Stochastic Modelling and Applied Probability 41

Terje Aven
Uwe Jensen

Stochastic Models in Reliability

Second Edition

Springer

Stochastic Mechanics
Random Media
Signal Processing and Image Synthesis
Mathematical Economics and Finance
Stochastic Optimization
Stochastic Control
Stochastic Models in Life Sciences

Stochastic Modelling and Applied Probability

(Formerly:
Applications of Mathematics)

41

For further volumes:
http://www.springer.com/series/602

Terje Aven • Uwe Jensen

Stochastic Models in Reliability

Second Edition

Springer

Terje Aven
University of Stavanger
Stavanger, Norway

Uwe Jensen
Fak. Naturwissenschaften
Inst. Angewandte Mathematik u. Statistik
Universität Hohenheim
Stuttgart, Germany

ISSN 0172-4568
ISBN 978-1-4614-7893-5 ISBN 978-1-4614-7894-2 (eBook)
DOI 10.1007/978-1-4614-7894-2
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013942488

Mathematics Subject Classification (2010): 60G, 60K, 60K10, 60K20, 90B25

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

In this second edition of the book, two major topics have been added to the original version. The first one relates to copula models (Sect. 2.3), which are used to study the effects of structural dependencies on system reliability. We believe that an introduction to the fundamental ideas and concepts of copula models is important when reviewing basic reliability theory. The second new topic we have included is maintenance optimization models under constraints (Sect. 5.5). These models have been addressed in some recent publications to meet the demand for models that adequately balance economic criteria and safety. We consider two specific models. The first is the so-called delay time model where the aim is to determine optimal inspection intervals minimizing the expected discounted costs under some safety constraints. The second model is also about optimal inspection, but here the system is represented by a monotone (coherent) structure function. In addition, we have made a number of minor adjustments to increase precision and we have also corrected misprints.

We received positive feedback to the first edition from friends and colleagues. Their hints and suggestions have been incorporated into this second edition. We thank all who contributed, by whatever means, to preparing the new edition.

Stavanger, Norway — Terje Aven
Stuttgart, Germany — Uwe Jensen

Preface to the First Edition

As can be seen from the files of the databases of *Zentralblatt/Mathematical Abstracts* and *Mathematical Reviews,* about 1 % of all mathematical publications are connected to the keyword *reliability.* This gives an impression of the importance of this field and makes it clear that it is impossible to include all the topics connected to reliability in one book. The existing literature on reliability covers *inter alia* lifetime analysis, complex systems and maintenance models, and the books by Barlow and Proschan [31, 32] can be viewed as first milestones in this area. Since then the models and tools have been developed further. The aim of *Stochastic Models in Reliability* is to give a comprehensive up-to-date presentation of some of the classical areas of reliability, based on a more advanced probabilistic framework using the modern theory of stochastic processes. This framework allows the analyst to formulate general failure models, establish formulas for computing various performance measures, as well as to determine how to identify optimal replacement policies in complex situations. A number of special cases analyzed previously can be included in this framework. Our book presents a unifying approach to some of the key research areas of reliability theory, summarizing and extending results obtained in recent years. Having future work in this area in mind, it will be useful to have at hand a general set-up where the conditions and assumptions are formulated independently of particular models.

This book comprises five chapters in addition to two appendices.

Chapter 1 gives a short introduction to stochastic models of reliability, linking existing theory and the topics treated in this book. It also contains an overview of some questions and problems to be treated in the book. In addition Sect. 1.1.6 explains why martingale theory is a useful tool for describing and analyzing the structure of complex reliability models. In the final section of the chapter we briefly discuss some important aspects of reliability modeling and analysis, and present two real-life examples. To apply reliability models in practice successfully, there are many challenges related to modeling and analysis that need to be faced. However, it is not within the scope of this

book to discuss these challenges in detail. Our text is an introduction to the topic and of motivational character.

Chapter 2 presents an overview of some parts of basic reliability theory: the theory of complex (monotone) systems, both binary and multistate systems, as well as lifetime distributions and nonparametric classes of lifetime distributions. The aim of this chapter has not been to give a complete overview of the existing theory, but to highlight important areas and give a basis for the coming chapters.

Chapter 3 presents a general set-up for analyzing failure-prone systems. A (semi-) martingale approach is adopted. This general approach makes it possible to formulate a unifying theory of both nonrepairable and repairable systems, and it includes point processes, counting processes, and Markov processes as special cases. The time evolution of the system can also be analyzed on different information levels, which is one of the main attractions of the (semi-) martingale approach. Attention is drawn to the failure rate process, which is a key parameter of the model. Several examples of application of the set-up are given, including a monotone (coherent) system of possibly dependent components, and failure time and (minimal) repair models. A model for analyzing the time to failure based on risk reserves (the difference between total income and accumulated costs of repairs) is also covered.

In the next two chapters we look more closely at types of models for analyzing situations where the system and its components could be repaired or replaced in the case of failures, and where we model the downtime or costs associated with downtimes.

Chapter 4 gives an overview of availability theory of complex systems, having components that are repaired upon failure. Emphasis is placed on monotone systems comprising independent components, each generating an alternating renewal process. Multistate systems are also covered, as well as systems comprising cold standby components. Different performance measures are studied, including the distributions of the number of system failures in a time interval and the downtime of the system in a time interval. The chapter gives a rather comprehensive asymptotic analysis, providing a theoretical basis for approximation formulae used in cases where the time interval considered is long or the components are highly available.

