

# PROBABILISTIC RELIABILITY ENGINEERING

# PROBABILISTIC RELIABILITY ENGINEERING

**BORIS GNEDENKO**

Moscow State University and SOTAS, Inc.

**IGOR USHAKOV**

SOTAS, Inc. and George Washington University

Edited by **JAMES FALK**

George Washington University



A Wiley-Interscience Publication

**JOHN WILEY & SONS, INC.**

New York / Chichester / Brisbane / Toronto / Singapore

**This text is printed on acid-free paper.**

**Copyright © 1995 by John Wiley & Sons, Inc.**

**All rights reserved. Published simultaneously in Canada.**

**Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.**

**This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering professional services. If legal accounting, medical, psychological, or any other expert assistance is required, the services of a competent professional person should be sought.**

ISBN-0-471-30502-2

1 0 9 8 7 6 5 4 3 2 1



# CONTENTS

<b>Preface</b>	<b>xv</b>
<b>Introduction</b>	<b>xvii</b>
<b>1 Fundamentals</b>	<b>1</b>
1.1 Discrete Distributions Related to Reliability / 1	
1.1.1 Bernoulli Distribution / 1	
1.1.2 Geometric Distribution / 2	
1.1.3 Binomial Distribution / 5	
1.1.4 Negative Binomial Distribution / 6	
1.1.5 Poisson Distribution / 8	
1.2 Continuous Distributions Related to Reliability / 10	
1.2.1 Exponential Distribution / 10	
1.2.2 Erlang Distribution / 14	
1.2.3 Normal Distribution / 15	
1.2.4 Truncated Normal Distribution / 19	
1.2.5 Weibull-Gnedenko Distribution / 19	
1.2.6 Lognormal Distribution / 22	
1.2.7 Uniform Distribution / 24	
1.3 Summation of Random Variables / 25	
1.3.1 Sum of a Fixed Number of Random Variables / 25	



## Xii CONTENTS

1.3.2	Central Limit Theorem / 28
1.3.3	Poisson Theorem / 30
1.3.4	Random Number of Terms in the Sum / 31
1.3.5	Asymptotic Distribution of the Sum of a Random Number of Random Variables / 32
1.4	Relationships Among Distributions / 34
1.4.1	Some Relationships Between Binomial and Normal Distributions / 34
1.4.2	Some Relationships Between Poisson and Binomial Distributions / 35
1.4.3	Some Relationships Between Erlang and Normal Distributions / 36
1.4.4	Some Relationships Between Erlang and Poisson Distributions / 36
1.4.5	Some Relationships Between Poisson and Normal Distributions / 37
1.4.6	Some Relationships Between Geometric and Exponential Distributions / 38
1.4.7	Some Relationships Between Negative Binomial and Binomial Distributions / 38
1.4.8	Some Relationships Between Negative Binomial and Erlang Distributions / 39
i 4.9	Approximation with the Gram-Charlie Distribution / 39
1.5	Stochastic Processes / 42
1.5.1	Poisson Process / 42
1.5.2	Introduction to Recurrent Point Processes / 47
1.5.3	Thinning of a Point Process / 53
1.5.4	Superposition of Point Processes / 54
1.6	Birth and Death Process / 56
1.6.1	Model Description / 56
1.6.2	<b>Stationary Probabilities / 58</b>
1.6.3	Stationary Mean Time of Being in a Subset / 60
1.6.4	Probability of Being in a Given Subset / 61
1.6.5	Mean Time of Staying in a Given Subset / 64
1.6.6	Stationary Probability of Being in a Given Subset / 65
1.6.7	Death Process / 65
	Conclusion / 68
	References / 69

## Appendix: Auxiliary Tools / 70

1.A.1 Generating Functions / 70

1.A.2 Laplace-Stieltjes Transformation / 71

1.A.3 Generalized Generating Sequences / 74

Exercises / 79

Solutions / 81

**2 Reliability Indexes****86**

## 2.1 Unrepairable Systems / 87

2.1.1 Mean Time to Failure / 87

2.1.2 Probability of a Failure-Free Operation / 88

2.1.3 Failure Rate / 90

## 2.2 Repairable System / 92

2.2.1 Description of the Process / 92

2.2.2 Availability Coefficient / 93

2.2.3 Coefficient of Interval Availability / 94

2.2.4 Mean Time Between Failures and Related  
Indexes / 95

## 2.3 Special Indexes / 98

2.3.1 Extra Time Resource for Performance / 98

2.3.2 Collecting Total Failure-Free Time / 99

2.3.3 Acceptable Idle Intervals / 100

## 2.4 Choice of Indexes and Their Quantitative Norm / 100

2.4.1 Choice of Indexes / 100

2.4.2 Required Reliability Level / 102

Conclusion / 107

References / 107

Exercises / 107

Solutions / 108

**3 Unrepairable Systems****110**

## 3.1 Structure Function / 110

## 3.2 Series Systems / 111

## 3.3 Parallel Structure / 116

3.3.1 Simple Redundant Group / 116

3.3.2 " $k$  out of  $n$ " Structure / 121

## 3.4 Mixed Structures / 123





## viii CONTENTS

- 3.5 Standby Redundancy / 128
  - 3.5.1 Simple Redundant Group / 128
  - 3.5.2 " $k$  out of Redundancy / 131
  - 3.5.3 On-Duty Redundancy / 133
- 3.6 Switching and Monitoring / 136
  - 3.6.1 Unreliability Common Switching Device / 136
  - 3.6.2 Common Switching Device with Unreliable Switching / 139
  - 3.6.3 Individual Switching Devices / 140
  - 3.6.4 Periodic Monitoring / 143
- 3.7 Dynamic Redundancy / 144
  - 3.7.1 Independent Stages / 145
  - 3.7.2 Possibility of Transferring Units / 146
- 3.8 Systems with Dependent Units / 147
  - 3.8.1 Series Systems / 148
  - 3.8.2 Parallel Systems / 149
  - 3.8.3 Mixed Structures / 151
- 3.9 Two Types of Failures / 153
- 3.10 Mixed Structures with Physical Parameters / 156
  - Conclusion / 162
  - References / 163
  - Exercises / 163
  - Solutions / 164

## 4 Load - Strength Reliability Models

167

- 4.1 Static Reliability Problems of "Load-Strength" Type / 167
  - 4.1.1 General Expressions / 167
  - 4.1.2 Several Particular Cases / 168
  - 4.1.3 Numerical Method / 174
- 4.2 Models of Cycle Loading / 178
  - 4.2.1 Fixed Level of Strength / 179
  - 4.2.2 Deteriorating Strength / 180
- 4.3 Dynamic Models of "Strength-Load" Type / 181
  - 4.3.1 General Case / 181
  - 4.3.2 Gaussian Stochastic Process / 185
  - 4.3.3 Poisson Approximation / 186
  - Conclusion / 189
  - References / 190



Exercices / 191

Solutions / 191

## **5 Distributions with Monotone Intensity Functions** **193**

- 5.1 Description of the Monotonicity Property of the Failure Rate / 193
- 5.2 Unit with an IFR Distribution of TTF / 197
- 5.3 System of IFR Units / 206
  - 5.3.1 Series System / 207
  - 5.3.2 Parallel Systems / 211
  - 5.3.3 Other Monotone Structures / 213
- Conclusion / 213
- References / 214
- Exercices / 215
- Solutions / 215

## **6 Repairable Systems** **218**

- 6.1 Single Unit / 218
  - 6.1.1 Markov Model / 218
  - 6.1.2 General Distributions / 225
- 6.2 Repairable Series System / 228
  - 6.2.1 Markov Systems / 228
  - 6.2.2 General Distribution of Repair Time / 233
  - 6.2.3 General Distributions of TTF and Repair Time / 233
- 6.3 Repairable Redundant Systems of Identical Units / 235
  - 6.3.1 General Markov Model / 235
- 6.4 General Markov Model of Repairable Systems / 238
  - 6.4.1 Description of the Transition Graph / 238
  - 6.4.2 Nonstationary Coefficient of Availability / 239
  - 6.4.3 Probability of Failure-Free Operation / 244
  - 6.4.4 Determination of the MTTF and MTBF / 245
- 6.5 Time Redundancy / 248
  - 6.5.1 System with Instant Failures / 248
  - 6.5.2 System with Noninstant Failures / 249
  - 6.5.3 System with a Time Accumulation / 250
  - 6.5.4 System with Admissible Down Time / 251
- Conclusion / 252
- References / 253

Exercises / 254

Solutions / 254

**7 Repairable Duplicated System 256**

- 7.1 Markov Model / 256
  - 7.1.1 Description of the Model / 257
  - 7.1.2 Nonstationary Availability Coefficient / 258
  - 7.1.3 Stationary Availability Coefficient / 260
  - 7.1.4 Probability of Failure-Free Operation / 262
  - 7.1.5 Stationary Coefficient of Interval Availability / 264
  - 7.1.6 MTTF and MTBF / 265
- 7.2 Duplication with an Arbitrary Repair Time / 267
- 7.3 Standby Redundancy with Arbitrary Distributions / 275
- 7.4 Method of Introducing Fictitious States / 278
- 7.5 Duplication with Switch and Monitoring / 286
  - 7.5.1 Periodic Partial Control of the Main Unit / 286
  - 7.5.2 Periodic Partial Monitoring of Both Units / 288
  - 7.5.3 Unreliable Switch / 290
  - 7.5.4 Unreliable Switch and Monitoring of Main Unit / 292
- Conclusion / 293
- References / 294
- Exercises / 294
- Solutions / 296nn

**8 Analysis of Performance Effectiveness 298**

- 8.1 Classification of Systems / 298
  - 8.1.1 General Explanation of Effectiveness Concepts / 298
  - 8.1.2 Classes of Systems / 300
- 8.2 Instant Systems / 301
- 8.3 Enduring Systems / 308
- 8.4 Particular Cases / 312
  - 8.4.1 Additive Type of a System Unit's Outcome / 312
  - 8.4.2 Systems with a Symmetrical Branching Structure / 316
  - 8.4.3 System with Redundant Executive Units / 322
- 8.5 Systems with Intersecting Zones of Action / 323
  - 8.5.1 General Description / 323
  - 8.5.2 Additive Coefficient of Effectiveness / 325

## Xii CONTENTS

- 8.5.3 Multiplicative Coefficient of Effectiveness / 326
- 8.5.4 Redundant Coefficient of Effectiveness / 327
- 8.5.5 Boolean Coefficient of Effectiveness / 328
- 8.5.6 Preferable Maximal Coefficient  
of Effectiveness / 328
- 8.5.7 Preferable Minimal Coefficient  
of Effectiveness / 329
- 8.6 Aspects of Complex Systems Decomposition / 329
  - 8.6.1 Simplest Cases of Decomposition / 330
  - 8.6.2 Bounds for Regional Systems / 331
  - 8.6.3 Hierarchical Decomposition and Bounds / 332
- 8.7 Practical Recommendation / 335
  - Conclusion / 337
  - References / 337
  - Exercises / 338
  - Solutions / 338

## Two-Pole Networks

- 9.1 Rigid Computational Methods / 341
  - 9.1.1 Method of Direct Enumeration / 341
  - 9.1.2 Method of Boolean Function Decomposition / 343
- 9.2 Method of Paths and Cuts / 345
  - 9.2.1 Esary-Proschan Bounds / 345
  - 9.2.2 Litvak-Ushakov Bounds / 348
  - 9.2.3 Comparison of the Two Methods / 353
  - 9.2.4 Method of Set Truncation / 354
  - 9.2.5 Generalization of Cut-and-Path Bounds / 356
- 9.3 Methods of Decomposition / 359
  - 9.3.1 Moore-Shannon Method / 359
  - 9.3.2 Bodin Method / 362
  - 9.3.3 Ushakov Method / 364
- 9.4 Monte Carlo Simulation / 370
  - 9.4.1 Modeling with Fixed Failed States / 370
  - 9.4.2 Simulation Until System Failure / 372
  - Conclusion / 373

**Xii CONTENTS**

References	/	374
Exercises	/	375
Solutions	/	376

<b>10 Optimal Redundancy</b>	<b>378</b>
10.1 Formulation of the Problem / 378	
10.1.1 Simplest Problems / 379	
10.1.2 Several Restrictions / 380	
10.1.3 Practical Problems / 381	
10.2 Optimal Redundancy with One Restriction / 381	
10.2.1 Lagrange Multiplier Method / 381	
10.2.2 Steepest Descent Method / 384	
10.2.3 Approximate Solution / 388	
10.2.4 Dynamic Programming Method / 390	
10.2.5 Kettelle's Algorithm / 393	
10.2.6 Method of Generalized Generating Sequences / 399	
10.3 Several Limiting Factors / 402	
10.3.1 Method of Weighing / 402	
10.3.2 Method of Generating Sequences / 407	
10.4 Multicriteria Optimization / 409	
10.4.1 Steepest Descent Method / 409	
10.4.2 Method of Generalized Generating Function / 410	
10.5 Comments on Calculation Methods / 41)	
Conclusion / 415	
References / 417	
Exercises / 418	
<b>11 Optimal Technical Diagnosis</b>	<b>420</b>
11.1 General Description / 420	
11.2 One Failed Unit / 421	
11.2.1 Dynamic Programming Method / 423	
11.2.2 Perturbation Method for One-Unit Tests / 424	
11.2.3 Recursive Method / 426	
11.3 Sequential Searches for Multiple Failed Units / 428	
11.3.1 Description of the Procedure / 428	
11.3.2 Perturbation Method / 428	
11.3.3 Recursive Method / 431	
11.4 System Failure Determination / 432	
Conclusion / 434	
References / 435	



<b>12 Additional Optimization Problems in Reliability Theory</b>	<b>436</b>
12.) Optimal Preventive Maintenance / 436	
12.2 Optimal Periodic System Checking / 438	
12.3 Optimal Discipline of Repair / 439	
12.4 Dynamic Redundancy / 445	
12.5 Time Sharing of Redundant Units / 450	
Conclusion / 453	
References / 454	
<b>13 Heuristic Methods in Reliability</b>	<b>455</b>
13.1 Approximate Analysis of Highly Reliable Repairable Systems / 456	
13.1.1 Series System / 457	
13.1.2 Unit with Periodic Inspection / 458	
13.1.3 Parallel System / 459	
13.1.4 Redundant System with Spare Unit / 463	
13.1.5 Switching Device / 464	
13.2 Time Redundancy / 466	
13.2.1 Gas Pipeline with Underground Storage / 466	
13.2.2 Oil Pipeline with Intermediate Reservoirs / 468	
13.3 Bounds for the MTTF of a Two-Pole Network / 470	
13.4 Dynamic Redundancy / 472	
13.5 Object with Repairable Models and an Unrenewable Component Stock / 473	
13.5.1 Description of the Maintenance Support System / 473	
13.5.2 Notation / 475	
13.5.3 Probability of Successful Service / 476	
13.5.4 Availability Coefficient / 478	
13.6 Territorially Dispersed System of Objects with Individual Module Stock, Group Repair Shops, and Hierarchical Warehouse of Spare Units / 481	
13.6.1 Description of the MSS / 481	
13.6.2 Object with an Individual Repair Shop and a Hierarchical Stock of Spare Units / 483	
13.6.3 Set of $K$ Objects with Repair Shops and a Central Stock with Spare Units / 490	
13.7 Centralized Inventory System / 496	

## Xii CONTENTS

- 13.8 Heuristic Solution of Optimal Redundancy Problem with Multiple Restrictions / 499
  - 13.8.1 Sequence of One-Dimensional Problems / 499
  - 13.8.2 Using the Most Restrictive Current Resource / 501
  - 13.8.3 Method of "Reflecting Screen" / 502
- 13.9 Multifunction System / 502
- 13.10 Maximization of Mean Time to Failure / 505
  - Conclusion / 510
  - References / 511
  - General References / 513

### Index 515

The following Abbreviations are used frequently throughout this book:

#### Abbreviation Meaning

	birth and death process
BDP	distribution function
d.f.	generating function
g.f.	independent and identically distributed
i.i.d.	Laplace-Stieltjes transform
LST	moment generating function
m.g.f.	mean repair time
f.	mean time between failures
MR	mean time to failure
T	probability of failure-free operation
MTBT	pseudo-random variable
MT	random variable
TF	time to failure
PPF	
O	
p.r.v	
.	
r.v.	
TTF	

# PREFACE

This book was initially undertaken in 1987 in Moscow. We have found that the majority of books on mathematical models of reliability are very specialized: essentially none of them contains a spectrum of reliability problems. At the same time, many of them are overloaded with mathematics which may be beautiful but not always understandable by engineers. We felt that there should be a book covering as much as possible a spectrum of reliability problems which are understandable to engineers. We understood that this task was not a simple one. Of course, we now see that this book has not completely satisfied our initial plan, and we have decided to make it open for additions and a widening by everybody who is interested in it.

The reader must not be surprised that we have not touched on statistical topics. We did this intentionally because we are now preparing a book on statistical reliability engineering.

The publishing of this book became possible, in particular, because of the opportunities given by B. Gnedenko to visit the United States twice: in 1991 by George Washington University (Washington, DC) and in 1993 by SOT AS, Inc. (Rockville, Maryland). We both express our gratitude to Professor James E. Falk (GWU), Dr. Peter L. Willson (SOTAS), and Dr. William C. Hardy (MCI) for sponsoring these two visits of B. Gnedenko which permitted us to discuss the manuscript and to make the final decisions.

We would also like to thank Tatyana Ushakov who took care of all of the main problems in the final preparation of the manuscript, especially in dealing with the large number of figures.



**XVI PREFACE**

We are waiting for the readers' comments and corrections. We also repeat our invitation to join us in improving the book for the future editions.

*Professor of the Moscow State University  
and Consultant to SOTAS, Inc-  
Chief Scientist, SOTAS, Inc.  
and Visiting Researcher at the George Washington University*

**Moscow, Russia  
Rockville, Maryland  
December 1993**

BORIS GNEDENKO

**Igor** UsHAKOV



# INTRODUCTION

The term *reliability*, in the modern understanding by specialists in engineering, system design, and applied mathematics, is an acquisition of the 20th century. It appeared because various technical equipment and systems began to perform not only important industrial functions but also served for the security of people and their wealth.

Initially, reliability theory was developed to meet the needs of the electronics industry. This was a consequence of the fact that the first complex systems appeared in this field of engineering. Such systems have a huge number of components which made their reliability very low in spite of their relatively highly reliable components. This led to the development of a specialized applied mathematical discipline which allowed one to make an a priori evaluation of various reliability indexes at the design stage, to choose an optimal system structure, to improve methods of maintenance, and to estimate the reliability on the basis of special testing or exploitation.

Reliability is a rich field of research for technologists, engineers, systems analysts, and applied mathematicians. Each of them plays a key role in ensuring reliability. The creation of reliable components is a very complex chemical-physical problem of technology. The construction of reliable equipment is also a very complex engineering problem. System design is yet another very complex problem of system engineering and systems analysis. We could compare this process to the design of a city: someone produces reliable constructions, another design and builds buildings, and a third plans the location of houses, enterprises, services, and so on. We consider mainly reliability theory for solving problems of system design. We understand all of the limitations of such a viewpoint.





For the deficiency in this book, we could recommend some books which are dedicated to reliability in terms of equipment and components. References can be found in the list of general publications at the end of this book. We understand that the problem of engineering support of reliability is very serious and extremely difficult. Most of this requires a concrete physical analysis and sometimes relates very closely to each specific type of equipment and component.

We are strongly convinced that the main problem in applied reliability analysis is to invent and construct an adequate mathematical model. Modeling is always an art and an invention. The mathematical technique is not the main issue. Mathematics is a tool for solution of the task.

Most modern mathematical models in reliability require a computer. Usually, reports prepared with the help of a computer hypnotize: accurate format, accurate calculations... But the quality of the solution depends only on the quality of the model and input data. The computer is only a tool, not a panacea. A computer can never replace an analyst. The term "GIGO," which reminds one of FIFO and LIFO in queuing theory, was not conceived in vain. It means: garbage in, garbage out.

A mathematical model, first of all, must reflect the main features of a real object. But, at the same time, a model must be clear and understandable. It must be solvable with the help of available mathematical tools (including computer programs). It must be easily modified if a researcher can find some new features of the real object or would like to change the form of representation of the input data.

Sometimes mathematical models serve a simple purpose: to make a designed system more understandable for a designer. This use of modeling is very important (even if there are no practical recommendations and no numerical results) because this is the first stage of a system's testing, namely, a "mental testing." According to legend Napoleon, upon being asked why he could make fast and accurate decisions, answered that it is very simple: spend the night before the battle analyzing all conceivable turns of the battle—and you will gain a victory. The design of a mathematical model requires the same type of analysis: you rethink the possible uses of a system, its operational modes, its structure, and the specific role of different system's parts.

The reader will not find many references to American authors in this book. We agree that this is not good. To compensate for this deficiency, we list the main English language publications on the subject at the end of this book. We also supply a restricted list of publications in Russian which are close to the subject of this book.

As a matter of fact, we based our book on Russian publications. We also used our own practical experience in design and consulting. The authors represent a team of an engineer and a professional mathematician who have worked together for over 30 years, one as a systems analyst at industrial research and development institutes and the other as a consultant to the same institutes. We were both consultants to the State Committee of Stan-

INTRODUCTION xi  
dards of the former Soviet Union. For over 25 years we have been running the Moscow Consulting Center of Reliability and Quality Control which serves industrial engineers all over the country.

We had a chance to obtain knowledge of new ideas and new methods from a tide of contemporary papers. We have been in charge of the journal *Reliability and Quality Control* for over 25 years, and for more than 20 years we have been responsible for this section on reliability and queuing theory in the journal *Tekhnicheskaya Kibernetika* (published in the United States as *Engineering Cybernetics* and later as the *Soviet Journal of Computer and Systems Sciences*).

This activity in industry and publishing was fruitful for us. Together we wrote several papers including review on the state of reliability theory in Russia.

We hope that the interested reader meets with terra incognita—Russian publications in the field, Russian names, and, possibly, new viewpoints, ideas, problems, and solutions. For those who are interested in a more in-depth penetration into the state of Russian results in reliability theory, we can suggest several comprehensive reviews of Russian works in the field: Levin and Ushakov (1965), Gnedenko, Kozlov, and Ushakov (1969), Belyaev, Gnedenko, and Ushakov (1983), and Rukhin and Hsieh (1987).

We tried to cover almost the entire area of applied mathematical models in the theory of reliability. Of course, we might only hope that the task is fulfilled more or less completely. There are many special aspects of the mathematical theory of reliability which appear outside the scope of this book. We suggest that our readers and colleagues join us in the future: the book is open to contributions from possible authors. We hope that the next edition of the book will contain new contributors. Please send us your suggestions and/or manuscripts of proposed new sections and chapters to the address of John Wiley & Sons.

BORIS GNEDENKO  
IGOR USHAKOV

## REFERENCES

- Belyaev, Yu. K., B. V. Gnedenko, and I. A. Ushakov (1983). **Mathematical problems in queuing and reliability theory.** *Engrg. Cybernet. (USA)*, vol. 22, no. 6.
- Gnedenko, B. V., B. A. Kozlov, and I. A. Ushakov (1969). **The role of reliability theory in the construction of complex systems (in Russian).** In *Reliability Theory and Queuing Theory*, B. Gnedenko, ed. Moscow: Sovietskoe Radio.
- Levin, B. R., and I. A. Ushakov (1965). **Some aspects of the present state of reliability (in Russian).** *Radiotekhnika*, No. 4.
- Rukhin, A. L., and H. K. Hsieh (1987). **Survey of Soviet work in Reliability.** *Statist. Sci.*, vol. 2, no. 4.

# CHAPTER 1

## FUNDAMENTALS

We decided to begin with a brief discussion of the more or less standard subject of probability theory and the theory of stochastic processes. Of course, we are trying to review all this from a reliability standpoint. We not only give a formal description of the main discrete and continuous distribution functions usually used in reliability analysis, but explain as well the nature of their appearance and their mutual interrelationships.

A presentation of stochastic processes does not pretend to cover this branch of probability theory. It is rather a recollection of some necessary background for the reader.

With the same purpose we decided to include an appendix to the chapter with a very short overview of the area of generating functions and Laplace-Stieltjes transforms.

### 1.1 DISCRETE DISTRIBUTIONS RELATED TO RELIABILITY

#### 1.1.1 Bernoulli Distribution

In applications, one often deals with a very simple case where only two outcomes are possible—success or failure. For example, in analyzing the production quality of some production line, one may choose a criterion (an acceptable level or tolerance limit) to divide the entire sample into two parts: "good" and "bad."

Consider another example: during equipment testing one may predetermine some specified time and check if the random time-to-failure of the chosen item exceeds it or not. Thus, each event might be related to success or failure by this criterion. ■

}





## 2 FUNDAMENTALS

We will denote a successful outcome as 1, and a failure as 0. This leads us to consider a random variable (r.v.)  $X$  for which  $\Pr\{X = 1\} = p$  and  $\Pr\{X = 0\} = 1 - p = q$ . The value of  $p$  is called the parameter of the Bernoulli distribution. The distribution function (d.f.) of the r.v.  $X$  can be written in the form

$$f_B(x|p) = \begin{cases} p & x = 1 \\ q & x = 0 \end{cases} \quad (1.1)$$

where the subscript  $B$  signals the Bernoulli distribution. Clearly,  $f_B(1|p) = p$  and  $f_B(0|p) = 1 - p = q$ . For the Bernoulli r.v. we know

$$E\{X\} = 1 \cdot p + 0 \cdot q = p \quad (1.2)$$

and

$$E\{X^2\} = 1^2 p + 0^2 q = p$$

The variance is expressed through the first and second moments:

$$\text{Var}\{X\} = E\{X^2\} - [E\{X\}]^2 = p - p^2 = p(1 - p) = pq \quad (1.3)$$

The moment generating function (m.g.f.) of the r.v.  $X$  can be written as

$$M(s) = E\{e^{sX}\} = pe^s + q \quad (1.4)$$

The m.g.f. can also be used to obtain the moments of the distribution:

$$M^{(j)}(0) = E\{X^j\} = \frac{d^j}{ds^j} (pe^s + q) \Big|_{s=0} = p$$

which coincide with (1.2) and (1.3).

A sequence of independent identically distributed (i.i.d.) Bernoulli r.v.'s is called a *sequence of Bernoulli trials with the parameter  $p$* . For example, one may sequentially test  $n$  statistically identical items by setting  $X_i = 1$  if the  $i$ th item operates successfully during the time period  $t$ , and  $X_i = 0$  otherwise ( $i = 1, \dots, n$ ). Thus, one has a random sequence of 1's and 0's which reflects the Bernoulli trial outcomes.

### 1.1.2 Geometric Distribution

Consider a unit installed in a socket. The unit is periodically replaced by a new one after time  $t$ . Thus, the socket's operation is represented by a sequence of cycles, each of which consists of the use of a new unit. Let  $X$  denote the trial's outcome:  $X = 1$  if a unit has not failed during the time

interval  $t$ , and  $X = 0$  otherwise. The probability of a successful operation during one cycle equals  $p$ . All units are identical and stochastically independent. The socket operates successfully for a random number of cycles  $X$  before a first failure. The distribution of the r.v.  $X$  is the subject of interest. This distribution of the length of a series of successes for the sequence of Bernoulli trials is called a geometrical distribution:

$$\Pr\{X = x\} = f_g(x|p) = p^x q \quad (1.5)$$

where the subscript  $g$  denotes the geometrical distribution. For (1.5) the d.f. is

$$\Pr\{X \leq x\} = \sum_{k=0}^x p^k q \quad (1.6)$$

Since (1.6) includes the geometric series, it explains the origin of the distribution's name.

Everybody knows how to calculate (1.6) in a standard way, but we would like to show an interesting way which can be useful in other situations. Let

$$z = 1 + p + p^2 + \dots \quad (1.7)$$

and

$$y = 1 + p + p^2 + \dots + p^x$$

Then (1.7) can be rewritten as

$$z = y + p^{x+1} + p(l + p + p^2 + \dots)$$

$$+ p^x = 1 + py$$

and, finally, if the sum converges

$$\sum_{k=0}^{\infty} p^k = \frac{1 - p^{l+1}}{1 - p} = \frac{1}{1 - p} = \frac{1}{q}$$

Now returning to (1.6), we obtain

$$\Pr\{X \leq x\} = 1 - p^{x+1} \quad (1.8)$$

Thus, with the probability defined in (1.8), a failure has occurred before the  $x$ th cycle. The probability of a series of successes of length not less than  $x$ , that is,  $\Pr\{X > x\}$  is, obviously,

$$\Pr\{X > x\} = 1 - \Pr\{X \leq x\} = p^x \quad (1.9)$$

#### 4 FUNDAMENTALS

Of course, the last result can be obtained directly. The set of all events, consisting of series of not less than  $x$  successes, is equivalent to  $x$  first successes and any other outcome afterwards.

For the geometric distribution, the m.g.f.,  $g_p$ , can be written as

$$g_p(s) = \sum_{j=0}^{\infty} p^j e^{js} \quad (1.10)$$

This sum has a limit if  $0 < pe^s < 1$ . To compute (1.10), we can use the same procedure as above. With the same notation, we obtain

$$y = 1 + a + a^2 + a^3 + \dots = 1 + a(1 + a + a^2 + \dots) = 1 + ay$$

and then

$$y = \frac{1}{1 - a} = \frac{1}{1 - pe^s}$$

Thus,

$$(1.11)$$

$$g_p(s) = \frac{1}{1 - pe^s}$$

The mean and variance of the geometric distribution can be found in a direct way with the use of bulky transformations. We will derive them using (1.11):

and

$$\frac{d}{ds} \left( \frac{1}{1 - pe^s} \right) = \frac{pe^s}{(1 - pe^s)^2} \quad (1.12)$$

$$\frac{d^2}{ds^2} \left( \frac{1}{1 - pe^s} \right) = \frac{pe^s(1 + pe^s)}{(1 - pe^s)^3} \quad (1.13)$$

Thus, the variance by (1.11) is

$$\text{Var}(A) = \frac{pe^s(1 + pe^s)}{(1 - pe^s)^3} - \left( \frac{pe^s}{(1 - pe^s)^2} \right)^2 \quad (1.14)$$

Substituting  $e^s$  for  $z$ , we obtain the generating function (g.f.),  $g_p$ , of the distribution, that is, a sum of the form

$$f(z) = \sum_{k=0}^{\infty} p^k z^k = \frac{p}{1 - pz} \quad (1.15)$$



In conclusion, we should emphasize that the geometric distribution possesses the memoryless, or Markovian, property: the behavior of a sequence of Bernoulli trials, taken after an arbitrary moment, does not depend on the evolution of the trials before this moment. This statement can be written as

$$\Pr\{X = k + t | X \geq k\} = \Pr\{X = t\}$$

Of course, this property of the geometric distribution follows immediately from the definition of a Bernoulli trial. At the same time, (1.14) follows from (1.7) and the definition of the conditional probability:

$$\Pr\{X = k + t \text{ and } X \geq k\} = \Pr\{X = k\} \Pr\{X = t\} = q p^k$$

For example, in the case with cycles of successful operations of a socket, the reliability index of the socket at an arbitrary moment of time does not depend on the observed number of successful cycles before this moment.

### 1.1.3 Binomial Distribution

In a sequence of Bernoulli trials, one may be interested in the total number of successes in  $n$  trials rather than in the series of successes (or failures). In this case the r.v. of interest is

$$X = X_1 + X_2 + \dots + X_n, \quad \text{where } X_i \text{ is 1 or 0}$$

For example, consider a redundant group of  $n$  independent units operating in parallel. The group operates successfully if the number of operating (or functioning) units is not less than  $m$ . Let  $X_i$  be 1 if the  $i$ th unit is functioning at some chosen time, and 0 otherwise. Then  $X$  is the number of successfully operating units in the group. Thus, the group is operating successfully as long as  $X \geq m$ .

When considering the distribution of the r.v.  $X$ , one speaks of the *binomial distribution* with parameters  $n$  and  $p$ .

By well-known theorems of probability theory, for any set of r.v.'s  $X_i$ ,

$$E\{X\} = \sum_{i=1}^n E\{X_i\} \tag{1.16}$$

In this particular case

$$E\{X\} = np \tag{1.17}$$

## 6 FUNDAMENTALS

For independent r.v.'s the variance of  $X$  is expressed as

$$\text{Var}\{E - E \text{Var}\{*,.\}\} \quad (1.18)$$

For i.i.d. Bernoulli r.v.'s

$$\text{Var}\{X\} = npq \quad (1.19)$$

For this distribution the m.g.f. is

$$H(s) = (pe^s + q)^n \quad (1.20)$$

Both (1.17) and (1.19) can be easily obtained from (1.20).

Substituting  $e^s = z$  transforms (1.20) into the g.f. of a binomial distribution

$$f(z) = (pz + q)^n \quad (1.21)$$

The reader can see that (1.21) is a Newton binomial so the origin of the distribution's name is clear.

If one writes (1.21) in expanded form, the coefficients at  $z^k$  is the probability of  $k$  successes in  $n$  trials

$$f(z) \sim p^n z^n + \binom{n}{1} p^{n-1} q z^{n-1} + \dots + q^n \quad (1.22)$$

So the probability that there will be  $x$  successes in  $n$  trials equals the coefficient of  $z^x$

$$Pr\{X = x\} = \binom{n}{x} p^x q^{n-x} \quad (1.23)$$

Of course, (1.21) can be written in the form  $f(z) = (p + qz)^n$ . In this case the coefficient of  $z^x$  will be the probability that exactly  $x$  failures have occurred.

### 1.1.4 Negative Binomial Distribution

The negative binomial distribution arises if one considers a series of Bernoulli trials before the appearance of the  $r$ th event of a chosen type. In other words, the r.v. is a sum of a fixed number, say  $k$ , of geometric r.v.'s. This distribution is sometimes called the *Pascal distribution*.

As an illustrative example consider a relay. With each switching the relay performs successfully with probability  $p$ . With probability  $q = 1 - p$  the relay fails and then is replaced by another identical one. Let us assume that

CONTINUOUS DISTRIBUTIONS RELATED TO RELIABILITY 7

each switching is independent with a constant probability  $p$ , and the relay replaces the failed one is identical to the initial one. If there is one main and  $x - 1$  spare relays, the time to failure of the socket has a negative binomial distribution.

Thus, a negative binomially distributed r.v.  $X$  can be expressed as

$$X = \sum_{i=1}^x X_i \tag{1.23}$$

where each  $X_i$  has a geometric distribution. Of course, in a direct way one can easily find the mean and variance of the negative binomial distribution using the corresponding expressions (1.12) and (1.14) for the geometric distribution

$$E\{X\} = x E\{X_i\} = \frac{x}{p} \tag{1.24}$$

and

$$\text{Var}\{X\} = x \text{Var}\{X_i\} = \frac{xq}{p^2} \tag{1.25}$$

The m.g.f. of the negative binomial distribution can be easily written with the help of the m.g.f. of the geometric distribution:

$$1 - qe^{-s} \tag{1.26}$$

Obviously, the mean and variance can be obtained from (1.26) by a standard procedure, but less directly. The example above shows that the use of an m.g.f. can result in a more straightforward analysis.

Consider a geometric r.v. representing a series of successes terminating with a failure. Let us find the probability that  $n$  trials will terminate with the  $j$ th failure; that is, during  $n$  trials one observes exactly  $x$  geometric r.v.'s. This event can occur in the following way: the last event must be a failure by necessity (by assumption) and the remaining  $n - 1$  trials contain  $x - 1$  failures and  $(n - 1) - (x - 1) = n - x$  successes, in some order. But the latter is exactly the case that we had when we were considering a binomial distribution:  $x - 1$  failures (or, equivalently,  $n - x$  successes) in  $n - 1$  trials. The probability equals

$$\Pr\{X = n\} = \Pr\{x - 1 \text{ failures among } n - 1 \text{ trials}\} \cdot \Pr\{\text{the } n\text{th trial is a failure}\} \tag{1-27}$$

## 8 FUNDAMENTALS

The second term of the product in (1.27) equals  $q$  and the first term (considered relating to failures) is defined to be

$$f_b(x-1|p, n-1) = \binom{n-1}{x-1} p^{x-1} q^{n-x} \quad (1.28)$$

Now (1.28) can be rewritten as

$$\Pr\{X = n\} = \binom{n}{n} p^n q^0 = p^n \quad (1.29)$$

The expression (1.29) can be written in the following form:

$$\Pr\{X = n\} = p^n \quad (1.30)$$

[We leave the proof of (1.30) for Exercise 1.1.3

Equation (1.26) explains the name of the distribution.

We mention that the negative binomial and the binomial distributions are connected in the following manner. The following two events are equivalent:

- ◆ In  $n$  Bernoulli trials, the  $k$ th success occurs at the  $n$ th trial where  $n \leq k$ , and all remaining trials are unsuccessful.
- The negative binomially distributed r.v. is less than or equal to  $n$ .

The first and second events are described with the help of binomial and negative binomial distributions, respectively. In other words,

$$—k$$

Thus, in some sense, a binomial d.f. plays the role of a cumulative d.f. for an r.v. with a negative binomial d.f.

### 1.1.5 Poisson Distribution

The Poisson distribution plays a special role in many practical reliability problems. The role of the Poisson distribution will be especially clear when we consider point stochastic processes, that is, processes which are represented by a sequence of point events on the time axis.

Before we begin to use this distribution in engineering problems, let us describe its genesis and its formal properties.

Again, let us consider a sequence of Bernoulli trials. One observes experiments each with a probability of success of  $p$ , and a probability of

CONTINUOUS DISTRIBUTIONS RELATED TO RELIABILITY 9

failure of  $q_x$ . The probability of no failures occurring during the experiment is

$$\Pr\{\text{no failures}\} = p^n \tag{1.31}$$

Let the probability (1.31), that is, the probability that there are no failures in  $n$  trials, be equal to  $P$ . Now let us assume that each mentioned trial consists, in turn, of  $m$  identical and independent subtrials, or "trials of the second level." So now we consider  $n_2 = n \cdot m$  experiments at the second level. If at least one failure has occurred in this group of experiments at the second level, we will consider that a failure of the entire process has occurred. If the probability of success for this second level is  $p_2$ , then one has the obvious relationship  $p = p_2^m$  or, consequently,

$$\Pr\{\text{no failures}\} = p_2^{n_2} = P$$

We can continue this procedure of increasing the number of trials and correspondingly increasing the probability of success in such a manner that for any  $n$ th stage of the procedure

$$\Pr\{\text{no failures}\} = p^n = P$$

Now let us consider the probability of  $k$  failures for the same process at a stage with  $n$  trials and corresponding probabilities  $p$  and  $q$ . We can use the binomial distribution

$$\Pr\{k \text{ failures}\} = \binom{n}{k} p^k q^{n-k}$$

$$\frac{n!}{k!(n-k)!} p^k q^{n-k}$$

Now let us write the expression for the case when  $k$  is fixed but  $n \rightarrow \infty$  and  $p \rightarrow 0$  in correspondence with the above-described procedure:

$$\lim_{n \rightarrow \infty} \Pr\{k \text{ failures}\} = \lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$e^{-np} \frac{(np)^k}{k!} \tag{1.32}$$

Thus, the Poisson distribution can be considered as a limiting distribution for the binomial when the number of trials goes to  $\infty$  (or, in practice, is very large) and the value  $np$  is restricted and fixed.

For this case it is convenient to introduce a special parameter, say  $\lambda$ , which characterizes the intensity of a failure in a time unit for this limiting case.

For the limit (1.32) one can speak of the transformation of a discrete Bernoulli trials process into a continuous process. Then  $\lambda t$  is the mean number of failures during a time interval  $t$ . (The memoryless property of Bernoulli trials is independent of when this interval begins.) So one can

**10 FUNDAMENTALS**

substitute  $nq$  in (1.32) for  $\lambda t$  and obtain

$$\Pr\{X=k; A\} = \frac{(A t)^k}{k!} e^{-A t} \tag{1.33}$$

We will soon discuss the main applications of the Poisson distribution. Here we emphasize that this distribution is a very good approximation for the binomial distribution when the number of trials is very large and the probability of failure in a single trial is extremely small (but the mean number of events during a fixed time interval is finite).

Now let us consider different characteristics of this distribution. Based on the definition of the parameter  $A$ , one can directly find the mean, that is, the average number of failures during a time interval  $t$ ,

$$E\{X\} = A t = A \tag{1.34}$$

The equation for the m.g.f. can be easily obtained with the use of (1.33)

$$= E\{e^{Xz}\} = \sum_{k=0}^{\infty} \frac{(A t)^k}{k!} e^{-A t} e^{kz} = e^{A t(z-1)} \tag{1.35}$$

The expression can also be used to obtain the second moment

$$= A^2 + A = \frac{d^2}{dz^2} e^{A t(z-1)} \Big|_{z=0} \tag{1.36}$$

and hence from (1.34) and (1.36) we obtain

$$\text{Var}\{X\} = A \tag{1.37}$$

**1.2 CONTINUOUS DISTRIBUTIONS RELATED TO RELIABILITY**

**1.2.1 Exponential Distribution**

The exponential distribution is the most popular and commonly used distribution in reliability theory and engineering. Its extreme popularity usually generates two powerful "lobbies" among the community of reliability specialists: "exponentialists" and "antiexponentialists." Both groups have many pro's and con's. Sometimes these groups remind one of the two political parties of egg eaters described by Jonathan Swift in his famous book *Gulliver's Travels*

The "exponential addicts" in engineering will tell you that this distribution is very attractive because of its simplicity. This may or may not be a good reason! Many mathematical researchers love the exponential distribution

because they can obtain a lot of elegant results with it. If, in fact, the investigated problem has at least some relation to an exponential model, this is an excellent reason!

Antagonists of the exponential distribution maintain that it is an unreasonable idealization of reality. There are no actual conditions that could generate an exponential distribution. This is not a bad reason for criticism. But on the other hand, it is principally impossible to find a natural process that is exactly described by a mathematical model.

The real question that must be addressed is: under which conditions it is appropriate to use an exponential distribution. It is necessary to understand the nature of this distribution and to decide if it can be applied in each individual case. Therefore, sometimes an exponential distribution can be used, and sometimes not. We should always solve practical problems with a complete understanding of what we really want to do.

Consider a geometric distribution and take the expression for the probability that there is no failure during  $n$  trials. If  $n$  is large and  $p$  is close to 1, one can use the approximation

$$\Pr\{A' > n\} = (1 - q)^n \sim e^{-nq} \tag{1.38}$$

If we consider a small time interval  $dt$ , then the probability of failure for a continuous distribution must be small. In our case this probability is constant for equal intervals. Let

$$\Pr\{\text{failure during } A\} = A \lambda$$

Then, for the r.v.  $X$ , a continuous analogue of a geometric r.v., with  $n \rightarrow \infty$  and  $A \rightarrow 0$ , we obtain

$$\lim_{A \rightarrow 0} (1 - \lambda A)^{n/\lambda k} = e^{-k} \tag{1.39}$$

It is clear that the exponential distribution is a continuous analogue of the geometric distribution under the aforementioned conditions. Using the memoryless property, (1.39) can be obtained directly in another way. This property means that the probability of a successful operation during the time interval  $t + x$  can be expressed as

$$P(t + x) = P(t) P(x|t) = P(t) \cdot P(x)$$

$$f(t+x) - f(t) + f(x) \tag{1.40}$$

where  $f(y) = -\ln P(y)$ .

12 FUNDAMENTALS

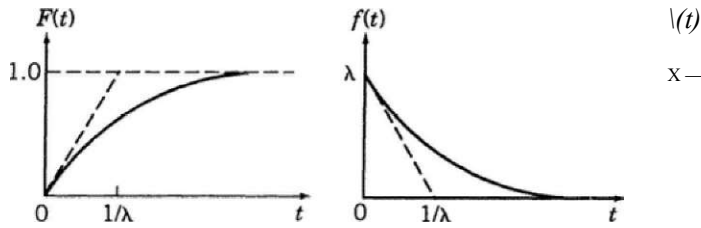


Figure 1.1. Exponential distribution  $F(t)$ , its density  $f(t)$ , and its hazard function  $\lambda(t)$ .

But the only function for which (1.40) holds is the linear function. Let  $f(y) = ay$ . Then  $P(y) = \exp(-ay)$ . Now one uses the condition that  $F(\infty) = 1 - P(\infty) = 1$  and finds that  $a = \lambda$ . Therefore, the probability of having no failure during the period  $t$  equals

$$P(t) = 1 - F(t) = \exp(-\lambda t) \tag{1.41}$$

The distribution function is

$$F(t) = 1 - \exp(-\lambda t)$$

and the density function is

$$f(t) = \lambda \exp(-\lambda t)$$

The exponential distribution is very common in engineering practice. It is often used to describe the failure process of electronic equipment. Failures of such equipment occur mostly because of the appearance of extreme conditions during their operation. We will show below that such events can be successfully described by a Poisson process. In turn, the Poisson process very closely relates to the exponential distribution.

In addition, we should emphasize that the exponential distribution appears in several practical important cases when one considers highly reliable repairable (renewal) systems.

Both of these cases are related to the case where a continuous (or discrete) stochastic process crosses a high-level threshold. Indeed, intuitively we feel that a level might be considered as "high" because it is very seldom reached.

Now let us find the main characteristics of the exponential distribution. The easiest way to find the mean of the exponential r.v. is to integrate the function  $P(t) = 1 - F(t)$ .

$$E\{A\} = \int_0^{\infty} \lambda t e^{-\lambda t} dt = \frac{1}{\lambda} \tag{1.43}$$



The second initial moment of the distribution can also be found in a direct way

$$E\{x^2\} = \int_0^{\infty} Ckt_j e^{-x'} dt = \frac{2}{A} \quad (1.44)$$

and, consequently, from (1.43) and (1.44)

$$V^{**} = \frac{2}{A} - \left(\frac{1}{A}\right)^2$$

that is, the standard deviation of an exponential distribution equals the mean

$$\sigma = \sqrt{V} = \frac{1}{A}$$

The m.g.f. for the density can also be found in a direct way

$$\int_0^{\infty} A e^{-s t} dt = \frac{A}{A + s} \quad (1.45)$$

For future applications it is convenient to have the Laplace-Stieltjes transform (LST) of a density function. For the density of an exponential distribution, the LST equals

$$\int_0^{\infty} A e^{-k t} e^{-s t} dt = \frac{A}{A + s} \quad (1.46)$$

As we considered above, the LST of the function  $P(t) = 1 - F(t) = e^{-A t}$ , taken at  $s = 0$ , equals the mean. In this case

$$\int_0^{\infty} A e^{-A t} dt = \frac{1}{A} \quad (1.47)$$

and, consequently,

$$\int_0^{\infty} A e^{-A t} dt = \frac{1}{A}$$

One very important characteristic of continuous distributions is the *intensity function* which, in reliability theory, is called the *failure rate*. This function is determined as the conditional density at a moment  $t$  under the

## 14 FUNDAMENTALS

condition that the r.v. is not less than  $t$ . Thus, the intensity function is

$$1 - F(t) = P(t)$$

For the exponential distribution the intensity function can be written as

$$\lambda e^{-\lambda t} = \lambda e^{-\lambda t} \quad (1.48)$$

that is, the failure rate for an exponential distribution is constant. This follows as well from the memoryless property. In reliability terms it means, in particular, that current or future reliability properties of an operating piece of equipment do not change with time and, consequently, do not depend on the amount of operating time since the moment of switching the equipment on. Of course, this assumption seems a little restrictive, even for "exponential addicts." But this mathematical description is sometimes practically sufficient.

### 1.2.2 Erlang Distribution

The Erlang distribution is the continuous analogue of a negative binomial distribution. It represents the sum of a fixed number of independent and exponentially distributed r.v.'s. The principal mathematical model for the description of queuing processes in a telephone system is a Markov one. Consider a multiphase stage, for example, a waiting line of messages. An observed message can stand in line behind several previous ones, say  $N$ . Then for this message the waiting time can be represented as a sum of the  $N$  serving times of the previous messages. By assumption, for a Markov-type model, each of these serving times has an exponential distribution, and so the resulting waiting time of the message under consideration has an Erlang distribution.

The sum of  $N$  independent exponential r.v.'s forms an Erlang distribution of the  $N$ th order. It is then clear that the mean of an r.v. with an Erlang distribution of the  $j$ Vth order is a sum of  $N$  means of exponential r.v.'s, that is,

CONTINUOUS DISTRIBUTIONS RELATED TO RELIABILITY 15

$$B\{X\} = j \quad (1.49)$$

## 16 FUNDAMENTALS

and so the variance equals  $N$  times the variance of a corresponding exponential distribution

$$\text{Var}\{X\} \tag{1.50}$$

Finally, the LST of the density of an Erlang distribution of the  $N$ th order is

$$* (*) - \tag{1.51}$$

The last expression allows us to write an expression for the density function of this distribution

(e.g., one can use a standard table of the Laplace-Stieltjes transforms). We will show the validity of (1.52) below when we consider a Poisson process.

Note that if the exponential r.v.'s which compose the Erlang r.v., are not identical, the resulting distribution is called a *generalized Erlang distribution*. Here we will not write the special expression for this case but one can find related results in Section 1.6.7 dedicated to the so-called death process.

### 1.2.3 Normal Distribution

This distribution occupies a special place among all continuous distributions because many complex practical cases can be modeled by it. This d.f. is often termed a *Gaussian distribution*.

The central limit theorem of probability theory states that the sum of independent r.v.'s under some relatively nonrestrictive conditions has an asymptotically normal distribution. This fundamental result has an intriguing history which has developed over more than two centuries.

A simple example of a practical application of the central limit theorem in engineering occurs in the study of the supply of spare parts. Assume that some unit has a random time to failure with an unknown distribution. We know only the mean and variance of the distribution. (These values can be estimated, even with very restricted statistical data.) If we are planning to supply spare parts over a long period of time, as compared to the *mean time to failure* (MTTS) of the unit, we can assume that the total time until exhaustion of  $n$  spare units has an approximately normal distribution. This approximation is practically irreproachable if the number of planned spare parts,  $n$ , is not less than 30.

In engineering practice the normal distribution is usually used for the description of the dispersion of different physical parameters. For example, the resistance or electrical capacity of a sample of units is often assumed to be normally distributed; the normal distribution characterizes the size of mechanical details; and so on. Incidentally, many mechanical structures exposed to wear are assumed to have a normal d.f. describing their time to failure.

The normal distribution of the random time to failure (TTF) also appears when the main parameter changes linearly in time and has a normal distribution of its starting value. (The latter phenomenon was mentioned above.) In this case the time to the exceedance of a specified tolerance limit will have normal distribution. We will explain this fact in mathematical terms below.

The normal distribution has the density function

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\} \quad (1.53)$$

where  $\mu$  and  $\sigma^2$  are the mean and the variance of the distribution, respectively. These two parameters completely characterize the normal distribution. The parameter  $\sigma$  is called the *standard deviation*. Notice that  $\sigma$  is always nonnegative. From (1.53) one sees that it is a symmetrical unimodal function; it has a bell-shaped graph (see Figure 1.2).

That  $\mu$  and  $\sigma^2$  are, respectively, the mean and the variance of the normal distribution can be shown in a direct way with the use of (1.53). We leave this direct proof to Exercises 1.2 and 1.3. Here we will use the m.g.f.

$$M(s) = \int_{-\infty}^{\infty} e^{-sx} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx = \exp\left(\mu s - \frac{\sigma^2 s^2}{2}\right) \quad (1.54)$$

(The proof of this is left to Exercise 1.4.)

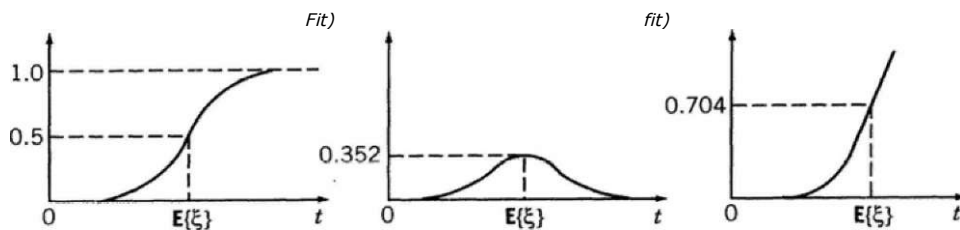


Figure 1.2. Normal distribution  $F(t)$ , its density  $f(t)$ , and its hazard function  $A(t)$ .

18 FUNDAMENTALS

From (1.54) one can easily find

$$E\{X^2\} = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx = a^2 + cr^2 \quad (1.55)$$

$$E\{X\} = \int_{-\infty}^{\infty} x \cdot f(x) dx = a \quad (1.56)$$

and

$$\text{Var}\{X\} = cr^2 \quad (1.57)$$

[The proof of (1.56) is left to Exercise 1.5.]

In applications one often uses the so-called standard normal d.f. In this case  $a = 0$  and  $a = 1$ . It is clear that an arbitrary normal r.v.  $X$  can be reduced to a standard one. Consider the new r.v.  $X' = X - a$  (obviously, the variances of  $X$  and  $X'$  are equal) and normalize this new r.v., by dividing by  $cr$ . In this way an arbitrary normal distribution can be reduced to the standard one (or vice versa) by means of a linear change of scale and changing the location of its mean to 0. The density of a normal d.f. is (see Figure 1.2)

<TLLT

The function (1.58) has been tabulated in different forms and over a very wide range (see Fig. 1.3). Using the symmetry of the density function, one can compile a table of the function

$$e^{-x^2/2} = \int_0^x f_n(x, \lambda) dx$$

The correspondence between the functions  $F_n(x)$  and  $F^*(x)$  is

Often one can find a standard table of the so-called *Laplace function*:  $I - 2F(x)$ . This kind of table is used, for instance, in artillery calculations to find the probability of hitting a target.

The distribution function of a normal distribution decreases very rapidly with increasing  $x$ . Most standard tables are, in fact, composed for  $|x| < 3$  or 4, which is enough for most practical purposes. But sometimes one needs

values for larger  $x$ . In this case we suggest the following iterative computational procedure.

20 FUNDAMENTALS

Consider the integral

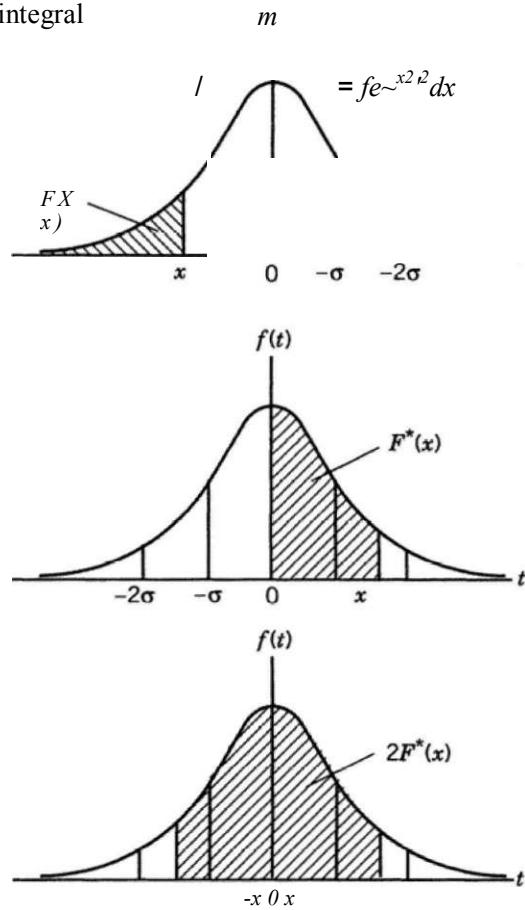


Figure 1.3. Three types of tabulated functions for the standard normal distribution.

It can be rewritten as

Using integration by parts, one obtains

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \delta(x-t) e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \delta(x-t) e^{-t^2/2} dt$$



Now we can evaluate I

$$\int_0^x \frac{e^{-t^2}}{t^3} dt = \frac{1}{2} \left( \frac{1}{x^2} - \frac{1}{r^2} \right) - \frac{1}{2} \int_r^x \frac{1}{t^3} dt$$

Thus at this stage of the iteration

and after integration by parts

$$I = \frac{1}{2} \left( \frac{1}{x^2} - \frac{1}{r^2} \right) - \frac{1}{2} \int_r^x \frac{1}{t^3} dt$$

More accurate approximations can be obtained in an analogous manner.

### 1.2.4 Truncated Normal Distribution

A normal d.f. ranges from  $-\infty$  to  $+\infty$ . But in reliability theory one usually focuses on the *lifetime* of some object, and so we need consider distributions defined over the domain  $[0, +\infty)$ . The new d.f. (see Fig. 1.4) is said to be "truncated (from the left)." The new density function,  $f(x|a, \sigma)$ , can be related to the initial one,  $f(x|a, \sigma)$ , as follows:

$$\tilde{f}(x|a, \sigma) = \frac{f(x|a, \sigma)}{\int_0^\infty e^{-(x-a)^2/2\sigma^2} dx}$$

In practical problems this truncation often has a negligible influence if  $a/\sigma$  is greater than 4 or 5.

The mean of a truncated distribution is always larger than the mean of its related normal distribution. The variance, on the other hand, is always smaller. We will not write these two expressions because of their complex form.

### 1.2.5 Weibull - Gnedenko Distribution

One of the most widely used distributions is the Weibull-Gnedenko distribution. This two-parameter distribution is convenient for practical applications because an appropriate choice of its parameters allows one to use it to describe various physical phenomena. One of the parameters,  $A$ , is called the scale parameter and another,  $\beta$ , is called the shape parameter of the

FUNDAMENTALS

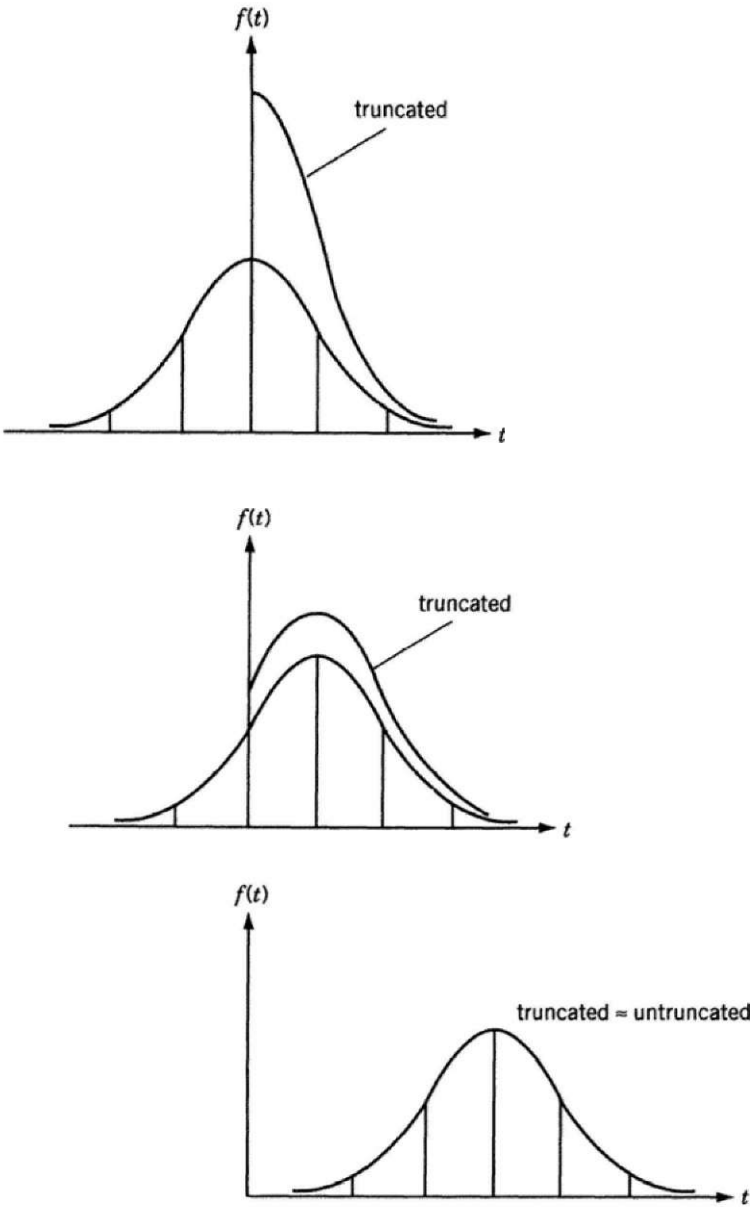


Figure 1.4. Normal distribution with different levels of truncation.

distribution. A Weibull-Gnedenko distribution has the form

$$F(t) = \begin{cases} h - e^{-(At)^p} & \text{for } t > 0 \\ 0 & \text{for } t < 0 \end{cases}$$

The density function is

$$f(t) = \begin{cases} pAt^{p-1} e^{-(At)^p} & \text{for } t > 0 \\ 0 & \text{for } t < 0 \end{cases}$$

The density function for several different parameter values is presented in Figure 1.5.

The failure rate of the distribution is

$$A(t) = A^{p-1} t^{p-1}$$

The behavior of the failure rate depending on the parameter values is depicted in Figure 1.6. For  $p = 1$ , the Weibull-Gnedenko d.f. transforms into a common exponential function (the failure rate is constant). For  $p > 1$ , one observes an increasing failure rate: for  $1 < p < 2$ , this is concave; for  $p \geq 2$ , this is convex. For  $0 < p < 1$ , the failure rate is decreasing.

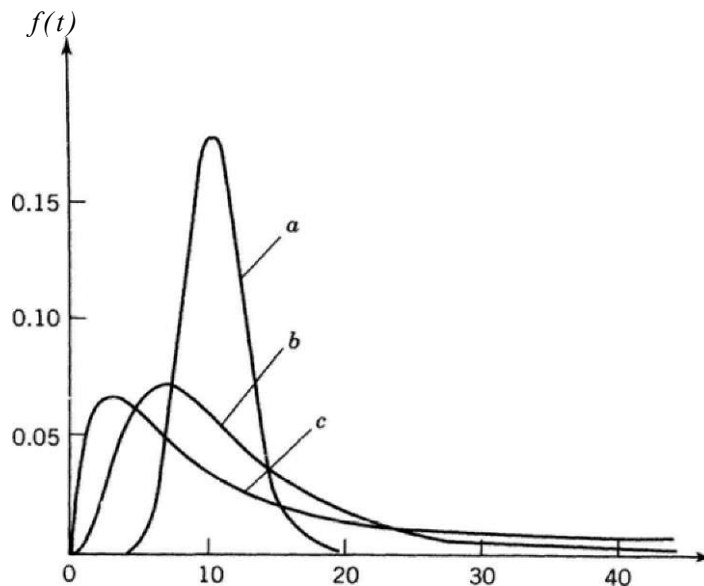


Figure 1.5. Density of the Weibull-Gnedenko distribution  $f(t)$  for the following parameters: (a)  $p = 1, A = 1$ ; (b)  $p = 2, A = 1$ ; (c)  $p = 4, A = 1; p = 2, A = 0.7$ .

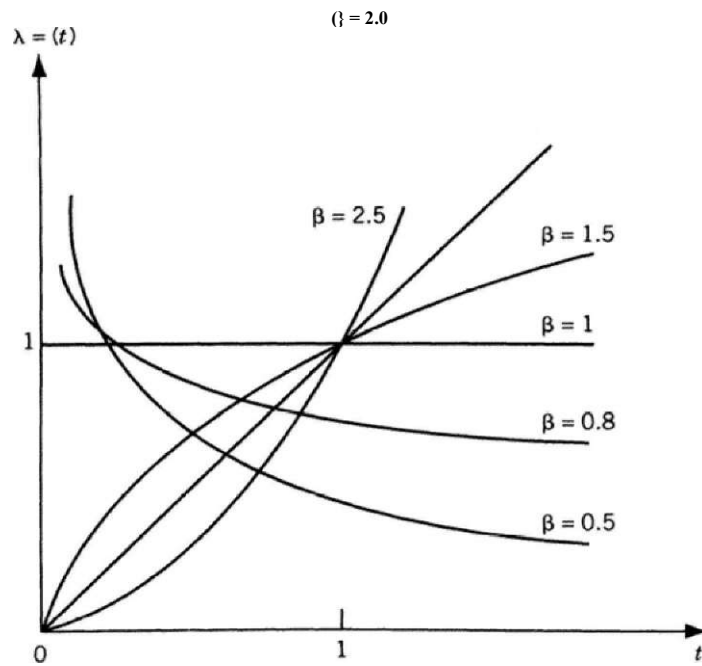


Figure 1.6. Hazard rate for the Weibull-Gnedenko distribution with  $A = 1$  and different parameter  $\beta$ .

The mean of this d.f. is

$$E\{T\} = \frac{\Gamma(1 + 1/\beta)}{\beta}$$

and the variance is

$$\text{Var}\{T\} = \frac{\Gamma(1 + 2/\beta) - (\Gamma(1 + 1/\beta))^2}{\beta^2}$$

where  $\Gamma(\cdot)$  is the gamma function.

### 1.2.6 Lognormal Distribution

In mechanics one often sees that material fatigue follows a so-called lognormal distribution. This distribution appears if the logarithm of the time to failure has a normal distribution. For  $f > 0$ , one has

$$F(t) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\log t - \mu}{\sigma} \right) \right]$$

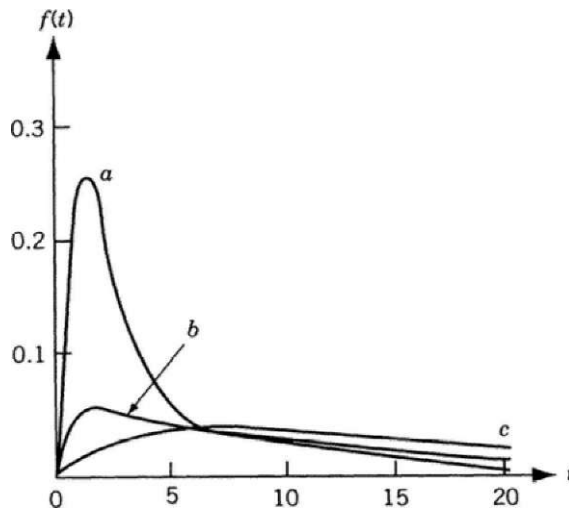


Figure 1.7. Density function for the log normal distribution with different parameters: (a)  $\mu = 1, \sigma = 1$ ; (b)  $\mu L = 3, a = 1.7$ ; (c)  $\mu = 3, a = 1$ .

and the density is

$$f(t) = \frac{1}{\sigma \sqrt{2\pi} t} \exp\left(-\frac{(\log f - \mu)^2}{2\sigma^2}\right) \quad \text{for } t > 0$$

A sample of a lognormal distribution for several parameter values is depicted in Figure 1.7. The mean and the variance have the following forms, respectively:

$$E\{t\} = e^{\mu + \frac{\sigma^2}{2}} \quad \text{and} \quad \text{Var}\{t\} = e^{2\mu} (e^{\sigma^2} - 1)$$

and

$$\text{Var}\{t\} = e^{2\mu} (e^{\sigma^2} - 1)$$

For a small coefficient of variation, one can use a normal approximation for a lognormal d.f.

### 1.2.7 Uniform Distribution

For this distribution the density function is constant over its domain  $[a, b]$ . The graphs of the density and distribution functions are presented in Figure 1.8. The density function is

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a < x < b \\ 0 & \text{for } x < a \text{ and } x > b \end{cases} \quad (1.59)$$

and the d.f. is

$$F(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a < x < b \\ 1 & \text{for } x > b \end{cases} \quad (1.60)$$

Because of the symmetry of the density function, the mean is  $(b + a)/2$ . The variance can be calculated as

$$\text{Var}\{X\} = \int_a^b \frac{(b-a)^2}{12} dx = \frac{(b-a)^2}{12} \quad (1.61)$$

The uniform distribution on the interval  $[0,1]$  plays an important role in reliability and its related applications. It is determined by the fact that an r.v.  $y = F^{-1}(x)$  [here  $F^{-1}$  is the inverse for  $F(x)$ ] has a uniform distribution. This fact is often used for the generation of r.v.'s with a desired distribution on the basis of uniformly distributed r.v.'s. For example, to generate an r.v.  $X$  with a specified d.f.  $F(x)$ , we must take the generator of a uniformly distributed r.v.  $Y_1, Y_2, \dots$  and arrange the inverse transforms:  $X = F^{-1}(Y_1)$ ,  $Y_2 = F^{-1}(Y_2)$ , ... ■

For computer simulations the so-called *pseudo-random variable* (p.r.v.) is usually generated. The first generator of uniformly distributed r.v.'s was

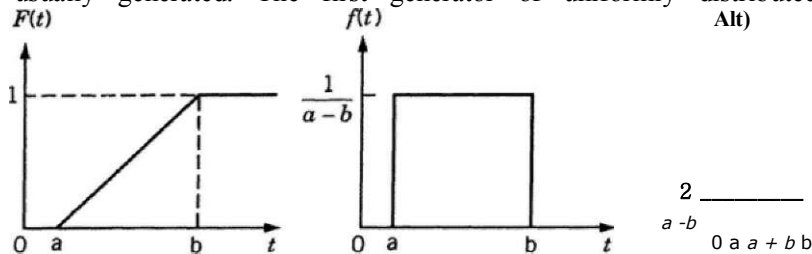


Figure 1.8. Uniform distribution  $F(t)$ , its density  $f(t)$ , and its hazard function  $AO$ .

introduced by John von Neuman. The principle consists in the recurrent calculation of some function.

