent if and only if* $\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$, *(2) State* j *is transient if and only if* $\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$.

2.3.2.7 Examples of Recurrent/Transient

Example 2.11. Consider a two-state Markov chain with states $\{0, 1\}$ and the transition probability matrix $P = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$. It is obvious that $\sum_{n=1}^{\infty} p_{00}^{(n)} = 1 + 1 + \cdots = \infty$ and $\sum_{n=1}^{\infty} p_{11}^{(n)} = \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \cdots = 1 < \infty$, which implies that state 0 is recurrent and state 1 is transient.

Example 2.12. Consider a three-state Markov chain with states $\{1, 2, 3\}$ and the transition probability matrix

$$P = \begin{bmatrix} \alpha & 1-\alpha & 0 \\ 0 & 0 & 1 \\ 1-\beta & 0 & \beta \end{bmatrix}.$$

Since $f_{11} = \sum_{n=1}^{\infty} f_{11}^{(n)} = \alpha + \sum_{n=3}^{\infty}(1-\alpha)(1-\beta)\beta^{n-3} = \alpha + (1-\alpha)\frac{(1-\beta)}{(1-\beta)} = 1$, then $f_{11} = 1$, and consequently the state 1 is recurrent.

2.3.2.8 Ergodic

Ergodic states are important in the classification of chains, and in proving the existence of limiting probability distributions.

Definition 2.2. The *mean recurrence time* μ_{jj} denotes the expected time of transitions needed to return to state j, starting from state j:

$$\mu_{jj} = E(T_j | X_0 = j) = \begin{cases} \sum_{n=1}^{\infty} n f_{jj}^{(n)} & \text{if } j \text{ is recurrent} \\ \infty & \text{if } j \text{ is transient.} \end{cases}$$

μ_{jj} may be infinite even if j is recurrent. The recurrent state j is called *positive recurrent* if $\mu_{jj} < \infty$ and it is *null recurrent* if $\mu_{jj} = \infty$.

It can be shown that for a finite-state Markov chain, all recurrent states are positive recurrent. Note that $\mu_{ij} = E(T_{ij})$ is the *mean first passage time*, where $T_{ij} = \min\{n \geq 0 : X_n = j | X_0 = i\}$.

Definition 2.3. Positive recurrent states that are aperiodic are called ergodic states. A Markov chain is said to be ergodic if all its states are ergodic states.

Example 2.13. Consider a two-state Markov chain $S = \{1, 2\}$ with transition probability matrix $\mathbf{P} = \begin{bmatrix} 1/3 & 2/3 \\ 1/4 & 3/4 \end{bmatrix}$. Given $X_0 = 1$, the chain enters state 2 as soon as it leaves state 1. Hence

$$P(T_{12} = n) = \frac{2}{3}\left(\frac{1}{3}\right)^{n-1}, \qquad n \geq 1,$$

and

$$\mu_{12} = \sum_{n=1}^{\infty} n f_{12}^{(n)} = \sum_{n=1}^{\infty} n \frac{2}{3}\left(\frac{1}{3}\right)^{n-1} = \frac{3}{2}.$$

Likewise, first return to state 1 at the nth step occurs after $n-2$ consecutive visits to state 2, so

$$P(T_1 = n) = \begin{cases} \frac{1}{3}, & n = 1 \\ \frac{2}{3}\left(\frac{3}{4}\right)^{n-2}\frac{1}{4}, & n \geq 2. \end{cases}$$

Hence

$$\mu_{11} = \sum_{n=1}^{\infty} n f_{11}^{(n)} = \frac{1}{3} + \sum_{n=2}^{\infty} n \frac{2}{3} \left(\frac{3}{4}\right)^{n-2} \frac{1}{4} = \frac{11}{3}.$$

Example 2.14. If the transition probability matrix is given by

$$\mathbf{P} = \begin{bmatrix} \alpha & 1-\alpha & 0 \\ 0 & 0 & 1 \\ 1-\beta & 0 & \beta \end{bmatrix}.$$

The state 1 is recurrent and its mean recurrence time is given by

$$\mu_{11} = \sum_{n=1}^{\infty} n f_{11}^{(n)} = \alpha + (1-\alpha)(1-\beta) \sum_{n=3}^{\infty} n \beta^{n-3} = \frac{3 - 2\alpha - 2\beta + \alpha\beta}{1-\beta} < \infty.$$

Hence state 1 is positive recurrent. Because $p_{11}^{(n)} > 0$ for $n \geq 3$, state 1 is aperiodic. Hence state 1 is ergodic (and also states 2 and 3).

Theorem 2.4. *An irreducible chain has a stationary distribution π if and only if all the states are positive recurrent; in this case, π is the unique stationary distribution and is given by $\pi_j = 1/\mu_{jj}$ for each $j \in S$, where μ_{jj} is the mean recurrence time of j.*

Example 2.15. Consider the Markov chain

$$P = \begin{bmatrix} 0 & 1/3 & 2/3 \\ 2/3 & 0 & 1/3 \\ 1/3 & 2/3 & 0 \end{bmatrix}.$$

The unique stationary distribution is $\pi = (1/3, 1/3, 1/3)$. Thus, $\mu_{11} = \pi_1^{-1} = 3$, $\mu_{22} = \pi_2^{-1} = 3$, and $\mu_{33} = \pi_3^{-1} = 3$. The mean recurrence time for all states is 3.

Theorem 2.5. *If we have an irreducible aperiodic chain:*

(1) If it is transient, then $p_{ij}^{(n)} \longrightarrow 0 \ \forall i, j$;

(2) If it is positive recurrent, then $p_{ij}^{(n)} \longrightarrow \pi_j = 1/\mu_{jj}$;

(3) If it is null recurrent, then $p_{ij}^{(n)} \longrightarrow 0 \ \forall i, j$, and the mean time to return to any state is infinite.

2.3.2.9 Limiting Probabilities in Finite-State Markov Chains

For a finite-state Markov chain: (i) not all states are transient, and (ii) there are no null recurrent states. An irreducible finite chain must be positive recurrent. If the irreducible finite-state Markov chain is aperiodic, there exists a unique and positive stationary distribution which is the limiting distribution.

For a Markov chain, all states can be classified into several recurrent classes $C_1, \ldots, C_m$, and remaining transient states which form a transient set T. Then all sets $C_1, \ldots, C_m, T$ are disjoint. Rearranging (or relabeling) all the states so that the irreducible classes are together and transient states are last, the Markov chain can be rewritten in the form

$$P = \begin{array}{c} \\ C_1 \\ C_2 \\ \vdots \\ C_m \\ T \end{array} \begin{array}{c} \begin{array}{ccccc} C_1 & C_2 & \ldots & C_m & T \end{array} \\ \begin{pmatrix} P_1 & 0 & \ldots & 0 & 0 \\ 0 & P_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & P_m & 0 \\ R_1 & R_2 & \ldots & R_m & Q \end{pmatrix} \end{array},$$

where $P_1, \ldots, P_m$ are the transition matrices for recurrent classes $C_1, \ldots, C_m$, Q is a square matrix for describing the transitions among all transient states for T, and $R_1, \ldots, R_m$ are (not necessarily square) matrices describing the transitions from all transient states to the corresponding recurrent classes $C_1, \ldots, C_m$. Let us consider the limiting probabilities $\lim_{n\to\infty} P_{ij}^{(n)}$ for all states. Assume that all positive recurrent states are *aperiodic*, we have

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\mu_{jj}} \qquad (i, j \in C_k; k = 1, \ldots, m),$$

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0 \qquad (i \in C_k, j \in C_l; k \neq l),$$

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0 \qquad (i \in C_k, j \in T),$$

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{f_{ij}}{\mu_{jj}} \qquad (i \in T, j \in C_k),$$

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0 \qquad (i, j \in T).$$

Theorem 2.6. *For a finite-state Markov chain, the eventual transition probability f_{ij} from state $i \in T$ to state $j \in C_k$ $(k = 1, \ldots, m)$ satisfies*

$$f_{ij} = \sum_{l \in C_k} P_{il} + \sum_{l \in T} P_{il} f_{lj} \qquad (i \in T, j \in C_k),$$

or in a matrix form

$$[f_{ij}] = [I - Q]^{-1} R_k \cdot \mathbf{1},$$

where I is a identity matrix and $\mathbf{1}$ is a column vector of all the components' unity.

Example 2.16. Consider a Markov chain with the state space $S = \{1, 2, 3, 4\}$ and the transition probability

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.3 & 0.7 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0.2 & 0 & 0.1 & 0.7 \end{bmatrix}.$$

We know that $\{1\}$ is an absorbing state, $\{2, 3\}$ are positive recurrent states, and $\{4\}$ is a transient state. For the recurrent class $\{1\}$, we have $P_1 = [1]$ and the limiting distribution $\pi_1 = 1/\mu_{11} = 1$. For the recurrent class $\{2, 3\}$, we have the submatrix $P_2 = \begin{bmatrix} 0.3 & 0.7 \\ 0.5 & 0.5 \end{bmatrix}$ and the stationary distribution is $(\pi_2, \pi_3) = (1/\mu_{22}, 1/\mu_{33}) = (5/12, 7/12)$. For the transient set $\{4\}$,

$$\begin{aligned} [I - Q]^{-1} &= [10/3], \\ [I - Q]^{-1} R_1 \cdot \mathbf{1} &= [10/3][2/10][1] = 2/3, \\ [I - Q]^{-1} R_2 \cdot \mathbf{1} &= [10/3][0 \quad 1/10] \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1/3. \end{aligned}$$

That is, $f_{41} = 2/3$ and $f_{42} = f_{43} = 1/3$. Thus, $P_{41}^{(\infty)} = f_{41}/\mu_{11} = 2/3$. $P_{42}^{(\infty)} = f_{42}/\mu_{22} = 5/36$, and $P_{43}^{(\infty)} = f_{43}/\mu_{33} = 7/36$. The limiting probability matrix is

$$P^{\infty} = \lim_{n \to \infty} P^n = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 5/12 & 7/12 & 0 \\ 0 & 5/12 & 7/12 & 0 \\ 2/3 & 5/36 & 7/36 & 0 \end{bmatrix}.$$

2.3.3 Poisson Processes

The Poisson process is the simplest stochastic process that arises in many applications, especially as a model for the 'counting process' (or 'arrival processes') that counts the number of failures through time. A stochastic process $\{N(t), t \geq 0\}$ is said to be a counting process if $N(t)$ represents the total number of 'events' that have occurred up to time t.

2.3.3.1 Poisson Process

Definition 2.4. A counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process with *rate* $\lambda > 0$ if

(i) $N(0) = 0$,
(ii) The process has independent increments.

(iii) The probability that k events take place for any interval t is Poisson distributed with mean λt. That is, $\forall\ s,t \geq 0, P(N(t+s)-N(s)=k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$, $k = 0,1,2,\ldots$.

It follows from condition (iii) that a Poisson process has stationary increments and also that $E[N(t)] = \lambda t$, which explains why λ is called the 'rate' of the process. To determine if an arbitrary counting process is actually a Poisson process, we must show that conditions (i)–(iii) are satisfied. Conditions (i) and (ii) can usually be verified from our knowledge of the process. However, it is not at all clear how we would determine that condition (iii) is satisfied, and for this reason we need an alternative definition of a Poisson process.

Definition 2.5. A counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process with rate $\lambda > 0$ if the following conditions are satisfied:

(i) $N(0) = 0$,
(ii) $\{N(t), t \geq 0\}$ has stationary (or homogeneous) independent increments,
(iii) $P(N(h) = 1) = \lambda h + o(h)$,
(iv) $P(N(h) \geq 2) = o(h)$.

From this definition, we derive the probabilities $P_n(t) = P(N(t) = n)$, $n = 0,1,2,\ldots$, which denotes the probability that n events occur in $[0,t]$, and $\sum_{n=0}^{\infty} P_n(t) = 1$. To find the distribution of $N(t)$, we first derive a differential equation for $P_0(t)$.

$$\begin{aligned} P_0(t+h) &= P(N(t+h)=0) \\ &= P(N(t)=0)P(N(t+h)-N(t)=0|N(t)=0) \\ &= P_0(t)P_0(h) = P_0(t)[1-\lambda h + o(h)], \end{aligned}$$

where the final two equations follow from (ii), (iii), and (iv). Hence

$$\frac{P_0(t+h)-P_0(t)}{h} = -\lambda P_0(t) + \frac{o(h)}{h}.$$

Letting $h \to 0$ yields $P_0'(t) = -\lambda P_0(t)$. Next, for $n \geq 1$,

$$\begin{aligned} P_n(t+h) &= P(N(t+h)=n) \\ &= P(N(t)=n, N(t+h)-N(t)=0) \\ &\quad + P(N(t)=n-1, N(t+h)-N(t)=1) \\ &\quad + P(N(t+h)=n, N(t+h)-N(t)\geq 2). \end{aligned}$$

By (iv), the last term in the above is $o(h)$; hence, by (ii), we obtain

$$\begin{aligned} P_n(t+h) &= P_n(t)P_0(h) + P_{n-1}(t)P_1(h) + o(h) \\ &= (1-\lambda h)P_n(t) + \lambda h P_{n-1}(t) + o(h). \end{aligned}$$

Thus,

$$\frac{P_n(t+h)-P_n(t)}{h} = -\lambda P_n(t) + \lambda P_{n-1}(t) + \frac{o(h)}{h}.$$

Letting $h \to 0$ yields $P_n^{'}(t) = -\lambda P_n(t) + \lambda P_{n-1}(t)$. Thus, we need to solve a family of differential equations:

$$\begin{aligned} P_0^{'}(t) &= -\lambda P_0(t), \\ P_n^{'}(t) &= -\lambda P_n(t) + \lambda P_{n-1}(t), \quad n = 1, 2, \ldots. \end{aligned}$$

Subject to (i), $P_0(0) = 1$ and $P_n(0) = 0$ for all $n \geq 1$, solving recursively we obtain

$$P_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \ldots.$$

Thus, for a fixed t, $N(t) \sim \text{Poisson}(\lambda t)$.

From the above results, Definition 2.5 implies Definition 2.4. In fact, two definitions are equivalent. It is obvious that the process has stationary increment from (iii) of Definition 2.4. Moreover, from (iii) of Definition 2.4, we have

$$\begin{aligned} P\{N(h) = 1\} &= \lambda h e^{-\lambda h} = \lambda h + o(h), \\ P\{N(h) = k\} &= \frac{(\lambda h)^k}{k!}e^{-\lambda h} = o(h), \quad k = 2, 3, \ldots, \end{aligned}$$

where we use

$$e^{-\lambda h} = 1 - \lambda h + \frac{(\lambda h)^2}{2!} - \frac{(\lambda h)^3}{3!} + \cdots = 1 - \lambda h + o(h).$$

Thus Definition 2.4 implies Definition 2.5. A Poisson process implies all properties in Definitions 2.4 and 2.5.

2.3.3.2 Interarrival Time and Arrival Time Distributions

For a Poisson process with rate $\lambda > 0$, the interarrival times X_n $(n = 1, 2, \ldots)$ are i.i.d. exponential r.v.s with parameter λ:

$$\begin{aligned} P(X_1 > t) &= P(N(t) = 0) = e^{-\lambda t}, \\ P(X_2 > t | X_1 = s) &= P(0 \text{ event in } (s, s+t] | X_1 = s) \\ &= P(0 \text{ event in } (s, s+t]) = e^{-\lambda t}, \end{aligned}$$

i.e., the conditional distribution $P(X_2 > t | X_1 = s)$ is independent of X_1, and $X_2 \sim \text{Exp}(\lambda)$. Repeating the same argument yields the desired results.

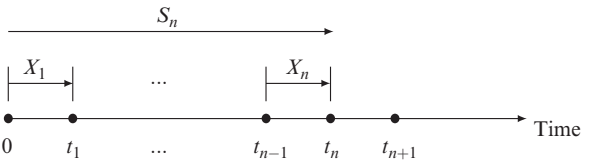


Fig. 2.9 Interarrival times and waiting times

The assumption of stationary and independent increments is equivalent to asserting that, at any point in time, the Poisson process probabilistically restarts itself. In other words, the process has no memory, and hence exponential interarrival times are to be expected. Thus, we have a third definition for the Poisson process.

Definition 2.6. A Poisson process with rate $\lambda > 0$ is a *renewal process* with the exponential interarrival distribution $F(t) = 1 - e^{-\lambda t}$.

The arrival time of the nth event, also called the waiting time until the nth event (from the origin), is $S_n = \sum_{i=1}^{n} X_i$, where $S_0 = 0$ (Fig. 2.9). It can be shown that S_n has a gamma distribution with shape parameter n and scale parameter λ. Use the independent increment assumption as follows:

$$\begin{aligned} P(t < S_n < t + dt) &= P(N(t) = n - 1) \cdot P(1 \text{ event in } (t, t + dt)) + o(dt) \\ &= \frac{e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}\lambda dt + o(dt). \end{aligned}$$

Dividing by dt and then letting it approach 0, we have $f_{S_n}(t) = \lambda e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}$.

2.3.3.3 Some Examples of the Poisson Process

Example 2.17 (Superposition). The pooled process $\{N_1(t) + N_2(t), t \geq 0\}$ of two independent Poisson processes $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ with respective rates λ and μ is again a Poisson process with rate $\lambda + \mu$:

$$\begin{aligned} P(N_1(t) + N_2(t) = n) &= \sum_{k=0}^{n} P(N_1(t) = k, N_2(t) = n - k) \\ &= e^{-(\lambda+\mu)t}\frac{[(\lambda + \mu)t]^n}{n!}. \end{aligned}$$

Example 2.18 (Decomposition). Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ and let the consecutive events follow the Bernoulli trials. For instance, each arrival of failures at an assembly line can be classified as a major and a minor with probability p and probability $1 - p$, respectively. The classified processes

$\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ of major failures and minor failures are independent Poisson processes with rates $p\lambda$ and $(1-p)\lambda$, respectively:

$$P(N_1(t) = k, N_2(t) = n-k) = P(N_1(t) = k, N_2(t) = n-k | N(t) = n) \cdot P(N(t) = n)$$
$$= e^{-p\lambda t} \frac{(p\lambda t)^k}{k!} \cdot e^{-(1-p)\lambda t} \frac{((1-p)\lambda t)^{n-k}}{(n-k)!}.$$

Example 2.19. Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ. For $s < t$ and $k < n$

$$P(N(s) = k | N(t) = n) = \frac{P(N(s) = k, N(t) - N(s) = n-k)}{P(N(t) = n)}$$
$$= \binom{n}{k} \frac{s^k (t-s)^{n-k}}{t^n} \sim \text{Binomial}\left(n, \frac{s}{t}\right).$$

Example 2.20. Given two independent Poisson process $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ with rates λ_1 and λ_2

$$P(N_1(t) = k | N_1(t) + N_2(t) = n) = \frac{P(N_1(t) = k, N_2(t) = n-k)}{P(N_1(t) + N_2(t) = n)}$$
$$= \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k} \sim \text{Binomial}\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right).$$

2.3.3.4 The Order Statistic Property

Suppose that exactly one event of a Poisson process has taken place by time t. The time at which the event occurred is a uniform distribution over $[0, t]$: For $s \leq t$,

$$P(X_1 \leq s | N(t) = 1) = \frac{P(N(s) = 1, N(t) - N(s) = 0)}{P(N(t) = 1)} = \frac{s}{t}.$$

This result is generalized as the following order statistic property.

Theorem 2.7. *Suppose $\{N(t), t \geq 0\}$ is a Poisson process, and let $S_k = \sum_{i=1}^{k} X_i$ be the kth arrival time. Given that $N(t) = n$, the n arrival times $S_1, S_2, \ldots, S_n$ have the same distribution as the order statistics corresponding to n independent random variables uniformly distributed on the interval $[0, t]$.*

Example 2.21 (A Counter Model). Suppose that a device is subject to shocks that occur in accordance with a Poisson process having rate λ. The ith shock gives rise to a damage D_i. The D_i, $i \geq 1$, are assumed to be i.i.d. and also to be independent of $\{N(t), t \geq 0\}$, where $N(t)$ denotes the number of shocks in $[0, t]$. The damage due to a shock is assumed to decrease exponentially in time. That is, if a shock

has an initial damage D, then a time t later its damage is $De^{-\alpha t}$. If we suppose that the damage are additive, then $D(t)$, the damage at t, can be expressed as $D(t) = \sum_{i=1}^{N(t)} D_i e^{-\alpha(t-S_i)}$, where S_i represents the arrival time of the ith shock. The question is to find $E[D(t)]$.

Conditioning on $N(t)$ yields

$$E[D(t)|N(t) = n] = E\left[\sum_{i=1}^{n} D_i e^{-\alpha(t-S_i)}|N(t) = n\right]$$
$$= E[D]\, e^{-\alpha t} E\left[\sum_{i=1}^{n} e^{\alpha S_i}|N(t) = n\right].$$

Letting $U_1, \ldots, U_n$ be i.i.d. $U(0, t)$ r.v.'s, then, by Theorem 2.7,

$$E\left[\sum_{i=1}^{n} e^{\alpha S_i}|N(t) = n\right] = E\left[\sum_{i=1}^{n} e^{\alpha U_{i:n}}\right] = E\left[\sum_{i=1}^{n} e^{\alpha U_i}\right] = \frac{n}{\alpha t}(e^{\alpha t} - 1).$$

Hence,

$$E[D(t)|N(t)] = \frac{N(t)}{\alpha t}(1 - e^{-\alpha t})E[D],$$

and, taking expectations,

$$E[D(t)] = \frac{\lambda E[D]}{\alpha}(1 - e^{-\alpha t}).$$

2.3.3.5 Nonhomogeneous Poisson Process

Definition 2.7. The counting process $\{N(t), t \geq 0\}$ is said to be a nonstationary or nonhomogeneous Poisson process (NHPP) with *intensity function* $\lambda(t)$, $t \geq 0$ if:

(i) $N(0) = 0$,
(ii) $\{N(t), t \geq 0\}$ has independent increments,
(iii) $P(N(t + h) - N(t) = 1) = \lambda(t)h + o(h)$,
(iv) $P(N(t + h) - N(t) \geq 2) = o(h)$.