Chapter 5 presents a framework for models of maintenance optimization, using the set-up described in Chap. 3. The framework includes a number of interesting special cases dealt with by other authors.

By allowing different information levels, it is possible to extend, for example, the classical age replacement model and minimal repair/replacement model to situations where information is available about the underlying condition of the system and the replacement time is based on this information. Again we illustrate the applicability of the model by considering monotone systems.

Chapters 3–5 are based on stochastic process theory, including theory of martingales and point, counting, and renewal processes. For the sake of completeness and to help the reader who is not familiar with this theory,

two appendices have been included summarizing the mathematical basis and some key results. Appendix A gives a general introduction to probability and stochastic process theory, whereas Appendix B gives a presentation of results from renewal theory. Appendix A also summarizes basic notation and symbols.

Although conceived mainly as a research monograph, this book can also be used for graduate courses and seminars. It primarily addresses probabilists and statisticians with research interests in reliability. But at least parts of it should be accessible to a broader group of readers, including operations researchers and engineers. A solid basis in probability and stochastic processes is required, however. In some countries many operations researchers and reliability engineers now have a rather comprehensive theoretical background in these topics, so that it should be possible to benefit from reading the more sophisticated theory presented in this book. To bring the reliability field forward, we believe that more operations researchers and engineers should be familiar with the probabilistic framework of modern reliability theory. Chapters 1 and 2 and the first part of Chaps. 4 and 5 are more elementary and do not require the more advanced theory of stochastic processes.

References are kept to a minimum throughout, but readers are referred to the bibliographic notes following each chapter, which give a brief review of the material covered and related references.

Acknowledgments

We express our gratitude to our institutions, the Stavanger University College, the University of Oslo, and the University of Ulm, for providing a rich intellectual environment, and facilities indispensable for the writing of this book. The authors are grateful for the financial support provided by the Norwegian Research Council and Deutscher Akademischer Austauschdienst. We would also like to acknowledge our indebtedness to Jelte Beimers, Jørund Gåsemyr, Harald Haukås, Tina Herberts, Karl Hinderer, Günter Last, Volker Schmidt, Richard Serfozo, Marcel Smith, Fabio Spizzichino and Rune Winther for making helpful comments and suggestions on the manuscript. Thanks for TEXnical support go to Jürgen Wiedmann.

We especially thank Bent Natvig, University of Oslo, for the great deal of time and effort he spent reading and preparing comments. Thanks also go to the three reviewers for providing advice on the content and organization of the book. Their informed criticism motivated several refinements and improvements. Of course, we take full responsibility for any errors that remain.

We also acknowledge the editing and production staff at Springer for their careful work. In particular, we appreciate the smooth cooperation of John Kimmel.

Stavanger, Norway — Terje Aven
Ulm, Germany — Uwe Jensen

Contents

1

Introduction

This chapter gives an introduction to the topics covered in this book: failure time models, complex systems, different information levels, maintenance and optimal replacement. We also include a section on reliability modeling, where we draw attention to some important factors to be considered in the modeling process. Two real life examples are presented: a reliability study of a system in a power plant and an availability analysis of a gas compression system.

1.1 Lifetime Models

In reliability we are mainly concerned with devices or systems that fail at an unforeseen or unpredictable (this term is defined precisely later) random age of $T > 0$. This random variable is assumed to have a distribution F, $F(t) = P(T \leq t), t \in \mathbb{R}$, with a density f. The hazard or failure rate λ is defined on the support of the distribution by

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)},$$

with the survival function $\bar{F}(t) = 1 - F(t)$. The failure rate $\lambda(t)$ measures the proneness to failure at time t in that $\lambda(t) \triangle t \approx P(T \leq t + \triangle t | T > t)$ for small $\triangle t$. The (cumulative) hazard function is denoted by Λ,

$$\Lambda(t) = \int_0^t \lambda(s)\, ds = -\ln\{\bar{F}(t)\}.$$

The well-known relation

$$\bar{F}(t) = P(T > t) = \exp\{-\Lambda(t)\} \tag{1.1}$$

establishes the link between the cumulative hazard and the survival function. Modeling in reliability theory is mainly concerned with additional information

T. Aven and U. Jensen, *Stochastic Models in Reliability*, Stochastic Modelling and Applied Probability 41, DOI 10.1007/978-1-4614-7894-2_1,

about the state of a system, which is gathered during the operating time of the system. This additional information leads to updated predictions about proneness to system failure. There are many ways to introduce such additional information into the model. In the following sections some examples of how to introduce additional information and how to model the lifetime T are given.