For example, one takes an exponential function with some two-digit power, chooses, say the 10th and 11th digits as the next power, and repeats the procedure from the beginning. Of course, such a procedure leads to the formation of a cycle: as soon as the same power appears, the continuation of the procedure will be a complete repetition of one of the previous links of p.r.v.'s. At any rate, it is clear that the cycle cannot be larger than 100 p.r.v.'s if the power of the exponent consists of two digits. Fortunately, modern p.r.v. generators have practically unrestricted cycle lengths.

At the same time, p.r.v.'s are very important for different numerical simulation experiments designed for comparison of different variants of a system design. Indeed, one can completely repeat a set of p.r.v.'s by starting the procedure from the same initial state. This allows one to put different system variants into an equivalent pseudo-random environment. This is important to avoid real random mistakes caused by putting one system in a more severe "statistical environment" than another.

### 1.3 SUMMATION OF RANDOM VARIABLES

The summation of random variables often comes up in engineering problems involving a probabilistic analysis. The observation of a series of time sequences or the analysis of the number of failed units arriving at a repair shop are examples. At the same time, the number of terms in the sum is not always given—sometimes it is random. Asymptotic results are also of practical interest.

#### 1.3.1 Sum of a Fixed Number of Random Variables

**General Case** Consider a repairable system which is described by cycles as "a period of operation" and "a period of repair." Each cycle consists of two r.v.'s  $\xi$  and  $\eta$ , a random time to failure (TTF) with distribution  $F(t)$ , and a random repair time with distribution  $G(t)$ , respectively. If the distribution of the complete cycle is of interest, we would analyze the sum  $\sigma = \xi + \eta$ . The distribution of this new r.v., denoted as  $D(t) = \Pr\{0 < t\}$ , is the convolution of the initial d.f.'s:

$$\begin{aligned}
 D(t) &= \Pr\{\xi + \eta \leq t\} = F * G(t) = \int_0^t F(t-x) dG(x) \\
 &= G * F(t) = \int_0^t G(t-x) dF(x)
 \end{aligned}$$

28 FUNDAMENTALS

If the Laplace-Stieltjes transforms (LSTs) of these d.f.'s

$$\langle P_F(s) = \int_0^\infty F(t) e^{-st} dt$$

and

$$\langle P_G(s) = \int_0^\infty G(t) e^{-st} dt$$

are known, the LST of the d.f.  $D(t)$  is

$$\{P_D(s) =$$

If one considers a sum of  $n$  i.i.d. r.v.'s, the convolution  $F^{*n}(t)$  is

$$\Pr\{E \leq x\} = \int_0^x \Pr\{E_1 + \dots + E_n \leq x - t\} dF(t)$$

$$\int_0^x$$

where all  $F^{*n}$ 's are determined recurrently. For a sum of i.i.d. r.v.'s each of which has LST equal to  $\langle p(s)$ ,

$$\langle P_N(s) = [p(s)]^n$$

For the sum of  $n$  r.v.'s with arbitrary distributions, one can write

$$E\{X^n\} = E\{E_1 + \dots + E_n\}^n = E\{E_1^n + \dots + E_n^n\} \tag{1.62}$$

and, for independent r.v.'s.

$$E\{X^n\} = E\{E_1^n\} + \dots + E\{E_n^n\} \tag{1-63}$$

Now we begin with several important and frequently encountered special cases.

**Sum of Binomial Random Variables** Consider two binomially distributed r.v.'s and  $v_2$  obtained, respectively, by and  $n_2$  Bernoulli trials with the same parameter  $q$ . From (1.21), the g.f. of the binomial distribution is

$$\mathcal{L}(s) = (pz + q)^{n_1} \tag{1.64}$$



Thus,

$$S(z) = \langle p_1(z) \rangle \langle p_2(z) \rangle = (pz + q)^n (pz + q)^{n_2} = (pz + q)^{n+n_2} \quad (1.65)$$

In other words, the sum of two binomially distributed r.v.'s with the same parameter  $p$  will produce another binomial distribution. This can be easily explained: arranging a joint sample from two separate samples of sizes  $n_1$  and  $n_2$  from the same population is equivalent to taking one sample of size  $n_1 + n_2$ .

Obviously, an analogous result holds for an arbitrary finite number of binomially distributed r.v.'s. Thus, we see that the sum of binomially distributed r.v.'s with the same parameter  $p$  produces a new binomially distributed r.v. with the same  $p$  and corresponding parameter

$$n = \sum_{i=1}^N n_i$$

For different binomial distributions, the result is slightly more complicated (see Exercise 1.8).

**Sum of Poisson Random Variables** Consider the sum  $X_z$  of two independent Poisson r.v.'s  $X_1$  and  $X_2$  with corresponding parameters  $A_i$ ,  $i = 1, 2$ . The m.g.f.'s for the two Poisson distributions are written as

$$V_i(z) = e^{W_i(z)} \quad i = 1, 2 \quad (1.66)$$

The m.g.f. for the distribution of the sum  $X_z$  can be written as

$$\langle p(z) \rangle = \langle p_1(z) \rangle \langle p_2(z) \rangle = e^{(A_1 + A_2)(z-1)} \quad (1.67)$$

that is, the resulting m.g.f. is the m.g.f. of a new Poisson d.f. with parameter  $A_z = A_1 + A_2$ .

An analogous result can be obtained for an arbitrary finite number of Poisson r.v.'s. In other words, the sum of  $N$  Poisson r.v.'s is again a Poisson r.v. with parameter equal to the sum of the parameters:

$$A = \sum_{i=1}^N A_i \quad (1.68)$$

**Sum of Normal Random Variables** The sum of independent normally distributed r.v.'s has a normal distribution. Again consider a sum of two r.v.'s. Let  $X_j$  be a normal r.v. with parameters  $\mu_j$  and  $\sigma_j$ ,  $j = 1, 2$ , and let

$\hat{z} \sim + X_2$ - Then the m.g.f. for  $X$  can be expressed as

$$\begin{aligned} &= \text{Vi}(2)p_2(z)^{\text{TM}} \exp(a, z + \backslash a\} z^2) \exp(a, z + \backslash a\} z^2) \\ &= \exp[ + a_2) + \pm z^2 (<rf + cr_2^2)] \end{aligned} \quad (1.69)$$

Therefore, the sum of two normal r.v.'s produces an r.v. with a normal distribution. For  $n$  terms, the parameters of the resulting normal distribution are

$$a_z - \text{£} \text{Of} \quad (1.70)$$

*J s i s N*

and

$$\langle r_N = \quad (1.71)$$

*V ts/sAf*

### 1.3.2 Central Limit Theorem

Many statisticians have worked on the problem of determining the limit distribution of a sum of r.v.'s. This problem has practical significance because, when a sum includes a large number of r.v.'s, the direct calculation of some characteristic of the sum becomes very complicated. The problem itself has aroused theoretical interest even outside of applications.

Above we showed that a sum of different normally distributed independent r.v.'s has a normal distribution, independent of the number of terms in the sum. The new resulting normal distribution has a mean equal to the sum of the means of the initial distributions and a variance equal to the sum of the variances. It is obvious that this property is preserved with the growth of  $n$ .

But what will be the limiting distribution of a sum of r.v.'s whose distributions are not normal? It turns out that, with increasing  $n$ , such a sum has a tendency to converge to a normally distributed r.v.

In simple engineering terms it appears that if we consider a sum of a large number  $n$  of independent r.v.'s then this sum has approximately a normal distribution. If we consider the sum of independent arbitrary distributed r.v.'s  $\text{£}$  with mean  $a = E\{\text{£}\}$  and variance  $v = \text{Var}\{\text{£}\}$ , then the normal distribution of the sum will have mean  $A = an$  and variance  $V = un$ . (Of course, some special restrictions on the independence and properties of distributions must be fulfilled.)

Historically, limit theorems developed over several centuries. Different versions of them pertain to different cases. One of the first attempts in this direction is contained in the following theorem.

**DeMoivre Local Theorem** Consider a sequence  $n$  of Bernoulli trials with a probability of success  $p$ . The probability of  $m$  successes  $P_{n,m}$  satisfies the relationship

$$\frac{e^{-y^2/2t}}{\sqrt{2\pi t}}$$

uniformly for all  $m$  such that

$$\frac{m - np}{\sqrt{npq}}$$

belongs to some finite interval.

This theorem, in turn, is the basis of the following theorem.

**Integral DeMoivre - Laplace Theorem** If  $v$  is the random number of successes among  $n$  Bernoulli trials, then for finite  $a$  and  $b$  the following relationship holds:

The next step in generalizing the conditions under which the sum of a sequence of arbitrary r.v.'s converges to a normal distribution is formulated in the following theorem.

**Liapounov Central Limit Theorem** Suppose that the r.v.'s  $X_i$  are independent with known means  $a_i$  and variances  $\sigma_i^2$ , and for all of them,  $\sum_{i=1}^n \sigma_i^3 / (\sum_{i=1}^n \sigma_i^2)^{3/2} \rightarrow 0$ . Also, suppose that

$$\lim_{n \rightarrow \infty} \frac{E \left[ \frac{(X - \sum_{i=1}^n a_i)^3}{(\sum_{i=1}^n \sigma_i^2)^{3/2}} \right]}{\sum_{i=1}^n \sigma_i^2} = 0$$

Then, for the normalized and centered (with zero mean) r.v.,

for any fixed number  $x$ ,

$$\lim_{n \rightarrow \infty} \Pr \left\{ \frac{\bar{X}_n - a}{\sqrt{2\pi}} < y \right\} = \int_{-\infty}^y e^{-z^2/2} dz$$

Thus, this theorem allows for different r.v.'s in the sequence and the only restrictions are in the existence of moments of an order higher than 2. As a matter of fact, this statement is true even under weaker conditions (the restriction of a variance is enough) but all r.v.'s in the sum must be i.i.d. For the sample mean, the related result is formulated in the following theorem.

**Lindeberg - Levy Central Limit Theorem** If the r.v.'s  $X_t$  are chosen at random from a population which has a given distribution with mean  $a$  and finite variance  $a^2$ , then for any fixed number  $y$ ,

$$\lim_{n \rightarrow \infty} \Pr \left\{ \frac{\bar{X}_n - a}{\sqrt{2\pi}} < y \right\} = \int_{-\infty}^y e^{-z^2/2} dz$$

where  $\bar{X}_n$  is the sample mean.  
Because

this theorem may be interpreted in the following way: the sum of i.i.d. r.v.'s approximately has a normal distribution with mean equal to  $na$  and variance equal to  $na^2$ .

A detailed historical review on the development of probability theory and statistics can be found in Gnedenko (1988).

### 1.3.3 Poisson Theorem

Considering the local DeMoivre theorem, we notice that this result works well for binomial distributions with  $p$  close to 1/2. But the normal approximation does not work well for small probabilities or on the "tails" of a binomial distribution. An asymptotic result for small  $p$  (for the "tails" of the binomial distribution) is formulated in the following theorem.

**Poisson Theorem** If  $p_n \rightarrow 0$  with  $n \rightarrow \infty$  then

$$\binom{n}{m} p_n^m (1 - p_n)^{n-m} - \frac{a_n^m}{m!} e^{-a_n} \rightarrow 0$$

where  $a_n = np$ ,

### 34 FUNDAMENTALS

This means that for small  $p$ , instead of calculating the products of astronomically large binomial coefficients with extremely small  $p^n$ , we can use a simple approximation. A standard table of the Poisson distribution can be used.

#### 1.3.4 Random Number of Terms in the Sum

Only a very general result can be given for the d.f. of the sum of r.v.'s, or for its LST when a random number of terms is distributed arbitrarily. Further, let us assume that  $v$  is geometrically distributed. Then the distribution of the sum of arbitrarily distributed r.v.'s is

$$P\{f_0^* \} - E\{pW E L z t\}$$

Consider a continuous d.f. The LST can be written as

$$\langle Pz\{s\} = \sum_{t \in \mathbb{N}} p^k q [ \langle p(s) ]^k$$

In general, both of the latter expressions are practically useful only for numerical calculation.

To find the mean of we may use the Wald equivalence:

$$E\{ \mathbf{E} f^* \} = E\{ \mathbf{E} \} \mathbf{E} U \} \quad (1-72)$$

Below we consider two cases where the sum of finite r.v.'s will lead to simple results.

**Geometrically Distributed Random Variables** We can investigate this case without using a mathematical technique. Consider an initial sequence of Bernoulli trials. The probability of success equals  $p$  and the probability of failure equals  $q = 1 - p$ . Now construct the new process consisting of only failures of the initial process and corresponding spaces between them. Consider a new procedure: each failure in the initial Bernoulli process creates a possibility for the appearance of a failure in the final process. (Failure cannot appear in the space between failures of the initial process.) A special moment concerning the "possibility" of a new process failure is considered. Let a failure of the initial process develop into a failure of the new (final) process with probability  $Q$ . Thus, if we consider the initial process, failure of the final process occurs there with probability  $Q^* = qQ$ . We have obtained this result using only verbal arguments. Of course, it can be derived in strict mathematical terms.

**Exponentially Distributed Random Variables** Consider the sum of a random number of exponentially distributed identical and independent r.v.'s, with parameter  $\lambda$ . Assume that the number of terms in the sum has a geometric d.f. with parameter  $p$ . We will express the LST of the resulting density function through the LST of the density function of the initial d.f. From the formula for the complete mathematical expectation, we have

$$\begin{aligned}
 &= q \cdot \frac{\lambda}{\lambda + s} + pq \frac{\lambda^2}{(\lambda + s)^2} + p^2 q \frac{\lambda^3}{(\lambda + s)^3} + \dots \\
 &= \lambda + s \cdot \frac{p \lambda^k}{(\lambda + s)^k} = \frac{\lambda + s}{1 - p \frac{\lambda}{\lambda + s}} = \frac{\lambda + s}{1 - p + p \frac{\lambda}{\lambda + s}} = \frac{\lambda + s}{1 - p + p \frac{\lambda}{\lambda + s}} \quad \cdot J_{73}
 \end{aligned}$$

Thus, we have an expression which represents the LST for an exponential distribution with parameter  $\lambda = \lambda q$ .

We illustrate the usefulness of this result by means of a simple example. Imagine a socket with unit installed. Such a unit works for a random time, distributed exponentially, until a failure occurs. After a failure, the unit is replaced by a new one. The installation of each new unit may lead to a socket failure with probability  $q$ .

This process continues until the first failure of the socket. This process can be described as the sum of a random number of exponentially distributed random variables where the random number has a geometrical distribution.

Of course, in general, the final distribution of the sum strongly depends on the distribution of the number of terms in the sum. The distribution of the number of terms in the sum is the definitive factor for the final distribution.

**1.3.5 Asymptotic Distribution of the Sum of a Random Number of Random Variables**

In practice, we often encounter situations where, on the average, the random number of terms in the sum is very large. Usually, the number of terms is assumed to be geometric. If so, the following limit theorem is true.

**Theorem 1.1** Let  $\{X_n\}$  be a sequence of i.i.d. r.v.'s whose d.f. is  $F(x)$  with mean  $a > 0$ . Let  $\nu$  be the number of discrete r.v.'s of a sequence with a geometric distribution with parameter  $p$ :  $\Pr\{\nu = k\} = (1-p)^{k-1} p$  where  $0 < p < 1$ . Then, if  $p \rightarrow 1$ , the d.f. of the normalized sum

$$\frac{X_1 + \dots + X_\nu}{\nu a}$$

converges to the exponential d.f.  $1 - e^{-x}$ .

36 FUNDAMENTALS

*Proof.* Consider the normalized r.v.

$$E \{ \sum_{i=1}^k \xi_i \}$$

By the Wald equivalency,

$$E \{ L \} = E \{ f \} E \{ \xi \}$$

Without loss of generality, we can take  $E \{ \xi \} = 1$ . Because  $v$  has a geometric distribution,  $E \{ p \} = 1/q$ . Hence,

$$E \{ p^k \} = \sum_{s=1}^{\infty} p^k q e^{-qs}$$

The LST of is

$$= E \{ e^{-rL} \} = E \{ \exp \{ -qs \} \}$$

$$E \{ p^k \} = \sum_{s=1}^{\infty} p^k q e^{-qs}$$

Note that

$$\exp \{ -qs \} = [e^{-qs}]$$

Then

$$= E \{ p^k \} = \sum_{s=1}^{\infty} p^k q e^{-qs}$$

$$= E \{ [e^{-qs}]^k \} =$$

Now with some simple transformations

$$\dots = \frac{q e^{-qs}}{1 - p e^{-qs}} = \frac{q e^{-qs}}{1 - p e^{-qs} + q e^{-qs} - q e^{-qs}}$$

$$= \frac{q e^{-qs}}{1 - p e^{-qs} + q e^{-qs} - q e^{-qs}}$$

$$= \frac{q e^{-qs}}{1 - p e^{-qs} + q e^{-qs} - q e^{-qs}}$$



Notice that  $\sum_{j=0}^n p_j = 1$ . Hence,

$$\lim_{n \rightarrow \infty} \sum_{j=0}^n p_j = 1$$

and, consequently,

$$\lim_{n \rightarrow \infty} \sum_{j=0}^n p_j = \lim_{n \rightarrow \infty} \sum_{j=0}^n p_j = 1$$

$$\varphi_X(s) = \frac{1}{s+1}$$

Taking into account that  $E\{X\} = 1$ , we have finally

that is, has an exponential distribution with parameter  $\lambda = 1$

### 1.4 RELATIONSHIPS AMONG DISTRIBUTIONS

Various distributions have common roots, or are closely related. As we discussed previously, the normal and exponential distributions serve as asymptotic distributions in many practical situations. Below we establish some connections among different distributions that are useful in reliability analysis.

#### 1.4.1 Some Relationships Between Binomial and Normal Distributions

The De Moivre-Laplace theorem shows that, for large  $n$  when  $\min(np, nq) \gg 1$ , the binomial distribution can be approximated by the normal distribution.

**Example 1.1** A sample consists of  $n = 1000$  items. The probability that the item satisfies some specified requirement equals  $\Pr\{\text{success}\} = p = 0.9$ . Find  $\Pr\{880 < \text{number of successes}\}$ .

*Solution.* For the normal d.f. which approximates this binomial distribution, we determine that  $a = np = 900$  and  $a^2 = npq = 90$ , that is,  $a \sim 9.49$ . Thus,

$$\Pr\{880 < X\} = \Phi\left(\frac{880 - 900}{\sqrt{90}}\right)$$

**38 FUNDAMENTALS**

$$= 1 - \Phi(-2.11) = 1 - 0.0175 = 0.9825$$

**Example 1.2** Under the conditions of the previous example, find the number of good items which the producer can guarantee with probability 0.99 among a sample of size  $n = 1000$ .

*Solution.* Using a standard table of the normal distribution, from the equation

$$\Pr(m \geq x) = \Pr\left(\frac{x - 900 + 0.5}{\sqrt{900}} \leq \frac{x - 900.5}{9.49}\right) = 0.01$$

we find

$$\frac{x - 900.5}{9.49} \approx -2.33$$

or  $x = 978.6$ . Thus, the producer can guarantee not less than 978 satisfactory items with the specified level of 99%.

We must remember that such an approximation is accurate for the area which is more or less close to the mean of the binomial distribution. This becomes clear if one notices that the domain of a normal distribution is  $(-\infty, \infty)$ , while the domain of a binomial distribution is restricted to  $[0, n]$ .

In addition, there is an essential difference between discrete and continuous distributions. Thus, we must use the so-called "correction of continuity":

$$\Pr\{X \geq j\} = \Pr\left\{X \geq j - \frac{1}{2}\right\}$$

**1.4.2 Some Relationships Between Poisson and Binomial Distributions**

By the Poisson theorem, a Poisson distribution is a good approximation for a binomial distribution when  $p$  (or  $q$ ) is very small.

**Example 1.3** A sample consists of  $n = 100$  items. The probability that an item is defective is equal to  $p = 0.005$ . Find the probability that there is exactly one defective item in the sample.

*Solution.* Compute  $\lambda = 100 \cdot (0.005) = 0.5$ . From a standard table of the Poisson distribution, we find  $P(1; 0.5) = 0.3033$ . The computation with the use of a binomial distribution gives

$$P_h(1, 0.005, 100) = \binom{100}{1} (0.005)^1 (0.995)^{99} \approx 0.5e^{-0.5} \approx (0.5) \cdot (0.6065) = 0.3033$$

### 1.4.3 Some Relationships Between Erlang and Normal Distributions

The normal approximation can be used for the Erlang distribution when  $k$  is large, for instance, when  $k$  is more than 20. This statement follows from the "Lindeberg form" of the central limit theorem.

Let  $Y$  be an r.v. with an Erlang distribution of the  $k$  th order. In other words,  $Y = X_1 + X_2 + \dots + X_k$  where all AT/s are i.i.d. r.v.'s with an exponential distribution and parameter  $\lambda$ . Then, if  $k \gg 1$ ,  $Y$  approximately has a normal distribution with mean  $\mu = k/\lambda$  and standard deviation  $\sigma = \sqrt{k}/\lambda$ .

**Example 1.4** Consider a socket with 25 units which replace each other after a failure. Each unit's TTF has an exponential distribution with parameter  $\lambda = 0.01$  [1/hour], Find the probability of a failure-free operation of the socket during 2600 hours. (Replacements do not interrupt the system operation.)

*Solution.* The random time to failure of the socket approximately has a normal d.f. with parameters  $\mu = 25 \cdot 100 = 2500$  hours and  $\sigma = \sqrt{2500} = 50$  hours. The probability of interest is

$$\Phi\left(\frac{2600 - 2500}{50}\right) = \Phi(2) = 0.9773$$

### 1.4.4 Some Relationships Between Erlang and Poisson Distributions

Consider the two following events:

- (a) We observe a Poisson process with parameter  $\lambda$ . The probability that during the interval  $[0, t]$  we observe  $k$  events of this process is

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

- (b) We observe an r.v.  $\xi_k$  with an Erlang distribution of order  $k$  with parameter  $\lambda$ . Consider the event that  $\xi$  is smaller than  $f$  and, at the same time,  $\xi + 1$  is larger than  $t$ . The probability of the latter event equals

$$\begin{aligned} & \int_0^f \lambda^k x^{k-1} e^{-\lambda x} e^{-\lambda(f-x)} dx \\ & = \lambda^k \int_0^f x^{k-1} e^{-\lambda f} dx \\ & = \lambda^k e^{-\lambda f} \frac{f^k}{k} \\ & = \frac{\lambda^k f^k}{k} e^{-\lambda f} \end{aligned}$$

Thus, events (a) and (b) are equivalent. It is important to remark that both the Erlang r.v. and the Poisson process are formed with i.i.d. r.v.'s with exponential d.f.'s.

Notice that the unconditional event  $\mathcal{E}_k > t$  is equivalent to the set of the following events in the Poisson process: {no events are observed} or {one event is observed} or {two events are observed} or... or {k - 1 events are observed}. This leads to the following condition:

$$\Pr \{ \mathcal{E} > * \} = \sum_{n=0}^{k-1} \frac{(A t)^n}{n!} e^{-A t} = \Pr \{ X \leq k-1 \} = \Pr \{ X \leq k-1 \} = \Pr \{ X \leq k-1 \}$$

or, for the probability of the event  $\mathcal{E}_k < t$ , that is, for the d.f. of the Erlang r.v. of the kth order, we have

Therefore, in some sense, the Poisson d.f. is a cumulative function for an r.v. with an Erlang distribution.

**1.4.5 Some Relationships Between Poisson and Normal Distributions**

Note that a high-ordered Erlang r.v. can be approximated by a normal r.v. and, at the same time, it has a Poisson distribution as its cumulative distribution. This fact can be used as a heuristic justification for the possibility of approximating a Poisson distribution with the help of a normal distribution. The strict proof of this statement can be obtained with the help of a Gram-Charlie set (see below).

Here we take without proof that a Poisson distribution can be approximated by a normal distribution. For a Poisson d.f. with a large mean a, the approximation can be written as

$$\Pr \{ m < x \} \approx \Phi \left( \frac{x - a - 0.5}{\sqrt{a}} \right)$$

Notice that this approximation is accurate in an area close to the mean and may be very bad for the "tails" of the Poisson distribution. This is explained by the fact that these two distributions have different domains: the normal distribution is defined on  $(-\infty, \infty)$ , while the Poisson d.f. has no meanings for  $m < 0$ .

Example 1.5 Assume that the number of failures of some particular unit of equipment has a Poisson distribution. The expected number of failures

## 42 FUNDAMENTALS

during a specified period of time equals 90. One has decided to supply the

equipment with 100 spare units of this type. With what probability will there be no deficit of spare units?

*Solution.*

$$Pr\{H \leq 100\} = \frac{100 - 90}{90} = 1.111$$

From a standard table of the normal distribution, we find that this probability equals 0.853.

Example 1.6 Under the conditions of the previous example, find how many spare units should be supplied so that the probability exceeds 0.995.

*Solution.*

$$Pr\{m > x\} \sim \frac{100 - 90 + 0.5}{90} = 1.111$$

From a standard table of the normal distribution, we find that

$$z = \frac{100 - 90 + 0.5}{\sqrt{90}} = 2.576$$

or  $x = 114.9$ . This means that one should have 115 spare units.

### 1.4.6 Some Relationships Between Geometric and Exponential Distributions

It is clear that an exponential distribution is an approximation to a geometric distribution with  $q = A \Delta t$

$$\lim_{\Delta t \rightarrow 0} [(1 - A \Delta t)^{1/\Delta t}] = e^{-A}$$

where  $A \Delta t < 1$ .

### 1.4.7 Some Relationships Between Negative Binomial and Binomial Distributions

The relationship between these distributions is similar to the relationship between the Erlang and Poisson distributions. Consider a sequence of Bernoulli trials that forms a negative binomially distributed r.v.  $v_k$  consisting of the sum of  $k$  geometrically distributed r.v.'s. Let us pay attention to the

first  $n$  trials where  $n > k$ . The event  $\{v_k > n\}$  means that, in the first  $n$  trials, there are 0, or 1, or 2, ... or  $k - 1$  failures, that is,

$$\Pr \{K > n\} = \sum_{j=0}^{k-1} \binom{n}{j} p^j (1-p)^{n-j}$$

**1.4.8 Some Relationships Between Negative Binomial and Erlang Distributions**

We noticed that the geometric distribution is related to the exponential distribution. In the same sense, the convolution of geometric distributions is related to the convolution of exponential distributions. No other comments are needed: the negative binomial and Erlang distributions are these convolutions.

**1.4.9 Approximation with the Gram-Charlie Distribution**

Because of the wide applications of the normal distribution, many attempts were made to use various compositions of this distribution to express other distributions. Below is one of them.

Let  $f(x)$  be the density function of a distribution other than the normal distribution. The mean  $a$  and the variance  $\sigma^2$  of this distribution are known. Introduce a new variable

$$t = \frac{x - a}{\sigma} \tag{1.74}$$

The density function  $f(x)$  can be represented with the help of the Gram-Charlie series

$$f(x) \sim A_0 \phi(t) + A_1 \phi'(t) + A_2 \phi''(t) + \dots$$

where  $\phi(t)$ ,  $\phi'(t)$ ,  $\phi''(t)$ , ... are the density of the normal distribution and its subsequent derivatives. The standard normal density is expressed as

Introduce the Chebyshev-Hermit polynomials:

$$H_n(t) = (-1)^n e^{t^2} \frac{d^n}{dt^n} e^{-t^2} \tag{1.75}$$



where  $c p^{(n)}(t)$  is the  $n$ th derivative of the normal density. By direct calculations we find

$$\begin{aligned}
 H_0(t) &= 1 \\
 H_1(t) &= 2t \\
 H_2(t) &= 4t^2 - 2 \\
 H_3(t) &= 8t^3 - 12t \\
 H_4(t) &= 16t^4 - 48t^2 + 12
 \end{aligned} \tag{1.76}$$

Usually, for practical problems, we do not need more than four terms of the Gram-Charlier set. From (1.75) it follows that

These functions go to 0 for all  $n$  when  $t \rightarrow \pm\infty$ . The functions  $H_0, H_2,$  and  $H_4$  are even, and the functions  $H_1, H_3,$  and  $H_5$  are odd, so from (1.77) it follows that

$$\begin{aligned}
 \int_{-\infty}^{\infty} p(t) H_0(t) dt &= \int_{-\infty}^{\infty} p(t) dt = 1 \\
 \int_{-\infty}^{\infty} p(t) H_1(t) dt &= 0 \\
 \int_{-\infty}^{\infty} p(t) H_2(t) dt &= -2 \\
 \int_{-\infty}^{\infty} p(t) H_3(t) dt &= 0 \\
 \int_{-\infty}^{\infty} p(t) H_4(t) dt &= 12
 \end{aligned}$$

The Chebyshev-Hermit polynomials are orthogonal, that is,

This fact can be proven by direct calculation. Now substitute the Chebyshev-Hermit polynomials into (1.75)

$$f(t) - A_0 H_0(t) - A_1 H_1(t) - A_2 H_2(t) - \dots \sim 0 \tag{1.78}$$

To find  $A_n$ , multiply both sides of (1.78) by  $H_n(t)$  and integrate from  $-\infty$  to  $\infty$ . Because of the above-mentioned orthogonality property of the Chebyshev-Hermit polynomials, we have

$$A_n = \frac{(-1)^n}{\int_{-\infty}^{\infty} p(t) H_n^2(t) dt} \int_{-\infty}^{\infty} f(t) H_n(t) p(t) dt \tag{1.79}$$

After substituting (1.76) into (1.79), we obtain

$$\begin{aligned}
 A_0 &= 1 \\
 A_1 &= -m \\
 A_2 &= -\sigma^2 \\
 A_3 &= -i[m_3 - 3m\sigma^2] \\
 &= \dots + 3]
 \end{aligned}
 \tag{1.80}$$

where  $m_i$  is the central moment on the  $i$ th order of the r.v.  $t$ . Thus,  $m_0 = 1$ ,  $m_1 = m$ , and, consequently, all initial moments are equal to centered moments. Then from (1.80)

$$\begin{aligned}
 A_1 &= -m \\
 A_2 &= -\sigma^2 \\
 A_3 &= -i[m_3 - 3m\sigma^2] = -i k_3 \sigma^3 \\
 A_4 &= \dots = i k_4 \sigma^4
 \end{aligned}$$

where  $k_3$  and  $k_4$  are known as the *coefficient of asymmetry* and the *coefficient of excess*, respectively,

$$\begin{aligned}
 k_3 &= \frac{m_3 - 3m\sigma^2}{\sigma^3} \\
 k_4 &= \frac{m_4 - 4m\sigma^3 - 3m^2\sigma^2}{\sigma^4}
 \end{aligned}$$

$k_3$  defines the deviation of the density function under consideration from a symmetrical function, and  $k_4$  defines the sharpness of the mode of the density function. All symmetric densities have  $k_3 = 0$ , and a normal density has  $k_4 = 0$ .

Finally, we obtain

or, after integration from  $-\infty$  to  $t$ ,

$$F(t) \sim P(t) - ik_3 \sigma^3 p'(t) + \frac{1}{2} k_4 \sigma^4 p''(t)$$

Notice that  $t$  is the linear function of  $x$ . And so,  $f(x)$  and  $F(x)$  can be

expressed as

$$f(x) = \frac{e^{-a} a^x}{x!}$$

and

Example 1.7 With the help of the Gram-Charlie series, the Poisson distribution can be approximately expressed as

$$\Pr\{1 < c\} m < t \rangle (f) - + -24\% <^3 \rangle (l) \quad (1.81)$$

where

$$r = \frac{x - a}{\sqrt{a}}$$

and  $a$  is the parameter of the Poisson distribution.

It is clear that for a large  $a$  one can disregard the last two terms of the right side of (1.81), and, consequently, the Poisson distribution can be approximated by the normal distribution for large  $a$ .

**REMARK.** The Gram-Charlie distribution can be successfully applied to the evaluation of d.f.'s. This takes place, for instance, in analyses of the distribution of a parameter of a piece of electronic equipment when the distributions of its components are known.

## 1.5 STOCHASTIC PROCESSES

Stochastic processes are used for the description of a system's operation over time. There are two main types of stochastic processes: discrete and continuous. Among discrete processes, point processes in reliability theory are widely used to describe the appearance of events in time (e.g., failures, terminations of repair, demand arrivals, etc.).

A well-known type of point process is the so-called *renewal process*. This process is described as a sequence of events, the intervals between which are

- 
- 1 Stationarity
    - Memorylessness (Markov property)
    - Ordinarity

## 48 FUNDAMENTALS

i.i.d. r.v.'s. In reliability theory this kind of mathematical model is used to describe the flow of failures in time.

A generalization of this type of process is the so-called *alternating renewal process* which consists of two types of i.i.d. r.v.'s alternating with each other in turn. This type of process is convenient for the description of *renewal*

*systems*. For such systems, periods of successful operation alternate with periods of idle time.

The more complex process is a process describing a system transition from state to state. The simplest kind of such a process is a *Markov process*. If the times that the process may change states are assumed to be discrete, the process is called the *Markov chain*.

We start with simplest cases and move in the direction of more complex mathematical models.

### 1.5.1 Poisson Process

In the theory of stochastic processes, the Poisson process plays a special role, comparable to the role of the normal distribution in probability theory. Many real physical situations can be successfully described with the help of a Poisson process. A classical example of an application of the Poisson process is the decay of uranium: radioactive particles from a nuclear material strike a certain target in accordance with a Poisson process of some fixed intensity.

In practice, the Poisson process is frequently used to describe the flow of failures of electronic equipment. In inventory control, the flow of random requests for replacement of failed units is also often assumed to be described by a Poisson process, especially if the system which generates these requests is large.

Sometimes the Poisson process is called "a process of rare events." Of course, the meaning of the word "rare" should be carefully defined in each particular case. Usually, we speak about rare events if they appear with a frequency which is lower than the frequencies of other accompanying processes. The Poisson process appears as the interaction of a large number of these processes and, consequently, has a frequency lower than the other processes.

In reliability, such "rare" events appear, for instance, when one considers a highly reliable *renewal redundant system* or a multicomponent *renewal series system*. This process also successfully describes the fluctuation over a high-level threshold.

This process is so named because the number of events in any fixed interval of length  $t$  has a Poisson distribution:

$$\Pr\{k \text{ events during } t\} = p_k(t) = \frac{(A t)^k}{k!} e^{-A t}$$

where  $A$  is called the parameter of the Poisson process.

First of all, note that the Poisson process possesses the three following properties that are often referred to as *characterization properties*:

The first property means that the d.f. of the number of observed events in a time interval depends only on the length of the interval and not on its position on the time axis.

The second property means that the d.f. of the number of observed events does not depend on the previous history of the process.

The third property means that the probability of an appearance of more than one event in an infinitesimally small interval  $h$  goes to 0:

$$\lim_{h \rightarrow 0} \Pr\{k \text{ events appear during } h, k > 1\} = 0$$

or, in another notation,

$$\Pr\{k \text{ events appear during } h, k > 1\} = o(h) \quad (1.82)$$

In practical problems, these properties are often assumed. These properties, which seem to be—at a first glance—purely qualitative, allow us to obtain strict mathematical results.

First, for a better understanding, we present a semiintuitive proof of the fact that these properties generate a Poisson process. Consider a Bernoulli process with probability of success  $p$  and a sufficiently large number of trials  $n$ . The Bernoulli process satisfies the first two properties (and trivially satisfies the third one because of its discrete nature). As we considered in Section 1.1, the number of successes in a series of  $n$  Bernoulli trials has a binomial distribution. As we have shown in Section 1.1, for large  $n$  the binomial distribution can be successfully approximated by a Poisson distribution.

We now return to the exact mathematical terms. First, add one extra property to the above three properties, namely, assume that the probability that there is exactly one event in a time interval  $h$ :

$$P_1(h) = Ah + o(h) \quad (1.83)$$

where  $A$  is some constant and  $o(h)$  was introduced in (1.82). As a matter of fact, (1.83) follows from the three properties characterizing a Poisson process.

Consider the probability of the appearance of  $k$  events in a time interval  $t + h$ . The formula for the probability can be easily written as

$$P_k(t + h) = \sum_{j=0}^k P_j(t) P_{k-j}(h) \quad (1.84)$$

Let

$$= E P_{k-j}(h) \quad (1.85)$$

$$0 \leq z_j < k-2$$

Obviously, 
$$\sum_{j=0}^{k-2} P_{k-j}(h) < L P_k(h) \quad 2 \lambda h + o(h) \tag{1.86}$$

because all  $P_s(t) < 1$ . We only reinforce the inequality (1.86) by changing the limits of summation

$$P_k < L W) = \Pr\{\text{two or more events appear during interval } h\} \tag{1.87}$$

At the same time, by assumption, this probability equals  $o(h)$ . As a result, we have the equality

$$P_k(t+h) = P_k(t)P_0(h) + P_{k-1}(t)P_1(h) + o(h) \tag{1.88}$$

In this equality, we can substitute  $P_1(h) = \lambda h + o(h)$ . Also,  $P_0(h) + P_1(h) + o(h) = 1$ , that is,  $P_0(h) = 1 - \lambda h + o(h)$ . Now (1.88) can be rewritten as

$$P_k(t+h) = P_k(t)(1 - \lambda h) + P_{k-1}(t)\lambda h + o(h) \tag{1.89}$$

and from (1.89) we obtain

and, after  $h \rightarrow 0$ ,

$$\frac{dP_k(t)}{dt} = -\lambda P_k(t) + \lambda P_{k-1}(t) \tag{1.90}$$

Thus, a system of equalities for  $P_k(t)$ ,  $k = 0, 1, \dots$ , has been obtained. We need to add one more equation to determine  $P_0(t)$ . Using the memoryless property, we can write

$$P_0(t+h) = P_0(t)P_0(h) = P_0(t)[1 - \lambda h + o(h)]$$

or, finally,

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) \tag{1.91}$$

To solve the system, we must determine the initial condition. Of course, at  $t = 0$ , the probability of no events equals 1; that is, the initial condition is  $P_0(0) = 1$ .



The system of differential equations (1.90) and (1.91) with the above initial condition can be solved by several different methods. We solve this system of equations with the use of the LST. Let  $\langle p_0(s) \rangle$  be the LST of the function  $P_0(t)$ :

$$\langle p_0(s) \rangle = \int_0^\infty p_0(t) e^{-st} dt \tag{1.92}$$

Applying (1.92) to (1.91) and keeping in mind the properties of the LST, one obtains

$$- \langle p_0(s) \rangle + s \langle p_0(s) \rangle = - \langle p_0(s) \rangle \tag{1.93}$$

which has the solution

$$\langle p_0(s) \rangle = \frac{\lambda}{\lambda + s} \tag{1.94}$$

As it follows from a table of Laplace-Stieltjes transforms, the function  $P_0(t)$  corresponding to (1.94) is exponential with parameter  $\lambda$ :

$$P_0(t) = e^{-\lambda t} \tag{1.95}$$

For arbitrary  $k > 0$ , from (1.90) the system of recurrent equations follows:

$$s \langle p_k(s) \rangle = -\lambda \langle p_k(s) \rangle + \lambda \langle p_{k-1}(s) \rangle \tag{1.96}$$

or

Finally, using (1.94) systematically, we have

$$\langle p_k(s) \rangle = \left( \frac{\lambda}{\lambda + s} \right)^k \tag{1.98}$$

From a table of LSTs, the latter transformation corresponds to a Poisson distribution

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{1.99}$$

For the Poisson distribution the mean number of events in a fixed interval of time is proportional to its length. The parameter  $\lambda$  is the mean number of events in a time unit, or, equivalently, it equals the inverse of the mean time

between events. Also, as known (see the Appendix), a convolution of Poisson distributions produces a Poisson distribution. Thus, for several disjoint intervals of lengths  $t_1, t_2, \dots, t_m$ , the distribution of the total number of events is Poisson with parameter  $\lambda(t_1 + t_2 + \dots + t_m)$ . In other words, the Poisson process is a point stochastic process with exponentially distributed intervals between neighboring events.

### 1.5.2 Introduction to Recurrent Point Processes

We often encounter situations where some events occur sequentially in such a way that the times between occurrence (interarrival times) can be successfully described by a sequence of independent r.v.'s. For instance, consider a socket with an installed unit which is instantly replaced upon failure by a new unit; the times between replacement moments form such sequence. In general, the length of each interval might depend on the number of the event because of a changing environment, a wearing out of the socket, and so on. Here we ignore such phenomena. A process of this type is called a *point process with restricted memory*.

A point process with restricted memory is a sequence of r.v.'s. It is called a *renewal (recurrent) point process* if all interarrival intervals are i.i.d. r.v.'s with identical d.f.'s  $F_k(t) = F(t)$ ,  $k \geq 2$ , with only the first interval having its own distribution  $F_1(t)$ .

The Poisson process represents a particular case of such a process in that the intervals between arrivals are independent and exponentially distributed.

We assume that a flow of failures is represented by a recurrent point process. This assumption is acceptable in many practical situations. At the same time, it allows us to obtain simple and understandable results.

For a renewal point process, there are two main characteristics: (1) the process intensity defined to be the mean number of process events arriving in a time unit and (2) the process parameter defined to be the limit probability of the arrival of at least one event.

Let  $N(t)$  be the number of events arriving during an interval of length  $t$ . Then, for the stationary process,

$$\lambda = \lim_{T \rightarrow \infty} \frac{E\{N(T)\}}{T} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \lambda(t) dt$$

The parameter of the process is defined as

$$\lambda = \lim_{T \rightarrow \infty} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P_j(t)$$

For an arbitrary stationary point process with a single arrival at a time and without so-called "points of condensation" (infinitesimally small intervals in

which an infinite number of discrete events might appear), we have

$$A^* < A$$

For a stationary and memoryless point process, the parameter coincides with the intensity. We can give an explanation of the parameter of a point process based on a more physical consideration:

$$A^*(t) = \Pr\{\text{at least one failure occurs in } [t, t + A]\}$$

Let  $f^{*k}(t)$  stand for a convolution of the  $k$ th order of the function  $f(t)$ :

$$f^{*k}(t) = \int_0^t f^{*(k-1)}(x) f(t-x) dx$$

it is clear that at least one failure might occur if

- The first failure occurs with probability  $f(t)A$ .
- The second failure occurs with probability  $f^{*2}(t)A, \dots,$
- The  $k$ th failure occurs with probability  $f^{*k}(t)A$  and so on.

Thus, the probability that a failure will occur for any of these reasons is

$$\Pr\{\text{at least one failure occurs in the interval } [t, t + A]\} = A^*(t)A =$$

where we use the conditional notation  $f^{*0}(t) = 1$ . Hence,

$$A^*(t) = \sum_{k=0}^{\infty} f^{*k}(t)A$$

The function  $A^*(t)$  allows us to express the so-called *characterization point process function* which we denote by  $A^*(f)$ :

$$A^*(M + t) = \int_t^{M+t} A^*(t) dt$$

Using this function, we can write

$$\Pr\{\text{no failures in } [t, t + t^*]\} = \exp$$

$$= -A^*(t + t^*) = -[A^*(t + t^*) - A^*(t)] =$$

(1.100)

The function  $A(t)$  is defined to be the "instant" conditional density of the failure distribution  $F(t)$ . We emphasize that the functions  $A(t)$  and  $A^*(t)$  are quite different.

Now consider the main characteristics of a renewal process. One of the characteristics of a renewal process is the mean number of events occurring up to a moment  $t$ . Denote the random number of events by  $N(t)$  and the mean number by  $H(t) = E\{N(t)\}$ .  $H(t)$  is called the *renewal function*. The derivative  $h(t) = H'(t)$  is called the *renewal density*. Consider a renewal process composed of i.i.d. r.v.'s with distribution  $F(t)$ . We can write

$$\Pr\{N(t) \leq k\} = \Pr\{N(t) \leq k\} = \Pr\{t \leq A = F^{**k}(t)\}$$

where  $F^{**k}(t)$  is the  $k$ th-order convolution of  $F(t)$ :

$$F^{**k}(t) = \int_0^t F^{**k-1}(t-x) dF(x)$$

$$F^{**1} = F(t)$$

The expression can be easily written as

$$\begin{aligned} \Pr\{\text{any event occurs in interval } [t, t + dt]\} \\ = E \Pr\{\text{the } k\text{th event occurs in interval } [t, t + dt]\} \\ = h(t) dt = \int_0^t f^{**k}(t) dt \quad (1.101) \end{aligned}$$

Integrating (1.101) allows us to write an expression for  $H(t)$ :

$$H(t) = F(t) + \int_0^t F^{**k}(t) dt \quad (1.102)$$

Of course,  $H(t)$  can be found in a standard way as the mean number of events during time  $t$ . The probability that exactly  $k$  events happen up to moment  $t$  is expressed as

$$\begin{aligned} \Pr\{N(t) = k\} &= \Pr\{N(t) \leq k\} - \Pr\{N(t) \leq k-1\} \\ &= F^{**k}(t) - F^{**k-1}(t) \end{aligned}$$

Thus, the distribution of  $N(t)$  is defined.  $H(t)$  can be found by

$$\begin{aligned}
 H(t) &= E\{N(t)\} = E \sum_{f \in \mathcal{C}} \Pr\{N(r)=f\} \\
 &= E \sum_{f \in \mathcal{C}} [f^* \cdot \mathbf{0} \\
 &= \mathbf{F}(\mathbf{r}) + E \sum_{k \geq 1} k F^{*k}(t) - E \sum_{k \geq 1} (k-1) F^{*k}(t) \\
 &\quad - E \sum_{k \geq 1} F^{*k}(\mathbf{0})
 \end{aligned}$$

For  $h(t)$  we can write

$\Pr\{\text{any event occurs in interval } [t, t + \Delta t]\}$

$$\begin{aligned}
 &= \Pr\{\text{the first event occurs in interval } [t, t + \Delta t]\} \\
 &\quad + \Pr\{\text{the last event happens in interval } [t, t + \Delta t]\} \tag{1.103}
 \end{aligned}$$

and the following random time  $\xi$  is such that  $x < \xi < x + \Delta x$

With  $\Delta t \rightarrow 0$ , (1.103) can be rewritten in differential form as

$$\lambda(r) - \lambda(r) + f' h(t-x) dF(x) \tag{1.104}$$

Naturally, the renewal density function at time  $t$  is the sum of the densities of the occurrence of all possible events of the renewal process: the first, or the second, or ..., or the  $k$ th and so on. From (1.104), by integration, we obtain

$$H(t) = F(t) + \int_0^t h(t-x) dF(x) \tag{1.105}$$

It is important to note that  $F^{*n} \geq [\mathbf{F}(\mathbf{r})\mathbf{F}]^n$ . Indeed,

$$F^{*n}(t) \geq \int_0^t \mathbf{F}(\mathbf{r})^n dF(x)$$

$$\sum_{i=1}^n \Pr\{x_i \leq t\} \geq \Pr\{\sum_{i=1}^n x_i \leq t\}$$

This states the simple fact that the sum of  $n$  nonnegative values is not less than the maximal one. (Equality occurs only if at least  $n - 1$  values are equal

to 0.) From this fact it follows that

$$0 - L \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} \dots \int_0^{\infty} [n(t)]^k - m L \int_0^{\infty} F(t) dt < \infty$$

Using (1.102) and observing that the integral on the right side of the equation is positive, we obtain two-sided bounds

$$F(t) < H(t) < y^{-\lambda - r y} \tag{1-106}$$

The next interesting bounds for a renewal process, built with "aging" r.v.'s  $i$ , can be obtained if we consider the following natural condition. Let  $N(t)$  events be observed up to a moment  $t$ . Thus,

$$E[N(t)] < \frac{L}{1 - sM(t)}$$

Using the Wald equivalency, we write

$$f < E U \{ H(t) + 1 \}$$

which produces

$$t^{-1} \ln r^1$$

For an aging r.v, the residual time  $f(t)$  is decreasing, which allows us to write

$$m H(t) <$$

$E\{U\}$

Thus, for a renewal process with aging r.v.'s we can write the two-sided bounds

$$-T - r - 1 < H(t) < \frac{t}{\mu} \quad (1.107)$$

in practical reliability problems we are often interested in the behavior of a renewal process in a stationary regime, that is, when  $t \rightarrow \infty$ . This interest is understandable because repairable systems enter an "almost stationary" regime very quickly (see Section 6.1). Several important facts are established for this case.

**Theorem 1.2** For any  $F(t)$ ,

$$H(t) \sim \frac{1}{\lambda} \quad (1.108)$$

In a mathematical sense this theorem is close to the Wald theorem. In a physical sense it means that, for a large interval of size  $t$ , the mean number of events is inversely proportional to the mean interarrival time.

**Theorem 1.3** If  $\lambda$  is continuous, then

$$\lim_{t \rightarrow \infty} h(t) = \lambda \quad (1.109)$$

This theorem reflects the fact that with increasing  $t$  the renewal becomes stationary and its characteristics become independent of the current time.

**Theorem 1.4 (Blackwell's Theorem)** For a continuous r.v.  $\lambda$  and an arbitrary number  $r$ ,

$$\lim_{t \rightarrow \infty} [H(t+r) - H(t)] = r \quad (1.110)$$

It is clear that this theorem is a simple generalization of the first one.

**Theorem 1.5 (Smith's Theorem)** If  $\lambda$  is a continuous r.v. and  $V(t)$  is a monotone nonincreasing function, integrable on  $(0, \infty)$ , then

$$\lim_{t \rightarrow \infty} \int_0^t V(t-x) dH(x) = \frac{1}{\lambda} \int_0^{\infty} V(x) dx \quad (1.111)$$

The function  $V(t)$  can be chosen arbitrarily between those which have a probabilistic nature. The choice of this function depends on the concrete applied problem. An interpretation of this theorem is provided in the following particular case.

**Corollary 1.1** The stationary probability of a successful operation (the stationary interval availability coefficient) equals

Here  $t_0$  is the time needed for a successful operation. The proof of this is left to Exercise 1.10.



### 1.5.3 Thinning of a Point Process

We often encounter where a unit failure leads to a system failure only if several additional random circumstances happen. For instance, in a system of a group of redundant units, a unit failure is the cause of a system failure if, at a particular moment, all of the remaining units have failed. Such a coincidence of random circumstances may be very rare. We may consider the flow of "possibilities" which generate a relatively rare flow of system failures. This procedure is called a *thinning procedure*.

**Poisson Process** The thinning of a Poisson process produces a Poisson process. To prove this fact, we consider the sum of a geometrically distributed random number of exponentially distributed r.v.'s. Indeed, thinning means that with some probability  $q$  an event remains in the final process and with probability  $p = 1 - q$  it is removed from it. Thus, we have a sequence of Bernoulli trials.

Of course, in the particular case of the Poisson process, we can apply a simple deduction based on its three characteristic properties. Indeed, stationarity is not violated by the Bernoulli-like exclusion of events from the initial process: all  $p$ 's are constant over the entire time axis. Ordinarity is also preserved because we only exclude events. The memorylessness property of the resulting process follows from the independent character of the event exclusion from the Bernoulli trial sequence.

**General Case** Consider a stationary recurrent point process for which the intervals between events have a distribution  $Fit$ ). Sometimes this distribution is called a "forming distribution." Apply the thinning procedure to this process. According to this procedure, each event remains in the process with probability  $q$  or is deleted with probability  $p = 1 - q$ . Thus, after such a procedure, the average number of points which remain in the newly formed process is  $1/q$  times less than in the initial process. In other words, the time interval between points in the new process is  $1/q$  times larger. The explanation of the procedure is depicted in Figure 1.9.

Each interval between events represents the sum of a random number of r.v.'s. Thus, the problem of renewal process thinning is equivalent to the

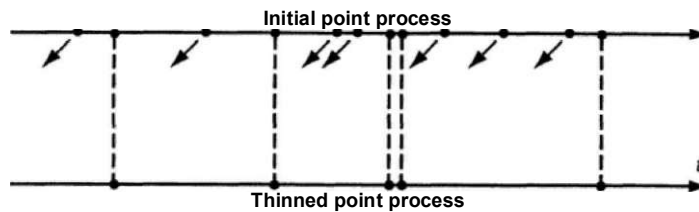


Figure 1.9. Example of the thinning procedure for a point process.

summation of a geometrically distributed random number of r.v.'s. (This was considered in Section 1.4.) In particular, in Section 1.4.5 we developed some asymptotic results. Here we use the standard terminology and methods of renewal theory, because this helps us to obtain some additional results.

Consider a special transformation of the renewal process,  $T_q$ : events are deleted from the process with probability  $p$ , and, simultaneously, the time scale is shrinking by a factor of  $1/q$ . This normalization of time keeps the length of the average interarrival interval the same as in the initial process. It is clear that the  $T_q$  transformation is equivalent to the summation of a geometrically distributed random number of r.v.'s with a d.f.  $F(t)$  and the further normalization of the resulting r.v.

Sequential applications of the transformations  $T_{q_1}$  and  $T_{q_2}$  to the process are equivalent to the single transformation  $T_{q_1 q_2}$ . We ask the reader to prove this is Exercises 1.11 and 1.12.

**Limit Renyi Theorem** The Renyi theorem is very important in many applications. These asymptotic results can be used if the thinning procedure is intensive enough. They are also very useful in developing heuristic approaches (see Chapter 13).

**Theorem 1.6** If transformations  $T_{q_1}, T_{q_2}, \dots$ , are such that, for  $n \rightarrow \infty$ ,

$$q_n = Q \ln 2 \cdot \ln n \rightarrow 0 \text{ as } n \rightarrow \infty$$

then their application to some point renewal process with an initial finite intensity  $A$  leads the resulting limit process to a Poisson process with the same intensity  $A$ .

We omit the proof because, in general, it coincides with the corresponding proof of Section 1.3.5.

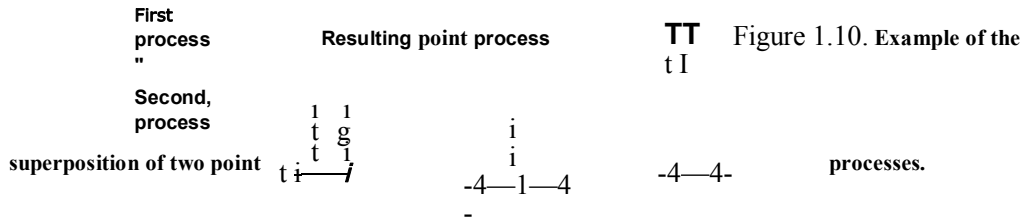
Later this result was generalized for the superposition of  $n$  different renewal processes with different thinning procedures.

We remark that a Poisson process is sometimes called "a process of rare events." From formulations of the above results, one can see that the  $T_q$  transformation generates the flow of "rare" events from a dense initial process.

### 1.5.4 Superposition of Point Processes

In reliability practice we frequently encounter a situation which might be described as the formation of a common point process from the superposition of several point processes (see Figure 1.10).

For example, consider the flow of failures of different units in a series system. Each unit generates its own renewal point process of failures: a failed



unit is replaced and the process continues. Unfortunately, even if we consider a small number of renewal processes, their superposition cannot be analyzed in terms of renewal processes! (The only exception is the superposition of Poisson processes.)

At the same time, fortunately, if the number of superimposed point processes is very large, the superposition of these processes produces a point process that is very close to being Poisson. In the theory of stochastic processes, the Poisson process plays a role which is analogous to that of the normal distribution in probability theory.

**Poisson Process** For the superposition of  $n$  Poisson processes, the resulting process is Poissonian. If the initial processes have parameters  $A_1, A_2, \dots, A_n$ , the resulting process has the parameter  $A = \sum_{k=1}^n A_k$ . To show this fact, consider an arbitrary moment of time  $t$ . Let  $\xi_k$  denote the residual time of the  $k$ th process, that is, the time from an arbitrary but fixed  $t$  until the appearance of the next event in the process. The memoryless property says that two r.v.'s  $\xi_k$  and  $\xi_k$ , which represent the time between events for the  $A$ th process, are statistically equivalent. Thus, for the  $k$ th initial process we can write the distribution of the residual time

$$\Pr\{\xi_k > r\} = \Pr\{e^{-Ar}\} = \exp(-Ar).$$

If we consider  $n$  processes, then, from a fixed (albeit arbitrary) moment  $t$  until the next arriving event, we observe an r.v.

$$U = \min_{1 \leq k \leq n} \xi_k$$

with d.f.

$$\Pr\{U > t\} = \Pr\{\min_{1 \leq k \leq n} \xi_k > t\} = \prod_{k=1}^n \Pr\{\xi_k > t\} = \prod_{k=1}^n \exp(-t A_k) = \exp(-t \sum_{k=1}^n A_k) = \exp(-t A)$$

Thus, the distribution of the time interval between neighboring events is exponential. As we know, only the Poisson process is characterized by such a property.

Of course, as above, we can prove this fact by checking that all three characteristic properties of the Poisson process are satisfied. Indeed, stationarity is kept because of the stationarity of all initial processes. Ordinarily is also preserved, because, for a continuous process, the probability of a coincidence of events equals 0. The memorylessness property of the resulting process follows from the independence of all the initial processes and their original memorylessness property.

**General Case** The proofs and even the formulation of the strict conditions of the related theorems are complex and lie outside the scope of this book. We only formulate the main results, sometimes even on a verbal level. The first strict result was formulated in the following theorem.

**Khinchine- Ososkov Theorem** If the limit

$$\lim_{n \rightarrow \infty} \Lambda_{nr} = \Lambda$$

exists, a necessary and sufficient condition that the process  $J_{r, \{r\}}$  converges to a Poisson process with parameter  $\Lambda$  is that, for any fixed  $t$  and  $n \rightarrow \infty$

$$\sum_{1 \leq r \leq k_n} \lambda_{nr} \int_0^t \varphi_{nr}(0, x) dx \Rightarrow \Lambda t$$

Later general results, relating to the superposition of stochastic point processes, are contained in the Grigelionis-Pogozhev theorem. On a qualitative level, the theorem states that a limit point process which is formed by the superposition of independent "infinitesimally rare" point processes converges to a Poisson process. The parameter of this resulting process is expressed as a sum of the parameters of the initial processes.

## 1.6 BIRTH AND DEATH PROCESS

The *birth and death process* is an important branch of Markov processes. We will not give details for the general Markov processes, but we will consider some special models of renewal systems in a later chapter. This approach is also very useful for the analysis of renewal systems.

### 1.6.1 Model Description

The behavior of a number of practical systems can be portrayed with the help of the birth and death process (BDP). Birth and death processes are widely

used for the construction of mathematical models in microbiology, zoology, and demography. They are also used in reliability and queuing theory.

Let us explain the nature of the BDP with three simple examples.

Consider a queueing system with one service unit and an unlimited number of input call sources. Suppose there are  $k$  calls on the line waiting for service at a moment  $t$ . We say that at this moment, the system is in state  $H_k$ . At the moment  $t + \Delta t$ , where  $\Delta t$  is infinitesimally small, the state changes to  $H_{k+1}$  if an additional call arrives during the interval  $\Delta t$ . If during  $\Delta t$  a call service in the system has been completed, at the moment  $t + \Delta t$  the system changes its state to  $H_{k-1}$ . Recall that, for a Markov process, the probability of more than one state change is  $o(\Delta t)$ , which means that  $o(\Delta t) \rightarrow 0$  as  $\Delta t \rightarrow 0$ . From our assumption of an unlimited number of input sources of calls, it follows the line length could be infinite. In other words, a BDP may have an infinite number of states. Suppose there is a specified criterion of system effectiveness; for example, a line of length more than  $m$  is considered to be inadmissible. Then the set of all system states may be divided into two subsets: the "up states"  $H_0, \dots, H_m$  and the "down states"  $H_{m+1}, H_{m+2}, \dots$ .

As another example, consider a parallel system with one main unit and  $m$  identical active redundant units. There is one repair facility. This system may be thought of as a queueing system with a limited number (namely,  $n + 1$ ) of input "call" sources. Indeed, if there are  $k$  units under repair, then only  $n + 1 - k$  of the remaining units may fail. The state  $H_{m+i}$  corresponds to system failure, and only a transition from this state to state  $H_m$  is possible. State  $H_{m+i}$  is called *reflecting*.

As the final example, consider a parallel system with  $n$  main units and  $m$  identical active redundant units. Again, there is one repair facility. It is clear that the mathematical description of this system is very close to that of the above example. In this case there are, in total,  $n + m$  sources of failure. The states  $H_{m+1}, \dots, H_{n+m}$  correspond to system failure states.

The last two examples are of BDPs with a finite number of states. In reliability theory it is sometimes reasonable to consider separately these two very similar cases.

The transition graphs for all three examples are shown in Figure 1.11.

If the system is in state  $H_j$  at a moment  $t$ , there are three possibilities during the next interval  $\Delta t$ :

• The process passes to state  $H_{j+1}$  with probability:

$$\lambda_j \Delta t + o(\Delta t)$$

• The process passes to state  $H_{j-1}$  with probability:

$$\mu_j \Delta t + o(\Delta t)$$

• The process remains in state  $H_j$  with probability:

$$1 - (\lambda_j + \mu_j) \Delta t + o(\Delta t)$$

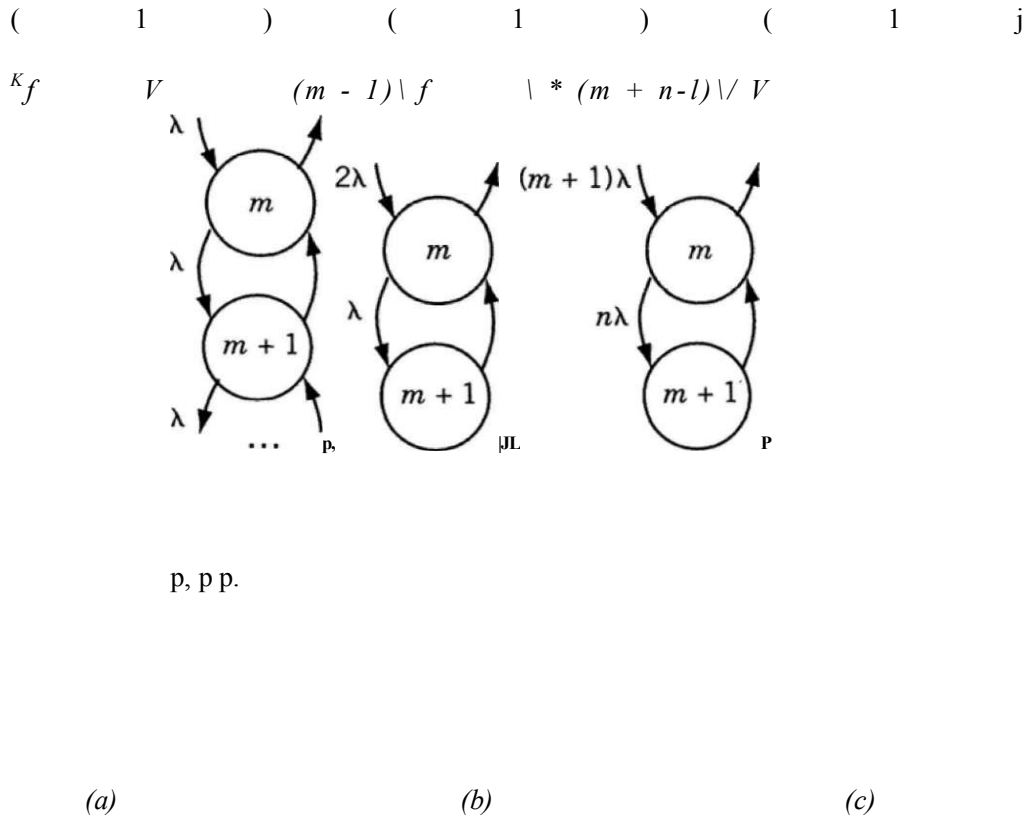


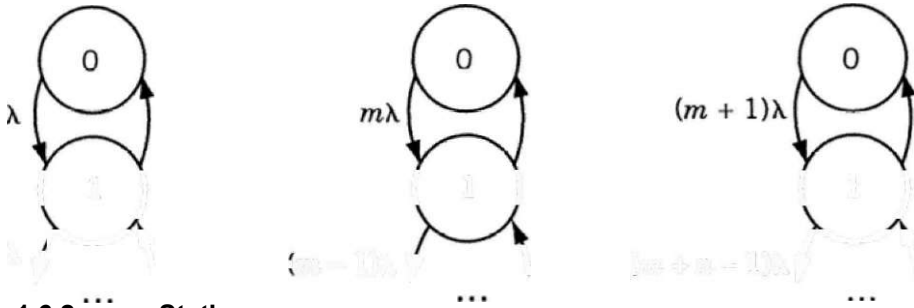
Figure 1.11. Three examples of death and birth processes: (a) a queuing system with an infinite source of demands; (b) a unit with  $m$  active redundant repairable units; (c) a series system of  $n$  units with  $m$  active redundant repairable units.

The failure state with the smallest index, say  $m + 1$ , may be considered absorbing. The other system failure states are of no interest because there are no transitions from them to the set of up states. In this case the process behavior in the subset of up states can be used to find the probability of a failure-free operation, the MTTF, and the mean time between failure (MTBF). Notice that if we are interested in finding repair (idle time) indexes, the state  $H_m$  must be chosen to be absorbing. In this case we consider the behavior of the process in the subset of system failure states.

If there are no absorbing states, the process is considered for finding the availability coefficients, both stationary and nonstationary. For a finite set of states, the state with the largest index is reflecting.

In reliability problems the state  $H_0$  is always considered as reflecting (if the process is not of a special type with states  $H_{-2}, \dots$ ).

For reliability problems it suffices to consider only BDPs with a finite number of states.



**1.6.2 Stationary Probabilities**

Consider a finite BDP with  $N + 1$  states (see Figure 1.11tr). For each state  $k$  and for two infinitesimally close time moments  $t$  and  $(t + \Delta t)$ , we can write

the expression

$$p_k(t + \Delta t) = p_k(t) + \Delta t [A_k - (A_k + M_k)p_k(t) + M_{k+1}p_{k+1}(t) + o(\Delta t)] + o(\Delta t)$$

$$p_k(t + \Delta t) \sim p_k(t) + \Delta t [A_k - (A_k + M_k)p_k(t) + M_{k+1}p_{k+1}(t)] + o(\Delta t)$$

In the limit as  $\Delta t \rightarrow 0$ , we obtain

$$\lim_{\Delta t \rightarrow 0} \frac{p_k(t + \Delta t) - p_k(t)}{\Delta t} = A_k - (A_k + M_k)p_k(t) + M_{k+1}p_{k+1}(t) \quad (1.112)$$

Because we are considering a finite process, we must set  $A_N = A_{N+1} = A_N = 0$ . In other words,

$$A_N = A_{N+1} = 0$$

and

$$A_0 = A_{-1} = 0$$

We add to this system of equations the normalizing equation

$$\sum_{k=0}^N p_k(t) = 1$$

and exclude any one of the above.

We note that (1.112) represents the equation of dynamic equilibrium. In other words, state  $k$  "loses each unit of its mass"  $p_k(t)$  with intensity  $A_k + M_k$  and "receives" a corresponding mass from states  $k - 1$  and  $k + 1$ . If there is no absorbing state, the process has stationary states. Thus, there are limits

$$\lim_{t \rightarrow \infty} p_k(t) = p_k$$

Moreover, if for any  $k$ ,

$$(1.113)$$

the stationary probabilities of these states do not depend on the initial state



at  $t = 0$ . The condition (1.113) means that there are no separated groups of states.