The NHPP has independent increments but does not have stationary property. Thus, there is the possibility that events may be more likely to occur at certain times than at other times. It can be shown that $N(t + s) - N(t) \sim$ Poisson$(m(t + s) - m(t))$, where $m(t) = \int_0^t \lambda(u)du$. Thus, $N(t)$ is Poisson distributed with mean $m(t)$,

and for this reason $m(t)$ is called the *mean value function* of the process. We use Definition 2.7 to find the probability $P_n(s) = P(N(t+s) - N(t) = n)$. For fix t,

$$\begin{aligned} P_0(s+h) &= P(N(t+s+h) - N(t) = 0) \\ &= P(N(t+s) - N(t) = 0) \cdot P(N(t+s+h) - N(t+s) = 0) \\ &= P_0(s)[1 - \lambda(t+s)h + o(h)]. \end{aligned}$$

Hence

$$\frac{P_0(s+h) - P_0(s)}{h} = -\lambda(t+s)P_0(s) + \frac{o(h)}{h}.$$

Letting $h \to 0$ yields

$$P_0'(s) = -\lambda(t+s)P_0(s). \tag{2.21}$$

By integration and setting $P_0(0) = 1$, we have $P_0(s) = e^{-[m(t+s)-m(t)]}$. Similarly, by the same procedure used in the Poisson process, we have

$$P(N(t+s) - N(t) = n) = e^{-[m(t+s)-m(t)]} \frac{[m(t+s) - m(t)]^n}{n!}, \quad n = 0, 1, 2, \ldots.$$

In particular,

$$P(N(t) = n) = e^{-m(t)} \frac{[m(t)]^n}{n!}, \quad n = 0, 1, 2, \ldots.$$

The entire course of a NHPP is determined by the distribution of first failure time:

$$P(X_{n+1} > x | S_1 = s_1, \ldots, S_n = s_n) = P(X_1 > s_n + x | X_1 > s_n), \ n = 0, 1, 2, \ldots,$$

where X's are interarrival times and S's are arrival times. In addition, setting $t = 0$ in (2.21), the intensity function of a NHPP satisfies $\frac{d}{dt} EN(t) = \frac{d}{dt}[-\ln \overline{F}(t)]$, the failure rate of first failure time. For example, if the first failure time has a Weibull distribution $F(t) = 1 - \exp(-\alpha t^\beta)$ with a shape parameter β and a scale parameter α, then the mean value function of the corresponding NHPP is $m(t) = \alpha t^\beta$. Such a NHPP is called a *Weibull process*, because first failure time has the Weibull distribution.

A NHPP model also corresponds to what is called *minimal repairs*, meaning that the system after repair is only as good as it was immediately before the failure. This means that the intensity function $\lambda(t)$ of the failure process immediately after the failure is the same as it was immediately before the failure, and hence is exactly as it would be if no failure had ever occurred. Thus we must have $\lambda(t) = h(t)$, where $h(t)$ is the failure rate of the first failure time. That is, the intensity function $\lambda(t)$ of

the failure process is completely defined by the failure rate $h(t)$, and the mean value function of a NHPP is $m(t) = \int_0^t h(w)dw$. Formally,

$$P(X_{n+1} > x|S_n = t) = P(X_1 > t + x|X_1 > t) = \frac{\overline{F}(t+x)}{\overline{F}(t)} = e^{-[m(t+x)-m(t)]}. \tag{2.22}$$

Thus, the conditional distribution of time to failure, given $S_n = t$, is $F_i(x|S_n = t) = 1 - e^{-[m(t+x)-m(t)]}$. The conditional expectation is given by

$$\int_0^\infty [1 - F_i(x|S_n = t)]dx = \int_0^\infty \exp\left[-\int_t^{t+x} \lambda(w)dw\right] dx.$$

For example, if $\lambda(t) = t$ then the conditional expectation is

$$\int_0^\infty \exp\left[-\int_t^{t+x} wdw\right] dx = \int_0^\infty \exp\left[-\frac{1}{2}\left((t+x)^2 - t^2\right)\right]$$
$$dx = (2\pi)^{1/2} e^{t^2/2}[1 - \Phi(t)],$$

where Φ is the standard normal distribution function.

Parzen [81] gives the unconditional distributions of the nth failure time S_n:

$$P(S_n > t) = P(N(t) \leq n-1) = \sum_{j=0}^{n-1} e^{-m(t)} \frac{(m(t))^j}{j!} = \int_{m(t)}^\infty \frac{x^{n-1}}{\Gamma(n)} e^{-x} dx. \tag{2.23}$$

The p.d.f. of S_n is

$$f_{S_n}(t) = \lambda(t) e^{-m(t)} \frac{[m(t)]^{n-1}}{(n-1)!}. \tag{2.24}$$

From (2.22) and (2.24),

$$P(X_{n+1} > x) = \int_0^\infty \lambda(s) e^{-m(x+s)} \frac{[m(s)]^{n-1}}{(n-1)!} ds. \tag{2.25}$$

By comparing (2.25) with (2.22) we see explicitly that the interarrival times are not independent, as is stated in Cox and Lewis [20].

From (2.23), the expectation of S_n is

$$E S_n = \int_0^\infty P(S_n > t)dt = \int_0^\infty m^{-1}(x) \frac{x^{n-1}}{\Gamma(n)} e^{-x} dx,$$

where m^{-1} is the inverse function of m. For example, the Weibull process where $m(t) = \alpha t^{\beta}, \alpha > 0$, then $m^{-1}(x) = \left(\frac{x}{\alpha}\right)^{1/\beta}$ and

$$ES_n = \int_0^{\infty} \left(\frac{x}{\alpha}\right)^{1/\beta} \frac{x^{n-1}}{\Gamma(n)} e^{-x} dx = \frac{\Gamma(n+\frac{1}{\beta})}{\alpha^{1/\beta} \cdot \Gamma(n)}.$$

2.3.3.6 Some Properties of NHPP

1. Like homogeneous Poisson process, NHPPs are closed under superposition, that is, the sum of a number of NHPPs is also a NHPP. Generally, we may mix the failure time data from different failure processes assumed to be NHPP and obtain an overall NHPP with a mean value function which is the sum of the mean value functions of the underlying NHPP models.
2. If events occur at $0 < t_1 < t_2 < \cdots < t_n < t$, the likelihood function of a NHPP is

$$f(t_1, \ldots, t_n) = \left\{\prod_{i=1}^{n} \lambda(t_i)\right\} \exp\left\{-\int_0^t \lambda(u) du\right\}.$$

   Assume that at $[s_{i-1}, s_i)$, the actual observed cumulative number of detected faults is n_i. The likelihood function is

$$L(n_1, \ldots, n_k) = \prod_{i=1}^{n} e^{-[m(s_i)-m(s_{i-1})]} \frac{[m(s_i) - m(s_{i-1})]^{n_i}}{n_i!}.$$

3. Any NHPP can be transformed to a homogeneous Poisson process through an appropriate time-transformation, and vice versa. Specifically, let $m(t)$ be a continuous nondecreasing function of t. Then $S_1, S_2, \ldots$ are arrival times in a NHPP $\{N(t), t \geq 0\}$ with $EN(t) = m(t)$ if and only if $m(S_1), m(S_2), \ldots$ are the arrival times in a homogeneous Poisson process with rate one. Lindqvist et al. [55] study the *trend renewal process*, by allowing above rate one homogeneous Poisson process to be any renewal process.
4. In software reliability modeling, we are interested in the cumulative number of failures $X(t)$ experienced up to time $t \geq 0$. We assume that the number of initial faults in a software is a finite unknown constant N, and $\{X(t), t \geq 0\}$ is a pure linear birth process with birth rate ϕ. If we write $Y(t) = N - X(t)$, then $\{Y(t), t \geq 0\}$ is a linear death process representing the number of remaining faults or undiscovered faults up to time t. If N has a Poisson distribution with parameter θ, then $X(t)$ and $Y(t)$ are independent and the unconditional distribution of $X(t)$ is a NHPP with the mean value function $m(t) = \theta(1-e^{-\phi t})$.
5. (M/G/∞ queue) Suppose that customers arrive at a service station in accordance with a Poisson process with rate λ. Upon arrival the customer is immediately

served by one of an infinite number of possible servers, and the service times are assumed to be independent with a common distribution G. We want to find the distribution of the number of customers that have completed service by time t and the distribution of the number of customers that are being served at time t. Consider an entering customer a type-I customer if it completed its service by time t and a type-II customer if it does not complete its service by time t. If the customer enters at time s, $s \leq t$, then it will be a type-I customer if its service time is less than $t-s$, and since the service time distribution is G, the probability of this will be $G(t-s)$. Similarly, a customer entering at time s will be a type-II customer with probability $1-G(t-s)$. Let $N_i(t)$ represents the number of type-i events that occur by time t ($i = 1, 2$), then $N_1(t)$ and $N_2(t)$ are independent Poisson random variables having respective mean $E[N_1(t)] = \lambda \int_0^t G(y)dy$ and $E[N_2(t)] = \lambda \int_0^t [1-G(y)]dy$, respectively. Details of this example, and further examples of this type, we refer the readers to Ross [93].

2.3.4 Birth and Death Processes

2.3.4.1 Chapman-Kolmogorov Equations

Definition 2.8. A stochastic process $\{X(t), t \geq 0\}$ is a continuous-time Markov chain if it takes on values in the set S (state space) of nonnegative integers and satisfies the Markov property:

$$\begin{aligned} P(X(t_{n+1}) = x_{n+1} | X(t_1) = x_1, X(t_2) = x_2, \ldots, X(t_n) = x_n) \\ = P(X(t_{n+1}) = x_{n+1} | X(t_n) = x_n), \end{aligned}$$

for $0 \leq t_1 < t_2 < \cdots < t_n < t_{n+1}$ and $x_1, x_2, \ldots, x_n, x_{n+1} \in S$.

Definition 2.9. The continuous-time Markov chain is called time-homogeneous or stationary if the transition probability $p_{ij}(s,t) = P(X(t+s) = j | X(s) = i)$ is independent of s, for all i, j, s, t, and we write

$$p_{ij}(t) = P(X(t+s) = j | X(s) = i) = P(X(t) = j | X(0) = i),$$

where $p_{ij}(t)$ is the probability that a Markov chain, presently in state i, will be in state j after an additional time t.

By the *Markov property* and the *homogeneity*, the amount of time a continuous-time Markov chain spends in each state is exponentially distributed.

If we put $p_{ij}(t)$ into the ith row, the jth column of a matrix, then the resulting matrix $\mathbf{P}(t) = (p_{ij}(t))$ is called the transition matrix of $\{X(t), t \geq 0\}$. The family $\{\mathbf{P}(t), t \geq 0\}$ is a stochastic semigroup; that is, it satisfies:

(i) $\mathbf{P}(t)$ is stochastic ($0 \leq p_{ij}(t) \leq 1$ and row sums equal to 1),

(ii) $\mathbf{P}(0) = \mathbf{I}$ (the identity matrix), (2.26)

(iii) The Chapman-Kolmogorov equation: $\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s)$.

2.3.4.2 Infinitesimal Generator

A matrix $\mathbf{Q}$, called the *infinitesimal generator* (or generator, Q-matrix, transition rate matrix), plays a role similar to that of a 1-step transition matrix in discrete-time Markov chains. Later in (2.30), we will see that there is a one-to-one correspondence between the infinitesimal generator $\mathbf{Q}$ and the transition matrix $\mathbf{P}(t)$.

Assume that the probability of 2 or more transitions in the interval $(t, t+h)$ is $o(h)$. We are interested in the behavior of $p_{ij}(h)$ for small h. Following (2.26), it turns out that $p_{ij}(h)$ is approximately linear in h when h is small. That is, there exist constants $\{q_{ij}; i, j \in S\}$ such that

$$p_{ij}(h) \approx \begin{cases} q_{ij}h & \text{if } i \neq j \\ 1 + q_{ii}h & \text{if } i = j. \end{cases}$$

It can be shown that each transition probability $p_{ij}(t)$ is uniformly continuous in $t \geq 0$, and $p_{ij}(t)$ is differentiable with respect to $t \geq 0$ [45]. Let us define

$$\mathbf{Q} \equiv \mathbf{P}'(0+) \qquad \text{or} \qquad \mathbf{Q} = \lim_{h \to 0+} \frac{\mathbf{P}(h) - \mathbf{I}}{h}.$$

Since $\mathbf{P}(0) = \mathbf{I}$, we have

$$q_{ij} = \begin{cases} \lim_{h \to 0+} \frac{p_{ij}(h)}{h} = p'_{ij}(0) & i \neq j \\ \lim_{h \to 0+} \frac{p_{ii}(h)-1}{h} = -\sum_{j \neq i} p'_{ij}(0) & i = j, \end{cases}$$

with $q_{ij} \geq 0$ for $i \neq j$ and $q_{ii} = -\sum_{j \neq i} q_{ij} \leq 0$. We call q_{ij} the *transition rate* from state i to state j. State i is an *absorbing* state if $q_{ii} = 0$.

$(-q_{ii})$ is the expected number of times that the process leaves state i per unit of time spent in state i. It is also called the *mean sojourn rate* of state i. Thus $(-q_{ii})$ is the reciprocal of the expected *holding time* that the process spends in state i per visit to state i. Similarly, q_{ij} is the transition rate from state i to state j in the sense that q_{ij} is the expected number of times that the process leaves from state i to state j per unit of time spent in state i. Thus $(-q_{ii}) = \sum_{j \neq i} q_{ij}$.

The information of the generator is a combination of the *embedded Markov chain* and the mean sojourn rates. Not only can the generator be obtained from the embedded Markov chain and mean sojourn rates, but the reverse is also true. For the generator matrix, the absolute value of the diagonal elements gives the mean sojourn rates. The transition matrix for the embedded Markov chain is then obtained by dividing the off-diagonal element by the absolute value of that row's diagonal element. The diagonal elements of the transition matrix for the embedded Markov

chain are zero (except an absorbing state). $q_{ij}/(-q_{ii})$, for example, is the conditional probability of a transition from state i to state j given that a transition from state i has taken place. If the holding time in each state follows an arbitrary distribution other than exponential, the involved process is a *semi-Markov process*.

Example 2.22. Suppose a Markov process with state space $\{a, b, c\}$ has a generator

$$\mathbf{Q} = \begin{bmatrix} -2 & 2 & 0 \\ 2 & -4 & 2 \\ 1 & 4 & -5 \end{bmatrix}.$$

The mean sojourn rates are $\lambda_a = 2$, $\lambda_b = 4$, and $\lambda_c = 5$, and the transition matrix for the embedded Markov chain is

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0.2 & 0.8 & 0 \end{bmatrix}.$$

2.3.4.3 Steady-State Probabilities

The steady-state probabilities are obtained directly from the generator $\mathbf{Q}$. Let $\boldsymbol{\pi}$ be a vector of steady-state probability matrix, then $\boldsymbol{\pi}$ is the solution to $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$, with row sum of $\boldsymbol{\pi}$ is 1. This equation is similar to $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$. Since Markov matrices have row sum of 1, the right-hand side of $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ is *one times* $\boldsymbol{\pi}$. Since the generator has row sums of 0, the right-hand side of $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$ is *zero times* $\boldsymbol{\pi}$.

Example 2.23. A 3-state Markov process $S = \{0, 1, 2\}$ with generator $\mathbf{Q} = \begin{bmatrix} -2 & 2 & 0 \\ 2 & -3 & 1 \\ 0 & 2 & -2 \end{bmatrix}$ has the steady-state distribution $(\pi_0, \pi_1, \pi_2) = (2/5, 2/5, 1/5)$.

2.3.4.4 Eigenvalue Approach

The eigenvalue approach is useful for finding the transition probabilities of a finite states Markov process. In principle, if we can calculate all the eigenvalues and their associated column eigenvectors for the generator $\mathbf{Q}$, then we can calculate $\mathbf{P}(t)$ analytically. Assume that $\mathbf{w}_1, \ldots, \mathbf{w}_n$ are eigenvalues of $\mathbf{Q}$ and $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are their associated right eigenvectors. We also assume that all the eigenvalues are different. We can write $\mathbf{Q} = \mathbf{U}\mathbf{W}\mathbf{U}^{-1}$, where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$, $\mathbf{U}^{-1}$ is the inverse matrix of $\mathbf{U}$, and $\mathbf{W}$ is the orthogonal matrix with eigenvalues on the main diagonal. Then $\mathbf{Q}^n = \mathbf{U}\mathbf{W}^n\mathbf{U}^{-1}$. It follows thats

$$\mathbf{P}(t) = \mathbf{I} + \sum_{i=1}^{\infty} \frac{(\mathbf{Q}t)^i}{i!} = \mathbf{I} + \sum_{i=1}^{\infty} \frac{\mathbf{U}(\mathbf{W}t)^i\mathbf{U}^{-1}}{i!} = \mathbf{U}e^{\mathbf{W}t}\mathbf{U}^{-1}. \tag{2.27}$$

Example 2.24. Consider an electronic device that is either "on standby" or "in use." The on-standby period is exponentially distributed with a mean of 20 s, and the in-use period is exponentially distributed with a mean of 12 s. Because the exponential assumption is satisfied, a Markov process $\{Y(t); t \geq 0\}$ with state space $E = \{0, 1\}$ can be used to model the electronic device as it alternates between being on standby and in use. State 0 denotes on standby, and state 1 denotes in use. The generator is given by $\mathbf{Q} = \begin{bmatrix} -3 & 3 \\ 5 & -5 \end{bmatrix}$, where the time unit is minute. The long-run probability of being in use is $\frac{3}{8}$. Now, we are interested in the time-dependent probabilities. Two eigenvalues are $\mathbf{w}_1 = 0$ and $\mathbf{w}_2 = -8$, and the corresponding eigenvectors are $(1 \;\; 1)'$ and $(3 \;\; -5)'$, respectively. Therefore, from (2.27), we have

$$\mathbf{P}(t) = \mathbf{U}e^{\mathbf{W}t}\mathbf{U}^{-1} = \begin{bmatrix} 1 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & e^{-8t} \end{bmatrix} \begin{bmatrix} 0.625 & 0.375 \\ 0.125 & -0.125 \end{bmatrix}$$

$$= \begin{bmatrix} 0.625 + 0.375e^{-8t} & 0.375 - 0.375e^{-8t} \\ 0.625 - 0.625e^{-8t} & 0.375 + 0.625e^{-8t} \end{bmatrix}, \quad t \geq 0.$$

The probability that the electronic device is in use at time t given that it started in use at time 0 is

$$P(Y(t) = 1 | Y(0) = 1) = 0.375 + 0.625e^{-8t}.$$

Example 2.25. Consider an *alternating renewal process* whose infinitesimal generator is

$$\mathbf{Q} = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}.$$

Two eigenvalues are $\mathbf{w}_1 = 0$ and $\mathbf{w}_2 = -(\lambda + \mu)$, and the corresponding eigenvectors are $(1 \;\; 1)'$ and $(\lambda \;\; -\mu)'$, respectively. Thus, from (2.27), we have

$$\mathbf{P}(t) = \begin{bmatrix} \dfrac{\mu}{\lambda+\mu} + \dfrac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} & \dfrac{\lambda}{\lambda+\mu} - \dfrac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} \\ \dfrac{\mu}{\lambda+\mu} - \dfrac{\mu}{\lambda+\mu}e^{-(\lambda+\mu)t} & \dfrac{\lambda}{\lambda+\mu} + \dfrac{\mu}{\lambda+\mu}e^{-(\lambda+\mu)t} \end{bmatrix}, \quad t \geq 0.$$

2.3.4.5 Kolmogorov Backward/Forward Equations

When the eigenvalue approach is not applicable, we need an alternative method to find transition probabilities $p_{ij}(t)$. Applying the Chapman-Kolmogorov equations $\mathbf{P}(t + s) = \mathbf{P}(t)\mathbf{P}(s)$, then

$$\frac{\mathbf{P}(t+h) - \mathbf{P}(t)}{h} = \frac{\mathbf{P}(t)\mathbf{P}(h) - \mathbf{P}(t)}{h} = \mathbf{P}(t)\frac{\mathbf{P}(h) - \mathbf{I}}{h}, \quad h > 0.$$

This implies

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q} \qquad \text{or} \qquad p'_{ij}(t) = \sum_k p_{ik}(t) q_{kj}. \tag{2.28}$$

This is Kolmogorov *forward* differential equation. Similarly, using $\mathbf{P}(t+s) = \mathbf{P}(s)\mathbf{P}(t)$, we have

$$\frac{\mathbf{P}(t+h) - \mathbf{P}(t)}{h} = \frac{\mathbf{P}(h)\mathbf{P}(t) - \mathbf{P}(t)}{h} = \frac{\mathbf{P}(h) - \mathbf{I}}{h}\mathbf{P}(t), \quad h > 0.$$

This implies

$$\mathbf{P}'(t) = \mathbf{Q}\mathbf{P}(t) \qquad \text{or} \qquad p'_{ij}(t) = \sum_k q_{ik} p_{kj}(t). \tag{2.29}$$

This is Kolmogorov *backward* differential equation. Subject to the boundary condition $\mathbf{P}(0) = \mathbf{I}$, the unique solution for both forward and backward equations is given by

$$\mathbf{P}(t) = e^{\mathbf{Q}t} = \sum_{i=0}^{\infty} \frac{(\mathbf{Q}t)^i}{i!}, \quad t \geq 0. \tag{2.30}$$

2.3.4.6 Birth Process

Poisson Process: A Poisson process discussed in Sect. 2.3.3.1 is a pure birth process. The generator of a Poisson process with state $S = \{0, 1, 2, \dots\}$ is

$$\mathbf{Q} = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & \dots \\ \vdots & & & \ddots & \end{bmatrix}.$$

From the Kolmogorov forward equations (2.28), we have

$$p'_{ij}(t) = p_{ij}(t) q_{jj} + p_{i,j-1}(t) q_{j-1,j} = -\lambda p_{ij}(t) + \lambda p_{i,j-1}(t).$$

Pure Birth Process: The pure (or simple) birth process is a natural and the simplest generalization of the Poisson process. It permits the chance of an event occurring at a given instant of time to depend upon the number of events which have already occurred.