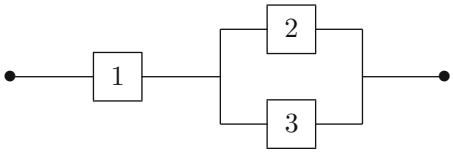
1.1.1 Complex Systems

As will be introduced in detail in Chap. 2, a complex system comprises n components with positive random lifetimes $T_i, i = 1, 2, \ldots, n, n \in \mathbb{N}$. Let $\Phi : \{0,1\}^n \to \{0,1\}$ be the structure function of the system, which is assumed to be monotone. The possible states of the components and of the system, "intact" and "failed," are indicated by "1" and "0," respectively. Then $\Phi_t = \Phi(\mathbf{X}_t)$ describes the state of the system at time t, where $\mathbf{X}_t = (X_t(1), \ldots, X_t(n))$ and $X_t(i)$ denotes the indicator function

$$X_t(i) = I(T_i > t) = \begin{cases} 1 & \text{if} \quad T_i > t \\ 0 & \text{if} \quad T_i \leq t, \end{cases}$$

which is 1, if component i is intact at time t, and 0 otherwise. The lifetime T of the system is then given by $T = \inf\{t \in \mathbb{R}_+ : \Phi_t = 0\}$.

Example 1.1. As a simple example the following system with three components is considered, which is intact if component 1 and at least one of the components 2 or 3 are intact:



In this example $\Phi_t = X_t(1)\{1 - (1 - X_t(2))(1 - X_t(3))\}$ is easily obtained with $T = \inf\{t \in \mathbb{R}_+ : \Phi_t = 0\} = T_1 \wedge (T_2 \vee T_3)$, where as usual $a \wedge b$ and $a \vee b$ denote $\min\{a, b\}$ and $\max\{a, b\}$, respectively. The additional information about the lifetime T is given by the observation of the state of the single components. As long as all components are intact, only a failure of component 1 leads to system failure. If one of the components 2 or 3 fails first, then the next component failure is a system failure.

Under the classical assumption that all components work independently, i.e., the random variables T_i, $i = 1, \ldots, n$, are independent, certain characteristics of the system lifetime are of interest:

- Determining the system lifetime distribution from the known component lifetime distributions or at least finding bounds for this distribution (see Sects. 2.1 and 2.2).

- Are certain properties of the component lifetime distributions like increasing failure rate (IFR) or increasing failure rate average (IFRA) preserved by forming monotone systems? One of these closure theorems states, for example, that the distribution of the system lifetime is IFRA if all component lifetimes have IFRA distributions (see Sect. 2.2).
- In what way does a certain component contribute to the functioning of the whole system? The answer to this question leads to the definition of several importance measures (see Sect. 2.1).

1.1.2 Damage Models

Additional information about the lifetime T can also be introduced into the model in a quite different way. If the state or damage of the system at time $t \in \mathbb{R}_+$ can be observed and this damage is described by a random variable X_t, then the lifetime of the system may be defined as

$$T = \inf\{t \in \mathbb{R}_+ : X_t \geq S\},$$

i.e., as the first time the damage hits a given level S. Here S can be a constant or, more general, a random variable independent of the damage process. Some examples of damage processes $X = (X_t)$ of this kind are described in the following subsections.

Wiener Process

The damage process is a Wiener process with positive drift starting at 0 and the failure threshold S is a positive constant. The lifetime of the system is then known to have an inverse Gaussian distribution. Models of this kind are especially of interest if one considers different environmental conditions under which the system is working, as, for example, in so-called burn-in models. An accelerated aging caused by additional stress or different environmental conditions can be described by a change of time. Let $\tau : \mathbb{R}_+ \to \mathbb{R}_+$ be an increasing function. Then $Z_t = X_{\tau(t)}$ denotes the actual observed damage. The time transformation τ drives the speed of the deterioration. One possible way to express different stress levels in time intervals $[t_i, t_{i+1}), 0 = t_0 < t_1 < \ldots < t_k, i = 0, 1, \ldots, k-1, k \in \mathbb{N}$, is the choice

$$\tau(t) = \sum_{j=0}^{i-1} \beta_j (t_{j+1} - t_j) + \beta_i (t - t_i),\ t \in [t_i, t_{i+1}), \beta_v > 0.$$

In this case it is seen that if F_0 is the inverse Gaussian distribution function of $T = \inf\{t \in \mathbb{R}_+ : X_t \geq S\}$, and F is the distribution function of the lifetime $T_a = \inf\{t \in \mathbb{R}_+ : Z_t \geq S\}$ under accelerated aging, then $F(t) = F_0(\tau(t))$. A generalization in another direction is to consider a random time change, which means that τ is a stochastic process. By this, randomly varying environmental conditions can be modeled.

Compound Point Processes

Processes of this kind describe so-called shock processes where the system is subject to shocks that occur from time to time and add a random amount to the damage. The successive times of occurrence of shocks, T_n, are given by an increasing sequence $0 < T_1 \leq T_2 \leq \ldots$ of random variables, where the inequality is strict unless $T_n = \infty$. Each time point T_n is associated with a real-valued random mark V_n, which describes the additional damage caused by the nth shock. The marked point process is denoted $(T, V) = (T_n, V_n), n \in \mathbb{N}$. From this marked point process the corresponding compound point process X with

$$X_t = \sum_{n=1}^{\infty} I(T_n \leq t) V_n \tag{1.2}$$

is derived, which describes the accumulated damage up to time t. The simplest example is a compound Poisson process in which the shock arrival process is Poisson and the shock amounts (V_n) are i.i.d. random variables. As before, the lifetime T is the first time the damage process (X_t) hits the level S. If we go one step further and assume that S is not deterministic and fixed, but a random failure level, then we can describe a situation in which the observed damage process does not carry complete information about the (failure) state of the system; the failure can occur at different damage levels S.