Consider this stationary case. If we take the limit as  $t \rightarrow \infty$  in (1.112), we obtain the system of linear equations

$$\begin{aligned} 0 &= -A_0 p_0 + M_1 p_1 \\ 0 &= A_{k-1} p_{k-1} - (A_k + M_k) p_k + M_{k+1} p_{k+1} \\ 0 &= -A_{N-1} p_{N-1} + M_N p_N. \end{aligned} \tag{1U4}$$

We again must replace one of the above equations with the normalizing condition

$$\mathbf{E} p^* = \mathbf{1} \tag{1-H5}$$

Now let us recall that (1.114) represents an equilibrium. It means that if we consider any cut in the transition graph, for instance, a cut between states  $k-1$  and  $k$ , there is an equality of flows "up" and "down":

$$M_k p_k = A_{k-1} p_{k-1} \tag{1.116}$$

From (1.116) we obtain the recurrent relationship

$$p_k = \frac{A_{k-1}}{M_k} p_{k-1} \tag{1-117}$$

which allows us to obtain

$$p_k = \frac{A_0 A_1 \dots A_{k-1}}{M_1 M_2 \dots M_k} p_0 = A_k p_0 \tag{1.118}$$

From (1.115) it follows that

$$\sum_{k=0}^N p_k = \mathbf{1} \tag{1.119}$$

where  $A_0 = I$ .

### 1.6.3 Stationary Mean Time of Being in a Subset

Consider a BDP whose total set of  $n$  states is divided into two subsets; one subset of up states,  $E_+ = \{1, \dots, H_m\}$ , and another of down states,  $E_- = \{H_m+1, \dots, H_n\}$ .

To find the stationary mean time of the process present in a specified subset of states, we use a well-known result. Let us distinguish two subsets of so-called boundary states:  $e_+$  which is a subset of  $E_+$  and  $e_-$  which is a subset of  $E_-$ . The process may enter the subset leaving only a state belonging to the boundary subset  $e_+$ . The subset  $e_-$  plays an analogous role for the subset  $E_+$ . In considering case  $e_+$  consists of only one state  $H_m$ , Briefly repeat the idea of the conclusion. The process may leave subset  $E_+$  only from state  $H_m$ , and being in this state it leaves with intensity  $A_m^*$ . Hence, the intensity of leaving subset  $E_+$  equals

$$A_+ \ll P^* A_m \tag{1.120}$$

where  $p^*$  is the conditional probability that the process is in state  $H_m$  under the condition that this is in subset  $E_+$ :

$$p_l = \frac{P_m}{\sum_j L_{js} p_j}$$

where  $p_j$  can be found from (1.119). Finally, we find that the mean stationary time of the process being in subset  $E_+$  is  $T_+ = 1/A_+$ :

$$r_+ = \sum_m A_m P_m \tag{1.121}$$

Obviously, the mean stationary time of the process being in subset  $E_-$  is similar to the above except for the following notation:

where  $A_- = A_{m+1}^*$

$$r_- = \sum_{m+1} A_{m+1} P_{m+1}$$

and, finally,

$$r = \frac{E^*}{M_{m+1} p_{m+1}}$$

**1.6.4 Probability of Being in a Given Subset**

The BDP with an absorbing state  $H_{n+1}$  can be described with the help of the following system of differential equations:

$$\frac{d}{dt} p_j(t) = A_j - (A_{j+1} + M_j) p_j(t) + M_{j+1} p_{j+1}(t) \quad 0 \leq j < n$$

$$E_j p_j(t) = 1 \tag{1.123}$$

$$A_{-1} = A_{n+1} = M_0 = A_{n+1} = M_{n+2} = 0$$

where  $p_j(t)$  is the probability of state  $H_j$  at moment  $t$ . Let the probabilities  $P_j(t)$  satisfy the initial conditions:

$$P_j\{^* \sim 0\} = p_j(0) \quad 0 < j < n + 1 \tag{1.124}$$

Let  $\theta_{n+1}$  be the duration of time before the system has reached the absorbing state  $H_{n+1}$  for the first time. We need to find the distribution function  $\Pr\{\theta_{n+1} > t\} = p_{n+1}(t)$ . This is the probability that the system has *not* reached the absorbing state  $H_{n+1}$  at time  $t$ . Let us apply the LST to (1.123). Then we find the system of linear algebraic equations:

$$\begin{aligned} A - (A_{j,j} + M_j + A_{j+1,j})p_j(s) &= -p_j(0) \\ 0 & \leq j < n + 1 \\ A_{y,y} = A_{y,y+1} = M_0 = M_{n+1} = M_{n+2} &= 0 \end{aligned} \tag{1.125}$$

By using Cramer's rule,

$$\tag{1.126}$$

where

72 FUNDAMENTALS

$$\begin{vmatrix}
 (A_0 + s) & M_1 & 0 & 0 & 0 \\
 A_1 & -(A_1 + A_2 + s) & M_2 & 0 & 0 \\
 0 & A_2 & -(A_2 + A_3 + s) & 0 & 0 \\
 0 & 0 & 0 & 1 + H_1 + s & M_n \\
 0 & 0 & 0 & A_n & -(A_n + s)
 \end{vmatrix}
 \quad (1.12)$$

$(s + x^{n-k+1})$   
 $(-1)^{n+1}$   
 $\mathbf{1}$

$$\mathbf{1} \quad \text{for } k < n$$

$$\mathbf{1}$$

(s)

$$\begin{vmatrix}
 -(A_1 + s) & M_1 & 0 & 0 & 0 \\
 A_1 & -(A_1 + M_1 + s) & M_2 & 0 & 0 \\
 0 & A_2 & -(A_2 + M_2 + s) & 0 & 0 \\
 0 & 0 & 0 & 1 + H_1 + s & M_n \\
 0 & 0 & 0 & A_n & -(A_n + s)
 \end{vmatrix}
 \quad (1.128)$$

Expanding the determinant  $\Delta_{n+1}(s)$  along the last row yields the recurrent

equation:

$$A_{n+1}(s) = -\lambda_{n+1} + \sum_{k=0}^n \lambda_k A_n(j) \mathbf{n} \setminus_k \quad (1.129)$$

The probability  $p_{n+1}(t)$  is found with the help of the inverse LST:

$$P_{n+1}(t) = \int_0^t D_{n+1}(s) e^{s't} ds$$

where  $i = 0, \dots, n$ .

We are now faced with the problem of calculating the roots (eigenvalues) of (1.130). It can be shown that  $A_{n+1}(s)$  is a polynomial of power  $n$ , and all of its roots  $-1 < k < n + 1$ , are distinct and negative. Also, all roots of the polynomials of the neighboring orders  $n$  and  $n + 1$  are intermitten. This fact facilitates the computation of the recurrent equation (1.130).

We omit the cumbersome intermediate transformations and write the final result, taking into account that the probability of interest  $P(t) = 1 - p_{n+1}(t)$ :

$$P(t) = \sum_{i=0}^n p_i(0) \mathbf{n}_A \mathbf{E} \mathbf{V}^{\mathbf{r}}_{i+k}$$

In (1.131) we need to insert the roots which are usually calculated by numerical methods.

Two cases are of special interest. They are given without special explanations:

1. When all units are in up states at moment  $t = 0$ .
2. When the system has just come out of a failure state at  $r = 0$ .

If the system begins its operation at the state with all units completely operational:

$$p_0(0) = 1, p_i(0) = 0 \quad 1 < i < n + 1$$

then

$$P_{ij}(t) = A_0 A_1 \dots A_n Z_{i, m+n+1}^{s_i} \frac{e^{-\lambda t}}{\lambda} \sum_{k=0}^{i-1} \binom{i-1}{k} \left( \frac{\lambda}{\lambda + \mu} \right)^k \left( \frac{\mu}{\lambda + \mu} \right)^{i-1-k} \quad (1.132)$$

If the system begins its operation just after coming out of a failure state.

$$P_{ij}(0) = 1, p_{ij}(0) = 0 \quad i = 0, 1, \dots, n+1$$

then

$$P_{ij}(t) = (-\lambda)^{-1} \sum_{k=0}^{i-1} \binom{i-1}{k} \left( \frac{\lambda}{\lambda + \mu} \right)^k \left( \frac{\mu}{\lambda + \mu} \right)^{i-1-k} e^{-\lambda t} \quad (1.133)$$

### 1.6.5 Mean Time of Staying in a Given Subset

Now let us determine the mean time of the process staying in the subset  $\mathcal{L} = \{H_0, \dots, H_m\}$  starting from the state  $H_0$ . Of course, we may use a standard procedure for calculating this value: first, find the probability of staying in this subset with the initial condition  $H_0(0) = 1$  and then integrate the obtained expression. But this approach is too difficult. (Also, we did not obtain the result in a form which is easily integrated!) Thus, we choose another method.

A transition of the process from the initial state  $H_0$  to the absorbing state  $H_{n+1}$ , can be considered as consisting of  $n+1$  steps:

- 
- from  $H_0$  to  $H_1$  plus
  - from  $H_1$  to  $H_2$  (with the probability of going back and forth to the state  $H_0$ ), plus
  - from  $H_2$  to  $H_i$  (with the possibility of going back and forth to the states  $H_0$  and  $H_1$ ), plus
  - from  $H_n$  to  $H_{n+1}$ .

Let us find the mean time of passing from  $H_k$  to  $H_{k+1}$  where  $k > 0$ . Consider the auxiliary BDP with only  $k+1$  states where  $H_{k+1}$  is absorbing. We can use the stationary mean time of entrance in the absorbing state

found in (1.131). Thus, the value of interest can be found to be

$$\frac{T - V_T - V}{n, n + j} \sim \frac{Z-T * k, k + \dots}{0 \text{ sitsn}} \quad \frac{L A - \hat{f}c}{Osksm \text{ } ^a k Pk}$$

**1.6.6 Stationary Probability of Being in a Given Subset**

Let us again consider two subsets,  $E_+ = \{H_0, \dots, H_m\}$  and  $E_- = [H_{m+1}, \dots, H_N]$ . One can find the stationary probability of being in a given subset, say  $E_+$ , in two ways. Denote this probability  $K$ . The first way amounts to finding

$$A: = \lim \Pr\{t_f(r) \in E_+\} = \sum p_j$$

The second way uses the means  $T_+$  and  $T_-$  determined in (1.121) and (1.122), respectively. Let  $K - 1 = k$ . For the stationary process, the probability of being in a given state is proportional to the portion of time occupied by this state over the entire interval of observation on the time axis. This leads to the condition  $(K/k) = (T_+/T_-)$ . With the condition  $K + k = 1$  we find

$$T_+$$

**1.6.7 Death Process**

A particular class of birth and death processes is the so-called death process. From the name of the process, it is clear that this can be obtained from the BDP by putting all  $M_j, 1 < j \leq N$ , equal to 0. We consider this mathematical model because it is very useful when dealing with certain redundant systems without repair. For example, a redundant system consisting of  $n$  identical dependent units might be analyzed with this technique. The units can be dependent in a special way when the failure rate of each of them depends on the number of failed (or, equivalently, the number of operating) units at the moment.

This process can be described by the linear transition graph (see Figure 1.12). Let the process have  $N + 1$  states:  $H_0, \dots, H_N$ . Let  $A_k$  denote the transition rate from state  $H_k$  to state  $H_{k+1}$ . Using the same technique as above, we can write the equation:

$$p^k U - A_{k+1} p_{k+1} + A_k p_k = 0 \tag{1.134}$$

for all  $0 \leq k < N - 1$  and  $A_N = 0$ .

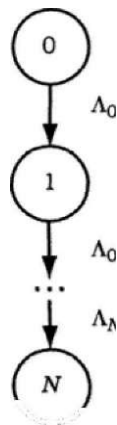


Figure 1.12. Example of a death process.

V\_\_'

We add the initial condition to the system of linear differential equations. In reliability problems this is usually  $p_0(0) = 1$ ; that is, the system is supposed to be in the state with all units up at the initial moment  $t = 0$ . Using the Laplace-Stieltjes transform, we can represent (1.134) in the form of the linear equations

$$\begin{aligned} -1 + s\hat{p}_0(s) &= -\Lambda_0 \hat{p}_0(s) \\ s \hat{p}_k(s) - \Lambda_k \hat{p}_{k-1}(s) &= -\Lambda_k \hat{p}_k(s) \end{aligned} \quad (1.135)$$

Solving (1.135) beginning with the first equation and sequentially substituting the obtained results in the next equation, we obtain

$$\hat{p}_0(s) = \frac{1}{s + \Lambda_0}$$

$$\hat{p}_k(s) = \frac{\Lambda_0 \Lambda_1 \dots \Lambda_{k-1}}{s + \Lambda_0} \quad (1.136)$$

$$\hat{p}_N(s) = \frac{\Lambda_0 \Lambda_1 \dots \Lambda_{N-1}}{s + \Lambda_0}$$

The solution for  $\hat{p}_k(s)$  is

$$\hat{p}_k(s) = \frac{\Lambda_0 \Lambda_1 \dots \Lambda_{k-1}}{(s + \Lambda_0)(s + \Lambda_1) \dots (s + \Lambda_{k-1})} \quad (1.137)$$



For different  $A_k$ 's the solution for  $p_n(t)$ , which is the probability that at moment  $t$  the process enters state  $H_N$ , can be found by using the inverse Laplace-Stieltjes transform

$$P_N(t) = \sum_{k=0}^{N-1} \frac{A_k}{\prod_{j=0, j \neq k}^{N-1} (A_j + A_k)} e^{-A_k t} \quad (1.138)$$

where  $\langle \langle * \rangle \rangle$  is a polynomial of the form

$$\langle \langle * \rangle \rangle = (* + A_0)(* + A_1) \dots (* + A_{N-1}) \quad (1.139)$$

and  $\langle \langle * \rangle \rangle'(-A_k)$  is the derivative with respect to  $x$  with the corresponding substitution.

If not all  $A_k$  are different, the expression for  $p_N(t)$  becomes more complicated. But even (1.138) is not particularly convenient for practical use.

In a very important practical case,  $A_k = A$  for all  $0 < k < N - 1$ . (Notice that this case corresponds to spare redundancy of identical units.) In this case (1.137) may be written in the form

$$p_N(t) = \frac{A^N t^{N-1}}{(N-1)!} e^{-At} \quad (1.140)$$

In this case we find (with the use of a table of the LSTs) that

$$\int_0^\infty e^{-st} p_N(t) dt = \frac{A^N}{s^N (s + A)} \quad (1.141)$$

This fact becomes clear if we consider a sequence of  $N$  identical exponentially distributed r.v.'s which represents a sample of the process until the entrance into state  $H_N$  (see Figure 1.13). As we mentioned above, a sum of  $N$  such r.v.'s has an Erlang distribution. The Poisson distribution is the cumulative function for the Erlang density and the result follows immediately.

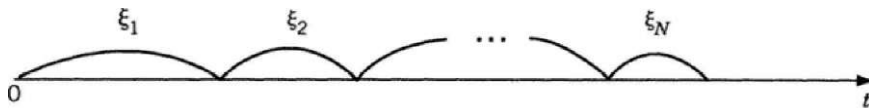


Figure 1.13. Time sequence of random values.

The mean time of the process entering into state  $H_N$  in the general case can easily be calculated as the sum of the time periods during which the process is remaining in each state

$$N \sim \sum_{k=1}^N T_k$$

Some details concerning death processes will be discussed later.

## CONCLUSION

Two distributions which are often used in engineering practice are the normal and the exponential. Each has its advantages and disadvantages. First of all, these distributions are very convenient for varied mathematical manipulations. But this argument is weak for practical applications. The question of their reasonable use, as with any modeling of real objects with the help of mathematical abstraction, always requires special "physical" verification based on experience and engineering intuition.

A Weibull-Gnedenko distribution is very convenient as a model for various physical phenomena because it is two parametrical. Besides, it has a clear physical sense as a distribution of extremal values. This distribution, as it relates to applied mechanical problems, was first mentioned in Weibull (1939). shortly after this, Gnedenko (1943) found classes of limit distributions of extreme values. A particular type of limit distribution has the form of the distribution discovered by Weibull

$$F(t) = 1 - \exp\left\{-\left(\frac{t}{A}\right)^b\right\}$$

where the new parameters are expressed as  $b = 1/\theta$  and  $a = \log A$ .

The reader interested in a deeper understanding of the probabilistic fundamentals of reliability theory should pay attention to special monographs. It is practically impossible to enumerate the books dedicated to this subject. An older, but nevertheless highly recommended book, is the book by Feller (1966). This book along with the book by Gnedenko (1967, 1988) were the main textbooks for several generations of statisticians and applied mathematicians.

For everyday use the books by DeGroot (1987) and Devore (1991) are recommended.

Concerning the limit theorems in the theory of stochastic processes, we must especially mention several works. Khinchine (1956a, 1956b, 1960) and Ososkov (1956) considered superposition of point processes, and later Grigelionis (1963) and Pogozhev (1964) generalized their result. Renyi (1962)

formulated the theorem on "thinning" of point processes which later was generalized by Belyaev (1962). Summary of all of these results can be found in Gnedenko and Kovalenko (1987).

The reader can find details concerning generalized generating sequences in Ushakov (1986, 1987, 1988a, 1988b).

## REFERENCES

- Belyaev, Yu. K. (1962).** Line Markov processes and their application to problems in reliability theory. *Trans. Sixth Conf. Probab. Statist.*, Vilnius.
- DeGroot, M. H. (1987).** *Probability and Statistics*, 2nd ed. Reading, MA: Addison-Wesley.
- Devore, J. L. (1991).** *Probability and Statistics for Engineering and the Sciences*, 3rd ed. Pacific Grove, CA: Brooks/Cole.
- Feller, W. (1966).** *An Introduction to Probability Theory and Its Applications*. New York: Wiley.
- Gnedenko, B. V. (1943).** Sur la distribution limit du terme maximum d'une serie aleatoire. *Ann. Math.*, no. 44.
- Gnedenko, B. V. (1967).** *The Theory of Probability*. New York: Chelsea.
- Gnedenko, B. V., Kovalenko, I. N. (1987).** *Introduction in Queuing Theory*, 2nd ed. (in Russian). Moscow: Nauka.
- Gnedenko, B. V. (1988).** *Course of Probability Theory*, 6th ed., revised (in Russian). Moscow: Nauka.
- Grigelionis, B. (1963).** On the convergence of sums of step stochastic processes to a Poisson process. *Theory Probab. Appl.*, vol. 8, no. 2.
- Grigelionis, B. (1964).** Limit theorems on sum of renewal processes. In *Cybernetics in the Service of Communism*, vol. 2, A. Berg, N. Bruevich, and B. Gnedenko, eds. Moscow: Energia.
- Khinchine, A. Ya. (1956a).** Streams of random events without aftereffect. *Theory Probab. Appl.*, No. 1.
- Khinchine, A. Ya. (1956b).** On Poisson streams of random events. *Theory Probab. Appl.*, no. 1.
- Khinchine, A. Ya. (1960).** *Mathematical Methods in the Theory of Queueing*. London: Charles Griffin.
- Ososkov, G. A. (1956).** A limit theorem for flows of similar events. *Theory Probab. Appl.*, vol. 1, no. 2.
- Pogozhev, I. B. (1964).** Evaluation of deviation of the equipment failure flow from a Poisson process. In *Cybernetics in the Service of Communism*, vol. 2, (in Russian) A. Berg, N. Brucvich, and B. Gnedenko, eds. Moscow: Energia.
- Renyi, A. (1956).** Poisson-folyamat egy jellemzese (in Hungarian). *Ann. Math. Statist.*, vol. 1, no. 4.
- Renyi (1962).**
- Ushakov, I. A. (1986).** A universal generating function. *Soviet J. Comput. Systems Sci.*, vol. 24, no. 5.

Ushakov, I. A. (1987). Optimal standby problem and a universal generating function. *Soviet J. Comput. Systems Sci.*, vol. 25, no. 4.

Ushakov, I. A. (1988a). Reliability analysis of multi-state systems by means of a modified generating function. *J. Inform. Process. Cybernet.*, vol. 24, no. 3.

Ushakov, I. A. (1988b). Solving of optimal redundancy problem by means of a generalized generating function. *J. Inform. Process. Cybernet.*, vol. 24, no. 4-5.

Weibull, W. (1939). A statistical theory of the strength of materials. *Ing. Vetenskaps Akad. Handl.*, no. 151.

Weibull, W. (1951). A statistical distribution of wide applicability. *J. Appt. Mech.*, no. 18.

## APPENDIX: AUXILIARY TOOLS

### 1.A.1 Generating Functions

Let  $v$  be a discrete random variable (r.v.) with distribution

$$p_r\{v = k\} \quad k \in 0, 1, 2, \dots$$

$$L P k^{-1}$$

$$V k$$

The *generating function* (g.f.) of  $v$  denoted by  $\langle p(z) \rangle$  is defined as

$$\langle p(z) \rangle = \sum_{k=0}^{\infty} p_k z^k$$

Thus, the coefficient of  $z^k$  equals the probability that  $v$  equals  $k$ . The g.f. is very convenient when one deals with the summation of discrete r.v.'s. Generating functions are especially effective when algorithms for computer calculations are involved.

For example, suppose we have two discrete r.v.'s  $a$  and  $b$ , with distributions  $a_k$  and  $b_k$ , respectively. We are interested in the distribution  $g_k$  of the new r.v.  $y = a + b$ . Of course, we could find the desired distribution directly:

$$\begin{aligned} p_r\{y = m\} &= \sum_{k=0}^m p_r\{a = k\} p_r\{b = m - k\} \\ &= \sum_{k=0}^m a_k b_{m-k} \end{aligned} \quad (1141)$$

But such an approach is not always simple or convenient. For computational purposes it is often better to use the g.f.'s of  $a$  and  $b$ . Let  $\langle p_a(z) \rangle$ ,  $\langle p_b(z) \rangle$ , and  $\langle p_y(z) \rangle$  be the g.f.'s of the respective distributions. Then we have

$$\langle p_y(z) \rangle = \langle p_a(z) \rangle \langle p_b(z) \rangle$$

In the new polynomial the coefficient of  $z^k$  is automatically equal to expression (1.141). This example does not exhibit all of the advantages of generating functions, but below we will show other cases where the use of g.f.'s is very effective.

Suppose we wish to find  $P\{v \leq k\}$ . We note that

$$P\{v \leq k\} = \sum_{z=0}^{\infty} z^k P(z)$$

where  $\sum_{z=0}^{\infty}$  operator that turns any negative power  $t$  of the term  $z^t$  into 1. Thus, after the substitution  $z = 0$ ,

$$P\{y \leq A\} = \sum P_k$$

Furthermore, it is clear that

$$\int_{-\infty}^{\infty} z^{k-1} E\{p_k z^{k-1}\} dz = E\{p_k\} = E\{v\}$$

To obtain higher moments, it is more convenient to use the so-called *moment generating function (m.g.f.)*  $\phi(s)$  of the r.v.  $v$ . This function can be written formally by simply substituting  $z = e^s$  into the generating function, that is,  $\phi(e^s) = \phi(z)$ . Then

$$\frac{d}{ds} \phi(e^s) = \sum_{k=0}^{\infty} k p_k e^{ks} = E\{k p_k e^{ks}\} = E\{v\}$$

$$\frac{d^2}{ds^2} \phi(e^s) = \sum_{k=0}^{\infty} k^2 p_k e^{ks} = E\{k^2 p_k e^{ks}\} = E\{v^2\}$$

### 1.A.2 Laplace - Stieltjes Transformation

With continuous r.v.'s the Laplace-Stieltjes transformation (LST) is often used. This transformation allows one to solve integral-differential equations with the "reduced" mathematical technique. The essence of the LST is depicted in Figure 1.14.

In this book we usually consider distributions of nonnegative r.v.'s. The transforms of such r.v.'s are defined for the distribution function (d.f.)  $F(t)$  as

$$\phi_F(s) = \int_0^{\infty} F(t) e^{-st} dt$$

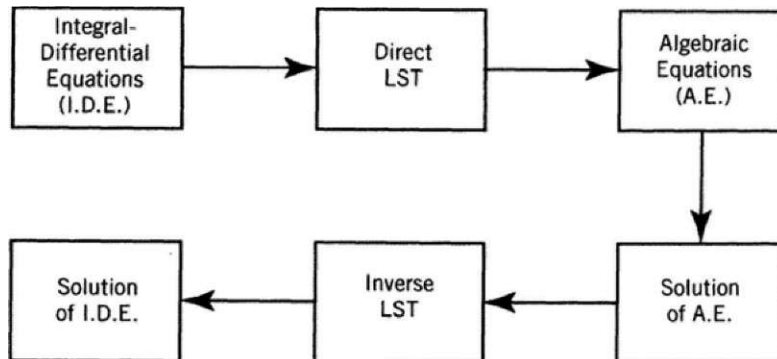


Figure 1.14. Scheme of the Laplace-Stieltjes transform usage.

and for the density function  $f(t)$  as

$$\langle p_f(s) \rangle = \int_0^{\infty} f(t) e^{-st} dt = \int_0^{\infty} e^{-st} dF(t)$$

If we consider the LST corresponding to the density function, the LST can be rewritten in the form

$$\langle p_f(s) \rangle = \int_0^{\infty} e^{-st} dF(t) = E\{e^{-st}\}$$

The correspondence between the original function  $f(t)$  and its LST  $\langle p(s) \rangle$  is usually denoted as

$$f(t) \leftrightarrow \langle p(s) \rangle$$

We now consider some properties of LSTs.

**Sum of Functions** the transformation of the sum of functions is the sum of the transforms:

This follows directly from the property of the integration. Obviously, (1.142) is true for any number of functions in the sum.

**Convolution of Functions** The convolution of two functions  $f_1(t)$  and  $f_2(t)$  is the function  $f(t)$  defined by

$$f(t) = \int_0^t f_1(t-x)f_2(x) dx = \int_0^t f_2(t-x)f_1(x) dx$$

This operation over the functions  $f_1(t)$  and  $f_2(t)$  is also denoted by

$$f(t) = f_1 * f_2(t)$$

The transform of the convolution of a pair of functions is the product of the transforms:

$$F(s) = F_1(s)F_2(s) \tag{1.143}$$

The proof of this is left as Exercise 1.14. Obviously, the correspondence is true for any number of functions in the convolution:

$$f_1 * f_2 * \dots * f_n = F_1(s)F_2(s)\dots F_n(s)$$

**Derivative of a Function** The transform of the derivative of a function can be expressed in terms of the transform of the function as

$$sF(s) = f(0) + sF(s) \tag{1.144}$$

The proof of this is left as Exercise 1.15.

**Integral of a Function** The transform of the integral of a function can be expressed by the transform of the function as

$$\int_0^t f(t-x) dx = \frac{F(s)}{s} \tag{1.145}$$

The proof of this is left as Exercise 1.16.

**Property of the LST of the Density Function** If the function  $f(t)$  is the density of the distribution of the r.v. that is,  $f(t) = [dF(t)]/dt$ , then

$$\int_0^\infty f(t)e^{-st} dt = F(s) \tag{1.146}$$

and

$$\int_0^\infty f(t)e^{-st} dt = -\frac{dF(s)}{ds} = -\frac{d}{ds} \int_0^\infty f(t)e^{-st} dt$$

**Property of the LST of the PFFO** If  $P(t)$  is the probability of a failure-free operation, that is,  $P(t) \sim 1 - F(t)$ , then the corresponding LST at 0 is

$$\int_0^{\infty} p(t) e^{-st} dt = \int_0^{\infty} p'(t) dt = T \quad (1.147)$$

where  $T$  is the mean of the distribution  $F(t) = 1 - P(t)$ . This value is called the *mean time to failure* (MTTF).

**Initial Moments of a Distribution** The Laplace-Stieltjes transformation of the density function allows us to obtain the moments as

$$\frac{d^k}{ds^k} \Big|_{s=0} \quad (1.148)$$

These moments are obtained more conveniently with the help of the continuous m.g.f. which coincides with the LST except in the sign of the power in the exponential:

$$\int_0^{\infty} f(t) e^{st} dt$$

In this case there is no change in the sign:

$$ds^*$$

The Laplace-Stieltjes transformation represents a very useful mapping from one functional space into a new one where the original functions are replaced with transformed ones. Operations over these new functions are often simpler in the transformed space. The general idea is reflected in Figure 1.14,

### 1.A.3 Generalized Generating Sequences

The method of *generalized generating sequences* (GGS) is based on a new approach which is genetically tied to generating functions. It is very convenient for a computerized realization of different enumeration problems which often arise in discrete optimization. We begin with a simple example to illustrate the main features of the GGS.

Consider a series connection of  $n$  resistors. Each unit in the series has a resistance which has a random value (for various reasons, e.g., manufacturing, storage, environmental influence, etc.). This random value of the unit's



resistance is characterized by some distribution. We assume that this distribution is discrete and the resistance of the  $i$ th resistor equals the value  $r_i$  with probability  $p_i$ , so that

$$\sum_{i=1}^L P_i r_i = 1$$

$$IZ/SM,$$

where  $M_i$  is the number of discrete values of the  $i$ th resistor. For each unit we can construct the generating function of the distribution of the resistance values:

$$G_i(Z) = \sum_{i=1}^{M_i} P_i Z^{r_i}$$

To find the distribution of the resistance of the entire series connection, we can compute its g.f.

$$G(z) = \prod_{i=1}^L G_i(z) = \sum_{i=1}^L P_i z^{r_i} \quad (1-149)$$

After simple algebraic transformations, we write the final expression in the form of a polynomial

$$C(z) = \sum P_s z^{R_s} \quad (1.150)$$

where the coefficient  $P_s$  of the term  $z^{R_s}$  equals the probability that the series system's resistance is  $R_s$ .

We remark that, in a computational sense, the introduction of the auxiliary variable  $z$  permits us to separate the variables of interest:  $p$  and  $r$ . (We omit other useful properties of the g.f. for this discussion because they are irrelevant here.) To compute  $P_s$  and  $R_s$ , one needs only to multiply the  $p$ 's and to add the  $r$ 's.

This example is very clear and contains no new information for those who know how to work with generating functions. Of course, if the problem is to calculate the resistance of a parallel connection of resistors, it is impossible to use (1.149) and (1.150) in any direct way. To use the g.f., one has to consider r.v.'s which measure conductivity (instead of resistance) and then find the desired result in terms of conductivity. Finally, the result can be transformed from units of conductivity to units of resistance.

Now suppose it is necessary to analyze the pipeline capacity of a set of pipes connected in series. In this example the collective capacity is the **minimum** of the capacities of the individual units. The usual generating function does not work here at all! We suggest a new approach which we call the *generalized generating sequence* (GGS).

To explain how the GGS works, we use the above example with resistors in series. First, we analyze the computations involved in moving  $G(z)$  as expressed by (1.149) to  $G(z)$  as expressed by (1.150). For the moment, consider a series system of two resistors labeled  $A$  and  $B$ . In terms of calculations, we perform the following operations.

1. The probability distributions of the resistances are stored as sequences of ordered pairs. We can associate these sequences with the symbols  $A$  and  $B$  and so write

$$A = \{[P \setminus u r_u], \{p_{x2} y r_l^{\wedge}, \dots, \{p_{let} r_{iv}\})$$

and

$$B = \{iP2l * r2 \setminus \} \{P22 > r22J > \dots \rightarrow \{ P2n > r2w \}$$

where, for example, the pair  $\{p_u, r_{iy}\}$  exhibits the probability that the resistance of resistor  $A$  will have the value  $r_{lj}$ ,

2. Now introduce a new operator  $CI$  which operates on the pair of sequences  $A$  and  $B$  and produces a new sequence  $C$  of ordered pairs  $(p_m, r_{3k})$ . The sequence  $C$  represents the probability distribution of the resistance of the series connection of  $A$  and  $B$ . Thus,

$$(I(A, B) = C$$

or, since each term of the sequence  $C$  is a pair of numbers, it can also be rewritten as

$$rI(A, B) = (n_p(A, B), Sl_r(A, B))$$

The sequence  $C$  is formed under  $\Pi$  from the pair  $(A, B)$  as follows:

- (a) For each pair  $(p_u, r_{ij})$  and  $(p_y, r_l^{\wedge})$  compute the pair  $(p_u p_2 f i r_u + r^2 A$
- (b) Order the obtained pairs according to increasing values of their second components.
- (c) When two or more pairs in the newly obtained sequence are tied in their second components, combine all such pairs into the single pair. The first component of the new pair is the *sum* of all first components of the tied pairs, and the second component of the new pair is the (common) *product* of the tied second components.

Note that the operators and have a very specific meaning in this example. But this meaning can be substituted by others in different situations. For example, for the pipeline consisting of a series connection of units with different capacities, one can write  $O_c(c_1, c_2) = \min(c_1, c_2)$  where  $c_i$  is

the capacity of the  $i$ th pipe. All of the remaining formal operations and the order of their performance are similar. Therefore, the above-described computational algorithm, in general, can be used with no restrictions on the polynomial form. The new approach can be used for enumeration problems involving different physical parameters. We will show the effectiveness of this operator for computational problems of complex system reliability analysis and discrete optimization problems.

Now let us describe the procedure in more general terms. Keeping in mind the use of a computer, we introduce a more formal description.

For a more vivid presentation we will use a special terminology to distinguish the GGS from the g.f. This will relieve us of having to use traditional terms in a new sense, which often leads to confusion. Moreover, we hope that this new terminology can help us, in a mnemonical sense, to remember and even to explain the procedure.

In an ancient Roman army, a *cohort* was the main combat unit. Each cohort consisted of *maniples* which were independent and sometimes specialized simple combat units. Several cohorts composed a *legion*. The use of this essentially military terminology appears to be convenient in this essentially peaceful applied mathematical field. We set up a one-to-one correspondence between the above-mentioned military units and the GGS with its attributes.

Consider a system consisting of  $n$  units. Each unit  $j$  is characterized by its GGS. Let the GGS of a unit be called a legion. Each legion  $j$  includes  $V_j$  cohorts:

Each cohort  $C_{jk}$  is composed of some set of the unit's parameters, special characteristics, and auxiliary attributes. We call these components of the cohort maniples. Therefore,

$$C_{JK} = (M_{JK,1}, M_{JK,2}, \dots, M_{JK,s})$$

where  $M_{JK,1}$  is the corresponding maniple and  $s$  is the number of different maniples (assumed to be the same for each cohort).

The operation of interaction between legions is denoted by  $Cl^L$ . This operator is used to obtain the resulting legion

$$L = N^L L,$$

The operator  $Cl^L$  denotes a kind of "n-dimensional Cartesian product" and a special "reformatting" of the resulting cohorts. This reformatting depends on

the specific nature of the problem [see, e.g., item (c) of the series resistors example].

As a result of this interaction of the legions, one obtains

$$N = \prod_{i=1}^N \text{OJ}$$

new cohorts. For each cohort the following notation is used where  $C_i^c$  denotes the cohort's interaction,  $k$  is the subscript of this cohort in the set obtained as a result of the procedure

$$c, = \prod_{i^k} C_{ij}$$

(before using the formatting procedure), and  $ij$  are corresponding subscripts of the cohorts taking part in the interaction (this fact is conditionally reflected in the notation  $ij^{**k}$ ). The new cohort can be represented as

$$C_k = (M_{k1}, M_{k2}, \dots, M_{ks})$$

Each new cohort is obtained as a result of a vector product-type interaction of maniples:  $n$  maniples of the first type interact between themselves,  $n$  maniples of the second type interact between themselves, and so on. The interaction between maniples of a specified type can be called a "natural" interaction because they involve a real physical sense of the corresponding parameters:

$$M_{kl}$$

$$- \text{Of } A\% \\ \prod_{i} \\ \text{i.efc}$$

Here the subscript  $l$  defines the type of maniple interaction.

The resulting legion consists of a set of cohorts obtained by using the formatting procedure. It can be written as

$$L = (C_{jij}, C_{(2)}, \dots,$$

where  $N^* < N$ . This formatting procedure can consist of special operations over  $N$  cohorts. For example, several cohorts can be joined into an equivalent one in which some specified maniple equals the sum of others: we have the same solution with the g.f. when we add the probabilities of the terms with the same power of  $z$ . It may also be the selection of a "priority" (or "domination") of one cohort over another. Such a formatting procedure will

be encountered in Chapter 10. The essential ideas of the proposed method of generating sequences can best be explained with the help of concrete examples. Such examples will be provided in Chapters 3, 8, and 10.

### EXERCISES

1.1 Prove the equivalency of expressions (1.29) and (1.30), that is, prove that

1.2 Prove that  $a$  is the mean of a normal distribution with density function

$$f_N(x|a, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-a)^2}$$

1.3 Prove that  $\sigma^2$  is the variance of a normal distribution with density function

$$f_N(x|a, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-a)^2}$$

1.4 Using the m.g.f. for the normal distribution, find the expression for the first moment (the mean).

1.5 Using the m.g.f. for the normal distribution, find the expression for the variance.

1.6 One observes two Bernoulli sequences with  $n_1$  and  $n_2$  trials, respectively. A successful trial appears at the first sequence with probability  $p_1$  and at the second sequence with probability  $p_2$ .

- Find the probability,  $R_k(n_1, n_2)$ , that there will be  $k$  successes in the entire  $n = n_1 + n_2$  trials.
- Show that for  $p_1 = p_2 = p$  the probability of interest equals

1.7 Prove that

90 FUNDAMENTALS

1.8 Prove that

$$\sum_{1 \leq k \leq n} (-1)^k \binom{n}{k} = 0$$

1.9 Prove that

$$\binom{n_1 + n_2}{k} = \sum_{0 \leq j \leq n} \binom{n_1}{k} \binom{n_2}{k - j}$$

- 1.10 There are two variants of equipment: one performs its operation during time  $t$  and another performs the same operation during time  $It$ . Both units have an exponentially distributed time to failure. The systems under consideration have different reliability: the first one has a failure rate equal to  $2A$ , and the second one has a failure rate equal to  $A$ . What variant of equipment will perform its task with larger probability?
- 1.11 A production line manufactures good quality items with probability 0.9. Find the probability that in a sample of size  $n = 500$  the number of failed items does not exceed 80.
- 1.12 The average portion of deficient items equals 0.01. Find the probability that in a sample of size  $n = 100$  the number of failed items does not exceed 2,
- 1.13 A flow of the equipment failures is formed by superposition of the flows of different types of units. Each type of unit produces a failure flow which can be described as a Poisson process. During a given period of time, the average number of failures of some specified type of unit, equals 36. How many spare units should be supplied for this period of time to support failure-free operation of this type of unit with probability 0.95?
- 1.14 Prove that the LST of the convolution of a pair of functions is the product of the LSTs of the transforms of the initial functions in convolution.
- 1.15 Prove that the LST of the derivative of a function can be expressed as  $f'(U) ** stpis) - fiox$
- 1.16 Prove that the LST of the integral of a function can be expressed as

$$\int_0^t f(t) dt \leftrightarrow \frac{1}{s} \varphi(s)$$

**SOLUTIONS**

1.1 For a binomial coefficient one can write the well-known expression

$$\binom{n}{m} = \frac{n!}{m!(n-m)!} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-m+1)}{1 \cdot 2 \cdot \dots \cdot m}$$

As one knows from mathematical combinatorics, the latter expression is true for any  $n$ —even for a negative noninteger. Thus, setting  $n$  negative, one obtains

$$\binom{-n}{m} = \frac{(-n) \cdot (-n-1) \cdot \dots \cdot (-n-m+1)}{1 \cdot 2 \cdot \dots \cdot m}$$

or, after trivial transformations,

$$\binom{-n}{m} = (-1)^m \binom{n+m-1}{m}$$

Because

one can finally write

1.2 Introduce a new variable

$$y = \frac{x-a}{ry/2}$$

Then the initial expression takes the form

$$\begin{aligned}
 & \int_{-\infty}^{\infty} f(y) e^{-y^2} dy \\
 & = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-y^2} dy + \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-y^2} dy
 \end{aligned}$$

The first term of the latter sum equals 0 because the function under integral is symmetrical in respect to  $y = 0$ . The second term is the well-known Euler-Poisson integral

$$\int_{-\infty}^{\infty} e^{-y^2} dy = \sqrt{\pi}$$

Thus, the final expression of the integral of interest equals  $\frac{1}{\sqrt{\pi}}$ .

Introduce a new variable

$$x = y^2$$

Then the initial expression takes the form

$$\frac{1}{\sqrt{\pi}} \int_0^{\infty} \frac{1}{\sqrt{x}} e^{-x} dx$$

which can be represented as

Taking the latter integral by parts, one obtains

$$\frac{1}{\sqrt{\pi}} \left( \sqrt{x} e^{-x} + \int_0^{\infty} e^{-x} dx \right)$$

The first term of the latter sum equals 0 because an exponential function grows faster than a linear one. The second term is the Euler-Poisson integral obtained above. Thus, the final expression of the integral of interest equals  $\frac{1}{\sqrt{\pi}}$ .



L4 Consider (1.54). Assume that the first derivative with the substitution  $j = 0$  derives the mean:

$$\frac{d}{ds} \ln \int_0^\infty A(j) e^{-js} dj \Big|_{s=0} = a$$

$$- \exp^{as} + -a s j = (a + cr sj \exp^{as} + \dots) \Big|_{s=0}$$

1.5 Using the intermediate

$$\frac{d^2}{ds^2} \ln \int_0^\infty A(j) e^{-js} dj = as + \frac{1}{T} a^2 s^2$$

result of Exercise 1.4, one

obtains

$$\begin{aligned} \frac{d^2}{ds^2} \ln \int_0^\infty A(j) e^{-js} dj &= \int_0^\infty (a + cr^2 s) \exp(-js) A(j) dj \\ &= a^2 \exp(-a) + \dots + (a + cr^2 s)^2 \exp(-as) + \dots \\ &= cr^2 + (as + a^2)^2 j \exp(-as) + \dots \end{aligned}$$

This gives the second initial moment which is equal to the sum of the variance  $a^2$  and the mean squared  $a^2$ . Substituting  $s = 0$  gives the desired result.

1.6 Denote by  $b(k, n)$  the probability that there will be  $k$  successes in  $n$  trials. Then

(a) 
$$\sum_{k=0}^n b(k, n) = 1$$

and, finally,

(b) If  $p_1 \sim p_2 \wedge p$  one can consider two experiments as one experiment with a total number of trials equal to  $n = n_1 + n_2$ . For this case one has

$$p^k q^{n-k} \sim \dots \tag{HI.2}$$

1.7 The solution follows immediately if one considers a binomial of the form  $(1 + 1)^n$ :

## 84 FUNDAMENTALS

$$(i + i)^n = \frac{1}{2} (W^{-1} - \frac{1}{2})$$

- 1.8 The solution follows immediately if one considers a binomial of the form  $(1 - 1)^n$ .
- 1.9 Compare solutions (E1.1) and (E1.2) obtained in Exercise 1.6. Substitution of  $P_i = p_2 = p$  into (E1.1) gives

$$R = p^k q^{n-k} \sum_{0 \leq j \leq k} \binom{n_1}{j} \binom{n_2}{k-j}$$

Comparison of the latter expression with (E1.2) leads to the desired result.

- 1.10 Both systems are equivalent in terms of the chosen criteria.
- 1.11 Apply the normal approximation with mean = 450 and a standard deviation  $\sigma = 745 = 6.7$ . Use a standard table of the normal d.f. for an argument =  $(420 - 450)/6.7 = -4.48$ .
- 1.12 Apply the Poisson approximation with parameter  $(0.01 \times 100) = 1$ . Use a standard table of the Poisson d.f.
- 1.13 Apply the normal approximation with mean = 36 and a standard deviation  $\sigma = \sqrt{36} = 6$ . Use a standard table of the normal d.f.
- 1.14 The convolution of two functions is defined as

$$f_1 * f_2(t) = \int_0^t f_1(x) f_2(t-x) dx$$

By definition, the LST is

$$s^{-1} (f_1 * f_2)(s) = \int_0^\infty \int_0^t f_1(x) f_2(t-x) e^{-st} dt dx$$

Using the Dirichlet formula, we obtain

$$\begin{aligned} s^{-1} (f_1 * f_2)(s) &= \int_0^\infty dx \int_x^\infty e^{-st} f_1(x) f_2(t-x) dt \\ &= \int_0^\infty dx \int_0^{t-x} e^{-s(x+y)} f_1(x) f_2(y) dy \end{aligned}$$

Substituting  $y = t - x$ , we obtain

$$s^{-1} (f_1 * f_2)(s) = \int_0^\infty f_1(x) e^{-sx} dx \int_0^\infty f_2(y) e^{-sy} dy = s^{-1} f_1(s) f_2(s)$$

Thus,  $f_1 * f_2(s) = f_1(s) f_2(s)$  which corresponds to (1.3).

1.15 By definition, the LST  $\langle p^*(s)$  of the derivative  $f(t) = df/dt$  is

$$p^*(s) = \int_0^{\infty} f(t) e^{-st} dt$$

The following simple transformations need no explanation:

$$\begin{aligned} \int_0^{\infty} f'(t) e^{-st} dt &= \int_0^{\infty} e^{-st} df(t) - f(t) e^{-st} \Big|_0^{\infty} - \int_0^{\infty} (-s) f(t) e^{-st} dt \\ &= -f(0) + s \int_0^{\infty} f(t) e^{-st} dt = -f(0) + s p^*(s) \end{aligned}$$

Thus, the desired equality is proven.

1.16 This relationship follows from the chain of simple transformations:

$$\begin{aligned} \int_0^{\infty} f'(x) dx e^{-st} &= - \int_0^{\infty} f'(x) e^{-sx} dx \\ &= - \int_0^{\infty} f(x) e^{-sx} dx - \int_0^{\infty} f(x) e^{-sx} dx \\ &= -f(0) + s \int_0^{\infty} f(x) e^{-sx} dx = -f(0) + s p^*(s) \end{aligned}$$

Thus, the validity of the equation is proven.

## CHAPTER 2

# RELIABILITY INDEXES

Reliability indexes are basically needed for the quantitative characterization of a system's ability to perform its operations. These indexes must reflect the most essential operating properties of the system, be understandable from a physical viewpoint, be simple to calculate at the design stage, and be simple to check at the test and/or usage stage.

Sometimes it is practically impossible to characterize a system with only one reliability index. But, at the same time, the number of reliability indexes has to be as small as possible. Psychologists say that more than three numerical characterizations of the quality of some object can only lead to confusion and misinterpretation of a situation. Those who deal with multicriteria optimization also know that the Pareto set should be of a small dimension. (One might recall the classical example from medieval French literature: the Buridan donkey died trying to solve a two-dimensional problem when he could not choose one bunch of hay from two!)

Simultaneously, one has to avoid the use of different "integrated" or "weighted" indexes: Such indexes generally have no clear physical sense and may mask an unacceptable level of one index by uselessly high levels of the others.

Reliability indexes may not only be used for the characterization of a system as a whole, but also some of the indexes may have an intermediate character. For example, the system, considered as an independent object, might be characterized by an availability coefficient. If the same system is part of a more complex structure, it may be more reasonable to characterize it separately with the mean time to failure (MTTF) index and the mean repair time index because they might be used to more accurately express the complex system's availability index. Moreover, the system as a whole can be

100 RELIABILITY INDEXES  
 characterized with indexes in the form of nondimensional real numbers. But for the system's subsystems, we sometimes need to know special functions (distribution functions, failure rates, etc.).

Almost all reliability indexes are of a statistical nature and depend on time.

We now make several points about unrepairable and repairable (renewal) units and systems. The distinction between repairable and unrepairable items is relative. The same system may be considered as repairable in one circumstance and as unrepairable in another. The main indicator is a system's ability to continue its operation after repair. For example, a computer used for routine calculations with no special time restrictions may be considered as repairable. The same computer used for a noninterruptible technological process or in a military action (which is almost always dangerous if interrupted!) may be considered as unrepairable. But if in the latter case the computer is used in an on-duty regime, it might be considered as repairable during the idle period.

Of course, some technical objects are essentially unrepairable. Some of them, such as a light bulb, cannot be repaired at all. As another example, a missile cannot be repaired during its mission. For convenience, in these cases we speak of a "renewal socket" in which unrepairable units are installed one after another in the case of a failure. Thus, after a first unsuccessfully launched missile, another one may be launched; after a first bulb has failed, another one may replace it.

## 2.1 UNREPAIRABLE SYSTEMS

### 2.1.1 Mean Time to Failure

If the criterion of a system's failure is chosen and perfectly well defined, we can determine its reliability indexes, in particular, the mean time to failure (MTTF).

After observing  $N$  failures of  $N$  unrepairable systems, there are records of nonnegative values:  $t_1, t_2, \dots, t_N$ . One of the most natural characteristics of this set of observations is the sample mean, or the mean time to failure (MTTF):

$$T = (t_1 + t_2 + \dots + t_N) / N \quad (2-1)$$

This reliability index means that the system, on average, works  $T$  time units before a failure.

Consider these values in increasing order, that is, present the observations as

$$t_1 < t_2 < \dots < t_N$$

In this new notation the following equivalent equation can be written:

$$T = Nt_1 + (N - 1)(t_2 - t_1) + \dots + (t_N - t_{N-1})$$

The equivalency of the formulas follows from consideration of Figure 2.1 where a histogram of the values  $t_k$  is presented.

If a prior distribution  $F(t)$  of a system's TTF is known, the expected value of  $T$  can be calculated in the standard way:

$$(2.2) \quad T = \int_0^{\infty} t dF(t)$$

For nonnegative random variables, the following equivalent expression can be written (see Exercise 2.1)

$$T = \int_0^{\infty} P(t) dt \quad (2.3)$$

where  $P(t) = 1 - F(t)$ .

The equivalency of (2.2) and (2.3) follows from the fact that we are only using different means of integrating the same function (see Figure 2.1). On a heuristic level this result may be explained by comparing this case with the analogous discrete case depicted by Figure 2.1.

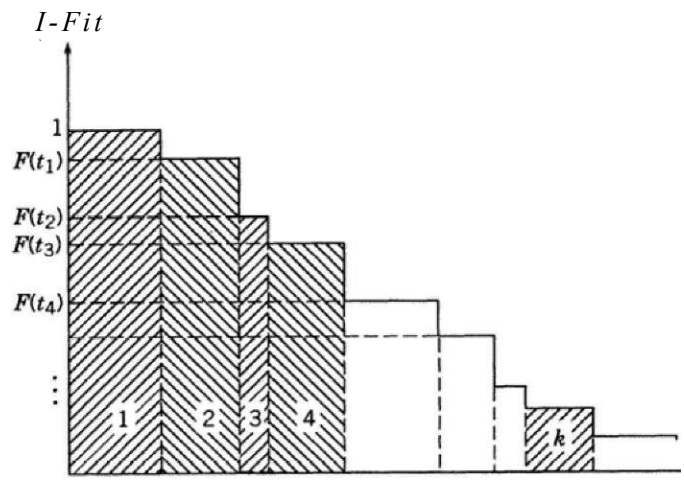
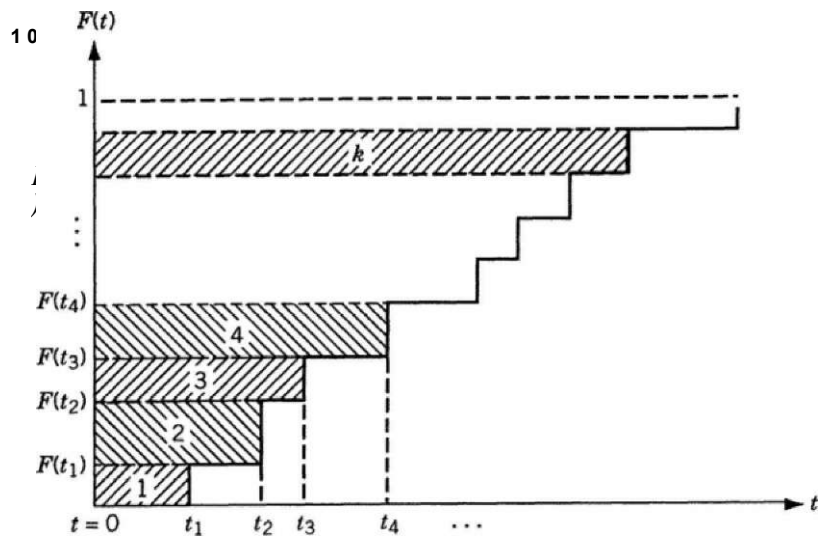
The reliability index MTTF is very convenient if the system's outcome linearly depends on the time of its successful performance. For example, consider a form for producing some plastic item. After a failure the form is replaced by a new one. Thus, this form can be considered as a socket in the above-mentioned sense. In this case the effectiveness of using the form depends only on the average time to failure. But in other cases this reliability index may prove to be inconvenient.

### 2.1.2 Probability of a Failure-Free Operation

Consider a system performing an operation with a fixed duration  $t_0$ . In this case each  $t_k < t_0$  corresponds to a system failure. A natural reliability index in this case is the probability of a failure-free operation, which reflects the frequency of appearance of the condition  $t_k > t_0$ . We introduce the so-called indicator function

$$d(x, y) = \begin{cases} 1 & \text{if } x > y \\ 0 & \text{otherwise} \end{cases}$$

In other words, we define a system failure in a new form: the system fails when  $d = 0$ .



- 1 -
- 1 -
- 1 -
- 1 -

Figure 2.1. (Explanation of two types of summation on a histogram of random variables.



For the same data we have to calculate the new reliability index as

$$P(t_0) = (l/N)(d_1 + d_2 + \dots + d_N)$$

where  $d_k = d(t_k, f_0)$ . If we know the distribution  $F(t)$  of the system TTF, the probability of a successful operation can be expressed as

$$P\{l_0\} = 1 - F(t_0)$$

Sometimes the duration of a task is a random variable itself, with a distribution  $H(t)$ . We may then speak about the expected probability of success for the random performance time. The general expression for this index is

$$P\{l_0\} = \int_0^{l_0} f_H(t) dH(t) \tag{2.4}$$

Several particular cases are considered in Exercise 2.2.

**2.1.3 Failure Rate**

As we mentioned above, we sometimes have to know some special functions in order to calculate the reliability indexes of a complex system. One such important function is the failure rate  $A(t)$ . In strict probabilistic terms this is the instant conditional density function at moment  $t$  under the condition that the random variable under consideration is not less than  $t$ , that is,

$$A(t) = \frac{f(t)}{1 - F(t)} \tag{2.5}$$

At first this function, called the hazard rate, appeared in demography connected to the insurance business. The physical sense of this function can be easily explained in the following simple terms. If we know the prior distribution  $F(t)$  with density  $f(t)$ , then an element of the conditional probability

$$Pr(dt|t) = A(t) dt$$

is the probability of the death of an individual of age  $t$  during the forthcoming time interval  $[t, t + dt]$ .

This function has exactly the same sense in reliability theory when one substitutes the corresponding terms. We refer to this function as the failure rate. To explain it, consider the uniform distribution  $F(t)$  on the interval  $[0,10]$ . In this case  $A(0) = f(0) = 0.10$  because  $P(0) = 1$  for a nonnegative r.v. Next, consider the moment  $t = 1$ . The area of the domain for the r.v. under

100 RELIABILITY INDEXES

the condition index is larger than 1 become smaller: now it is [1, 10]. So  $A(1) = 1/9$ . Of course, the same result can be obtained directly from (2.5) if we substitute  $f(1) = 0.10$  and  $P(1) = 0.90$ . Then for the next moment, say  $t = 5$ , we have  $A(5) = 0.20$ ; for  $t = 9$  we have  $A(9) = 1.0$ ; for  $t = 9.9$  we have  $A(9.9) = 10.0$ ; for  $t = 9.99$  we have  $A(9.99) = 100.0$ ; and so on. The function  $A(t)$  approaches  $\infty$  at  $t = 10$ .

For a normal distribution with mean  $\mu = 10$  and standard deviation  $\sigma = 1$ , we can calculate  $A(t)$  using a standard table of the normal distribution.

In both cases we observe that the function  $A(t)$  is increasing and unbounded. Thus, the unit's reliability for such a TTF distribution becomes worse in time. Such an aging process is very natural for most real objects. But this type of increasing function is not the only one. As we considered in Chapter 1 for the exponential distribution, the failure rate is constant in time. Moreover, the so-called "mixture of exponential distributions" has a monotonically decreasing function  $A(t)$ .

For the mixture of two exponential functions, we can write

$$F(t) = 1 - p[\exp(-a_1 t)] - (1 - p) [\exp(-a_2 t)] \quad (2.6)$$

The expression for  $A(t)$  can easily be obtained

$$A(t) = \frac{p a_1 \exp(-a_1 t) + (1 - p) a_2 \exp(-a_2 t)}{p \exp(-a_1 t) + (1 - p) \exp(-a_2 t)}$$

We will analyze this equation "on a verbal level" using only simple explanations. For  $t = 0$  we have

$$A(0) = p a_1 + (1 - p) a_2$$

that is,  $A(0)$  is a weighted hazard rate at this moment. Then note that the function  $A(t)$  is a monotonically decreasing function. If  $a_1 > a_2$ , then  $\lim_{t \rightarrow \infty} A(t) = a_2$ .

From (2.5) it follows that

$$A(t) = \frac{dF(t)}{P(t)} = \frac{dP(t)}{dt} \frac{1}{P(t)}$$

This immediately yields

$$P(t) = \exp \int_0^t -f(x) dx \quad (2.7)$$

From the condition follows that, for

$$0 < P(t) < 1, \text{ it any } t,$$

and

$$0 < \int_0^{\infty} f'(t) dt < \infty$$

Thus, the function possesses (2.8).

$$\lim_{t \rightarrow \infty} f'(t) dt = \infty$$

is such that  $A(f) > 0$  and properties (2.7) and

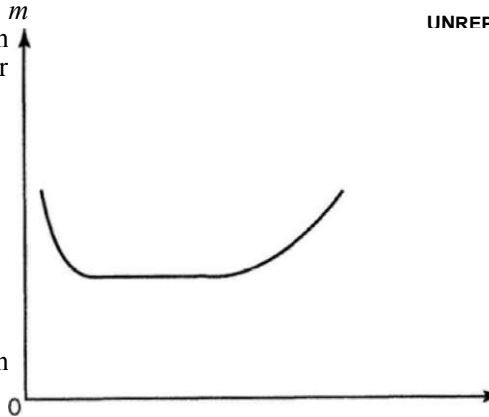


Figure 2.2. U-shaped function of A(f).

In most practical so-called

cases we observe a "U-shaped form" of the depicted in Figure 2.2.

function  $A(t)$ , as During the first period of time, we observe a "burning-out" process. This process consists in the early failing of weak or defective items. Then follows a period of "normal" operation during which the failure rate is constant. During this period a failure only occurs "completely incidentally," or, as one sometimes says, "as a brick fallen from the roof." It is a period of wearing-out, fatigue, and other normal phenomena of aging.

We will show below that qualitative knowledge about the failure rate is very important for reliability analysis.

## 2.2 REPAIRABLE SYSTEM

### 2.2.1 Description of the Process

During the observation of a repairable system, we can record the sequence of periods, each of which consists of a successful performance time plus an idle

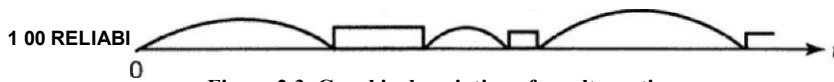


Figure 2.3. Graphic description of an alternating process.

time. Such a process is illustrated in Figure 2.3. Let denote the random time from the completion of the  $(k - 1)$ th repair to the  $k$  th failure, and let  $\tau_k$  denote the duration of the  $k$ th repair (renewal).

In the simplest case with a socket (a one-unit system), we suppose that a repair is equivalent to the replacing of a failed unit. This corresponds to a complete renewal. In this case we consider an alternating stochastic process with i.i.d. r.v.'s  $X$  and  $Y$ , having distributions  $F(x)$  and  $G(y)$ , respectively. We denote this alternating stochastic process by  $\{X, Y\}$ .

Of course, the corresponding process for a system consisting of several renewal units may be much more complicated. Almost all the following explanations will be—for simplicity—presented for a renewal unit, or a socket.

All indexes used for an unreparable system can also be used in this case for the appropriate purpose. But for repairable units and systems we have to consider several special indexes. They are more complicated and need more explanation.

### 2.2.2 Availability Coefficient

Consider a system which has to work in a "waiting" regime and, at the same time, the duration of the task performance is negligibly small. In this case a natural reliability index is the so-called *availability coefficient*  $K(t)$ . This index is the probability that the system will be in an operating state at a specified moment (in the future).

The numerical value of  $K(t)$  depends on the specified moment of time  $t$ . For example, if we know that at  $t = 0$  the system is new and, consequently, is in an operating state, then at moment  $e$ , where  $e$  is small, the probability that the system is in an operating state is close to 1 and approximately equals  $K(e) = P(e)$ .

The behavior of  $K(t)$  in time can be periodically attenuating or strictly attenuating. This depends on the types of d.f.'s  $F(x)$  and  $G(y)$ . For illustrative purposes, consider the case where  $F(x)$  is a normal d.f. with a small coefficient of variation  $k$  and  $G(y)$  is a degenerate function (i.e.,  $Y$  is constant).  $K(t)$  for this case is presented in Figure 2.4.

It is clear that the first time to failure has a normal d.f. with some mean  $T$  and a relatively small  $k < 1$ . The renewal completion time has the same distribution biased on the time axis. If  $k > 3\alpha$ , there may be some interval between  $T$  and  $T + \tau$  where  $K(t) = 0$ . The second time to failure also has a normal

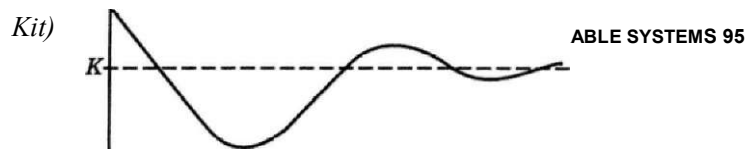


Figure 2.4. Example of the oscillating behavior of  $K(t)$  in time.

d.f. but with the standard deviation larger by times. Thus, the zone between  $2T$  and  $2(T + \tau_j)$ , where  $K(t) = 0$ , will be smaller. Finally, for  $t \gg T + \tau_j$ ,  $K(t)$  will be almost constant.

If both d.f.'s  $F(t)$  and  $G(t)$  are exponential, the function  $K(t)$  is strictly decreasing with exponential speed. We will consider this case later.

For large  $t$  the initial state has practically no influence on the behavior of  $K(t)$ . In this case the probability that the system is in an operating state equals the proportion of the total up time to the total operating time. Later we will show that for this case

$$K = \frac{EU}{E} \quad (2.10)$$

The index  $K$  is called the *stationary availability coefficient*, or simply, the *availability coefficient*.

Sometimes we are interested not in a "point" characteristic  $K(t)$  but in the average time spent in an operating state during some period of time, say  $t$ . We introduce the index

$$K^*(t) = \frac{1}{t} \int_0^t K(x) dx$$

If  $t \rightarrow \infty$ , both  $K(t)$  and  $K^*(t)$  have the same limit, namely,  $K$  defined in (2.10).

### 2.2.3 Coefficient of Interval Availability

If the duration of the system's task is not negligibly small, we speak about the coefficient of interval availability, that is, the probability that at a time  $t$  the

system is found in an up state and will not fail during the performance of a task of length, say  $t_n$ . Denote this index  $Rit, r_n$ ):

$$i?(M_0)=X(r)/\gg(/_0ir)$$

We will consider this index later in more detail but now we note that  $P(r_0[r] = P(t_{r}))$  only when  $Fit$  is exponential.

If the system is not a socket with renewal unit, the situation is more complicated. In Chapter 7 we will illustrate this statement on a duplicated system consisting of two identical renewal units.

## 2.2.4 Mean Time Between Failures and Related Indexes

**Mean Time Between Neighboring Failures** In general, the mean time to a first failure  $T_{(1)}$  differs from the mean time from the first repair termination to the second failure  $T_{(2)}$ , and so on. In other words, all intervals of failure-free operations  $T_{(k)}$ ,  $k = 1, 2, \dots$ , may be different. We consider several typical situations.

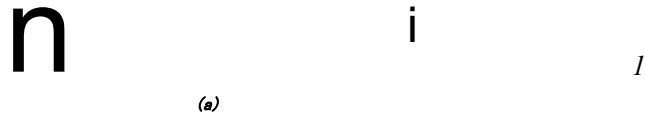
For a socket, or a one-unit system, the MTBF coincides with the MTTF because a new unit, put into the socket, is supposed to be identical to the failed one. But this equivalence of the MTTF and MTBF cannot be extended, even for the simplest two-unit system.

Consider a Markov model of a redundant system of independent and identical units [in other words, both  $F(t)$  and  $G(t)$  are exponential]. Assume that we know how to compute the mean time to a forthcoming failure for this system for the following two cases: starting with the state when two units are up ( $T^{(2)}$ ), and starting with the state when only one unit is up ( $T^{(1)}$ ) (see Figure 2.5).

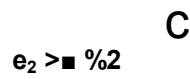
On average, the time to failure from state 2 is larger than the TTF from state 1. Indeed, in order to fail starting from state 2, the system must first enter state 1 and then from this state transit to the failure state. (Of course, from state 1 the system might even return to state 2 again.) In other words,  $T^{(2)} = T^{(1)} + T^*$  where  $T^*$  is the time of the system staying in state 2 until entering state 1.

In this particular case all of the remaining intervals of failure-free periods are i.i.d. r.v.'s, because for this particular Markov process all of the initial conditions are the same.

Now consider the behavior of a series system of two independent units. For simplicity, suppose that the repair time is negligible. Let each unit have a normal d.f. of TTFs with a mean  $T$  and a very small variation coefficient. If at the moment  $t = 0$  both units are new, then the first and second failures of the system are expected to appear close to each other and are around  $t = T$ . If the random TTFs of the units are denoted by  $\xi_1$  and  $\xi_2$ , respectively, then  $T_{(1)} = \min\{\xi_1, \xi_2\}$  and  $T_{(2)} = \max\{\xi_1, \xi_2\} - \min\{\xi_1, \xi_2\}$ . If the r.v.'s  $f_s$  are as



(a)



(b)

Figure 2.5. Examples of two time diagrams for a two-unit system: (a) a system of independent units; (b) a system working until both units have failed.

described,  $E\{t\}$ , and  $T_{[2]}$  has the order of  $\sqrt{t}$ . By assumption,  $A \ll T$  (see Figure 2.6).

The next couple of failures are expected to appear close to  $t = 2T$ , but the expected deviation is  $\sqrt{2}$  times larger than the initial deviation. For the case  $A \ll T$ ,

$$r_{pi} \approx \sqrt{t} \quad (2.11)$$

and

$$r_{(4)} = \max(\sqrt{t} + \sqrt{t}, \sqrt{t}) \approx \sqrt{t} \quad (2.12)$$

[Notice that strictly speaking (2.11) and (2.12) should be expressed in a more complicated way. We must take into account the mixture of the failure flows of the two sockets. This is explained by the appearance of extremely small and extremely large r.v.'s.]

Thus, if  $T_{[1]} > T_{[2]}$ , then  $T_{[2]} < T_{[3]}$  and  $r_{[4]} < r_{[3]}$ . At the same time,  $T_{[1]} > T_{[2]}$ , and  $r_{[2]} < r_{[4]}$ . The process continues in the same manner for larger numbers of interfailure intervals.

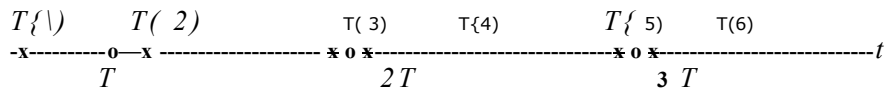


Figure 2.6. Example of the failure flows of a series system of two units.

With  $rt > 1$ , for any variance  $a^2$ , the value  $a$  begins to be larger than  $T$ . This leads to a strong mixture of moments of failure of both sockets of the system. In theoretical and practical terms, this means that  $T_j^{(j)} \sim +$  and, moreover,  $T_{[n]} \rightarrow \infty$  with  $n \rightarrow \infty$ . Thus, even for the simplest two-unit system, all MTBFs are different (though they may have the same asymptotical value). More complex cases appear when we consider a series system of more than two units. Notice that in reliability theory the term *MTBF* is usually used for the stationary regime, that is, as  $t \rightarrow \infty$ .

There are other indexes used in reliability theory which are connected with the time to failure. One of them is the instantaneous MTTF at time  $t$ . This is the mean time to failure from a specified moment  $t$  under the condition that a failure has happened just at this moment. From qualitative arguments it is clear that for  $t$  comparable with  $T$  this new index will differ from the MTTF. We remark that for the stationary regime, the values of both of these indexes coincide.

To conclude the discussion about the MTBF, we must emphasize that each time one should understand what kind of a particular TTF is under consideration. If we again regard a renewal series system of  $n$  units, each of them with a normally distributed TTF with a very small coefficient of variation, then we have the following:

1. For the MTTF

$$T \sim \min_{1 \leq i \leq n} T_i$$

2. The next  $n - 1$  MTTFs might be extremely small depending on the number of units and the smallness of the variation coefficient:

$$j(I)_{-E} \{t^{1+i}\}$$

where  $t^{(i)}$  is the  $i$ 'th-ordered statistic,  $1 < i < n - 1$ . A possible behavior of  $t^{(k)}$  for a series system is presented in Figure 2.7.