Definition 2.10. A pure birth process $\{X(t), t \geq 0\}$ is a continuous-time Markov chain with positive parameters $\{\lambda_k, k = 0, 1, 2, \dots\}$ called *birth rates* satisfying the following statements.

1. $P(X(t+h) - X(t) = 1 | X(t) = k) = \lambda_k h + o_{1,k}(h)$,
2. $P(X(t+h) - X(t) = 0 | X(t) = k) = 1 - \lambda_k h + o_{2,k}(h)$,
3. $P(X(t+h) - X(t) \geq 2 | X(t) = k) = o_{3,k}(h)$,
4. $X(0) = 0$.

The infinitesimal generator of a pure birth process $\{X(t), t \geq 0\}$ is

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 \dots \\ 0 & -\lambda_1 & \lambda_1 & 0 & 0 \dots \\ 0 & 0 & -\lambda_2 & \lambda_2 & 0 \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

To find the distribution of $X(t)$, we let $P_n(t) = P(X(t) = n)$. The probability $P_n(t)$ is the transition probability $P(X(t) = n | X(0) = 0)$. Using the Kolmogorov forward equations (2.28), we have

$$\begin{aligned} P_0'(t) &= -\lambda_0 P_0(t), \\ P_n'(t) &= -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t), \quad n = 1, 2, \dots, \end{aligned}$$

with the initial conditions $P_0(0) = 1$ and $P_n(0) = 0$, $n = 1, 2, \dots$. Solving recursively, we have

$$\begin{aligned} P_0(t) &= e^{-\lambda_0 t}, \\ P_1(t) &= \lambda_0 \left[\frac{e^{-\lambda_0 t}}{\lambda_1 - \lambda_0} + \frac{e^{-\lambda_1 t}}{\lambda_0 - \lambda_1} \right], \quad (\lambda_0 \neq \lambda_1), \\ P_2(t) &= \lambda_0 \lambda_1 \left[\frac{e^{-\lambda_0 t}}{(\lambda_1 - \lambda_0)(\lambda_2 - \lambda_0)} + \frac{e^{-\lambda_1 t}}{(\lambda_0 - \lambda_1)(\lambda_2 - \lambda_1)} + \frac{e^{-\lambda_2 t}}{(\lambda_0 - \lambda_2)(\lambda_1 - \lambda_2)} \right], \\ &\vdots \\ P_n(t) &= \lambda_0 \lambda_1 \dots \lambda_{n-1} \left[\psi_{0,n} e^{-\lambda_0 t} + \psi_{1,n} e^{-\lambda_1 t} + \dots + \psi_{n,n} e^{-\lambda_n t} \right], \end{aligned}$$

where

$$\psi_{k,n} = [(\lambda_0 - \lambda_k) \dots (\lambda_{k-1} - \lambda_k)(\lambda_{k+1} - \lambda_k) \dots (\lambda_n - \lambda_k)]^{-1}$$

($\lambda_j \neq \lambda_k$ unless $j = k$).

To find a general transition probability $P_{ij}(t)$, we again use the Kolmogorov forward equations (2.28) and yield the differential equations

$$\begin{aligned} P_{ii}^{'}(t) &= -\lambda_i P_{ii}(t), \\ P_{ij}^{'}(t) &= P_{i,j-1}(t)q_{j-1,j} + P_{i,j}(t)q_{jj} \\ &= \lambda_{j-1}P_{i,j-1}(t) - \lambda_j P_{ij}(t), \quad t \geq 0 \quad (j > i;\ \ P_{ij}(t) = 0 \text{ if } j < i), \end{aligned}$$

with $P_{ii}(0) = 1$ and $P_{ij}(0) = 0$, $j \neq i$. Solving recursively we get

$$\begin{aligned} P_{ii}(t) &= e^{-\lambda_i t}, \\ P_{ij}(t) &= \lambda_{j-1}\int_0^t e^{-\lambda_j(t-s)}P_{i,j-1}(s)ds, \quad j > i. \end{aligned} \tag{2.31}$$

Yule Process: The Yule process is a well-known example of a pure birth process in biology. It characterizes the reproduction of living organisms, in which under certain conditions (sufficient food, no mortality, no migration, etc.) the probability of a birth at a given instant is proportional to the population size at that time. That is, the Yule process is a linear birth process with birth rate $\lambda_k = k\lambda$ $(k = 0, 1, 2, \dots)$.

We first start with a single individual at time 0, i.e., $X(0) = 1$. The infinitesimal generator with $S = \{0, 1, 2, \dots\}$ is

$$\mathbf{Q} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \dots \\ 0 & -\lambda & \lambda & 0 & 0 \dots \\ 0 & 0 & -2\lambda & 2\lambda & 0 \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

Let $P_n(t) = P(X(t) = n)$, given $X(0) = 1$. The Kolmogorov forward equations (2.28) is

$$\begin{aligned} P_0^{'}(t) &= 0, \\ P_n^{'}(t) &= -n\lambda P_n(t) + (n-1)\lambda P_{n-1}(t), \quad n = 1, 2, \dots, \end{aligned}$$

with the initial conditions $P_1(0) = 1$ and $P_n(0) = 0$ $(n = 2, 3, \dots)$. Solving recursively we get $P_0(t) = 0$ and

$$P_n(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}, \quad n = 1, 2, \dots. \tag{2.32}$$

Thus, starting with a single individual at time 0, $X(t)$ is geometric distributed with mean $e^{\lambda t}$ and variance $(1 - e^{-\lambda t})e^{2\lambda t}$.

Example 2.26. Applying the binomial theorem into (2.32), we write

$$P_n(t) = e^{-\lambda t}(1-e^{-\lambda t})^{n-1} = e^{-\lambda t} \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k e^{-\lambda k t}.$$

Let $\tilde{P}_n(s) = \int_0^\infty P_n(t) e^{-st} dt$ be the Laplace transform of $P_n(t)$ for the linear birth process with $X(0) = 1$. Then

$$\tilde{P}_n(s) = \int_0^\infty e^{-\lambda t} \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k e^{-\lambda k t} e^{-st} dt = \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} \frac{1}{s+(k+1)\lambda}.$$

Next, if the population starts with γ individuals, its size at time t will be the sum of γ i.i.d. geometric r.v.'s, and thus have a negative binomial distribution. Let $P_n^\gamma(t) = P(X(t) = n | X(0) = \gamma)$, then

$$P_n^\gamma(t) = \binom{n-1}{\gamma-1} \left(e^{-\lambda t}\right)^\gamma \left(1-e^{-\lambda t}\right)^{n-\gamma}, \quad n = \gamma, \gamma+1, \ldots.$$

To obtain a general transition probability $P_{ij}(t)$, we substitute $\lambda_j = j\lambda$ into (2.31) and solve them recursively to get the desired results.

The following is an example of the Yule process as a epidemic model, which may be used to model faults propagation in network or software reliability.

Example 2.27. Consider a population of m individuals that at time 0 consists of one "infected" and $m-1$ "susceptibles". Once infected an individual remains in that state forever and suppose that in any time interval h any given infected person will cause, with probability $\alpha h + o(h)$, any given susceptible to become infected. Let $X(t)$= # of infected individuals at time t, the $\{X(t), t \geq 0\}$ is a pure birth process with

$$\lambda_n = \begin{cases} (m-n)n\alpha, & n = 1, \ldots, m-1 \\ 0, & \text{otherwise.} \end{cases}$$

Let T be the time until the total population is infected, then $T = \sum_{i=1}^{m-1} T_i$, where T_i is the time from i infectives to $i+1$ infectives ($i = 1, \ldots, m-1$). Because $T_i \sim \text{Exp}((m-i)i\alpha)$, the mean and variance of T are, respectively,

$$E[T] = \sum_{i=1}^{m-1} \frac{1}{(m-i)i\alpha} = \frac{1}{m\alpha} \sum_{i=1}^{m-1} \left(\frac{1}{m-i} + \frac{1}{i}\right) \approx \frac{2\log(m-1)}{m\alpha},$$

$$\text{Var}(T) = \frac{1}{\alpha^2} \sum_{i=1}^{m-1} \left(\frac{1}{(m-i)i}\right)^2.$$

2.3.4.7 Death Process

Pure Death Process:

Definition 2.11. A pure (or simple) death process $\{X(t), t \geq 0\}$ is a continuous-time Markov chain with parameters $\{\mu_k, k = 1, 2, \ldots, n\}$ called *death rates* satisfying the following properties:

(i) $P(X(t+h) - X(t) = -1 | X(t) = k) = \mu_k h + o_{1,k}(h)$,
(ii) $P(X(t+h) - X(t) = 0 | X(t) = k) = 1 - \mu_k h + o_{2,k}(h)$,
(iii) $P(X(t+h) - X(t) \leq -2 | X(t) = k) = o_{3,k}(h), \ k \geq 0$,
(iv) $X(0) = n$.

It is natural that in the pure birth process, the events take place infinitely often for an infinite interval of time. However, in the pure death process, at most n events take place for any interval of time, since no event take place once the process reaches state 0, an absorbing state.

Let $P_k(t) = P(X(t) = k | X(0) = n)$, $k = 0, 1, \ldots, n$, be the transition probability with the initial condition $X(0) = n$. The Kolmogorov forward equations (2.28) for the pure death process are

$$\begin{aligned}
P_n'(t) &= -\mu_n P_n(t), \\
P_k'(t) &= -\mu_k P_k(t) + \mu_{k+1} P_{k+1}(t), \quad k = 1, 2, \ldots, n-1, \\
P_0'(t) &= \mu_1 P_1(t),
\end{aligned}$$

subject to the initial conditions $P_n(0) = 1$ and $P_k(0) = 0$ $(k = 1, 2, \ldots, n-1)$. Solving recursively we obtain

$$\begin{aligned}
P_n(t) &= e^{-\mu_n t}, \\
P_{n-1}(t) &= \mu_n \left[\frac{e^{-\mu_{n-1} t}}{\mu_n - \mu_{n-1}} + \frac{e^{-\mu_n t}}{\mu_{n-1} - \mu_n} \right], \\
P_{n-2}(t) &= \mu_n \mu_{n-1} \left[\frac{e^{-\mu_n t}}{(\mu_{n-1} - \mu_n)(\mu_{n-2} - \mu_n)} + \frac{e^{-\mu_{n-1} t}}{(\mu_n - \mu_{n-1})(\mu_{n-2} - \mu_{n-1})} \right. \\
&\qquad \left. + \frac{e^{-\mu_{n-2} t}}{(\mu_n - \mu_{n-2})(\mu_{n-1} - \mu_{n-2})} \right], \\
&\vdots \\
P_0(t) &= \mu_n \mu_{n-1} \ldots \mu_1 \left[\psi_{n,0} e^{-\mu_n t} + \psi_{n-1,0} e^{-\mu_{n-1} t} + \cdots + \psi_{0,0} e^{-\mu_0 t} \right],
\end{aligned}$$

where

$$\psi_{k,0} = [(\mu_n - \mu_k) \ldots (\mu_{k+1} - \mu_k)(\mu_{k-1} - \mu_k) \ldots (\mu_0 - \mu_k)]^{-1}$$

$(\mu_j \neq \mu_k$ unless $j = k)$.

Linear Death Process: If $\mu_k = k\mu$ $(k = 1, 2, \ldots, n)$, i.e., the death rate is proportional to the number of members alive in a population. With the initial conditions $P_n(0) = 1$ and $P_k(0) = 0$ $(k = 1, 2, \ldots, n-1)$, we have

$$P_k(t) = \binom{n}{k}(e^{-\mu t})^k(1 - e^{-\mu t})^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

That is, $X(t) \sim \text{Binomial}(n, e^{-\mu t})$. The mean and variance are, respectively,

$$E[X(t)] = ne^{-\mu t} \quad \text{and} \quad \text{Var}[X(t)] = ne^{-\mu t}(1 - e^{-\mu t}).$$

2.3.4.8 Birth and Death Processes

Birth and Death Process

Definition 2.12. A birth and death process $\{X(t), t \geq 0\}$ is a continuous-time Markov chain on the states $\{0, 1, 2, \ldots\}$ with stationary transition probability $P_{ij}(t) = P(X(t+s) = j | X(s) = i)$ satisfying

1. $P_{i,i+1}(h) = \lambda_i h + o_{1,i}(h)$,
2. $P_{i,i-1}(h) = \mu_i h + o_{2,i}(h)$,
3. $P_{ii}(h) = 1 - (\lambda_i + \mu_i)h + o_{3,i}(h)$,
4. $P_{ij}(0) = \sigma_{ij}$, (i.e., $P_{ii}(0) = 1$, $P_{ij}(0) = 0$ if $j \neq i$),
5. $\mu_0 = 0, \lambda_0 > 0, \lambda_i, \mu_i > 0, i = 1, 2, \ldots,$

where λ_i are called the birth rates and μ_i are called the death rates.

The infinitesimal generator of a birth and death process $\{X(t), t \geq 0\}$ is

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 \ldots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 \ldots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 \ldots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

The amount of time the process in state k before making a transition into a different state is $\text{Exp}(\lambda_k + \mu_k)$. From the Kolmogorov forward equations (2.28), we have

$$\begin{aligned} P'_{ij}(t) &= P_{i,j-1}(t)q_{j-1,j} + P_{ij}(t)q_{jj} + P_{i,j+1}(t)q_{j+1,j} \\ &= \lambda_{j-1}P_{i,j-1}(t) - (\lambda_j + \mu_j)P_{ij}(t) + \mu_{j+1}P_{i,j+1}(t), \quad j \geq 1, \\ P'_{i0}(t) &= -\lambda_0 P_{i0}(t) + \mu_1 P_{i1}(t), \quad (\mu_0 = 0, \lambda_{-1} = 0). \end{aligned}$$

From the Kolmogorov backward equations: $P_{ij}^{'}(t) = \sum_{k=0}^{\infty} q_{ik} P_{kj}(t)$, we have

$$\begin{aligned} P_{ij}^{'}(t) &= q_{i,i-1} P_{i-1,j}(t) + q_{ii} P_{ij}(t) + q_{i,i+1} P_{i+1,j}(t) \\ &= \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t) + \lambda_i P_{i+1,j}(t), \quad i \geq 1, \\ P_{0j}^{'}(t) &= -\lambda_0 P_{0j}(t) + \lambda_0 P_{1j}(t). \end{aligned}$$

For both cases, the initial conditions are $P_{ij}(0) = \delta_{ij}$. The transition probabilities $P_{ij}(t)$ may in principle be calculated from a knowledge of the birth and death rates, although in practice these functions rarely have nice forms. It is an easier matter to determine the asymptotic behavior of the process as $t \to \infty$. There is a theoretical interest involving the transition probabilities $P_{ij}(t)$ called Karlin-McGregor representation, see Doom [109].

A Linear Growth Model with Immigration: A birth-death process is called a linear growth process with immigration if

$$\begin{aligned} \mu_n &= n\mu, \quad n \geq 1, \\ \lambda_n &= n\lambda + \theta, \quad n \geq 0. \end{aligned}$$

Such processes occur naturally in the study of biological reproduction and population growth. Each individual is assumed to give birth at an exponential rate λ; in addition, there is an exponential rate of increase θ of the population due to an external source such as immigration. Hence, $\lambda_n = n\lambda + \theta$. Deaths are assumed to occur at an exponential rate μ for each member of the population, and hence $\mu_n = n\mu$. Substitute $\lambda_n = n\lambda + \theta$ and $\mu_n = n\mu$ into the Kolmogorov forward equations (2.28):

$$\begin{aligned} P_{i0}^{'}(t) &= -\theta P_{i0}(t) + \mu P_{i1}(t), \quad (\lambda_0 = 0), \\ P_{ij}^{'}(t) &= (\lambda(j-1) + \theta) P_{i,j-1}(t) - ((\lambda + \mu) j + \theta) P_{ij}(t) \\ &\quad + \mu(j+1) P_{i,j+1}(t), \quad j \geq 1. \end{aligned}$$

If $X(0) = i$ and we multiply the jth equation by j and sum, it can be shown that the expected value $M(t) = E[X(t)] = \sum_{j=1}^{\infty} j P_{ij}(t)$ satisfies the differential equation

$$M^{'}(t) = \theta + (\lambda - \mu) M(t),$$

with the initial condition $M(0) = i$.

2.4 Software Reliability Models

2.4.1 Introduction

Musa et al. [76] pointed out that three of the most significant software product characteristics are *quality*, *cost*, and *time of delivery*. It seems unlikely that anyone can create a software product simultaneously satisfying high quality, rapid delivery, and low cost, so trade-offs are required among the characteristics. Quantitative measures exist for cost and schedule, but the quantification of quality is more difficult. The quality of a software product can be represented by *software quality attributes* which are multidimensional properties and even hierarchical structures. Before a software product is released for service, it is developed through software development life cycle (SDLC): requirements, design, implementation, testing, and validation. Determination and measurement of user desirable software quality attributes is a subject of *software engineering* which concerns the study of the SDLC in various aspects.

Software reliability is defined as the probability of failure free software operation for a specified period of time in a specified environment [76]. Software reliability is an attribute of software quality and it has proved to be the most readily quantifiable of the attributes of software quality. Many software reliability models are applied to the testing or debugging phase of software development. There is an implicit assumption in most works on software reliability that the failure data obtained during the test and development phase represents its behaviour in the use environment [56].

Early software reliability models were deterministic models which are also called *software metrics* or *complexity metrics*. Software metrics are defined by analyzing the structure of the program or the flowgraph of the program. These models usually empirically measure the software quality attributes. They are used in the early phases of the SDLC to predict the number of faults, or are used in the maintenance phase for assessing and controlling the quality of a software product [37, 46, 68, 91]. Nowadays software reliability models refer to probabilistic ones and they have become the mainstream of software reliability study. The subject of software reliability has rapidly developed and spawned some original probability models of its own [22].

A *fault* is defined as a defect in a system that may cause an *error* during its operation. If an error affects the service to be provided by a system, a *failure* occurs. Because of the characteristics of software products, the operational environments of software, and the variety of development processes, the definition of software failure differs from application to application and should be defined clearly in the specifications. For example, a response time of 30 s could be a serious failure for an air traffic control system, but acceptable for an airline reservation system.

There is no physical process that determines when a program should fail, rather there occur in the program certain faults of logic or syntax that might cause the program to crash or produce results contrary to the specifications. Upon execution

of a program, an input state is translated into an output state. Hence, a program can be regarded as a function mapping the input space (the set of all input states) to the output space (the set of all output states). Different users have different operational profiles which will determine the probabilities of selection of different inputs during execution. Software failures may not be 'generated' stochastically but may be 'detected' in a random way [42]. It justifies the use of stochastic models.

A large number of statistical software reliability models have been proposed over the past decade for evaluating the reliability of computer software. For reviews of the existing models, see [2, 33, 52, 64, 76, 83, 102, 104]. One of the main approaches of those models is to obtain results from assumptions about the stochastic behavior of how failures occur during software's running time. In this approach, the methods deal with the failure process by characterizing the time pattern of failure occurrences.

It is of great interest to classify or unify the large amount of software reliability models. Model classification and unification are useful in research in a sense that similar models are related to each other, and new and useful models can be built based on this. Section 2.4.2 reviews classification and unification of software reliability models. The most cited software reliability models are the Jelinski-Moranda model [42], henceforth the JM model, and the Goel-Okumoto model [34], henceforth the GO model. Essentially, the JM model corresponds to a (bounded) pure birth process, while the GO model corresponds to an NHPP process. In Sect. 2.4.3, we review the assumptions of the JM model. Sections 2.4.4–2.4.6 introduce several types of SRGMs by relations between the transition rate functions and the conditional failure rates, including i.i.d. order statistic models, time-between-failure models, and NHPP models. In Sect. 2.4.7, a more general setting of self-exciting processes is discussed.

2.4.2 Taxonomy of Software Reliability Models

Software reliability models can be broadly divided into the so-called white-box and black-box models. The white-box (or architecture-based) models incorporates information on the structure of the software in the models, while the black-box models disregard the internal structure of the system and represent solely its external failure behavior. A good review of architecture-based models is given by Goševa-Popstojanova and Trivedi [36]. The black-box approach is the most prevalent approach, and they can be further classified as *static* and *dynamic*. Static models include input-domain related models, complexity models (deterministic), and error seeding models (probabilistic). Error seeding models are also called capture-recapture models, and hypergeometric distribution is the basic tool [14]. The dynamic models consider the failure process as a stochastic process, and the reliability of the software changes at different stages of testing and debugging. Classifications of black-box software reliability models have been discussed by Goel [33], Gokhale et al. [35], Musa et al. [76], Pham [83], and Singpurwalla

and Wilson [102]. Recently, Sharma et al. [98] have provide a figure which thoroughly classifies software reliability model according to SDLC. Models are classified into six categories: early prediction models, software reliability growth models (SRGMS), input domain based models, architecture based models, hybrid black box models, and hybrid white box models.

By unification we mean a mechanism which can generate new models and accommodate some existing models. Several such unifications had been proposed. For example, Langberg and Singpurwalla [51] unified some earlier models by adopting Bayesian method, Miller [70] discussed general order statistic models, Chen and Singpurwalla [18] unified software reliability models by self-exciting processes, Pham [86] generalized NHPP by extending the differential equations of the GO model, Wang [113] generalized existing software reliability models by considering both properties of mixture and self-exciting, and Huang et al. [40] presented how several existing software reliability growth models based on NHPPs can be derived according to the concept of weighted arithmetic, weighted geometric, or weighted harmonic means.