Another way to describe the failure mechanism is the following. Let the accumulated damage up to time t be given by the shock process X_t as in (1.2). If the system is up at $t-$ just before t, the accumulated damage equals $X_{t-} = x$ and a shock of magnitude y occurs at t, then the probability of failure at t is $p(x + y)$, where $p(x)$ is a given $[0, 1]$-valued function. In this model failures can only occur at shock times and the accumulated damage determines the failure probability.

1.1.3 Different Information Levels

It was pointed out above in what way additional information can lead to a reliability model. But it is also important to note that in one and the same model different observation levels are possible, i.e., the amount of actual available information about the state of a system may vary. The following examples will show the effect of different degrees of information.

1.1.4 Simpson's Paradox

This paradox says that if one compares the death rates in two countries, say A and B, then it is possible that the crude overall death rate in country A is higher than in B although all age-specific death rates in B are higher than in A. This can be transferred to reliability in the following way. Considering a two-component parallel system, the failure rate of the system lifetime may increase

although the component lifetimes have decreasing failure rates. The following proposition, which can be proved by some elementary calculations, yields an example of this.

Proposition 1.2. *Let $T = T_1 \vee T_2$ with i.i.d. random variables T_i, $i = 1, 2$, following the common distribution F,*

$$F(t) = 1 - e^{-u(t)},\ t \geq 0, u(t) = \gamma t + \alpha(1 - e^{-\beta t}),\ \alpha, \beta, \gamma > 0.$$

If $2\alpha e^{\alpha} < \left(\frac{\gamma}{\beta}\right)^2 < 1$, then the failure rate λ of the lifetime T increases, whereas the component lifetimes T_i have decreasing failure rates.

This example shows that it makes a great difference whether only the system lifetime can be observed (aging property: IFR) or additional information about the component lifetimes is available (aging property: DFR). The aging property of the system lifetime of a complex system does not only depend on the joint distribution of the component lifetimes but also, of course, on the structure function. Instead of a two-component parallel system, consider a series system where the component lifetimes have the same distributions as in Proposition 1.2. Then the failure rate of $T_{\text{ser}} = T_1 \wedge T_2$ decreases, whereas $T_{\text{par}} = T_1 \vee T_2$ has an IFR.

1.1.5 Predictable Lifetime

The Wiener process $X = (X_t), t \in \mathbb{R}_+$, with positive drift μ and variance scaling parameter σ, is a popular damage threshold model. The process X can be represented as $X_t = \sigma B_t + \mu t$, where B is standard Brownian motion. If one assumes that the failure level S is a fixed known constant, then the lifetime $T = \inf\{t \in \mathbb{R}_+ : X_t \geq S\}$ follows an inverse Gaussian distribution with a finite mean $ET = S/\mu$. One criticism of this model is that the paths of X are not monotone. As a partial answer, one can respond that maintenance actions also lead to improvements and thus X could be decreasing at some time points. A more severe criticism from the point of view of the available information is the following. It is often assumed that in this model the paths of the damage process can be observed continuously. But this would make the lifetime T a predictable random time (a precise definition follows in Chap. 3), i.e., there is an increasing sequence $\tau_n, n \in \mathbb{N}$, of random time points that announces the failure. In this model one could choose $\tau_n = \inf\{t \in \mathbb{R}_+ : X_t \geq S - 1/n\}$, and take n large enough and stop operating the system at τ_n "just" before failure, to carry out some preventive maintenance, cf. Fig. 1.1. This does not usually apply in practical situations. This example shows that one has to distinguish carefully between the different information levels for the model formulation (complete information) and for the actual observation (partial information).

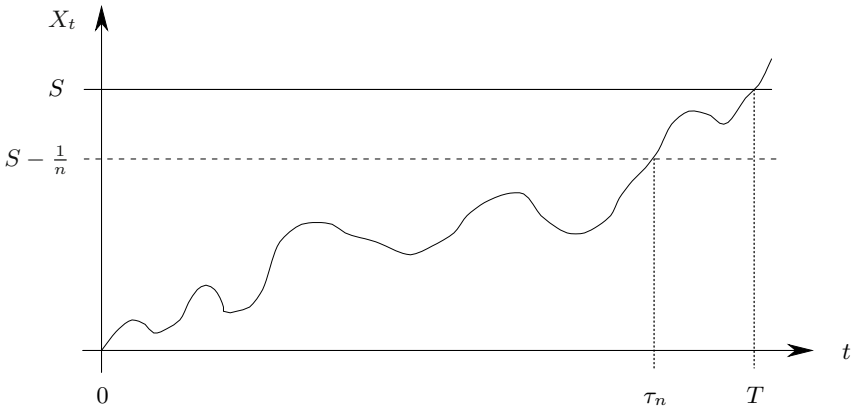


Fig. 1.1. Predictable stopping time

1.1.6 A General Failure Model

The general failure model considered in Chap. 3 uses elements of the theory of stochastic processes and particularly some martingale theory. Some of the readers might wonder whether sophisticated theory like this is necessary and suitable in reliability, a domain with engineering applications. Instead of a comprehensive justification we give a motivating example.