3. The stationary MTTF for any recurrent point process with continuous distributions of TTFs is

$$T = \lim_{k \rightarrow \infty} T^{(k)}$$

This value is the limit for  $T^{(k)}$  when  $k \rightarrow \infty$ .

In practice, we are often interested in the mean time of a failure-free operation starting from some specified moment  $t$ . In the theory of renewal



$\lambda$  ----- $\lambda$  -----  $\lambda$  ----- $\lambda$  -----  $\lambda$  ----- $\lambda$  -----  $\lambda$  ----- $\lambda$  -----  $\lambda$  ----- $\lambda$  -----  $\lambda$  ----- $\lambda$  -----  $\lambda$  ---  
 $J<0) -jil) j(2) J(3) jii] fi 5) J<6] y(7) y<8> j«9) jilO)$

Figure 2.7. Example of changing the mean time between failures depending on the current number of failure for which this value is evaluated.

processes, this value is called the mean residual time. In general, this index differs from the MTTF and any version of the MTBF.

If  $t \rightarrow \infty$ , for a recurrent process this index differs from both of those mentioned above. The exception is a Poisson process for which all three indexes coincide.

## 2.3 SPECIAL INDEXES

Now we consider some special reliability indexes for repairable systems. These indexes are nontraditional: they describe not a failure-free operation but rather a successful operation during a specified time. In some sense, there is no "local" failure criterion. The determination of a successful or unsuccessful operation is made, not at the moment of a current failure, but only after the completion of the entire system's performance during an acceptable operational time. This means that some interruptions of the system operation might be considered as not being destructive.

### 2.3.1 Extra Time Resource for Performance

Sometimes a system has some reserve time to perform its task; that is, the interval of time  $\theta_0$  given for the performance of the system operation is more than the time  $t_0$  required for a successful operation.

Examples of such situations can be taken from different areas of applications: conveyer production lines, electronic equipment with special power supplies, a computer performing routine calculations not in real time, and so forth. (Other detailed examples will be provided below.) In all of these cases not all failures of the system lead to the failure of the overall system's performance.

Consider a computer performing a computational task whose duration is  $t_0$ . The computer has a resource of time  $\theta_0$  for its performance which is

larger than the required time  $t_0$ . Random negligibly short interruptions (errors) may appear, each of which will destroy the results of all of the performed calculations.

In this case the probability of success can be written as

$$\Pr\{\text{at least one } t_k > t_0 \mid k: t_k \in \theta_0\} \quad (2.13)$$

Let the computer's task be segmented into phases and assume that the computer works in a restarting regime. After the completion of each phase, all intermediate results are put into the computer's memory. Each short failure has destroyed only the very last phase of the overall solving task. After the error has been found, the calculations for this particular phase are repeated. We do not give the formal definition of the corresponding reliability index but it is understandable that this index may be defined with the help of (2.13). We will consider this case in the special section dedicated to time redundancy.

### 2.3.2 Collecting Total Failure-Free Time

Suppose a system is required to accumulate some given amount of successful operating time during some given period  $\theta_0$  for the successful performance of the task. The probability of success can be written as

$$(2.14) \quad \Pr\left\{\sum_k t_k > \theta_0 \mid k: t_k \in \theta_0\right\}$$

As an example, we may again consider a computer system with a restarting regime for which each failure takes some time to repair but the operation of the system can be continued without loss of the previously obtained results. This situation may be observed if the restarting phases are very short, so there is practically no loss of the intermediate results but the system needs some time for restoration.

A close phenomenon occurs when one considers the transportation of some load. Failures and consequent repairs may only delay the termination of the task, but will not lead to a total failure. Of course, if the total idle time exceeds some limit, the task should be considered as not fulfilled (e.g., in the transportation of fresh food).

This index can obviously be written as

$$(2.15) \quad \Pr\left\{\sum_k x_k < \Psi_0 \mid k: x_k \in \Psi_0\right\}$$

where is the specified during the period  $\theta_0$ .

allowable total down time



### 2.3.3 Acceptable Idle Intervals

100 RELIABILITY INDEXES

Some systems possess the property of *time inertia*: they are insensitive to short breakdowns. As an example, consider a responsible computer system which has an independent power supply to prevent the system from occasional short failures of the common power system. In this case, if a failure of the main power system occurs, the computer system can operate with the help of this special power supply. Another example can be represented by a multistage conveyor system with an intermediate storage of spare subproducts in the case of a breakdown in the previous stages.

Thus, roughly speaking, an operational interruption of any such system can be noticed only if the duration of the down time  $x_k$  exceeds some specified value  $j_{c0}$ . In this case the reliability index is

$$Pr\{x_k < j_{c0}\} = e^{-\lambda j_{c0}} \quad (2.16)$$

In real life we meet more complicated situations. For example, a redundant power supply may demand substantial time to recharge, and this fact must be taken into account.

Of course, some combinations of the listed criteria for a system's failure may be considered. Some of them will be presented later.

## 2.4 CHOICE OF INDEXES AND THEIR QUANTITATIVE NORM

### 2.4.1 Choice of Indexes

The problem of choosing a reliability index arises before an operations research analysis. The solution of this problem depends on the nature of the object to be analyzed, its operations, and its expected results.

Depending on the operational level of the system, reliability indexes can be divided into two groups: operational and technical. If we deal with a system performing its individual and independent operation with a concrete final output, the reliability indexes should characterize the system's ability to perform its operation successfully. Such indexes are called *operational*.

If we deal with an object that is a subsystem and only performs some functions that are necessary to fulfill the operation of the system as a whole, the reliability indexes may be auxiliary. We can express the operational indexes of the system as a whole through these indexes. Such indexes are called *technical*. They are used to describe the reliability of a system's components and parts.

Starting with operational indexes, consider a computer that can be used to perform several quite different operations. The computer which is used for routine calculating tasks may be characterized with the help of an average percentage of useful operational time. The availability coefficient is the appropriate reliability index in this case.

The same computer used for supporting a long and noninterrupted technological process can naturally be characterized by the *probability of a failure-free operation* (PFFO).

If the computer is used for an automated landing system in an airport, and the duration of each operation is negligibly small in comparison with the computer's mean time to failure (MTTF), the reliability index should reflect the number of successfully served airplanes. In this case the availability coefficient is also the most appropriate reliability index.

If the same computer is unrepairable, for instance, its task consists of collecting and processing information in a spy satellite, the best characterization of it is the MTTF.

In all of these cases the reliability index corresponds to the system predestination and to the nature of its use.

Now consider an example when a reliability index is used to characterize "inner" technical abilities. Consider two identical computers connected in parallel. The natural reliability index for this duplicated system is the PFFO. In this case each computer is only a subsystem taking part in the performance of a system's operation. What should we know about each separate computer to characterize this system? To compute the complex PFFO, one needs to know the probability distributions of both the time to failure and the repair time of the computer, as well as the parameters of these distributions. The distribution itself is *not* an index; it is a function. But parameters of the distribution can be considered as *technical* reliability indexes of the computer. These parameters have no relation to a system's operation, they only reflect an ability to work in general.

Note that the type of reliability index chosen does not depend on the responsibility of the performed operation. The responsibility of the system's operation defines the level of requirements but is not part of the nomenclature of reliability indexes.

When choosing reliability indexes we should take into account the following simple recommendations based on common sense:

## 2.4.2 Required Reliability Level

100 RELIABILITY INDEXES

The problem of choosing the level of reliability is a very complex one. In practice, this problem is usually solved on the basis of engineering experience. For the purposes of determining reliability requirements, equipment may be divided into three groups by its "level": systems, subsystems, and units (components).

A system is considered as an object with its own goals of performance. A system performance is evaluated with the help of operational reliability indexes which are a measure of its success.

A subsystem is a more or less independent part of the system. It is considered to be an assembly of objects within a system. Each subsystem performs functions that are necessary for the operation of the system as a whole. The system's subsystems can be characterized by operational indexes if their functions can be measured with independent indexes or by technical indexes if these indices are used to express the system's performance effectiveness index.

A unit, or a component, is the smallest indivisible part of an object. The term *unit* is sometimes also used as a generic term for one physically separate item. In general, the term *component* is usually used for the smallest technological part of an object: electronic components, mechanical details, and so on.

The only problem which can be formulated as a mathematical problem is the assignment of reliability requirements among subsystems (parts of the system) when the requested level of reliability is known for the system as a whole. In this case the problem is reduced to the problem of the optimal allocation of some resources used for the improvement of reliability. The technical aspects of the problem will be considered in Chapter 10. Here we explain the nature of the problem.

Consider a system consisting of  $N$  independent subsystems. Assume that the probability of successful operation of the system as a whole must not be less than  $R_0$ . Each subsystem can be designed with different levels of reliability. Such levels depend on the expenditure of some kind of resource, for example, money. Suppose that we know all functions  $P_k(c)$  which reflect how the reliability index of the  $k$ th subsystem increases as a function of the expenditure of the resource  $c$ .

If the system's reliability index can be represented as

$$p(c_0) = \prod_{k=1}^n p_k\{c_k\}$$

and the value of  $C_0$  is specified, then the problem is to find the optimal allocation of the total resource  $C_0$  in such a way that the resulting system index is maximal, that is, find  $C^3 = \{C_1, C_2, \dots, C_n\}$  such that

---

3 The chosen indexes must allow one to analyze them with the help of analytic or computer models at the stage of system design.

- The total number of reliability indexes chosen for a system's characterization should be as small as possible.
- The chosen indexes should be easily interpreted.

CHOICE OF INDEXES AND THEIR QUANTITATIVE NORM 103

$$P(C^*) = \max_{I} \min_{k} \{ P_k(C_k) \mid E Q < C_0 \} \quad \& \quad N \quad I$$

- 
- The indexes should allow one to formulate clear requirements on reliability.
  - The indexes must allow one to estimate the achieved reliability level after field tests or experimental exploitation.
  - Complex "integrated" indexes must be avoided: various "convolutions" and "weightings" of different indexes usually have no real meaning.

Of course, we might also formulate the inverse problem: to design a system with a required level of reliability, for example,  $R_0$ . Then the problem may be reformulated as

$$P(C^*) = \min_{1 \leq k \leq N} \left( \sum_{1 \leq k \leq N} P_k(C_k) zR \right)$$

A solution to both of these problems is presented in Chapter 10.

Unfortunately, even such simple problems cannot usually be solved in practice because the functions  $P_k(C_k)$  are often unknown. In such cases we usually use heuristic methods based on a proportional distribution of reliability "quotas" among the system's units.

What must one do in the general case when it is necessary to assign a reliability level to the system as a whole? In our opinion, there is only one thing to do: perform an evaluation based on engineering experience. Prototypes can be used for comparison with the designed system and, on the basis of this, the decision about a possible or desirable reliability level might be made.

Naturally, if one fixes the amount of available resources for the production of some type of technical system, then we not only have to solve the problem of an optimal reliability level, but we must also answer the question of how many such systems we intend to produce? In turn, the number of systems of some chosen type depends on the number of other "competing" systems in the same area of use or utility. Assume that we are considering the design of a new type of jet. First of all, too high a level of reliability will demand a high level of expenses for the production of each jet and, as a consequence, will lead to a decrease in the total number of jets produced. It is clear that it is useless to have only one extremely highly reliable jet and it is equally unreasonable to have a large number of jets, each of which has a very low reliability. To choose "a golden middle" is a problem which lies outside the scope of mathematics and even outside the scope of engineering. The only way to solve this problem is to rely on expert's opinions and traditions.

But the experts' opinions are also not isolated. Taking into account all considerations concerning this particular type of a jet, experts have to think about the number and reliability of other jets owned by the airline. But this total number depends on a specific situation, considering the transportation system of the country as a whole. In turn, it depends on the level of the national economy. The level of the national economy depends on a number of unformulated and nonformulated factors: the political stability of the country, the external situation in the world, and so forth. Thus, we are convinced that any attempt to try to solve this problem in some "precise" sense is doomed.

But then one may ask: Why use mathematical methods at all? Why not rely on experts' opinions to solve all problems of this kind? The answer is that



mathematical methods of analysis of situations help one to make logically strong decisions; mathematical models of technical systems help one to understand the nature of systems being designed. We begin to make local solutions in optimal ways. This leads to a kind of process of "natural selection." As in nature this process allows for the survival of only those who have best adapted to existing environments. And, in this situation, those technical systems which are "locally optimally designed" have a better chance to "survive" under currently existing circumstances.

We now consider possible methods of establishing reliability.

**System Level** Consider two principal cases. One of them consists in the use of practical experience and engineering intuition. Mostly it is based on an analysis of prototypes of the new technical system to be investigated. This method needs no special comments.

Practically, the only time a system's reliability requirement appears is if:

- The system's outcome can be measured in cost units, that is, in the same units as the costs of the system's design, production, and maintenance.
- The system's structure and operational modes are well known in advance.
- Necessary statistical data for all components are determined with a satisfactory degree of confidence.

In this case the designer has an opportunity to compare  $M$  different variants of the system's design and to choose the most profitable one. The objective function of the system's performance for the  $k$ th variant can be written in the form

$$F_k(R) = E_k(R) - \gamma C_k(R)$$

where  $R$  is the system's reliability index,  $E_k(R)$  is the outcome of the  $k$ th variant of the designed system, and  $C_k(R)$  is the expenditure needed to design, produce, and maintain the system with index  $R$ ,  $1 < k < M$  and  $\gamma$  is a dimensional coefficient analogous to a Lagrange multiplier. The value of  $R$  depends on the structure of the  $k$ th variant,  $S_k$ , and on the reliability indexes of the subsystems used,  $r_j^k$ ,  $1 < i < n_k$ , where  $n_k$  is the number of subsystems in the  $k$ th variant of the system. Thus,  $R$  itself can be written in a general form as

$$R = R(S_k, r_j^k) \quad (1 < k < M, \quad 1 < n_k)$$

For simplicity, suppose that all functions are differentiable. Then the optimal level of the reliability index  $R$  can usually be determined by solving the equation

$d$

## 1 00 RELIABILITY INDEXES

or, equivalently,

$$\frac{dE_k(R)}{dR} = \frac{dC_k(R)}{dR}$$

Each optimum can be evaluated and then the variant  $k$  with the highest value of  $F_k(R_k^{opt})$  is selected. Unfortunately, such an ideal situation appears extremely rare in engineering practice.

**Subsystem Level** Suppose that the system's reliability requirement is specified. Then the problem is to distribute the given value of the index over the subsystems. We consider several cases, each of them representing different information concerning the system's structure and the availability of statistical input data.

**Uniform Allocation of Requirements** This method is usually used when one can imagine only the approximate size of a subsystem of the main system. A reliability index  $R$  of a probabilistic nature (e.g., the probability of success or the availability coefficient) is specified for the system as a whole. The simplest assumption is that the system has a series structure and consist of  $n$  subsystems. The reliability requirement for each subsystem is then given by

$$R_i = R^{1/n} \quad 1 < i < n$$

Clearly, if subsystem indices are chosen in such a way, the system reliability index equals  $R$ .

If requirements can be specified as the system's MTTF  $T$ , we can choose for each  $i$ th subsystem

$$r_i = nT$$

This means that we additionally assume that the TIT of any subsystem has an exponential distribution.

**Allocation in Proportion to the Number of Units** Assume that the same conditions exist as before, but in addition subsystem  $i$  consists of  $n_i$  units which are essentially similar in their complexity. In this case the requirement (in terms of the probability of success) should be chosen to be

$$R_i = R^{n_i/n} \quad 1 \leq i < n \quad (2.17)$$

where

$$a_i = \frac{n_i}{n}$$



1 00 RELIABILITY INDEXES

When all distributions are assumed to be exponential, the requirements can be formulated in terms of the MTTF  $T_i = \frac{1}{a_i}$

This method can be useful if different subsystems are designed by different subcontractors. It is reasonable to specify "softer" demands for more complex subsystems.

**Allocation in Proportion to the Expected Failure Rate** Suppose a designer has more complete information about the system: the unit failure rates are known (perhaps from previous experience), and the hypothesis about the exponential distribution can be considered valid. In this case the previous method can be improved. We can use (2.17) but we substitute  $a_i$  defined by

$$a_i = \frac{\sum_{j=1}^n \lambda_j f_{ij}}{\sum_{j=1}^n \lambda_j V_{ij}}$$

$\sum_{j=1}^n \lambda_j < M$

where  $M$  is the number of types of units and  $n$  is the number of units of the  $i$ th type in subsystem  $i$ .

**Optimal Allocation of Reliability Requirements** This method is applied if we know the system's structure  $S$  and can predict the cost-reliability trade-off for each subsystem. The problem is to find the values of  $1 < r < n$ , that yield the required reliability index at the lowest cost. This problem can be written in mathematical terms as

$$\min_{\lambda_i} \sum_{i=1}^n C_i(\lambda_i) \mid \sum_{i=1}^n \lambda_i f_{ij} \geq R_j(S); \sum_{i=1}^n \lambda_i \leq M$$

where  $C_i(\lambda_i)$  represents the subsystem's costs as a function of its reliability and  $S$  is the conditional notation of the system structure. For instance, if we consider a series system, the reliability function can be represented as

$$R(S) = \prod_{i=1}^n R_i(\lambda_i)$$

In other words, the optimal allocation of reliability requirements between subsystems is a type of optimal redundancy problem (see Chapter 10).

**Reliability Requirements for a Component** Almost all equipment components in engineering are of general usage. The only method in this case is based on the "natural selection" principle. In other words, the better and cheaper components among existing ones survive the competition in a technical and economic environment. And, at the same time, new components appear and replace technical "dinosaurs."

## CONCLUSION

This chapter does not need any special comments. In one form or another, reliability parameters are discussed in any book on reliability engineering or theory. As examples, we refer the reader to the wide list of general references at the end of this book.

The nomenclature of reliability indexes in a systematic and structured form can be found in Kozlov and Ushakov (1970) and Ushakov (1985, 1994). The methodological problems of choosing indexes and quantitative requirements in reliability engineering are discussed in Gnedenko, Kozlov, and Ushakov (1969).

## REFERENCES

- Gnedenko, B. V., B. A. Kozlov, and I. A. Ushakov (1969). The role and place of reliability theory in engineering. In *Theory of Reliability and Queuing Systems* (in Russian). Moscow: Nauka.
- Kozlov, B. A., and I. A. Ushakov (1970). *Reliability Handbook*. New York: Holt, Rinehart, and Winston.
- Kozlov, B. A., and I. A. Ushakov (1975). *Handbook on Reliability of Radio and Automation Systems* (in Russian). Moscow: Soviet Radio.
- Ushakov, I. A., ed. (1985). *Reliability of Technical Systems: Handbook* (in Russian). Moscow: Radio i Sviaz.
- Ushakov, I. A., ed. (1994). *Handbook of Reliability Engineering*. New York: Wiley.

## EXERCISES

- 2.1 Prove that the mean value of a nonnegative r.v.  $v$  with distribution  $Fit$  can be expressed in the form of (2.4) which is equivalent to (2.3).
- 2.2 A system has an exponentially distributed TTF with parameter  $A$ . The operation to be performed also has a random duration. Find the

1 00 RELIABILITY INDEXES

- probability that the system successfully performs its operation if
- (a) the operation duration is distributed exponentially with parameter  $a$ ;
  - (b) the operation duration is distributed normally

$$f_N(x|a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

with mean equal to  $a$  and variance equal to  $a^2$ . We also assume that  $A_{CT} = 1$ .

2.3 Build the graph of the failure rate for the mixture of two exponential distributions (2.6) with the following parameters, respectively,

- (a)  $A_1 \ll 1$  [1/hour],  $A_2 \ll 1$  [1/hour];
- (b)  $A_1 = 0.5$  [1/hour],  $A_2 = 1$  [1/hour];
- (c)  $A_1 = 2$  [1/hour],  $A_2 = 1$  [1/hour],

SOLUTIONS

1.1

$$r = f \cdot t \cdot dF(t) = f \cdot t \cdot d[1 - e^{-\lambda t}] = -f \cdot t \cdot dP(t)$$

and, after integrating by parts,

$$- \int t dP(t) = -tP(t) + \int P(t) dt = r \cdot t \cdot O dt = A^1 \sim A^0 \cdot 3 dt$$

1.2

(a) Using (2.4), one writes

$$\int_0^\infty \lambda e^{-\lambda x} dx = \int_0^\infty \lambda e^{-\lambda x} dx = 1$$

(b) First of all, consider the given conditions. Almost all "probabilistic mass" is concentrated in a relatively very compact area related to the MTTF of the system. This means that in this area the exponential function can be successfully approximated by the set with at most two terms:

$$1 - Ax + \frac{A^2 x^2}{2} \quad (A^*)$$

Thus, one has

SOLUTIONS 109

$$\begin{aligned}
 & \int_0^{\infty} e^{-\lambda x} f_N(x|a, <r) dx = r \int_0^{\infty} e^{-\lambda x} f_N(x|a, tr) dx \\
 & = \int_0^a f_N(x|a, a) dx - \int_a^{\infty} f_N(x|a, cr) dx + \int_0^{\infty} f_N(x|a, a) dx \\
 & = 1 - aX + \dots
 \end{aligned}$$

In the first case there is no mixture at all; the second and third cases differ only by the scale. In general, one can write

and 
$$f(x) = p_1 e^{-\lambda_1 x} + p_2 k_2 e^{-\lambda_2 x}$$

$$\begin{aligned}
 & p_1 k_1 e^{-\lambda_1 x} + p_2 k_2 e^{-\lambda_2 x} \\
 & \mathbf{A}(l) \\
 & p_1 e^{-\lambda_1 x} + p_2 e^{-\lambda_2 x}
 \end{aligned}$$

(The numerical solution is left to the reader.)

## CHAPTER 3

# UNREPAIRABLE SYSTEMS

In this chapter we will consider the main types of unrepairable systems. The only type that we will not address is a general network structure, which will be considered in a later chapter.

### 3.1 STRUCTURE FUNCTION

For convenience in future mathematical explanations, let us introduce the so-called *indicator function*  $x_i$  for unit  $i$ :

$$x_i = \begin{cases} 1 & \text{if the } i\text{th unit is operating} \\ 0 & \text{otherwise} \end{cases} \quad (3.1)$$

Let us introduce a similar function for the system as a whole. This new function depends on all of the  $x_i$ 's, the system's structure, and the criterion of system failure that has been chosen:

$$f(x_1, x_2, \dots, x_n) \sim \begin{cases} 1 & \text{if the system is operating} \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

In reliability theory this function is called the *structure function* of a system. If each unit has two states—up and down—then a system of  $n$  units may have  $2^n$  different states determined by states of the individual system's units. The function (3.2) is determined by the system failure criterion.

Of course, system states may differ from each other by their level of operational effectiveness. This case will be considered in Chapter 8. Here we



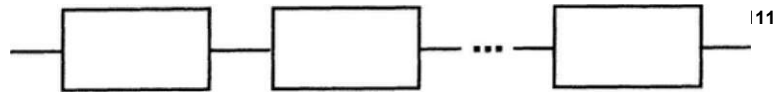


Figure 3.1. System with a series structure.

restrict ourselves to the case where a system has only two possible states: up and down.

From the definition of (3.2) it is clear that the  $x$ 's are Boolean variables and  $f(x_1, x_2, \dots, x_n)$  is a Boolean function. We also denote this function by  $f(X)$  where  $X = (x_1, x_2, \dots, x_n)$ . System reliability structures are often displayed as a two-pole network. One of the simplest examples of such a network is presented in Figure 3.1. The connectedness of this two-pole network is equivalent to the equality of the Boolean function (3.2) to 1.

Each unit  $i$  may be in state  $x_i = 1$  or  $x_i = 0$  in random. If each Boolean variable  $x_i$  is considered as a Bernoulli r.v., then  $E\{x_i\}$  is interpreted as the probability that unit  $i$  is in an up state, and  $E\{f(X)\}$  is defined as the probability of the system's successful operation:

$$P_{\text{sys}} = E\{f(X)\}$$

We consider only monotone functions  $f(X)$  for which  $f(X) > f(X')$  if  $X > X'$ . Here the inequality  $X > X'$  means that  $x_i > x'_i$  for all  $i$  and there is a strict inequality at least for one  $i$ . This assumption is very natural. Indeed, a unit failure generally will not improve a system's operational state. Therefore, if a system is down in state  $X$ , it cannot be up in state  $X'$  with some additionally failed units. (Of course, it is correct under the assumption that the system was correctly designed.) We emphasize that it relates only to systems whose operation can be described in terms of Boolean functions.

### 3.2 SERIES SYSTEMS

The series structure is one of the most common structures considered in engineering practice. A system with such a structure consists of units which are absolutely necessary to perform the system's operation: a failure of any of one of them leads to a system failure. Schematically, this structure is represented in Figure 3.1.

Of course, the series system in a reliability sense does not always correspond to a real physical series connection of the system units. For example, the parallel connection of capacities (Figure 3.2) subjected to failures of a shortage type corresponds to a series structure in reliability terms.

Let us denote the structure function of a series system as  $f(x_1, x_2, \dots, x_n)$ .

This function is

$$f(X) = \prod_{i=1}^n x_i \quad (3.3)$$

11 Figure 3.2. A parallel connection of capacitors which represents a series structure in a reliability sense.

where the symbol  $\cup$  denotes the Boolean product (disjunction). The same expression can be written in an equivalent form

$$a(X) = \min x_t$$

In reliability theory systems consisting of independent units are usually considered. In this case the computation of the probability of a successful system operation is easy. We are interested in the probability

$$\Pr\{a(x_1, x_2, \dots, x_n) = 1\} = E\{a(x_1, x_2, \dots, x_n)\} \quad (3.4)$$

For independent units (3.4) might be rewritten in two equivalent ways

$$\Pr\{a(x_1, \dots, x_n) = 1\} = \prod_{i=1}^n \Pr\{x_i = 1\} = \prod_{i=1}^n p_i \quad (3.5)$$

$$E\{a(x_1, \dots, x_n)\} = \prod_{i=1}^n E\{x_i\} = \prod_{i=1}^n p_i \quad (3.6)$$

Expressions (3.5) and (3.6) make the following statements true:

1. A series system's reliability decreases (increases) if the reliability of any unit decreases (increases).
2. A series system's reliability decreases (increases) if the number of units increases (decreases).
3. A series system's reliability is worse than the reliability of any of its units.

The first two statements reflect the monotonicity property.

Above we have considered the static case where probabilities are specified as constant. But the process of a system's operation develops over time, so it is reasonable to consider a random function  $x_i(t)$ :

$$x_i(t) = \begin{cases} 1 & \text{if the } i\text{th unit is operating at moment } t \\ 0 & \text{otherwise} \end{cases} \quad (3.7)$$

This function is monotone and nonincreasing over time for unrepairable units; that is, after a failure the unit cannot return to state 1. In other words,  $x_i(t + \Delta) \leq x_i(t)$  for any  $\Delta > 0$ . Thus, for the system as a whole, it follows that  $X(t + \Delta) \leq X(t)$ . From (3.5) it follows that

$$P(t) = \prod_{i=1}^n x_i(t) \quad (3.8)$$

Obviously, (3.5) and (3.8) can be written in a direct way from the verbal definition of a series system's successful operation:

$$\begin{aligned} P(t) &= \Pr\{\text{a series system operates successfully}\} \\ &= \Pr\{\text{all system's units are up}\} \\ &= \Pr\{\text{unit 1 is up, AND unit 2 is up, ..., AND unit } n \text{ is up}\} \\ &= \Pr\{\text{unit 1 is up}\} \Pr\{\text{unit 2 is up}\} \dots \Pr\{\text{unit } n \text{ is up}\} \end{aligned}$$

In a more general case, direct calculations must be used for obtaining the function  $P(t)$  for the system. But in one important particular case, when each  $P_j(t)$  is an exponential function, we can write a very simple expression

$$P(t) = \prod_{i=1}^n \exp(-\lambda_i t) = e^{-\sum_{i=1}^n \lambda_i t} \quad (3.9)$$

where

$$\lambda = \sum_{i=1}^n \lambda_i$$

Suppose a system consists of highly reliable units:  $\lambda_j(t) = 1 - e^{-\lambda_j t}$  where  $\lambda_j(t)$  is very small; for example,

$$\lambda_j(t) \leq \frac{1}{n}$$

11 ~~UNREPAIRABLE SYSTEMS~~ smallness of for each particular case depends on the number of units.] Then, for a system with units having arbitrary distributions of time to failure,

$$n_{vs},(0 = n \int_0^{\infty} [1 - e^{-\lambda t}]^{-1} dt \ll \frac{1}{U_{isn}} \quad (3.10)$$

The error of the calculation in this case will not exceed the value

$$\Delta, (0 - n \int_0^{\infty} [1 - e^{-\lambda t}]^{-1} dt < L e, (0 \ll i(0) < (?) [ \max e, (0)^2 \quad V^{\wedge} / t | S_{ii} S_n J \quad (3.11)$$

Let us consider a particular case. Suppose a series system consists of  $n$  units, each of which has a continuous failure distribution with a nonzero first derivative at  $t = 0$ . Suppose the system is operating during a small period of time  $t_0$ . The Taylor series restricted to the first term is

$$dt \frac{F(t)}{(-)} \quad 'o = l(0)' o$$

Note that, at  $t = 0$ ,

$$m - m - m \quad l(0)$$

Then, for a series system consisting of a large number of highly reliable identical units with an arbitrary d.f.  $F(t)$ ,

$$" [1 - A(0) <_o]^n - \exp[-nA(0) *_o] \quad (3.12)$$

If the units are different but some of them have distribution functions  $F_i(t)$ ,  $i \in a$ , with nonzero first derivatives at  $t = 0$  equal to  $A_i(0)$ , then for small  $t_0$

$$P^{\wedge} t C_o \sim n(1 - A_i(0) r_o) = \exp \quad \blacksquare 'o L A_i(0) \quad (3.13)$$

Of course, we assumed that  $|a| \gg 1$ ; that is, the number of distributions with a nonzero derivative is large. Therefore, we see one more example of an exponential distribution in reliability theory.

If the distribution of a unit's TTF is such that  $dF_i/dt' = 0$  for  $i < k$  and  $d^k F_i/dt'^k = a$ , then

$$P(t) = \left[ 1 - \left( \frac{3}{4} - \frac{1}{4} \right) \right]$$

For large  $n$  one can write the approximation

where  $A = na$ . Thus, this series system has a Weibull-Gnedenko distribution of time to failure. One practical example of such a system concerns a set of bearings in a machine. Another example will be presented in Section 3.4. In the ideal case, if all of the series system's units have a constant TTF, that is, a degenerate distribution of the type

$$f(t) = \begin{cases} 1 & \text{if } t \leq T_i \\ 0 & \text{otherwise} \end{cases}$$

then  $P(t)$  coincides with the  $p_i(t)$  of the worst unit, that is,

$$P(t) = \begin{cases} 1 - \prod_{i=1}^n (1 - p_i(t)) & \text{if } t \leq \min_i T_i \\ 0 & \text{otherwise} \end{cases} \quad (3.16)$$

Of course, such a distribution does not exist in a real life. (Mathematics always deals with ideal objects!) But normally distributed r.v.'s with very small coefficients of variation can be considered as "almost constant" or "almost nonrandom."

Now consider the MTTF of a series system. For any series system the random TTF, say  $Y$ , can be expressed through the random TTFs of its units  $\{y_i\}$  in the following way:

$$Y = \min_{i=1, \dots, n} y_i \quad (3.17)$$

)

The MTTF can be found in a standard way as

$$r_{sys} = E\{y\} = \int_0^\infty P(y > t) dt \quad (3.18)$$

where  $P(y > t)$  is determined above. For an exponential distribution,  $p(t) = \exp(-At)$ ,

11 6 UNREPAIRABLE SYSTEMS

$$T_{sys} = \int_0^{\infty} \exp(-r \frac{E}{V} A_i) dt = \frac{L_i}{L_i - V - I} \bar{T}_i \quad (3.19)$$

where  $L_i$  is the MTTF of the  $i$ th unit.



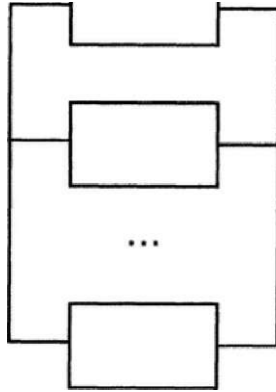


Figure 3.3. System with a parallel structure.

For units with a degenerate distribution

$$T_{\text{sys}} = \min T, \tag{3.20}$$

that is, the MTTF of the system equals the MTTF of the worst unit.

### 3.3 PARALLEL STRUCTURE

#### 3.3.1 Simple Redundant Group

Another principal structure in reliability theory is a parallel connection of units (Figure 3.3). This system consists of one main unit and  $m - 1$  redundant units. We call such a system a *simple redundant group*. A system failure occurs if and only if all of the system's units have failed. In other words, the system is operating as long as at least one of its units is operating. Sometimes parallel systems are called *systems with an active (or loaded) redundancy*. Thus, the redundant units are in a working regime during the entire time of the system's operation. A main feature of active redundancy is that all of the reliability characteristics of the redundant units are assumed to be the same as the system's operational units.

The structure function of a parallel system,  $\phi(X)$ , is

$$\phi(X) \sim \bigvee_{i=1}^m \phi_i(x_1, x_2, \dots, x_m) \tag{3.21}$$

where the symbol  $\bigvee$  denotes Boolean summation (conjunction). The same



expression can be written in an equivalent form:

$$S(X) = \max_j t_i$$

**ISI'sm**

For further discussion we need to acknowledge the following result.

**DeMorgan's Rule** For two Boolean variables  $x$  and  $y$ , the following equivalences are true (see the exercises):

$$x \wedge y = (x \vee y)' \tag{3.22 a}$$

$$x \vee y = (x' \wedge y')' \tag{3.22 b}$$

$$x \vee y = (x' \wedge y')' \tag{3.22 c}$$

$$x \wedge y = (x' \vee y')' \tag{3.22 d}$$

All of these equivalences express the same property but in slightly different form. The most important one for us is (3.22a). If one considers a series system of two units  $x$  and  $y$ , and "1" means an up state, then  $x \vee y = 1$  means unit  $x$  and/or unit  $y$  are in an up state; that is, the system is in an up state. At the same time,  $x \wedge y = 0$  means unit  $x$  and unit  $y$  are in a down state; that is, the system is in a down state. It is clear that these two events are complementary. To prove (3.22), one may use a *Venn diagram*. This diagram graphically depicts random events, their sum and intersection, complementary events, and so on. A simple case with two events  $A$  and  $B$  is presented in Figure 3.4. The proof of (3.22c) one can find in Figure 3.5.

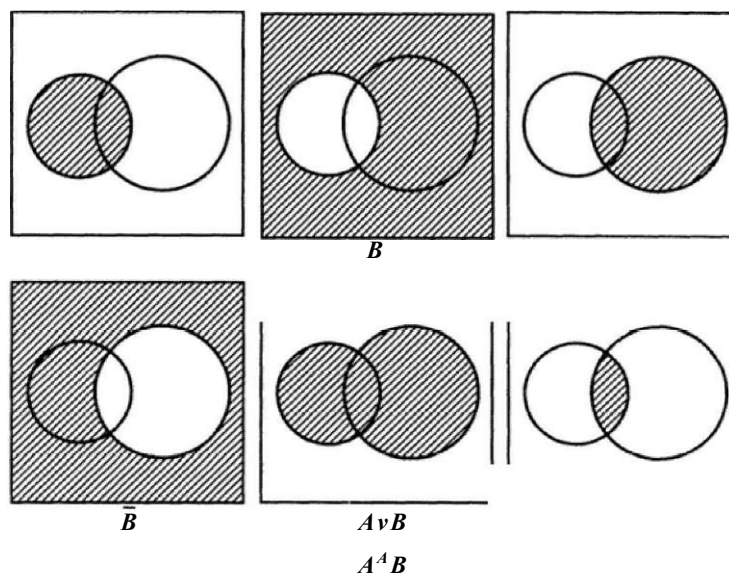


Figure 3.4. Samples of main Venn diagrams.

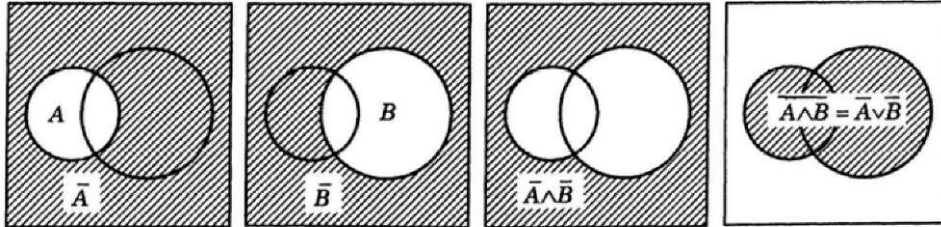


Figure 3.5. The proof of the DeMorgan Rule with the help of the Venn Diagram.

From the above-given particular forms of DeMorgan's rule, the following generalizations can be easily obtained:

$$u^*, \quad \frac{-n}{1 \text{ SiSn}} \quad (3.23a)$$

$$n^{**} \quad \frac{-u}{1 \wedge \text{JSn}} \quad (3.23b)$$

$$U \quad \frac{-n}{1 \text{ sisw}} \quad (3.23c)$$

$$\circ = U \quad (3.23d)$$

**ISiSH**

These latter statements can be proved by induction and we leave their proofs to the exercises.

Another (almost purely verbal) explanation of (3.23) follows from the definition of a parallel system's failure which was given at the very beginning of this section:

$$\begin{aligned} & \Pr\{\text{a parallel system operates successfully}\} \\ &= \Pr\{\text{at least one unit operates successfully}\} \\ &= \Pr\{\text{unit } x_1 \text{ is up, OR unit } x_2 \text{ is up, ..., OR unit } x_m \text{ is up}\} \\ &= \Pr\{U^* = 1\} \end{aligned}$$

At the same time,

$$\begin{aligned} & \Pr\{\text{a parallel system has failed}\} \\ &= \Pr\{\text{all of its units have failed}\} \\ &= \Pr\{\text{unit } x_1 \text{ is down, AND unit } x_2 \text{ is down, ..., AND unit } x_m \text{ is down}\} \\ &= \Pr\{n \text{ Xj} = e | n \text{ ei}^{\wedge} = n\} \end{aligned}$$

We note that if two events, say  $z$  and  $\bar{z}$ , are complementary, then

$$\Pr\{z = 1\} + \Pr\{z = 0\} = 1$$

Consequently,

$$\Pr\{U \text{ system} = 1\} = 1 - \Pr\{n \times A = 1 - \prod_{i=1}^n Q_i\} \quad (3.24)$$

Now the equivalence of (3.23) can be confirmed in an inverse way by the equality of the probabilities.

We repeat that a detailed inference was done above only from a methodological viewpoint to provide further discussion. Of course, it was enough to use a verbal definition of a parallel system of independent units and to write the final expression. Sometimes a different form equivalent to (3.24) is used

$$P_{\text{sys}} = P_1 + Q_1 P_2 + Q_1 Q_2 P_3 + \dots + Q_1 \dots Q_{m-1} P_m$$

$$= P_1 + h(P_2 + h(P_3 + \dots))$$

This expression can be explained as follows:

Pr{a parallel system operates successfully}

= Pr{the first unit is up, OR if the first unit has failed, the second one has not failed; OR if both of these units have failed, then the third one has not failed, OR ... }

If each of the system's units has an exponential TTF distribution,  $p_i(t) = \exp(-t/A_i)$ , for a highly reliable system where  $\max_i \lambda_i \ll 1/m$ , one can write  $\lambda_i(t) = \lambda_i A_i$ , and, finally,

$$\lambda_{\text{sys}} = 1 - \prod_{i=1}^m \exp(-\lambda_i A_i) \quad (3.25)$$

If each unit of a parallel system has a constant time to failure (a degenerate distribution of TTF), then

$$\lambda_{\text{sys}} = \begin{cases} \lambda_i & \text{for } i = 1, 2, \dots, m \\ 10 & \text{otherwise} \end{cases} \quad (3-26)$$

Now consider a parallel system's MTTF. For this system the random TTF ( $T_{\text{sys}}$ ) is expressed through the random TTFs of its units ( $T_i$ ) as

$$T_{\text{sys}} = \max_i T_i \quad (3.27)$$

Thus, this is equivalent to the statement that a parallel system operates successfully until the last failure of its units.

When each unit has an exponential distribution of TTF, an analytic expression can be derived. For this purpose write the probability of failure-free operation in the form

$$P(t) = \prod_{i=1}^m (1 - e^{-\lambda_i t}) \quad (3.28)$$

Integrating (3.28) gives

$$T_{\text{sys}} = \int_0^{\infty} P(t) dt = \int_0^{\infty} \prod_{i=1}^m (1 - e^{-\lambda_i t}) dt \quad (3.29)$$

If, at the same time, all units are identical

In this case the MTTF has the form

$$(3.30)$$

where  $T$  is the MTTF of a single unit. For large  $m$ , a well-known approximation for a harmonic set can be applied:

$$T_{\text{sys}} \approx T(\ln m + C) \quad (3.31)$$

where  $\mathbf{C}$  is Euler's constant:  $\mathbf{C} = .57712$ .

**TABLE 3.1 Dependence of a System's MTTF on Its Units' MTTF**

Number of Redundant Units	Relative Growth of the System MTTF
0	1
9	2.88
99	5.18
999	7.48
10 <sup>6</sup>	23.6

Formula (3.30) can be explained in a simple and understandable way with the use of the memoryless property of the exponential distribution. At the moment  $t = 0$  the system of  $m$  active redundant units has a failure rate  $A_m = mA$ . The first failure occurs in a random time  $Z_m$  with an exponential distribution with parameter  $A_m$ . After this failure the system consists of  $m - 1$  units, so its failure rate is now  $A_{m-1} = (m - 1)A$ . The second failure occurs in a random time  $Z_{m-1}$ , with an exponential distribution with parameter  $A_{m-1}$ . And so on, until the last unit has failed.

The total time of a successful system's operation consists of the sum of all these intervals, that is,  $T_{\text{sysl}} = E\{Z_1 + Z_2 + \dots + Z_m\}$ . Obviously, this result coincides with (3.30).

From (3.30) and (3.31) it follows that, at least theoretically, the use of active redundancy potentially allows one to construct a system with an arbitrarily large MTTF value. Of course, one needs to understand that such a mathematical model is strongly idealized. First of all, one must take into account the necessity to use a switching device which itself possesses a nonideal reliability. On the other hand, even with absolutely reliable switching devices, the growth of the system's MTTF is very slow. Several examples are shown in Table 3.1.

Hardly anybody would ever use such redundancy (even with absolutely reliable switches!) to improve the MTTF. But this kind of redundancy can be successfully used if one considers other indexes of reliability, for example, the probability of a system's successful performance. In this case if the initial value of  $q(t)$  is much less than 1, each new parallel unit decreases the system's unreliability level by the order  $q$ .

Note that for a nonexponentially distributed TTF with an increasing failure rate (i.e., for "aging" units), the growth of a system's MTTF" is even slower.

### 3.3.2 "k out of n" Structure

For some technical schemes one sometimes considers a special structure—the so-called "k out of n" structure, or voting system. In engineering practice such a system almost always consists of identical units. In this case the system

operates successfully if at least  $k$  out of its total  $n$  units are operating. The

structure function of the system is illustrated here by a simple example with "2 out of 3":

$$f(x_1, x_2, X_3) = (x_1 \vee x_2) \wedge (x_2 \vee X_3) \wedge (x_1 \vee X_3)$$

(This case is most often encountered in engineering practice.)  
 In general, the structure function of a "k out of n" structure can be written in the form

$$f(X) = \prod_{i \in E} x_i^k$$

where

$$x_i^k = \begin{cases} 1 & \text{if } x_i = 1 \\ 0 & \text{otherwise} \end{cases}$$

We will use an explanation based on combinatorial methods, avoiding the structure function. Considering a "k out of n" structure corresponds to the binomial test scheme, so

$$\Pr\{v=j\} = \binom{n}{j} p^j q^{n-j}$$

and, consequently, the probability of a system's failure-free operation equals

$$P_{\text{sys}}(0) = \Pr\{M \geq k\} = \sum_{j=k}^n \binom{n}{j} p^j q^{n-j} \tag{0-32}$$

$$P_{\text{sys}}(0) = \sum_{j=k}^n \binom{n}{j} p^j q^{n-j}$$

For a highly reliable system where  $q \ll 1/n$  from (3.33) one can easily write

The task of finding the MTTF of such a system for arbitrary unit failure distributions is not simple. One may use any numerical method. The only case where it is possible to obtain an analytical result is the case of the exponential distribution  $p(t) = \exp(-\lambda t)$ .

We will not integrate (3.32), but we will use the method described above. The system starts with  $n$  operating units and operates, on the average,  $1/\lambda$



units of time until the first failure. After the first failure, there are  $n - 1$  operating units in the system. They work until the next failure, on the average,  $1 / (n - 1)A$  units of time, and so on until the  $(k + 1)$ th failure has occurred. Thus, the system's MTTF equals

For an arbitrary distribution  $p(t)$  one should use a direct numerical integration of  $\Pr_V M$  or a Monte Carlo simulation.

### 3.4 MIXED STRUCTURES

Pure series or pure parallel systems are rarely encountered in practice. Indeed, mixed structures with series and parallel fragments are common. For example, a duplicate computer system may be used for monitoring a production line. Each of these two computers, in turn, is represented by a series structure, and so on.

A combination of series and parallel structures can generate various mixed structures. First, let us consider "pure" series-parallel and parallel-series types of structures (see Figures 3.6a and 3.7a) because they will be of interest in further discussions.

For these structures the following expressions can be easily written. For a parallel-series structure, one has

$$P_{PS}(t) = E \left\{ \bigcup_{1 \leq i \leq M} A_i(\mathbf{X}_i) \right\}$$

where

$$\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{iN})$$

or

$$\begin{aligned}
 P_{PS}(t) &= 1 - E \left\{ \bigcap_{1 \leq i \leq M} \overline{x_{i1} \wedge x_{i2} \wedge \dots \wedge x_{iN}} \right\} \\
 &= 1 - \prod_{1 \leq i \leq M} \left( 1 - \prod_{1 \leq j \leq N} p_{ij} \right) \tag{3.34}
 \end{aligned}$$

where  $N$  is the number of units in a series subsystem.

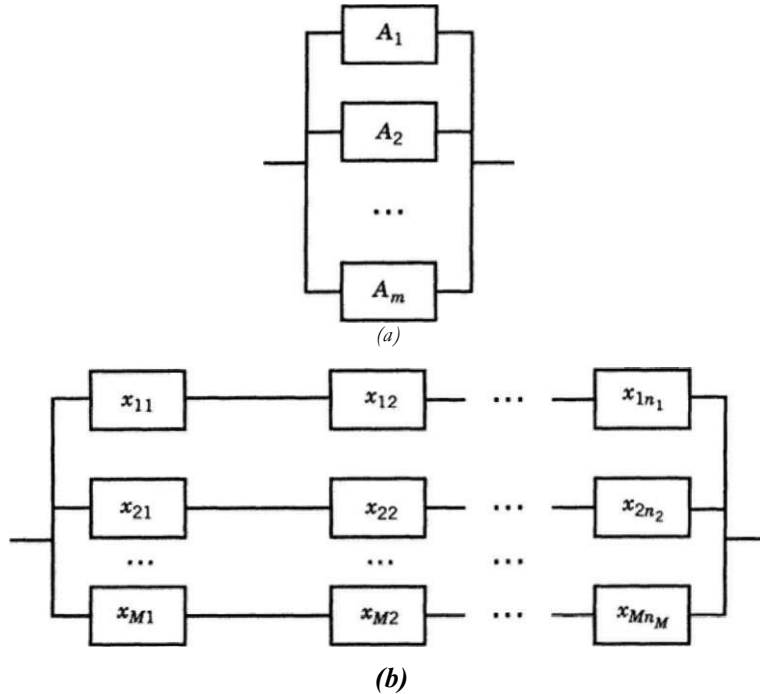


Figure 3.6. Parallel-series structure: (a) in an aggregate form; (b) in a detailed form.

For a series-parallel structure, one writes

$$/v_o = E\{n^*, -(*/)\}$$

$$jvp(o-e \sum_{i=1}^n A_{x_{2i}} a_{x_{Mi}} \backslash \text{hsisN} > - n (i- n \ll \#) \text{ I I s i s M } \quad (3.35)$$

where  $M$  is the number of units in a parallel subsystem and  $= 1 - p$ . In conclusion, we make the following remark. If we would like to improve the reliability of a series system of  $N$  units using redundancy, there are two ways to do so. The first way is to use  $M$  redundant systems as a whole. The second way is to use  $M$  redundant units for each of the main units (see Figure 3.8). Comparing (3.34) and (3.35), one can find that it is more effective to use a series-parallel structure rather than a parallel-series structure. In particular,

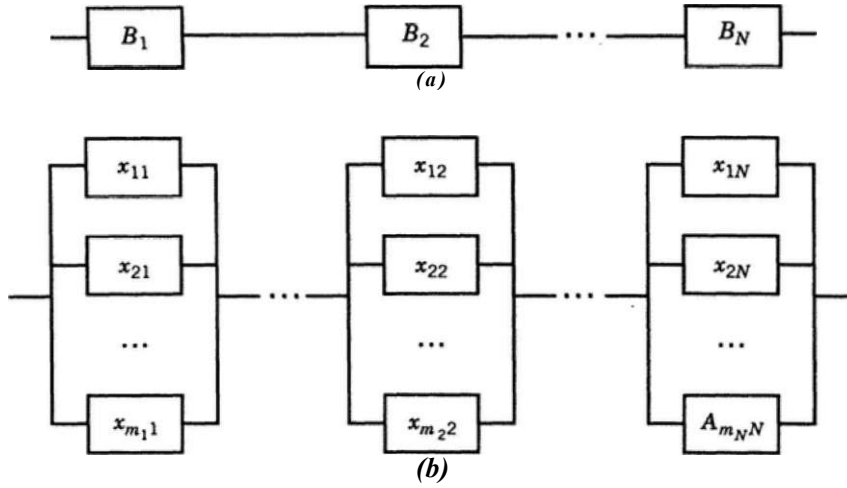


Figure 3.7. Series-parallel structure: (a) in an aggregate form; (b) in a detailed form.

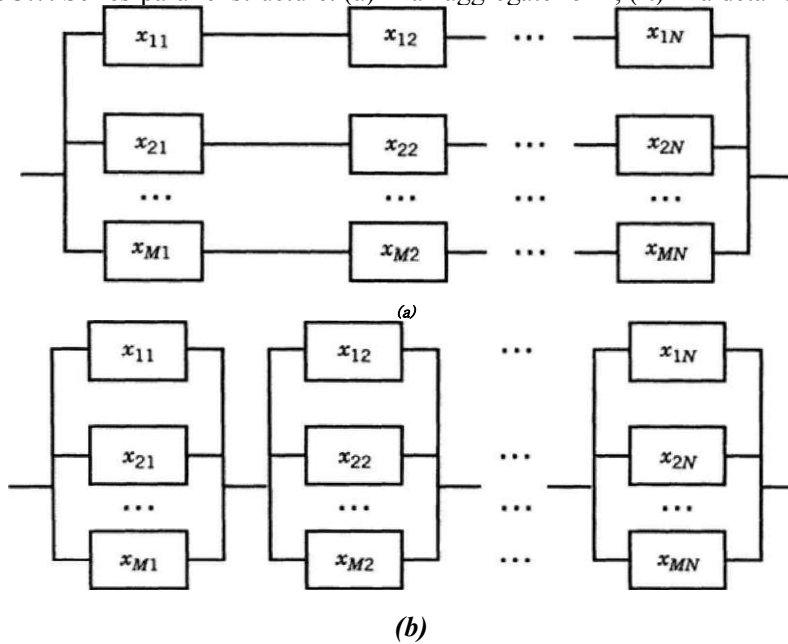


Figure 3.8. Parallel-series (a) and series-parallel (b) structures of size  $N \times M$ .

for identical units

$$1 - (1 - p^N)^M \leq (1 - q^M)^N \tag{3.36}$$

From Figure 3.8 one sees that in a series-parallel system there are more "degrees of freedom," more possibilities to avoid failures. To check this statement, we suggest the extremely simple and clear proof of the statement based on the inequality

$$\max_{I \subseteq S, |I|=M} \min_{i \in I} x_i < \min_{I \subseteq S, |I|=M} \max_{i \in I} x_i \tag{3.37}$$

This inequality means that under any splitting of the set of  $x_i$ 's by subsets, the minimal value among the maximal values for all these subsets is always larger (not smaller) than the maximal value among the minimal values.

Now, using this fact, one can prove the statement. Notice that if  $f_i$  is the random TTF of the  $i$ th unit in the  $s$ th subsystem of series units, then the random TTF of this subsystem is

$$f_s = \min_{i \in S} f_i \tag{3.38}$$

and, consequently,

$$f_{ps} = \max_{I \subseteq S, |I|=J} f_I \tag{3-39}$$

is the random TTF of the parallel-series system as a whole.

Consider the same set divided in such a way that  $f_{ij}$  is the random TTF of the  $j$ th unit in the  $i$ th subsystem of parallel units. Then the random TTF of this subsystem is

$$\max_{I \subseteq S, |I|=i} f_I, \tag{3.40}$$

$$f_{is} = \min_{I \subseteq S, |I|=i} f_I,$$

and, consequently,

$$f_{sp} = \min_{I \subseteq S, |I|=N} f_I \tag{3.41}$$

is the random TTF of the series-parallel system as a whole.

A substitution of (3.38) to (3.41) in (3.37) gives, for any sample of r.v.'s  $f_i$  that  $f_{sp} > f_{ps}$ . From this it automatically follows that

$$T_{ps} = E\{U_{ps}\} \wedge^E \{f_{sr}\} = T_{sp}$$

and

$$W) = M f_{ps} P^{U_{sp}} - P S P W$$

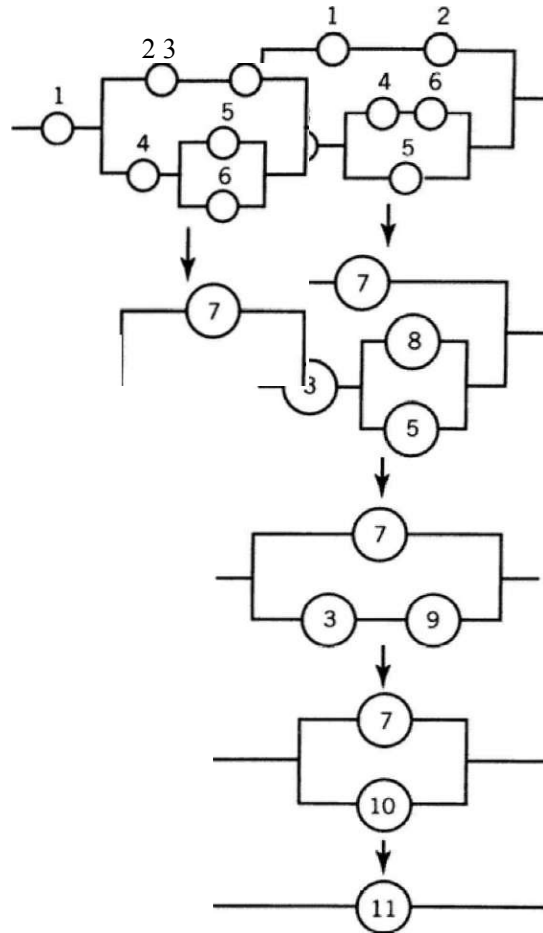
For a "long" series-parallel system (when  $N > 1$ ), the Weibull-Gnedenko distribution might be applied if the system's reliability is relatively high. Consider a system of independent identical units. The distribution of the TTF of each parallel subsystem is such that  $M$  is the first order of the derivative which differs from 0. As we considered in Section 3.1, in this case

for small  $t_0$  and relatively large  $N$ , the Weibull-Gnedenko distribution can be used for the description of the TTF of the system as a whole.

Thus, any series-parallel or parallel-series system can be understood as a two-pole network of a special type. This network possesses the so-called *reducible structure*. A sequential application of the following procedures—(a) replacement of each series connection by a single equivalent unit and (b) replacement of each parallel connection (or "k out of n" structure) by a single equivalent unit—allows one to transform any reducible structure into a single equivalent unit.

Such a reduction is very convenient for the calculation of the probability of a system's successful operation. For instance, consider the structure shown in Figure 3.9. This figure depicts the sequential steps of the system reduction. We hope that the figure is self-explanatory.

Using a similar procedure in an



L<iK!>J

L

!

<D-Gs>

I

—⊙—

Figure 3.9. Examples of the reduction of a complex structure with parallel and series inner substructures to the simplest kinds of structures.

inverse way, one can construct various reducible structures from a single equivalent unit by a detailed disaggregation at each step. Examples of irreducible structures (different arbitrary networks) are presented in a later chapter.

### 3.5 STANDBY REDUNDANCY

A number of systems have standby redundant units. These units are not included in an "active" system's structure. Moreover, these redundant units cannot fail before they occupy an active position. A standby unit instantly replaces its corresponding main unit after the latter's failure. Generally, such a replacement does not occur instantly, but most mathematical models assume that the time of the standby unit's switching into a main position is 0.

The spare parts supply problem is usually analyzed in terms of standby redundancy. For this problem the sufficiency of the stock of spare parts for replacement (or repair performance) is considered rather than the system's successful operation. (Usually, in this case, one considers the inventory system itself. But for simplicity of terminology we will use the terms *system failure* and *instant replacement* in further discussion.)

For standby redundancy it is convenient to speak about "a socket" and a set of units which are put into it one by one after each failure. All units for the socket are considered to be identical. In reality, a standby unit of some type must be able to replace the main unit only with a unit of the same type.

#### 3.5.1 Simple Redundant Group

If a system consists of a single main unit and  $m - 1$  standby units, we call it a simple redundant group. In this case the random time of the system's successful operation  $0_m$  equals

$$\sum_{i=1}^{m-1} Z_i$$

Thus, a system's MTTF can immediately be written as

$$T_w = E\{9J - E \left( \sum_{1 \leq i \leq m} Z_i \right) = E m) = E 7\} \quad (3.42)$$

The MTTF does not depend on the switching order of the steady units. If all

units of the system are identical,

$$T_{\text{yst}} = m T$$

where  $T$  is the single unit's MTTF.

It is clear that standby redundancy is much more effective than active redundancy: here the growth of  $r_{\text{yst}}$  is linear, and for the active redundancy case the growth is only logarithmic. But again we would like to emphasize that this mathematical model is a very idealized picture of reality. Remember the well-known property of the mean: formula (3.42) is valid even if the standby units are dependent.

The probability of a system's successful operation  $P_t(r)$  can be written as

$$P_t(r) = \Pr\{e_m \leq r\} = \Pr\{E \leq mT - r\} = e^{-\lambda(mT - r)}, \quad (3.43)$$

The system's TTF,  $T_{\text{yst}}$ , represents the sum of independent r.v.'s. As we know from Chapter 1, in this case

$$T_{\text{yst}} = \int_0^{\infty} t f_{\text{yst}}(t) dt = \int_0^{\infty} t \lambda^m e^{-\lambda t} dt$$

where  $P^k(t)$  is the probability of a failure-free operation of the system with  $k - 1$  standby units ( $k$  units in the redundant group).

As known from Section 1.3.1, only a very restricted number of d.f.'s allow one to find convolutions in convoluted form. The reader can use the above-mentioned results for probability calculations. Of course, if the number of standby units in the redundant group is large, a normal approximation based on the central limit theorem (see Section 1.3.2) can be used.

In engineering practice, especially in electronics, the most frequently used distribution  $F(t)$  is exponential. The standby group's random TTF has an Erlang d.f. of the  $m$ th order, and the probability of a failure-free operation is

$$P_t(r) = \frac{\lambda^m}{(m-1)!} \int_0^r (r-x)^{m-1} e^{-\lambda x} dx = e^{-\lambda r} \sum_{k=0}^{m-1} \frac{(\lambda r)^k}{k!} \quad (3.44)$$

For  $m \gg 1$  the approximation can be written as

$$P_t(r) \approx \frac{1}{\sqrt{2\pi m}} e^{-\lambda r} e^{\frac{(\lambda r - m)^2}{2m}} \quad (3.45)$$



If  $\lambda t$  is not too small, the following inequality is true:

$$\sum_{m \leq i < \infty} \frac{(\lambda t)^i}{i!} e^{-\lambda t} = \frac{(\lambda t)^m}{m!} e^{-\lambda t} \left[ 1 + \frac{\lambda t}{n+1} + \frac{(\lambda t)^2}{(n+1)^2} + \dots \right]$$

$$< \frac{(\lambda t)^m}{m!} e^{-\lambda t} \left[ 1 + \frac{\lambda t}{n+1} + \frac{(\lambda t)^2}{(n+1)^2} + \dots \right]$$

$$= \frac{(\lambda t)^m}{m!} \left( 1 - \frac{\lambda t}{n+1} \right)^{-1} e^{-\lambda t}$$

(3.46)

This can be used for approximate computations. The substitution of (3.46)

into (3.45) produces an approximation of  $P_{sm}(t)$ :

$$P_{\text{sys}}(t) \approx 1 - \frac{(\lambda t)^m}{m!} \left( 1 - \frac{\lambda t}{n+1} \right)^{-1} e^{-\lambda t}$$

Note that this value is smaller than the exact value; that is, it delivers a "guaranteed result."

In conclusion, note that standby redundancy is more effective than active redundancy (at least in theory!). This follows from the simple fact that

$$f(\text{standby redundancy}) = f_1 \text{ a max } ft = \mathcal{L}(\text{active redundancy})$$

The equality is never attained because of the strongly positive values of the  $f$ 's. (Of course, we are considering the case where  $rn > 1$ .) Of course, the reader should never forget that standby redundancy, in practice, requires some time to switch a unit into an active regime.

Finally, we would like to point out the relationship between the MTTFs for series and parallel systems of two units. (The result can easily be expanded by induction to an arbitrary number of units.) Suppose that one unit has a random TTF  $f_1$  and another has  $f_2$ . It is clear that

$$f_1 + f_2 = \min(f_1, f_2) + \max(f_1, f_2)$$

because one of these r.v.'s is obviously larger and another is smaller. Taking the mean of both sides of the equality, one gets

## 11 6 UNREPAIRABLE SYSTEMS

$$E\{t_1} = E\{\min(f, e_2)\} + E\{\max(f, t_2)\}$$

Now we can see that these values are, in order:

- \* The first is the MTTF of a duplicated system of two standby redundant units which are working sequentially one after another.
- \* The second is the MTTF of a series connection of these units.
- \* The third is the MTTF of a parallel connection of these units.

In particular, when both  $\lambda$ 's are exponentially distributed, one can obtain a convenient expression for the MTTF of a parallel system of two different units;

$$T_{\text{parallel}}^{-k_2} = \frac{1}{\lambda_1} + \frac{1}{\lambda_2}$$

or, in final form,

$$T_{\text{parallel}} = \frac{1}{\frac{1}{\lambda_1} + \frac{1}{\lambda_2}}$$

Of course, the latter expression has such a simple form only because both distributions are exponential,

### 3.5.2 "r out of n" Redundancy

The use of standby units for several main units is very common. For example, consider a system which includes  $k$  main units. To support the system, there are  $n - k$  spare units which can replace any main unit of the group. This method of redundancy is very efficient because of the large number of "degrees of freedom" in the usage of standby units. Indeed, no unit is predetermined to replace some specified main unit. (We repeat that this mathematical model is mainly used to describe a spare units supply system.)

For standby redundancy the formulas for  $P^{\wedge}(t)$  and  $r_{\text{sys}}$  cannot be written in a convoluted form except for the case of an exponentially distributed random TTF of the units. We may write the result basing our explanation on simple arguments.

Recall again that we assume that the units are independent. The system consists of  $k$  identical units and has  $n - k$  standby units. The system failure rate equals  $k\lambda$ . After a first failure the failed unit is replaced by a redundant unit and the system continues its operation. The random TTF equals  $r$ , and has an exponential distribution with parameter  $k\lambda$ . The MTTF in this case equals  $T/k$  where  $T$  is the MTTF of a single unit. The memoryless property of the exponential distribution and the independence of the units ensure the exponentiality of the system's random TTF. Hence, a random period of a system's successful operation consists of the sum of  $n - k + 1$  i.i.d. TTFs.

(One period to the first system failure and then  $n - k$  replacements of spare units.)

Therefore, the system's MTTF is

$$T_{\text{sys}} = E(\sum_{i=1}^n U_i - (rt - k + I)) \quad (3.47)$$

The probability of a system's successful operation when its units have exponential distributions is

$$P_{\text{sys}}(t) = \Pr\{E_{\min U_i} > t\} = \sum_{i=1}^n \Pr\{E_i > t\} = \sum_{i=1}^n e^{-\lambda t} = n e^{-\lambda t} \quad (3.48)$$

In general, the problem is very complicated. The most reasonable way to calculate accurate values of the reliability indexes  $P_{\text{sys}}(t)$  and  $T_{\text{sys}}$  is via Monte Carlo simulation.

Below we give a simple method for obtaining lower and upper bounds for these reliability indexes. It is clear that the best use of the standby units would be in a so-called "time-sharing" regime. Here the MTTF of the "/c out of  $n$ " structure could be calculated as the total operation time of all units divided by  $k$ . The upper bound for  $r_{\text{sys}}$ , follows:

$$T^{\wedge} < E \&\} = \backslash T \quad (3.49)$$

Comparison of (3.47) and (3.49) shows the difference between the accurate value of  $T_{\text{sys}}$  for the exponential distribution and its upper bound. An upper bound for  $P_{\text{sys}}(t)$  can be obtained via the use of similar explanations:

$$P_{\text{sys}}(t) \leq \sum_{i=1}^n \Pr\{E_i > t\} = n e^{-\lambda t} \quad (3.50)$$

To obtain lower bounds, we use the fact that the joint use of redundant units is more effective than an individual one. Let us equally allocate all redundant units among  $k$  initially operational units of the system. Then we have  $k$  series subsystems, each with  $n/k$  redundant units.

If  $n/k$  is not an integer, the procedure will be slightly more difficult. Denote the integer part of  $n/k$  by  $m^* = \lfloor n/k \rfloor$ . Then  $a = n - km^*$  subsystems have  $m^* + 1$  redundant units and all of the remaining  $b = k - (n - km^*)$  ones have  $m^*$  standby units. Thus, a lower bound for  $P_{\text{sys}}(t)$  is

$$> 1 - [1 - F^{m^*}(0)]^b [1 - F^{m^*+1}(t)]^a \quad (3.51)$$

where  $F(t)$  is the d.f. of the random TTF of a single unit. A lower bound of the MTTF can be found by integrating (3.51). If the coefficient of variation of  $F(t)$  is small, (3.51) can be reduced to

where the previous notation is preserved.

### 3.5.3 On-Duty Redundancy

The use of standby redundant units in an operation requires a special regime on duty. This regime is intermediate between the two previously considered types: active and standby.

We illustrate the subject with several examples. An electronic monitor needs at least a portion of a second to be ready to display information. A redundant computer in a control system must be supplied with current information before it is switched to an operational regime. Usually, the unit on duty has a regime which is lighter than the working unit but harder than a total standby one. Practically, there is no realistic input data for this on-duty regime and, moreover, even a confident knowledge about the process is absent. Even with the appropriate input data, the problem of a reliability evaluation in this case is hardly solvable analytically under general assumptions. As a rule, Monte Carlo simulation allows one to obtain numerical results. But even in this case a lack of input data makes the result very problematic.

The only mathematically acceptable model arises when all units have an exponential distribution of their TTFs: the main ones with parameter  $\lambda$ , and the redundant ones with parameter  $a\lambda$  where  $a < 1$ . In general, a system may have on-duty units which are used for the replacement of failed main units and standby units which are switched into an on-duty position.