For many SRGMs, it is usually assumed that a fault is removed immediately on detection and correction, and no new faults are introduced. The so-called perfect debugging. In addition, the removal time is always assumed to be negligible. However, it is more practical to consider effect of debugging and timing of debugging as the imperfect removals happen in practice.

Effect of debugging:

- Perfect removal: detected faults are removed successfully,
- Imperfect removal: detected faults are not removed successfully and no new fault is introduced,
- Fault introduction: new faults may be introduced by incorrect debugging.

Timing of debugging:

- Immediate: all detected faults instantaneously repaired at time of first manifestation,
- Time-delayed: a fault may manifest itself repeatedly before being debugged.

The term *imperfect debugging* is used for both imperfect removal and fault introduction. In terms of the number of remaining faults, under certain assumptions (see Sect. 2.4.3), the imperfect removal process can be described by a simple death process [13, 33], and the imperfect debugging process can be modeled by a birth-death process [49].

Intuitively, we may assume a probability p for a fault to be removed successfully, and consider that there are two types of failures. The first type of failures are caused by inherent faults and the second type of failures are caused by introduced faults. A simple way for describing introduced faults is to assume that new faults are introduced by a homogeneous Poisson process with some parameter θ. For examples, Tokuno and Yamada [107] assume that failures caused by inherent faults follow Moranda's geometric model [72], while Yamada [118] assumes that inherent

Table 2.1 Assumptions from the JM model

	Initial fault content:
(A1)	The number of initial faults is a finite unknown constant, say N
	Independent of faults:
(A2)	Faults are assumed to be statistically independent
	Constant failure rates:
(A3)	Each fault has the same constant detection rate, say ϕ
	Fault removal (debugging) process:
(A4)	A fault is removed immediately on detection and correction, and no new faults are introduced. The removal time is negligible

faults follows NHPP models. Zeephongsekul et al. [124], based on NHPPs, consider that primary failures generate secondary-faults under imperfect debugging. Pham et al. [86] propose generalized NHPP models for modeling imperfect debugging.

An intrinsic assumption in most models is that a fault is removed immediately on detection. Consequently, the failure detected process and the fault corrected process are regarded as the same. An alternative approach, using queueing theorem, studies the detection process and the correction process separately, see Yang [122], Dohi et al. [23], and Huang and Huang [39].

We will not pursue further about imperfect debugging and time-delay correction models. However, in Sect. 2.5.3, a bivariate counting process is applied to a 2-version programming system to illustrating the effect of imperfect debugging.

2.4.3 The JM Model

The JM model [42] is the first software reliability model to be widely known and used. Assumptions of the JM model are important but seems not realistic. Many models are proposed hereafter to challenge the JM model. However, most of them are propagated from the JM model by extending one or more assumptions. Assumptions from the JM model are listed in Table 2.1. Assumption (A4) is known as *perfect debugging*. A program will exhibit the phenomenon of reliability growth after fault detection and correction, and thus perfect debugging models are usually called software reliability growth models (SRGMs). Assumption (A3) is a basic property of the exponential distribution. Hence the time to failure for each fault has the exponential distribution with parameter ϕ. From assumption (A2), given N and ϕ, the T_i are independent exponential distributions with parameters $(N - i + 1)\phi$. The p.d.f. of T_i is given by $f_i(\tau|\phi, N) = (N - i + 1)\phi \cdot e^{-(N-i+1)\phi\tau}$, $\tau > 0$. Let $h_i(\tau)$ be the conditional failure rate of T_i, then $h_i(\tau) = (N - i + 1)\phi$. Thus the failure rates of the program at any time are proportional to the number of remaining faults, and the sequence of failure rates shows a decreasing trend. The MTTF of T_i is $\{(N - i + 1)\phi\}^{-1}$, an increasing function of i.

Table 2.2 Five types of software reliability models

	$S_{i-1} \le t < S_i$	
(a)	Specify $r_i(t) = (N - i + 1)\psi(t)$	Then $h_i(\tau) = (N - i + 1)\psi(\tau + S_{i-1})$
(b)	Specify $r_i(t) = (N - i + 1)\psi_i(t)$	Then $h_i(\tau) = (N - i + 1)\psi_i(\tau + S_{i-1})$
(c)	Specify $r_i(t) = \lambda(t)$	Then $h_i(\tau) = \lambda(\tau + S_{i-1})$
	$0 \le \tau < S_i - S_{i-1}$	
(d)	Specify $h_i(\tau) = (N - i + 1)\psi(\tau)$	Then $r_i(t) = (N - i + 1)\psi(t - S_{i-1})$
(e)	Specify $h_i(\tau) = \psi_i(\tau)$	Then $r_i(t) = \psi_i(t - S_{i-1})$

Let $\{X(t), t \ge 0\}$ be the software failure process where $X(t)$ is the cumulative number of failures experienced up to time $t \ge 0$. For the JM model, the failure process $\{X(t), t \ge 0\}$ is a linear birth continuous-time Markov chain with failure state space $\{0, 1, \ldots, N\}$. This failure process satisfies two properties: *homogeneity* and *conditional orderliness*. The property of conditional orderliness prevents the process having multiple simultaneous failures.

For a conditional orderly failure process $\{X(t), t \ge 0\}$ with state space $\{0, 1, \ldots\}$, the *transition rate* $r_i(t)$ is the probability of encountering a new fault in the next instant of time dt, given that there are $(i - 1)$ failures in $[0, t)$, i.e., for $i = 1, \ldots,$

$$r_i(t) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} P(X(t + \Delta t) - X(t) = 1 | X(t) = i - 1), \quad S_{i-1} \le t < S_i. \tag{2.33}$$

Recall that the conditional *failure rate* of T_i is defined by

$$h_i(\tau) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} P(\tau < T_i \le \tau + \Delta t | T_i > \tau), \quad \tau \ge 0. \tag{2.34}$$

$h_i(\tau)$ is assumed to exist for all the T_i. Relations between $r_i(t)$ and $h_i(\tau)$ can be established in the following theorem.

Theorem 2.8. *Let* $0 \le S_1 \le S_2 \le \ldots$ *be the indexed failure times and let* $T_i = S_i - S_{i-1}, i = 1, \ldots$. *The relations between the transition rate* $r_i(t)$ *given in Eq. (2.33) and the conditional failure rate* $h_i(\tau)$ *given in Eq. (2.34) are, for* $i = 1, \ldots,$

$$r_i(t) = h_i(t - S_{i-1}), \quad S_{i-1} \le t < S_i, \tag{2.35}$$

$$h_i(\tau) = r_i(\tau + S_{i-1}), \quad 0 \le \tau < S_i - S_{i-1}. \tag{2.36}$$

Thus, if we specify $r_i(t)$ then we have $h_i(\tau)$, and vice versa. Five types of software reliability models are given in Table 2.2. The JM model is a special example of all cases with $r_i(t) = h_i(\tau) = (N - i + 1)\phi$.

Type (a) is the class of *i.i.d. order statistic models*, type (c) is the class of NHPP models, types (d) and (e) are the class of *time-between-failure models*

(or *concatenated failure rates models*), and type (b) seems not be considered by the existing models. In type (b) we retain the explicit factor $(N - i + 1)$ rather than subsuming it into ψ_i, because it is more intelligible to do so. Note that types (a), (b) and (d) have finite fault content while types (c) and (e) may have infinite fault content.

Theorem 2.9. *Let* $0 \leq S_1 \leq S_2 \leq \ldots$ *be the indexed failure times and let* $T_i = S_i - S_{i-1}, i = 1, \ldots$. *The joint density function of* $T_1, \ldots, T_k$ *is given by*

$$f(t_1, \ldots, t_k) = \prod_{j=1}^{k} \left\{ h_j(t_j) \exp\left[- \int_0^{t_j} h_j(\tau) d\tau \right] \right\},$$

and the joint density function of $S_1, \ldots, S_k$ *is*

$$f(s_1, \ldots, s_k) = \prod_{j=1}^{k} \left\{ h_j(s_j - s_{j-1}) \exp\left[- \int_0^{s_j - s_{j-1}} h_j(\tau) d\tau \right] \right\},$$

where the conditional failure rates $h_i(\tau)$ *are defined by Eq. (2.34). When the* $h_j(\tau)$ *are given as types (a), (b) and (d) of Table 2.2,* k *must be less than or equal to the initial number of faults.*

Theorem 2.9 is useful to obtain the likelihood functions of parameters. We can then find estimates of parameters by applying the usual estimation techniques, such as the maximum likelihood approach or the least squares approach.

Example 2.28. The JM model [42]: $h_j(\tau) = (N - j + 1)\phi$

$$f(t_1, \ldots, t_k) = \frac{N!}{(N-k)!} \phi^k \exp\left[-\phi \sum_{j=1}^{k} (N - j + 1) t_j \right], \quad k \leq N,$$

or

$$f(s_1, \ldots, s_k) = \frac{N!}{(N-k)!} \phi^k \exp\left[-\phi \left(\sum_{j=1}^{k} s_j + (N-k) s_k \right) \right], \quad k \leq N.$$

Example 2.29. The Littlewood model [57]: $h_j(\tau) = (N - j + 1)\alpha / (\beta + \tau + S_{j-1})$

For $k \leq N$,

$$f(t_1, \ldots, t_k) = \frac{N!}{(N-k)!} \alpha^k \beta^{N\alpha} \left[\prod_{j=1}^{k-1} \left(\frac{1}{\beta + t_1 + \cdots + t_j} \right)^{\alpha+1} \right] \times \left(\frac{1}{\beta + t_1 + \cdots + t_k} \right)^{(N-k+1)\alpha+1},$$

or

$$f(s_1,\ldots,s_k)=\frac{N!}{(N-k)!}\alpha^k\beta^{N\alpha}\left[\prod_{j=1}^{k-1}\left(\frac{1}{\beta+s_j}\right)^{\alpha+1}\right]\left(\frac{1}{\beta+s_k}\right)^{(N-k+1)\alpha+1}.$$

Example 2.30. The Schick-Wolverton model [95]: $h_j(\tau)=(N-j+1)\phi\tau$

For $k\le N$,

$$f(t_1,\ldots,t_k)=\frac{N!}{(N-k)!}\phi^k\left[\prod_{j=1}^{k}t_j\right]\exp\left[-\frac{\phi}{2}\sum_{j=1}^{k}(N-j+1)t_j^2\right],$$

or

$$f(t_1,\ldots,t_k)=\frac{N!}{(N-k)!}\phi^k\left[\prod_{j=1}^{k}(s_j-s_{j-1})\right]\exp\left[-\frac{\phi}{2}\sum_{j=1}^{k}(N-j+1)(s_j-s_{j-1})^2\right].$$

Example 2.31. The Littlewood-Verrall model [60]: $h_j(\tau)=\alpha/(\tau+\psi_j)$

$$f(t_1,\ldots,t_k)=\prod_{j=1}^{k}\left[\frac{\alpha}{\psi_j}\left(\frac{\psi_j}{t_j+\psi_j}\right)^{\alpha+1}\right],$$

or

$$f(t_1,\ldots,t_k)=\prod_{j=1}^{k}\left[\frac{\alpha}{\psi_j}\left(\frac{\psi_j}{s_j-s_{j-1}+\psi_j}\right)^{\alpha+1}\right].$$

2.4.4 I.I.D. Order Statistic Models

Let assumptions (A1), (A2) and (A4) in Table 2.1 hold. Assumption (A3) states that every fault has the same deterministic effect on the overall failure rate. However, faults may have different failure rates. To generalize this, we consider two approaches. First, assumption (A3) can be interpreted by a Poisson shock model. A fault causes a failure to occur according to a Poisson process with rate ϕ. We may assume that shocks follow a Poisson process with rate $\psi(t)$ representing that faults have different failure rates. Second, recall that failure rates of T_i, time between failures, in the JM model are $h_i(\tau)=(N-i+1)\phi$. Each failure rate of T_i is a summation of failure rates of remaining faults that are equally ϕ. It has been argued that the detection rates $\{h_i\}$ depend on the size of faults subsets which may not be of the same size [56]. Faults with larger size tend to be removed first. Thus we may assume that the sequence $\{h_i\}$ is a stochastic process.

Table 2.3 Type (a):$r_i(t) = (N - i + 1)\psi(t)$; i.i.d. order statistics models

(M1)	Exponential	$\psi(t) = \phi$	JM model [42]
		$F(t) = 1 - \exp(-\phi t)$	$(\phi > 0)$
(M2)	Pareto	$\psi(t) = \alpha/(\beta + t)$	Littlewood [57]
		$F(t) = 1 - [\beta/(\beta + t)]^{\alpha}$	$(\alpha, \beta > 0)$
(M3)	Gamma(2,ϕ)	$\psi(t) = \phi^2 t/(1 + \phi t)$	
		$F(t) = 1 - (1 + \phi t)\exp(-\phi t)$	$(\phi > 0)$
(M4)	Logistic	$\psi(t) = \phi/[1 + c\ \exp(-\phi t)]$	
		$F(t) = [1 - \exp(-\phi t)]/[1 + c\ \exp(-\phi t)]$	$(\phi, c > 0)$
(M5)	Weibull	$\psi(t) = \alpha\beta t^{\beta-1}$	Wagoner [112]
		$F(t) = 1 - \exp(-\alpha t^{\beta})$	$(\alpha > 0, \beta > 0)$
(M6)	Rayleigh	$\psi(t) = \phi t$	
		$F(t) = 1 - \exp(-\phi t^2/2)$	$(\phi > 0)$
(M7)	Parabola	$\psi(t) = \phi(-at^2 + bt + c)$	
		$F(t) = 1 - \exp[\phi(-at^3/3 + bt^2/2 + ct)]$	$(a, b, c > 0)$

In the first approach, we assume that shocks follow a Poisson process with rate $\psi(t)$. This in turn generalizes the exponential distribution by a distribution F such that

$$F(t) = 1 - \exp\left\{-\int_0^t \psi(s)ds\right\}. \tag{2.37}$$

One way to obtain a reasonable common distribution F is using the doubly stochastic process. Assume that time to failures caused by faults are independent exponential distributions with rates $\phi_1, \phi_2, \ldots, \phi_N$. If the rates ϕ have a mixing distribution G, then we have i.i.d. mixture distributions given by

$$F(t) = \int_0^{\infty} (1 - e^{-\phi t})dG(\phi) = 1 - \int_0^{\infty} e^{-\phi t}dG(\phi). \tag{2.38}$$

So that the unconditional failure rates are

$$\psi(t) = \frac{f(t)}{\overline{F}(t)} = \frac{\int_0^{\infty} \phi e^{-\phi t}dG(\phi)}{\int_0^{\infty} e^{-\phi t}dG(\phi)}.$$

Thus the models become the *i.i.d. order statistic models* [70]. From Eqs. (2.37) and (2.38), the Laplace transform of the mixing distribution G is given by

$$G^*(t) = \int_0^{\infty} e^{-\phi t}dG(\phi) = \exp\left\{-\int_0^t \psi(s)ds\right\}.$$

Let $X(t)$ be the cumulative number of failures up to time t, then $X(t)$ has a binomial distribution similar to that in the JM model. The mean value function of $X(t)$ is $m(t) = NF(t)$. Models in this class, listed in Table 2.3, correspond to type (a) of Table 2.2.

Table 2.4 Type (e): $h_i(\tau) = \psi_i(\tau)$; time-between-failure-models

(M8)	Pareto	$h_i(\tau) = \alpha/(\tau + \psi_i)$	Littlewood-Verrall [60]
		$F_i(\tau) = 1 - [\psi_i/(\tau + \psi_i)]^\alpha$	$(\alpha > 0, \psi \uparrow i)$
(M9)	Geometric	$h_i(\tau) = Dk^{i-1}$	Moranda [72]
		$F_i(\tau) = 1 - \exp[-Dk^{i-1}\tau]$	$(D > 0, 0 < k < 1)$
(M10)	Power	$h_i(\tau) = \phi(N - i + 1)^\gamma$	Xie [116]
		$F_i(\tau) = 1 - \exp[-\phi(N - i + 1)^\gamma \tau]$	$(\phi > 0, \gamma > 1)$

Model (M2) is proposed by Littlewood [57]. He assumed the mixing distribution G to be a gamma distribution with the scale parameter β and the shape parameter α, namely, $dG(\phi) = \frac{\beta^\alpha}{\Gamma(\alpha)}\phi^{\alpha-1}e^{-\beta\phi}$, $\alpha, \beta > 0$. Models (M3) and (M4) can be viewed as variations of Yamada et al. [120, 121] in Table 2.6. $\psi(t)$ in model (M5) is decreasing for $0 < \beta < 1$, constant for $\beta = 1$, and increasing for $\beta > 1$. In model (M6), the distribution function F is known as the Rayleigh distribution, a Weibull distribution with shape parameter 2. In model (M7), $\psi(t)$ is a parabola which increases and reaches a maximum, then declines. Both models (M6) and (M7) are variations of Schick and Wolverton [95, 96] in Table 2.5.

2.4.5 Time Between Failures Models

In the second approach of extension from assumption (A3), partially due to the difficulty in defining what a program fault is, the failure rates of the T_i, times between successive failures, are given by $h_i(\tau) = \Lambda_i$, $i = 1, 2, \ldots$, which form a stochastic process. To describe reliability growth, we may choose $\{\Lambda_i\}$ such that the Λ_i are stochastically decreasing with i, i.e., $P(\Lambda_{i+1} < \lambda) \geq P(\Lambda_i < \lambda)$, $\forall\lambda \geq 0$, or equivalently such that the T_i are stochastically increasing with i, i.e., $P(T_{i+1} > t) \geq P(T_i > t)$, $\forall t \geq 0$. If the Λ_i have a mixing distribution G_i, then the mixture distribution is

$$F_i(\tau) = 1 - \int_0^\infty e^{-\lambda_i \tau} dG_i(\lambda_i).$$

The posterior failure rate is

$$h_i(\tau) := \psi_i(\tau) = \frac{\int_0^\infty \lambda_i e^{\lambda_i \tau} dG_i(\lambda_i)}{\int_0^\infty e^{-\lambda_i \tau} dG_i(\lambda_i)}.$$

In some cases, one can simply specify a particular function $\psi_i(\tau)$. This class of models, listed in Table 2.4, corresponds to type (e) of Table 2.2.

The above approach, however, does not depend on the number of remaining faults $(N - i + 1)$. An alternative way is to assume that $h_i(\tau) = (N - i + 1)\psi(\tau)$. This can

Table 2.5 Type (d): $h_i(\tau) = (N - i + 1)\psi(\tau)$; time-between-failure-models

(M11)	Rayleigh	$h_i(\tau) = (N - i + 1)\phi\tau$	Schick-Wolverton [95]
		$F_i(\tau) = 1 - \exp\left(-\int_0^\tau h_i(s)ds\right)$	$(\phi > 0)$
(M12)	Parabola	$h_i(\tau) = (N - i + 1)\phi(-a\tau^2 + b\tau + c)$	Schick-Wolverton [96]
		$F_i(\tau) = 1 - \exp\left(-\int_0^\tau h_i(s)ds\right)$	$(a, b, c > 0)$

be explained as that faults are i.i.d. within each T_i with common failure rate $\psi(\tau)$. The distribution functions of T_i are

$$F_i(\tau) = 1 - \exp\left(-\int_0^\tau h_i(s)ds\right).$$

This class of models, listed in Table 2.5, corresponds to type (d) of Table 2.2.

In model (M8), Littlewood and Verrall [60, 61] assumed that the Λ_i are independent gamma random variables with the shape parameter α and the scale parameter ψ_i, namely, $g(\lambda_i|\alpha, \psi_i) = \frac{\psi_i^\alpha}{\Gamma(\alpha)}\lambda_i^{\alpha-1}\exp(-\psi_i\lambda_i)$. Model (M9) is the JM geometric de-eutrophication model suggested by Moranda [72]. The failure rates are constant for each T_i, but they decrease geometrically in i after each failure. D is the program failure rate during the first interval and k is a constant such that the rates form a converging geometric series. When a failure occurs and a fault is removed, the drop in the failure rates is $D(1 - k)k^{i-1}$. Since $h_i(t) \to 0$ when $i \to \infty$, thus the fault content is infinite. Model (M9) has been generalized to the so-called *proportional models*, see Moranda [73], Gaudoin and Soler [30], and Gaudoin et al. [31].

Model (M10) is a power-type function of the number of remaining faults where $0^\gamma = 0$ and $\gamma > 1$ such that $h_i(t)$ is a convex function denoting that early stage of debugging has higher detection probability than later. Models (M11) and (M12) in Table 2.5 are different from models (M6) and (M7) in Table 2.3, although with the same name of distributions. For example, in model (M11) the failure rate drops to 0 immediately after the $(i-1)$th failure, while in model (M6) the failure rate drops to $(N-i+1)\phi\tau$ where τ is the time to the $(i-1)$th failure. In both models (M11) and (M6), the failure rate increases linearly with slope $(N - i + 1)$ until the ith failure.