Example 1.3. We consider a simple two-component parallel system with independent $\mathrm{Exp}(\alpha_i)$ distributed component lifetimes $T_i, i = 1, 2$. The system lifetime $T = T_1 \vee T_2$ has distribution function

$$F(t) = P(T_1 \le t, T_2 \le t) = (1 - e^{-\alpha_1 t})(1 - e^{-\alpha_2 t})$$

with an ordinary failure rate

$$\lambda(t) = \frac{\alpha_1 e^{-\alpha_1 t} + \alpha_2 e^{-\alpha_2 t} - (\alpha_1 + \alpha_2) e^{-(\alpha_1 + \alpha_2)t}}{e^{-\alpha_1 t} + e^{-\alpha_2 t} - e^{-(\alpha_1 + \alpha_2)t}}.$$

This formula is rather complicated for such a simple system and reveals nothing about the structure of the system.Using elementary calculus it can be shown that for $\alpha_1 \neq \alpha_2$ the failure rate is increasing on $(0, t^*)$ and decreasing on (t^*, ∞) for some $t^* > 0$. This property of the failure rate, however, is neither obvious nor immediate to see. We also know that F is of IFRA type.

But is it not more natural and simpler to say that a failure rate (process) should be 0 as long as both components work (no system failure can occur) and, when the first component failure occurs, then the rate switches to α_1 or α_2 depending on which component survives? We want to derive a model that allows such a simple failure rate process and also includes the ordinary failure rate. Of course, this simple failure rate process, which can be expressed as

$$\lambda_t = \alpha_1 I(T_2 \le t < T_1) + \alpha_2 I(T_1 \le t < T_2),$$

needs knowledge about the random component lifetimes T_i. Now the failure rate λ_t is a stochastic process and the information about the status of the components at time t is represented by a filtration. The model allows for changing the information level and the ordinary failure rate can be derived from λ_t on the lowest level possible, namely no information about the component lifetimes.

The modern theory of stochastic processes allows for the development of a general failure model that incorporates the above aspects: time dynamics and different information levels. Chapter 3 presents this model. The failure rate process λ_t is one of the basic parameters of this set-up. If we consider the lifetime T, under some mild conditions we obtain the failure rate process on $\{T > t\}$ as the limit of conditional expectations with respect to the pre-t-history (σ-algebra) $\mathcal{F}_t$,

$$\lambda_t = \lim_{h \to 0+} \frac{1}{h} P(T \le t + h | \mathcal{F}_t),$$

extending the classical failure rate $\lambda(t)$ of the system. To apply the set-up, focus should be placed on the failure rate process (λ_t). When this process has been determined, the model has basically been established. Using the above interpretation of the failure rate process, it is in most cases rather straightforward to determine its form. The formal proofs are, however, often quite difficult.

If we go one step further and consider a model in which the system can be repaired or replaced at failure, then attention is paid to the number N_t of system failures in $[0, t]$. Given certain conditions, the counting process $N = (N_t), t \in \mathbb{R}_+$, has an "intensity" that as an extension of the failure rate process can be derived as the limit of conditional expectations

$$\lambda_t = \lim_{h \to 0+} \frac{1}{h} E[N_{t+h} - N_t | \mathcal{F}_t],$$

where $\mathcal{F}_t$ denotes the history of the system up to time t. Hence we can interpret λ_t as the (conditional) expected number of system failures per unit of time at time t given the available information at that time. Chapter 3 includes several special cases that demonstrate the broad spectrum of potential applications.

1.2 Maintenance

To prolong the lifetime, to increase the availability, and to reduce the probability of an unpredictable failure, various types of maintenance actions are being implemented. The most important maintenance actions include:

- Preventive replacements of parts of the system or of the whole system
- Repairs of failed units

- Providing spare parts
- Inspections to check the state of the system if not observed continuously

Taking maintenance actions into account leads, depending on the specific model, to one of the following subject areas: Availability Analysis and Optimization Models.

1.2.1 Availability Analysis

If the system or parts of it are repaired or replaced when failures occur, the problem is to characterize the performance of the system. Different measures of performance can be defined as, for example,

- The probability that the system is functioning at a certain point in time (point availability)
- The mean time to the first failure of the system
- The probability distribution of the downtime of the system in a given time interval.

Traditionally, focus has been placed on analyzing the point availability and its limit (the steady-state availability). For a single component, the steady-state formula is given by $MTTF/(MTTF + MTTR)$, where $MTTF$ and $MTTR$ represent the mean time to failure and the mean time to repair (mean repair time), respectively. The steady-state probability of a system comprising several components can then be calculated using the theory of complex (monotone) systems.

Often, performance measures related to a time interval are used. Such measures include the distribution of the number of system failures, and the distribution of the downtime of the system, or at least the mean of these distributions. Measures related to the number of system failures are important from an operational and safety point of view, whereas measures related to the downtime are more interesting from a productional point of view. Information about the probability of having a long downtime in a time interval is important for assessing the economic risk related to the operation of the system. For production systems, it is sometimes necessary to use a multistate representation of the system and some of its components, to reflect different production levels.