Consider a system of  $k$  main (operational) units,  $l$  on-duty units, and  $m$  standby units. Let  $N = k + l + m$ . For this case one can build the transition graph in Figure 3.10. On the basis of this graph, the following system of

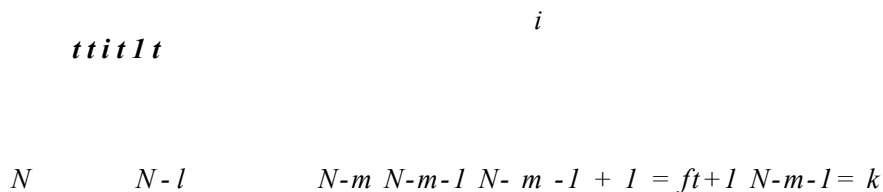


Figure 3.10. Transition graph for an unrepairable system of  $k$  main,  $l$  on-duty, and  $m$  standby units.

differential equations can be written:

$$P'_N(t) = \dots + \lambda A$$

$$PN-M = \dots (H^*A + \lambda A)$$

$$PN-m(0) = \dots [^*A + \lambda a]$$

$$PN-m-1(0) = \dots (\lambda - 1)A$$

$$= \dots + \dots = \dots -P^* + \dots [^*A + \lambda A]$$

$$PN-m-1(0) = P^*(0) = -Pk(O \wedge A)$$

$PN-m-liO =$  The initial state for this process is the system state with all operating units, so  $pf = 1$ . The system MTTF can be found immediately from the transition

graph in Figure 3.10:

1 1

$$T_{\text{mean}} = \dots A(A + \lambda a) + \dots A(* + (\lambda - 1)A)$$

The method described in Chapter 2 can be used to find the probability of a failure-free operation. But we avoid writing bulky and boring expressions. The obvious upper- and lower-bound models can be written with common sense: one for an upper bound (see the transition graph in Figure 3.11), and another for a lower bound (see Figure 3.12). The first graph contains all transition intensities equal to the maximal one,  $A(k + a)$ . The second one is constructed under the observation that as soon as at least one on-duty redundant unit has been spent, all of the remaining units become standby units.

**§ 3.3**

$N \quad N-1 \quad N-m \quad N-m-1 \quad N-m-1+1 \quad N-m-1$

Figure 3.11. Transition graph for obtaining the lower bound.

$N$              $N-1$              $N-m$   $N-m-1$              $N-m-1+1$   $N-m-1$

Figure 3.12. Transition graph for obtaining the upper bound evaluation.

For the graph of Figure 3.11, we can write the result immediately without solving the system of differential equations. Indeed, we have the sum of  $m + 1$  i.i.d. exponential r.v.'s with means equal to  $1/A(k + a)$ . Therefore,

$$JWO = L^{-\frac{1}{A(k+a)}} e^{-\frac{1}{A(k+a)} t}$$

For the second case we have the sum of  $m$  exponential r.v.'s with means  $1/A(k + a)$  and of one exponential r.v. with mean  $1/Ak$ :

where

$$F(x) = \int_0^x \frac{1}{A(k+a)} e^{-\frac{1}{A(k+a)} t} dt + \frac{1}{Ak} e^{-\frac{1}{Ak} t}$$

and

$$F(0) = 0$$

$$F(1) = 1$$

These boundary estimates are given not as essential results but rather as examples of the possible thinking involved in finding simple solutions to complex problems.

We repeat that on-duty redundancy, in general, is a real problem in practice, not because of the solution difficulties but because of the lack of information. Usually, nobody has the slightest idea of the kind of distribution parameters (or even the distributions of the TTFs themselves!) which a unit in an on-duty regime has.

### 3.6 SWITCHING AND MONITORING

Traditional mathematical models of redundant systems reflect an idealized redundancy: no switching device, no operational monitoring, and no maintenance are investigated. These idealized mathematical models help a researcher to understand the redundancy phenomenon but they could also lead to very harmful mistakes if these models are used without corrections.

For example, the mathematical formulas show that it is possible to reach any specified level of system reliability with the use of redundant units. But engineering practice convinces us that real system reliability improvement depends on the reliability and quality of the monitoring and switching devices. Both of them are usually far from ideal themselves.

Of course, taking into account all of these various factors will lead to more complex and less elegant mathematical results. But in engineering, the results must be not only elegant but also useful!

#### 3.6.1 Unreliable Common Switching Device

In practice, active redundant units are not really operating in parallel. In the case of electronic equipment, the simultaneous presence of several output signals from all redundant units could lead to a real mess. In the case of an information system, the superposition of output information from several computers can produce false signals. Usually, active redundant units are operating in an on-duty regime although their reliability characteristics may not be distinguished from the main unit. All functions of monitoring, switching, and special interface duties are performed by some special device which we call a switching device (SD). Of course, a model of such a group of redundant units almost coincides with the model of a group of active redundant units. But at the same time one needs to take into account the presence of the SD.

Consider a group of  $m$  redundant units which uses a common SD for switching from a failed unit to an active redundant one. First of all, note that the SD itself might be one of two main types:

1. The SD is always necessary for the normal operation of the redundant group as a whole.
2. The SD is necessary only at the moment of switching performance.

In the first case, the SD can be, for example, an interface between the redundant group and the remaining equipment. It can be of a various physical nature (electrical, mechanical, hydraulic, etc.). The successful operation of the redundant group depends directly on a failure-free operation of the SD.



In the second case, the SD becomes necessary only at the moment of switching. Even if the SD has failed, the system can successfully operate until the main unit fails. But then the system will have failed even if there are available redundant units.

**Necessary Switching Device** Denote the random TTF of the  $i$ th unit of the redundant group of  $m$  units by  $f_i$  and the random TTF of the SD by  $\theta$ . For the first case, we can write that the random TTF of the redundant group  $\xi$  is  $\min(\theta, \max f_i)$ . For the redundant group as a whole,

$$= P_{SD}(t)P_m(t) \tag{3.52}$$

where  $P_m(t)$  is the probability of a failure-free operation of the redundant group. In other words, for such an SD, the redundant group can be investigated as a simple series-parallel system (see Figure 3.13). The MTTF for this case can be generally found by integrating  $P^{\wedge}O$ ) as determined in (3.52).

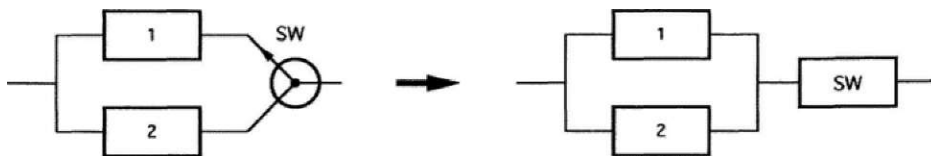


Figure 3.13. Approximate representation of the duplicated system with the switch as a series connection of the two redundant units and a non-ideal switch.

**Switching Device Using Only for Switching** The second case should be considered in more detail. There are two possibilities for an SD failure:

1. The SD may fail during the time of the system's operation.
2. The SD may fail with probability  $Q$  only at the moment of switching, independent of the time when it has occurred.

**Switching Device Failure Depending on Time** Consider a redundant group of identical and independent units with an exponential distribution of their random TTFs. The probability of a successful operation of the redundant group is denoted by  $P_{RG}(t)$ . The distribution of the switching device TTF  $F_{SD}(t)$  is arbitrary. There are two possibilities for the system's operation:

Notice that both cases—active and standby redundancy—are equivalent, in the sense of the general formula (3.53), under the assumption of exponentially of a unit's TTF. These conditions can be written as

$$P_{SD}(0) = 1 - F_{SD}(0) \quad (3.53)$$

where  $P_{SD}(0) = 1 - F_{SD}(0)$ .

The exponentiality of a unit's TTF permits us to use (3.53) for both active and standby redundancy. The expressions for the redundant group,  $P_{RG}(t)$ , are different, but the residual time of the remaining unit after an SD failure is exponential in both cases.

From (3.53) one sees how the reliability of an SD influences the reliability of the system as a whole.

A system's MTTF  $T_{sys}$  can be obtained only by integrating the corresponding  $P(t)$ . Now we would like to consider some limiting cases. If  $T_{RG} \gg T_{SD}$ , then  $T_{sys}$  equals  $T_{RG}$  for all practical purposes. If  $T_{RG} \ll T_{SD}$ , then  $T_{sys}$  approximately equals  $T_{SD} + T$ , where  $T$  is the MTTF of a single unit.

**Example 3.1** Consider a duplicate system with an active redundant unit. The distributions of the random TTFs for the unit and for the SD are exponential with parameters  $\lambda$  and  $\lambda_{SD}$ , respectively. Find the probability of a failure-free operation during a time interval  $t$ :

$$P(t) = 1 - \int_0^t (1 - e^{-\lambda(x)}) e^{-\lambda_{SD}(t-x)} \lambda e^{-\lambda x} dx$$

This solution can be easily obtained in closed form.

For a unit with an arbitrary TTF distribution, the solution is not so simple, though its general form does not seem especially awkward. (Very often—including this case—a "simple" form of a formula hides numerical difficulties which arise during computation!)

Consider active redundancy. Assume that the current operating unit of the redundant group is chosen randomly. The residual unit's TTF begins from the moment  $t$  (the SD failure). Equation (3.53) transforms into

$$W(t) = \int_0^t P_{SD}(t-x) f(x) dx + P_{SD}(t) F(t)$$

---

4 The SD fails at some moment  $x < t$ , the redundant group has not failed until  $x$ , and, after this moment, the current main unit does not fail during the remaining time  $t - x$ .

**SWITCHING AND MONITORING 143**

(3.54)

where  $p(t)$  is the probability of a unit's failure-free operation and  $q(t) = 1 - p(t)$ .

For a standby redundant group, the expression is slightly bulky: the conditional distribution of the residual time of a unit which appears in the operational position at the moment  $x$  depends on the number of failures which have occurred before that moment. We obtain an approximation by considering the process of failures before  $x$  as renewal. Then we can write

$$\hat{F}_{\text{syst}}(0) \sim \hat{F}_{\text{standby}}^{(5)}(x)$$

---

5 The first unit chosen at random operates without a failure during period  $t$ .

$$\int_0^x p(x-y)p(t-x|y) dH(y)$$

^Sd( 0 ^standby Rg( 0

$$\int_0^+ p\{t-y\} dH(y) \quad ( \quad dF_{so}(x) \quad (3,55)$$

**SWITCHING AND MONITORING 145**

where  $H(t)$  is the renewal function, In other words,  $H(t) dt$  is the probability that some failure has occurred in the time interval  $[t, t + dt]$ . We use  $H(t)$  though we observe a finite sequence of r.v.'s but not the point renewal process. We should remark that, for highly reliable systems, this approximation is quite good.

Both cases also allow the following approximation:

$$P^{\wedge} = p(t) + q\{t\}P_{so}(t)P^{\wedge-n}U \quad (3.56)$$

where  $P^{\wedge}$  is the system with an SD and a redundant group of size  $k$ .

This approximation gives a lower bound on the probability of interest because we assume that the SD operates successfully during the entire period  $t$ . As a matter of fact, an SD failure may not lead to a system failure.  $\int_0^+ p\{t-y\} dH(y)$  and by RC,  $\int_0^+ p\{t-y\} dH(y)$

**3.6.2 Common Switching Device with Unreliable Switching**

Consider an active redundant group of  $n$  independent and identical units. The system's successful operation is possible in two situations:

This permits one to write a recurrent relationship in the form

$$= p(t) + R \int_0^t p(t-x) dq(x) \quad (3.57)$$

where  $R$  is the probability of a successful switching.

For a standby redundant group of  $m$  independent and identical units, the system's successful operation is possible in two situations:

1. The first unit chosen at random operates without a failure during period
2. The first unit chosen at random fails at some moment  $x < t$ , the SD performs a successful switch, and from this moment on the new system of  $m - 1$  redundant units and SD perform successfully until time  $t$ .

This description permits one to write the recurrent relationship

$$- x) dq(x) \quad (3.58)$$

where  $P^{\wedge}(t)$  is the probability of a successful operation of the active redundant group of  $k$  units during a time interval  $t$ .

### 3.6.3 Individual Switching Devices

We assume that each unit of a redundant group may be chosen to replace a failed main unit at random. After the main unit's failure, an individual SD associated with the next unit in the redundant group may successfully perform the next connection, or it may fail. Assume also that a unit's failure leads to a corresponding SD's failure. The absence of an operating unit with an operating SD leads to the system's failure. As follows from the above description, the SD is necessary for the unit to operate.

**Switching Devices That Fail with Time** For active redundancy a system operates successfully in the following situations:

The probability of a system's successful operation can be written as

1

$$P_{i0} + \sum_{j=1}^m \binom{m}{j} p^{j-1}(x) q^{m-j}(x) M^j \quad (3.59)$$

where  $p(t) = p_{SD}(t)p()$ .

For a standby redundant group, the expression is simpler. The following are situations wherein a system operates successfully:

1. The first unit operates successfully.
2. After its failure there is a group of  $m - 1$  redundant units with a random number of operating SDs; this new system operates successfully during the remaining time. The random number of operating SDs appears because of SD failures during the on-duty regime.

The probability of a system's successful operation can be written as

$$+ \sum_{k=1}^m \binom{m}{k} [p_{SD}(*)]^k [1 - p_{SD}(*)]^{m-k} \quad (3.60)$$

Of course, (3.59) and (3.60) can be practically utilized only with the aid of a computer.

**Switching Devices That Fail at the Time of Switching** First, consider an active redundant group. There are again two situations where the system can successfully perform its operation:

---

6 The first unit operates successfully.

that the new system of —  $k$  redundant units must successfully operate during the remaining time period  $t - x$ .

This verbal description corresponds to the expression

$$\ll (o - p(o + * / ^f E ("rMipwi'irf*]))^{m-1} \cdot \{ E Q^k P \} \& \{ t-x \setminus x \} \setminus dq \{ x \} \tag{3-61}$$

where  $Q = 1 - R$ .

When we consider standby redundancy, the system can successfully operate if:

1. The first unit operates successfully.
2. After its failure at some moment  $x$ , there is a group of  $m - 1$  standby units. In some order we try to switch each of these units to the main position until a first successful switching occurs. The number of attempts before success is distributed geometrically with parameter  $R$ . After  $k$  SDs have failed during switching, a successful attempt occurs ( $k$  is random). The new system of  $m - k - 1$  redundant units must successfully operate during the remaining time interval  $t - x$ . The appropriate probability is

$$PZ^k P O + R f' E Q^k P^{\sim k \sim n} (t - X) dq(x) \tag{3.62}$$

The MTTF for both systems can only be found numerically. Note that for the exponential distribution and large  $m$ , (3.62) can be approximated by

$$= \tag{3.63}$$

and the MTIT can be approximated by

$$7_{,ys,} = \ \tag{3-64}$$

Both (3.63) and (3.64) are obtained under the assumption of the correctness of the application of the result of the random summation to exponentially distributed r.v.'s. (Note that in our case we consider a fixed number of Bernoulli trials.)



This brief review does not, of course, cover the entire theme. There are various cases in practice which lie outside the scope of this material. Our main purpose was to display some inferences in this area and not to give the reader a "cook book."

**3.6.4 Periodic Monitoring**

In this section we do not try to examine the monitoring problem but rather give a simple example of the possible influence of monitoring on a system's reliability. Above we considered a redundant system with the possibility of an instant replacement of a failed main unit by a redundant one. In many practical cases such a situation is unrealistic. In many cases the state of the units, main or redundant, can be checked only at some prespecified moments, usually at periodic intervals.

Consider a simple system consisting of two parallel units and one standby redundant unit which cannot be switched immediately to either of the parallel units. This redundant unit can replace either failed parallel unit only at some predetermined moments  $t_s = sA, s = 1, 2, \dots$ . At these moments the state of the two parallel units is checked and a failure may be detected. (In other words, the monitoring of the units is not continuous.) All units are assumed identical and independent, and their TTF distributions are assumed exponential with parameter  $\lambda$ .

The system is considered to have failed if:

1. Both parallel units have failed inside a period between two neighboring check points, even if there is a standby unit.
2. There are no units operating at some moment.

Consider the probability of a system's failure-free operation during  $N$  cycles. For this case the following discrete recurrent equation can be written:

$$P_{\text{sys}}(N) = P_{\text{sys}}(N-1) + 2pq[1 - Q(N-1)] \quad (3.65)$$

with  $P_{\text{sys}}(0) = 1$ . Here  $Q(K)$  is the probability of a failure-free operation of two parallel units during  $K$  cycles,  $Q(K) = 1 - [1 - p^K]^2$ , and  $p = 1 - e^{-\lambda A}$ .

Equation (3.65) can be solved numerically.

For the system's MTTF, one can write

$$T_{mt} = p^2 [A + 3 + 2pq[A + T_2]] + q^2 A^* \quad (3.66)$$

where after a successfully operating cycle of length  $A$  the system starts its failure-free operations from the beginning. A cycle with two failures contains a portion of useful time which is denoted by  $A^*$ . Setting  $A^* = A$ , we can write

the approximation

$$- A + p^2 T^{\wedge} + 2pqT_2 \tag{3.67}$$

where  $T_2$  is the average number of successful cycles of the two parallel units.

We essentially use here the Markov nature of the model. Even in this simple case we have no strong results in a simple form. But we see that monitoring essentially changes the operation process and, consequently, changes the reliability indexes of the system.

Equations (3.65) and (3.66) are complicated enough to make some quantitative conclusions, but we consider two simple limiting cases. It is clear that  $A \rightarrow 0$  leads to the continuous monitoring model, and, hence, the system reverts to a system composed of two active redundant units and one standby redundant unit. Incidentally, the MTTF of such a system equals

$$\frac{1}{2A} + \frac{1}{2A} + \frac{1}{A} = \frac{3}{2A}$$

If, on the contrary, we assume that  $A \rightarrow \infty$ , it means that factually the system has no redundant units at all because they will never be used. In this case  $r_{sys} = 3/2A$ . Thus, for intermediate  $A$ 's, the value of  $T_{mt}$  lies somewhere between the mentioned values.

It is clear that for a series system of units with an exponentially distributed TTF, it is totally unreasonable to have any redundant group which can be switched only after periodically checking the system's state. (We consider reliability indexes such as the probability of failure-free operation or the MTTF.)

### 3.7 DYNAMIC REDUNDANCY

This interesting redundancy class is very close to the classical problems of inventory control. Consider a redundant group of  $n$  units. A part of them might be used as active and another part as standby. There is a possibility of checking the current state of the active units only at some predetermined moments. Thus, there is no feedback information within the interval between two neighboring check points. The system can be found to be failed at a moment of checking even if there are some standby units available to be used. The following questions arise: How many units should be reasonably switched between two checking moments? How does one refill the active redundant group?

From a qualitative point of view, it is clear that it is not reasonable to switch all redundant units to an active state: the reliability at the first stage of operation will be high but the redundant units will be rapidly spent. To switch a small number into the active redundant group is unreasonable because a system failure can, with a high probability, appear before a current check.

This kind of problem may appear in connection with nonmonitoring technical systems, for example, space vehicles designated for an investigation of the solar system. The time of response can be excessively long, and it becomes impossible to control the situation, so that one needs to find some prior rule of autonomous switching of the redundant units over time without external signals,

A possible solution to the problem is to choose moments of switching spare units into acting positions. We will discuss the problem of finding the optimal moments of switching in Chapter 11. Now we only consider how to calculate the reliability indices for such a system.

We call such a kind of redundancy a *dynamic redundancy*. We will only investigate dynamic redundancy with exponentially distributed unit ITFs. All units are also supposed to be identical and independent.

### 3.7.1 Independent Stages

The system possesses  $n$  identical and independent units to perform its function up to some time  $t_0$ . Then an initial group of units,  $n_0$ , is installed as an active redundant group and all remaining units are placed in a standby regime. The duration of the system's operation is divided into  $k$  stages. There are moments  $0 < \tau_1 < \tau_2 < \dots < \tau_k < t_0$  when the new group of standby redundant units are to be switched into an active regime. When we consider independent stages, such switching is performed at some predetermined moments. Such a procedure is called a *programmed controlled switching*. The previous group of units is expelled from the operation with no consideration of their real state. (As a matter of fact, no active units may fail before the beginning of the next stage.) In this case all stages are independent. Such a situation arises if the deployment of previously used units for use at the next stage is a difficult or even impossible engineering task.

In this simple case the probability of a system's successful operation during a time interval equals

$$n \{1 - kWP\} \tag{3-68}$$

where  $Q(t_j)$  is the probability of a failure-free operation of a single unit.

The calculation of a system's MTTF is not very simple in this case. Assume that a failure of the system occurs at some stage  $k$ . This means that the system operates successfully for  $k - 1$  stages and during some random time within the last stage. The  $k$ -th stage duration equals  $A_k = \tau_k$  — The conditional value of a failure-free operational time during stage  $k$  (denote this by  $C_k$ ) is

$$f^* = \{ \max t, A_k \}$$

This conditional mean time can be found in the standard way,

$$\frac{\int_0^\infty p(x) dx}{p(A)} = \frac{h}{m^*}$$

and only the group of units at the last stage operates until complete exhaustion of all redundant units.

$$E\{n_P(\wedge)\} = E\{A_k + E\{J\}\} = E\{\max_{i=1, \dots, r} t_i\}$$

Finally, for the system we have

(3.69)

Once can use an approximation by replacing  $\lambda_k$  with  $A_k$  or one can obtain lower and upper bounds by substituting  $\lambda_k = 0$  and  $\lambda_k = A^*$ , respectively.

### 3.7.2 Possibility of Transferring Units

A more interesting and more complicated case arises if one considers the possibility of using all nonfailed units at some stage for the next stage. Of course, in this case it is possible to analyze only the systems whose units have an exponentially distributed TTF. If stage  $j$  has a duration  $A$  and there are  $m$  units in the active redundant group (including those from the previous stage), then the probability of a failure-free operation is given by

$$P_j(A) = 1 - [1 - \exp(-AAj)]^m$$

After the first stage of a successful operation, the system has a random number, say  $y$ ,  $0 \leq y \leq m$ , of operating units. These units can be used at the second stage, starting at the moment  $\tau$ , with  $n_2$  new units switched in by the prior rule. Thus, the total number of units acting at this stage equals  $n_2 + j$ . The probability of exactly  $j$  units ( $j > 0$ ) being transferred to the second stage is

$$P_{Wi} = \binom{m}{j} p^j q^{m-j}$$

where  $p = 1 - q$ . If the system performed successfully during the first stage,  $j$  units ( $j > 0$ ) leave to operate at the second stage. At the same time, at moment  $\tau$ , new  $n_2$  units are switched into the system. So, for a two-stage process with  $m \sim \text{bin}(j + n_2)$ , one can write the probability of interest as

$$P_{U*2} = E \left( \binom{m_2 + j}{j} p^j q^{m_2 + j - j} \right) \tag{3-70}$$

$\} I$



### 3.8.1 Series Systems

Consider a series system of two units. Let  $x_i$ ,  $i = 1, 2$ , be the indicator function of the  $i$ th unit. Suppose the  $x_i$ 's are associated. For instance, both of them are dependent on the same environmental factor  $f$ ; that is, notationally, they are  $x_i(f)$  and  $x_j(f)$  for some specified  $f$ . Then for these two units one can write

$$\Pr\{x_1 = 1, x_2 = 1\} = \Pr\{x_1 = 1\}\Pr\{x_2 = 1\} + p(x_1, x_2) \quad (3.73)$$

where  $p(x_1, x_2)$  is the correlation coefficient

This normalized value satisfies the condition  $-1 < p < 1$ . For  $n$  associated r.v.'s it is possible to consider only the case  $p > 0$ . From this condition and (3.73), for a series system of two associated units, it follows that

$$\Pr\{x_1 = 1, x_2 = 1\} > \Pr\{x_1 = 1\}\Pr\{x_2 = 1\} \quad (3.74)$$

This result can be immediately generalized for a series system of  $n$  associated units:

$$\Pr\{X = 1\} > \prod P_i \quad (3.75)$$

From (3.75) it automatically follows that for a series system of  $n$  associated units:

$$\Pr\{X = 1\} > \prod_{i=1}^n P_i \quad (3.76)$$

Consider the example when each unit of a series system depends on the same factor  $p$ , for instance, the temperature. The system is designed for use at two different temperatures,  $p_1$  and  $p_2$ . The designer decides to check the probability of a failure-free operation of the series system of  $n$  units. For this purpose, the designer arranges for a unit testing under these two conditions.

The probabilities of the unit's failure-free operation under these two conditions are  $p_1$  and  $p_2$ , respectively. The average unit failure-free operation probability equals  $p = (1/2)(p_1 + p_2)$ . At a first glance, it is very attractive to try to compute the system reliability index as  $p^n$  if we know nothing about the real conditions of the system's use.

But let us assume that the first condition appears in practice with frequency  $R$ , and the second condition appears with frequency  $Q = 1 - R$ . (Of course, the frequency  $R$  can be considered to be a probability.) Then a

realistic value of the index is  $P_{\text{yst}} = R\rho^n + Q\rho$ . It is easy to check that  $\hat{P}_{\text{yst}} = \lambda \text{Avst}$  (To convince yourself in a particular case, do it with  $n = 2$  and  $R = 1.2$ .) Of course, the same phenomenon will be observed if one considers more than two different environmental conditions.

Another example of a system with associated units is a system operating in a changing regime. Assume that a system operates with probability  $p_k$  at the  $k$ th regime. Under this regime the system's units have a failure rate equal to  $\lambda_k$ . It may happen if the system switches from regime to regime periodically (or, perhaps, randomly). In this case

$$UO = \sum_k p_k e^{-\lambda_k t}$$

This is larger than

$$\sum_k p_k e^{-\lambda_k t}$$

So, for a series system we can use the hypothesis of the unit's independence to obtain a conservative bound on the reliability index of types  $P(t)$  or  $T$ .

### 3.8.2 Parallel Systems

Now consider a parallel system of two associated units. For this system we have

$$\begin{aligned} Pr\{x_x, V^*_2 = 1\} &= 1 - Pr\{A_i = 1\} = 1 - [Pr\{S_i = 1\} + p(x_u, x_2)] \quad (3.77) \\ &< 1 - Pr\{J_{E_i} = 1\} Pr\{J_{t_i} = 1\} = 1 - \text{fl}(r)^*_2(0) \end{aligned}$$

where  $p(x_t, x_2)$  is the correlation coefficient for the indicator functions. It is easy to show that

$$-\text{Var} \wedge \text{Var} K \quad \wedge$$

But  $\text{Cov}(x_i, x_2) = \text{OM}xy, j_2)$  and  $\text{Var}x_i = \text{Var}\{j_{E_i}\}, i = 1, 2$ .

Equation (3.78) can immediately be generalized for a parallel system of  $m$  associated units:

$$P_{\text{yst}} = Pr\{3(X) - 1\} \leq 1 - n \lambda, \quad (3.79)$$

1 sism





Figure 3.14. Transition graph for obtaining the upper and lower bounds for a parallel system with dependent units.

The reader can consider the previous examples applied to a parallel system. We consider several examples connected with the death process which, as we mentioned above, can successfully be used for describing unreparable redundant units. We consider a special type of dependence.

For simplicity, consider a parallel system of three units. All units operate in the system in a nominal regime. For such a regime, each unit has a failure rate  $A$ . If the units are independent, the transition graph is presented in Figure 3.14a. Assume that, after the failure of the first active redundant unit, the two remaining units are forced to operate in a harsher regime. For example, in an electrical parallel circuit, as the flow through each resistor becomes larger, the resistors produce more heat, the surrounding temperature increases, and, consequently, the failure rate increases. In a hydraulic circuit, after one of the parallel pipes is closed, the remaining are under a higher pressure and, consequently, can fail with higher probability. Thus, a unit's failure rate often depends on the state of the other units.

Assume that  $A$  of each unit is an increasing function of the flow through the unit. In this case  $A_3 = 3A$ , but  $A_2 = 2(A + A_2)$  and  $A_1 = A + A_1$ , where  $A_1 > A$ ;  $A_2 > A$ . The transition graph for this case is presented in Figure 3.14b. It is clear that this system of associated units is less reliable than the initial system of independent units.

Now let us consider a case which is, in some sense, the inverse of the previous one. All parallel units are operating in a restricted room. A single unit operating in this room has a nominal failure rate  $A$ . Each working unit generates a heat which accumulates in the room and influences all of the remaining units. Thus, the more units that are operating, the higher the temperature, which leads to the decreasing reliability of each unit. (At

the same time, remember that the system is a parallel system!) We begin the analysis of the transition graph for this case with state 1.

A single unit operates with  $A_1 = A$ . When two units are operating, the temperature in the room is higher and, consequently,  $A_2 = 2(A + A^*)$ . When all three units are operating,  $A_3 = 3(A + A^*)$ . Under the conditions of the example,  $A^* < A$ . The transition graph for this case is shown in Figure 3.14c.

A comparison of these two cases with the initial system of independent units shows that both of them possess a less favorable reliability index than the initial system: in each case the transition intensities are higher than in the initial case. Thus, on average, systems with associated units reach a failure state more quickly. We finish this comparison with a comparison of the

MTTFs:

$$T_{\text{sys}}^- = \frac{1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1}{3A + 2A + A \quad 3A + 2(A + A_2) + A + A}$$

and

$$T_{\text{sys}}^+ = \frac{1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1}{3A + 2A + A \sim 3\{A + A^*\} + 2(A + A^*) + A}$$

### 3.8.3 Mixed Structures

We consider this concept in more detail in Chapter 9 when we address two-pole network bounds. Here we only illustrate some kind of dependence between the system's units when the system has a mixed structure. Consider the simplest series-parallel and parallel-series systems with different forms of unit dependence. To perform its operation, the system should have at least one unit of type  $A$  and at least one unit of type  $B$ . Assume that we analyze a system whose two units (say functional blocks) have their own power supply (PS). The power supply is not absolutely reliable. Of course, a power supply failure leads to an immediate failure of both units which are supplied by this PS.

First, consider a series-parallel system. There are two possibilities of switching the power supply (see Figure 3.15). Denote the probabilities of a successful operation by  $p_A$ ,  $p_B$ , and  $p_{PS}$ . Then for structure (a) we can write

$$P_a = p_{PS} [1 - (1 - p_A p_B)^2] + 2 p_{PS} q_{PS} p_A p_B$$

and for structure (b) we can write

$$P_b = p_{PS}^2 [1 - (1 - p_A p_B)^2]$$

It is obvious that structure (a) is better than structure (b).

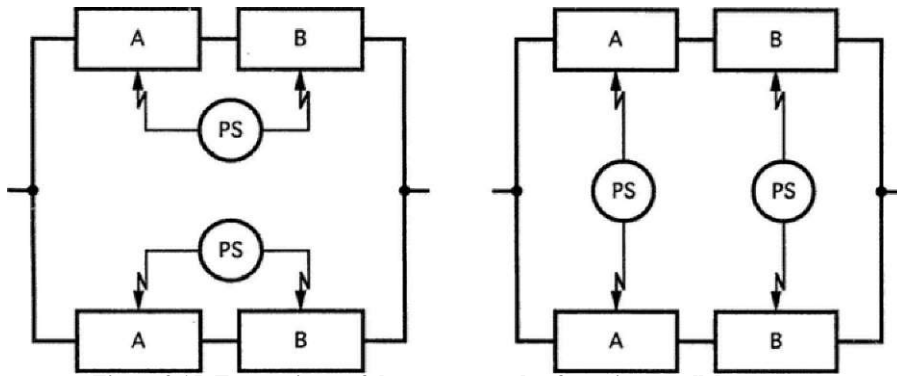


Figure 3.15. Two variants of the power supply of a series-parallel structure.

Now consider two variants of the parallel-series structure (see Figure 3.16). For structure (c) we have

$$K = P_{PS} \cdot O \sim \sim <ll) + {}^2P_{PS}IPSPAPB$$

and for structure (d) we have

Again, we can deduce that  $P_c > P_d$  without calculation, based only on our previous knowledge about the reliability of a series system with associated units.

A consideration of these examples shows us that the reliability of some auxiliary units may have an influence on other system units in such a way that the reliability of a parallel-series structure might be worse than the reliability

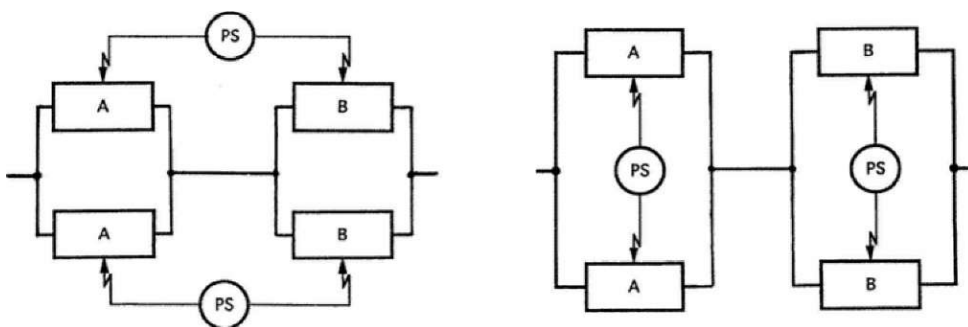


Figure 3.16. Two variants of the power supply of a parallel-series structure.

of a series-parallel structure. Indeed, it is possible that for some fixed  $p_A$ ,  $p_B$ , and  $p_{PS}$ , for instance, the inequality  $p_a > p_d$  is true:

$$P p s I^1 - (.^1-PAPB)^*| + IPpsIPSPAPB > PP^{\wedge 1} " " 1B)$$

Avoiding general deductions, take  $p_A = p_B = p$ . Denote for simplicity  $p_{PS} = /i$  and  $<2 = 1 - /?$ . Then the condition

$$/ ?^2 [1 - (1 - p^2)^2] + 2P < 2P^2 > \ll^2 (1 " Q^1)^2$$

is equivalent to

$$R^2 > \frac{(i - 0^2 - [i - 0 - P^2]^2)}{2p^2}$$

The right part of the inequality is restricted by 1 for any  $p$ . Thus if  $Q > R^2$  this inequality holds for any  $p$ . The solution of the corresponding equality gives

$$Q \ll \frac{3 \pm \sqrt{5}}{2} \approx 0.382$$

In other words, for an unreliable common unit (in our case the power supply), with a reliability index lower than approximately 0.6, one should choose a parallel-series structure rather than a series-parallel one.

For some additional examples of the analysis of systems consisting of dependent units, see Gnedenko, Belyaev, and Solov'yev (1969).

### 3.9 TWO TYPES OF FAILURES

Some units have two types of failures. For instance, a resistor may be disconnected (leaving an open circuit) in an electric circuit, and in this case no flow goes through the unit. Or it may burn out, and so will not provide any resistance at all (a short circuit). One observes an analogous effect with capacitors: no capacity at all (a short circuit) or infinite capacity (disconnection). In pipelines, holes allow the leaking of pumped liquid, which decreases the user's consumption and, simultaneously, decreases the hydraulic resistance. Rubbish in the pipe results in the same decrease in user consumption, but, at the same time, increases the hydraulic resistance.

In a most essential way, this phenomenon appears in relay circuits. These circuits are assigned for connection and disconnection and the nature of their failure can be one of two kinds: they may fail to connect or they may fail to disconnect. Each unit (relay) itself is subjected to two similar failures. It

makes the problem of redundancy of such systems more difficult: a parallel structure of relays fails if at least one unit makes a false connection when it should be disconnected, and a series structure fails if at least one unit causes a false disconnection when it should be connected. As a matter of fact, mixed series-parallel (parallel—series) structures are more effective in this case. Moreover, for a relay with known probabilities of failure of both types, there is an optimal mixed structure. We consider this problem separately in Chapter II.

Consider a parallel-series relay system. This system can be considered as a two-pole network with an input on the left and an output on the right. Each unit of the system at any moment of time can be in one of three jointly exclusive possible states: failure-free with probability  $p$ , failed in a "connected" state with probability  $c$ , or failed in a "disconnected" state with probability  $d$ . First, we consider the case where the system must provide a connection between the input and output. For each series circuit of  $n$  units, the probability of a successful connection  $R_{con}$  can be written as

$$P_{con} = (p + c)^n \tag{3.80}$$

and, for the system as a whole,

$$P_{con} = 1 - (1 - p - c)^m \tag{3-81}$$

If the system must provide a disconnection, the corresponding probabilities

$$P_{discon} = 1 - (1 - c)^m \tag{3.82}$$

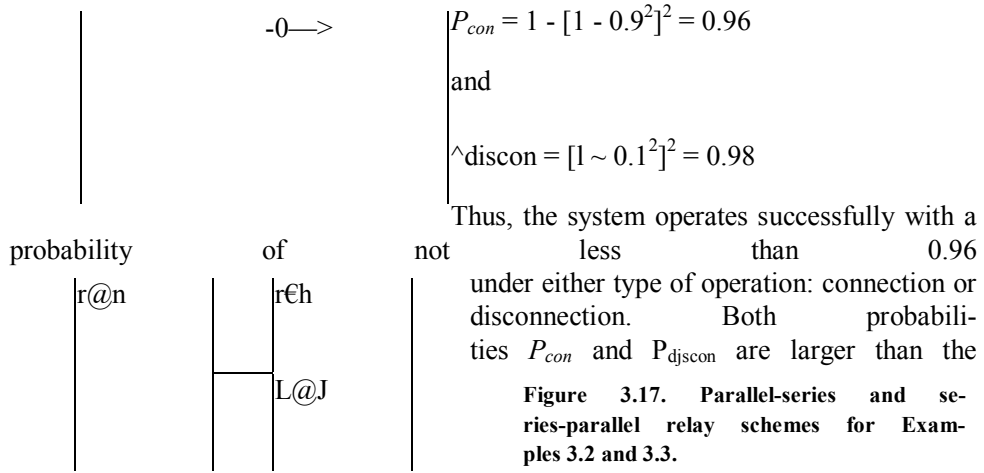
and

$$P_{discon} = 1 - cT \tag{3-83}$$

A relay system operation consists of alternating cycles of connections and disconnections. It seems that for this system it is reasonable to choose a reliability index in the form

$$P_{syst} = \min(F_{C(m)}, F_{Jiscon}) \tag{3.84}$$

It is clear that a single relay with the same parameters can perform successfully in both cases only with probability  $p$ . (Any kind of failure makes one of the operations totally impossible.)



Example 3.2 Consider a parallel-series system with  $n - m = 2$ ,  $p = 0.8$ , and  $c - d = 0.1$  (see Figure 3.17a). this system, corresponding initial probability of a single unit (under the condition that it equal 0.9). Now let us consider a general series-parallel relay system. Again consider first the case when the system must provide a connection between the input and output. For each parallel circuit of  $m$  units, the probability of a successful connection  $R_{con}$  is

$$P_{con} = 1 - (1 - p - c)^m = 1 - d^m \tag{3.85}$$

and for the system as a whole

$$P_{con} = 1 - (1 - P_{con})^n \tag{3-86}$$

If the system must provide a disconnection, the corresponding probabilities are

$$P_{discon} = (1 - P_{con})^m = (1 - d)^m \tag{3.87}$$

and

$$P_{discon} = 1 - [1 - (c + d)]^n$$

Again we can use (3.84) to characterize the system as a whole. Again in this case a single relay with the same parameters can perform successfully in both cases only with probability  $p$ .

Example 3.3 Consider a system with  $n = m = 2$ ,  $p = 0.8$ , and  $c = d = 0.1$  (see Figure 3.17). For this system,

$$P_{con} = 1 - [1 - 0.1^2]^2 \approx 0.98$$

and

$$P_{discon} = [1 - 0.9^2]^2 = 0.96$$

Thus, the system operates successfully with a probability of not less than 0.96 under either type of operation: connection or disconnection. Again both probabilities  $P_{con}$  and  $P_{discon}$  are larger than the corresponding initial probability of a single unit.

One can notice that the structures of Figures 3.17a and b are "mirror images" with respect to the probabilities  $c$  and  $d$ . Thus, both structures are equivalent for the relay with  $c = d$ .

### 3.10 MIXED STRUCTURES WITH PHYSICAL PARAMETERS

A unit presented with an indicator function  $x_t$  reflects a "dichotomic" object which can only be one of two states: for reliability problems they are termed "success" and "failure." But sometimes we need to analyze systems consisting of units with physical parameters whose particular value plays an essential role.

In Chapter 1 we introduced the *generalized generating sequence* (GGS). Here we use it and make some concrete additions to the general method. These additions are helpful for the designing of appropriate computer algorithms.

We present the discussion via simple examples.

**Series System** This case has been considered in Chapter 1. Thus, we consider only simple examples.

Example 3.4 Consider an oil pipeline consisting of  $n$  pipes (units) connected in series. Each unit has a random capacity which decreases for different reasons: the accumulation of so-called "heavy" fractions on the pipe

walls, a deformation of pipes, and so forth. The distribution of the capacity for each pipe is supposed to be known. We also assume that the distributions are determined by a finite number of values, say  $v_i$ , for the  $i$ th pipe. The GGS for the  $i$ th pipe is presented in the form of the legion

$$L_i = \{(c_i * P_i) \gg \langle V \rangle_{i(s_i)}\}$$

Here we try to avoid the complexity of a general notation and so denote  $M_{ik1} = c_{kk}$  and  $M_{ik2} = p_{,k}$ , which corresponds to their natural notation as a capacity and a probability.

The interaction of  $n$  legions produces  $N = n v_i$  different cohorts  $C_k$ ,  $1 < k \leq N$ ,

$$c_k = (c_k > P_k)$$

The capacity and the probability are determined by the rules of the cohort interactions: the "cells" with the values of capacities and the "cells" with the values of probabilities are considered separately. The capacity is determined by

$$= \text{ft}^c c_u = \min_{j \leq k} c_m \gg \frac{1}{j} \frac{S}{S_k} \gg \frac{?}{k} J_i$$

and the probability is determined by

$$P_k \ll \bigwedge_i P_{Ui}, = n_{P,J}, \gg_{S_i < n}$$

The operator  $ft^c$  in this particular case possesses the following property. If for two terms of the final GGS there are  $C_k$  and  $C_{k+i}$  with  $c_k = c_{k+i}$ , then these two terms form a new term with parameters  $c^*$  and  $p_k$  determined by

$$C^* = C_{k+i} \text{ and } p_k = p_{k+i}$$

Let us call this the *absorption property*.

Now assume that there is a known failure criterion for this pipeline, for example, suppose it is considered to be failed when  $c_k < c^0$ . In this case, to obtain the resulting reliability index, one has to revise the operators  $ft^c$  and  $Q^c$  in an appropriate way.

If  $c_k$  must be larger than  $c^0$ , the actual capacity does not play any role. The operator  $ft^c$  must be determined in such a way that any  $c_k \geq c^0$  might be considered as some  $c_{accept}$  and the remaining  $c_k$  s are set equal to 0. In this case one has cohorts of two types: the ones with  $c_{accept}$  and the other with 0. Incidentally, a computer procedure for finding the minimal value may



be solved in a sequential way:

$$c^* = \min(c_1, c_2)$$

$$c^* = \min(c^*, c_3)$$

$$c^* = \min(c^*, c_n) - \min(c, 0)$$

One may stop the procedure as soon as the value  $c^* < c^0$  appears at some intermediate step of the calculation.

Now let possess the above-mentioned absorption property and, additionally, the *preference property*. In our case the latter means that if two cohorts have different sets of maniples, then under some specified conditions the one which possesses the "better" maniple is kept for further consideration and the one with the "worse" maniple is excluded.

In the case of cohort interaction we, at first, use the absorption property and obtain the final legion in an intermediate form

$$\hat{P} = ((C_{accept}, P), (0, p^0))$$

where  $p$  is the sum of all  $p$ 's of the cohorts with  $c_{accept}$ . The resulting legion,  $P$  after applying the preference operation, will have the form

$$L - (c)$$

It is clear that  $R_{xyst} = p$ .

**A Parallel System** The formal technique used for parallel systems completely coincides with the above-described method. But, for convenience, we will use the corresponding operators  $13^L$ ,  $U^c$ , and  $13^M$ . We use these new symbols to distinguish the operations over the maniples. Indeed, for instance, for resistance

$$t Sn \quad \sum_{1 \leq i \leq n} \frac{1}{r_i}$$

$$O'' = \sum_{1 \leq i \leq n} r_i$$

and

for the time to failure caused by a short connection

$$n^M \quad \max_{1 \leq i \leq n} SiSn$$

**TABLE 3.2 Main Kinds of Maniple Interactions**

Physical Nature of Maniple	Series Structure	Parallel Structure $f_{1<w>}$
Probability of success (a cutoff failure type)	$w_1, w_2$	$1 - (1 - W_1)(1 - W_2)$
Probability of success (a short-connection failure type)	$1 - (1 - W_1)(1 - W_2)$	
Probability of failure (a cutoff failure type)	$1 - (1 - w_1)(1 - W_2)$	$W_1 w_2$
Probability of failure (a short-connection failure type)	$w_1 w_2$	$1 - (1 - W_1)(1 - W_2)$
Random TTF for a cutoff	$\min\{w_1, w_2\}$	$\max\{W_1, W_2\}$
Random TTF for a short connection	$\max\{w_1, w_2\}$	$\min\{W_1, W_2\}$
Electrical capacity	$[w_1^{-1} + w_2^{-1}]^{-1}$	$W_1 + W_2$
Ohmic resistance	$W_1 + W_2$	$[w_1^{-1} + w_2^{-1}]^{-1}$
Ohmic conductivity	$W_1 + W_2$	$[w_1 + w_2]$
Capacity of communication channel	$\min\{w_1, w_2\}$	$W_1 + W_2$
Cost of transportation through network	$W_1 + W_2$	$\min\{w_1, w_2\}$

and

$$U^M f_t = \min_{i \in \{1, 2, \dots, n\}} f_{t_i}$$

and so on.

In Table 3.2 we present the principal kinds of maniple interaction operators for systems with series and parallel structures (the considered parameters are denoted in the table by  $w$ ).

Of course, the functions listed in Table 3.2 do not exhaust all possibilities.

After this brief review of the possible interactions between maniples, we might begin with a consideration of the system with a reducible structure.

**Mixed Structures** Here we illustrate how to use the GGS for the analysis of a mixed structure with a simple example. We will not produce detailed transformations and calculations because for us and for the reader it is more reasonable to leave it to a computer. We ignore the physical nature of the system and its units for the moment.

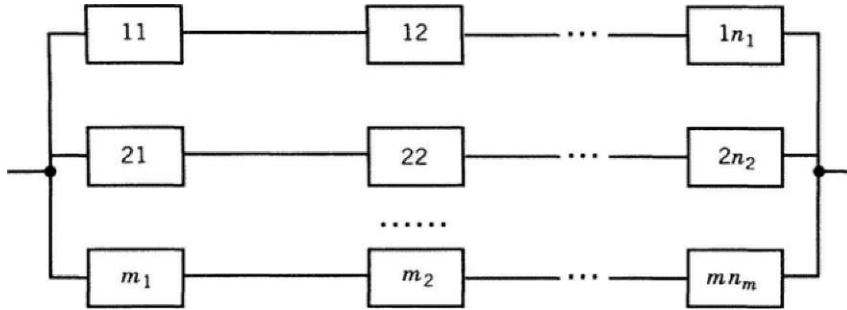


Figure 3.18. Structure of the system considered in Example 3.5.

Example 3.5 A parallel-series system is presented in Figure 3.18. For the system

$$L = U^L a^L L_{jj}$$

U/s m 5 in,

The remaining interactions depend on the concrete nature of the system.

Example 3.6 The series-parallel system is presented in Figure 3.19. For the system

$$L = a^L U^L L_{jj}$$

1 £ i £ n 1 i/ifSm,

The remaining interactions again depend on the concrete nature of the system.

Example 3.7 The system with a mixed structure is presented in Figure 3.20. For this system the following chain of operations for obtaining the resulting

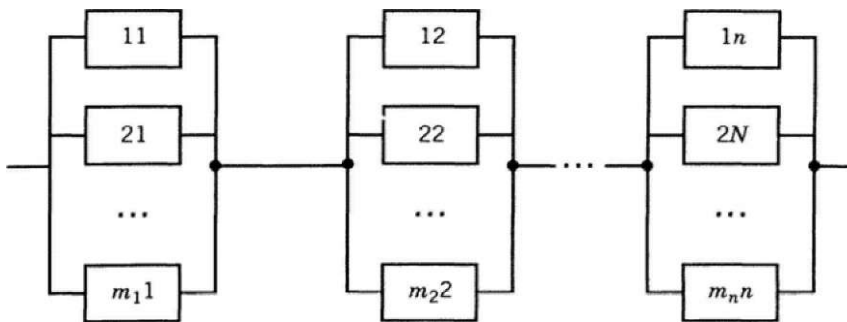


Figure 3.19. Structure of the system considered in Example 3.6.

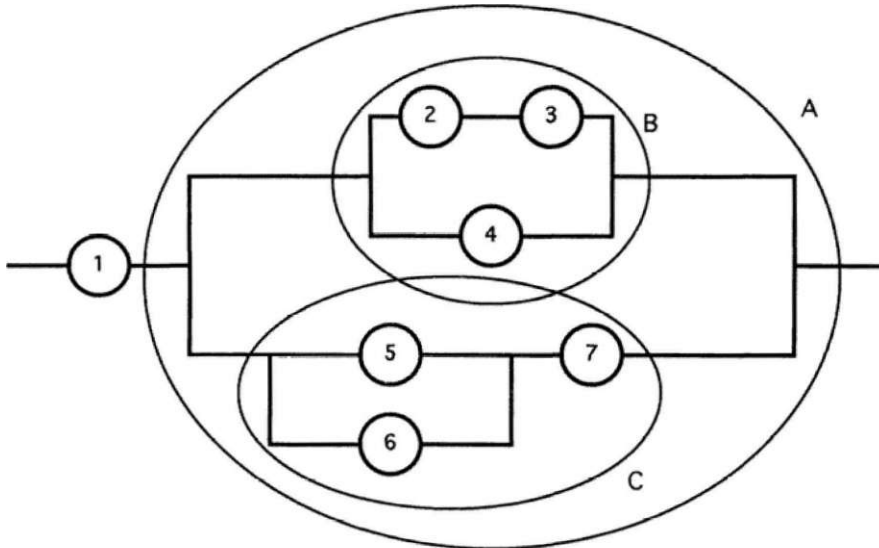


Figure 3.20. Structure of the system considered in Example 3.7.

legion can be written: For the system as a whole, Figure 3.21a, we obtain

$$L = fI^j(L, L^{(A)})$$

Subsystem (A) can be presented itself as Figure 3.216. Then

$$D^A = U_i(U^b L^{(C)})$$

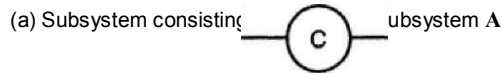
We will write expressions for  $L^{(B)}$  and  $L^{(C)}$ , again with no explanations (use the ancient rule: "see the sketch"):

$$L^{(B)} = 13^L(L_4, CI^L(L_2, L_3))$$

and

$$L^{(C)} = CI^L(L_7, U^L(L_5, L_6))$$

Therefore, the final macroalgorithm for the system GGS computation can now be written in the final form



(6) Subsystem consisting of subsystem B and subsystem C

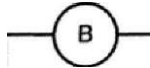


Figure 3.21. Sequence of the system structure transformation for Example 3.7.

(c) Subsystem B

**CONCLUSION**

In general, the investigation of unrepairable systems—series and parallel—can be reduced to combinatorial problems. It is almost impossible to find a track to the first works ~~CD~~ in this area. We suspect that if one finds such a work it would be (in terms of the terminology) a work of one of the three Bernoullis—Jacob, Daniel, or Nicholas! Seriously speaking, almost all of the first works and reports on reliability contained such types of analysis. The methods of analysis of unloaded redundancy have the same long history.

Therefore, we restrict ourselves to the following comments. We would only like to mention that some special problems (aging systems, systems with an irreducible structure) will be considered in the following chapters. The reader can find material dedicated to this problem in almost any book on reliability theory or engineering (see the list of general references at the end of this book). We find that for general purposes it is enough to refer to handbooks.

## REFERENCES

- Barlow R. E., and F. Proschan (1975). *Statistical Theory of Reliability and Life Testing*. New York; Holt, Rinehart and Winston.
- Gnedenko, B. V., Yu. K. Belyayev, and A. D. Solovyev (1969). *Mathematical Methods in Reliability Theory*. San Diego: Academic Press.
- Kozlov, B. A., and I. A. Ushakov (1970). *Reliability Handbook*. New York: Holt, Rinehart, and Winston.
- Ushakov, I. A., ed. (1985). *Reliability of Technical Systems: Handbook (in Russian)*. Moscow: Radio i Sviaz.
- Ushakov, I. A., ed. (1994). *Handbook of Reliability Engineering*. New York: Wiley.

## EXERCISES

- 3.1 Prove (3.22a) using the Venn diagram.
- 3.2 Prove (3.22i>), (3.22c), and (3.22\*/) on the basis of the result of Exercise 3.1. {Hint: Use the "double rejection" rule of Boolean algebra:  $x = x \cdot x$ .}
- 3.3 Prove identities from (3.23 a) to (3.23 d).
- 3.4 Write the Boolean function  $\pi(X)$  for the scheme depicted in Figure E.3.2.

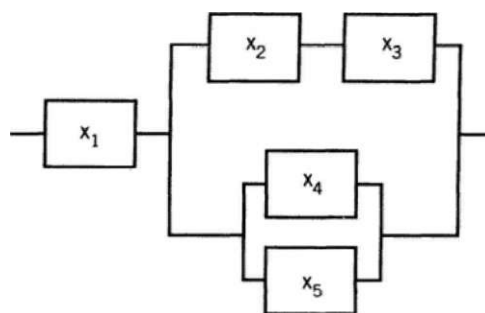


Figure E3.2.

- 3.5 Write the Boolean function  $\pi(X)$  for the scheme depicted in Figure E3.3.

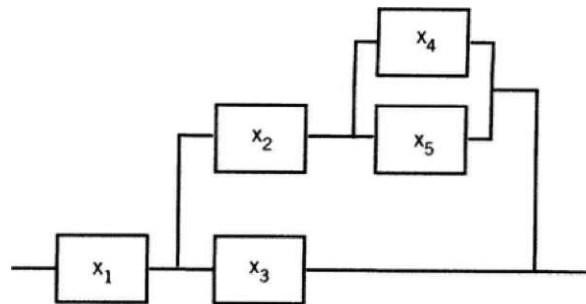


Figure E3-3.

- 3.6 A system consists of 10 identical and independent units connected in series. The requirement of the probability of a failure-free operation equals 0.99. What reliability level must a system unit have to satisfy the system requirements?
- 3.7 A system consists of three identical units connected in parallel. The requirement of the probability of a failure-free operation equals 0.999. What reliability level must a system unit have to satisfy the system requirements?

**SOLUTIONS**

- 3.1 For given sets  $X$  and  $Y$  (see the shadowed areas in Figures E3.1a and  $b$ , the union is a set of elements belonging to at least one of them (see the shadowed area in Figure E3.1c). Then  $X$  and  $Y$  are depicted in Figures E3.1  $d$  and  $e$  (see the shadowed areas). In Figure E3.1/ one finds the area  $X \cap Y$  shadowed. Consequently, the complementary area is  $X \cup Y$ . Obviously, the latter area coincides with the shadowed area in Figure E3.1c. Thus, the desired result is obtained.
- 3.2 For example, let us prove identity (3.22b)

$$X \cap Y = X \cup Y$$

Take a rejection operation from both sides of the identity which does not violate it

$$X \cap Y = X \cup P = X \cup Y$$

Now use a rejection operation to all arguments which also does not

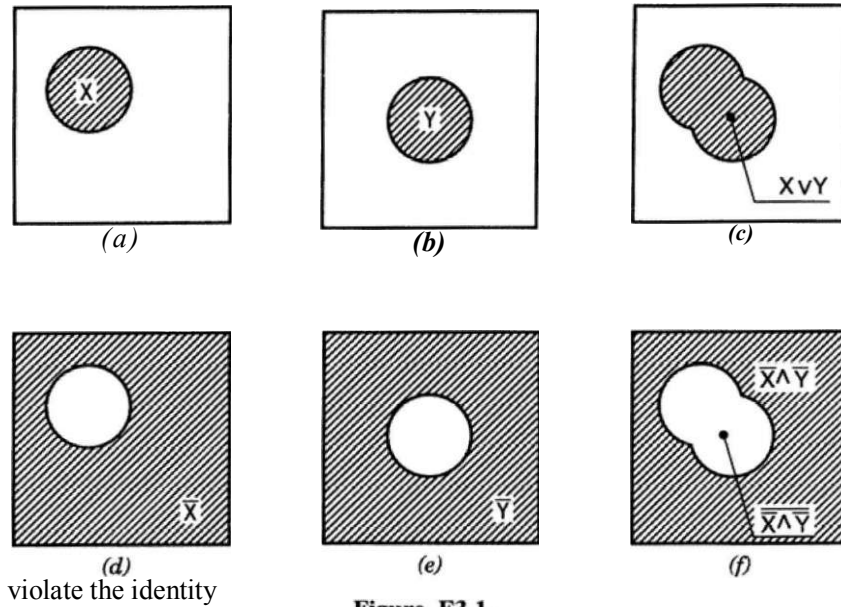


Figure E3.1.

$$X \cap Y = \overline{X \cap Y} \cap \overline{X \cap Y}$$

Thus, this identity is reduced to the first one, (3.22a), which was proven in the previous exercise.

3.3 Using (3.22a), one has for the three arguments

$$X_1 \cap X_2 \cap X_3 = (X_1 \cap X_2) \cap X_3 = (X_1 \cap X_2) \cap X_3$$

Using the identity

$$X_1 \cap X_2 = X_1 \cap X_2$$

one finally obtains

$$X_1 \cap X_2 \cap X_3 = X_1 \cap X_2 \cap X_3$$

Now if for  $n = 1$  arguments we have

$$\bigcap_{i=1}^n X_i = 0$$





then, using the previous rule, one finally obtains

$$1 \text{ SiSn-1} \cup X_i = \left( \begin{array}{l} u \vee \\ F M - 1 \quad ' I \wedge i V \ll I. \\ 1 \text{ SiSn-1} \quad M < iSB - I \end{array} \right.$$

$$1 \text{ Si SB - 1} \quad \begin{array}{l} I < i < n \\ J \text{ SI SB - 1} \end{array} \quad A X_n$$

$$1 \text{ SiSB-1} \quad u x_j \wedge x_n \quad \bigvee_{i=1}^n X_i$$

$$\bigvee_{i=1}^n \text{ SiSB - i} \quad \begin{array}{l} n x \setminus ax_j \sim n x_i \\ 1 < /Sn \end{array}$$

This completes the proof.

Denote  $Y_1 = X_1 \wedge X_2$  and  $Y_2 = X_4 \vee A \wedge$ . In this notation

$$1 \text{ SiSB-1} \quad A \left[ \langle *, A * \rangle \vee (*_4 \vee *_5) \right]$$

In reliability computational practice, one usually uses such expressions without "ORs"; that is, one reduces an initial form in a special way using DeMorgan's rule. In the example under consideration, one has

$$\langle P(X_i) = J f, A \left[ (A \cup \dots) \wedge A \right]$$

Denote  $X_4 \vee X_5 = Y_1$ ,  $X_2 \wedge Y_1 = Y_2$ ,  $Y_2 \vee X_3 = Y_4$ . In this notation  $\text{tpi}(X_i) = X_i \wedge Y_4$  or in open form

$$- x_j \wedge \{ X_3 \vee [X_2 \wedge (X_4 \vee X_5)] \}$$

The final expression using only logic AND and rejection operators has the form

$$X_x \wedge \overline{\left\{ \overline{X_3} \wedge \left[ \overline{X_2} \wedge \overline{(\overline{X_4} \wedge \overline{X_5})} \right] \right\}}$$

A unit has to have  $p = 1 - y^{199} \sim 0.999$ .

Let an unknown probability of failure of a unit be denoted by  $q$ . For the system under consideration, one can write

$$1 - 0.999 = q^3$$

Thus,  $q = \text{kt.OOI} = 0.1$ , that is,  $p = 0.9$ .

## CHAPTER 4

# LOAD - STRENGTH RELIABILITY MODELS

For many reliability engineering applications, one needs to investigate the ability of a structure or a piece of equipment to survive under extreme conditions. For a mechanical construction, one speaks of the probability that it can withstand a specified external load (a shock, vibration, etc.) or internal tension. For electronic equipment, one is concerned with the probability that it is able to withstand a specified voltage jump in its power supply or a significant change in its input signals.

Both an external load and a construction strength might be considered as random. The first is random in a very natural way, as it depends on environmental factors. The second is random because of the inherent instability of any technological process.

### 4.1 STATIC RELIABILITY PROBLEMS OF "LOAD - STRENGTH" TYPE

#### 4.1.1 General Expressions

Generally, the construction strength  $X$  and the applied load  $Y$  are random. The problem is to find  $R$ , the probability of the system's successful operation, that is, the probability that the applied load does not exceed the actual level of construction strength:

$$= \Pr\{X > r\} \quad (4.1)$$

167



Let  $\Pr\{A < j\} = F(x)$  and  $\Pr\{Y < x\} = G(x)$ . Then the probability of a successful operation of the system can be calculated as

$$\begin{aligned} R &= \int_{-\infty}^{\infty} \Pr\{Y > x\} dG(x) = \int_{-\infty}^{\infty} [1 - F(x)] dG(x) \\ &= \int_{-\infty}^{\infty} [1 - F(x)] g(x) dx = \int_{-\infty}^{\infty} g(x) dx - \int_{-\infty}^{\infty} F(x) g(x) dx \end{aligned} \quad (4.2)$$

If both distributions are continuous, then

$$R = \int_{-\infty}^{\infty} [1 - F(x)] g(x) dx$$

where  $f(x)$  is the density of  $F(x)$  and  $g(x)$  is the density of  $G(x)$ .

If  $X$  and  $Y$  are considered to be independent r.v.'s, it is convenient to introduce a new random variable,  $Z = X - Y$ , with distribution  $H(x)$ . Then (4.1) can be rewritten in the form

$$R = \Pr(Z > 0) = \int_0^{\infty} dH(x) \quad (4.3)$$

#### 4.1.2 Several Particular Cases

***Fix) and G(x) Are Normal*** In this case

$$f(x) = \frac{1}{\sigma_f \sqrt{2\pi}} e^{-\frac{(x - \mu_f)^2}{2\sigma_f^2}} \quad (4.4)$$

where  $\mu_f$  and  $\sigma_f$  are, respectively, the mean and the standard deviation of the strength's distribution  $F(x)$ , and

$$g(x) = \frac{1}{\sigma_g \sqrt{2\pi}} e^{-\frac{(x - \mu_g)^2}{2\sigma_g^2}} \quad (4.5)$$

where  $\mu_g$  and  $\sigma_g$  are, respectively, the mean and the standard deviation of the load's distribution  $G(x)$ .

Notice that we consider the area of domain of both distributions to range from  $-\infty$  to  $\infty$ . Of course, one should consider truncated distributions such that their r.v.'s cannot be negative. But, in practice,  $\mu_f > 3\sigma_f$  and  $\mu_g > 3\sigma_g$ , so that such a truncation does not lead to any crucial numerical errors.

Now introduce the new r.v.,  $Z = X - Y$ . The mean of this new r.v. equals  $E\{Z\} = S - L$  and

$$\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2}$$

which immediately gives the required result

$$R = \Pr(Z > 0) = \int_0^{\infty} \frac{1}{\sigma_Z \sqrt{2\pi}} \exp\left\{-\frac{(x - E\{Z\})^2}{2\sigma_Z^2}\right\} dx = \Phi\left(\frac{S - L}{\sigma_Z}\right) \quad (4.6)$$

Numerical results can be found from a standard table of the normal distribution.

From (4.6) one can see that the reliability of the construction decreases if the variances of  $X$  and/or  $Y$  increase. Roughly speaking, the more uncertain the conditions of use and the more unstable the quality of the construction, the lower is the reliability of the construction.

**Example 4.1** The span of a bridge has a safety coefficient  $c_s$  equal to 5. The safety coefficient is determined as  $c_s = S/L$ . The coefficient of variation of the strength  $K_s$  equals 0.05 and that of the load  $K$ , equals 0.2. (a) What is the probability of a successful operation of the construction? (b) What is the probability of a successful operation of the construction if the coefficient of variation of the strength is twice as large?

*Solution.* (a) Assume that  $L = 1$ . (By an appropriate normalizing this is always possible.) Then, taking into account the value  $c_s$ , we obtain  $S = 5$ . By definition, the coefficient of variation of the r.v.  $Z$  is the ratio  $\text{Var}\{Z\}/E\{Z\}^2$ . Therefore,  $\sigma_Z = 0.2$  and  $\sigma_Z^2 = (0.05 \times 2.5)^2 = 1.25$ . The probability of a successful operation equals

$$\Phi\left(\frac{5}{\sqrt{1.25}}\right) = \Phi(4.472) = 0.999517$$

(b) In this case the value of the variance is 2.50 and the probability of a successful operation equals

**F(x) and G(x) Are Exponential** In this case

$$(4.7)$$

where  $S$  is the main strength, and

$$g(x) \ll -e^{-x/L}$$

where  $L$  is the mean load. Using (4.2), we obtain

$$R = \int_0^{\infty} \int_0^{\infty} e^{-x/L} e^{-y/S} dx dy = \int_0^{\infty} \int_0^{\infty} e^{-(1/L + 1/S)x} dx dy = \frac{S}{S+L} \quad (4.9)$$

In this case the variances do not influence the resulting probability. Of course, it should be mentioned that exponential distributions in problems such as this are very seldom encountered in practice (especially for a distribution of strength). One can find an example of this in Exercise 4.2.

**F(x) Is Normal and G(x) Is Exponential** Let us use the expression

$$(4.8)$$

$$R = \int_0^{\infty} f(x) \int_0^{\infty} g(y) dy dx$$

Notice that

$$\int_0^{\infty} g(y) dy = \int_0^{\infty} e^{-Ay} dy = 1/A$$

Therefore, we can write

$$R = \int_0^{\infty} f(x) \left[ \int_0^{\infty} e^{-Ax} e^{-\frac{1}{2} \left( \frac{x-S}{\sigma_f} \right)^2} dx \right] dx$$

$$= \int_0^{\infty} f(x) e^{-Ax} \left[ \int_0^{\infty} e^{-\frac{1}{2} \left( \frac{x-S}{\sigma_f} \right)^2} dx \right] dx$$

$$= \int_0^{\infty} f(x) e^{-Ax} \left[ \int_0^{\infty} e^{-\frac{1}{2} \left( \frac{x-S}{\sigma_f} \right)^2} dx \right] dx$$

Combine the powers of the exponential functions of the second terms to get



STATIC RELIABILITY PROBLEMS OF "LOAD - STRENGTH" TYPE 171

*f*

*Jn*

172 LOAD - STRENGTH RELIABILITY MODELS

the complete square form and the free term:

$$(x - S)^2 + \frac{x}{T} - 2af(x - S) + \frac{2S}{Lj} - \frac{iL}{L}$$

Then

$$\int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2} \left[ \frac{(x - S)^2}{T} - 2af(x - S) + \frac{2S}{Lj} - \frac{iL}{L} \right]\right\} dx$$

---


$$\int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2} \left[ \frac{(x - S)^2}{T} - 2af(x - S) + \frac{2S}{Lj} - \frac{iL}{L} \right]\right\} dx$$

Change the variables as

$$t = -5 + A\sigma y^2) \text{ and } \langle r_f dt = dx$$

Now the final expression becomes

STATIC RELIABILITY PROBLEMS OF "LOAD - STRENGTH" TYPE 173

$$1 - d \int_0^L L^2$$

$$\int_0^L \exp$$

Notice that for most practical problems such as this, the strength  $S$  should be located "far" from the point 0. This means that the value  $S/L \gg 1$ . Incidentally, this corresponds well to the assumption that we do not take into account the truncation of the normal distribution in  $x = 0$ . In this case, of

course,

$$R = 1 - \exp\left(-\frac{2}{L} \int_0^L L^2 dt\right) \quad (4.11)$$

$$s \sim t^i \quad (4.10)$$

o>

If, in addition,  $Aoy$  is small, say of order 1, then it is possible to write the next approximation

$$2 \frac{1}{L^2} L^2$$

Therefore, if one takes into account that the mean load equals  $L = 1/A$ , (4.12) can be rewritten as

$$R \sim 1 - \exp\left[-SA + \frac{1}{2}(Aoy)^2\right] \quad (4.13)$$

**$F(x)$  Is Normal and  $G(x)$  Is Biased Exponential** The biased exponential distribution with parameter  $A = 1/L$  and bias  $I^*$  is presented in Figure 4.1. In this case

$$f(x) = \begin{cases} 0 & \text{for } x < I^* \\ \frac{A \exp\{-A(x - I^*)\}}{1 - \exp\{-A(S - I^*)\}} & \text{for } x > I^* \end{cases} \quad G^*(x) = \int_0^x f(x) dx$$

where

$$G^*(x) = \begin{cases} 0 & \text{for } x < I^* \\ \frac{1 - \exp\{-A(x - I^*)\}}{1 - \exp\{-A(S - I^*)\}} & \text{for } x > I^* \end{cases} \quad (4.14)$$

After changing variables, (4.14) becomes

$$f(y) = \begin{cases} 0 & \text{for } y < S - I^* \\ \frac{\exp\{-Ay\}}{1 - \exp\{-A(S - I^*)\}} & \text{for } y > S - I^* \end{cases} \quad [1 - \exp\{-Ay\}] dy \quad (4.15)$$

Omitting the transformations which are quite similar to the above, we

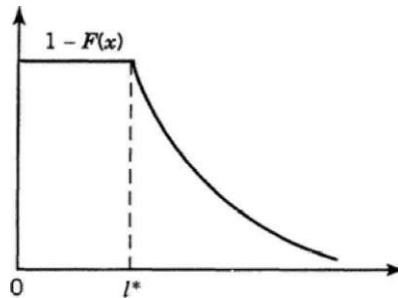


Figure 4.1. Sample of a biased exponential distribution.

present the final result

$$R = 1 - \frac{(S - I^*)}{a_f} \exp \left[ -\frac{2(S - I^*)A + AV}{(S - I^*) - A} \right] \quad (4.16)$$

In this case the similar approximate expressions become

$$R = 1 - \exp \left[ -\frac{2(S - I^*)A + AV}{(S - I^*) - A} \right] \quad (4.17)$$

for  $(S - I^*)/a_f \gg 1$  and

$$R = 1 - \exp \left[ -\frac{2(S - I^*)A + AV}{(S - I^*) - A} \right] \quad (4.18)$$

for small values of  $Aoy$ .