2.4.6 NHPP Models

Assume that assumptions (A2) and (A4) in Table 2.1 hold, and that the number of initial faults N in assumption (A1) has a Poisson prior distribution with finite mean $\theta > 0$. One way of constructing an NHPP is to assume that N is Poisson distributed in the i.i.d. order statistics models discussed in Sect. 2.4.4. Let F be the common distribution of faults, usually continuous with positive support. For each fault, given that we haven't found it by time s, the conditional probability that we find it in time

Table 2.6 Type (c): $r_i(t) = \lambda(t)$; NHPP models with $m(\infty) < \infty$

(M13)	Exponential	$\lambda(t) = \theta\phi e^{-\phi t}$	GO model [34]
		$m(t) = \theta(1 - e^{-\phi t})$	$(\phi > 0)$
(M14)	Pareto	$\lambda(t) = \theta\alpha\beta^{\alpha}/(\beta + t)^{\alpha+1}$	Miller [70]
		$m(t) = \theta[1 - (\beta/(\beta + t))^{\alpha}]$	$(\alpha, \beta > 0)$
(M15)	Gamma(2,ϕ)	$\lambda(t) = \theta\phi^2 t e^{-\phi t}$	Yamada et al. [120]
		$m(t) = \theta\left[1 - (1 + \phi t)e^{-\phi t}\right]$	$(\phi > 0)$
(M16)	Logistic	$\lambda(t) = \theta(1 + c)\phi e^{-\phi t}(1 + ce^{-\phi t})^2$	Yamada et al. [121]
		$m(t) = \theta(1 - e^{-\phi t})/(1 + ce^{-\phi t})$	$(\phi, c > 0)$
(M17)	Weibull	$\lambda(t) = \theta\alpha\beta t^{\beta-1}\exp(-\alpha t^{\beta})$	Goel [33]
		$m(t) = \theta[1 - \exp(-\alpha t^{\beta})]$	$(\alpha > 0, 0 < \beta < 1)$
(M18)	Logistic	$\lambda(t) = \theta\phi\alpha e^{-\phi t}/(1 + \alpha e^{-\phi t})^2$	Yamada-Osaki [119]
		$m(t) = \theta/(1 + \alpha e^{-\phi t})$	$(\alpha, \phi > 0)$

interval $(s, t]$ is $P(s < \tau \leq t | \tau > s) = 1 - \bar{F}(t)/\bar{F}(s)$. Then for a fixed number n of faults and $n \geq u$, we have, for $k = 0, 1, \ldots, n - u$,

$$P(X(t) - X(s) = k | N = n, X(s) = u) = \binom{n-u}{k} \left(1 - \frac{\bar{F}(t)}{\bar{F}(s)}\right)^k \left(\frac{\bar{F}(t)}{\bar{F}(s)}\right)^{n-u-k},$$

which is the probability that we find k of $n - u$ independent faults in time interval $(s, t]$. Thus, for $0 < s < t$,

$$X(t) - X(s) | N, X(s) \sim \text{Binomial}(N - X(s), 1 - \bar{F}(t)/\bar{F}(s)).$$

In particular, $X(t)|N \sim \text{Binomial}(N, 1 - \bar{F}(t))$. If N has a Poisson distribution with parameter θ, then it can be shown that $\{X(t), t \geq 0\}$ is a NHPP with the mean value function $m(t) = \theta F(t)$.

In Table 2.6, we list models in which the expected number of initial faults are finite, $m(\infty) = \lim_{t\to\infty} m(t) = \theta < \infty$. Model (M14), mentioned by Miller [70], is the NHPP version of Littlewood model (M2). Models (M15) and (M16) are discussed by Yamada et al. [120, 121], and are referred to as the delayed S-shaped and the inflection S-shaped SRGMs, respectively. Both models have increasing failure rates $\psi(t)$. Model (M17) is proposed by Goel [33] to describe the situation that the intensity function first increases and then decreases. Model (M18) is the logistic growth curve model suggested by Yamada and Osaki [119]. However, we have $F(0) = 1/(1 + \alpha) > 0$. If we modify $F(t)$ by

$$F'(t) = \left(F(t) - \frac{1}{1 + \alpha}\right)\frac{1 + \alpha}{\alpha},$$

such that $F'(0) = 0$ and $F'(\infty) = 1$, then we have the same model as (M16).

Table 2.7 Type (c): $r_i(t) = \lambda(t)$; NHPP models with $m(\infty) \to \infty$

(M19)	Poisson	$\lambda(t) = \lambda$	Musa-Ackerman [74]
	($N \to \infty$ in (M1))	$m(t) = \lambda t$	($\lambda > 0$)
(M20)	Power law process	$\lambda(t) = \alpha\beta t^{\beta-1}$	Crow [21]
	($N \to \infty$ in (M5))	$m(t) = \alpha t^{\beta}$	($\alpha, \beta > 0$)
(M21)	Logarithmic Poisson	$\lambda(t) = \lambda_0/(\lambda_0\phi t + 1)$	Musa-Okumoto [75]
	($N \to \infty$ in (M2))	$m(t) = \ln(\lambda_0\phi t + 1)/\phi$	($\phi, \lambda_0 > 0$)

Another class of NHPP models can be generated directly from various mean value functions $m(t)$ or intensity functions $\lambda(t)$. Some of them are originally used in hardware system reliability, and are referred to as *reliability growth models*. A growth function (the intensity function or the mean value function) is used to represent the growth of system reliability after improvement. For the models listed in Table 2.7, the expected number of initial faults are infinite, i.e., $m(\infty) = \lim_{t\to\infty} m(t) \to \infty$.

Model (M19) is the homogeneous Poisson process appeared in Musa and Ackerman [74]. Model (20) is the Duane model or the Duane process. In light of Duane's discovery [24], Crow [21] proposes the model (M20) where the intensity function $\lambda(t) = \alpha\beta t^{\beta-1}$ is the failure rate of a Weibull distribution and the time to the first failure is Weibull distributed. Hence this model is commonly referred to as the Weibull process. As Ascher and Feingold [8] point out, the name of the Weibull process is improper. We refer this model as the *power law process* according to them. The logarithm of the mean value function is $\ln m(t) = \ln\alpha + \beta \ln t$ which is consistent with Duane's empirical observation that the cumulative number of failures versus cumulative operation hours is approximately linear on the log-log plot. It is clear that $0 < \beta < 1$ corresponds to the decreasing intensity function. In this case, the intensity function $\lambda(t) \to \infty$ as $t \to 0$ is an unrealistic property for real systems. See Littlewood [58] for a modified Duane model. Model (M14), with inverse power law intensity function $\lambda(t) = \theta\alpha\beta^{\lambda}(\beta + t)^{-\alpha-1}$, can be viewed as a modified model of (M20).

Model (M21) is Musa and Okumoto's logarithmic Poisson model [75]. They postulate that the intensity function $\lambda(t)$ is a exponentially decreasing function of the mean value function $m(t)$ given by $\lambda(t) = \lambda_0 \exp(-\phi m(t))$, where λ_0 represents the initial value of $\lambda(t)$, and ϕ measures the relative decrease in the failure intensity that follows every failure. Since $\lambda(t) = dm(t)/dt$ and $m(0) = 0$, we have $m(t) = \ln(\lambda_0\phi t + 1)/\phi$ and $\lambda(t) = \lambda_0/(\lambda_0\phi t + 1)$. By the fact that $m(t)$ is an NHPP and a logarithmic function of t, model (M21) is then called the logarithmic Poisson model.

In fact, it can be shown that Model (M21) is a limit case of the Littlewood model (M2), the Poisson model (M19) is a limit case of the JM model (M1), and that the power law process (M20) is a limit case of the Wagoner model (M5).

2.4.7 Self-exciting Point Processes

2.4.7.1 Self-exciting Point Process

Let $\{X(t), t \geq 0\}$ be a counting process, which characterizes failure occurrences of a computer software, defined on a *filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathcal{P})$ with $\{X(t), t \geq 0\}$ adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ where $\mathcal{F}_t := \sigma\{X(s), 0 \leq s \leq t\}$. A *filtration* $\{\mathcal{F}_t, t \geq 0\}$ is a family of sub-σ-fields of $\mathcal{F}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$ in $[0, \infty)$. A probability space endowed with such a filtration is called a filtered probability space. We assume that the filtration is *standard*. That is, the filtration satisfies *right-continuous*: $\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$, $\forall t \geq 0$, and *complete*: $\mathcal{F}_0$ contains all of the $\mathcal{P}$-null sets in $\mathcal{F}$. A counting process $\{X(t), t \geq 0\}$ is said to be *adapted* to the *filtration* $(\mathcal{F}_t)_{t\geq 0}$ if $X(t)$ is $\mathcal{F}_t$-measurable for each t. Thus $X(t)$ is known when $\mathcal{F}_t$ is known at time t, i.e., $\mathcal{F}_t$ represents the history or information available up to time t.

A counting process is said to be self-exciting if it depends on the entire or some fractions of its history $(\mathcal{F}_t)_{t\geq 0}$ that affects or excites the intensity function $\Lambda(t|\,\mathcal{F}_t)$ of the process defined as

$$\Lambda(t|\,\mathcal{F}_t) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} E[X(t + \Delta t) - X(t)|\,\mathcal{F}_t].$$

Assume that the process does not have failures which occur simultaneously. In other words, the process satisfies the conditional orderliness property [19]: for any $\mathcal{G}_t \subseteq \mathcal{F}_t$ and $t \geq 0$, as $\delta \downarrow 0$,

$$P(X(t + \delta) - X(t) \geq 2|\,\mathcal{G}_t) = o(\delta) P(X(t + \delta) - X(t) = 1|\,\mathcal{G}_t).$$

When such is the case, the intensity function becomes

$$\Lambda(t|\,\mathcal{F}_t) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} P(X(t + \Delta t) - X(t) = 1|\,\mathcal{F}_t), \quad S_{X(t)} \leq t < S_{X(t)+1}, \tag{2.39}$$

where S_i is the time of occurrence of the ith failure with $S_0 \equiv 0$. Let $T_i = S_i - S_{i-1}$ be the interarrival time, for $i = 1, 2, \ldots$.

A self-exciting point process takes account of the points generated by the failure times so that, for example, $\mathcal{F}_t = \{X(t), S_1, \ldots, S_{X(t)}\}$. A point process is said to have no-memory if $\Lambda(t|\,\mathcal{F}_t) = \Lambda(t)$; it has 0-memory if $\Lambda(t|\,\mathcal{F}_t) = \Lambda(t|X(t))$; and it has m-memory if $\Lambda(t|\,\mathcal{F}_t) = \Lambda(t|X(t), S_{X(t)}, \ldots, S_{X(t)-m+1})$, for $m = 1, 2, \ldots$. It has been shown that if $\{X(t), t \geq 0\}$ is m_2-memory then it is m_1-memory, for $0 \leq m_1 \leq m_2 < \infty$, and a counting process is 0-memory if and only if it is a Markov process. In addition, a self-exciting point process is 1-memory with intensity function

$$\Lambda(t|X(t), S_{X(t)}) = f(X(t), t - S_{X(t)}),$$

for some function f if and only if the times between failures $T_1, T_2, \ldots$ are independent [103]. Note that the intensity function $\Lambda(t|\,\mathcal{F}_t)$ is continuous from the left, and we assume for $t \geq 0$ that $E[\Lambda(t|\,\mathcal{F}_t)] < \infty$.

Recall that the expectation of $X(t)$ is called the mean value function, and we denote it by $m(t) = E[X(t)]$. The derivative of the mean value function is called the *rate of occurrence of failures* (ROCOF) representing approximate probability that a failure occurs in $(t, t + dt]$, and we denote it by $v(t) = \frac{d}{dt}E[X(t)]$. If the counting process $\{X(t), t \geq 0\}$ does not have failures which occur simultaneously, the ROCOF is equivalent to

$$v(t) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} P(X(t + \Delta t) - X(t) = 1). \tag{2.40}$$

$\Lambda(t|\,\mathcal{F}_t)$ is the probability of encountering a failure within the next instant of time dt, given the history available up to the present time t. For Poisson processes, due to the property of independent increments, the ROCOF (2.40) and the intensity function (2.39) are the same, but in general these two functions are different. When a system has only one r.v. T and we define $X(t) := \mathbb{I}\{T \leq t\}$, then the intensity function (2.39) is equivalent to the failure rate function (2.3). Thus, failure rate function is a special case of intensity function.

2.4.7.2 Classification of SRGMs

Another function which is used in reliability modeling (to include software reliability), is the concatenated failure rate function of the failure process. Given $\mathcal{F}_t$, it is defined by Chen and Singpurwalla [18],

$$H(\tau|\,\mathcal{F}_t) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} P(\tau + S_{X(t)} < S_{X(t)+1} \leq S_{X(t)} + \tau + \Delta t|\,\mathcal{F}_t), \quad \tau \geq 0. \tag{2.41}$$

The intensity function (2.39) and the concatenated failure rate function (2.41) can be viewed as a wilder sense of the transition rate function (2.33) and the conditional failure rate function (2.34).

Let us consider that a system consists of only one component whose lifetime is T, and define $X(t) = \mathbb{I}\{T \leq t\}$; thus, for a 1-component system $\Lambda(t|\,\mathcal{F}_t)$ is equivalent to $H(\tau|\,\mathcal{F}_t)$. As might be expected, there is a close relation between $\Lambda(t|\,\mathcal{F}_t)$ and $H(\tau|\,\mathcal{F}_t)$, similar to Theorem 2.8, as is shown in the following:

$$\begin{aligned} \Lambda(t|\,\mathcal{F}_t) &= H(t - S_{X(t)}|\,\mathcal{F}_t), \quad S_{X(t)} \leq t < S_{X(t)+1}, \\ H(\tau|\,\mathcal{F}_t) &= \Lambda(\tau + S_{X(t)}|\,\mathcal{F}_t), \quad 0 \leq \tau < S_{X(t)+1} - S_{X(t)}. \end{aligned}$$

Table 2.8 Relations between $\Lambda(t|\,\mathcal{F}_t)$ and $H(\tau|\,\mathcal{F}_t)$

(a)	$\Lambda(t\|\,\mathcal{F}_t) = (N - X(t))\psi(t)$	$\Longrightarrow$	$H(\tau\|\,\mathcal{F}_t) = (N - X(t))\psi(\tau + S_{X(t)})$
(b)	$\Lambda(t\|\,\mathcal{F}_t) = (N - X(t))\psi_{X(t)+1}(t)$	$\Longrightarrow$	$H(\tau\|\,\mathcal{F}_t) = (N - X(t))\psi_{X(t)+1}(\tau + S_{X(t)})$
(c)	$\Lambda(t\|\,\mathcal{F}_t) = \lambda(t)$	$\Longrightarrow$	$H(\tau\|\,\mathcal{F}_t) = \lambda(\tau + S_{X(t)})$
(d)	$H(\tau\|\,\mathcal{F}_t) = (N - X(t))\psi(\tau)$	$\Longrightarrow$	$\Lambda(t\|\,\mathcal{F}_t) = (N - X(t))\psi(t - S_{X(t)})$
(e)	$H(\tau\|\,\mathcal{F}_t) = \psi_{X(t)+1}(\tau)$	$\Longrightarrow$	$\Lambda(t\|\,\mathcal{F}_t) = \psi_{X(t)+1}(t - S_{X(t)})$

Accordingly, this relation enables us to indicate a method to classify software reliability models. The implication is that if we can specify $\Lambda(t|\,\mathcal{F}_t)$ then we have $H(\tau|\,\mathcal{F}_t)$, and vice versa. Here we give five cases as examples in Table 2.8, similar to Table 2.2, where $\lambda(\cdot)$ and $\psi(\cdot)$ are continuous functions.

Obviously, $\Lambda(t|\,\mathcal{F}_t)$ is no-memory in case (c), 0-memory in cases (a) and (b), and 1-memory in cases (d) and (e). Except for case (b), the given cases correspond to some categories of software reliability models in the context of Chen and Singpurwalla [18]. Thus, case (a) corresponds to the class of i.i.d. order statistic models, case (c) corresponds the class of NHPP models, and cases (d) and (e) correspond to the concatenated failure rate models. For $m \geq 2$ memory models, we refer the readers to Singpurwalla and Soyer [101] and Al-Mutairi et al. [3].

Observe that the relation between the intensity function (2.39) and the concatenated failure rate function (2.41) is analogous to the relation of the transition rate function (2.33) and the conditional failure rate function (2.34). The difference is on the filtration $\mathcal{F}_t = \sigma\{X(s) : 0 \leq s \leq t\}$. Based on these relations, a software reliability model can be represented simply by the intensity function according to various filtrations we might choose. Thus, a more general class of models can be constructed from a more wilder filtration.

2.4.7.3 A Self-exciting and Mixture Model

Apart from the self-exciting, another property for the model in our discussion is mixture which is assumed that individual faults come with i.i.d. random failure rates λ and that failure times have intensity function $\phi(t|\lambda, \mathcal{F}_t)$. Let $X_i(t)$ be the indicator function of failure due to fault i up to time $t \geq 0$. Therefore, the cumulative number of software failures up to time $t \geq 0$ is given by $X(t) = \sum_{i=1}^{N_0} X_i(t)$, where N_0 is the unknown initial number of faults in the software.

Now we can discuss an intensity function which have both properties of self-exciting and mixture. Assumptions are given in the following.

(i) The X_i are mutually independent, given N_0, $\{\lambda_i\}_{i\geq 0}$, and $\mathcal{F}_t$,

(ii) $$\phi(t|\lambda_i,\ \mathcal{F}_t) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} P(X_i(t + \Delta t) - X_i(t) = 1|N_0, \{\lambda_i\}_{i\geq 0},\ \mathcal{F}_t), \tag{2.42}$$

(iii) The failure process satisfies the conditional orderliness property,

where λ_i is the random failure rate of faults i. The intensity function of the software can be determined and

$$\Lambda(t|N_0,\ \{\lambda_i\}_{i\geq 0},\ \mathcal{F}_t) = \lim_{\Delta t\to 0^+} \frac{1}{\Delta t} P(X(t+\Delta t) - X(t) = 1|N_0, \{\lambda_i\}_{i\geq 0},\ \mathcal{F}_t). \tag{2.43}$$

The following results are provided by Wang [113].

Theorem 2.10. *Under the conditions of (2.43), if the λ_i have a mixing distribution function G, then the intensity function of $\{X(t), t \geq 0\}$, given N_0 and $\mathcal{F}_t = \sigma\{X(s), 0 \leq s \leq t\}$, is*

$$\Lambda(t|N_0, \mathcal{F}_t) = (N_0 - X(t)) \cdot \nu(\lambda, \mathcal{F}_t, G) \tag{2.44}$$

where

$$\nu(\lambda, \mathcal{F}_t, G) = \frac{\int_0^\infty \phi(t|\lambda,\ \mathcal{F}_t) \exp[-\Phi(t|\lambda,\ \mathcal{F}_t)] dG(\lambda)}{\int_0^\infty \exp[-\Phi(t|\lambda,\ \mathcal{F}_t)] dG(\lambda)},$$

and

$$\Phi(t|\lambda) = \int_0^t \phi(s|\lambda,\ \mathcal{F}_s) ds.$$

Example 2.32 (Littlewood [57]). Assume that the mixing distribution G is gamma distributed with the scale parameter β and the shape parameter α. If $\phi(t|\lambda,\ \mathcal{F}_t) \equiv \lambda$, then the intensity function is

$$\Lambda(t|N_0) = (N_0 - X(t)) \frac{\alpha}{\beta + t}.$$

This is an example corresponds to case (a) of Table 2.8.

Example 2.33 (Goel and Okumoto [34]; Miller [70]). Assume that $N_0 \sim Poisson$ (θ) and $\phi(t|\lambda,\ \mathcal{F}_t) \equiv \lambda$. If the λ_i degenerate to a constant λ_0, then the intensity function is

$$\Lambda(t) = \theta\lambda_0 e^{-\lambda_0 t}.$$

Furthermore, if the mixing distribution G is gamma distributed with the scale parameter β and the shape parameter α then the intensity function is

$$\Lambda(t) = \frac{\theta\alpha\beta^\alpha}{(\beta + t)^{\alpha+1}}.$$

These are examples correspond to case (c) of Table 2.8.

Example 2.34 (Schick and Wolverton [95]). Assume that the λ_i degenerate to a constant λ_0 and $\phi(t|\lambda_0,\ \mathcal{F}_t) = \lambda_0(t - S_{X(t)})$, then the intensity function is

$$\Lambda(t|N_0,\ \mathcal{F}_t) = (N_0 - X(t))\lambda_0(t - S_{X(t)}).$$

The concatenated failure rate function is

$$H(\tau|N_0,\ \mathcal{F}_t) = (N_0 - X(t))\lambda_0\tau, \quad \tau \geq 0.$$

This is an example corresponds to case (d) of Table 2.8.

2.5 Examples

2.5.1 A Model for Load-Sharing System with Common-Cause Systems

In Sect. 2.2.2, we have discussed the MVE and FMVE models. The idea here is to combine both models together. Consider a system consisting of n components whose lifetimes have the occurrence of common-cause failures and the increase of failure rates for the surviving components. Such phenomena happen especially in redundant systems [76]. A series of Poisson shocks can be used to describe such failures, where each shock may destroy components separately or simultaneously. Accordingly, failures of the system can be considered to be n possible transitions of a Markov chain. One (or more) components fails first, then one (or more) of the surviving components fails, and so on. Initially there is $2^n - 1$ independent Poisson processes that govern the occurrence of the shocks. After the system receives a shock, which destroys one (or more) of the n components, the surviving components are subject to shocks governed by another cluster of independent Poisson processes which may have different parameters (failure rates) from the previous ones due to an increased load. At this stage the number of processes will depend on the number of surviving components. The system may continue to function, depending on the structure of the system (e.g., a k-out-of-n:F system), and is waiting for receiving another shock.

Let $X = (X_t)_{t\geq 0}$ be the failure process of the system where X_t denotes the number of failed components up to time t. The state space E consists of 2^n states and we partition it into $n+1$ classes as $E = \{E_0, E_1, \ldots, E_n\}$, where the subscript of E_i denotes number i of failed components and the elements of E_i indicate $\binom{n}{i}$ combinations of failed components. We denote these classes by the notations: $E_0 = \{0\}$, $E_1 = \{1_1, \ldots, 1_n\}$, $\ldots$, $E_n = \{n\}$, where the permutation of subscript in each class is in lexicographical order and the cardinality of E_i is $|E_i| = \binom{n}{i}$. Class E_n is the absorbing class containing only one state, and other classes constitute a partition of the set of transient states. One of the simplest example is a birth process which passes through all $n+1$ classes by $E_0 \to E_1 \to \cdots \to E_n$. In our discussions, after visiting class E_i, the next visited class may be any class E_j with $i < j \leq n$.