Compared to the steady-state availability, it is of course more complicated to compute the performance measures related to a time interval, in particular the probability distributions of the number of system failures and of the downtime. Using simplifications and approximations, it is however possible to establish formulas that can be used in practice. For highly available systems, a Poisson approximation for the number of system failures and a compound Poisson approximation for the downtime distribution are useful in many cases.

These topics are addressed in Chap. 4, which gives a detailed analysis of the availability of monotone systems. Emphasis is placed on performance

measures related to a time interval. Sufficient conditions are given for when the Poisson and the compound Poisson distributions are asymptotic limits.

1.2.2 Optimization Models

If a valuation structure is given, i.e., costs of replacements, repairs, downtime, etc., and gains, then one is naturally led to the problem of planning the maintenance action so as to minimize (maximize) the costs (gains) with respect to a given criterion. Examples of such criteria are expected costs per unit time and total expected discounted costs.

Example 1.4. We resume Example 1.3, p. 6, and consider the simple two-component parallel system with independent $\text{Exp}(\alpha_i)$ distributed component lifetimes $T_i, i = 1, 2$, with the system lifetime $T = T_1 \vee T_2$. We now allow preventive replacements at costs of c units to be carried out before failure, and a replacement upon system failure at cost $c + k$. It seems intuitive that $T_1 \wedge T_2$, the time of the first component failure, should be a candidate for an optimal replacement time with respect to some cost criterion, at least if c is "small" compared to k. How can we prove that this random time $T_1 \wedge T_2$ is optimal among all possible replacement times? How can we characterize the set of all possible replacement times?

These questions can only be answered in the framework of martingale theory and are addressed in Chap. 5.

One can imagine that thousands of models (and papers) can be created by combining the different types of lifetime models with different maintenance actions. The general optimization framework formulated in Chap. 5 incorporates a number of such models. Here the emphasis is placed on determining the optimal replacement time of a deteriorating system. The framework is based on the failure model of Chap. 3, which means that rather complex and very different situations can be studied. Special cases include monotone systems, (minimal) repair models, and damage processes, with different information levels.

1.3 Reliability Modeling

Models analyzed in this book are general, in the sense that they do not refer to any *specific* real life situation but are applicable in a number of cases. This is the academic and theoretical approach of mathematicians (probabilists, statisticians) who provide tools that can be used in applications.

The reliability engineer, on the other hand, has a somewhat different starting point. He or she is faced with a real problem and has to analyze this problem using a mathematical model that describes the situation appropriately.

Sometimes it is rather straightforward to identify a suitable model, but often the problem is complex and it is difficult to see how to solve it. In many cases, a model needs to be developed. The modeling process requires both experience on the part of the practitioner and knowledge on the part of the theorist.

However, it is not within the scope of this book to discuss in detail the many practical aspects related to reliability modeling and analysis. Only a few issues will be addressed. In this introductory section we will highlight important factors to be considered in the modeling process and two real life examples will be presented.

The objectives of the reliability study can affect modeling in many ways, for example, by specifying which performance measures and which factors (parameters) are to be analyzed. Different objectives will require different approaches and methods for modeling and analysis. Is the study to provide decision support in a design process of a system where the problem is to choose between alternative solutions; is the problem to give a basis for specifying reliability requirements; or is the aim to search for an optimal preventive maintenance strategy? Clearly, these situations call for different models.

The objectives of the study may also influence the choice of the computational approach. If it is possible to use analytical calculation methods, these would normally be preferred. For complex situations, Monte Carlo simulation often represents a useful alternative, cf., e.g., [13, 64].

The modeling process starts by clarifying the characteristics of the situation to be analyzed. Some of the key points to address are:

Can the system be decomposed into a set of independent subsystems (components)? Are all components operating normally or are some on stand-by? What is the state of the component after a repair? Is it "as good as new"? What are the resources available for carrying out the repairs? Are some types of preventive maintenance being employed? Is the state of the components and the system continuously monitored, or is it necessary to carry out inspections to reveal their condition? Is information available about the underlying condition of the system and components, such as wear, stress, and damage?

Having identified important features of the system, we then have to look more specifically at the various elements of the model and resolve questions like the following:

- How should the deterioration process of the components and system be modeled? Is it sufficient to use a standard lifetime model where the age of the unit is the only information available? How should the repair/replacement times be modeled?
- How are the preventive maintenance activities to be reflected in the model? Are these activities to be considered fixed in the model or is it possible to plan preventive maintenance action so that costs (rewards) are minimized (maximized)?
- Is a binary (two-state) approach for components and system sufficiently accurate, or is multistate modeling required?

- How are the system and components to be represented? Is a reliability block diagram appropriate?
- Are time dynamics to be included or is a time stationary model sufficient?
- How are the parameters of the model to be determined? What kind of input data are required for using the model? How is uncertainty to be dealt with?

Depending on the answers to these questions, relevant models can be identified. It is a truism that no model can cover all aspects, and it is recommended that one starts with a simple model describing the main features of the system.

The following application examples give further insight into the situations that can be modeled using the theory presented in this book.

1.3.1 Nuclear Power Station

In this example we consider a small part of a very complex technical system, in which safety aspects are of great importance. The nuclear power station under consideration consists of two identical boiling water reactors in commercial operation, each with an electrical power of 1,344 MW. They started in 1984 and 1985, respectively, working with an efficiency of 35%.