***F(x) Is Biased Exponential and G(x) Is Normal*** Consider the biased distribution of the strength. We do this because it is unreasonable to consider any construction with a strength equal to 0. By assumption, the strength might not be less than  $s^*$ , so in this case

$$R = \int_0^x g(x) dx + \int_{s^*}^x f(y) dy \quad (4.19)$$

Notice that we again use the lower limit of 0 in the integral. As we pointed out above, for numerical calculations the truncation of the normal distribution at  $f = 0$  to the left can be neglected.

A simple transformation leads to

$$R = \int_{s^*}^x g(x) dx + \int_{s^*}^x p(x + s^*) e^{-x} dx \quad (4.20)$$

176 LOAD - STRENGTH RELIABILITY MODELS

*jj*

$$R = \int_a^{\infty} \frac{1}{s^* - L} \exp\left\{-\left(\frac{y-L}{s^*}\right)^n\right\} dy$$

exp

$$\left\{ \frac{y-L}{s^*} \right\}^n$$

$$R = \int_a^{\infty} \frac{1}{s^* - L} \exp\left\{-\left(\frac{y-L}{s^*}\right)^n\right\} dy$$

or, in detailed form,

$$\exp\left\{-\left(\frac{y-L}{s^*}\right)^n\right\} dy \quad (4.21)$$

Avoiding repetition of the transformation which completely coincide with those above, we write the final result directly as

$$R = \int_a^{\infty} \frac{1}{s^* - L} \exp\left\{-\left(\frac{y-L}{s^*}\right)^n\right\} dy \quad (4.22)$$

Additional results for some other important particular cases can be found in Kapur and Lamberson(1977), We would like to mention that this reference contains useful formulas for the Weibull-Gnedenko distribution which is important for description of the strength of mechanical construction.

### 4.1.3 Numerical Method

In general, it is reasonable to use an approximate numerical method. This method is good for calculations using histograms as well as standard statistical tables. In the first case, the approximation is defined by restricted statistical data and their inevitably discrete nature. In the second case, the approximate nature of the solution is explained by a discrete representation of continuous distributions. Because of the approximate nature of these calculations, it is sometimes reasonable to consider the upper and lower bounds of the calculated values.

First, assume that a set of statistical input data is given. The set of observed values of the material strength is  $X_1, \dots, X_n$ , and the set of observed values of the load is  $Y_1, \dots, Y_m$ . Arrange the ordered set  $W_1, \dots, W_{n+m}$  where each  $W_s$  is one of the  $Y$ 's or one of the  $X$ 's.

#### 178 LOAD - STRENGTH RELIABILITY MODELS

For each  $W_s = X$ , calculate the number of  $W_r$ 's where  $W_r = Y_j$  and  $r < s$ . Denote this value by  $k_s$ . This value means that, on the average, in  $k_s$  cases of  $m$  possible observations of the r.v.  $Y$ , the load will be smaller than the given strength  $X$ . In other words, we might say that with conditional probability  $k_s/m$ , the investigated system with fixed strength  $X$ , will operate successfully if the load will take on one of the possible values of  $Y$ . Thus, the complete probability of success is

Obviously, the same numerical result can be obtained if we consider  $W_s = Y_j$  and calculate the number of  $W_r$ 's where  $W_r = X_L$  for each  $W_s = Y_j$ ,  $r < s$ . Denote this value by  $k^*$ . This value means that in  $k^*$  cases of  $m$



STATIC RELIABILITY PROBLEMS OF "LOAD - STRENGTH" TYPE 179

possible observations of the r.v.  $X$ , the strength will be smaller than the specified load  $Y$ . It means that with conditional probability  $k^*/m$ , the investigated system will fail under the load  $Y$ ; that is, the complete probability of success is

$$R = 1 - \sum_{m=1}^{\infty} Z_k t$$

Example 4.2 The following data are available:  $X_1 = 98.1, X_2 = 98.2, X_3 = 99.4, X_4 = 100.3, X_5 = 101.2, X_6 = 103.5, X_7 = 103.9, X_8 = 104.1, \dots, X_{16} = 110.2$ ;  $Y_1 = 79.1, Y_2 = 82.4, \dots, Y_{16} = 98.0, Y_{17} = 98.3, Y_{18} = 98.5, Y_{19} = 99.5$ . Calculate the probability that the construction will operate successfully.

Solution. We find from the data that  $k_1 = 18/21, k_2 = 18/21, k_3 = 20/21, k_4 = \dots = k_{15} = 1$ . Thus, the result taken by the first expression is

$$1 - \frac{18}{21} - \frac{18}{21} - \frac{20}{21} - \dots - 1 = 1 - \frac{18 + 18 + 20 + \dots + 15 \cdot 21}{21^{16}} \approx 0.9792$$

From the same data, one finds that  $k^* = \dots = k_{15} = 1, k_{16} = 0, k_{17} = 2/16, k_{18} = 20/16, k_{19} = 2/16, k_{20} = 3/16$ .

Using the second expression, one has the following result:  $1 - \frac{2}{16} - \frac{2}{16} - \frac{3}{16} - \dots = 1 - \frac{2 + 2 + 3 + \dots}{336} \approx 0.9792$

If tables of the distributions  $F(x)$  and  $G(y)$  are available, the numerical calculation of the index  $R$  can be performed using the following formulas:

$$R = \sum_{m=1}^{\infty} \frac{1}{SM} [F(mA) + G((m+1)A) - G(mA)] \quad (4.23)$$

$$R = \sum_{m=1}^{\infty} \frac{G((m+1)A) - G(mA)}{SM} \quad (4.24)$$

where  $A$  is the chosen increment and  $M$  is the number of increments. It is clear that the summation can only be performed in the area of the distribution's domain where the corresponding values of the product terms are significant. For practical purposes, the increments may be chosen to have a value ranging from 0.5 to 0.05 of the smallest standard deviations of the distributions  $F(x)$  and  $G(y)$ . Obviously, the more accurate the result that is

#### 180 LOAD - STRENGTH RELIABILITY MODELS

needed, the smaller the increments must be. For practical calculations, the left bound  $m$  of the summation must begin with the value  $k = \{m: F(-mA) < e\}$  where  $e$  is chosen in correspondence with the needed accuracy.

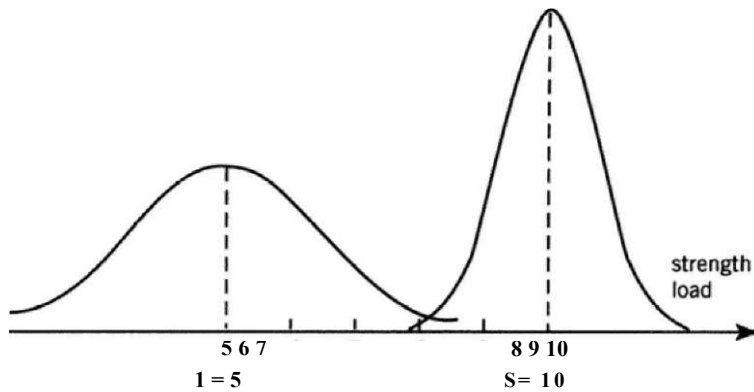


Figure 4.2. Explanation of the solution of Example 4.3.

**Example 4.3** The strength has a normal distribution with  $S = 10$  and  $ar_f = 1$ , and the load also has a normal distribution with  $L = 5$  and  $cr_g = 2$  (all values are measured in some conditional scales). Calculate the probability of success R using a standard table of the normal distribution.

*Solution.* We present Figure 4,2 to illustrate the solution. This figure helps us to see that, for example, the point 7 corresponds to  $L + cr$ , and, at the same time, corresponds to  $S - 3\sigma$ , and the point 9 corresponds to  $L + 2\sigma$  and to  $S - \sigma$ , and so on. Use a standard table of the normal distribution and arrange (only for illustrative purposes) the new Table 4.1 with the input data for numerical calculation. Thus, the probability of failure equals 0.98385. A calculation with the use of the strong formula gives

$$\Phi\left(\frac{10 - 5}{\sqrt{1^2 + 2^2}}\right) = \Phi(\sqrt{5}) = 0.9878$$

TABLE

4.1

	Value of $G(k + 1) - G(k)$	Argument $k + 1/2$	Value of $F(k + 1/2)$	Intermediate
[7,8]			0.0062	0.00057
[8,9]	0.0920	7.5	1	0.00295
Interval	0.0441	8.5	0.0668	0.01114
[k, k + 1]	0.0164	9.5	0.692	0.00149
1]	0.0049	10.	0.308	
		5		

182 LOAD - STRENGTH RELIABILITY MODELS

0.01615

The relatively large error is explained by the use of excessively large increments.

**REMARK.** The unreasonably high level of accuracy of the calculations is presented only to compare the obtained solution with the exact solution. Once more we would like to emphasize that for practical purposes the use of "too accurate" a solution can be considered almost incorrect because of the very rough statistical data which we usually have in practice.

Sometimes it might be more useful to obtain lower and upper bounds on the value  $R$  because this allows one to evaluate the accuracy of the result. Lower bounds can be written as

$$R = \int_{1 \leq m < M} [1 - F(m + 1)A][G((m + 1)A) - G(mA)] \quad (4.25)$$

$$R = \int_{1 \leq m < M} G(m)A[F((m+1)A) - f(mi)] \quad (4.26)$$

and upper bounds as

$$r = \int_{1 \leq m < M} [1 - F(m)A][G((m + 1)A) - G(mA)] \quad (4.27)$$

$$R = E G\{(m + 1)A\} \{F((m + 1)A) - F(mA)\} \quad (4.28)$$

**Example 4.4** Suppose the construction has a truncated exponential distribution of the strength with parameters  $p = 1/S = 0.5$  and  $s^* = 10$ , and a normal distribution of the load with parameters  $L = 6$  and  $a_g = 2$ . Find the probability of success  $R$  for this construction.

*Solution.* Find upper and lower bounds on the probability  $R$ . For the purpose of numerical calculation, construct a special table (see Table 4.2) based on standard tables of the normal and exponential distributions. Table 4.2 contains the meaning of the corresponding distribution  $G(x)$  and the

**TABLE 4.2**

$m$	$x$	$ z $	$O(m)$		$Fim)$	$Aim)$
1	10.0	2.00	0.9773	0.00	0.000	0.221
2	10.5	2.25	0.9878	0.25	0.221	0.173
3	11.0	2.50	0.9938	0.50	0.394	0.134
5	11.5	2.75	0.9970	0.75	0.528	0.104
7	12.0	3.00	0.9987	1.00	0.632	

increments of  $Fix$  in the area of interest. The calculation of lower and upper bounds is performed using formulas (4.26) and (4.28).

In Table 4.2,  $m$  is the number of the term in the sum,  $x$  is the absolute value,  $z_1$  is the argument of the standard normal distribution, and  $z_2$  is the argument of the standard exponential function  $A^*(m) = A i m + I) - A(m)$ . Using (4.26), we obtain

$$R = (0.9773)(0.221) + (0.9878)(0.173) \\ + (0.9938)(0.134) + (0.9970)(0.104) + r = 0.9917$$

where  $r$  is the probability of the "tail" of the strength's distribution with an insignificant influence of the load (all this area must be considered as the area of the "practically absolute" reliability). Using (4.28), we obtain

$$R = (0.09878)(0.221) + (0.9938)(0.173) \\ + (0.9970)(0.134) + (0.9987)(0.104) + r = 0.9956$$

The difference between the two values is significant—about 100%. (Notice that if the probabilities are close to 1, one should consider the complementary probabilities, i.e., 0.0083 and 0.0044 in the investigated case.) This means that the values of  $A$  are chosen too large.

## 4.2 MODELS OF CYCLE LOADING

The static models of the "strength-load" type which we considered in the previous section may be referred to as one-cycle loading models. Moreover, this single cycle is assumed to be short enough (but not a shock!), so the strength of the material is assumed to be constant in time. In other words, there is no time for any deterioration or fatigue effects to appear. For practical tasks, such a consideration is important even if the cycles are considered to be independent and identical. Anyway, this more accurately reflects the physical process than a totally static situation. A consideration of the cycle loading is supported by the results of the previous section: the probability determined there is considered as a characteristic of the ability of the chosen construction to withstand a specified fixed load during one cycle.

In real life, the strength of a mechanical construct might monotonically change in time, due to deterioration, fatigue and aging processes, environmental influences and so on. (For electronic equipment, the "strength" can fluctuate: the actual tolerance limits can change in time depending on the temperature, humidity, and other environmental influences. Below we often refer to mechanical systems.) The load can also change in time for various obvious reasons. We consider only a simple case: a sequence of shock-type (practically instantaneous) loading. Notice that an investigation of continuous

loading with a simultaneous changing of the load and strength is a very sophisticated physical problem which can only be solved for some particular cases.

We will discuss a very particular case of cycle loading when the strength  $X$  and the load  $Y$  are independent random variables with known distributions  $F(x)$  and  $G(y)$ , respectively. For compactness, the mean value of the strength  $E\{X\}$  will be denoted by  $S$  and the mean value of the load  $E\{Y\}$  by  $L$ . The strength is assumed to be fixed (known or unknown) or monotonically changing and the load can be represented by a sequence of independent r.v.'s from cycle to cycle.

#### 4.2.1 Fixed Level of Strength

**Known Fixed Level** Suppose that the level of strength is known and equals some value  $s^0$ , while the load is random with distribution function  $G(y)$ . The values of the load at each cycle are mutually independent r.v.'s. Denote the random number of failure-free cycles by  $v$ .

The probability that exactly  $k$  cycles will be successful equals  $\Pr\{v = k | s^0\} = p^k q$ ; that is, the r.v.  $v$  has a geometrical d.f. with  $p = G(s^0)$  and  $q = 1 - p$ . The probability of success during  $K$  or more cycles equals

$$\Pr\{v \geq K | s^0\} = p^K \text{ where } p = \Pr\{Y \leq s^0\} = G(s^0)$$

The mean number of cycles before failure equals  $EM = 1/q$ . If  $q \ll 1$ , an approximation in exponential form can be written as

$$\Pr\{v \geq K | s^0\} = e^{-qK}$$

**Unknown Fixed Level** Now assume that  $s^0$  is unknown but constant during the total period of the system's operation. The only thing we know is the prior distribution  $F(x)$ . In this case

$$\Pr\{v \geq K\} = \int_C [G(x)]^K dF(x)$$

where  $C$  is the domain of the distribution  $F(x)$ .

Let the probability  $q(x) = 1 - G(x)$  be small "on average." In practice, this corresponds to the condition

$$\frac{S - L}{\sqrt{\sigma_S^2 + \sigma_L^2}} \gg 1$$

where  $\sigma_S$  is the standard deviation of the d.f.  $F(x)$ . We can conclude that the



right "tail" of the distribution  $G(x)$  is concave in the essential area of the domain of the distribution  $F(x)$ . Then the following simple bound is true:

$$\Pr\{f > A\} < [G(S)]^*$$

The mean number of cycles is

$$EH = \int_c^{\infty} [1 - G(x)]^* dx / F(x)$$

The corresponding approximation for small  $q$  is

$$EH < [1 - G(S)]^{-1}$$

#### 4.2.2 Deteriorating Strength

If the expected number of successful cycles is very large, that is, the operational time is sufficiently large, we might assume that the level of the system's strength decreases in time. Indeed, most materials deteriorate with time and, consequently, the system strength becomes weaker and weaker. We will consider several simple models.

1. Assume that the material strength decreases from cycle to cycle in such a way that  $p_{k+i} = ap_k$  where  $p_k$  is the probability of success at the  $k$ th cycle and  $a$  is constant,  $0 < a < 1$ . Then

$$\Pr\{i \leq K | s_0, a\} = p(pa)(pa^2) \dots \{p < a^K\} \\ = p^K \int_0^{\infty} a^k p^k a^{k^2/2}$$

and the mean number of successful cycles is

$$EH = \int_0^{\infty} pV^{*-1} dx$$

2. Again consider the case where the level of the strength is known and the deterioration is described by an exponential decrease of this level: at the  $k$ th cycle, the level of the strength is  $x_k = s_0 a^k$  where  $0 < a < 1$ . The probability of success over at least  $K$  cycles equals

$$\Pr\{y > X^* | s_0, a\} = \int_0^{\infty} G(s_0 a^k)$$

Note that  $G(s_0 a^k) > \dots$  and for the most commonly used distributions, this discrete function is concave. Then

$$\Pr\{y > K | s_0, a\} < G(s_0 a^{K/2})$$

The mean number of cycles until the system fails equals

$$W = \frac{1}{p_1} + \frac{2}{p_1 p_2} + \frac{3}{p_1 p_2 p_3} + \dots$$

$$= \frac{1}{p_1} (1 + \frac{p_2}{p_1} + \frac{p_2^2}{p_1^2} + \dots)$$

3. For a known distribution  $F(x)$  of the initial value of the strength, the probability of success equals

$$Pr\{K \leq F(x), a\} = \int_0^a G(x^k) dF(x)$$

We do not have a simple approximation for this case,

#### 4.3 DYNAMIC MODELS OF "STRENGTH - LOAD" TYPE

In the previous sections we considered a simple version of the dynamic loading process, that is, the cycling process. That scheme is sufficiently good to describe some specific mechanical systems. But for most electronic systems the process of "loading" should be described as a continuous stochastic process. Indeed, in this case one considers a process of randomly changing the system parameters inside the tolerance zone. We will consider only a simple case where one-dimensional stochastic process crosses a specified level.

**4.3.1 General Case** Consider a differentiable stochastic process  $x(t)$ . We are interested in the distribution of intervals between neighboring intersections of a specified level  $a$  by the process. At first, we find the probability that the process will intersect the level  $a$  at moment  $t$ . This event happens if the two following events have occurred:

$$\{x(t) < a\} \text{ and } \{x(t+dt) > a\}$$

In other words, the probability of the event equals

$$Pr\{x(t) < a \text{ and } x(t+dt) > a\} \quad (4.29)$$

Let  $v(t)$  be the speed of the process, that is,  $v(t) = dx(t)/dt$ . Now we can rewrite (4.29) in the new form

$$Pr\{x(t) - v(t) dt < x(t) < a\} \quad (4.30)$$

To find this probability, we need to know the density function of the joint distribution  $f(x, v|t)$  of the ordinate  $x$  of the process  $x(t)$  and its derivative for the same moment of time  $t$ . Using these terms, we can write

$$Pr\{x(t) - v(t) dt < x(t) < a\} = \int_0^a \int_{a-v dt}^{\infty} f(x, v|t) dx dt \quad (4.31)$$

The internal integral can be computed instantly because of its special limits

$$\int_{x-vdt}^x f(x, v|t) dx = dtvf(a, v|t)$$

Substitution of (4.32) into (4.31) gives us

$$\Pr\{x - v(t) dt < x(t) a\} = dt \int_0^a Cf(a, v|t) v dv \quad (4.33)$$

This formula shows that the probability of the intersection of the specified level by the stochastic process during the infinitesimally small time interval  $dt$  is proportional to the length of the interval. This allows one to introduce the time density for this probability  $p(a|t)$ . Using (4.33) gives

$$\Pr\{a - u(t) dt < x(t) a\} = p(a|t) dt \quad (4.34)$$

and, consequently,

$$p(a|t) = \int_0^a f(a, u|t) v dv \quad (4.35)$$

Analogously, one can find the derivative of the probability  $p(a|t)$ :

$$\frac{d}{dt} p(a|t) = - \int_0^a f(a, v|t) v dv \quad (4.36)$$

Adding and subtracting (4.35) and (4.36), one can easily obtain the two following equations:

$$p(a|t) + \frac{d}{dt} p(a|t) = \int_0^a f(a, u|t) v du \quad (4.37)$$

and

$$p(a|t) - \frac{d}{dt} p(a|t) = \int_0^a f(a, v|t) v dv \quad (4.38)$$

It is clear that

$$f(a, v|t) = f(v|a, t) f(a|t)$$

Then one can rewrite (4.37) and (4.38)

$$p(a|t) + \frac{d}{dt} p(a|t) = \int_0^a f(v|a, t) f(a|t) v dv \quad (4.39)$$

$$p(a|t) - \frac{d}{dt} p(a|t) = \int_0^a f(v|a, t) f(a|t) v dv \quad (4.40)$$

Using (4.35) for any time interval  $T$ , one can obtain the mean time of  $x(t)$  being over the specified level  $a$ . To obtain the result, we use the following simple arguments. Let us divide the total period  $T$  into  $n$  small nonoverlapping intervals located around the points  $t_j, 1 < j < n$ . For some  $t_j$ , we can write

$$\Pr\{A - (f_y) > a\} = \int_a^{\infty} f(x|t_j) dx \tag{4.41}$$

Assume that the intervals  $S/s$  are chosen so small that changing signs by the function  $x(t) - a$  can be neglected. Next, introduce the indicator function  $\{dt_j\}_+$  such that

$$t_j = \begin{cases} 1, & \text{if } x(t) - a > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.42}$$

Using this notation, one can write that the total time for which the function  $y(t)$  exceeds the level  $a$  equals

$$T_a \sim \sum_{j=1}^n A_j \tag{4.43}$$

and the mean time when the function  $y(t)$  exceeds the level  $a$  equals

$$E\{T_a\} = \sum_{j=1}^n E\{A_j\} \tag{4.44}$$

At the same time,

$$E\{A_j\} = \int_a^{\infty} f(x|t_j) dx \tag{4.45}$$

Using (4.45) and taking the limit in (4.44), we obtain the expression for the total time for which the process  $y(t)$  exceeds the level  $a$ :

$$E\{T_a\} = \int_a^{\infty} f(x) dx \tag{4.46}$$

If one is interested in the average number of intersections  $n_a$  during a time interval  $T$ , the same simple arguments can be used. Now introduce another indicator function

$$N_j = \begin{cases} 1 & \text{if } y(t) - a > 0 \text{ in the interval at least once} \\ 0 & \text{otherwise} \end{cases}$$

**DYNAMIC MODELS OF "STRENGTH - LOAD" TYPE 191**

The total number of intersections during period  $T$  equals

$$K = \sum_{j=1}^n N_j$$

Again the mean value can be expressed as

$$E(N_n) = \int_0^T E\{W_y\} dt \quad (4.47)$$

where

$$E\{N_j\} \sim p(a|t_j) \delta_j \quad (4.4)$$

8)

Taking the limit of (4.48) with the substitution of (4.35), we obtain

$$E\{AU - \int_0^T r_v f(a, v|t) dv dt\} \quad (4.49)$$

In addition to (4.46), the last expression permits us to write the expression for the mean time  $t_a$  for which the process  $x(t)$  exceeds the level  $a$  during a single intersection. Indeed,

$$* \text{BJAU} \quad (4.50)$$

or, using the corresponding complete expressions,

$$t_a = \frac{\int_0^T \int_a^\infty f(x|t) dx dt}{\int_0^T \int_0^\infty u f(a, u|t) dv dt} \quad (4.51)$$

#### 192 LOAD - STRENGTH RELIABILITY MODELS

All of these results are essentially useful for stationary processes because in this case all of the functions do not depend on the current time, that is,  $f(x|t) = f(x)$  and  $f(x, v|t) = f(x, v)$ . Then all of the previous results can be

rewritten in the simpler form, namely,

$$E\{r_{fl}\} = T \int_a^{\infty} f(x) dx \quad (4.52)$$

$$ETO = T \int_0^{\infty} v f(a, v) dv \quad (4.53)$$

$$\int_a^{\infty} f(x) dx \quad (4.54)$$

$$\int_0^{\infty} v f(a, v) dv$$

Naturally, for the stationary process the values of  $E\{TJ}$  and  $E\{V_{,,}\}$  depend only on the length of the period  $T$ . More precisely, they are proportional to  $T$ . The mean time  $E\{f_a\}$  for which the process exceeds the level  $a$  does not depend on  $T$ . For a stationary process one can also introduce the mean number of intersections per unit of time  $\lambda_a$ :

$$\lambda_a = \int_0^{\infty} v f(a, v) dv \quad (4.55)$$

that is, the probability of a level crossing in a unit of time.

#### 4.3.2 Gaussian Stochastic Process

To calculate all of the above-mentioned parameters of the specified level intersection, one needs to know the characteristics of the stochastic processes  $f(x|t)$  and  $f(x, v|t)$ . For stationary processes, one needs to know  $f(x)$  and  $f(x, v)$ . Fortunately, for the most important practical case—the *Gaussian stochastic process* (GSP)—sufficiently simple formulas can be obtained.

Note that the Gaussian process is often taken as the mathematical model of the random change of electrical parameters over time. There are many physical reasons to use this model because the influence of the number of internal and external factors leads to the formation of conditions for the validity of such a model. Indeed, these various factors might often be considered as relatively independent, and the influence of each of them on the resulting process is relatively small. Of course, the correctness of these hypotheses should be checked or verified each time.

We consider only a stationary process for which we know the mean  $E\{X\}$  and the variance  $\sigma_x^2$ . For a normal process in a stationary regime, the ordinate distribution is

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left[ -\frac{(x - \bar{x})^2}{2\sigma_x^2} \right]$$

where  $\bar{x} = K_x i(0)$  and  $K_x i(r)$  is itself the correlation function.

## 194 LOAD - STRENGTH RELIABILITY MODELS



DYNAMIC MODELS OF "STRENGTH - LOAD" TYPE 195

It is known from the theory of stochastic processes that the ordinate of the GSP and its derivative for the same moment of time are noncorrelated. Thus, the joint density function can be presented as the product of the two separated densities

$$f(x, v) = f(x) f(v) \quad (4.56)$$

or

$$f(x, v) = \exp \left[ -\frac{(x - E\{X\})^2}{2\sigma^2} \right] \exp \left[ -\frac{v^2}{2\sigma^2} \right]$$

Note that the variance  $\sigma^2$  can be expressed through the correlation function of the process as

$$(4.58) \quad \tau=0$$

and  $v(t)$  equals 0 because the stationary process is considered.

An expression for  $A_a$  can be obtained from (4.55) after the substitution of (4.57) **196 LOAD STRENGTH RELIABILITY MODELS**

$$P(a) = \frac{cr}{\mu T} \exp\left(\frac{a - E\{A\}}{\mu T}\right) \quad (4.59)$$

The expression for  $E\{r_a\}$  can be obtained in an analogous way:

$$r = \int_0^a \exp\left(-\frac{x}{\mu T}\right) \phi(x) dx$$

where  $\phi(x)$  is the normal distribution function.

### 4.3.3 Poisson Approximation

The crossing of a "high level" threshold by a stochastic process is of great interest for reliability analysis. It is clear that the probability of the crossing in this case should be sufficiently small; that is, such intersections are "rare events." As we mentioned above, the sequence of rare events forms a Poisson stochastic process. We omit the proof of the fact that in this particular case this hypothesis is also valid. Here we accept this as a known fact.

In general, we may assume that the mean number of intersections  $E\{N_a\}$  of level  $a$  for a specified period  $T$  approximately equals the mean number of events  $AT$  for some Poisson process with parameter  $A$ . Thus, this parameter

$$1 - \langle \dots \rangle \quad (4.60)$$

$$A \text{-----} (4.61)$$

where  $E[NJ]$  is determined by (4.51) or (4.55), depending on the type of stochastic process under consideration.

We will not write the expressions for the Poisson probabilities. The reader can do it easily him/herself. We only write the expressions for the probability of a failure-free operation (i.e., no intersection during the time  $T$ ) for the nonstationary and stationary cases using the corresponding values of  $E\{N_a\}$ . For the nonstationary process one has

$$P_0 = \exp \left\{ - \int_0^T \int_{Jc} v f(a, v|t) dt \right\} \quad (4.62)$$

and for the stationary process one has

$$P_0 = \exp \left\{ - \int_0^T \int v f(a, v) dt \right\} \quad (4.63)$$

For Gaussian processes, the probability  $P_0(T)$  can easily be written with the use of  $A_0$  from (4.59).

It is difficult to estimate the error obtained via the use of such an approximation for the Gaussian process. The only simple physical explanation lies in the fact that there is practically no correlation between neighboring moments of intersections of the specified "high level." (To check this fact, one should take into consideration the mean time between two neighboring intersections: "too much water has passed under the bridge" after the previous intersection!)

Example 4.5 Consider equipment characterized by a two-dimensional parameter with components  $X$  and  $Y$ . Both components  $X$  and  $Y$  are fluctuating in time. Their fluctuations are described as the identical independent stationary Gaussian processes with means equal to 0 and correlation functions

$$K_x(t) = K_y(r) = \langle r^2 e^{-\alpha|r|} [\cos \beta|r| + \hat{\wedge} \sin \beta|r|] \rangle \quad (4.64)$$

The tolerance limit area of the equipment parameter is represented by a sphere with radius  $a$ . Find the mean time that the system's parameter is spending inside the tolerance limit area if, at the moment  $t = 0$ , both  $X(t)$  and  $Y(t)$  are in the center of the tolerance sphere.

can be easily expressed as

*Solution.* Let

$$R(t) = y/x \setminus t + Y \setminus t \tag{4.65}$$

and

$$= dR(t) \tag{4-66}$$

Using (4.56), we can write

$$T = \frac{\int f(r) dr}{\int v_r f(a, v_r) dv_r} \tag{4.67}$$

where  $f(a, v_r)$  is the joint density of the distribution of the two r.v.'s  $R$  and  $v_r$  for  $R = a$ .

Consider an arbitrary period of time  $T$ . There are, on average,  $k_a T$  intersections, and the vector parameter represented by the point  $(X, Y)$  is outside the specified tolerance zone during the mean time  $E\{r_a\} A_a$ . Consequently, the system parameter will be inside the tolerance zone, on average, during the time  $T[1 - E\{r_a\} A_a]$ . The mean time that the parameter spends inside the tolerance zone is

$$r = \frac{1}{T} \int_0^T r dt = E\{U\} \tag{4.68}$$

Using (4.56), we can write

$$r =$$

$$\frac{\int_0^J \dots}{\int_{-\infty}^{\infty} v_r f(a, v_r) dv_r}$$

DYNAMIC MODELS OF "STRENGTH - LOAD" TYPE 199

Now we find the corresponding densities and compute the final result in a compact and constructive form. At first, notice that the r.v.'s  $X$  and  $Y$  are independent and have normal distributions, so

$$f(x, y) = \frac{1}{2\pi\sigma^2} \exp \left[ -\frac{1}{2\sigma^2} (x^2 + y^2) \right] \quad (4.69)$$

Further,

$$\Pr\{r < R < r + dr\} = \int_{x^2+y^2 < r+dr} f(x,y) dx dy$$

n

that is,  $f(r)$  is a Rayleigh density.

To determine the density  $f(r, v_r)$ , we need to consider a system of four normally distributed r.v.'s:  $X, Y, v_x = dX(t)/dt$ , and  $v_y = dY(t)/dt$ .

For a Gaussian process, all of these r.v.'s are independent. The variances of  $v_x$  and

$v_y$  are identical and equal

$$Dv_x = Dv_y = a^2(a^2 + p^2) \quad (4.71)$$

Thus,  $v_x$  and  $v_y$  do not depend on the coordinates of  $(X, Y)$  and have a circle normal distribution; that is, the projection of the vector  $(v_x, v_y)$  on the direction of  $R$  has a normal distribution with variance (4.71). Thus, the

two-dimensional density  $f(r, v_r)$  can be expressed in the following way:

$$f(r, v_r) = f(r) f(v_r) = \dots$$

$$\exp\{-\frac{1}{2} a^2 (a^2 + p^2) \dots\}$$

After substitution of (4.70) and (4.72) into (4.68), we obtain the final result

$$a y a_2 + p^2 f t ^ \wedge$$

This example shows that the use of stochastic process theory to find reliability indexes is not a simple task. But difficult practical problems always need the use of more or less complicated mathematical tools. Note also that besides the technical complexity of the solution there are also some special needs concerning the input data. Such data are not always available.

## **CONCLUSION**

We presented only a very brief description of the problem which could be explained by our intention to consider, primarily, system reliability. The problem related to the degradation of mechanical constructions under the random load and fluctuation of their physical parameters and the strength

## 202 LOAD - STRENGTH RELIABILITY MODELS

are special branches of modern reliability theory. Each of these branches must occupy a separate book.

The reader can find some appropriate formulations of the problem of reliability of mechanical systems and many useful related results, for example, in Kapur and Lamberson (1977). We would also like to mention Bolotin (1975). In Becker and Jensen (1977) one finds an analysis of a similar mathematical problem related to the reliability of electrical equipment under a stochastic fluctuation of parameters. Mechanical problems in reliability engineering are considered in Konyonkov and Ushakov (1975). Some results concerning the reliability of mechanical systems are contained in Ushakov (1985, 1994).

Interesting results concerning accumulations of random shocks can be found in Barlow and Proschan (1975). Elegant mathematical results can be obtained with the use of the Kolmogorov equations if the process of the parameter fluctuation can be described as a Markov process.

One can find a lot of interesting results in the extensive literature on noise analysis in radio equipment. This powerful branch of applications was stimulated by the pioneering work of Rice (1944, 1945).

At last, we would like to mention that this problem must be considered on a serious physical level. This chosen mathematical model must correspond to a real object, either electronic equipment or a mechanical construct. Writing a set of abstract models covering this subject area seems to be a hopeless task. Besides, it is not a simple task to find the appropriate statistical data for the models dealing with the random behavior of real parameters.

## REFERENCES

- Barlow, R. E., and F. Proschan (1975). *Statistical Theory of Reliability and Life Testing*. New York: Holt, Rinehart, and Winston.
- Becker, P. W., and F. Jensen (1977). *Design of Systems and Circuits for Maximum Reliability and Maximum Production Yield*. New York: McGraw-Hill.
- Bolotin, V. V. (1975). *Application of Methods of Probability Theory and Reliability Theory for Construction Design (in Russian)*. Moscow: Stroiizdat.
- Gertsbakh, I. B., and Kh. B. Kordonsky (1969). *Models of Failure*. Berlin: Springer.
- Kapur, K. C., and L. R. Lamberson (1977). *Reliability in Engineering Design*, New York: Wiley.
- Konyonkov, Yu. K., and I. A. Ushakov (1975). *Aspects of Electronic Equipment Reliability Under Mechanical Stress (in Russian)*. Moscow: Sovetskoye Radio.
- Rice, S. O. (1944, 1945). **Mathematical analysis of random noise**. *Bell Syst. Tech. J.* vol, 23, no. 3, and vol, 24, no. 1.
- Sveshnikov, A. A. (1968). *Applied Methods of the Stochastic Function Theory*. Moscow: Nauka.



- Ushakov, I. A., ed. (1985). *Reliability of Technical Systems: Handbook (in Russian)*. Moscow: Radio i Sviaz.
- Ushakov, I. A. (1994). *Handbook of Reliability Engineering*. New York: Wiley.
- Vinogradov, O. G. (1991). *Introduction to Mechanical Reliability: A Designer's Approach*. New York: Hemisphere.

## EXERCISES

- 4.1 The distributions of both a strength and a load are normal. The mean of the load is known:  $L = 10$  conditional units and the standard deviation  $a_g = 2$ . Find the parameters of the distribution of the strength  $S$  and  $\sigma$  which deliver a probability of failure-free operation equal to  $R = 0.995$ .
- 4.2 The distributions of both the strength and the load are exponential with parameters  $1/5$  and  $1/L$ , respectively.  $L = 1$  conditional unit. Find  $S$  which delivers a probability of failure-free operation equal to  $R = 0.995$ ,
- 4.3 The strength's distribution is normal with unknown parameters  $S$  and  $\sigma$  and a known coefficient of variation  $k = 0.04$ . The distribution of the load is exponential with  $L = 1$  conditional unit. Find the parameter  $S$  which delivers  $R = 0.999$ .

## SOLUTIONS

- 4.1 First of all, notice that the problem as formulated here is incorrect: one should know in advance the mean of the strength  $a$  or its standard deviation or the coefficient of variation  $k = \sigma/S$ . Without this correction the problem has no unique answer.

Let us assume that one knows  $k = 0.04$ . The problem can be solved by sequential iterations. For choosing a first value of  $S$ , notice that because of the requirement,  $R = 0.995$ , there must be at least more than  $L + 2.5\sigma$ . Choose  $S^{(1)} = L + 3a_g = 16$ . Then

$$\sigma = \sqrt{Jk(S^{(1)})^2 - \sqrt{10704}} = 256 = 3.2$$

Now it is clear that this level of strength is unacceptable. Choose the next value, for instance,

$$S^{(2)} = L + 3\sigma + 3\sigma^{(1)} = 26$$

204 LOAD - STRENGTH RELIABILITY MODELS

This value of  $S^{(2)}$  leads to

$$o^n = k (S^{(2)})^2 = (0.04) (676) \ll 5.2$$

Check the above obtained result

$$(MT53)_{-1, (2,22)_{0.987}} P = 0 \quad \begin{matrix} 26 - 10 \\ \mathbf{I 26 - 10} \end{matrix} \setminus$$

Thus, the value 5 is still smaller than one needs to deliver  $R = 0.995$ . The procedure continues. (We leave it to the reader to obtain the final numerical result.)

4.2 From (4.9) one can write

$$S = \frac{LR}{1 - R} = \frac{1 \cdot (0.995)}{0.005} = 199 \text{ conditional units}$$

This coefficient of safety is too large. The assumption that both strength and load distributions are exponential is unrealistic in practice. At the least, this is quite unreasonable as a distribution of strength.

4.3 For a highly reliable construction, one can use (4.12) or (4.13). This gives

$$R = 1 - \exp \left( - \frac{1}{2} \left( \frac{S}{U} \right)^2 \right) = 0.999$$

or

$$\exp \left( - \frac{0.025}{S} \right) = 0.001$$

The latter can be rewritten as

$$0.025^2 - S = 6.9$$

We leave to it to the reader to complete the solution.

## CHAPTER 5

# DISTRIBUTIONS WITH MONOTONE INTENSITY FUNCTIONS

For a quantitative characterization of reliability, we must know the failure distributions. Such detailed and complete information is not always available in engineering practice. Fortunately, in some cases we do not need to know the particular type of distribution, it is enough to know only some parameters of the distribution and the fact that this distribution belongs to some special class of distributions. In this case we can often obtain bounds on the reliability indexes based, for example, on the known mean and variance or other similar parameters of the distribution. Concerning the distributions, we need only know that they belong to the class of distributions with a monotone failure rate. There are several main classes of such distributions and these are described below.

### 5.1 DESCRIPTION OF THE MONOTONICITY PROPERTY OF THE FAILURE RATE

A very natural phenomenon of reliability as it changes over time is often encountered: the longer an item is functioning, the worse the residual reliability properties become. For many practitioners this phenomenon seems almost to be unique. Indeed, deterioration, fatigue, and other similar physical processes lead to a worsening reliability. Such phenomena (and their associated distributions) are called *aging*.

206 DISTRIBUTIONS WITH MIXTURES OF EXPONENTIALS  
 Distributions with mixtures of exponentials  
 If an item works for a long period of time, we become more sure of its reliability. Sometimes this property follows from a physical phenomenon connected with a change in the chemical and mechanical features of an item: a penetration of one material into another through contacting surfaces, a strengthening of the joining materials, a "self-fitting" of frictional parts, and so forth. Sometimes this is connected with a "burning-out" effect. This phenomenon is called *younging*. As an example of the latter property, consider a mixture of two equal parts of items: one with a constant M'ITF equal to 100 hours and another with a constant MTTF equal to 900 hours. We observe an item chosen at random from this mixed group of items. At the moment  $t = 0$  the MTTF equals

$$T = T_1 p + T_2 (1 - p) = 100(0.5) + 900(0.5) = 500 \text{ hours}$$

and the probability of a failure-free operation, say during 200 hours, equals

$$\Pr\{X \geq 200 \mid \text{starting at } t = 0\} = 0.5$$

But if it is known that at  $t = 101$  hours the item is still functioning, the values of both reliability indexes under the condition that the new trial starts at  $t = 101$  hours change:

$$T^* = T_1^* p + T_2^* (1 - p) = 799 \text{ hours and } \Pr\{X \geq 200 \mid \text{starting at } t = 101\} = 1$$

Both values for the used item are larger than for the new item on the average. Of course, there is no change in the item itself. We have only new information which allows us to make a posteriori a new conclusion about the item's reliability. An analogous example was considered in Chapter 1 when the mixture of exponential distributions was analyzed.

Notice that we observe a similar effect in "burning-out." It is normal practice to use some stress tests (temperature shocks, accelerated vibration, etc.) for selecting technologically weak items. The same effect is observed when weak units and manufacturing defects are eliminated during a high-failure "infant mortality" period under normal conditions.

It is the appropriate time to recall the ancient Greek myth about the Spartans who killed their weak and ill infants by throwing them from a high rock into a canyon. They did this to ensure that their remaining children would be healthy and strong. (We must state that this is only a myth: it was not a custom in the ancient democracy. As new sources claim, rich, free citizens of Greece replaced their weak and ill infants with the healthy babies of poor families.)

DESCRIPTION OF THE MONOTONICITY PROPERTY OF THE FAILURE RATE

Of course, there are no "immortal" items. First of all, if a failure rate decreases in the initial phase, an increase in the failure rate at some point is inevitable. Many items have the failure rate function of a "U-shaped form (see Figure 2,2). Second, even a probability distribution with a decreasing

208 DISTRIBUTIONS WITH MONOTONE INTENSITY FUNCTIONS (Of course, this puts a special condition on the decreasing failure rate function.) The exponential distribution is the boundary distribution between distributions with increasing and decreasing rates.

One of the basic characteristics in further analysis is the "conditional instantaneous density" of the time-to-failure distribution. For this conditional density, we usually use the terms *failure rate* or *failure intensity*. The strict mathematical definition of this, as we mentioned above, is

$$A(t) = \frac{f(t)}{F(t)} \quad (5.1)$$

Thus, in reliability terms, this is the instantaneous failure distribution density at time  $t$  under the condition that the item has not failed until  $t$ . A better explanation can be presented in terms of an "element of probability."  $A(t)$  is the probability of an unrepairable unit failure in the interval of time  $[t, t + \Delta t]$  under the condition that the unit has not failed by moment  $t$ . This conditional density changes continuously with time.

Sometimes it is useful to consider the function:

$$A(t) = \frac{f(t)}{F(t)} \quad (5.2)$$

Integration of (5.1) and (5.2) yields

$$P(t) = \exp \left( - \int_0^t f(x) dx \right) = e^{-M(t)} \quad (5.3)$$

In this chapter we consider only the simplest properties of distributions with a monotone failure rate. A more detailed analysis of the subject can be found in Barlow and Proschan (1975).

We do not consider the U-shaped  $A(t)$ 's or the nonmonotonic ones. Notice that a nonmonotonic  $A(t)$  is not very unusual at all. The following example from Barlow and Proschan (1975) can be analyzed in very simple terms.

**Example 5.1** Consider an unrepaired system consisting of two different units in parallel. Each unit has an exponentially distributed TTF. For this system

$$P(t) = 1 - (1 - e^{-\lambda_1 t})(1 - e^{-\lambda_2 t})$$

and

$$A(t) = \lambda_1 e^{-\lambda_1 t} + \lambda_2 e^{-\lambda_2 t} - (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)t} \quad (5.4)$$

To find the maximum of (5.4) directly by differentiation is a boring problem. We may analyze it in more simple terms. From a physical viewpoint, if the parallel system is functioning for a very long period of time, the most probable situation is that there is only one unit which has survived. If so, this unit, on average, is the most reliable one. Moreover, the longer the period of observation, the higher the conditional probability that the survivor is the best unit. Suppose that, in our case,  $A_1 < A_2$ . Thus, for the system A it  $\rightarrow A_1$ . At the same time, for any parallel system  $A(0) = 0$ . Show that at some f the function  $A(t)$  is larger than  $A_1$ :

$$HO = \frac{A_1 e^{-x_1} + A_2 e^{-x_2}}{e^{-x_1} + e^{-x_2}} - (A_1 + A_2) e^{-x_1 - x_2} > A_1 \quad (5.5)$$

The inequality (5.5) easily transforms into

$$+ A_2 e^{-A_1 t} - (A_1 + A_2) e^{-(A_1 + A_2)t} > A_1 e^{-A_1 t} + A_2 e^{-(A_1 + A_2)t}$$

and after the simple transformations

$$> e^{-A_1 t}$$

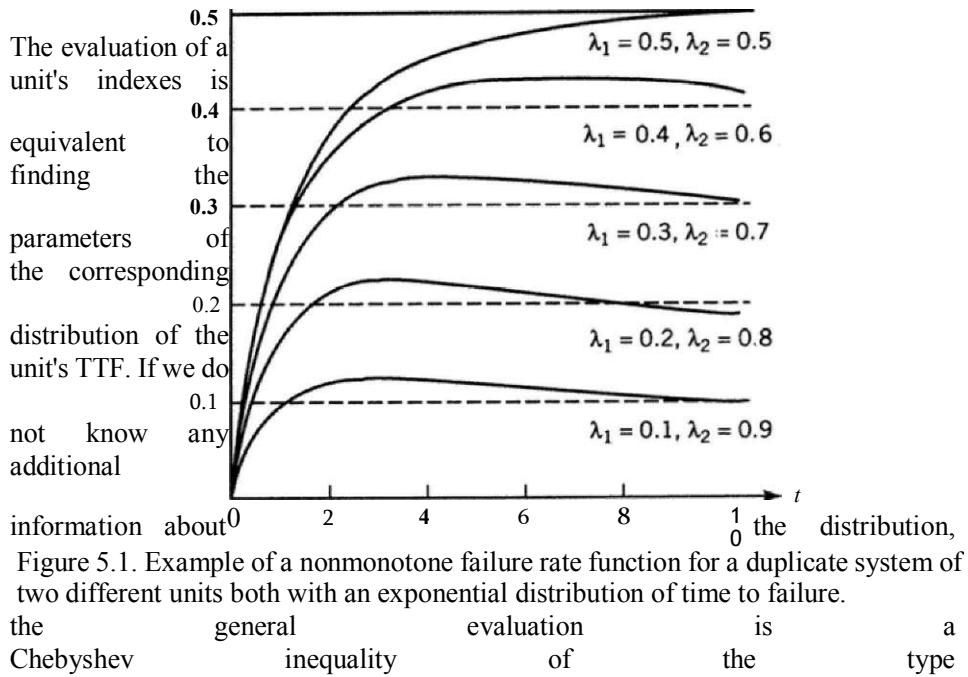
The last inequality is valid starting from

$$t_n = \frac{1}{A_1 - A_2} \ln \frac{A_1 + A_2}{A_1} \quad (5.6)$$

Thus, the function  $A(t)$  for the system starts from 0, then intersects the level  $A_1$  from below, and after this reaches its maximum and exceeds the limit value of  $A_1$  from above. From (5.5) one can see that  $A(t)$  is monotonically increasing if  $A_1 = A_2 = A$ . Figure 5.1 presents the  $A(t)$  behavior over time for three characterizing proportions between  $A_1$  and  $A_2$ .

Below we consider the distributions with an increasing failure rate (IFR d.f.'s) though this is only one (and the most narrow class) of the "aging" distributions. The reader can find other subclasses of the "aging" distributions, as well as the "younging" distributions, in the original interpretation, in the excellent book by Barlow and Proschan (1975).

5.2 UNIT WITH IFR DISTRIBUTION OF TTF



$$\Pr\{|\xi - E(\xi)| \geq \varepsilon\} \leq \frac{\varepsilon^2}{e^2} \quad (5.7)$$

where  $e$  is an arbitrary positive value.

This inequality is very well known in probability theory. To give the reader a sense of the result, we follow the proof given in Gnedenko (1988). By definition,

$$\Pr\{|\xi - E\{\xi\}| \geq \varepsilon\} = \int_{|x - E\{\xi\}| \geq \varepsilon} dF(x)$$



Because in the domain of integration  $(1/e) \leq x - E\{t\} < 1$ ,

$$\int_0^1 dF(x) < \int_0^1 (x - E\{t\}) f dF(x)$$

$$\frac{1}{2} \leq \frac{1}{c} \frac{\text{Var}\{t\}}{c^2}$$

This completes the proof.

Inequality (5.7) is universal and so is not too constructive for practical purposes (as with any universal tool). For instance, one sees that (5.7) only makes sense when  $e > tr$ . In other words, this estimate is not true in some area around the mean. But notice that, at the same time, at the distribution's tails, the estimate is very rough. Suppose that additional information is available. Then we can obtain narrower bounds. Consider the class of IFR d.f.'s. We first prove several additional statements.

**Theorem 5.1** The graph of an IFR d.f.  $P(t)$  crosses the graph of an arbitrary exponential function  $e^{-\lambda t}$  at most once from above. If these two functions do not cross,  $P(t)$  lies strictly under this exponential d.f. (see Figure 5.2).

*Proof.* The intensity function  $A(t)$  for the IFR distribution might increase infinitely or be bounded above by some number  $A^*$ . In the first case, there exists a moment  $t_0$  when

$$A(t_0) = \int_0^{t_0} A(t) dt < 1 - e^{-\lambda t_0}$$

and, for any  $y > t_0$ ,

$$\frac{A(y)}{A(y)}$$

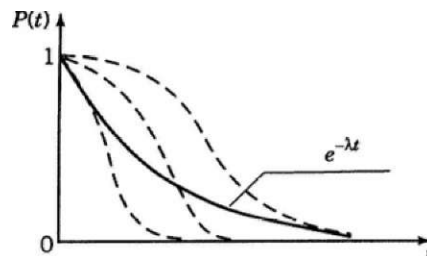


Figure 5.2. Explanation of the contents of Theorem 5.1: possible types of relationships between the exponential function and different IFR distributions of time to failure.

because  $A(t)$  increases. Thus,  
 (5.8)

In the second case, for  $P(t) \begin{cases} \geq e^{-\lambda t} & \text{for } t \leq t_0 \\ \leq e^{-\lambda t} & \text{for } t > t_0 \end{cases}$  any  $t > 0$ , we have  $A(t) > \lambda t$  and, consequently,  $A(t) > \lambda t$  for any  $t$ , never crosses  $\lambda t$ . Thus,

$$P(t) < e^{-\lambda t}$$

From (5.8) and (5.9) it follows that the "right tail" of the IFR distribution decreases faster than the corresponding tail of the exponential function. (5.9)

$$\left. \frac{d}{dt} P(t) \right|_{t=0}$$

**Corollary 5.1** If an IFR d.f.  $P(t)$  has a first derivative different from 0 at

$$t = 0, \text{ say } \lambda,$$

then  $P(t)$  lies everywhere below  $e^{-\lambda t}$ .

*Proof.* The proof follows from the fact that  $A(t) > \lambda t$  for all  $t$ . Some hint of a graphical explanation can be found in Figure 5.2.

**Corollary 5.2** An IFR d.f.  $P(t)$  necessarily crosses  $e^{-\lambda t}$  from above once if both distributions have the same MTTF equal to  $T$ .

*Proof.* By Theorem 5.1 both d.f.'s have to intersect once or not intersect at all. The second statement contradicts the corollary, so we need to check this. Suppose that both d.f.'s do not intersect. This means that

$$\int_0^\infty P(t) dt < \int_0^\infty e^{-\lambda t} dt = \frac{1}{\lambda} = T$$

which contradicts the statement concerning the equality of the MTTFs. For a graphical explanation, refer to Figure 5.2.

**Theorem 5.2** For an IFR d.f.  $P(t)$ , the function

$$7P(T)$$

decreases with increasing  $t$ .

*Proof.*  $A(t)$  is convex, so  $\log P(t) = -A(t)$  is concave. But then

$$\frac{\log P(t) - \log P(0)}{t - 0}$$

decreases with increasing  $t$ . Consequently,

$$1 - \frac{\log P(t)}{t}$$

decreases in  $t$ . After substituting  $P(0) = 1$  and using an exponential transformation, the proof of the theorem follows.

This theorem produces the following interesting corollaries.

**Corollary 5.3** For an IFR d.f.  $P(t)$ ,

$$P(x) < [P(t)]^{x/t}$$

for all  $x > t$ .

This allows us to predict (i.e., to compute a lower bound) the probability of a failure-free operation of an IFR unit for a specified time, if we know the value of  $P(t)$  for a smaller interval of time. This corollary can be of great use for an application in testing IFR units during a short testing period.

**Corollary 5.4** For an IFR d.f. the initial moments of all orders are finite.

*Proof.* Indeed, for any  $t$ ,

$$\int_0^\infty x^r P(x) dx < \int_0^\infty x^r [P(t)]^{x/t} dx = \int_0^\infty x^r e^{-x/t} dx < \infty$$

where  $[P(t)]^{x/t}$  is replaced by  $e^{-x/t}$ . The reader knows that the exponential d.f. has the moments of all orders.

The last corollary shows that arguments about the properties of "aging" units, which seem to be just qualitative statements, have led to very strong restrictions on the moments of an IFR d.f. Incidentally, note that the coefficient of variation of the IFR d.f. is always less than 1.

Now, using all of the above results, the following important characteristics of IFR d.f.'s can be obtained.

**Theorem 5.3** If  $i_p$  is the quantile of an IFR d.f.  $P(t)$ ,  $\Pr\{U \geq p\}$ , then

$$P(t) \leq e^{-at} \text{ for } t < i_p, \quad a = -\frac{\ln(1-p)}{i_p}$$

$$P(t) \geq e^{-ar} \text{ for } t > i_p$$

where

$$a = -\frac{\ln(1-p)}{i_p}$$

*Proof.* An exponential function can be found which goes through the point  $(i_p, 1 - p)$ . The parameter of the exponent can be found from the equation

$$e^{-a i_p} = 1 - p$$

We then use Theorem 5.1 to complete the proof.

**Theorem 5.4** A lower bound for an IFR d.f. is determined

by 
$$P(t) \geq \begin{cases} e^{-rt} & \text{for } t < T \\ e^{-rT} & \text{for } t > T \end{cases}$$

$$r = -\frac{\ln(1-p)}{i_p}$$

where  $t$  is the MTTF

$$r = \int_0^{\infty} f_p(t) dt$$

*Proof.* We first present a rigorous proof. For an IFR distribution the function  $A(r)$  is an increasing convex function, so by Jensen's inequality

$$E\{A(\xi)\} < A(E\{\xi\}) = A(T) \tag{5.10}$$

Denote  $P(t) = y$  and rewrite

$$E\{A(\xi)\} - E\{-\ln y(\xi)\} = E\{-\ln y\} - \int_0^1 \ln y dy = 1 \tag{5.11}$$

From (5.10) and (5.11)

$$E\{A(\xi)\} = 1 - A(T) = -\ln P(T)$$

immediately follows where

$$P(iT) > e^{-1} \quad (5.12)$$

or, equivalently,  $[P(t)]^{1/t} > [P(T)]^{1/T} > e^{-V^T}$

$$[P(T)]^{1/T} > e^{-1/T} \text{ and, finally, for } t < T \text{ the required result is obtained}$$

Now from Corollary 5.3 for  $t < T$  we can write

$$P(t) > e^{-t/T}$$

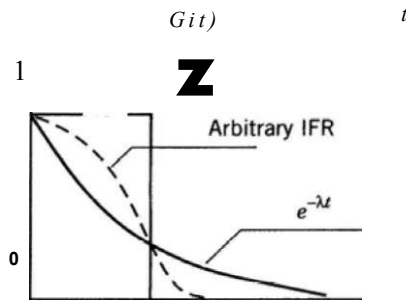
The same result can be derived from simple explanations based on a graphical presentation (see Figure 5.3).

The first inequality follows immediately from a comparison of the exponential function  $e^{-t/T}$  and a degenerate function  $G(t)$  with the same MTTF

$$G(t) = \begin{cases} 1 & \text{for } t < T \\ 0 & \text{for } t > T \end{cases}$$

The degenerate function (i.e., a distribution of a constant value) is the boundary distribution for the class of IFR d.f.'s. By Theorem 5.1 the degenerate function crosses the exponential function from above at point  $t = T$ . All strictly IFR d.f.'s may cross the graph of a given exponent only for  $t > T$  which follows from the equality of the MTTFs. The second inequality is trivial because  $P(t)$  is a nonnegative function. Notice that a lower bound is reached by the exponential d.f. for  $t < T$  and by the degenerate d.f. for  $t > T$ .

Figure 5.3. Explanation of the proof of Theorem 5.4: relationships among IFR, exponential, and degenerate reliability functions.



**Theorem 5.5** An upper bound for an IFR d.f. is determined as

$$P(t) \leq \begin{cases} P(t) & \text{for } t \leq T \\ e^{-\lambda(t-T)} P(T) & \text{for } t > T \end{cases} \quad (5.13)$$

where  $\lambda$  depends on  $f^*$  and is found from the condition

$$\int_0^{t^*} f^*(t) dt = \int_0^T f(t) dt$$

or, equivalently,

$$1 - e^{-\lambda T} = \int_0^{t^*} f^*(t) dt$$

*Proof.* The first inequality in (5.13) is trivial and follows from the definition of a d.f. The second inequality is equivalent to the statement that for  $t > T$  the IFR function  $P(t)$  crosses the graph of the function  $E^*(t)$  from above, which is the exponential function truncated from the right at point  $t^*$ ,

$$E^*(t) = \begin{cases} P(T) e^{-\lambda(t-T)} & \text{for } t < t^* \\ 0 & \text{for } t > t^* \end{cases}$$

at some point  $t = t^*$  if both  $P(t)$  and  $E^*(t)$  have the same MTTF. This fact can be proved immediately by assuming the contrary. Suppose that there is no such crossing. Then  $P(t)$  lies above  $E^*(t)$  everywhere, but then

$$\int_0^{\infty} P(t) dt > \int_0^{\infty} E^*(t) dt$$

which contradicts our suggestion about the equality of their MTTFs. A graphical explanation of (5.13) is given in Figure 5.4.

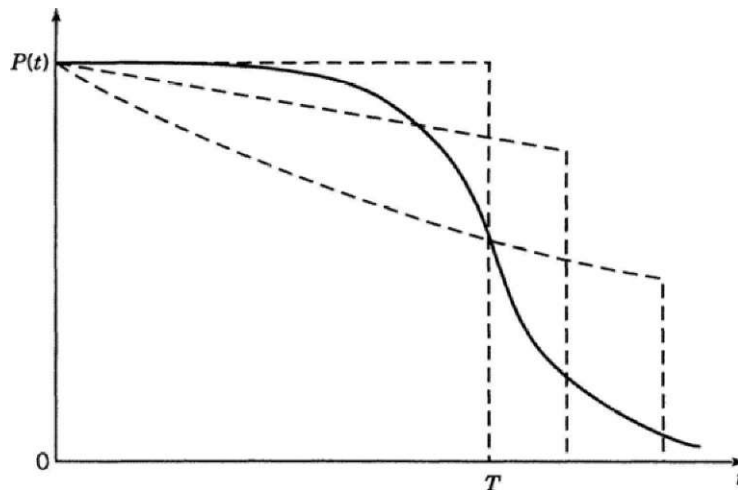


Figure 5.4. Explanation of the proof of Theorem 5.5: finding  $t^*$  by constructing the exponential truncated from the right.



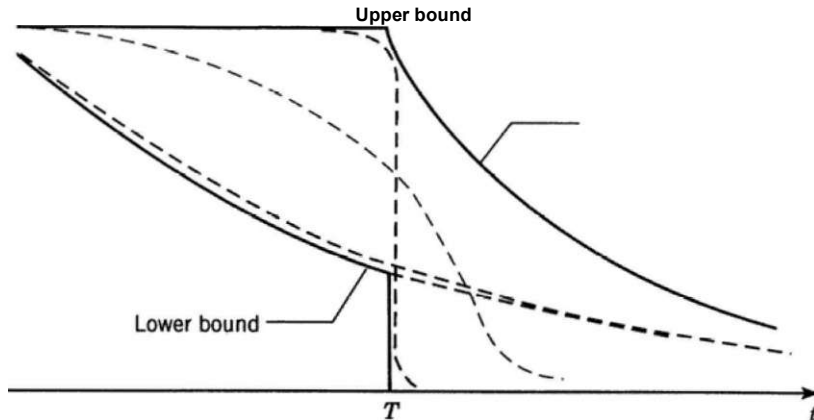


Figure 5.5. Area of possible values of IFR reliability functions with the same MTTF and samples of different IFR reliability functions.

As a result, we have lower and upper bounds for the IFR function  $P(t)$  which are represented in Figure 5.5. In this figure  $P_x(t)$  is the function with a coefficient of variation close to 0, and  $P_y(t)$  is the function with a coefficient of variation close to 1.

**Theorem 5.6** An upper bound for the quantile of the IFR distribution is expressed by its MTTF,  $T$ , and corresponding probability  $p$ ,  $p = 1 - \Pr\{f > f_p\}$ , as

$$\ln(1 - p)$$

*Proof.* Notice first that, from  $\Pr\{f > f_p\} = P(f_p) = 1 - p$ ,

$$HQ_{1-p} = -\ln(1 - p)$$

Now the chain of obvious inequalities based on the previous results can be written as

$$\begin{aligned} T &= \int_0^{\infty} p\{t\} dt * \int_0^{\infty} f''(t) p(t) dt > \int_0^{\infty} [p(t_p)] \\ &= \int_0^{\infty} \frac{f(t) [1 - P(t)]}{f'(t)} \cdot x dx \end{aligned}$$

**220 DISTRIBUTIONS WITH MONOTONE INTENSITY FUNCTIONS**

and simple integration gives us

$$\int_{-\infty}^x \frac{e^{-x}}{\ln(1-p)} dx = \frac{e^{-x}}{\ln(1-p)} (\exp[\ln(1-p)] - 1)$$

$$= \frac{e^{-x}}{\ln(1-p)} [1 - (1-p)] = \frac{e^{-x} p}{-\ln(1-p)}$$

which produces the desired result.

**Theorem 5.7** A lower bound for the quantile  $\xi$  of the IFR distribution is expressed by  $T$  and  $p$  as

$$T \leq \xi_p \leq \frac{(-\ln(1-p))T}{p} \quad \text{for } 1 - p \geq e^{-T}$$

*Proof.* We prove the first of these inequalities separately for  $\xi_p \geq T$  and  $\xi_p < T$ . For the first case

$$1 - p = P(\xi_p) > e^{-\xi_p} \wedge \exp(-\xi_p)$$

For the second case with the use of Theorem 5.2, we immediately write

$$1 - p > e^{-\xi_p} \wedge \exp(-\xi_p)$$

Thus, the desired inequality is valid in both cases.

The second inequality, which is valid for the condition

$$1 - p = P(\xi_p) < e^{-\xi_p} \wedge \exp(-\xi_p)$$

follows immediately if we recall (5.12). Thus,

which corresponds to the desired condition  $\xi_p > T$ .

**Corollary 5.5** For the median  $M$  of an IFR d.f., the following bounds are valid:

$$(-\ln)7 < M < (-2\ln) T$$

$$\frac{M}{2 \ln 2} < 7 < \frac{M}{\ln 2}$$

*Proof.* The proof follows automatically from Theorems 5.6 and 5.7 after the substitution  $p = 1/2$ .

If instead of the MTTF, we know the variance of the IFR distribution, the bounds can be improved. We do not consider these more complex cases and advice the reader to refer to Barlow and Proschan (1975). for an excellent discussion of the subject.

**5.3 SYSTEM OF IFR UNITS**

As we saw above, an IFR type of distribution of a unit TTF leads to interesting and constructive results. An extension of these results appears when we consider systems consisting of units with IFR types of TTF d.f.'s.

We will formulate all of the results in the form of theorems because each of them requires a mathematical proof. First of all, we prove a lemma, which is simple but very important for future considerations.

**Lemma 5.1** If (1) the function  $f(x)$  is monotonic, restricted, and nonnegative on the positive semiaxis, (2) the function  $g(x)$  is absolutely integrable on the positive semiaxis, (3) the latter function is such that  $g(x) > 0$  for  $x < a$  and  $< 0$  for  $x > a$ , and (4)

$$\int_0^a g(x) dx = 0$$

then, if  $f(x)$  decreases (increases), the following inequality is true:

$$\int_0^a f(x) g(x) dx > 0$$

*Proof.* The proof can be presented as a chain of simple transformations:

$$\begin{aligned} \int_0^{\infty} f(x)g(x)dx &= \int_0^a f(x)g(x)dx + \int_a^{\infty} f(x)g(x)dx \\ &< (\max_{0 \leq x \leq a} f(x)) \int_0^a g(x)dx + \int_a^{\infty} \min_{x > a} f(x)g(x)dx \\ &= f(a) \int_0^a g(x)dx + \int_a^{\infty} f(x)g(x)dx = \int_0^{\infty} f(x)g(x)dx \end{aligned}$$

The sense of the lemma is clear from Figure 5.6 where an increasing function  $f(x)$  is shown. Obviously, the square  $S_1$  is taken in the resulting expression with less "weight" than the square  $S_2$ , and thus the sum turns out to be negative.

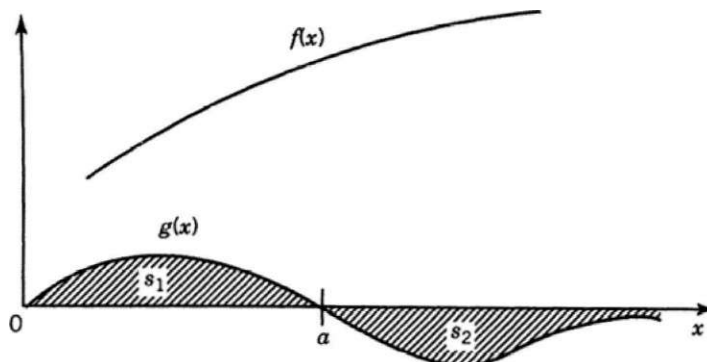


Figure 5.6. Graphical explanation of the proof of Lemma 5.1.

### 5.3.1 Series System

**Theorem 5.8** A lower bound on the probability of a failure-free operation of a series system of IFR units with known MTTF is

$$P(t) \geq \prod_{i=1}^n \left( 1 - \frac{t}{T_i} \right) \quad \text{for } t > t^*$$

where  $t^* = \min T_r$

*Proof.* The proof immediately follows from Corollary 5.2 by a simple substitution.

This lower bound is very important in practice because it gives a guaranteed estimate of the real but unknown value of the reliability index.

**Theorem 5.9** An upper bound on the probability of a failure-free operation of a series system of IFR units with known MTTF is

$$P(t) < \begin{cases} 1 & \text{for } T^{\wedge} t < T_2 \\ \exp[-\sum_{i=1}^n \lambda_i t] & \text{for } T_2 \leq t < T_3 \end{cases}$$

for  $t \notin T$

$$\exp[-\sum_{i=1}^n \lambda_i t] \quad \text{for } t \leq \min T_i = T, \quad 1 \leq i \leq n \tag{5.15}$$

where  $T_i, 1 \leq i \leq n$ , are ordered MTTFs and each  $\lambda_i$  is found from the equation

$$1 - \lambda_i T_i = \exp(-\lambda_i T_i)$$

*Proof.* The proof follows immediately from (5.13) of Theorem 5.5.

**Theorem 5.10** An upper bound on the probability of a failure-free operation of a series system of IFR units with known  $a = A(0)$  is

$$P(t) < \exp(-\sum_{i=1}^n \lambda_i t) \tag{5-16}$$

for  $t < \min T_i$

*Proof.* The proof follows immediately from Corollary 5.1.