Let Q be an infinitesimal generator of the failure process X, and the submatrix from class E_i to class E_j be denoted by $Q_{i,j}$ which is defined by

$$Q_{i,j} = \lim_{h \downarrow 0^+} \frac{1}{h} P\{X_{t+h} \in E_j | X_t \in E_i\}, \quad \text{if } i < j,$$

$$Q_{i,j} = \lim_{h \downarrow 0^+} \frac{1}{h} (P\{X_{t+h} \in E_j | X_t \in E_i\} - I), \quad \text{if } i = j,$$

where I is an identity matrix. The submatrix $Q_{i,j}$ is a $\binom{n}{i} \times \binom{n}{j}$ constant matrix which is a zero matrix if $i > j$. The entries of $Q_{i,j}$ are denoted by ${}^k\lambda^l_{i,j}$, where $k = 1, \ldots, \binom{n}{i}$ and $l = 1, \ldots, \binom{n}{j}$, and ${}^k\lambda^k_{i,i} = -\sum_{i<j} \sum_{l \neq k} {}^k\lambda^l_{i,j} < 0$.

The permutation of subscripts of the $Q_{i,j}$'s are again arranged in lexicographical order. Such arrangement guarantees that Q is a upper triangular matrix. For each state s, we let $P_s(t) = P\{X_t = s | X_0 = 0\}$ as usual. We assemble the functions $P_s(t)$ for $s \in E_j$ into a row vector with $\binom{n}{j}$ entries in lexicographical order and denote it by $P_{E_j}(t)$.

$$Q = \begin{array}{c} \\ E_0 \\ E_1 \\ E_2 \\ \vdots \\ E_{n-1} \\ E_n \end{array} \begin{array}{c} \begin{array}{ccccccc} E_0 & E_1 & E_2 & \cdots & E_{n-2} & E_{n-1} & E_n \end{array} \\ \begin{pmatrix} Q_{0,0} & Q_{0,1} & Q_{0,2} & \cdots & Q_{0,n-2} & Q_{0,n-1} & Q_{0,n} \\ 0 & Q_{1,1} & Q_{1,2} & \cdots & Q_{1,n-2} & Q_{1,n-1} & Q_{1,n} \\ 0 & 0 & Q_{2,2} & \cdots & Q_{2,n-2} & Q_{2,n-1} & Q_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & Q_{n-1,n-1} & Q_{n-1,n} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix} \end{array}.$$

In particular, $Q_{0,0} = {}^0\lambda^0_{1,1}$, $Q_{n,n} = 0$, and for $j = 1, \ldots, n-1$,

$$Q_{j,j} = \begin{pmatrix} {}^1\lambda^1_{j,j} & 0 & \cdots & 0 & 0 \\ 0 & {}^2\lambda^2_{j,j} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & {}^{|E_j|-1}\lambda^{|E_j|-1}_{j,j} & 0 \\ 0 & 0 & \cdots & 0 & {}^{|E_j|}\lambda^{|E_j|}_{j,j} \end{pmatrix}$$

is a $|E_j| \times |E_j|$ diagonal matrix where $|E_j| = \binom{n}{j}$ the cardinality of the E_j class.

The transient state probabilities $P_{E_j}(t)$ can be shown to be

$$P_{E_0}(t) = \exp(Q_{0,0}t),$$

$$P_{E_j}(t) = \sum_{k=0}^{j-1} \sum_{0=i_0<i_1<\cdots<i_k<j} W_{0,i_1} \circ W_{i_1,i_2} \circ \cdots \circ W_{i_k,j}(t) \exp(Q_{j,j}t), \quad (2.45)$$

for $j = 1, \ldots, n-1$, where

$$W_{i,j}(t) = \int_0^t \exp[Q_{i,i}(\tau)] Q_{i,j} \exp[-Q_{j,j}(\tau)] d\tau,$$

and the function $A \circ B(t)$ is a convolution defined by $A \circ B(t) := \int_0^t A(\tau) dB(\tau)$.

The proof of (2.45) and related reliability properties are presented by Wang [114]. To illustrate the usefulness of the infinitesimal generator representation, we investigate the following 2- and 3-component systems.

For a 2-component system, state 0 denotes no failure, and State 1_1, 1_2, 2 denote the failure of components 1, 2, and both components, respectively. The infinitesimal generator of the given model is

$$\mathbf{Q} = \begin{array}{c} \\ 0 \\ 1_1 \\ 1_2 \\ 2 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1_1 & 1_2 & 2 \end{array} \\ \begin{pmatrix} -(\lambda_1 + \lambda_2 + \lambda_{12}) & \lambda_1 & \lambda_2 & \lambda_{12} \\ 0 & -(\lambda_2' + \lambda_{12}) & 0 & \lambda_2' + \lambda_{12} \\ 0 & 0 & -(\lambda_1' + \lambda_{12}) & \lambda_1' + \lambda_{12} \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array}. \quad (2.46)$$

Let $P_n(t) = P(X_t = n | X_0 = 0)$ be the transient state probabilities, given the initial condition $X_0 = 0$. From the Kolmogorov forward equation (2.28) with initial condition $P_0(0) = 1$, we have the solution

$$\begin{aligned} P_0(t) &= e^{-\lambda t}, \\ P_{1_1}(t) &= \begin{cases} \frac{\lambda_1}{\lambda_1 + \lambda_2 - \lambda_2'} \left[e^{-(\lambda_2' + \lambda_{12})t} - e^{-\lambda t} \right], & \lambda_1 + \lambda_2 \neq \lambda_2' \\ \lambda_1 t e^{-\lambda t}, & \lambda_1 + \lambda_2 = \lambda_2', \end{cases} \\ P_{1_2}(t) &= \begin{cases} \frac{\lambda_2}{\lambda_1 + \lambda_2 - \lambda_1'} \left[e^{-(\lambda_1' + \lambda_{12})t} - e^{-\lambda t} \right], & \lambda_1 + \lambda_2 \neq \lambda_1' \\ \lambda_2 t e^{-\lambda t}, & \lambda_1 + \lambda_2 = \lambda_1', \end{cases} \end{aligned}$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$.

$P_0(t)$ is the probability of no component failing up to time t, and it is equivalent to the probability that the system is considered as a series structure so that $R_{1:2}(t) = P_0(t) = e^{-\lambda t}$. The system reliability for the parallel structure is the probability that at most one component fails in this 2-component system. It follows that

$$\begin{aligned} R_{2:2}(t) &= P_0(t) + P_{1_1}(t) + P_{1_2}(t) \\ &= \begin{cases} \kappa_1 e^{-(\lambda_2' + \lambda_{12})t} + \kappa_2 e^{-(\lambda_1' + \lambda_{12})t} + (1 - \kappa_1 - \kappa_2) e^{-\lambda t}, & \lambda_1 + \lambda_2 \neq \lambda_1', \lambda_2' \\ \kappa_1 e^{-(\lambda_2' + \lambda_{12})t} + \lambda_2 t e^{-(\lambda_1' + \lambda_{12})t} + (1 - \kappa_1) e^{-\lambda t}, & \lambda_1 + \lambda_2 = \lambda_1' \neq \lambda_2' \\ \kappa_2 e^{-(\lambda_1' + \lambda_{12})t} + \lambda_1 t e^{-(\lambda_2' + \lambda_{12})t} + (1 - \kappa_2) e^{-\lambda t}, & \lambda_1 + \lambda_2 = \lambda_2' \neq \lambda_1' \\ [1 + (\lambda_1 + \lambda_2)t] e^{-\lambda t}, & \lambda_1 + \lambda_2 = \lambda_1' = \lambda_2', \end{cases} \end{aligned}$$

where $\kappa_1 = \frac{\lambda_1}{\lambda_1+\lambda_2-\lambda_2'}$ and $\kappa_2 = \frac{\lambda_2}{\lambda_1+\lambda_2-\lambda_1'}$. When $\lambda_{12} = 0$, i.e., no common-cause occurs, this model reduces to the FBVE model. If there is no change of failure rate on first failure, i.e., $\lambda_1' = \lambda_1$ and $\lambda_2' = \lambda_2$, the model reduces to the BVE model.

Let $T_1(T_2)$ be the time at which component 1(2) fails. If the generator is given as (2.46), then the joint density function of T_1 and T_2 can be shown to be

$$f(t_1,t_2) = \begin{cases} \lambda_1(\lambda_2' + \lambda_{12})e^{-(\lambda_1+\lambda_2-\lambda_2')t_1-(\lambda_2'+\lambda_{12})t_2}, & 0 < t_1 < t_2 \\ \lambda_2(\lambda_1' + \lambda_{12})e^{-(\lambda_1'+\lambda_{12})t_1-(\lambda_1+\lambda_2-\lambda_1')t_2}, & 0 < t_2 < t_1, \end{cases} \tag{2.47}$$

for the absolutely continuous part, and

$$g(t) = \lambda_{12}e^{-\lambda t}, \tag{2.48}$$

for the singular part. Equations (2.47) and (2.48) lead to the joint reliability function

$$\bar{F}(t_1,t_2) = \begin{cases} \kappa_1 e^{-(\lambda_1+\lambda_2-\lambda_2')t_1-(\lambda_2'+\lambda_{12})t_2} + (1-\kappa_1)e^{-\lambda t_2}, & t_1 < t_2 \\ \kappa_2 e^{-(\lambda_1+\lambda_2-\lambda_1')t_2-(\lambda_1'+\lambda_{12})t_1} + (1-\kappa_2)e^{-\lambda t_1}, & t_2 < t_1, \end{cases} \tag{2.49}$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}, \kappa_1 = \frac{\lambda_1}{\lambda_1+\lambda_2-\lambda_2'}$ and $\kappa_2 = \frac{\lambda_2}{\lambda_1+\lambda_2-\lambda_1'}$. For convenience, we call the distribution of (2.49) the bivariate exponential shared-load distribution (BVESL).

For a 3-component system, state 0 denotes no failure, and states 1_1, 1_2 and 1_3 denote respectively the failure of component 1, 2 and 3, states 2_{12}, 2_{13} and 2_{23} denote respectively the failure of two components 1 and 2, 1 and 3, and 2 and 3, and state 3 denotes the failure of all three components. The infinitesimal generator of the given model is

$$\mathbf{Q} = \begin{array}{c} \\ 0 \\ 1_1 \\ 1_2 \\ 2_{12} \\ 1_3 \\ 2_{13} \\ 2_{23} \\ 3 \end{array} \begin{array}{c} \begin{array}{cccccccc} 0 & 1_1 & 1_2 & 2_{12} & 1_3 & 2_{13} & 2_{23} & 3 \end{array} \\ \begin{pmatrix} -\alpha_0 & \lambda_1 & \lambda_2 & \lambda_{12} & \lambda_3 & \lambda_{13} & \lambda_{23} & \lambda_{123} \\ 0 & -\alpha_1 & 0 & \lambda_2' + \lambda_{12} & 0 & \lambda_3' + \lambda_{13} & 0 & \lambda_{23} + \lambda_{123} \\ 0 & 0 & -\alpha_2 & \lambda_1' + \lambda_{12} & 0 & 0 & \lambda_3' + \lambda_{23} & \lambda_{13} + \lambda_{123} \\ 0 & 0 & 0 & -\alpha_{12} & 0 & 0 & 0 & \alpha_{12} \\ 0 & 0 & 0 & 0 & -\alpha_3 & \lambda_1' + \lambda_{13} & \lambda_2' + \lambda_{23} & \lambda_{12} + \lambda_{123} \\ 0 & 0 & 0 & 0 & 0 & -\alpha_{13} & 0 & \alpha_{13} \\ 0 & 0 & 0 & 0 & 0 & 0 & -\alpha_{23} & \alpha_{23} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array},$$

where

$$\begin{aligned} \alpha_0 &= \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{12} + \lambda_{13} + \lambda_{23} + \lambda_{123}, \\ \alpha_1 &= \lambda_2' + \lambda_3' + \lambda_{12} + \lambda_{13} + \lambda_{23} + \lambda_{123}, \\ \alpha_2 &= \lambda_1' + \lambda_3' + \lambda_{12} + \lambda_{23} + \lambda_{13} + \lambda_{123}, \\ \alpha_3 &= \lambda_1' + \lambda_2' + \lambda_{13} + \lambda_{23} + \lambda_{12} + \lambda_{123}, \end{aligned}$$

$$\alpha_{12} = \lambda_3'' + \lambda_{13} + \lambda_{23} + \lambda_{123},$$
$$\alpha_{13} = \lambda_2'' + \lambda_{12} + \lambda_{23} + \lambda_{123},$$
$$\alpha_{23} = \lambda_1'' + \lambda_{12} + \lambda_{13} + \lambda_{123}.$$

Note that the upper $(2^2 \times 2^2)$ submatrix denotes the failure process of components 1 and 2. In order to obtain the reliability functions of the system, we need to find the transient state probabilities $P_i(t) = P(X_t = i | X_0 = 0)$, $i = 0, 1_1, 1_2, 1_3, 2_{12}, 2_{13}, 2_{23}$. It follows that the reliability functions are $R_{1:3}(t) = P_0(t)$, $R_{2:3}(t) = P_0(t) + \sum_{i=1}^{3} P_{1_i}(t)$, and $R_{3:3}(t) = P_0(t) + \sum_{i=1}^{3} P_{1_i}(t) + \sum_{i<j} P_{2_{ij}}(t)$.

Example 2.35. Assume that the failure rates of a 3-component system are

$\lambda_1 = \lambda_2 = \lambda_3 = 0.01$ failure/h, $\lambda_{12} = \lambda_{13} = \lambda_{23} = 0.002$ failure/h, $\lambda_{123} = 0.001$ failure/h
$\lambda_1' = \lambda_2' = \lambda_3' = 0.012$ failure/h, $\lambda_1'' = \lambda_2'' = \lambda_3'' = 0.015$ failure/h.

The infinitesimal generator of the system is

$$\mathbf{Q} = \begin{array}{c} \\ 0 \\ 1_1 \\ 1_2 \\ 2_{12} \\ 1_3 \\ 2_{13} \\ 2_{23} \\ 3 \end{array} \begin{array}{c} \begin{array}{cccccccc} 0 & 1_1 & 1_2 & 2_{12} & 1_3 & 2_{13} & 2_{23} & 3 \end{array} \\ \begin{pmatrix} -0.037 & 0.01 & 0.01 & 0.002 & 0.01 & 0.002 & 0.002 & 0.001 \\ 0.0 & -0.031 & 0.0 & 0.014 & 0.0 & 0.014 & 0.0 & 0.003 \\ 0.0 & 0.0 & -00.031 & 0.014 & 0.0 & 0.0 & 0.014 & 0.003 \\ 0.0 & 0.0 & 0.0 & -0.020 & 0.0 & 0.0 & 0.0 & 0.020 \\ 0.0 & 0.0 & 0.0 & 0.0 & -0.031 & 0.014 & 0.014 & 0.003 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & -0.020 & 0.0 & 0.020 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & -0.020 & 0.020 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{pmatrix} \end{array}.$$

The transient state probabilities are given by

$$P_{00}(t) = e^{-0.037t},$$
$$P_{01_1}(t) = p_{01_2}(t) = p_{01_3}(t) = -1.6667e^{-0.037t} + 1.6667e^{-0.031t},$$
$$P_{02_{12}}(t) = p_{02_{13}}(t) = p_{02_{23}}(t) = 2.6275e^{-0.037t} - 4.2424e^{-0.031t} + 1.615e^{-0.02t}.$$

It follows that the system reliability functions are given by

$$R_{1:3}(t) = e^{-0.037t},$$
$$R_{2:3}(t) = 5e^{-0.031t} - 4e^{-0.037t},$$
$$R_{3:3}(t) = 4.8449e^{-0.02t} - 7.7272e^{-0.031t} + 3.8824e^{-0.037t}.$$

By integrating system reliability functions over the time interval $[0, \infty]$, we obtain the MTTFs for i-out-of-3:F systems, $i = 1, 2, 3$:

$$\text{MTTF}_{1:3} = 27.027 \text{ h}, \qquad \text{MTTF}_{2:3} = 53.182 \text{ h}, \qquad \text{MTTF}_{3:3} = 97.91 \text{ h}.$$

2.5.2 A Shared-Load k-out-of-n:G Repairable System

Shao and Lamberson [97] proposed an approach to compute and analyze a shared-load k-out-of-n:G repairable system. The assumptions of the model are given as follows:

Assumptions:

1. The failure rate of all functioning components in the system is the same and constant in each success state. When a component fails, the system state changes and the failure rate for every functioning component changes. For i components functioning, every functioning component has the constant failure rate, λ_i, $i = k, \ldots, n$. And $\lambda_n < \lambda_{n-1} < \cdots < \lambda_k$.
2. A failed component must be detected and disconnected by a controller. The probability of successful detection and switching of the controller is α, for each failure event. If the controller cannot detect and disconnect a failed unit or the controller itself has failed, the system fails. The controller failure rate is a constant, λ_c. The controller is never repaired or replaced during a mission.
3. At most r components can be in repair at one time; the constant repair rate is μ. The repair rate for j components failed is: $\mu_j = \mu \cdot \min\{j, r\}$.
4. A repaired component is as good as new and is immediately re-connected to the system.
5. The switch-over time and the time for re-energizing a repaired component are negligible.

Based on the above assumptions, a continuous-time Markov chain can be used to create a set of differential equations. The state space for the system is given below.

State Space:

- State j $(j = 0, 1, \ldots, n-k)$: j components have failed and have been disconnected from the network, the remaining $(n-j)$ components and the controller are functioning.
- State $n-k+1$: The system fails because only $(k-1)$ components are functioning. However, the system can return to the working state $(n-k)$ at a repair rate μ_{n-k+1}.
- State f: The system fails because the controller cannot detect and disconnect a failed component (Fig. 2.10).

The system state space $E = \{0, 1, \ldots, j, \ldots, n-k, n-k+1, f\}$, and $E = S \cup F$, where the failed state space F consists of failed states $n-k+1$ and f, while the success state space S includes the remaining states.

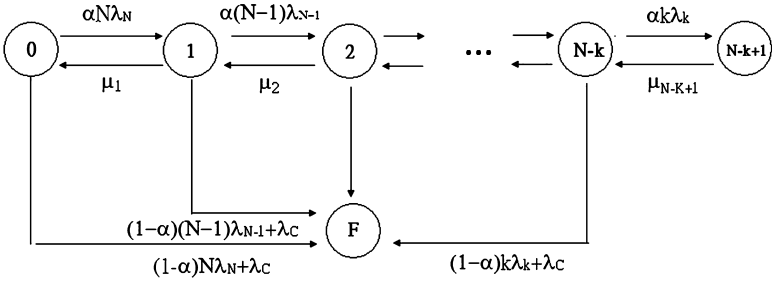


Fig. 2.10 Markov transition diagram for a shared-load k-out-of-n:G system

Generator: The infinitesimal generator of the given model is

$$Q = \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & \dots & n-k & n-k+1 & f \end{array} \\ \begin{pmatrix} -n\lambda_n - \lambda_c & \alpha n\lambda_n & & & & (1-\alpha)n\lambda_n + \lambda_c \\ \mu_1 & -\mu_1 - (n-1)\lambda_{n-1} - \lambda_c & & & & (1-\alpha)(n-1)\lambda_{n-1} + \lambda_c \\ & \mu_2 & \ddots & & & \vdots \\ & & \ddots & & & \vdots \\ & & & \alpha(k+1)\lambda_{k+1} & & (1-\alpha)(k+1)\lambda_{k+1} + \lambda_c \\ & & & -\mu_{n-k} - k\lambda_k - \lambda_c & \alpha k\lambda_k & (1-\alpha)k\lambda_k + \lambda_c \\ & & & \mu_{n-k+1} & -\mu_{n-k+1} & \\ & & & & & 0 \end{pmatrix} \end{array}.$$

Differential Equations: The set of differential equations based on the generator is

$$\begin{aligned} P_0'(t) &= [-n\lambda_n - \lambda_c]P_0(t) + \mu_1 P_1(t), \\ P_j'(t) &= [\alpha(n-j+1)\lambda_{n-j+1}]P_{j-1}(t) + [-\mu_j - (n-j)\lambda_{n-j} - \lambda_c]P_j(t) \\ &\quad + [\mu_{j+1}]P_{j+1}(t), \quad j = 1, 2, \dots, n-k, \\ P_{n-k+1}'(t) &= [\alpha k\lambda_k]P_{n-k}(t) + [-\mu_{n-k+1}]P_{n-k+1}(t), \\ P_f'(t) &= \sum_{j=0}^{n-k} [(1-\alpha)(n-j)\lambda_{n-j} + \lambda_c]P_j(t). \end{aligned}$$

The initial conditions for this set of linear differential equations are: $P_0(0) = 1$ and $P_i(0) = 0$ for $i \in E - \{0\}$.

System Availability & Reliability: The system availability $A(t)$ is the probability that the system is surviving at time t:

$$A(t) = \sum_{i \in S} P_i(t) = \sum_{j=0}^{n-k} P_j(t).$$

The $P_j(t)$ can be determined by integrating the set of linear differential equations numerically. Availability can be considered as the generalized reliability when the system is repairable or component-replaceable.