Nuclear power plants have to shut down from time to time to exchange the nuclear fuel. This is usually performed annually. During the shutdown phase a lot of maintenance tasks and surveillance tests are carried out. One problem during such phases is that decay heat is still produced and thus has to be removed. Therefore, residual heat removal (RHR) systems are in operation. At the particular site, three identical systems are available, each with a capacity of 100%. They are designed to remove decay heat during accident conditions occurring at full power as well as for operational purposes in cooldown phases.

One of these RHR systems is schematically shown in Fig. 1.2. It consists of three different trains including the closed cooling water system. Several pumps and valves are part of the RHR system. The primary cooling system can be modeled as a complex system comprising the following main components:

- Closed cooling water system pump (CCWS)
- Service water system pump (SWS)
- Low-pressure pump with a pre-stage (LP)
- High-pressure pump (HP)
- Nuclear heat exchanger (RHR)
- Valves (V_1, V_2, V_3)

For the analysis we have to distinguish between two cases:

1. The RHR system is not in operation.
   Then the functioning of the system can be viewed as a binary structure of the main components as is shown in the reliability block diagram in

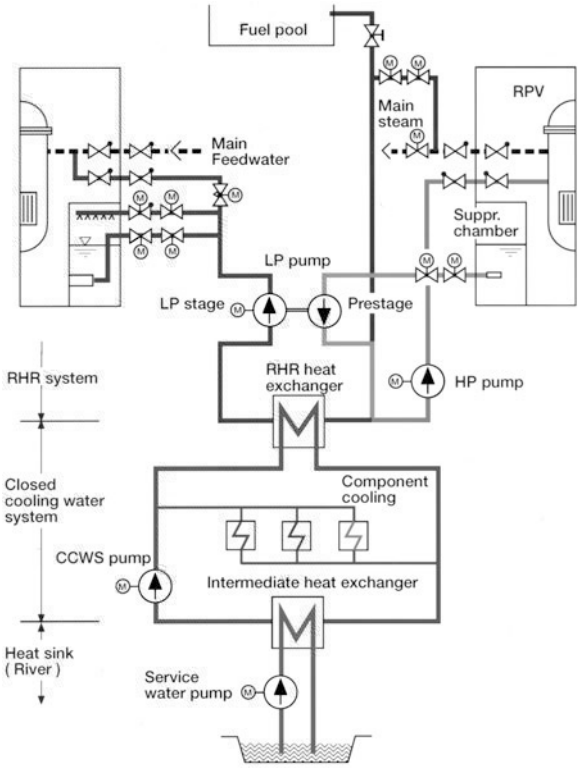


Fig. 1.2. Cooling system of a power plant

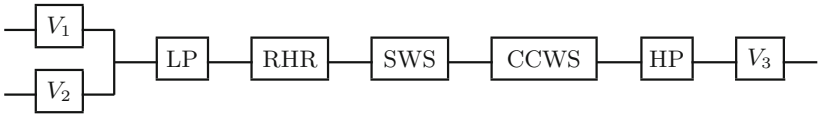


Fig. 1.3. Reliability block diagram

Fig. 1.3. When the system is needed, it is possible that single components or the whole system fails to start on demand. In this case, to calculate the probability of a failure on demand, we have to take all components in the reliability block diagram into consideration. Two of the valves, V_1 and V_2, are in parallel. Therefore, the RHR system fails on demand if either V_1 *and* V_2 fail or at least one of the remaining components LP,..., HP, V_3 fails. We assume that the time from a check of a component until a failure in the idle state is exponentially distributed. The failure rates are $\lambda_{v_1}, \lambda_{v_2}, \lambda_{v_3}$ for the valves and $\lambda_{p_1}, \lambda_{p_2}, \lambda_{p_3}, \lambda_{p_4}, \lambda_h$ for the other components. If the check (inspection or operating period) dates t time units back, then the probability of a failure on demand is given by

$$1 - \{1 - (1 - e^{-\lambda_{v_1} t})(1 - e^{-\lambda_{v_2} t})\} e^{-(\lambda_{p_1} + \lambda_{p_2} + \lambda_{p_3} + \lambda_{p_4} + \lambda_h + \lambda_{v_3})t}.$$

2. The RHR system is in operation.
   During an operation phase, only the pumps and the nuclear heat exchanger can fail to operate. If the valves have once opened on demand when the operation phase starts, these valves cannot fail during operation. Therefore, in this operation case, we can either ignore the valves in the block diagram or assign failure probability 0 to V_1, V_2, V_3. The structure reduces to a simple series system. If we assume that the failure-free operating times of the pumps and the heat exchanger are independent and have distributions $F_{p_1}, F_{p_2}, F_{p_3}, F_{p_4}$, and F_h, respectively, then the probability that the system fails before a fixed operating time t is just

$$1 - \bar{F}_{p_1}(t)\bar{F}_{p_2}(t)\bar{F}_{p_3}(t)\bar{F}_{p_4}(t)\bar{F}_h(t),$$

   where $\bar{F}(t)$ denotes the survival probability.

In both cases the failure time distributions and the failure rates have to be estimated. One essential condition for the derivation of the above formulae is that all components have stochastically independent failure times or lifetimes. In some cases such an independence condition does not apply. In Chap. 3 a general theory is developed that also includes the case of complex systems with dependent component lifetimes. The framework presented covers different information levels, which allow updating of reliability predictions using observations of the condition of the components of the system, for example.