**Theorem 5.11** The MTTF of a series system of independent IFR units has the following bounds:

$$1 \tag{5.17}$$

$$r_1 - \min T >$$

$$E r$$

*Proof.* An upper bound follows trivially from the obvious statement that for any  $t$  the system is less reliable than any of its units:

$$p_1(t) \leq \min p_i(t)$$

Therefore,

$$\int_0^{\infty} p_n(t) dt \leq \int_0^{\infty} P_n(t) dt = r_{\text{sys}}$$

We may use Lemma 5.1 to obtain a lower bound. We show that the replacement of an arbitrary unit with an IFR distribution of TTF with a unit with an exponentially distributed TTF, which has the same MTTF, leads to a decrease of the series system's MTTF. Suppose that such a replacement is done for the  $i$ th unit of the system. We need to prove that

$$\int_0^{\infty} P_i(t) dt \geq \int_0^{\infty} e^{-t/T_i} dt$$

or, equivalently,

$$A = \int_0^{\infty} [P_i(t) - e^{-t/T_i}] dt > 0$$

Note that, by Theorem 5.1,  $P_i(t)$  crosses  $e^{-t/T_i}$  once and from above and, by assumption, both these functions have the same MTTF. Thus,

$$P_i(t) - e^{-t/T_i}$$

corresponds to the function of Lemma 5.1. At the same time, the function

$$e^{-t/T_i}$$

corresponds to the decreasing function  $f(x)$  in Lemma 5.1. Thus, by Lemma 5.1,  $A > 0$ , and the desired intermediate statement is proved.

The systematic replacement of a system units with an IFR distribution with units with a corresponding exponential distribution produces

$$r_{\text{sys}} \leq \int_0^{\infty} e^{-t/T_i} dt = \frac{1}{T_i} \prod_{j=1}^n \frac{1}{\lambda_j}$$

Thus, the theorem is proved.

The upper bound can be improved if we possess additional information about the distribution  $P_i(t)$  for example, if we know the first derivatives in  $t = 0$ .

**Theorem 5.12** The upper bound of the series system MTTF can be written as

$$T_{\text{system}} < \frac{1 - \exp\left(-\min_{i=1, \dots, n} \lambda_i(0) t\right)}{\sum_{i=1}^n \lambda_i(0)} \quad (5.18)$$

where  $a_i$  is determined from the condition

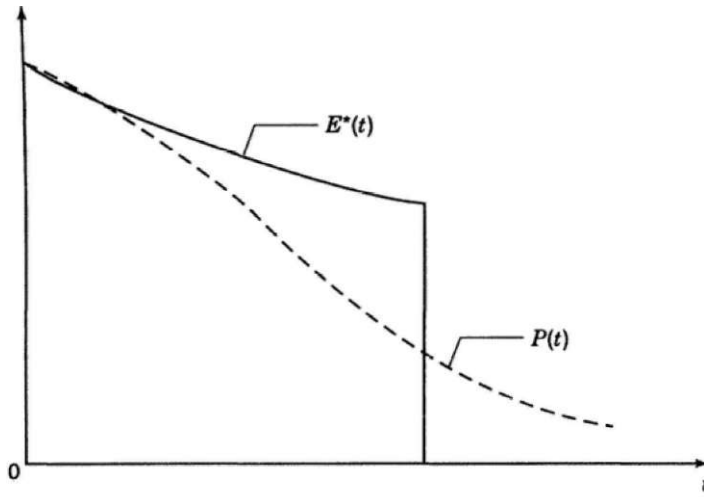
$$\int_0^{a_i} e^{-\lambda_i(0)t} dt = T_i$$

*Proof.* Consider an exponential distribution truncated from the right:

$$E^*(t) = \begin{cases} \frac{1 - e^{-\lambda_i(0)t}}{\lambda_i(0) a_i} & \text{for } t < a_i \\ 0 & \text{for } t > a_i \end{cases}$$

This distribution  $E^*(t)$  has the same MTTF as the initial distribution  $P_n(t)$ . Hence,  $E^*(t)$  crosses  $P_n(t)$  from above (see Figure 5.7).





**Figure 5.7.** Graphical explanation of the proof of Theorem 5.12: intersection of the IFR reliability function and the exponential function truncated from the right where their derivatives in  $t = 0$  are equal.

In the expression for the system MTTF, replace the unit with distribution  $P_{n,}(t)$  by  $E^*(t)$ . The new system MTTF is

$$\int_0^\infty n P_{n,}(t) dt = \int_0^\infty n P_{n,}(t) dt$$

We now find the value of A:

$$A = \int_0^\infty [E^*(t) - P_{n,}(t)] n dt - r \int_0^\infty P_{n,}(t) dt$$

$$= \int_0^\infty [E^*(t) - P_{n,}(t)] n dt$$

Again we can use Lemma 5.1, noticing that  $E^*(t) - P_{n,}(t)$  corresponds to the function  $g(x)$  from Lemma 5.1 and  $P_{n,}(t)$  corresponds to the decreasing function from Lemma 5.1. Thus, by Lemma 5.1,  $A > 0$ , that is, the replacement of any IFR unit, say the  $n$ th, with a unit with distribution  $E^*(t)$  might only increase the system's MTTF.

Thus, the systematic replacement of units in the above-described manner leads to the final result

$$A \geq \int_0^\infty [E^*(t) - P_{n,}(t)] n dt$$

$$= \int_0^\infty [E^*(t) - P_{n,}(t)] n dt$$

and this completes the proof.

### 5.3.2 Parallel Systems

**Theorem 5.13** An upper bound for the probability of failure of a parallel system of IFR units can be expressed as

$$Q(t) = \prod_{i=1}^n Q_i(t) \quad \text{for } t \leq t^*$$

$$Q(t) = \prod_{i=1}^n Q_i(t) \quad \text{for } t > t^*$$
(5.19)

where

$$t^* = \min T_j$$

*Proof.* The proof follows directly from Theorem 5.4,

For the probability of a failure-free operation, the following lower bound follows from (5.19):

$$P(t) > \begin{cases} 1 - \prod_{i=1}^m (1 - \exp(-\lambda_i t)) & \text{for } t < t^* \\ 0 & \text{for } t > t^* \end{cases} \quad (5.20)$$

**Theorem 5.14** A lower bound for the probability of failure of a parallel system of IFR units has the form

$$Q(t) > \begin{cases} 1 - \prod_{i=1}^m (1 - \exp(-\lambda_i t)) & \text{for } t < t^* \\ 0 & \text{for } t > t^* \end{cases} \quad (5.21)$$

where

$$t^* = \max_{1 \leq i \leq m} T_i$$

*Proof.* The proof follows immediately from Theorem 5.5.

For the probability of a failure-free operation, the following upper bound follows from (5.21):

$$P(t) \leq \begin{cases} \prod_{i=1}^m (1 - \exp(-\lambda_i t)) & \text{for } t < t^* \\ 1 & \text{for } t > t^* \end{cases} \quad (5.22)$$

**Theorem 5.15** The MTTF of a parallel system of independent IFR units has the bounds

$$\max_{1 \leq i \leq m} T_i < T_{\text{sys}} < \sum_{i=1}^m T_i$$

$$T_{\text{sys}} \leq \frac{1}{\sum_{i=1}^m \lambda_i} \quad (5.23)$$

*Proof.* This proof is analogous to the proof of Theorem 5.11 and so we omit it.

### 230 DISTRIBUTIONS WITH MONOTONE INTENSITY FUNCTIONS

Again note that the lower bound is trivial and can be instantly found for a degenerate distribution, that is, for the case when all  $T/s$  are constant.

We now mention that, for a parallel system, the MTTF is larger if the unit failure distributions have larger variances. (In qualitative terms, this result is close to that obtained for dependent units.) It seems paradoxical that an unstable production of some units for a parallel structure is better than a stable production: we used to think that stability is almost always better than instability. But there is no enigma at all if one notices that the random time to failure of a parallel system is the maximum of the unit's random time to failure.

**Theorem 5.16** A lower bound for the MTTF of a system consisting of units with an IFR distribution, for which we know the first derivative in  $t = 0$ , is

$$T_{jst} - E \int_0^\infty \frac{Z}{1 - \exp(-\min\{7, 7\} [A_i(0) + A_j(0)])} dt + (-1)^j \sum_{i=1}^n \frac{A_i(0)!}{i!} \exp \left( - \sum_{j=1}^n \lambda_j A_j(0) \right)$$

*Proof.* The proof here is analogous to that of Theorem 5.12. We only need to notice that the probability of a failure-free operation for this case after all substitutions of  $P_i(t)$  for  $E_i^*(t)$  has the form

$$p(t) = \prod_{i=1}^n [1 - e^{-\lambda_i t}]$$

where  $\lambda_i^*(t)$  is defined in Theorem 5.12.

### 5.3.3 Other Monotone Structures

Instead of writing detailed formulas with the simple substitution of IFR d.f.'s  $P_i(t)$  for degenerate or exponential d.f.'s, we mention only that one can obtain a lower bound for the system reliability by substituting lower bounds of the corresponding units' failure-free probabilities. Analogously, after the substitution of the upper bounds of the units' probabilities, a lower bound for the system probability is obtained. The reader can find some related results in Barlow and Proschan (1975).

## CONCLUSION

This relatively new branch of reliability theory was initiated by Barlow and Proschan and became widely known after their book [Barlow and Proschan (1975)] was published. First papers on the properties of distributions with a

monotone failure rate appeared in the previous decade [Barlow, Marshall, and Proschan (1963); Barlow and Marshall (1964); Solovyev and Ushakov (1967); Gnedenko, Belyaev, and Solovyev (1969); among others].

We would like to present here a simple but important result concerning repairable systems [Ushakov (1966)]. The stationary interval availability coefficient of a system with an "aging" distribution  $F(t)$  of TTF has lower and upper bounds of the form

$$K \left( 1 - \frac{1}{T} \int_0^{\infty} F(x) dx \right) \leq R(t_0) \leq K \left( 1 - \frac{1}{T} \int_0^{\infty} F(x) dx \right) + \frac{1}{T} \int_0^{\infty} F(x) dx$$

where  $K$  is the stationary availability coefficient and  $T$  is the mean of the distribution  $F(t)$ . These bounds can be easily obtained with the help of Lemma 5.1. Indeed,  $R(t_0)$  for any distribution  $F(t)$  can be written as

$$R(t_0) = \int_0^{\infty} P^*(t_0) f(x) dx$$

where  $P^*(t_0)$  is the distribution of a stationary residual time:

$$P^*(t) = \int_0^{\infty} f(x) dx$$

Substitution of degenerate and exponential d.f.'s into the latter expression and application of Lemma 5.1 produce the necessary result. This result and some others can be found in Gnedenko, Belyaev, and Solovyev (1969). Some new results can be found in Gnedenko (1983).

A collection of practical results on aging units and systems consisting of aging units is presented in Ushakov (1985, 1994). This problem is especially important in practice when one possesses only very restricted statistical information but has some reasonable physical arguments about the possible behavior of a time-to-failure distribution.

## REFERENCES

- Barlow, R. E., and A. W. Marshall (1964). Bounds for distributions with monotone hazard rate. I and II. *Ann. Math. Statist.*, vol. 35.
- Barlow, R. E., and A. W. Marshall (1965). Tables of bounds for distributions with monotone hazard rate. *J. Amer. Statist. Assoc.*, vol. 60.
- Barlow, R. E., and F. Proschan (1975). *Statistical Theory of Reliability and Life Testing*. New York: Holt, Rinehart, and Winston.
- Barlow, R. E., A. W. Marshall, and F. Proschan (1963) Properties of probability distributions with monotone hazard rate. *Ann. Math. Statist.*, vol. 34.
- Gnedenko, B. V., ed. (1983). *Mathematical Aspects of Reliability Theory (in Russian)*. Moscow: Radio i Sviaz.

Gen den ko, B. V., Yu. K. Belyaev, and A. D. Soiovyev (1969). *Mathematical Methods of Reliability Theory*. New York: Academic.

Soiovyev, A. D., and I. A. Ushakov (1967). Some bounds on a system with aging elements (in Russian). *Automat. Comput. Sci.*, no. 6.

Ushakov, I. A. (1966). An estimate of reliability of a system with renewal for a stationary process (in Russian), *Radiotekhnika*, no. 5.

Ushakov, I. A., ed. (1985). *Reliability of Technical Systems: Handbook* (in Russian). Moscow: Radio i Sviaz.

Ushakov, I. A., ed. (1994). *Handbook of Reliability Engineering*. New York: Wiley.

**EXERCISES**

- 5.1 Two units have the same mean  $T$ . One unit has a uniform d.f. and another has an exponential d.f. Which one will deliver the larger probability of failure-free operation at moment  $t = T$ ? At moment  $t = 2T$ ?
- 5.2 Consider the Erlang d.f. of a high order (e.g.,  $n = 10$ ). Explain (without exact proof) how  $A(t)$  behaves.
- 5.3 One has to choose a system for the continuous performance of an operation during 100 hours. There are two possibilities: to choose for this purpose a system with an MTTF of 200 hours or to choose another system with an MTTF of 300 hours. Which system should be chosen.
- 5.4 What kind of  $A(t)$  has the system depicted in Figure E5.3?

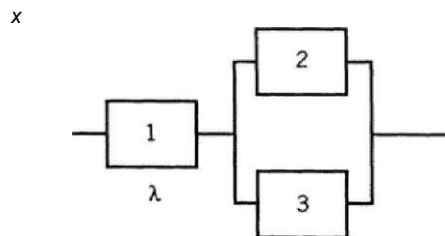


Figure E5.3.

**SOLUTIONS**

5.1 See Figure E5.1.

5-2 Consider a clear physical example where such a distribution appears: A standby redundancy group of  $n$  identical and independent units. One

knows that a large number of random variables has an approximately normal distribution (at least far from the "tails"). One knows (see the approximation for a highly reliable redundant group) that  $A(0) = 0$  and  $A(t)$  is increasing by  $t$  and convex near 0. Then consider large  $t$ . If the redundant group is still operating, the probability that there is only one up unit is increasing in  $t$ . But one unit with an exponentially distributed TTF has a constant failure rate. So, the time diagram for  $A(t)$  has the form shown in Figure E5.2.

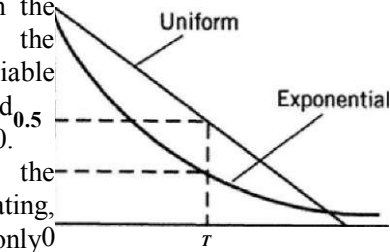
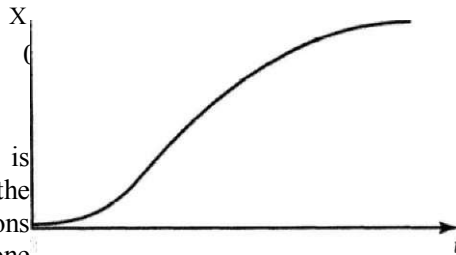


Figure E5.1. the probability that there is only one up unit is increasing in  $t$ .

X(t)A

Figure E5.2.



5.3 The problem as formulated here is incorrect: everything depends on the kind of distribution. If both distributions are exponential, then one should choose the second system. If both systems have an almost constant TTF, there is no difference between them although, from a common viewpoint (with no particular sense in this case!), everybody will again choose the second system. This might be, as a matter of fact, unreasonable if the first system is, for instance, cheaper. But if the first system has any "realistic" distribution of TTF (exponential, normal, etc.) and the second one has a "two-mass" distribution, that is,

$$TTF_2 = \begin{cases} 0 \\ \tau_2 \end{cases}$$

with probability  $p$  and with probability  $1 - p$

where  $\tau_2 > 300$ , the solution is not unique.



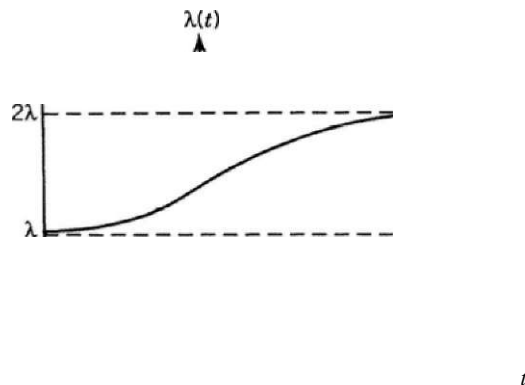
Consider an exponential d.f. with  $\lambda = 1/200$ . For this distribution  $PFFO_2(t) = e^{-\lambda t}$ . For the second case let  $p = 0.9$  and  $r_2 = 3000$  hours. This corresponds to  $MTTF_2 = 300$  hours. In this case  $PFFO_2(100 \text{ hours}) = 0.1$  which is worse than the exponential distribution considered above.

Now assume that  $p = 0.5$  and  $r_2 = 600$  hours. Then  $PFFO_2(100 \text{ hours}) = 0.5$ , which is better than the previous case. For other distributions one can obtain similar conclusions (with other numerical results).

**5.4** One should repeat all of the arguments used in the solution of Exercise 5.2 taking into account that:

- \* Unit 1 might be the cause of the system failure during all periods of time.
- \* For a large time period, the parallel connection of units 2 and 3 with probability close to 1 will consist of only one unit; thus the entire system also will almost surely consist of two series units: unit 1 and one of the units 2 or 3.

The solution is represented graphically in Figure E5.4.



0  
**Figure E5.4.**

## CHAPTER 6

# REPAIRABLE SYSTEMS

In engineering practice, one of the most important objects under investigation is a *repairable system*. In general, repairable systems might be analyzed with the help of Monte Carlo simulation. There are no essential analytical results for the most general mathematical models except for some very particular cases. The most important analytical models frequently used in practice are Markov models. For these models all the system units' random TTFs and repair times are assumed to be exponential. (More accurately, each random duration of being in any state has an exponential distribution.) These assumptions might be far from valid, and so each time their appropriateness must be carefully considered. Note that if the suggestion about exponentially distributed TTFs is admissible (especially for electronic equipment), it seems artificial for the repair time. Indeed, the residual repair time should depend on the time already spent. We have discussed this issue earlier. But as we will show below, sometimes the assumptions of a distribution's exponentiality produce acceptable numerical results that can be utilized in engineering design. At any rate, Markov models are very popular for practical engineering problems because of their clarity and mathematical simplicity.

### 6.1 SINGLE UNIT

#### 6.1.1 Markov Model

We first consider the simplest possible repairable system: a single unit. At any moment in time, the unit is in one of two states: it is either operating or

it has failed. The transition graph is presented in Figure 6.1. Here state 0 denotes an operating state, and state 1 corresponds to a failed state. This graph has a simple interpretation. When in state 0, the unit might go to state 1 or stay at the current state. Leaving state 0 occurs with an intensity  $\lambda$ , and leaving state 1 occurs with an intensity  $\mu$ .

The unit transition process can be described as an alternative renewal process. It is represented by a sequence of mutually independent r.v.'s  $\tau$  (a unit up time) and  $t$ ; (a unit repair time). Both  $\tau$  and  $t$  have exponential distributions with parameters  $\lambda$  and  $\mu$ , respectively. A sample time diagram is presented in Figure 6.2.

Using the graph of Figure 6.1, we can easily construct the following formula:

$$P_0(t + \Delta t) = (1 - \lambda \Delta t) P_0(t) + \mu \Delta t P_1(t) \quad (6.1)$$

This expression means that the transition process may appear in state 0 at moment  $t + \Delta t$  under the following conditions:

- \* It was there at moment  $t$  and did not leave during the interval  $\Delta t$ .
- At moment  $t$  it was in state 1 and moved to state 0 during the interval  $\Delta t$ .

The conditional probability of leaving state 0 equals  $\lambda \Delta t$ , and the conditional probability of leaving state 1 equals  $\mu \Delta t$ .

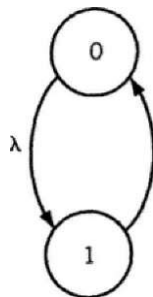


Figure 6.1. Transition graph for a renewable unit.

STATES

UP

DOWN

Figure 6.2. Time diagram for a renewable unit.

From (6.1) we obtain

$$\frac{p_0(t) - p_0(t_0)}{\Delta t} = -\lambda P_0(t) + \mu P_1(t) \quad (6.2)$$

In the limit as  $\Delta t \rightarrow 0$ , we obtain

$$\dot{P}_0(t) = -\lambda P_0(t) + \mu P_1(t) \quad (6.3)$$

This represents the simplest example of Kolmogorov's equation. This equation expresses a condition of dynamic equilibrium. To solve it with respect to any  $P_k(t)$ , we need to have one more equation. It is clear that another equation cannot be obtained in the same manner: it would be linearly dependent on the first one and, consequently, would not be useful in obtaining the solution. The second equation which should be chosen is the so-called *normalization equation*:

$$P_0(t) + P_1(t) = 1 \quad (6.4)$$

which means that at any moment the unit must be in one of two possible states.

We also need to determine the initial condition for the solution of the system of differential equations. In this simple case the problem can easily be solved in general when  $P_0(t_0) = p$  and, consequently,  $P_1(t_0) = q, p + q = 1$ .

This problem can be solved with the help of different methods. We will use the Laplace-Stieltjes transform (LST) to make the presentations in the book uniform.

Recall that the LST  $\phi(s)$  of the function  $f(t)$  is defined as

$$\phi(s) = \int_0^{\infty} f(t) e^{-st} dt \quad (6.5)$$

(In this context we consider functions  $f(t)$  defined over the positive axis.)

**Nonstationary Availability Coefficient** The system (6.2) and (6.4) for this case has the LST:

$$s\phi_0(s) - p_0(0) = -\lambda\phi_0(s) + \mu\phi_1(s) \quad (6.6)$$

$$\phi_0(0) + \phi_1(0) = 1 \quad (6.7)$$

or, in canonical form,

$$(A + s) < P_{0is} - wis) = p$$
$$s < P_0(s) + s < P_s(0 = 1$$

Thanks to the LST, the system of linear differential equations turns into a system of algebraic equations. To solve (6.7), we can use Cramer's rule:

$$\langle Po(s) \rangle = \frac{\begin{vmatrix} P & \\ 1 & 5 \end{vmatrix}}{\begin{vmatrix} A + s & -p \\ s & s \end{vmatrix}} = \frac{ps + p}{s^2 + (A + ft)s} \quad (6.8)$$

To invert this LST, we have to present it in the form of a sum of terms of type  $a/s$  or  $b/(s + a)$ . The inverse functions for these terms are a constant and an exponential function, respectively.

To present the solution (6.8) in the desired form, we should find the roots of the denominators of (6.8). They are:  $s_1 = 0$  and  $s_2 = -(A + p)$ . Now we can write

$$\langle Po(s) \rangle = \frac{A}{s} + \frac{B}{s + A + p} \quad (6.9)$$

where  $A$  and  $B$  are the unknown constants to be determined. To find them, we should note that two polynomials with similar denominators are equal if and only if the coefficients of their numerators are equal. Thus, we set the

$$Ps + p = A(s + A + p) + Bs \quad (6.10)$$

representations

$$\frac{A}{s} + \frac{B}{s + A + p} = \frac{Ps + p}{s(s + A + p)}$$

equal:

And so we obtain a new system for  $A$  and  $B$  by equalizing the coefficients of the polynomials:

$$A + B - p = 0 \quad (6.11)$$

$$A(A + p) - p = 0$$

It is easy to find

$$A = \frac{p}{A + p} \quad (6.12)$$

$$B = \frac{p - p(A + p)}{A + p}$$

$$A + p = \frac{p}{A + p}$$

Thus, the LST of interest can be written as

$$\begin{aligned} & \text{M} \\ <Po(s) & \text{A} + p \cdot s \quad \text{A} + p. \\ & \quad \quad \quad \text{£ } Ap - - p) \\ & \text{A} + p + s \end{aligned} \quad (6.13)$$

Finally, the nonstationary availability coefficient, that is, the inverse LST of (6.10), is

$$K(0 = p_0(t) = \frac{IL}{fjL} + \frac{Ap \sim u(1-p)}{A + T fJL} \quad (6.14)$$

If the original system state is operational, that is, if  $P_0(t) = 1$ , the solution is

$$= \frac{A}{A + fJL} + \frac{1}{A + T fJL}$$

The function  $K(t)$  showing the time dependence of the system availability is presented in Figure 6.3.

**Stationary Availability Coefficient** It is clear that if  $t \rightarrow \infty$ ,  $K(t)$  approaches the stationary availability coefficient  $K$ :

$$K = \frac{A}{A + fJL} + \frac{1}{1 + T} \quad (6.16)$$

where  $T = 1/A$  is the unit's MTTF and  $r = 1/n$  is the unit's mean time to repair (MTR).

We should notice that, in general, such a method of obtaining a stationary availability coefficient is not excusable in a computational sense. For this purpose, one should write a system of linear algebraic equations in a direct way without the use of the system of differential equations. It is important to realize that the stationary regime represents static equilibrium. This means that all derivatives  $dP_k(t)/dt$  are equal to 0 because no states are changing in

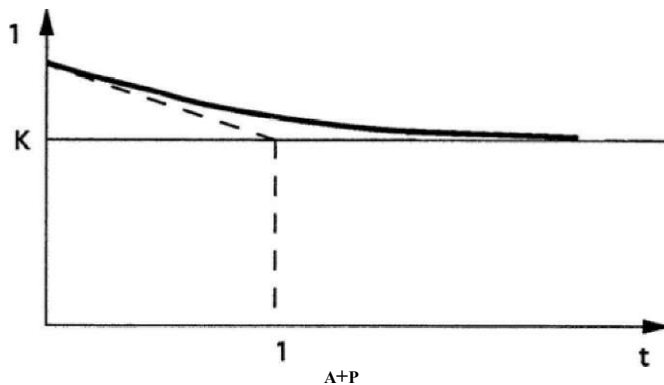


Figure 6.3. Time dependence of nonstationary availability coefficient  $K(t)$  for exponential distributions of TTF and repair time.



time, "on the average." Consequently, all  $P_k$ 's must be constant. It is also clear that the initial conditions (the original state of the unit at moment  $t = 0$ ) also will not make any sense. This assumption leads directly to the following system of algebraic equations:

$$\begin{aligned}
 & -\lambda P_0 + \lambda = 0 \\
 P_0 + P_1 & = 1
 \end{aligned} \tag{6.17}$$

where

$$P_k - \lambda m P_k(t) \tag{6.18}$$

are the stationary probabilities of interest. Again, the solution can be obtained with Cramer's rule

$$P_0 = \frac{\begin{vmatrix} 0 & \lambda \\ 1 & 1 \end{vmatrix}}{\begin{vmatrix} -\lambda & \lambda \\ 1 & 1 \end{vmatrix}} \tag{6.19}$$

Of course, we mention Cramer's rule not as a computational tool, but rather as a methodological reference. Everyone might choose his or her own method for this particular computational task.

**Probability of a Failure-Free Operation** Considering previous reliability indexes, we assumed that both unit states are transient. But if one needs indexes such as the probability of a failure-free operation during a specified time interval, or the MTTF, the transition graph should be reconstructed. In these cases the unit failure state has to be absorbing.

The transition graph for this case is presented in Figure 6.4. There is no transition from state 1 back to state 0, that is,  $f_{10} = 0$ . In this case we have the

# 6

O

Figure 6.4. Nontransitive graph for computation of the MTTF of a renewable unit with an exponentially distributed TTF.

equation

$$\sim P_0(t) = \sim X P_0(t) \tag{6.20}$$

This differential equation can again be solved with the help of the LST. First, we write the following algebraic equation with the natural initial condition  $P_0(0) = 1$ :

$$-1 + s \sim P_0(s) = -A \sim P_0(s) \tag{6.2}$$

1)

and then solve it to obtain

$$W \tag{6.22}$$

)

**Mean Time to Failure** To find the unit's MTTF, we should recall that the mean of nonnegative r.v.'s can be found as

$$E\{X\} = \int_0^\infty \sim P(x) dx \tag{6.23}$$

where  $P(t) = 1 - F(t)$ . Using the previous notation, we take  $P(t) = P_0(t)$ . At the same time, we can write

$$T^* = \int_0^\infty P(t) dt = \int_0^\infty \sim P(t) dt$$

It follows that, to find the MTTF, we can use the solution for  $P_0(t)$  in terms

of the LST and substitute  $s = 0$ . In fact, it is even sometimes simpler to solve a corresponding system of equations directly with the substitution  $s = 0$ .

Considering a single unit, there is no technical difference:  $\frac{1}{j - 0 A} \tag{6.25}$

Notice that if we need to find the MTR, it is necessary to start from state 1 and choose state 0 as absorbing.

We present an in-depth analysis of this simple case in order to make future explanations of more complex models more understandable. We do this to avoid explanations below with unnecessary additional details. The same purpose drives us to use a homogeneous mathematical technique for all routine approaches (though, in general, we try to use various methods because our main purpose is to present ideas and not results).

**6.1.2 General Distributions**

Many results can be obtained for a renewal unit. We remark that it might be very useful for the reader to review Section 1.6.5.

Consider an alternative renewal process  $\{ft_{\tau}\}$  starting with a subinterval of type  $ft$  that is, at  $t = 0$  an r.v.  $ft$  starts. This process can be considered as a model of the operation of a socket with installed units which is replaced after failure. In this case  $\xi$  is the random TTF and  $\eta$  is the random repair time. Let  $F(t)$  and  $G(t)$  be the distributions of the r.v.'s  $\xi$  and  $\eta$ , respectively. Let us call  $\theta_k = (ft + \eta)_k$  the  $k$ th cycle of operation of the socket. The distribution of  $\theta$  can be written as

$$B_{\theta}(t) = \int_0^t f'F(t-x)dG(x) = \int_0^t f'G(t-x)dF(x) \tag{6.26}$$

**Nonstationary Availability Coefficient** This reliability index means that at moment  $t$ , a unit is in an up state; that is, one of the r.v.'s  $\xi$  covers the point  $t$  on the time axis (see Figure 6.5): Consider a renewal process formed with  $\{0\}$  and denote a renewal function of this process by  $H(t)$ . Then, using the results of Section 1.5.2, we immediately obtain the following integral equation:

$$K(t) = 1 - F(t) + \int_0^t (1 - F(t-x))dH(x) \tag{6.27}$$

In other words, (6.27) means that either no failures have occurred or—if failures have occurred—the last cycle  $\theta$  is completed by moment  $\tau$ ,  $0 < x < t$ , and a new r.v.  $\xi$  is larger than the remaining time,  $\xi > t - x$ . The function  $H(t)$  in this case is

$$H(t) = \int_0^t f l^{**}(\tau) d\tau$$

where  $B^{*k}(t)$  is the border convolution of  $B(t)$ . Thus, in general,  $K(t)$  could be found with the help of (6.27).

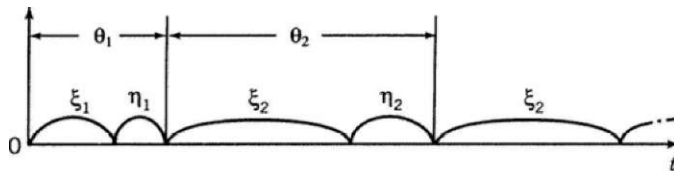


Figure 6.5. Time diagram for an alternative renewal process describing a unit operation.

**Stationary Availability Coefficient** Intuitively, it becomes clear that (6.27) has a limit when time is increasing:  $K(t) \rightarrow K$ . (Strictly speaking, the involved distributions must be continuous.) Indeed, applying the Smith theorem (Section 1.5.2), we obtain

$$K = \frac{E\{L\}}{E\{L\} + \sum_{i=1}^n \frac{f_i(t)}{n}}$$

On a heuristic level, this result can be explained by the following arguments. Consider some interval of time  $L$  such that the number of cycles on it  $n$  is sufficiently large. Then

The index  $K$  is the probability that an arbitrary moment will be covered by an interval of type  $f$ . It is clear that this probability is proportional to the total portion of time occupied by all intervals of type  $f$ :

$$K = \frac{\sum_{i=1}^n \int_0^{\infty} f_i(t) dt}{\sum_{i=1}^n \int_0^{\infty} f_i(t) dt + L}$$

and, if  $n$  is large, one may replace each sum with the coefficient  $1/n$  for the mean of the respective r.v.

$$K = \frac{E\{L\}}{E\{L\} + \sum_{i=1}^n \frac{f_i(t)}{n}} \tag{6.29}$$

**Nonstationary Interval Availability Coefficient** Again, we can write the integral equation

$$f_i(t_0) = 1 - F(t_0) + \int_0^{t_0} [1 - F(t_0 - x)] dH(x) \tag{6.30}$$

The explanation of (6.30) is similar to the explanation of (6.27).

**Stationary Interval Availability Coefficient** Again, we use the Smith theorem and write

$$R(t_0) = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m f_i(t_0) \tag{6.31}$$

It is conv $R(t_0) = K \Pr\{\xi \geq t_0\} = K \frac{1}{E\{\xi\}} \int_{t_0}^{\infty} P(x) dx$  (6.32)

where  $\xi$  is the residual time of the renewal process formed with the r.v.'s  $\{\xi\}$ . From (6.32) it becomes clear that  $R U_0$  differs from  $ff_{\text{wrong}}(f_0) = KP\{t_0\}$ . In engineering practice, nevertheless,  $i?_{\text{wroi,g}}(f_0)$  is often erroneously used. We should emphasize that  $\xi$  and its residual time  $\xi$  are statistically equivalent only for an exponentially distributed r.v. Consequently, in this case (and only in this case!),

$$R(t_a) \sim KP(t_0) = R(t_0) = Ke$$

For a highly reliable unit, (6.32) can be written in the convenient form of two-sided bounds if  $Fit$  is "aging." For this purpose we use a result from Chapter 5. Recall that

$$\int_t^{\infty} D_T(x) dx \leq \int_t^{\infty} F(x) dx \leq \int_t^{\infty} e^{-x/T} dx$$

where  $F(t)$  is an "aging" distribution with mean  $T$  and  $D_T U$  is a degenerate distribution, that is, a constant  $T$ . Then it follows that

where  $f$  is the residual value of the renewal process formed with  $\{\xi\}$ . For a highly reliable unit, we can write a very simple and very convenient approximation

$$P(* o) = 1 \sim T$$

Thus, for the  $\frac{1}{1 + \frac{\tau}{T}} \left(1 - \frac{t_0}{T}\right) \leq R(t_0) \leq \frac{1}{1 + \frac{\tau}{T}} e^{-t_0/T}$  index of interest, we write

(6.33)

and for a highly reliable unit

250 REPAIRABLE SYSTEMS

$$R(t_0) \approx 1 - \frac{\tau + t_0}{T}$$

(6.34)



## 6.2 REPAIRABLE SERIES SYSTEM

### 6.2.1 Markov Systems

Consider a series system of  $n$  independent units. Assume that distributions of the TTF,  $F_j(t)$ , and distributions of the repair time,  $G_j(t)$ , are exponential:

$$F_j(t) = 1 - e^{-\lambda_j t} \quad G_j(t) = \int_0^t e^{-\mu_j x} dx$$

Here  $\lambda_j$  and  $\mu_j$  are the parameters of the distributions, or the intensities of failure and repair, respectively.

Reliability indexes depend on the usage of the system's units during the system's idle time. We consider two main regimes of system units in this situation:

1. After a system failure, a failed unit is shipped to a repair shop and all of the remaining system units are switched off. In other words, the system failure rate equals 0 during repair. In this case only one repair facility is required and there is no queue for repair.
2. After a system failure, a failed unit is shipped to a repair shop but all the remaining system units are kept in an operational state. Each of them can fail during the current repair of the previously failed unit (or units). In this case several repair facilities might be required. If the number of repair facilities is smaller than the number of system units, a queue might form at the repair shop.

**System with a Switch-Off During Repair** The transition diagram for this system is presented in Figure 6.6. We will not write the equations for this case. As much as possible, we will try to use simple verbal explanations.

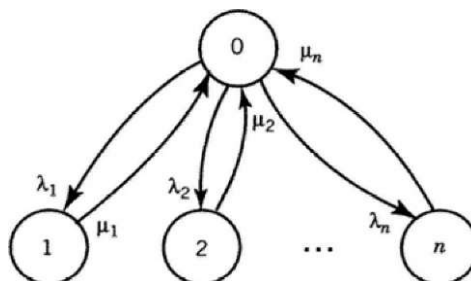


Figure 6.6. Transition graph for a series system which is switching off during idle time.

1. Probability of Failure-Free Operation

Any exit from state 0 leads to failure. Hence,

$$P_0(t) = \exp(-\lambda t) \quad (6.35)$$

where

$$\lambda = \sum_{i=1}^n \lambda_i$$

Thus, by this reliability characteristic, the system is equivalent to a single unit with a failure rate  $\lambda$ .

2. MTTF

If  $P_0(t) = e^{-\lambda t}$ , the MTTF of the system equals  $T_{sys} = 1/\lambda$ . No comments are needed.

3. Mean Repair Time

Let us consider a general case where all units differ by their repair time  $1/\mu_k$ . The current repair time of the system depends on which unit has failed. The distribution of the system's repair time can be represented in the form

$$Pr\{T = t\} = \sum_{k=1}^n p_k e^{-\lambda t} \quad (6.36)$$

where  $p_k$  is the probability that the  $k$ th unit is under repair. The probability  $p_k$  can be easily found as

$$p_k = \frac{\lambda_k}{\lambda} \quad (6.37)$$

Notice that the distribution of the system's repair time has a decreasing intensity function; that is, with the growth of the current repair time, the residual repair time becomes larger and larger.

We consider this phenomenon in more detail. Write (6.36) in the form

$$r \int_0^t \sum_{k=1}^n p_k(x) dx = \exp(-\dots)$$

From here we find

$$M \int_0^t r(t) dt = \sum_{k=1}^n \dots$$

Now we note that  $f_i(t)$  is a monotone function. For  $t = 0$ , a simple qualitative analysis gives us

$$\dots = \dots$$

Now, as  $t \rightarrow \infty$ ,

$$\lim_{t \rightarrow \infty} \dots = \dots$$

where  $\dots$  corresponds to the subscript of a minimal  $f_{ik}$ . Obviously, the average value is larger than the minimum. This function is never below the minimal. Hence,  $n(t)$  decreases from the average value of  $p$  at  $t = 0$  to the minimal value among all  $f_i$ 's. It can be shown that this decrease is monotone.

Of course, from (6.36) and (6.37), it follows immediately that

$$T_{\text{system}} \sim \dots$$

where  $\tau_i = 1/r_i$  is the MTTF of the  $i$ th units.

#### 4. Nonstationary Availability and Interval Availability Coefficients

We are able to find these reliability indexes only with the help of general methods of renewal process theory, in spite of the exponentiality of a TTF distribution. One can also use standard Markov methods applied to the

## 254 REPAIRABLE SYSTEMS

transition graph presented in Figure 6.6. The corresponding system of equa-

tions for the availability coefficient is

$$\begin{aligned}
 & \frac{dP_0(t)}{dt} + \lambda P_0(t) = \mu \sum_{k=1}^n P_k(t) \\
 & = 1 \text{ taken for } 0 \leq t < \infty
 \end{aligned}$$

and the initial condition  $P_0(0) = 1$ . We will not solve these equations here. But if  $K(t)$  is found, then—because of the exponentiality of the TTF distribution— $R(t, t_0) = K(t)e^{-\lambda t}$ .

### 5. Stationary Availability Coefficient

With known  $\lambda_{\text{sys}}$  and  $r_{\text{sys}}$ , this index can be found in a standard way as  $K = (T_{\text{sys}}) / (T_{\text{SL}} + r_{\text{sys}})$ . Note that in this particular case it is convenient to write

$$K = \frac{\lambda_{\text{sys}}}{1 + \lambda_{\text{sys}} r_{\text{sys}}} \tag{6.38}$$

### 6. Stationary Interval Availability Coefficient

Because of the exponential distribution of the system TTF, we can use the expression  $R(t_0) = K P(t_0)$  where  $P(t_0)$  is defined in (6.35).

Notice that if all are constant (equal to  $\lambda$ ), the above-described system is transformed into a single repairable unit with an intensity of failure equal to

$$\lambda_{\text{sys}} = \lambda \sum_{i=1}^n A_i$$

and an intensity of repair  $\mu$ . In most practical cases it is enough for the first stages of design to put  $\lambda_{\text{sys}} = E\{X\}$  in this model and to use this approximation instead of using the exact model. We remark that in most practical cases, when the equipment has a modular construction, the mean time of repair might be considered almost equal for different system units. But an even more important argument for such a suggestion is that a mathematical model must not be too accurate at the design stage when one does not have accurate input data.

**System Without Switch-Off During Repair** First, consider a series system of  $n$  different repairable units when there are  $n$  repair facilities in a workshop; that is, each unit might be repaired independently. The units' failures are assumed independent. In this case the system can be considered



**REPAIRABLE SERIES SYSTEM 257**

realistic practical case, by the way). In this case there is no way other than to construct a transition graph, to write a system of linear differential equations, and then to solve them.

**6.2.2 General Distribution of Repair Time**

If the TTFs of all the units remain exponentially distributed, the main simple results can be obtained practically in the same form as for the Markovian model.

**System With Switch-Off During Repair** First of all,  $f_{\text{sys}}(t_0)$  and  $T_{\text{sys}}$  remain the same as in the previous case. The mean repair time is defined with the help (6.36). Consequently, the stationary availability and interval availability coefficients can be expressed in standard form. At the same time, nonstationary indexes can be found with the help of the general methods of renewal process theory. The model of the investigated operation process forms an alternative process  $\{\lambda^*, \tau_j^*\}$ . Each is an exponential r.v. with parameter  $\lambda$  and  $\tau_j^*$  is an r.v. with a complex "weighted" d.f.

$$G^*(s) \cdot Pr\{T < t\} \dots E A, G_A(0) \sum Z_i Z_n$$

For analytical purposes it is more reasonable to use Monte Carlo simulation. We would like to emphasize again that a detailed exploration of a nonstationary regime is usually a task far removed from practical needs because of the insufficiency of the input data.

**System Without Switch-Off During Repair** In this case  $P_{\text{sys}}(f_0)$  and  $T_{\text{sys}}$  remain the same as in the previous cases. Even such a stationary index as  $K$  can be found only for the case when the number of repair facilities equals the number of system units, that is, when all system units are totally independent. In this case  $K$  is defined as

$$K = \frac{1}{1 \wedge_{\text{sys}} \wedge_{\text{sys}}}$$

If the system units are dependent through the lack of repair facilities, we recommend the use of Monte Carlo simulation for the computation of nonstationary indexes.

But if  $K$  or  $K(t)$  is known, to find  $R(t_0)$  or  $R(t, t_0)$  is a simple task because of the exponentiality of the system TTF:  $R(t_0) = KP(t_0)$  and  $R(t, t_0) = K(t)P(t_0)$ .

**6.2.3 General Distributions of TTF and Repair Time**

This case is especially difficult if one considers nonstationary indexes. They can only be found with the help of Monte Carlo simulation. Let each unit of the system be described with the help of an alternative renewal process. The



superposition of these processes is not an alternative renewal stochastic process. The new process has more sophisticated structure: it does not have the regeneration moments that appeared when one considered a system of units with exponential TTF.

But to find the stationary coefficient  $K$ , we can use the idea that stationary probabilities do not depend on the distribution of the repair time. Therefore, one can use a Markov model with the following parameters for each unit:

$$Z = \frac{\int_0^{\infty} [1 - F_i(t)] dt}{\int_0^{\infty} f_i(t) dt} \text{ and } \tau_i = \frac{1}{\int_0^{\infty} f_i(t) dt} dt$$

If the number of repair facilities equals the number of system units (all units are totally independent), the system stationary availability coefficient can be found as

$$I_{sist} = \prod_{i=1}^n \frac{1}{1 + \tau_i \lambda_i}$$

The stationary interval availability coefficient can also be found with the help of the following arguments. For each unit we can easily find the conditional stationary probability of a failure-free operation under the condition that a unit is in an operational state at the starting moment:

$$= \Pr\{t \leq f_0\} = \int_0^{\infty} P(x) dx$$

Then for the system

$$I_{S/Sn} = \prod_{i=1}^n \frac{1}{1 + \tau_i \lambda_i}$$

Now it is possible to write  $J^*(r_0)$  as

$$J^*(r_0) = \int_0^{\infty} P(x) dx \prod_{i=1}^n \frac{1}{1 + \tau_i \lambda_i}$$

Also, we can again use the two-sided bounds (6.33) if the unit TTF distributions are "aging":

$$\prod_{i=1}^n \frac{1}{1 + \tau_i \lambda_i} \leq J^*(r_0) \leq \prod_{i=1}^n \frac{1}{1 + \tau_i \lambda_i} \quad (6.41)$$

## 260 REPAIRABLE SYSTEMS

Naturally, using (6.34) for a highly reliable system, we can write

$$P_{\text{sys}}^*(0) = 1 - \sum_{i=1}^k \lambda_i \tau_i \quad (6-42)$$

Of course, analogous approximations could be written for all of the above-considered cases in this section.

## 6.3 REPAIRABLE REDUNDANT SYSTEMS OF IDENTICAL UNITS

### 6.3.1 General Markov Model

Let us consider a redundant system consisting of  $k$  main operating units and  $n = n_1 + n_2 + n_3$  redundant units. Here we use the following notation:

- $n_1$  is the number of active redundant units in the same regime as the main operating units; each unit has a failure rate  $\lambda$ ;
- $n_2$  is the number of units in an underloaded on-duty regime; each unit has a failure rate  $\lambda'$ ,  $\lambda' = \nu \lambda$ , where  $\nu$  is the so-called loading coefficient,  $0 < \nu < 1$ ;
- $n_3$  is the number of standby units; each unit has  $\lambda'' = 0$ .

A failed unit is shipped to the repair shop. A failed operating unit is replaced with an active redundant unit. Instantly, this unit is replaced by a unit which is in an underloaded regime, and, in turn, the latter is replaced by a standby unit. An analogous chainlike procedure is performed with a failed unit of other levels of redundancy. There are  $I$  repair facilities,  $1 \leq I \leq n + k$ . All units have an exponential repair time distribution with the same parameter  $\mu$ .

Let  $H_j$  denote a system state with  $j$  failed units. Obviously, the system can change its state  $H_j$  only for one of two neighboring states:  $H_{j-1}$  after the repair of a failed unit or  $H_{j+1}$  after a new failure. Hence, this process is described with the help of a linear transition graph (see Figure 6.8) and belongs to the birth and death process (see Section 1.6).

The transition from state  $H_j$  to state  $H_{j+1}$  within a time interval  $[t, t + \Delta t]$  occurs with probability  $\lambda_j \Delta t + o(\Delta t)$ . The transition to state  $H_{j-1}$  in the same time interval occurs with probability  $\mu_j \Delta t + o(\Delta t)$ . With probability  $1 - \lambda_j \Delta t - \mu_j \Delta t - o(\Delta t)$ , no change occurs. For underloaded units, the coefficient of loading is  $\nu$ ,  $0 < \nu < 1$ .

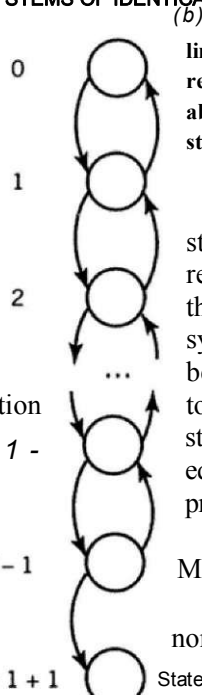
A system with  $n$  redundant units has  $n + k + 1$  states  $H_0, H_1, H_2, \dots, H_{n+k}$ . States  $H_{n+j}$  with  $n+j$  failed units,  $1 \leq j \leq k$ , are states corresponding to a system failure. After an exit into the first system failure state,  $H_{n+1}$ , the process develops further: it may move to the next

Figure 6.8. Two graphs for a system failure on, or it may state,  $H_n$ . If absorbing, the equations must absorbing states

used for successful availability MTBF. If we stationary state  $H_{n+k}$  is equations can be



must have no transition operational new system of calculating the coefficient, the consider the States of > the system's failure



linear transition redundant system: absorbing state; (ft) state.

state,  $H_{n+2}$ , and so return to the up the state  $H_{n+1}$  is system of be changed: all to the set of states. This equations can be probabilities of interval MTTF, and the nonstationary and State of the system's failure

availability coefficients, the reflecting. The corresponding system of used

for calculating the nonstationary availability and/or interval availability coefficients.

262 REPAIRABLE SYSTEMS

Considering the transition graph without an absorbing state, for a state  $H_j$ ,  $0 \leq j \leq n + \alpha_2 + n_i + k$ , one may write  $A_y$  and  $Mf$ .

$$\begin{aligned}
 A_0 &= kX + \alpha_1 A + n_2 vX \\
 A_{i+1} &= \lambda A + n_i A + n_2 vX = A_0 \\
 &= A_0 \\
 A_{i+2} &= kX + n_i A + (\alpha_2 - 1) \mu A \\
 A_{i+2} &= \lambda A + \mu A + (\alpha_2 - 2) uX \\
 A_{n+\alpha_2+2} &= \lambda A + \alpha_1 A \\
 A_{n+\alpha_2+2} &= kX + (n_i - 1) A \\
 A_{n+\alpha_2+2} &= \lambda A + (\alpha_1 - 2) A \\
 A_{n+\alpha_2+n} &= \lambda A \\
 A_{n+\alpha_2+n+1} &= (k - 1) A \\
 A_{n+\alpha_2+n+2} &= (\alpha_1 - 2) A \\
 A_{n+\alpha_2+n+3} &= \lambda A
 \end{aligned}$$

and for all  $M_j$ ,  $0 \leq j \leq n + k = n_x + n_2 + n_3 + k$ ,

$$M_j = n, M_2 = 2/JL, \dots, M_n = In, M_{n+1} = ln, \dots, M_{n+k} = lp$$

The system with the absorbing  $H_{n+x}$  state is the system which operates until a first failure. This system can be analyzed with the following system of differential equations:

$$\begin{aligned}
 dp_j(t) &= - (A_{j-1} + \lambda_j) p_j(t) + M_{j+1} p_{j+1}(t) \quad 0 \leq j \leq n + 1 \\
 & \hspace{15em} (6.43)
 \end{aligned}$$

$$A_{n+1} - A_{n+2} = A_0 = A_1 = \dots = A_{n+1} = 0 \quad (6.44)$$

where  $p_j(t)$  is the probability that the system is in state  $H_t$  at moment  $t$ . The normalization condition is

$$\begin{aligned}
 \mathbf{E} PAO &= \mathbf{i} \\
 0s/sn+1 &
 \end{aligned}$$

The system with the reflecting  $H_{n+k}$  state can be described by the following system of differential equations:

$$\frac{dp_i(t)}{dt} = \lambda_i p_{i-1}(t) - (\lambda_i + \mu_j) p_i(t) + \mu_j p_{i+1}(t) \quad (0 < j < n+k)$$

$$A^* = A_{n+k} = M_0 - M_{n+k+1} \ll 0$$

with normalization equation

$$\sum p_i = 1$$

Because our goal is not to write down formulas for very general models but to show the methodology and methods, we hope that the reader can use the corresponding equations from Section 1.6 dedicated to the death and birth process.

Precise formulas for such a general case are almost always long and complicated. If one deals with highly reliable systems, we recommend the reader refer to Chapter 12. (If one deals with an unreliable system, we recommend a redesign of the system, not a useless calculation!)

The next section is devoted to general methods of analysis of repairable systems.

## 6.4 GENERAL MARKOV MODEL OF REPAIRABLE SYSTEMS

### 6.4.1 Description of the Transition Graph

From the very beginning, we would like to emphasize that a Markov model is an idealization of a real process. Our main problem is not to solve the system of mathematical equations but rather to identify the real problem, to determine if the real problem and the model are an appropriate fit to each other. If, in fact, they are a good fit, then a Markov model is very convenient.

Now let us assume that we can construct the transition graph which describes a system's operation. This graph must represent a set of mutually exclusive and totally exhaustive system states with all of their possible one-step transitions. Using some criterion of system failure, all of these states can be divided into two complementary disjoint subsets, *up states* and *down states*. A transition from the subset of up states to the subset of down states may occur only when an operating unit fails. An inverse transition may occur only if a failed unit is renewed by either a direct repair or by a replacement. Let us consider a system with  $n$  units. Any system state may be denoted by a binary vector

$$s = (s_1, \dots, s_n)$$

where  $s_i$  is the state of the  $i$ 'th unit. We set  $s_i = 1$  if the unit is operational and  $s_i = 0$  otherwise. The transition from  $(s_1, \dots, s_i = 1, \dots, s_n)$  to  $(s_1, \dots, s_i = 0, \dots, s_n)$  means that the  $i$ 'th unit changes its state from up to down. The transition rate (or the transition intensity) for this case equals the  $i$ 'th unit's failure rate.

A transition from system state  $(s_1, \dots, s_i = 0, \dots, s_n)$  to state  $(s_1, \dots, s_i = 1, \dots, s_n)$  means that the  $i$ 'th unit was in a failed state and was renewed. The transition rate for this case equals the  $i$ 'th unit's repair rate. These kinds of transitions are most common. For Markovian models we assume that only one unit may fail (or be renewed) at a time. (If several units may change states simultaneously, for example, under a group repair, we will consider this separately.) Of course, there are other possible interpretations of states and transitions. For instance,  $s_i = 1$  may be a state before monitoring or switching, and  $s_i = 0$  is the same state after the procedure. We denote these transitions from state to state on transition graphs with arrows. The rates (intensities) are denoted as weights on the arrows. The graph structure is determined by the operational and maintenance regime of the system's units and the system itself. After the transition graph has been constructed, it can be used as a visual aid to determine different reliability indexes. An example of such a transition graph for a system consisting of three different units is presented in Figure 6.9.

#### 6.4.2 Nonstationary Coefficient of Availability

Let  $E(k)$  denote the subset of the entire set of system states which includes states from which a direct transition to state  $k$  is possible, and let  $e(k)$  denote the subset to which a direct transition from state  $k$  is possible. The union  $E(k) \cup e(k)$  is the subset of system states that have a direct connection to or from state  $k$  (see Figure 6.10). For each state  $k$  of the transition graph, we can write the following differential equation:

$$\frac{d}{dt} p_k(t) = \sum_{i \in E(k)} A_{ik} p_i(t) - p_k(t) \sum_{j \in e(k)} A_{kj} \quad (6.45)$$

where  $A_{ik}$  is the intensity of the transition from  $i$  to  $k$ , and  $p_k(t)$  is the probability that the system is in state  $k$ , at moment  $t$ .

The transition graph and the system of differential equations can be interpreted as those which describe a dynamic equilibrium. Indeed, imagine that each node  $i$  is a "basin" with "liquid" which flows to each other node  $j$  (if there is an arrow in the corresponding direction). The intensity of the flow is proportional to  $A_{ij}$  (specified in each direction) and to a current amount of "liquid" in the source  $p_i(t)$ .

If there are  $n$  states, we can construct  $n$  differential equations. To find the nonstationary coefficient of availability, we take any  $n - 1$  equations, add the

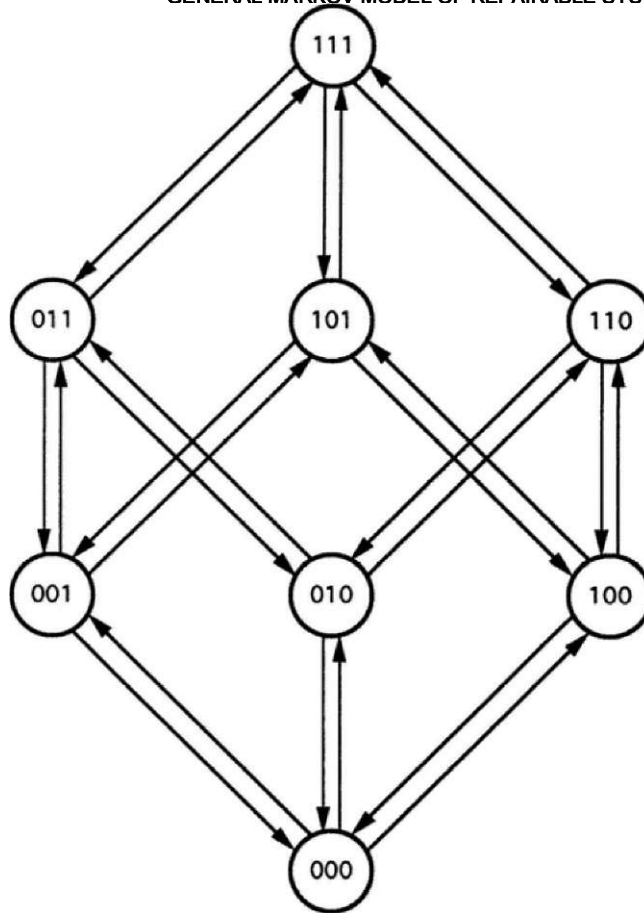


Figure 6.9. Transition graph for a system consisting of three different renewable units.

normalization condition

$$\sum_i L P_i(t) = 1 \tag{6-46}$$

and add initial conditions of the type  $p_j(0) = p_j$  where  $p_j(0)$  is the probability that the system is in state  $j$  at  $t = 0$ . In turn, the  $p_j$ 's are probabilities that conform to a normalization condition similar to (6.46). If  $p_i = 1$  for some  $i$ , then  $p_j = 0$  for all  $j, j \neq i$ . In most problems the initial system state is the state when all units are up.

To find the nonstationary availability coefficient, we can use the Laplace-Stieltjes transform (LST). Then the system of  $n$  linear differential

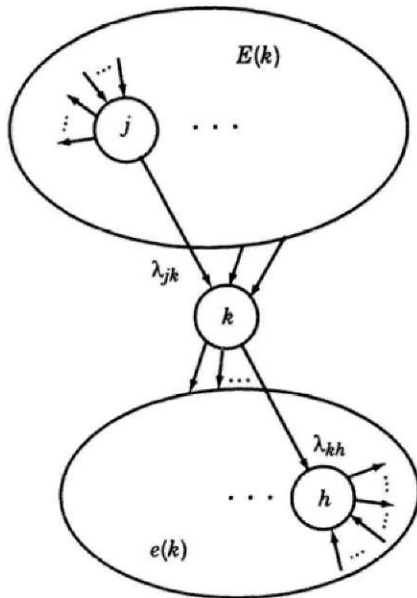


Figure 6.10. Fragment of a transition graph for compiling a system of differential equations.

equations transforms into the system of linear algebraic equations:

$$\sum_{i \in E(k)} \lambda_{jk} p_i(s) = -\sum_{i \in E(k)} \lambda_{jk} p_i(s) + \sum_{i \in E(k)} \lambda_{jk} p_i(s)$$

$$\sum_{i \in E(k)} \lambda_{jk} p_i(s) = 1$$

$i \in E(k)$

where  $\langle p, (s) \rangle$  is the LST for  $p, (t) :$

$$f_i^*(s) = \int_0^\infty p_i(t) e^{-st} dt$$

For writing a system of algebraic equations directly in terms of the LST, one can construct a special graph which is close to the one depicted in Figure 6.11. This new graph includes a state (distinguished by shadowing) which "sends" to each state  $i$  of the graph a "flow" equal to the value of  $p_i(0)$ . Recall that this is the probability that the system is in state  $i$  at time  $t = 0$ . At the same time, each state "sends" to this special state a "flow" equal to  $s$  (argument of the LST). The construction of this graph can be clarified by a comparison with the previous one depicted in Figure 6.10.

In (6.47) we use any  $n - 1$  equations of the total number  $n$ , because the entire group of equations is linearly dependent. This is always true when we consider a transition graph without absorbing states. In this case, in order to find all  $n$  unknown  $p_i(0)$ 's, we must use the normalization equation (6.46).



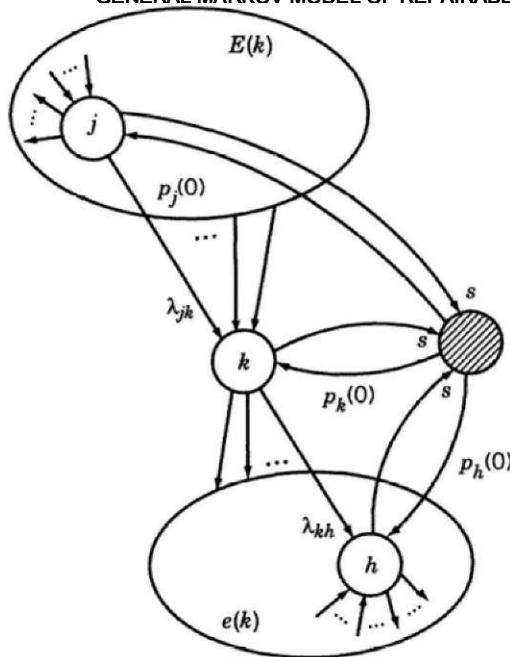


Figure 6.11. Fragment of a transition graph for compiling a system of algebraic equations in LST terms.

This system of equations may be written in canonical form:

$$\begin{aligned}
 \lambda_{jn} p_j(s) + \lambda_{jn} p_2(s) + \dots + b_{ln}(p_{..}(s)) &= c_l \\
 \lambda_{2l} p_l(s) + \dots + b_{2n} p_n(s) &= c_2
 \end{aligned}
 \tag{6.50}$$

$$\lambda_{nl} p_l(j) + b_{n2} p_2(s) + \dots + b_{nn} p_n(s) = c_{..}$$

where  $b_u$  is the coefficient of the  $y$ th term of the  $r$ th row and  $c_l$  is the corresponding constant.

To solve this system of linear equations, we can apply Cramer's rule:

$$p_j(s) = \frac{D_j(s)}{D(s)}
 \tag{6-51}$$

where  $D(s)$  is the determinant of the system and  $D_j(s)$  is the determinant of the matrix formed by replacing the  $i$ th column of  $D$  by the vector  $(c_1, c_2, \dots, c_n)$ . Once more, we repeat that the reference to Cramer's rule is made for explicit explanations, not as a recommendation for computation.

We then find the LST of the availability coefficient:

$$\langle p(s) = \mathbf{E} V_i(s) - 75777 \mathbf{E} D, (s) \tag{6.52}$$

where  $\mathcal{L}$  is a subset of up states. We can use the following procedure to invert this LST.

1. Write  $\langle p(s)$  in the form

$$\langle p(s) = \frac{A_0 + A_1 s + A_2 s^2 + \dots + A_n s^n}{B_0 + B_1 s + B_2 s^2 + \dots + B_{n+l} s^{n+l}} \tag{6.53}$$

where  $A_j$  and  $B_j$  are known coefficients.

2. Find the polynomial roots:

$$B_0 + B_1 s + B_2 s^2 + \dots + B_{n+l} s^{n+l} = 0$$

Let these roots be  $b_1, b_2, \dots, b_{n+l}$ . Thus,

$$B_0 + B_1 s + \dots + B_{n+l} s^{n+l} = \prod_{j=1}^{n+l} (s - b_j)$$

3. Write  $\langle p(s)$  in the form of a sum of simple fractions:

$$\langle p(s) = \frac{P_1}{s - b_1} + \frac{P_2}{s - b_2} + \dots + \frac{P_{n+l}}{s - b_{n+l}} \tag{6.54}$$

where the  $P_j$ 's are coefficients to be found.

4. Rewrite in the form

$$\langle p(s) = \frac{\mathbf{E}}{(s - b_1)(s - b_2) \dots (s - b_{n+l})}$$

After elementary transformations, we obtain

$$f + a_2 s^2 + \dots = \sum_{j=1}^{n+l} \frac{a_j}{(s - b_j)}$$

where the  $a_j$ 's are expressed through different and  $b_j$ 's.

5. Polynomials of the form  $\langle p(s)$  and of the form of (6.53) are equal if and only if

$$A_0 = \dots = A_i = a_i, \dots, A_{n+l} = a_{n+l}$$

The  $a_j$ 's are defined from these equations.

6. After we have found  $a/s$ , the inverse LST is applied to  $\langle p(s) \rangle$  in the form of (6.54)

$$f(0) - \int_0^\infty T \sim K(t) - \int_0^\infty H_j e^{b_j t} dt + I$$

REMARK. If  $a(r)$  has multiple roots for the denominator, that is, if several  $b/s$  are equal, then (6.54) may be rewritten as

$$\sum_{i=1}^n \frac{T h V}{(s - b_j)^{k_j}}$$

where  $k$  is the number of roots equal to  $b$ , and  $n'$  is the number of different roots. To all terms of the form

$$\frac{P_i}{(s - b_j)^{k_j}}$$

the corresponding inverse LST is applied:

$$P_i \cdot \frac{t^{k_j-1} e^{-b_j t}}{(k_j - 1)!}$$

### 6.4.3 Probability of Failure-Free Operation

To determine the probability of a failure-free operation, absorbing states are introduced into the transition graph. They are the system's failure states. Transitions from any absorbing state are impossible, which means that all transition intensities out of an absorbing state are 0. We can change the domain of summation in the previous equations in a way which is equivalent to eliminating the zero transition rates. Using the previous notation, we can write for an operational state  $k$ :

$$\sum_{a \in E} -R_{T,PK} \{ * \} - \sum_{j \in E} P_{*j}^* ( \sum_{l \in E} E_{j,l} A_{kl} + \sum_{i \in E} E_{i,k} A_{ik} ) I(0)$$

If the transition graph has  $m$  operational states, we can construct  $m$  differential equations. (In this case the equations are not linearly dependent. Of course, we may use the normalization condition as one of the equations in this new system, excluding any one of the differential equations.) These equations and initial conditions are used to find the probability of a failure-free operation of the system.

## 270 REPAIRABLE SYSTEMS

We again use the LST to find the following system of linear differential equations:

$$s \langle p_k(s) - p_k = - \sum_{i \in E(k)} A_{ki} \langle p_k(s) + \sum_{i \in E(k)} A_{ik} V_i(s) \quad (6.55)$$

for all  $k \in E$ . The solution of this system of equations can be found with the help of the same methodology as before.

### 6.4.4 Determination of the MTTF and MTBF

Recall that

$$T = \int_0^{\infty} p(t) dt$$

If  $\langle p_i(s) \rangle$  is the LST for the probability of failure-free operation of the system, then

$$T = \int_0^{\infty} \langle p_i(t) \rangle dt$$

Thus, we can find the MTTF (or MTBF) by solving the following system:

$$-p^* = \langle p^*(0) \rangle L L A_{\text{eff}}(0) \quad i \in E^*$$

for all  $k \in E$ . Note once more that this system was derived from (6.55) by the substitution of  $s \rightarrow 0$ . To find the MTTF, one sets the initial conditions as  $\langle p_i(0) \rangle = I$ , where  $i$  is the subscript of a state in which the system is totally operable. Obviously,  $\langle p_j(0) \rangle = 0$  for all the other states. To find the MTBF, we set the initial conditions in the form  $\langle p^*(i) \rangle = p_f$  where the  $i$ 's in this case are the conditional stationary probabilities of the states  $i$  that belong to  $E^*$ . The latter is a subset of the up states which the process visits first after the system renewal.

The conditional stationary probabilities  $p_f$ 's can be obtained from the unconditional ones as

$$\langle p_i(0) \rangle = P_f =$$

Example 6.1 Consider a repairable system of two different units in parallel (Figure 6.12). The parameters of the units are  $\lambda_1, \lambda_2$ , and  $\mu_1, \mu_2$ . Both units can be repaired independently. The transition graph is presented in Figure 6.13. Here  $P_0$  is the state with both units operational;  $H_1$  ( $H_2$ ) is the state where the first (second) unit failed;  $H_n$  is the state where both units failed.

Let  $p_k(t)$  equal the probability of the  $k$ th state at moment  $t$ . There are two systems of equations to calculate the reliability indexes. If the system's failure state  $H_n$  is reflecting, the system of equations is

$$\frac{d}{dt} p_0(t) = -(\lambda_1 + \lambda_2)p_0(t) + \mu_1 p_1(t) + \mu_2 p_2(t)$$

$$\frac{d}{dt} p_1(t) = \lambda_1 p_0(t) - (\lambda_1 + \mu_1)p_1(t) + \mu_2 p_2(t)$$

$$\frac{d}{dt} p_2(t) = \lambda_2 p_0(t) - (\lambda_2 + \mu_2)p_2(t) + \mu_1 p_1(t)$$

$$p_0(t) + p_1(t) + p_2(t) + p_n(t) = 1$$

Hi

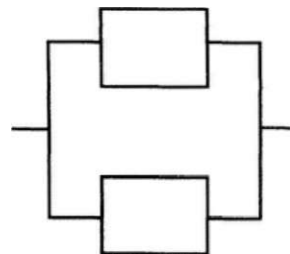
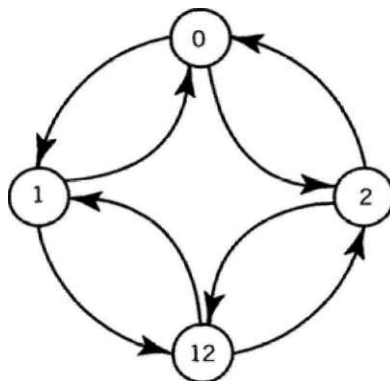


Figure 6.12. Repairable system of two different independent units connected in parallel.