System reliability is the availability of the same system without repair and replacement. Let $\mu_j = 0$, for $j = 1, 2, \ldots, n-k+1$, then the set of linear differential equations becomes

$$\begin{aligned} P_0'(t) &= -(\beta_n + \lambda_c) P_0(t), \\ P_j'(t) &= (\alpha\beta_{n-j+1}) P_{j-1}(t) - (\beta_{n-j} + \lambda_c) P_j(t), \quad j = 1, 2, \ldots, n-k, \\ P_{n-k+1}'(t) &= (\alpha\beta_k) P_{n-k}(t), \end{aligned} \tag{2.50}$$

where $\beta_j = j\lambda_j$, $j = k, k+1, \ldots, n$. Taking the Laplace transform of (2.50), we have the following expressions:

$$\begin{aligned} sP_0^*(s) - 1 &= -(\beta_n + \lambda_c) P_0^*(s), \\ sP_1^*(s) &= \alpha\beta_n P_0^*(s) - (\beta_{n-1} + \lambda_c) P_1^*(s), \\ sP_j^*(s) &= (\alpha\beta_{n-j+1}) P_{j-1}^*(s) - (\beta_{n-j} + \lambda_c) P_j^*(s), \quad j = 2, \ldots, n-k. \end{aligned}$$

Thus,

$$P_j^*(s) = \frac{\alpha\beta_{n-j+1}}{s + \beta_{n-j} + \lambda_c} P_{j-1}^*(s) = \frac{\prod_{i=1}^{j} \alpha\beta_{n-i+1}}{\prod_{i=0}^{j} (s + \beta_{n-i} + \lambda_c)}, \quad j = 1, 2, \ldots, n-k.$$

Next, we use an inverse Laplace transform to solve the above equations. The approach is to expand the product fraction into a partial fraction form:

$$\frac{1}{\prod_{i=0}^{j} (s + c_i)} = \sum_{i=0}^{j} \left[\prod_{k=0, k\neq i}^{j} \frac{1}{(c_k - c_i)} \right] \frac{1}{s + c_i}, \quad j = 1, 2, \ldots, n-k.$$

Thus, the result is $P_0(t) = e^{-(\beta_n + \lambda_c)t}$ and

$$P_j(t) = \prod_{i=1}^{j} \alpha\beta_{n-i+1} \cdot \sum_{i=0}^{j} \frac{e^{-(\beta_{n-i} + \lambda_c)t}}{\prod_{k=0, k\neq i}^{j} (\beta_{n-k} - \beta_{n-i})}, \quad j = 1, 2, \ldots, n-k. \tag{2.51}$$

The reliability function of the shared-load system is $R(t) = \sum_{j=0}^{n-k} P_j(t)$.

Special cases: (1) If the absorbing state is removed from the chain by considering perfect fault coverage ($\alpha = 1$ and $\lambda_c = 0$), the result is a birth-death process. (2) If we further assumed that the components are nonrepairable and $\lambda_i = \lambda$ for $i = 0, 1, \ldots, n$, then the shared-load system reduces to a regular k-out-of-n system.

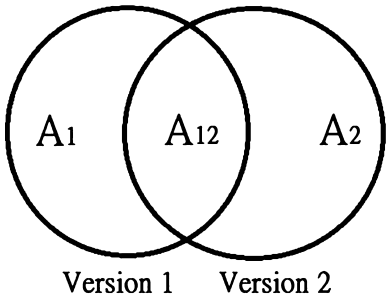


Fig. 2.11 Software faults in the 2-version software system

2.5.3 A Dependent Model for Fault-Tolerant Software Systems During Debugging

N-version programming (NVP) is one of the most important software fault-tolerance techniques. Numerous researchers have studied the effect of NVP for the improvement of software reliability on fault correlation theoretically [25, 27, 53, 59, 80, 87, 106] as well as empirically [15, 16, 26, 44, 47, 65, 66]. However, only a few of them consider this issue during the testing and debugging part of the software development life cycle. During testing and debugging, faults may not be successfully removed. Imperfect debugging may result in unsuccessful removal, and the introduction of new faults. In this example we review a model proposed by Wang [115], who developed a bivariate counting process for a 2VP by assuming positive dependency among versions. Considering imperfect debugging, this bivariate process characterizes dynamic changes of fault contents for each version during testing and debugging. Further explanations for the assumptions and proofs of the given model, see Wang [115].

2.5.3.1 Model Assumptions

The fault content of a 2VP system is divided into three components, according to their independence. That is, the system is composed of three mutually exclusive components, as depicted in Fig. 2.11. Here A_1 and A_2 represent independent faults which do not cause failing of the other version, and A_{12} represents related faults that affect both versions.

Assumptions:

1. Both software versions can fail on the sample input, which can be caused by either the related faults (A_{12}) or the independent faults (A_1 and A_2).
2. Related and independent faults are mutually exclusive.
3. When a failure occurs at a version, a debugging effort occurs immediately. The fault content of that version is instantaneously reduced by one, increased by one, or remains the same.
4. An independent fault can be successfully removed with high probability.

5. Debugging of a related fault may result in successful removal (reduce size of A_{12} by one), introduction of a new related fault (increase size of A_{12} by one), or unsuccessful removal (size of A_{12} has no change).
6. In terms of unsuccessful debugging of a related fault in A_{12}, it can become an independent fault in either A_1 or A_2.
7. The version-failure detection rate at any time is proportional to its current fault content (the number of faults remaining in the version at that time).
8. The fault removal rate of version 1 (2) is proportional to the current fault content of version 1 (2).
9. The fault introduction rate of version 1 is proportional to the current fault content of version 2, and vice versa. This mainly reflects the common assumption of positive correlated failures.
10. The removal rate per fault and introduction rate per fault are constant.
11. The simultaneous occurrences of multiple failures in a version are not allowed (i.e., the 'orderly property').

2.5.3.2 The Proposed Model

According to the assumptions, the model developed here is a bivariate counting process. Let $N_i(t)$ be the fault content (total number of faults) of version i at time t. The transition probabilities in $(N_1(t), N_2(t))$ during the time interval $(t, t+dt)$ are

$$\begin{aligned} &P(N_1(t+dt) - N_1(t) = i, N_2(t+dt) - N_2(t) = j | N_1(t) = x, N_2(t) = y) \\ &\quad = f_{ij}(x, y)dt, \end{aligned}$$

where i and j are not both zero and $f_{ij}(x, y)$ are suitable non-negative functions of x and y. Assume that

$$\begin{aligned} &f_{-1,0}(x, y) = \alpha_1 x, \ f_{0,-1}(x, y) = \alpha_2 y, \\ &f_{1,0}(x, y) = \beta_1 y, \quad f_{0,1}(x, y) = \beta_2 x, \end{aligned} \tag{2.52}$$

where $\alpha_1, \alpha_2, \beta_1$ and β_2 are unknown constants, although it is possible to let them be functions of time. Let the joint state probability of $N_1(t)$ and $N_2(t)$ at time t be given by $P_{m,n}(t) = P(N_1(t) = m, N_2(t) = n)$. The fundamental equations governing this bivariate counting process are the following Kolmogorov forward equations [9]

$$\begin{aligned} P'_{m,n}(t) &= \alpha_1(m+1)P_{m+1,n} + \alpha_2(n+1)P_{m,n+1} + \beta_1 n P_{m-1,n} \\ &\quad + \beta_2 m P_{m,n-1} - [(\alpha_1 + \beta_2)m + (\alpha_2 + \beta_1)n]P_{m,n}, \quad \text{for } m, n \geq 0. \end{aligned}$$

Let $M(\theta, \phi, t)$ be the moment generating function and it is given by

$$M(\theta, \phi, t) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} e^{m\theta + n\phi} P_{m,n}(t).$$

It can be shown that [9]

$$\frac{\partial M(\theta,\phi,t)}{\partial t} = \sum_{i,j} \left(e^{i\theta+j\phi} - 1\right) f_{ij}\left(\frac{\partial}{\partial\theta}, \frac{\partial}{\partial\phi}\right) M(\theta,\phi,t), \tag{2.53}$$

where i and j are not both zero. From (2.52) and (2.53), the partial differential equation of the moment generating function is given by

$$\frac{\partial M}{\partial t} = \left[\alpha_1(e^{-\theta} - 1) + \beta_2(e^{\phi} - 1)\right]\frac{\partial M}{\partial\theta} + \left[\alpha_2(e^{-\phi} - 1) + \beta_1(e^{\theta} - 1)\right]\frac{\partial M}{\partial\phi},$$

with initial condition $M(\theta,\phi,0) = e^{b_1\theta+b_2\phi}$. It is sometimes easier to work with the cumulant generating function, $K(\theta,\phi,t)$, which is defined by $K = \log M$. The partial differential equation of the cumulant generating function is

$$\frac{\partial K}{\partial t} = \left[\alpha_1(e^{-\theta} - 1) + \beta_2(e^{\phi} - 1)\right]\frac{\partial K}{\partial\theta} + \left[\alpha_2(e^{-\phi} - 1) + \beta_1(e^{\theta} - 1)\right]\frac{\partial K}{\partial\phi}, \tag{2.54}$$

with initial condition $K(\theta,\phi,0) = b_1\theta + b_2\phi$. Express the cumulant generating function into a Taylor expansion

$$K(\theta,\phi,t) = \sum_{i,j\geq 0} K_{ij}(t)\frac{\theta^i}{i!}\frac{\phi^j}{j!},$$

where $K_{ij}(t)$ are cumulants and $K_{00} \equiv 0$. It can be shown that the first and second order cumulants correspond to the expected number, variances, and covariance of fault contents as [9]

$$\begin{aligned} K_{10}(t) &= E[N_1(t)], & K_{01}(t) &= E[N_2(t)],\\ K_{20}(t) &= \mathrm{Var}[N_1(t)], & K_{02}(t) &= \mathrm{Var}[N_2(t)],\\ K_{11}(t) &= \mathrm{Cov}[N_1(t), N_2(t)]. \end{aligned}$$

To obtain the desired cumulants, we use the following procedure. Taking the first and second order cumulants and equating coefficients of $\theta, \phi, \theta^2, \theta\phi$, and ϕ^2 on both sides of (2.54) give the following differential equations

$$\begin{aligned} \frac{dK_{10}}{dt} &= -\alpha_1 K_{10} + \beta_1 K_{01},\\ \frac{dK_{01}}{dt} &= \beta_2 K_{10} - \alpha_2 K_{01},\\ \frac{dK_{20}}{dt} &= -2\alpha_1 K_{20} + \alpha_1 K_{10} + \beta_1 K_{01} + 2\beta_1 K_{11}, \end{aligned}$$

$$\frac{dK_{11}}{dt} = \beta_2 K_{20} - (\alpha_1 + \alpha_2) K_{11} + \beta_1 K_{02},$$
$$\frac{dK_{02}}{dt} = \beta_2 K_{10} + \alpha_2 K_{01} + 2\beta_2 K_{11} - 2\alpha_2 K_{02}. \tag{2.55}$$

These equations can be solved by the general ordinary differential equation method, using the initial conditions $K_{10}(0) = b_1$ and $K_{01}(0) = b_2$, and all other cumulants being zero. Solving $dK_{10}/dt = -\alpha_1 K_{10} + \beta_1 K_{01}$ and $dK_{01}/dt = \beta_2 K_{10} - \alpha_2 K_{01}$, we have

$$K_{10}(t) = A_1 e^{\lambda_1 t} + A_2 e^{\lambda_2 t}, \tag{2.56}$$
$$K_{01}(t) = B_1 e^{\lambda_1 t} + B_2 e^{\lambda_2 t}, \tag{2.57}$$

where A_1, A_2, B_1 and B_2 are constants and

$$\lambda_1 = \frac{-(\alpha_1 + \alpha_2) + \sqrt{(\alpha_1 - \alpha_2)^2 + 4\beta_1\beta_2}}{2}, \tag{2.58}$$
$$\lambda_2 = \frac{-(\alpha_1 + \alpha_2) - \sqrt{(\alpha_1 - \alpha_2)^2 + 4\beta_1\beta_2}}{2}. \tag{2.59}$$

Let $C = \lambda_1 - \lambda_2 = \sqrt{(\alpha_1 - \alpha_2)^2 + 4\beta_1\beta_2}$, then coefficients of A_1, A_2, B_1 and B_2 are

$$A_1 = \frac{b_1(C - \alpha_1 + \alpha_2) + 2\beta_1 b_2}{2C}, \tag{2.60}$$
$$A_2 = \frac{b_1(C + \alpha_1 - \alpha_2) - 2\beta_1 b_2}{2C}, \tag{2.61}$$
$$B_1 = \frac{\alpha_1 - \alpha_2 + C}{2\beta_1} A_1, \tag{2.62}$$
$$B_2 = \frac{\alpha_1 - \alpha_2 - C}{2\beta_1} A_2. \tag{2.63}$$

Parameters $\alpha_1, \alpha_2, \beta_1$ and β_2 can be expressed as, by substituting (2.56) and (2.57) into (2.55),

$$\alpha_1 = \frac{A_1 B_2 \lambda_1 - A_2 B_1 \lambda_2}{A_2 B_1 - A_1 B_2}, \tag{2.64}$$
$$\alpha_2 = \frac{A_2 B_1 \lambda_1 - A_1 B_2 \lambda_2}{A_1 B_2 - A_2 B_1}, \tag{2.65}$$
$$\beta_1 = \frac{A_1 A_2 C}{A_2 B_1 - A_1 B_2}, \tag{2.66}$$

$$\beta_2 = \frac{B_1 B_2 C}{A_1 B_2 - A_2 B_1}. \tag{2.67}$$

Thus, if we have numerical estimates for values of A_1, A_2, B_1, B_2 and λ_1, λ_2, then $\alpha_1, \alpha_2, \beta_1$ and β_2 are given by (2.64)–(2.67), respectively.

Since β_1 and β_2 are all positive, the roots λ_1 and λ_2 given in (2.58) and (2.59) are always real. If $\alpha_1\alpha_2 < \beta_1\beta_2$, then one of the λ_i is positive and the other is negative. If $\alpha_1\alpha_2 > \beta_1\beta_2$, both λ_1 and λ_2 are negative. This suggests that the result of eventually extinguishing the faults is certain, if $\alpha_1\alpha_2 > \beta_1\beta_2$.

Special case 1. Let $\alpha_1 = \alpha_2 \equiv \alpha$ and $\beta_1 = \beta_2 \equiv \beta$. This case describes the situation that both versions affect each other in a 'symmetric' fashion. It can be shown that the expected number of faults are $K_{10}(t) = \frac{b_1+b_2}{2} e^{-(\alpha-\beta)t} + \frac{b_1-b_2}{2} e^{-(\alpha+\beta)t}$ and $K_{01}(t) = \frac{b_1+b_2}{2} e^{-(\alpha-\beta)t} - \frac{b_1-b_2}{2} e^{-(\alpha+\beta)t}$.

Special case 2. Let $\beta_1 = 0$ and $\alpha_1 \neq \alpha_2$. In this case, only version 2 is affected by version 1. The expected number of faults are $K_{10}(t) = b_1 e^{-\alpha_1 t}$ and $K_{01}(t) = b_2 e^{-\alpha_2 t} + \frac{b_1\beta_2}{\alpha_1-\alpha_2} [e^{-\alpha_2 t} - e^{-\alpha_1 t}]$. Further, if $\alpha_1 = \alpha_2 \equiv \alpha$, the expected number of faults are $K_{10}(t) = b_1 e^{-\alpha t}$ and $K_{01}(t) = b_2 e^{-\alpha t} + b_1\beta_2 t e^{-\alpha t}$.

Special case 3. Let $\beta_1 = 0$ and $\beta_2 = 0$. This case describes that the failure behavior of both versions are s-independent. As such the expected number of faults are $K_{10}(t) = b_1 e^{-\alpha_1 t}$ and $K_{01}(t) = b_2 e^{-\alpha_2 t}$. In fact, this is the JM model [42] for two independent versions.

2.5.3.3 Expectation of Failures

The bivariate counting process by itself cannot fully describe the debugging process, for example it cannot provide the total number of failures up to time t. We need to consider both failure content and fault content. A failure occurs when the actual output deviates from the expected output as a result of executing the faulty program. For each version, when a failure has occurred it can reduce one fault, increase one fault, or give no change of the fault content. The associated fault content and failure content can also be studied by a multivariate processes. Let

$N_i(t) \equiv$ the fault content of version i at time t, $i = 1, 2$,

$M_i(t) \equiv$ the number of observed failures of version i up to time t, $i = 1, 2$.

Set

$$P(N_1(t+dt) - N_1(t) = i, N_2(t+dt) - N_2(t) = j, M_1(t+dt) - M_1(t) = k,$$
$$M_2(t+dt) - M_2(t) = l \mid N_1(t) = x, N_2(t) = y) = f_{i,j,k,l}(x, y)dt.$$

Then we assume that

$$f_{-1,0,1,0}(x,y) = \alpha_1 x,\ f_{1,0,1,0}(x,y) = \beta_1 y,$$
$$f_{0,-1,0,1}(x,) = \alpha_2 y,\quad f_{0,1,0,1}(x,y) = \beta_2 x.$$

Follow the similar procedure discussed in Sect. 2.5.3.2, we get the partial differential equation of the cumulant generating function $K(\theta_1, \theta_2, \theta_3, \theta_4)$

$$\frac{\partial K}{\partial t} = \left[\alpha_1(e^{-\theta_1+\theta_3} - 1) + \beta_2(e^{\theta_2+\theta_4} - 1)\right]\frac{\partial K}{\partial \theta_1} + \left[\beta_1(e^{\theta_1+\theta_3} - 1) + \alpha_2(e^{-\theta_2+\theta_4} - 1)\right]\frac{\partial K}{\partial \theta_2}.$$

In addition to (2.56) and (2.57), the expected numbers of failures are

$$E[M_1(t)] = \int_0^t \left[(\alpha_1 A_1 + \beta_1 B_1)e^{\lambda_1 \tau} + (\alpha_1 A_2 + \beta_1 B_2)e^{\lambda_2 \tau}\right] d\tau,$$
$$E[M_2(t)] = \int_0^t \left[(\beta_2 A_1 + \alpha_2 B_1)e^{\lambda_1 \tau} + (\beta_2 A_2 + \alpha_2 B_2)e^{\lambda_2 \tau}\right] d\tau,$$

where A_1, A_2, B_1 and B_2 are expressed in (2.60)–(2.63), respectively. Moreover,

$$\lim_{t\to\infty} E[M_1(t)] = \begin{cases} \frac{\alpha_1 A_1 + \beta_1 B_1}{(-\lambda_1)} + \frac{\alpha_1 A_2 + \beta_1 B_2}{(-\lambda_2)} & \text{if } \lambda_1 < 0 \text{ and } \lambda_2 < 0 \\ \infty & \text{otherwise,} \end{cases} \tag{2.68}$$

and

$$\lim_{t\to\infty} E[M_2(t)] = \begin{cases} \frac{\beta_2 A_1 + \alpha_2 B_1}{(-\lambda_1)} + \frac{\beta_2 A_2 + \alpha_2 B_2}{(-\lambda_2)} & \text{if } \lambda_1 < 0 \text{ and } \lambda_2 < 0 \\ \infty & \text{otherwise.} \end{cases} \tag{2.69}$$

2.5.3.4 Reliability

Suppose that n_1 faults of version 1 and n_2 faults of version 2 still remain in the program at time t_e. The overall failure rate at that time would be $n_1(\alpha_1(t_e) + \beta_2(t_e)) + n_2(\alpha_2(t_e) + \beta_1(t_e))$. If the system fails when one of its versions fail, then the reliability of the entire system as a function of time t takes the form

$$R_{Dep}(t|t_e) = [R_1(t|t_e)]^{n_1}\,[R_2(t|t_e)]^{n_2},$$

if the system fails only when both its versions fail, the reliability of the entire system is

$$R_{Dep}(t|t_e) = 1 - \left(1 - [R_1(t|t_e)]^{n_1}\right)\left(1 - [R_2(t|t_e)]^{n_2}\right), \tag{2.70}$$

where $R_1(t|t_e) = \exp\{-(\alpha_1 + \beta_2)t\}$ and $R_2(t|t_e) = \exp\{-(\alpha_2 + \beta_1)t\}$. The function $R_1(t|t_e)$ and $R_2(t|t_e)$ can be viewed as reliabilities attributable to each of the individual faults of versions 1 and 2, respectively.

2.5.3.5 Estimation of Parameters

In fact, the actual fault contents for each version cannot be observed. However, the number of failures can be observed. Thus, we use the number of observed failures to estimate the unknown parameters, $\alpha_1, \alpha_2, \beta_1$ and β_2. We can minimize the objective function

$$S = \sum_{i=1}^{m} \left\{ \left(\widehat{M_1}(t_i) - EM_1(t_i)\right)^2 + \left(\widehat{M_2}(t_i) - EM_2(t_i)\right)^2 \right\},$$

i.e., the sum of two least squares functions, where $\widehat{M_1}(t_i)$ and $\widehat{M_2}(t_i)$ represent the number of observed failures of versions 1 and 2 up to time t_i, respectively. However, the estimates for $\alpha_1, \alpha_2, \beta_1$ and β_2 may not be very reliable since a small change in λ_1 and λ_2 may magnify the effect in these estimates.

2.5.3.6 Example

We illustrate the proposed model by using the observed failure data of the water reservoir control (WRC) system [106]. Water is supplied via a source pipe controlled by a source valve and removed via a drain pipe controlled by a drain valve. The WRC system achieves fault-tolerance and high reliability by using NVP software control logic with $N = 2$. This 2VP system application assumes that the reliability of the voter is 1, and that the voter can identify exactly which version(s) is failed when a failure occurs [106]. Following these assumptions, the 2VP system fails only when both its software versions fail at the same input data. This WRC system data set is listed in Table 2.9.