1.3.2 Gas Compression System

This example outlines various aspects of the modeling process related to the design of a gas compression system.

A gas producer was designing a gas production system, and one of the most critical decisions was related to the design of the gas compression system.

At a certain stage of the development, two alternatives for the compression system were considered:

(i) One gas train with a maximum throughput capacity of 100%
(ii) Two trains in parallel, each with a maximum throughput capacity of 50%.

Normal production is 100%. For case (i) this means that the train is operating normally and a failure stops production completely. For case (ii) both trains are operating normally. If one train fails, production is reduced to 50%. If both trains are down, production is 0.

Each train comprises compressor–turbine, cooler, and scrubber. A failure of one of these "components" results in the shutdown of the train. Thus a train is represented by a series structure of the three components compressor–turbine, cooler, and scrubber.

The following failure and repair time data were assumed:

Component	Failure rate (unit of time: 1 year)	Mean repair time (unit of time: 1 h)
Compressor–turbine	10	12
Cooler	2	50
Scrubber	1	20

To compare the two alternatives, a number of performance measures were considered. Particular interest was shown in performance measures related to the number of system shutdowns, the time the system has a reduced production level, and the total production loss due to failures of the system. The gas sales agreement states that the gas demand is to be met with a very high reliability, and failures could lead to considerable penalties and loss of goodwill, as well as worse sales perspectives for the future.

Using models as will be described in Chap. 4, it was possible to compute these performance measures, given certain assumptions.

It was assumed that each component generates an alternating renewal process, which means that the repair brings the component to a condition that is as good as new. The uptimes were assumed to be distributed exponentially, so that the component in the operating state has a constant failure rate. The failure rate used was based on experience data for similar equipment. Such a component model was considered to be sufficiently accurate for the purpose of the analysis. The exponential model represents a "first-order approximation," which makes it rather easy to gain insight into the performance of the system. For a complex "component" with many parts to be maintained, it is known that the overall failure rate exhibits approximately exponential nature. Clearly, if all relevant information is utilized, the exponential model is rather crude. But again we have to draw attention to the purpose of the analysis: provide decision support concerning the choice of design alternatives. Only the essential features should be included in the model.

A similar type of reasoning applies to the problem of dependency between components. In this application all uptimes and downtimes of the components were assumed to be independent. In practice there are, of course, some dependencies present, but by looking into the failure causes and the way the components were defined, the assumption of independence was not considered to be a serious weakness of the model, undermining the results of the analysis.

To determine the repair time distribution, expert opinions were used. The repair times, which also include fault diagnosis, repair preparation, test and restart, were assessed for different failure modes. As for the uptimes, it was assumed that no major changes over time take place concerning component design, operational procedures, etc.

Uncertainty related to the input quantities used was not considered. Instead, sensitivity studies were performed with the purpose of identifying how sensitive the results were with respect to variations in input parameters.

Of the results obtained, we include the following examples:

- The gas train is down 2.7% of the time in the long run.
- For alternative (i), the average system failure rate, i.e., the average number of system failures per year, equals 13. For alternative (ii) it is distinguished between failures resulting in production below 100% and below 50%. The average system failure rates for these levels are approximately 26 and 0.7, respectively. Alternative (ii) has a probability of about 50% of having one or more complete shutdowns during a year.
- The mean lost production equals 2.7% for both alternatives. The probability that the lost production during 1 year is more than 4% of demand is approximately equal to 0.16 for alternative (i) and 0.08 for alternative (ii).

This last result is based on assumptions concerning the variation of the repair times. Refer to Sect. 4.7.1, p. 162, where the models and methods used to compute these measures are summarized.

The results obtained, together with an economic analysis, gave the management a good basis for choosing the best alternative.

Bibliographic Notes. There are now many journals strongly devoted to reliability, for example, the *IEEE Transactions on Reliability* and *Reliability Engineering and System Safety*. In addition, there are many journals in Probability and Operations Research that publish papers in this field.

As mentioned before, there is an extensive literature covering a variety of stochastic models of reliability. Instead of providing a long and, inevitably, almost certainly incomplete list of references, some of the surveys and review articles are quoted, as well as some of the reliability books.

From time to time, the *Naval Research Logistics Quarterly* journal publishes survey articles in this field, among them the renowned article by Pierskalla and Voelker [130], which appeared with 259 references in 1976, updated by Sherif and Smith [144] with an extensive bibliography of 524 references in 1981, followed by Valdez-Flores and Feldman [158] with 129 references in 1989. Bergman's review [39] reflects the author's experience in industry and emphasizes the usefulness of reliability methods in applications. Gertsbakh's paper [75] reviews asymptotic methods in reliability and especially investigates under what conditions the lifetime of a complex system with many components is approximately exponentially distributed. Natvig [125] gives a concise overview of importance measures for monotone systems. The surveys of Arjas [4] and Koch [108] consider reliability models using more advanced mathematical tools as marked point processes and martingales. A guided tour for the non-expert through point process and intensity-based models in reliability is presented in the article of Hokstad [89]. The book of Thompson [155] gives a