Apply the method in Sect. 2.5.3.5 to the WRC data set, the LSEs of $\alpha_1, \alpha_2, \beta_1$ and β_2 are

$$\hat{\alpha}_1 = 0.00568, \qquad \hat{\beta}_1 = 0.000399,$$
$$\hat{\alpha}_2 = 0.00735, \qquad \hat{\beta}_2 = 0.000874.$$

Based on estimated values of α_i and β_i, related measurements can be obtained. Since $\hat{\alpha}_1\hat{\alpha}_2 > \hat{\beta}_1\hat{\beta}_2$, the perfect debugging (elimination of all faults) is certain. Substituting $\hat{\alpha}_i$ and $\hat{\beta}_i$ into (2.56) and (2.57) and setting $t = 0$, we have the initial fault contents for both versions

$$\hat{b_1} = 29.594 \qquad \text{and} \qquad \hat{b_2} = 25.004.$$

Table 2.9 Failure-time normalized-data of WRC 2VP system

	Failure time			Failure time	
Failure no.	Version 1	Version 2	Failure no.	Version 1	Version 2
1	1.2	3.6	14	39.2	34.8
2	2.8	8.4	15	40.0	36.4
3	8.4	12.8	16	44.0	36.8
4	10.0	14.4	17	44.8	38.0
5	16.4	17.2	18	54.0	39.2
6	20.0	18.0	19	56.0	41.6
7	24.4	20.0	20	62.4	42.0
8	28.0	23.2	21	80.0	46.4
9	29.2	25.2	22	92.0	59.6
10	31.2	28.0	23	99.6	62.4
11	34.0	28.4	24		98.8
12	36.0	30.8	25		99.6
13	36.8	31.2	26		100.0

This shows that the initial number of faults in version 1 is bigger than that in version 2. The expected numbers of failures eventually encountered for both versions are obtained by substituting estimated values of parameters into (2.68) and (2.69). The numbers are 32.826 and 34.608 for version 1 and version 2, respectively. This shows that we expect that more failures occur in version 2 than that in version 1. Initially the number of faults in version 1 is bigger than version 2. However, we expect that version 1 has less failures eventually encountered than version 2. This shows that version 1 may have a better debugging performance than version 2. Part of the reason may be attributed to dependency between versions. From (2.56) and (2.57), we have the expected numbers of remaining faults for versions at time t

$$E[N_1(t)] = (-2.162)e^{-0.00754t} + (31.755)e^{-0.00549t},$$
$$E[N_2(t)] = (10.083)e^{-0.00754t} + (14.921)e^{-0.00549t}.$$

If these two versions are assumed to be independent of each other, then the expected numbers of remaining faults at time t are given by $E_I[N_1(t)] = (31.697)e^{-0.00564t}$ and $E_I[N_2(t)] = (33.600)e^{-0.00629t}$. From (2.70), the reliability for the 2VP system when $t_e = 250$ is

$$R_{Dep}(t|250) = 1 - (1 - e^{-0.0506t})(1 - e^{-0.04117t}).$$

References

1. O.O. Aalen, Ø. Borgan, H.K. Gjessing, *Survival and Event History Analysis: A Process Point of View* (Springer, New York, 2008)
2. A.A. Abdel-Ghaly, P.Y. Chan, B. Littlewood, Evaluation of competing software reliability predictions. IEEE Trans. Softw. Eng. **12**, 950–967 (1986)

3. D. Al-Mutairi, Y. Chen, N.D. Singpurwalla, An adaptative concatenated failure rate model for software reliability. J. Am. Stat. Assoc. **93**, 1150–1163 (1998)
4. S.V. Amari, R.B. Misra, Closed-form expressions for distribution of sum of exponential random variables. IEEE Trans. Reliab. **46**(4), 519–522 (1997)
5. S.V. Amari, J.B. Dugan, R.B. Misra, Optimal reliability of systems subject to imperfect fault coverage. IEEE Trans. Reliab. **48**(3), 275–284 (1999)
6. S.V. Amari, K.B. Misra, H. Pham, Tampered failure rate load-sharing systems: status and perspectives, in *Handbook on Performability Engineering*, ed. by K.B. Misra (Springer, London, 2008)
7. P.M. Anderson, S.K. Agarwal, An improved model for protective-system reliability. IEEE Trans. Reliab. **41**(3), 422–426 (1992)
8. H. Ascher, H. Feingold, *Repairable Systems Reliability: Modeling, Inference, Misconceptions and Their Causes* (Marcel Dekker, New York, 1984)
9. N.T.J. Bailey, *The Elements of Stochastic Processes* (Wiley, New York, 1964)
10. N. Balakrishnan, C.D. Lai, *Continuous Bivariate Distributions* (Springer, New York, 2009)
11. R.E. Barlow, F. Prochan, *Statistical Theory of Reliability and Life Testing: Probability Models* (To Begin With, Silver Spring, 1981)
12. H.W. Block, A.P. Basu, A continuous bivariate exponential extension. J. Am. Stat. Assoc. **69**, 1031–1037 (1974)
13. P.J. Boland, N.N. Chuív, Optimal times for software release when repair is imperfect. Stat. Probab. Lett. **77**, 1176–1184 (2007)
14. L.C. Briand, K.E. Emam, B.G. Freimut, A comprehensive evaluation of capture-recapture models for estimating software defect content. IEEE Trans. Softw. Eng. **26**(6), 518–540 (2000)
15. S.S. Brilliant, J.C. Knight, N.G. Leveson, Analysis of faults in an N-version software experiment. IEEE Trans. Softw. Eng. **16**(2), 238–247 (1990)
16. X. Cai, M.R. Lyu, M.A. Vouk, An experimental evaluation on reliability features of N-version programming, in *Proceeding of the 16th International Symposium on Software Reliability Engineering*, Chicago, 2005
17. K.C. Chae, G.M. Clark, System reliability in the presence of common-cause failures. IEEE Trans. Reliab. **35**(1), 32–35 (1986)
18. Y. Chen, N.D. Singpurwalla, Unification of software reliability models by self-exciting point processes. Adv. Appl. Probab. **29**, 337–352 (1997)
19. D.R. Cox, V. Isham, *Point Processes* (Chapman & Hall, London, 1980)
20. D.R. Cox, P.A.W. Lewis, *The Statistical Analysis of Series of Events* (Chapman & Hall, London, 1966)
21. L.H. Crow, Reliability analysis for complex repairable systems, in *Reliability and Biometry: Statistical Analysis of Lifelength*, ed. by F. Proschan, R.J. Serfling (SIAM, Philadelphia, 1974)
22. M.J. Crowder, A.C. Kimber, R.L. Smith, T.J. Sweeting, *Statistical Analysis of Reliability Data* (Chapman & Hall, London, 1991)
23. T. Dohi, S. Osaki, K.S. Trivedi, An infinite server queueing approach for describing software reliability growth: unified modeling and estimation framework, in *Proceedings of the 11th Asia-Pacific Software Engineering Conference*, Busan, 2004
24. J.T. Duane, Learning curve approach to reliabilitymonitoring. IEEE Trans. Aerosp. **2**, 563–566 (1964)
25. D.E. Eckhardt, L.D. Lee, A theoretical basis for the analysis of multiversion software subject to coincident errors. IEEE Trans. Softw. Eng. **11**(12), 1511–1517 (1985)
26. D.E. Eckhardt, A.K. Caglavan, J.C. Knight, L.D. Lee, D.F. McAllister, M.A. Vouk, J.P. Kelly, An experimental evaluation of software redundancy as a strategy for improving reliability. IEEE Trans. Softw. Eng. **17**(7), 692–702 (1991)
27. M. Ege, M.A. Eyler, M.U. Karakas, Reliability analysis in N-version programming with dependent failures, in *Proceedings of the 27th Euromicro Conference*, Warsaw, 2001
28. K.N. Fleming, N. Mosleh, R.K. Deremer, A systematic procedure for incorporation of common cause events into risk and reliability models. Nucl. Eng. Des. **93**, 245–273 (1986)

29. J.E. Freund, A bivariate extension of the exponential distribution. J. Am. Stat. Assoc. **56**, 971–977 (1961)
30. O. Gaudoin, J.L. Soler, Statistical analysis of the geometric de-eutrophication software-reliability model. IEEE Trans. Reliab. **41**(4), 518–524 (1992)
31. O. Gaudoin, C. Lavergne, J.L. Soler, A generalized geometric de-eutrophication software-reliability model. IEEE Trans. Reliab. **43**(4), 536–541 (1994)
32. B.V. Gnedenko, Y.K. Belyayev, A.D. Solovyev, *Mathematical Methods of Reliability Theory* (Academic, New York, 1969)
33. A.L. Goel, Software reliability models: assumptions, limitations, and applicability. IEEE Trans. Softw. Eng. **11**(12), 1411–1423 (1985)
34. A.L. Goel, K. Okumoto, Time-dependent error-detection rate model for software reliability and other performance measures. IEEE Trans. Reliab. **28**(3), 206–211 (1979)
35. S.S. Gokhale, P.N. Marinos, K.S. Trivedi, Important milestomes in software reliability modeling, in *Proceedings of Software Engineering and Knowledge Engineering (SEKE) '96*, Lake Tahoe, 1996
36. K. Goševa-Popstojanova, K.S. Trivedi, Architecture-based approaches to reliability assessment of software systems. Perform. Eval. **45**, 179–204 (2001)
37. M.H. Halstead, *Elements of Software Science* (North-Holland, Amsterdam, 1977)
38. P. Hokstad, M. Rausand, Common cause failure modeling: status and trends, in *Handbook of Performability Engineering*, ed. by K.B. Misra (Springer, London, 2008)
39. C.Y. Huang, W.C. Huang, Software reliability analysis and measurement using finite and infinite server queueing models. IEEE Trans. Reliab. **57**(1), 192–203 (2008)
40. C.Y. Huang, M.R. Lyu, S.Y. Kuo, A unified scheme of some nonhomogenous Poisson process models for software reliability estimation. IEEE Trans. Softw. Eng. **29**(3), 261–269 (2003)
41. L. Huang, Q. Xu, Lifetime reliability for load-sharing redundant systems with arbitrary failure distributions. IEEE Trans. Reliab. **59**(2), 319–330 (2010)
42. Z. Jelinski, P.B. Moranda, Software reliability research, in *Statistical Computer Performance Evaluation*, ed. by W. Freiberger (Academic, New York, 1972)
43. H. Joe, *Multivariate Models and Dependence Concepts* (Chapman & Hall, London, 1997)
44. P.K. Kapur, A. Gupta, P.C. Jha, Reliability growth modeling and optimal release policy under fuzzy environment of an N-version programming system incorporating the effect of fault removal efficiency. Int. J. Autom. Comput. **4**(4), 369–379 (2007)
45. M. Kijima, *Markov Processes for Stochastic Modeling* (Chapman & Hall, London, 1997)
46. B.A. Kitchenham, B. Littlewood, *Measurement for Software Control and Assurance* (Elsevier, London, 1989)
47. J.C. Knight, N.G. Leveson, An experimental evaluation of the assumption of independence in multiversion programming. IEEE Trans. Softw. Eng. **12**(1), 96–109 (1986)
48. K.A.H. Kobbacy, D.N.P. Murthy, *Complex System Maintenance Handbook* (Springer, London, 2008)
49. W. Kremer, Birth-death and bug counting. IEEE Trans. Reliab. **32**(1), 37–47 (1983)
50. P.H. Kvam, E.A. Peña, Estimating load-sharing properties in a dynamic reliability system. J. Am. Stat. Assoc. **100**, 262–272 (2005)
51. N. Langberg, N.D. Singpurwalla, A unification of some software reliabilitymodels. SIAM J. Sci. Stat. Comput. **6**, 781–790 (1985)
52. J. Ledoux, Software reliability modeling, in *Handbook of Reliability Engineering*, ed. by H. Pham (Springer, London, 2003)
53. H.H. Lin, K.H. Chen, R.T. Wang, A multivariate exponential shared-load model. IEEE Trans. Reliab. **42**(1), 165–171 (1993)
54. B.H. Lindqvist, On the statistical modeling and analysis of repairable systems. Stat. Sci. **21**(4), 532–551 (2006)
55. B.H. Lindqvist, G. Elvebakk, K. Heggland, The trend-renewal process for statistical analysis of repairable systems. Technometrics **45**, 31–44 (2003)
56. B. Littlewood, Theories of software reliability: how good are they and how can they be improved? IEEE Trans. Softw. Eng. **6**(5), 489–500 (1980)

57. B. Littlewood, Stochastic reliability-growth: a model for fault-removal in computer-programs and hardware-designs. IEEE Trans. Reliab. **30**(4), 313–320 (1981)
58. B. Littlewood, Rationale for a modified Duane model. IEEE Trans. Reliab. **33**(2), 157–159 (1984)
59. B. Littlewood, D. Miller, Conceptual modeling of coincident failures in multiversion software. IEEE Trans. Softw. Eng. **15**(12), 1596–1614 (1989)
60. B. Littlewood, J.L. Verrall, A Bayesian reliability growth model for computer software. J. R. Stat. Soc. C **22**, 332–346 (1973)
61. B. Littlewood, J.L. Verrall, A Bayesian reliability model with a stochastically monotone failure rate. IEEE Trans. Reliab. **23**(2), 108–114 (1974)
62. H. Liu, Reliability of a load-sharing k-out-of-n:G system: non-iid components with arbitrary distributions. IEEE Trans. Reliab. **47**(3), 279–284 (1998)
63. J.C. Lu, Weibull extensions of the Freund and Marshall-Olkin bivariate exponential models. IEEE Trans. Reliab. **38**(5), 615–619 (1989)
64. M.R. Lyu, *Handbook of Software Reliability Engineering* (McGraw-Hill, New York/IEEE Computer Society, Los Angeles, 1996)
65. M.R. Lyu, Y. He, Improving the n-version programming process through the evolution of a design paradigm. IEEE Trans. Reliab. **42**(2), 179–189 (1993)
66. M.R. Lyu, Z. Huang, K.S. Sze, X. Cai, An empirical study on testing and fault tolerance of software reliability engineering, in *Proceedings of the 14th International Symposium on Software Reliability Engineering (ISSRE 2003)*, Denver, 2003
67. A.W. Marshall, I. Olkin, A multivariate exponential distribution. J. Am. Stat. Assoc. **62**, 30–44 (1967)
68. T.J. McCabe, A compexity measure. IEEE Trans. Softw. Eng. **2**(4), 308–320 (1976)
69. W.Q. Meeker, L.A. Escobar, *Statistical Methods for Reliability Data* (Wiley, New York, 1998)
70. D.R. Miller, Exponential order statistic models of software reliability growth. IEEE Trans. Softw. Eng. **12**(1), 12–24 (1986)
71. K.B. Misra, *Handbook of Performability Engineering* (Springer, London, 2008)
72. P.B. Moranda, Predictions of software reliability during debugging, in *Proceedings Annual Reliability and Maintainability Symposium*, Washington, DC, 1975
73. P.B. Moranda, Event-altered rate models for general reliability analysis. IEEE Trans. Reliab. **28**(5), 376–381 (1979)
74. J.D. Musa, A.F. Ackerman, Quantifying software validation: when to stop testing? IEEE Softw. **6**, 19–27 (1989)
75. J.D. Musa, K. Okumoto, A logarithmic Poisson execution time model for software reliability measurement, in *Proceedings 7th International Conference on Software Engineering*, Orlando, 1984
76. J.D. Musa, A. Iannino, K. Okumoto, *Software Reliability: Measurement, Prediction, Application* (McGraw-Hill, New York, 1987)
77. A. Myers, *Complex System Reliability: Multichannel Systems with Imperfect Fault Coverage* (Springer, London, 2010)
78. T. Nakagawa, *Maintenance Theory of Reliability* (Springer, London, 2005)
79. NEA, International common-cause failure data exchange. ICDE general coding guidelines. Technical note NEA/CSNI/R(2004)4. Nuclear Energy Agency (2004)
80. V.F. Nicola, A. Goyal, Modelling of correlated failures and community error recovery in multiversion software. IEEE Trans. Softw. Eng. **16**(3), 350–359 (1990)
81. E. Parzen, *Stochastic Processes* (Holden Day, San Francisco, 1962)
82. E.A. Peña, Dynamic modeling and statistical analysis of event times. Stat. Sci. **21**(4), 487–500 (2006)
83. H. Pham, *Software Reliability* (Springer, New York, 2000)
84. H. Pham, *Handbook of Reliability Engineering* (Springer, London, 2003)
85. H. Pham, *Recent Advances in Reliability and Quality in Design* (Springer, London, 2008)
86. H. Pham, L. Nordmann, X. Zhang, A general imperfect-software-debugging model with s-shaped fault-detection rate. IEEE Trans. Reliab. **48**(2), 169–175 (1999)

87. P.T. Popov, L. Strigini, J. May, S. Kuball, Estimating bounds on the reliability of diverse systems. IEEE Trans. Softw. Eng. **29**(4), 345–359 (2003)
88. P. Pukite, J. Pukite, *Modeling for Reliability Analysis* (IEEE, New York, 1998)
89. L. Rade, Expected time to failure of reliability systems. Math. Sci. **14**, 24–37 (1989)
90. A.E. Raftery, Inference and prediction for a general order statistic model with unknown population size. J. Am. Stat. Assoc. **82**, 1163–1168 (1987)
91. S. Ramani, S.S. Gokhale, K.S. Trivedi, in *Computer Performance Evaluation: Modelling Techniques and Tools, SREPT: Software Reliability Estimation and Prediction Tool,* ed. by R. Puigjaner et al. Lecture Notes in Computer Science (LNCS 1469) (Springer, New York, 1998)
92. M. Rausand, A. Høyland, *System Reliability Theory: Models, Statistical Methods, and Applications* (Wiley, Hoboken, 2004)
93. S.M. Ross, *Introduction to Probability Models*, 9th edn. (Academic, London, 2007)
94. E.M. Scheuer, Reliability of an m-out-of-n system when component failure induces higher failure rates in survivors. IEEE Trans. Reliab. **37**(1), 73–74 (1988)
95. G.J. Schick, R.W. Wolverton, Assessment of software reliability, in *Proceedings of the Operations Research* (Physica, Wirzberg-Wien, 1973). September 1972 in Hamburg
96. G.J. Schick, R.W. Wolverton, An analysis of competing software reliability models. IEEE Trans. Softw. Eng. **4**, 104–120 (1978)
97. J. Shao, L.R. Lamberson, Modeling a shared-load k-out-of-n:G system. IEEE Trans. Reliab. **40**(2), 205–209 (1991)
98. K. Sharma, R. Garg, C.K. Nagpal, R.K. Garg, Selection of optimal software reliability growth models using a distance based approach. IEEE Trans. Reliab. **59**(2), 266–276 (2010)
99. N.D. Singpurwalla, Foundational issues in reliability and risk analysis. SIAM Rev. **30**(2), 264–282 (1988)
100. N.D. Singpurwalla, *Reliability and Risk: A Bayesian Perspective* (Wiley, New York, 2006)
101. N.D. Singpurwalla, R. Soyer, Non-homogeneous autoregressive processes for tracking (software) reliability growth, and their Bayesian analysis. J. R. Stat. Soc. B **54**, 145–156 (1992)
102. N.D. Singpurwalla, S.P. Wilson, Software reliability modeling. Int. Stat. Rev. **62**, 289–317 (1994)
103. D.L. Snyder, M.I. Miller, *Random Point Processes in Time and Space*, 2nd edn. (Springer, New York, 1991)
104. H.S. Son, M.C. Kim, Software faults and reliability, in *Reliability and Risk Issues in Large Scale Safety-Critical Digital Control Systems*, ed. by P.H. Seong (Springer, London, 2009)
105. Z. Tang, J.B. Dugan, An integrated method for incorporating common cause failures in system analysis, in *Proceedings of the 50th Annual Reliability and Maintainability Symposium*, Los Angeles, 2004
106. X. Teng, H. Pham, A software-reliability growth model for N-version programming systems. IEEE Trans. Reliab. **51**(3), 311–321 (2002)
107. K. Tokuno, S. Yamada, An imperfect debugging model with two types of hazard rates for software reliability measurement and assessment. Math. Comput. Model. **31**, 343–352 (2000)
108. U.S. Nuclear Regulatory Commission, *Reactor Safety: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants*. WASH-1400, NUREG-75/014 (U.S. Nuclear Regulatory Commission, Washington, DC, 1975)
109. E.A. van Doom, On the α-classification of birth-death and quasi-bith-death processes. Stoch Model **22**, 411–421 (2006)
110. J.K. Vaurio, An implicit method for incorporating common-cause failures in system analysis. IEEE Trans. Reliab. **47**(2), 173–180 (1998)
111. A.K. Verma, S. Ajit, M. Kumar, *Dependability of Networked Computer-Based Systems* (Springer, London, 2011)
112. W.L. Wagoner, The final report on a software reliability measurement study. TOR-0074(4112)-1. Aerospace Corporation, El Segundo (1973)

113. R. Wang, A mixture and self-exciting model for software reliability. Stat. Probab. Lett. **72**, 187–194 (2005)
114. R.T. Wang, A reliability model for multivariate exponential distributions. J. Multivar. Anal. **98**, 1033–1042 (2007)
115. R.T. Wang, A dependent model for fault tolerant software systems during debugging. IEEE Trans. Reliab. **61**(2), 504–515 (2012)
116. M. Xie, *Software Reliability Modelling* (World Scientific, Singapore, 1991)
117. L. Xing, Fault-tolerant network reliability and importance analysis using binary decision diagrams, in *Proceedings of the 50th Annual Reliability and Maintainability Symposium*, Los Angeles, 2004
118. S. Yamada, Software reliability growth models incorporating imperfect debugging with introduced faults. Electron. Commun. Jpn. Part 3 **81**(4), 33–40 (1998)
119. S. Yamada, S. Osaki, Software reliability growth modeling: models and applications. IEEE Trans. Softw. Eng. **11**(12), 1431–1437 (1985)
120. S. Yamada, M. Ohba, S. Osaki, S-shaped reliability growth modeling for software error detection. IEEE Trans. Reliab. **32**(5), 475–478 (1983)
121. S. Yamada, M. Ohba, S. Osaki, S-shaped software reliability growth models and their applications. IEEE Trans. Reliab. **33**(4), 289–292 (1984)
122. K.Z. Yang, An infinite server queueing model for software readiness assessment and related performance measures. Ph.D. Dissertation, Syracuse University, 1996
123. S. Zack, *Introduction to Reliability Analysis: Probability Models and Statistical Methods* (Springer, New York, 1992)
124. P. Zeephongsekul, G. Xia, S. Kumar, Software reliability growth models: primary failures generate secondary-faults under imperfect debugging. IEEE Trans. Reliab. **43**(3), 408–413 (1994)
125. T. Zhang, M. Horigome, Availability and reliability of system with dependent components and time-varying failure and repair rates. IEEE Trans. Reliab. **50**(2), 151–158 (2001)



http://www.springer.com/978-1-4614-4192-2

Energy-Efficient Fault-Tolerant Systems

Mathew, J.; Shafik, R.A.; Pradhan, D.K. (Eds.)

2014, XIV, 335 p. 132 illus., 51 illus. in color., Hardcover
ISBN: 978-1-4614-4192-2