## 272 REPAIRABLE SYSTEMS

Figure 6.13. Transition graph for the system depicted in Figure 6.12.

$$= \sim(\Lambda I + \Lambda_2 > /' o(f) + M, P]( t) + M2P2O)$$

$$= \Lambda IPO(0 - (\Lambda_2 + M_1) / >, (')$$

$$^{TMP_2}(0 = A_2 P_0(R) - (A, + \Lambda_2) p_2(t)$$

$$\sim p_{12}(0 = A_2 p, (f) + A, p_2(/)$$

The corresponding solutions in the form of the Cramer determinants are

$$T = - \begin{matrix} 1 & 0 \\ & M_2 & & MO(0) \\ & & & 0 & & p, (0) \\ & 1 & & 1 & & \\ (A, + A_2) & & & Mi & & \\ & A, & & (\Lambda_0 + M1) & & M_2 \\ & & & 0 & & 0 \\ & & & & & + fi_2) \\ & & & -(A, + fi_2) & & p_2( \\ 0) & & & & & \\ -(A, + A_2) & & & Mi & & \\ A, & & & \sim(A_2 + Mj) & & \\ & & & 1 & & A, & & 0 \\ & & & \sim(\Lambda_0 + Mi) & & & & \\ & & & 0 & & & & \\ & & & & & & & 1 \\ & & & & & & & 0 \end{matrix}$$

$$Jt. 0) \_ \_ \_ \begin{matrix} (A, + A_2) & & & Mi & & \\ A] & & & -(A_2 + fLi) & & \\ A, & & & 0 & & \\ & & & & & & & -(A, + fi_2) \\ & & & & & & & M_2 \\ & & & & & & & 0 \\ & & & & & & & -(A, + m_2) \end{matrix}$$

If state  $H_n$  is absorbing,

274 REPAIRABLE SYSTEMS

1	1	1	1
$-(A_1 + A_2)$	M1	M2	0
$A_1$	$-(A_0 + M1)$	0	M2
$a_2$	0	$-(A_1 + n_2)$	M1
1	1	1	1
$-(A_1 + A_2)$	M1	M2	0
$A_1$	$-(A_2 + M1)$	0	M2
$A_2$	0	$-(A_1 + M_2)$	M1



The solutions are not presented in closed form because of their length and complexity.

## 6.5 TIME REDUNDANCY

Considering reliability indexes, we emphasize that so-called *time redundancy* might be a very effective measure of a system's reliability improvement. This type of redundancy can be used in two main cases:

- The required time of operation completion by an absolutely reliable system is less than the time admissible for operation performance.
- System failures leading to short idle periods might be ignored in the sense of successful operation performance.

These problems are solved with special mathematical methods differing from the usual ones used in other reliability problems. Let us consider several main types of systems with time redundancy.

### 6.5.1 System with instant Failures

Consider a system performing an operation of duration  $t_0$ . System failures are very short, practically instantaneous. The flow of these failures can be successfully described by a point renewal process. Each failure interrupts a system's successful operation, and the system is forced to restart its operation from the beginning. In other words, we assume that a failure destroys the result of an operation. For restarting the operation in an attempt to complete the required performance, the system must have a time resource.

Such situations are encountered in practice if one considers a computer operating with short errors which destroy a current result. A computer performs a task which requires  $t_0$  units of failure-free time for its successful performance. Thus, if there is a time resource, a computer can perform its operation even after the appearance of some error.

We assume that the total time for the system performance is  $T$ . Let us also assume that the system begins to operate at the moment  $t = 0$  when it is "new." The distribution of the TTF is  $F(t)$ . Let  $R_0(t) = P\{T > t\}$  denote the probability that during interval  $[0, T]$  there will be at least one period between failures exceeding the required value  $t_0$ , and let  $P(t) = 1 - F(t)$ .

The system performs its operation successfully during time  $T$  if two events occur:

- There are no failures during time interval  $[0, t_0]$ ,
- A failure has occurred at  $x < t_0$ , but, after this moment, the system successfully performs its operation during the remaining time  $T - x$ .

The latter event is complex. First, a failure might occur at any moment of time between 0 and  $t_0$ , and, second, at the moment of a failure the process

starts from the beginning but for a smaller time interval. This verbal explanation leads us to the recurrent expression

$$R_0(e \setminus T) = P(\theta) + \int_0^\theta R_0(e \setminus T - x) dF(x) \quad (6.56)$$

If the remaining interval is smaller than  $\theta$ , the operation cannot be performed successfully. This leads to the condition

$$R_0(\theta \setminus x < 0) = 0$$

Equations of such a recurrent type are usually solved numerically. We will not provide a mathematical technique for this solution.

Above we considered a situation where a system begins to operate at moment  $t = 0$ . Now let us assume that a system is in an on-duty regime and a request for starting the operation arrives in a random time. More exactly, we assume that we consider a stationary process, and a random time from the request arrival to a system failure is a residual time. Such a situation is typical of many military systems which must be ready at all times to perform their duties: no enemy in modern times informs you about the beginning of hostile actions.

Let the tilde denote a distribution of the residual time. In this case the expression of interest is not changed significantly. We give it without explanation because of its obviousness:

$$\tilde{R}_0(0 \setminus T) = P(0) + \int_0^\theta R_0(0 \setminus T - x) dF(x) \quad (6.57)$$

where the function  $R_0$  under the integral must be taken from (6.56) with the corresponding condition.

Of course, in this case we must again write the condition

$$K(0U < f, \cdot) = 0$$

which means that a system cannot successfully perform its operation if the time resource is smaller than the required time of operation.

### 6.5.2 System with Noninstant Failures

If failures are noninstant, one must take into account the lengths of idle periods between up periods. Let  $G(t)$  denote a distribution of idle time. If a failure has occurred within the first interval  $[0, f_0]$ , a random period of idle time is needed to restore the system. In general, there are no restrictions on the length of the idle time  $y$ . Thus, we must consider the possibility that this value changes within the entire interval  $[*, T - *]$ . At the same time, if the system spent  $x$  units of time for unsuccessful operation and then  $y$  units of

time for restoration, only  $T - x - y$  units of time remain to perform the operation.

This verbal description permits us to write a recurrent expression

$$R(d|T) = P(d) + \int_0^x R(t_0|T-x-y) dG(y) \quad dF(x) \quad (6.58)$$

where  $J_t(0|U < t_0) = 0$ .

Now we consider the above-analyzed system which is operating in an on-duty regime. In principle, the explanation of the equation remains similar to the previous case. We must additionally take into account the fact that the system at an arbitrary stationary moment of time can be found in one of two possible states: up or down. We only explain the situation where a system at the beginning of operation is in a down state. In this case one first observes a residual restoration time and after this a system is considered as "new." Again, we use a tilde to denote the distribution of a residual value.

The expression for this case can be written in the form

$$R(t_0|T) = K P(0) + \int_0^x \tilde{R}(t_0|T-x-y) dG(y) \quad dF(x) \\ + k \int_0^x R(e|T-x) dG(x)$$

where  $K$  is the availability coefficient and  $k = 1 - K$ . Recall that  $K = t/(T + r)$  where  $T$  is the MTTF and  $r$  is the MTR.

### 6.5.3 System with a Time Accumulation

Some systems must accumulate time of successful operation during a total period of performance. Of course, in this case we consider an alternating process of up and down periods. Denote the probability that a system will accumulate more than  $\delta$  units of successful operation during period  $T$  as  $S(0|r)$ . For this probability one can consider two events that lead to success:

- A system works without failures during time from the beginning.
- \* A system failed at moment  $x < \delta$ , was repaired during time  $y$ , and during the remaining interval of  $T - x - y$  tries to accumulate  $\delta - x$  units of time of successful operation. This description leads us to the recurrent expression

$$S(d|T) = P(0) + \int_0^\delta \int_0^{T-x-y} S(t_0-x|T-x-y) dG(y) \quad dF(x)$$

This expression is correct for the case where a system starts to perform at  $t = 0$ .

If a system is in an on-duty regime and begins to accumulate time of successful operation at a stationary arbitrary moment, one must take into account that a system may occur at an up or down state. Each of the corresponding periods is represented by a residual time. The expression for the probability that a system will accumulate more than  $x$  units of successful operation during period  $T$  as  $S(0|T)$  starts to perform at an arbitrary moment is

$$S(0|T) = K P(x) + \int_0^T S(0|T-x-y) dG(y) dF(x) + k \int_0^T R(0|T-y) dG(y) \tag{6.59}$$

where  $F(x) = 1 - P(x)$  is the distribution of a time of failure-free operation,  $G(x)$  is the distribution of a repair time, and  $S(0|T - Jt)$  is taken from (6.58). Expression (6.59) is correct with the additional condition  $S(jc|y < *) = 0$ .

**6.5.4 System with Admissible Down Time**

A system is considered to be successfully operating if during period  $T$  there will be no down time larger than  $T_1$ . This case in some sense is a "mirror" for that considered on page 249. We will omit the details and write the recurrent expression immediately:

$$Q(v|T) = P(T) + \int_0^T Q(v|T-x-y) dG(y) dF(x)$$

This expression is correct under an additional condition:

$$Q(v|x \wedge v) = 1$$

The same system may be considered in an on-duty regime. We again will omit the details and write the recurrent expression

$$Q(v|T) = K P(T) + \int_0^T Q(v|T-x-y) dG(y) dF(x) + k \int_0^T Q(v|T-y) dG(y)$$

This expression is correct under an additional condition:

$$Q(v|x < r, ) = 1$$

This subject as a whole requires a much more detailed discussion. There are many interesting detailed models concerning, for instance, computer systems. The reader who is interested in the subject can refer to Kredentser (1978), Cherkosov (1974), and Ushakov (1985, 1994). Some applications of these methods to oil and gas transportation systems can be found in Rudenko and Ushakov (1989).

## CONCLUSION

The models of repairable systems discussed in this chapter concern some ideal schemes: switches are supposed to be absolutely reliable; monitoring of the operation of the system's units is continuous; after repair, units are considered to be as good as new; and so forth. Besides, when using Markov models, one must assume that all distributions of failure-free intervals and repair times are exponentially distributed.

All of these assumptions seem to make such kinds of models practically useless. But the same can be said about *any* mathematical model: a mathematical model is only a reflection of a real object or real process. Each mathematical model may only be used if the researcher understands all of the model's limitations.

First of all, Markov models are very simple though simplicity is not a good excuse for their use. But using Markov models for highly reliable systems very often gives the desired practical results in reliability prediction.

Next, the lack of some realistic assumptions concerning switching and monitoring may be taken into account. (We try to show this in the next chapter.) This point is really very serious and must be taken into consideration. To demonstrate the importance of continuous monitoring of redundant units, let us consider a simple example.

A repairable system consists of  $n$  units in parallel (i.e., this is a group of one main and  $n - 1$  loaded redundant units). A system unit has an exponentially distributed TTF. Redundant units are checked only at the moment of failure of the main operating unit. At this moment all failed units are repaired instantaneously! If there is at least one nonfailed redundant unit, this unit replaces a failed main unit and the system continues to operate under its initial conditions. It seems that such a system with instant repair should be very reliable. But this system has no control over the system's unit states.

Find the MTTF of this system on the basis of simple explanations. A main unit has failed, on average, in  $T$  units of time, and with probability  $1/n$  up to this moment all of the remaining  $n - 1$  units have failed. It is clear that such

## 280 REPAIRABLE SYSTEMS

a system will work, on average,  $nT$  units of time until a failure. But as the reader will recall, a standby redundant group of  $n$  units without repair has the same MTTF!

It is difficult to find out who wrote the pioneering works in this area. The reader can find a review in the next chapter dedicated to renewal duplicated systems—a particular case of redundancy with repair. The reader can find general information on this question in a number of books on reliability, some of which are listed at the end of this book. For a brief review, we refer the reader to the *Handbook of Reliability Engineering* by Ushakov (1994).

Time redundancy represents a separate branch of renewal systems, closely related to the theory of inventory systems with continuous time. The reader can find many interesting models for reliability analysis of such systems in Cherkesov (1974) and Kredentser (1978). The reader can find applications of these methods to gas and oil pipelines with intermediate storage in Rudenko and Ushakov (1989). General methods of time redundancy are briefly presented in Ushakov (1985, 1994). An interesting discussion on repairable systems can be found in Ascher and Feingold (1984).

## REFERENCES

- Ascher, H., and H. Feingold (1984). *Repairable System's Reliability: Modelling, Inference, Misconceptions and Their Causes*. New York: Marcel Dekker.
- Cherkesov, G. N. (1974). *Reliability of Technical Systems with Time Redundancy*. Sovetskoe Radio: Moscow.
- Gertsbakh, I. B. (1984). Asymptotic methods in reliability theory: a review. *Adv. in Appl. Probab.*, vol 16.
- Gnedenko, B. V., Yu. K. Belyaev, and A. D. Solovyev (1969). *Mathematical Methods of Reliability Theory*. New York: Academic.
- Gnedenko, D. B., and A. D. Solovyev (1974). A general model for standby with renewal. *Engrg. Cybernet. (USA)*, vol. 12, no. 6.
- Gnedenko, D. B., and A. D. Solovyev (1975). Estimation of the reliability of complex renewable systems. *Engrg. Cybernet. (USA)*, vol. 13, no. 3.
- Kredentser, B. P. (1978). *Prediction of Reliability of Systems with Time Redundancy (in Russian)*. Kiev: Naukova Dumka.
- Rudenko, Yu. N., and I. A. Ushakov (1989). *Reliability of Energy Systems (in Russian)*. Novosibirsk: Nauka.
- Solovyev, A. D. (1972). Asymptotic distribution of the moment of first crossing of a high level by birth and death process. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.*, Issue 3.
- Ushakov, I. A., ed. (1985). *Reliability of Technical Systems: Handbook (in Russian)*. Moscow: Radio i Sviyaz.
- Ushakov, I. A., ed. (1994). *Handbook of Reliability Engineering*. New York: Wiley.

## EXERCISES

- 6.1 A system has an exponentially distributed TTF with a mean  $t = 100$  hours and a repair time having a general distribution  $G(t)$  with a mean  $r = 0.5$  hour. Find the system's stationary interval availability coefficient for the operation during 0.5 hour.
- 6.2 Construct a transition graph for a repairable system consisting of two main units, one loaded redundant unit which can replace instantaneously each of them, and three spare units. After a main unit has failed, a loaded redundant unit replaces it. In turn, one of the spare units replaces the redundant unit. Failed units are subjected to repair after which they become as good as new. All units are identical, each with a failure rate equal to  $\lambda$ . There are two repair facilities, each of which can repair only one failed unit at a time. The intensity of repair by a repair facility is equal to  $\mu$ . After a total exhaustion of all redundant units, repair is performed over the entire system with intensity  $M$ .
- 6.3 Construct a transition graph for the system depicted in Figure E6.2.

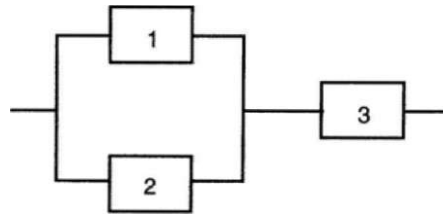


Figure E6.2. Structure diagram for the system described in Exercise 6.3.

## SOLUTIONS

- 6.1 The stationary availability coefficient depends only on the mean and not on the type of distribution of the TTF and repair time. Thus,  $K = (100)/(100 + 0.5) = 0.995$ . If the system is found within a failure-free interval, which is exponentially distributed, then the probability of successful operation of length  $t_0$  beginning at an arbitrary moment of time can be written as

$$\lim_{t \rightarrow \infty} P(t, t + t_0) = e^{-\lambda t_0}$$

Finally, after substituting the corresponding numerical data, one has

$$\exp(-0.5/100) = 0.995$$

and

$$R(t_0 = 0.5) = (0.995)(0.995) = 0.99$$

6.2 The solution is depicted in Figure E6.1

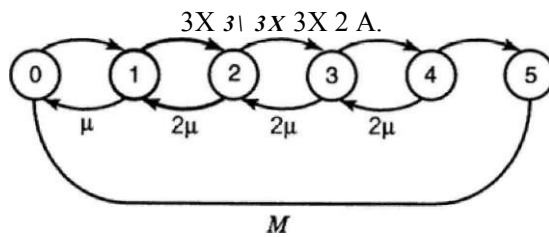


Figure E6.1. Transition graph for the system described in Exercise 6.2.

6.3 See Figure E6.3.



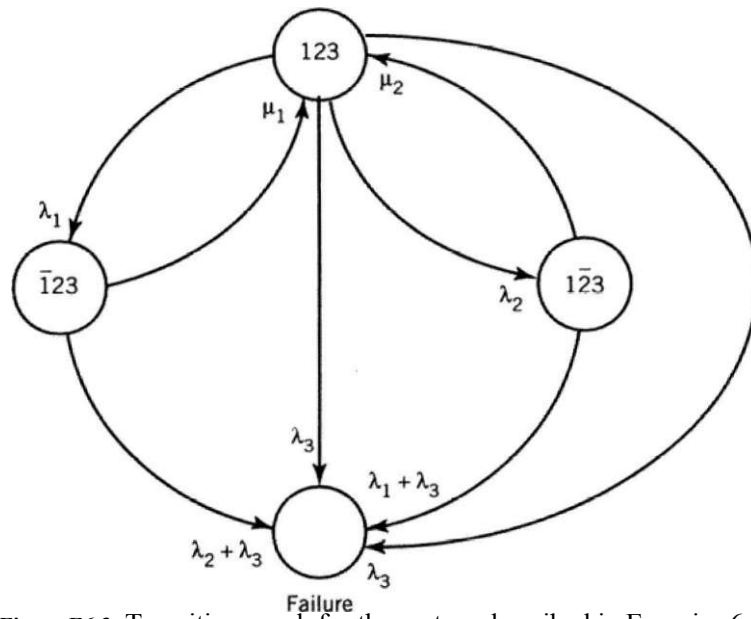


Figure E6.3. Transition graph for the system described in Exercise 6.3.

## CHAPTER 7

# REPAIRABLE DUPLICATED SYSTEM

Duplication refers to the particular case of redundancy where there is a single redundant unit to support a single working (main) unit. We distinguish this particular case for both practical and methodological reasons. First of all, when a designer feels that the reliability of some unit is low (sometimes this understanding may occur on a purely intuitive level), duplication is a simple way to improve it. Indeed, if a failure may occur with a relatively small probability, it is generally not necessary to have more than one redundant unit. In general, the number of redundant units depends on the desired value of the system's reliability index and/or on permissible economical expenditures.

From a methodological viewpoint, duplication presents the clearest way to explain certain special mathematical tools, their idiosyncrasies, and their ability to treat a real technical problem. It allows for the possibility of following mathematical transformations in detail. (Unfortunately, nobody has either the capacity or the desire to present similar detailed explanations for more complicated cases.)

### 7.1 MARKOV MODEL

As we have pointed out, a duplicated renewal system is one of the most frequently encountered structures in engineering practice. In the reliability analysis of electronic equipment (at least, in the first stages of design), the distributions of the time to failure and of the repair time are usually assumed exponential. In this case Markov models are adequate mathematical models

256

to describe such systems. We note that the final results obtained with Markov models are usually acceptable in a wide variety of practical cases (especially when applied to highly reliable systems).

### 7.1.1 Description of the Model

Consider a duplicated system consisting of two identical units. Usually, the following assumptions are made:

- The system units are mutually independent.
- After a failure of the operating unit, its functions are immediately assumed to be performed by the redundant unit.
- Repair (renewal) of a failed unit begins immediately.
- A repaired unit is considered to be a new unit.
- The switching device is considered absolutely reliable.

Two important aspects of a renewal system should also be taken into account: the regime of the redundant unit and the attributes of the repair workshop.

The following regimes of a redundant unit characterized by failure rate  $A'$  might be considered:

1. The redundant unit operates under the same conditions as an operational unit; that is, their failure rates are equal,  $A = A'$ .
2. The redundant unit is in a completely idle state, that is  $A' = 0$ .
3. The redundant unit is in an intermediate state between completely idle and operational, that is,  $0 < A' < A$ .

The first case is often referred to as *internal redundancy*, the second as *standby redundancy*, and the third as *waiting redundancy*.

The renewal regime might be distinguished by the number of repair facilities (places for repair, the number of technicians special equipment), that is, by the number of failed units which can be repaired simultaneously. We consider two cases:

1. An unrestricted renewal when the number of repair facilities equals the number of possible failed units (in this particular case, two facilities are enough).
2. An extremely restricted renewal with a single repair facility.

The transition graphs describing these models are presented in Figure 7.1 (there are two of them: with and without an absorbing state). Corresponding particular cases for different regimes of redundant units and different attributes of the repair shop are reflected in Figure 7.2.

Figure 7.1. Transition graph for a renewable duplicated system: (a) state 2 is reflecting; (b) state 2 is absorbing.

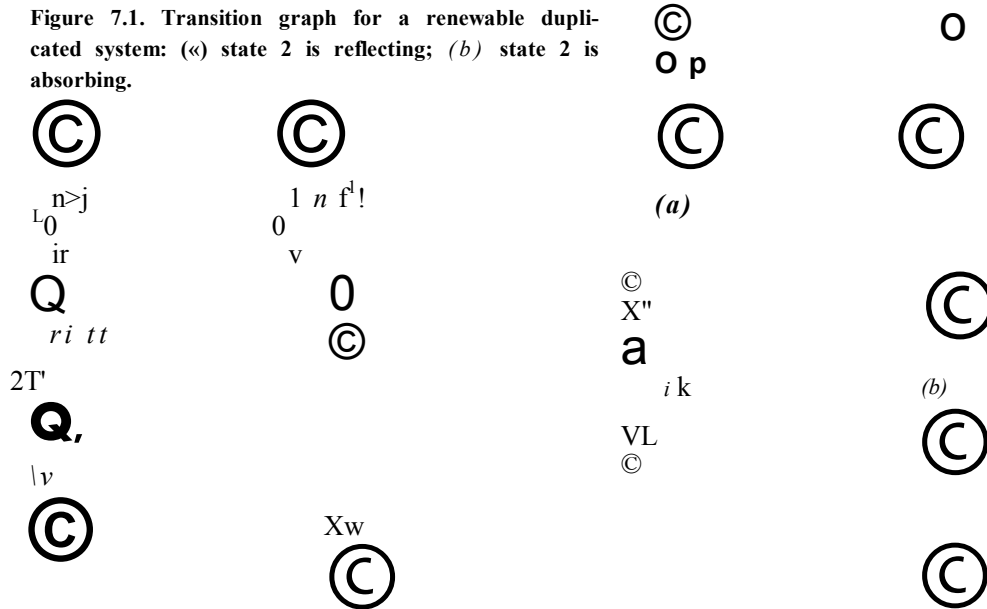


Figure 7.2. Transition graphs for four main models of a renewable duplicated system: (a) a loaded redundant unit, two repair facilities; (fc) a loaded redundant unit, one repair facility; (c) an unloaded redundant unit (spare unit), two repair facilities; (d) an unloaded redundant unit, one repair facility.

Using the above-described technique, corresponding systems of equations for obtaining the various reliability indexes can be easily written. In this particular simple case, the solutions can be obtained in a general form. The final results for particular cases can be derived easily.

### 7.1.2 Nonstationary Availability Coefficient

The system of differential equations (in canonical form) with the initial conditions  $P_a(0) = 0$  is

$$\frac{d}{dt} p_0(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t) + \mu_2 P_2(t) \quad (7.1)$$

$$P_0(0) = 1$$

The LST of (7,1) is

$$\begin{aligned}
 & (A_0 + j) < p_0(j) - M_i < P_i(s) - 1 \\
 & - A_0 < p_{,,}(s) + (5 + A, + M_i) < P_i(-0 + H_2 < P_2(s) = 0 \quad (7.2) \\
 & s < P_0(s) + 5 < p_{,,}(s) + s < P_z O = 1
 \end{aligned}$$

Notice that the availability coefficient equals

$$* < o = n c) + ^ ( 0 = 1 - ^ ( 0$$

Thus, to find the LST of  $K i t$ , we can find [see the last line in (7.2)]

$$\begin{aligned}
 & I \\
 & < P(i(s) + < P_i(s) = j \sim
 \end{aligned}$$

From (7.2) it is easy to write

$$\begin{aligned}
 & A_0 + \quad \quad \quad + f t i + s 0 \\
 & A_0 + 5 \quad - M I \quad 0 \\
 & \frac{A_n}{A_{,,}} \quad A_j + + s \quad \sim t^i \\
 & \quad \quad \quad s \quad 2 \\
 & \quad \quad \quad s \\
 & \quad \quad \quad A_0 A \\
 & \quad \quad \quad , \\
 & + s(A_0 + A_i + M_i + f t 2) + A_0 A_i + a_0 M_2 + M_1 M_2]
 \end{aligned}$$

Thus,

$$+ f i t^{-1} = 7 \sim \frac{f_2^2(s)}{s^2 + s(A_0 + A, + /a, + f i_2) + A_0 A [ + A + /i, /i_2]}$$

Now we should refer to the technique described in Chapter 6 in the section on Markov processes:

1. Represent the LST as the sum of simple fractions

$$\begin{aligned}
 & j^2 + i(A_{,,} + A_c + /i_t + n_2) + A_0 M_2 + M_1 A_2 \\
 & s[5^3 + s(A_0 + A_c + j u_{,,} + + A_0 A_t + \setminus_0 n_2 + \\
 & \quad \quad \quad A \quad B \quad C \\
 & 5 - 5 - 1 \quad s - s,
 \end{aligned}$$

where  $A$ ,  $B$ , and  $C$  are unknown.

2. Find the roots of the denominator. The first two roots of the denominator are conjugate, that is

$$s_{1,2} = -\frac{\alpha}{2} \pm \sqrt{\frac{\alpha^2}{4} - \beta}$$

where, in turn,

$$\begin{aligned} a &= A_0 + A, + \lambda L_x + \\ P &= A^2 \theta^2 l + A^2 \theta M^2 + M l^2 \end{aligned}$$

and  $s_3 = 0$ .

3. Find the unknown values  $A$ ,  $B$ , and  $C$  by equalizing the polynomial coefficients of the numerators.
4. Apply the inverse LST to obtain simple fractions with the numerators  $A$ ,  $B$ , and  $C$  found above to obtain the final result.

After these transformations the result is obtained

$$K(s) = 1 - \frac{A_0 A}{s^2 + \alpha s + \beta} \quad (7.3)$$

Obviously, if  $s = s_2$ , l'Hospital's rule must be used.

Now any result of interest can be obtained by substituting the appropriate values of  $A$  and  $\lambda$ . In general, the solution for a duplicate renewal system can be obtained in a closed form, but this solution is not very compact, even for the simplest case.

Of course, we should notice that, for active redundancy and unrestricted repair, the final result can be written immediately with the use of the appropriate result for a single unit:

$$K(s) = \frac{1 - K^*(f)}{s + \lambda A}$$

where  $K^*(t)$  is the nonstationary availability coefficient of a single unit. This result is obvious because both units are supposed to be mutually independent.

### 7.1.3 Stationary Availability Coefficient

The solution can be derived from (7.1) by putting the derivatives equal to 0. The same result can be directly obtained from the corresponding transition

graph by writing the equilibrium equations:

$$\begin{aligned}
 & \sim l_0 P_0 + - 0 \\
 Aq/q - (A, -Jtt,)?! + / * _2 P_2 = 0 & \qquad (7.4) \\
 + + ^ P I = 9 &
 \end{aligned}$$

The solution is

$$\begin{aligned}
& K = I - P_2^i = I - \frac{A_0 - (A_0 A_t + A_0^2 + M_1 M_2)}{A_0 + A_t + M_2} \\
& \quad = 1 - \frac{A_0 A_t + A_0^2 + M_1 M_2}{A_0 + A_t + M_2} \\
& \quad = \frac{A_0 + A_t + M_2 - A_0 A_t - A_0^2 - M_1 M_2}{A_0 + A_t + M_2} \\
& \quad = \frac{A_0(1 - A_t) + A_t(1 - A_0) + M_2(1 - M_1)}{A_0 + A_t + M_2} \quad (7.5)
\end{aligned}$$

To obtain values of this reliability index for different cases, the specific A's and /A's should be substituted. The results for the four most important cases depicted in Figure 7.2 are presented in Table 7.1. In this table we used the notation  $y = \frac{A}{I}$ . For highly reliable systems with  $A \ll I$ , all of the expressions in Table 7.1 can be easily transformed to obtain the approximations given in Table 7.2. The expressions in Table 7.2 allow one to give an understandable explanation of all of the effects. Naturally, the worst value of K gives the case of active redundancy and restricted repair (the failure rate

**TABLE 7.1 Availability Coefficient for Four Main Models of a Renewable Duplicated System**

	(a)	(b)
(A)	$1 + \frac{y}{I + 2y}$	$1 + \frac{y}{I + 4y}$
(B)	$1 + \frac{y^2}{2(1 + y)}$	$1 + y$

(A) Loaded redundant unit; (B) unloaded redundant unit; (a) two repair facilities; (b) one repair facility.



**TABLE 7.2 Approximation for Availability Coefficient for Highly Reliable Duplicated System**

	(a)	(b)
(A)	$I + 2y$	$1 + 4-y$
(B)	$I \sqrt{\frac{v^2}{2(1+y)}}$	$\frac{y^2}{I}$

(A) Loaded redundant unit; (B) unloaded redundant unit; (a) two repair facilities; (b) one repair facility.

$A_0$  is the largest and the repair rate  $p_2$  is the smallest). The case of unrestricted repair yields a mean repair time of less than one-half of the restricted repair time ( $1/2/a$  and  $\sqrt{fi}$ , respectively). Below in this section we will show that the MTTF of highly reliable systems of active redundancy is one-half of the MTBF for standby redundancy. Of course, for two independent units, that is, when the redundancy is active and the repair is unrestricted, we can write

$$K = i - (i - K * y = i - A + p$$

using the availability coefficient for the single unit

$K^*$ . Then

$$\frac{2A + n^2}{A^2 + 2A + \lambda} \left\{ 1 + \frac{1}{2A + M_2} \right\} = \frac{1}{1 + I + 2y}$$

The intermediate case with either the "underloaded" redundant unit (when  $A < A_0 < 2A$ ) or with the "dependent" repair when ( $1/a < p_2 <$  can be easily obtained numerically from the general expression (7.5). Of course, this index can be realized as

$$\lim_t K(t)$$

but this is not effective when  $K(t)$  is not available.

#### 7.1.4 Probability of Failure-Free Operation

## 292 REPAIRABLE DUPLICATED SYSTEM

To find this probability, it is necessary to construct a system of differential equations using the graph of Figure 7.1b with absorbing state 2. In this case the equations are not linearly dependent. For the initial conditions  $P_{0i}(0) = 1$ ,

the system of linear differential equations is

$$\begin{aligned}
 -p_0'(t) &= -\lambda_0 P_0(t) \\
 \frac{d}{dt} P_1(t) &= \lambda_0 P_0(t) - (\lambda_1 + \mu_1) P_1(t) \\
 P_0(0) &= I
 \end{aligned}
 \tag{7.6}$$

The LST of (7.6) is

$$\begin{aligned}
 (\lambda_0 + s) p_0(s) - \lambda_0 &= 1 \\
 -\lambda_0 p_0(s) + (\lambda_1 + \mu_1 + s) p_1(s) &= 0
 \end{aligned}
 \tag{7.7}$$

and the solution has the form

$$\langle P_0(s) \rangle = \frac{\begin{vmatrix} 1 & \lambda_0 \\ \lambda_0 & \lambda_1 + \mu_1 + s \end{vmatrix}}{\begin{vmatrix} \lambda_0 + s & \lambda_0 \\ -\lambda_0 & \lambda_1 + \mu_1 + s \end{vmatrix}}$$

$$\tag{7.8}$$

Applying the procedure that we used to obtain (7.3), we find

$$\langle P_0(s) \rangle = \frac{1}{s^2 + s(\lambda_0 + \lambda_1 + \mu_1) + \lambda_0 \lambda_1}$$

$$\tag{7-9}$$

where the superscript (0) stands for the initial conditions  $P_0(0) = I$  and also

$\langle P_0(s) \rangle$

$-P$

$$\begin{aligned}
 a^* &= \lambda_0 + \lambda_1 + \mu_1 \\
 \beta^* &= \lambda_0 \lambda_1
 \end{aligned}$$

If we are interested in the system's PFFO immediately after its repair, it is necessary to set  $P_1(0) = 1$ . The corresponding system of linear algebraic equations in the LST is

294 REPAIRABLE DUPLICATED SYSTEM

$$\begin{aligned} (\Lambda_0 + s)P_0(s) - M_1P_1(s) &= 0 \\ -\Lambda_0P_0(s) + (\Lambda_1 + M_1 + s)P_1(s) &= 1 \end{aligned}$$

$$\begin{array}{c}
 \begin{array}{|cc|c|}
 \hline
 0 & -M_1 & \\
 1 & A_0 + M_1 + J & \\
 \hline
 \end{array}
 +
 \begin{array}{|c|}
 \hline
 A_0 \\
 -A_0 \ 1 \\
 \hline
 \end{array}
 \\
 \\
 \begin{array}{|c|}
 \hline
 A_0 + S - M_1 \\
 -A_0 A_0 + M_1 + S \\
 \hline
 \end{array}
 \\
 A_0 + H_1
 \end{array}$$

$$s^2 + s(A_0 + A_0 + j\omega - j) + A_0 A_0,$$

Notice that the denominator in (7.10) is the same as in (7.8), so we can use the roots (eigenvalues) obtained above. Omitting routine transformations, we may write the final result for this case (7.10)

$$P_0^{(1)} = \tau^{\wedge} [(*) - A_0 - A_0]^{**} - (4 - A_0 - A_0 K_f] \quad (7.11)$$

where the superscript (1) indicates the corresponding initial conditions.

### 7.1.5 Stationary Coefficient of interval Availability

The task can be solved by setting the initial conditions:  $P_0(0) = P_0$  and  $P_1(0) = P_1$ , in (7.6) where  $P_0$  and  $P_1$  are the stationary probabilities obtained from (7.4). The  $p$ 's can be found from (7.4) separately as

$$P_n = \begin{array}{c}
 \begin{array}{|ccc|}
 \hline
 0 & M_1 & 0 \\
 0 & -(A_0 + M_1) & M_2 \\
 1 & 1 & 1 \\
 \hline
 \end{array}
 \\
 \\
 \begin{array}{|ccc|c}
 \hline
 A_0 & M_1 & 0 & A_0 A_0 + \frac{M_1 M_2}{A_0 M_1 + M_1 M_2} \\
 A_0 & -(A_0 + M_1) & M_2 & \\
 1 & 1 & 1 & \\
 \hline
 \end{array}
 \end{array}$$

and

$$P_1 = \begin{array}{c}
 \begin{array}{|cc|}
 \hline
 -A_0 & 0 \\
 A_0 & M_2 \\
 1 & 1 \\
 \hline
 \end{array}
 \\
 \\
 \begin{array}{|ccc|c}
 \hline
 -A_0 & M_1 & 0 & A_0 A_0 - A_0 M_2 + M_1 M_2 \\
 A_0 & -(A_0 + M_1) & M_2 & \\
 1 & 1 & 1 & \\
 \hline
 \end{array}
 \end{array}$$

and the solution

$$\begin{array}{|ccc|c}
 \hline
 -A_0 & M_1 & 0 & A_0 A_0 - A_0 M_2 + M_1 M_2 \\
 A_0 & -(A_0 + M_1) & M_2 & \\
 1 & 1 & 1 & \\
 \hline
 \end{array}$$



where  $\theta_1$  and  $\theta_2$  are values such that the  $MTTF = \theta_1 + \theta_2$ . The solution of this equation system yields

$$\theta_n = \frac{1}{A} \text{ and } \theta_j = \frac{M_j}{A_0 A_j}$$

Of course, this result coincides with (7.12).

The MTBF may be computed in the same manner, and we leave this as an exercise.

298 REPAIRABLE DUPLICATED SYSTEM

To find the system's MTTF, it is sometimes more convenient to use the following arguments. Consider the transition graph in Figure 7.1. Let us find the system's MTTF (this means that at  $t = 0$  the system is in state 0). Denote the mean time needed to reach the absorbing state 2 from the initial state 0 as  $T_{02}$  and from state 1 as  $T_{12}$  (Apropos,  $T_{02}$  = MTTF and  $T_{12}$ , « MTBF.) Obviously,

$$T_{02} = \frac{1}{\lambda_0 + \lambda_1} + T_{12} \tag{7.13}$$

because the process inevitably moves from state 0 to state 1, After this, based on the Markov property, the process can be considered to be starting from state 1.

The process stays in state 1 for an average time  $1/(\lambda_1 + \mu_1)$  and then moves either to state 2 or to state 0, It moves to state 0 with probability  $\mu_1/(\lambda_1 + \mu_1)$  and then starts traveling again from state 0. Hence, we can write

$$T_{02} = \frac{1}{\lambda_0 + \lambda_1} + \frac{\mu_1}{\lambda_1 + \mu_1} T_{02} \tag{7.14}$$

Substituting (7.14) into (7.13) yields

$$\begin{aligned} \text{MTTF} - T_{02} &= \frac{1}{\lambda_0 + \lambda_1} + \frac{\mu_1}{\lambda_1 + \mu_1} T_{02} \\ &= \frac{1}{\lambda_0 + \lambda_1} + \frac{\mu_1}{\lambda_1 + \mu_1} \left( \frac{1}{\lambda_0 + \lambda_1} + \frac{\mu_1}{\lambda_1 + \mu_1} T_{02} \right) \end{aligned}$$

From (7.13) it also immediately follows that

$$\text{MTBF} = r_{12} = \frac{1}{\lambda_1 + \mu_1} + \frac{\lambda_1}{\lambda_1 + \mu_1} T_{02}$$

Now, on a very understandable and almost verbal level, we can explain the difference between the MTBF (or the MTTF) for repaired duplicated systems of identical units which have a different regime for the redundant unit. For active redundancy  $A_u = 2A$  and for standby redundancy  $A_0 = A$ . In other words, in the first case, the system stays in state 0, on average, one-half the time that it stays in the second case. This fact can be seen more clearly from



the approximate expressions for a highly reliable system when  $y = A/x \ll 1$ :

$$MTTF_{active} = T + \frac{1}{\lambda} + \frac{1}{2\lambda} \frac{1}{1 + y} \quad (7-15)$$

and

$$MTTF_{total} = T + \frac{1}{\lambda} + \frac{1}{2\lambda} \frac{1}{1 + y} \quad (7-16)$$

Incidentally, (7.15) and (7.16) could be explained on the basis of the Renyi theorem. Consider an alternative renewal process describing the operation of a repaired duplicated system.

For a highly reliable system, this process can be approximately represented as a simple renewal process if one neglects small intervals of being at state 1. The system's successful operation period consists of the sum of a random number of intervals of the length  $1/A_0$  until the process has jumped to state 2. This random number has a geometrical distribution with parameter  $p = \lambda_i / (\lambda_i + \lambda_t) \sim 1 \sim T$ . Thus, the sum also has an exponential distribution with parameter  $A_0 y$ . This means that approximately

$$P(t) \approx \lambda_0 y \exp(-A_0 y t) \quad (7.17)$$

We should now remember that for active redundancy  $A_y = 2A$  and for standby redundancy  $A_0 = A$ .

We wrote all of these solutions in such a detailed form because the LST technique is very important in engineering applications. A certain amount of practice is needed to apply this to practical problem solutions. We believe that the best way to master these approaches is to work out simple exercises.

## 7.2 DUPLICATION WITH AN ARBITRARY REPAIR TIME

For repairable duplicated systems, models more complicated than the Markovian type can be analyzed. We first consider a model described in the following way:

---

10 Both units are independent and identical.

\* The operating unit has an exponential distribution of time to failure  $F(t)$  with parameter  $\lambda$ , and the redundant unit has a similar distribution  $F_r(t)$ , also exponential with parameter  $\lambda_r$ ,  $0 < \lambda_r < \lambda$ . (This condition means that the redundant unit might be, in general, in an underloaded regime.)

### 300 REPAIRABLE DUPLICATED SYSTEM

- \* The repair time of a failed unit has an arbitrary distribution  $G(t)$ .
- ' The repair of a failed unit begins immediately after a failure has occurred.
- \* After repair, the unit becomes completely new.
- \* The repaired unit is immediately installed into the system.

It is clear that after an operational unit has failed, the redundant unit replaces it and becomes operational. A system failure occurs if and only if the operating unit fails during the repair of the other unit, that is, when both of the system's units have failed.

Let us find the distribution of the system's time to failure  $R_s(t)$ . A failure-free operation of the duplicated system during a time period  $t$  can be represented as the union of the following mutually disjoint events:

1. The first failure in the system occurs after moment what happens with the redundant unit does not play any role. The probability of this event is  $\exp[-(A_1 + A_2)t]$ .
2. The first failure of either of the two units occurs at some moment  $z < t$ , the failed unit is not repaired during the interval  $(t - z)$ , but the unit in the operating position has not failed up to  $t$ . The probability of this event is

$$\int_0^t (A_1 + A_2) e^{-(A_1 + A_2)z} [1 - G(t - z)] e^{-A_1(t-z)} dz$$

3. The last event is the most complicated. In this case, as some moment  $x < t$ , the duplicated system comes to the initial state, that is, state 0, where both system's units are operational. This occurs if one of the units has failed during the interval  $[z, z + dz]$ , the repair has taken time  $x - z$ , and the operating unit has not failed during repair. After the completion of the repair, the system operates successfully during the remaining period of time  $(t - x)$  with probability  $R_s(t - x) \Rightarrow 1 - F_s(t - x)$ . The probability of this event is

$$\int_0^t \int_0^{t-x} R_s(t-x) dx [g(x-z)] e^{-A_1(x-z)} dz$$

where  $g(t)$  is the density function of the distribution  $G(t)$ .

Now it is easy to write the final equation for the probability of a system's failure-free operation:

$$R(t) = \int_0^t f R(t-x) e^{-\lambda x} dx + e^{-\lambda t} [1 - G(t)] \quad (7.18)$$

Thus, we have an integral equation with a kernel of the type

$$R(t) = A(t) + \int_0^t f R(t-x) B(x) dx \quad (7.19)$$

where, in the above case,

$$A(t) = \int_0^t f e^{-\lambda x} [1 - G(t-x)] dx$$

$$B(t) = \int_0^t e^{-\lambda z} g(t-z) dz \quad (7.20)$$

The recurrent equation (7.18) can be solved by the method of sequential iterations. But we prefer to obtain the solution in the form of the LST as it allows us to investigate the asymptotical behavior of  $R(t)$ .

If we denote

$$a(s) = \int_0^\infty e^{-s t} A(t) dt$$

$$b(s) = \int_0^\infty e^{-s t} B(t) dt$$

$$P(s) = \int_0^\infty e^{-s t} dG(t)$$

$$p(s) = \int_0^\infty e^{-s t} R(t) dt$$

then the solution can be represented in the form

$$R(s) = \frac{a(s)}{1 - f P(s)} \quad (7.21)$$

and, finally, the LST of interest is

$$a(s)$$

302 REPAIRABLE DUPLICATED SYSTEM

The functions  $a(s)$  and  $b(s)$  can easily be found from (7.20)

$$\frac{5 + A + (A + A_s)[1 - ip(s + A)]}{(s + A + A_s)(i + A)} \tag{7.23}$$

$$b(s) = \frac{\dots}{A + A[ + s]}$$

Thus, after substituting (7.23) into (7.22), we obtain

$$= \frac{5 + A + (A + A_s)[1 - \wedge(J + A)]}{(i + A)[i + (A + A_s)(1 - \wedge(A + j))]} \tag{7.24}$$

Therefore, the general case has been investigated. It is clear that for active redundancy, when  $A_s = A$ ,

$$\frac{5 + A + 2A[I - \Psi(A + J)]}{(A + s)[s + 2A(1 - \Psi(A + j))]}$$

For standby redundancy, when  $A_t = 0$ ,

$$\frac{s + A + A[1 - \wedge(A + 5)]}{(A + s)[s + A(1 - \Psi(A + j))]}$$

Since (7.24) is the LST of  $R(t) = 1 - F_s(t)$ , the MTTF can be derived from this expression directly with the substitution  $s = 0$ :

$$\frac{r - / M \frac{A + (A + A_s)[1 - \wedge(A)]}{1}}{T_s = \Psi > (5) 1, 0 = - 7 T_s, r r r, \dots} = - + \frac{1}{A(A + A_s)[1 - *(A)] A (A + \wedge[1 "**(*)]} \tag{7.25}$$

In deriving the latter expression, we use the memoryless property of the exponential distribution: if an object with an exponentially distributed random TTF has not failed until some moment  $t$ , then the conditional probability of the random residual TTF of the object is the same exponential distribution as the original one. Thus, the process of an operation of a duplicated system has the so-called *renewal moments*, that is, such Markov moments at which all of the prehistory of the process has no influence on the future development of the process starting from this moment.

The MTTF of the duplicated system without repair, as it was obtained above, equals

$$T = \frac{1}{A + A} + \frac{1}{A}$$

From (7.25) it follows that the effectiveness of redundancy with renewal increases very quickly as  $i/t(A) \rightarrow 1$ . Notice that  $0(A)$  is not more than the probability that the random TTF of a unit exceeds the duration of its repair. It means that

$$a = 1 - \int_0^\infty [1 - e^{-kt}] dG(t) \tag{7.26}$$

is the probability of an unsuccessful repair, denoted by  $a$ . Here  $\xi$  is a random

TTF and  $\tau$  is a random repair time. Notice that for the exponential distribution this probability equals  $A/(A + p) = y/(1 + y)$  where  $y = A/\lambda$ .

Let us make a final remark concerning the system MTTF. It is possible to write a clear and understandable recurrent equation to express  $T$ . The period of a system's successful operation can be represented by a sequence of cycles of the type "time to failure of any of the system's units + time of successful repair" which terminates with a system's failure (the cycle with an unsuccessful repair). Arguing similarly as in Section 7.1, the recurrent relationship can be written as

$$T = aT + (1-a)T_s \quad T = \frac{1}{A + A}$$

Finally, for  $T$ , one obtains

$$T = \frac{1}{(A + A)a} + \frac{1}{A}$$

It is clear that the MTTF of the duplicated repairable system depends on the distribution  $G(t)$ . Let us investigate this relationship in more detail. From (7.26) we can derive

$$\begin{aligned} & \int_0^\infty (1 - L) dG(t) \quad ("AO \\ & = \int_0^\infty t dG(t) \sim \frac{1}{2} \int_0^\infty t^2 dG(t) + \frac{1}{6 \cdot 0} \int_0^\infty t^3 dG(t) \end{aligned}$$

304 REPAIRABLE DUPLICATED SYSTEM

$$= A E\{T\} - |E\{T, \rho\} + \text{ME}\{TJ^3\} + \blacksquare \cdot \blacksquare$$

Such a representation is very useful if the system is highly reliable, that is, when  $\lambda T \ll 1$ . Then the following approximation is true:

$$A \approx \lambda E\{T\} - \lambda^2 E\{T^2\} = \lambda T - [\lambda^2 R^2 + \lambda^2 \text{Var}\{r\}]$$

where  $r = E\{r\}$ . Thus, between several distributions  $G(t)$  with the same mean, the probability  $a$  is smaller if the variance of the repair time is larger. From this statement it follows that the best repair is characterized by a (practically) constant duration.

Let us give several simple examples.

**Example 7.1** Find  $a$  when the repair time is constant,  $T = \tau$ . By direct calculations

$$a = \Pr\{\lambda \tau < T\} = 1 - e^{-\lambda \tau}$$

and, for the highly reliable system when  $\lambda T = \lambda \tau \ll 1$ ,

$$a \approx \lambda T - \frac{1}{2}(\lambda T)^2$$

**Example 7.2** Find  $a$  when the repair time distribution is exponential. By direct calculations

$$a = \Pr\{\lambda r < T\} = \frac{\lambda}{\lambda + \lambda} \ll 1 \quad \frac{\lambda}{\lambda + \lambda} = 1 - \frac{1}{1 + \lambda T}$$

approximately,

$$A \approx \lambda T - (\lambda T)^2$$

**Example 7.3** Find  $a$  when the repair time distribution is normal with mean equal to  $r$  and variance equal to  $\sigma^2$ . Find the LST for this distribution, using essentially the same technique that we applied in Section 1.3.3 for obtaining the m.g.f.

$$\lambda(s)$$

We remind the reader that the LST and the m.g.f. differ only by the sign of the argument  $s$ . Thus,

$$a = 1 - p(\lambda) = 1$$

An approximation has the form

$$A \sim \lambda T + \frac{1}{2} \lambda^2 \sigma^2$$

From these examples we see that with  $A_T \rightarrow 0$  the repair (renewal) effectiveness becomes higher and tends to be invariable with respect to the type of  $G(t)$ . This is true for most reasonable practical applications. For example, to replace a failed bulb may take some 10 seconds, but its lifetime may equal hundreds of hours; to change or even to repair a car's tire takes a dozen minutes which is incomparably less than its average lifetime. This fact leads to new methods of investigation, namely, to asymptotic methods.

Assume that the parameters  $A$  and  $A_i$  of the model are fixed and then consider a sequence of repair time distributions  $G_1, G_2, \dots, G_n, \dots$ , which changes in such a way that

$$\alpha_n = \int_0^\infty [1 - G_n(t)] dF(t) \rightarrow 0 \tag{7.27}$$

This means that the probability that the operational unit fails during repair goes to 0,

Under this condition the appearance of some limit distribution of a system's failure-free operation is expected. Of course, if  $a_M \rightarrow 0$  the system's MTTF goes to  $\infty$ . To avoid this, we must consider a normalized random TTF, namely,  $a\xi$ . It is clear that this new r.v. has a constant mean equal to 1 independent of the value of  $a$ . The distribution of this r.v. is

$$Pr\left\{ \frac{\xi}{R} > t \right\} = \left( \frac{t}{\alpha} \right)$$

The LST of this function is from (7.24)

$$\int_0^\infty e^{-st} R \left( \frac{t}{\alpha} \right) dt = \alpha \bar{\varphi}(\alpha s) = a \frac{as + A + (A + A_i)[1 - \exp(-as + A)]}{(ai + A)[as + (A + A_j)(1 - \exp(-as + A))]} \tag{7.28}$$

Now under the assumption that  $a, \rightarrow 0$ , we can write

$$*(A) - \exp(-as + A) = \int_0^\infty e^{-As} (1 - e^{-as'}) dG(t)$$

$as$   $as$



STANDBY  
REDUNDANCY  
WITH  $I_{te}$   
ARBITRARY  
DISTRIBUTION  
S

307

## REPAIRABLE

That is,

## DUPLICATE

## D SYSTEM

$$Y(A) - \frac{a^2 s}{A} + A = 0$$

where  $0 < \rho < 1$ . Therefore, if  $\rho \ll 0$  uniformly on any finite interval of the domain of  $s$ ,

$$a s(a s) = a - \frac{a s + A + (A + A_s) | a + \frac{a^2 s}{A}}{(a s + A) a s + (A + A_s) | a + \frac{a^2 s}{A}} \quad (7.29)$$

Because the LST (7.29) corresponds to an exponential d.f., we get the following asymptotic result:

$$\lim_{t \rightarrow \infty} \Pr\{a t > t\} = e^{-\rho t}$$

For practical problems, this means that for a small value of  $\rho$  the following approximate expression can be used:

$$R(t) \approx e^{-\rho t} \quad (7.30)$$

$$R(t) \approx e^{-T t} \quad (7.31)$$

where  $T$  has been defined in (7.25). Incidentally, (7.31) is more accurate than (7.30): it has an error of order  $\rho$  in comparison with  $\rho^2$  associated with the latter.

This can be explained in the following "half-verbal" terms. The random time to failure of a duplicated system consists of a random number of cycles of the type "totally operational system — successful repair" and a final cycle of the type "totally operational system - unsuccessful repair." Stochastically, all cycles of the first type are identical and, by the assumption of the exponentiality of the distribution, are mutually independent (the latter assumption is based on the memoryless property). The only cycle differing from these is the last one. But if to suggest that the number of cycles of the first

type is large, on average, the distribution of the system time to failure can be approximated by the exponential distribution.

The use of the approximations (7.30) and (7.31) requires the value of  $a$ . This value can be obtained easily in this case. Moreover, if we know that

### 310 REPAIRABLE DUPLICATED SYSTEM

$G(t)$  has a restricted variance, then in limit  $a \ll A\tau$ . (Of course, the condition  $\tau \rightarrow 0$  is necessary.) In turn, this means that the following can be obtained:

$$R(t) \sim \dots \quad (7.32)$$

If the conditions for the variance of  $G(t)$  do not hold, the latter expression, of course, is wrong. The reader can verify this with an example of a sequence of two-mass discrete distributions  $G(t), G_2(t), G_n(t)$ , each with two nonzero values of probability: at 0 and at some positive point. Let all of the distributions have the same mean but, with increasing  $n$ , the probability at the positive point becomes smaller with moving of the point to the right along the axis. The variance in this case is infinitely increasing with increasing  $n$ .

### 7.3 STANDBY REDUNDANCY WITH ARBITRARY DISTRIBUTIONS

For standby redundancy the results can be obtained for the most general case, namely, when both distributions—of a random TTF and of a random repair time—are arbitrary. Let us use the same notation for the distributions:  $F(t)$  and  $G(t)$ . The duplicated system's operation can be graphically represented in Figure 7.3.

The system's operation consists of the following random intervals. The first one endures until the main unit fails; its random length is  $\xi_0$ . The second interval and all of the remaining intervals,  $k = 2, 3, \dots$ , are successful if and only if each time a random failure-free time of the operating unit ( $\xi_k$ ) is longer than the corresponding random repair time of the failed unit  $\eta_k$ . The last interval when a system failure has occurred has a random duration different from all of the previous ones: this is the distribution of the random TTF under the condition that  $\xi < \tau$ . All of these explanations become transparent if one considers a constant repair time: the first failure-free interval has unconditional distribution of the r.v. all of the remaining intervals (except the last one) have a conditional distribution under the condition that  $\xi > \tau$ ; and the last one has a conditional distribution under the condition that  $\xi < \tau$ . In other words, the first of these distributions is positively biased and another is truncated from the right.

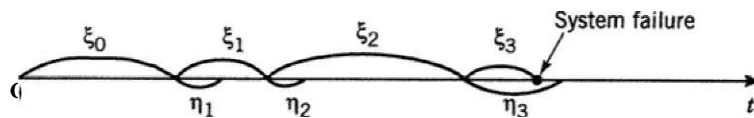


Figure 7.3. Time diagram for duplicated system operation with a standby redundant unit.

Let  $\xi^*$  denote a random value representing the system's time to failure starting from the moment  $\xi_0 + g_k$ . Here  $k$  is the number of the last cycle when a system failure has occurred. For the distribution of the r.v. the following recurrent equation can be easily written:

$$1 - F^*(t) = 1 - F(t) - \int_0^t [1 - F^*(t - x)]G(x) dF(x) \quad (7.33)$$

The first term of the sum reflects the fact that during time  $t$  (no failure occurs). The expression under the integral means that the first failure occurs in the interval  $x + dx$ , but the repair of the failed unit has been completed up to this moment, and from this moment on the system is in the same state as in the previous moment at moment  $\xi_0$ . Thus, this is the regeneration moment for the renewal process under consideration.

The final goal is to find the distribution of the random value  $\xi_0 + g^*$  and to express the probability of the system's successful operation  $P(t)$ :

$$P(t) = 1 - Q(t) = 1 - \int_0^t F^*(t - x) dF(x) \quad (7.34)$$

The numerical solution of (7.33) and (7.34) can be obtained by sequential iteration. But again we will use the LST which is useful for future asymptotic analysis.

Introduce the following notation:

$$\begin{aligned} \Phi(s) &= \int_0^\infty e^{-st} dF(t) \\ \psi(s) &= \int_0^\infty e^{-st} G(t) dF(t) \\ \langle p(s) \rangle &= \int_0^\infty f dF^*(t) \\ \langle p(s) \rangle &= \int_n C e^{-dQit} \end{aligned}$$

Then the LST of (7.33) can be written as

$$\langle p(s) \rangle = \langle D(i) \rangle + \langle A(s) \rangle [1 - \langle p(s) \rangle]$$

and,

finally,

$$\langle p(s) \rangle = \frac{\langle D(i) \rangle}{1 - \langle A(s) \rangle} \quad (7.35)$$

Combining (7.35) with (7.34), we get

$$\langle p(s) \rangle_{\mathbb{R}(j)} = \mathcal{L}(s) \mathcal{L}(s) - \frac{\mathcal{L}(0)}{H f} - i^{-*}(\cdot) \quad (7.36)$$

From this LST, the system's MTTF can be found by setting  $s = 0$ . But in this case we prefer a more direct way:

$$= E\{I_0\} + E(\xi \wedge \dots - R + \dots)$$

where  $\nu$  is the random number of cycles which has a geometric distribution with parameter  $a$ :

$$a - A \int_0^{\infty} G(r) 1 < F < 0$$

which is small in practical cases.

Let us investigate the asymptotic behavior of  $\langle pit \rangle$ . Suppose that  $F(t)$  is fixed and the distribution of the repair time changes by some sequence in such a way that

$$\langle \dots \rangle = A \int_0^{\infty} (0 \int dF(t) \wedge 0$$

Let us introduce the corresponding distributions and LSTs:  $Q_n(t)$ ,  $\langle p_n(s) \rangle$ ,  $i \int \dots (s)$  and, additionally,  $x_n(s)$ :

$$x_n(s) = \langle J \rangle_n^* - \langle \dots \rangle^* - \int_0^{\infty} [1 - G_n(t)] dF(t)$$

Now we evaluate the difference

$$\begin{aligned} - x_n(a_n s) &= \int_0^{\infty} [1 - G_n(t)] dF(t) \\ &< a_n \int_0^{\infty} [1 - G_n(t)] dF(t) \\ &= \int_0^{\infty} [1 - G_n(t)] dF(t) + \int_0^{\infty} t dF_c(t) \\ &\mathcal{L} a_n s \end{aligned}$$

If in this inequality we let

$$C_n =$$

then both terms in the last set of square brackets go to 0. This leads to the statement:

$$\lim_{i \rightarrow \infty} \dots = 1$$

and the limit is uniformly exceeded on any finite area of domain of  $s$ . Now the normalized random variable  $a_n r$  is considered. The d.f. of this r.v. is

$$\Pr\{A_n, T < r\} = Q_n$$

The LST of this d.f. is

For  $a_n \rightarrow 0$ , from (7.36) it follows that

$$\begin{aligned} <p_n\{a_n s\} &= 4 >(a_n O \\ &1 - <!>(«,,\$) \\ &x_n(<X_n s) \\ &a \\ &\underline{1 - <j>(a_n, Q_+} \quad i + sT \end{aligned}$$

and the limit is uniformly exceeded on any finite area of domain of  $s$ . Consequently,

$$\lim_{n \rightarrow \infty} \Pr\{d_n r < />\} = !- e^{-nT} \tag{7.37}$$

From (7.37) it follows that for a small value of  $a$  the approximation

$$nO = \Pr\{\xi_{\text{sys}} > r\}$$

is true.

### 7.4 METHOD OF INTRODUCING FICTITIOUS STATES

As we considered in Chapter 1, some combinations of exponential distributions can produce distributions with both increasing and decreasing intensity

314 REPAIRABLE DUPLICATED SYSTEM

functions, or failure rates. This fact leads to the idea of an approximation of some arbitrary distributions. We will show that such an approximation can allow us to reduce semi-Markov processes to Markov processes.

A mixture of exponential distributions with different parameters leads to a distribution which has the decreasing intensity function

$$\begin{aligned}
 F(t) &= \sum_{i=1}^n P_i e^{-\lambda_i t} = \int_0^{\infty} P(u) e^{-\lambda u} e^{-\lambda t} du \\
 & \quad \lambda_i > 0, \sum_{i=1}^n P_i = 1
 \end{aligned}
 \tag{7-38}$$

A convolution of  $n$  identical exponential distributions  $e(t) = \exp(-\lambda t)$  leads to an Erlang distribution of the  $n$ th order which can be expressed in the following recurrent way:

$$F_n(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}
 \tag{7.39}$$

If there are exponential functions with different parameters  $e_k(t) = \exp(-\lambda_k t)$ , then the generalized Erlang d.f. holds

$$F_n(t) = \int_0^t e_1(x) e_2(t-x) \dots e_n(t-x) dx
 \tag{7.40}$$

Both the Erlang and the generalized Erlang distributions belong to the IFR class. Notice that the generalized Erlang d.f. can naturally approximate a wider subclass of distributions belonging to the IFR class.

It is reasonable to remember that the Erlang distribution represents an appropriate mathematical model for standby redundancy. Indeed, the process of a standby redundant group's operation can be described as a sequence of a constant number of periods of a unit's successful operation.

Thus, (7.38) can be used as a possible approximation of the IFR distributions, and (7.39) with (7.40) can be used for the DFR distributions. Of course, such an approximation leads to an increase in the number of states in the stochastic process under consideration. (Nothing can be obtained free, even



in mathematics!) But we should mention that the process itself becomes much simpler: it becomes purely Markov. At the same time, such an approximation is good only for systems of a very restricted size.

For simplicity of further illustrations, we will consider only cases where the initial distributions are approximated with the help of combinations of two exponential distributions. We should mention that the problem of determining an appropriate approximation of a distribution with monotone failure rates by the means of (7.38) to (7.40) is a special problem lying outside of the scope of this book.

Now we illustrate the main idea by means of simple examples.

**IFR Repair Time and Exponential Time to Failure** For some applied problems it is natural to use the exponential distribution for a random TTF. At the same time, to assume an exponential distribution for the repair time might seem strange: why should the residual time of repair not depend on the time already spent? If a repair involves a routine procedure, a more realistic assumption involves the IFR distribution of this r.v. To make this statement clearer, we consider a repair process as a sequence of several steps: if one step is over, the residual time of repair is smaller because now it consists of a smaller number of remaining steps.

In this case two failure states might be introduced for a unit: state 1 and state 1\*, both with an exponentially distributed time remaining in each of them. These sequential states represent the series sequence of two stages of repair (see Figure 7.4a). The total random time of staying in a failed state subset is the sum of two exponentially distributed random variables and, consequently, will have an IFR distribution. Incidentally, in this case (7.40) has the following expression:

$$A_1 = \begin{pmatrix} -\lambda & 0 \\ \lambda & -\mu \end{pmatrix}, \quad A_2 = \begin{pmatrix} -\lambda_1 & 0 \\ \lambda_1 & -\mu \end{pmatrix} \quad (7.41)$$

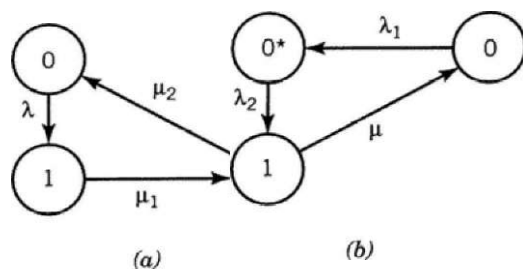


Figure 7.4. Transition graphs for a multistate model of renewable units: (a) with an IFR distributed repair time and an exponentially distributed failure-free time; (b) with an exponentially distributed repair time and an IFR distributed failure-free time.

**316 REPAIRABLE DUPLICATED SYSTEM**

(ft) with an IFR distributed failure-free time and an exponentially distributed repair time.

Suppose that for some reason a unit should be considered as having an IFR distribution for its TTF and an exponential distribution of repair time. Then a "dual" transition graph is considered with two operational states: state 0 and state 0\* (see Figure 1Ab).

**DFR Repair Time and Exponential Time to Failure** Sometimes a DFR repair time might be reasonably assumed. For example, a system may consist of two units: one of them takes more time for repair than another although both of them have exponentially distributed random repair times with different parameters. Thus, the system's repair time depends on which of the two units fails. In this case a "weighed" distribution could be a good mathematical model, and one more realistically assumes the DFR distribution of random time.

In this case two failure states are introduced: state 1 and state 1\*, both with an exponential distribution but with different parameters. Both states are separate and located on the same layer of the transition graph (see Figure 7.5a). Therefore, the process goes from operational state 0 to state 1 with probability

$$p_1 A$$

$$p_1 A + (1 - p_1) \lambda$$

and to state 1\* with probability

$$(1 - p) \lambda$$

$$p_1 A + (1 - p) \lambda$$

The total time of staying in a failed state subset has a DFR distribution.

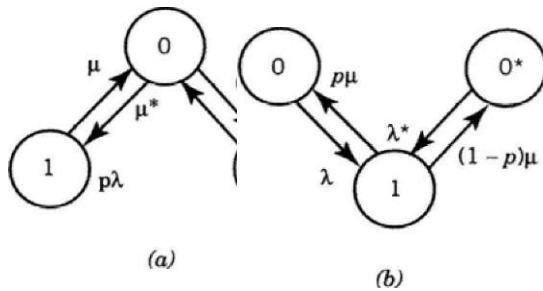


Figure 7.5. Transition graphs for a multistate model of renewable units: (a) with a DFR distributed repair time and an exponentially distributed failure-free time; (b) with a DFR distributed failure-free time and an exponentially distributed repair time.



- A DFR distribution of TTF and an IFR distribution of repair time (Figure 7.6c).
- Both distributions are DFR (Figure 1.6d).

It should be mentioned that this mathematical scheme allows one to create models for even more complex situations. For example, let us consider the following case. Both distributions are DFR, and the transition graph is close to that presented in Figure 1.6d, but with some differences. Let state 0 correspond to a long average operational time (a small value of intensity  $A_0$ ) and let state 011 correspond to a short average operational time (a large value of intensity  $A_0$ ). Thus,  $A_{011} > A_0$ . States 1 and 1\* have the same meanings: the first state is characterized by a repair intensity  $M_1$  the second by  $\mu_1^*$  where  $\mu_1^* > M_1$ . Let us assume that a "short" system repair time follows after a "long" time of a successful operation, and, on the contrary, after a "short" up time a repair time is usually "long." This can be explained on a physical level in the following way. A failure after a "normally long" failure-free operation is expected to be "normal" itself; that is, it requires, on average, a smaller time of repair. In the transition graph, it means that  $p_{01} > q_{01} = 1 - p_{01}$ . On the other hand, "short" periods of failure-free operation are sup-

$$\begin{aligned}
 l(t) &= \lambda_0 K_0(t) + p_2 K_2(t) \\
 K_1(t) &= \lambda_0 K_0(t) - \mu_1^* K_1(t)
 \end{aligned}
 \tag{7.42}$$

$$l = K_0(t) + K_1(t) + K_2(t)$$

$$K_0(0) = 1$$

'we put "long" and "short" in quotation marks because we consider r.v.'s with corresponding large and small means, but this does not mean that an r.v. with a larger mean cannot be less than an r.v. with a small mean, and vice versa. For simplicity, we use these terms for r.v.'s.

$$s(s - \mu_1^*)(s - s_2)$$

Equations (7.44) and (7.45) lead to the following system of equations:

$$\begin{aligned}
 A + B + C &= 1 \\
 M_1 + M_2 &= \lambda_0(s_1 + s_2) - B s_2 - C \mu_1^* \\
 A s_1 s_2 &= M_1 M_2
 \end{aligned}$$

or, taking into account that  $s_1 = s_2 = b$ ,  $A$  can be immediately expressed as

$$A = \frac{M_1 M_2}{\lambda_0 (M_1 + M_2) + M_1 M_2}$$

### 320 REPAIRABLE DUPLICATED SYSTEM

posed to be connected with some kind of "serious" failure, which leads to a "long" repair time. In the transition graph, this means that  $p_{0_0} > q_{0_0} = 1 - \lambda v^{-1}$

Of course, the inverse situation might be considered. Explanations also seem very reasonable: a "long" repair might follow a "long" period of successful operation. Indeed, we expect more failures of redundant units may appear during the longer period of time. As usual, a narrative of the system, which is taken to be a basis for the mathematical model, depends on the concrete actual nature of the system under investigation. For systems consisting of several units such an approximation may lead to difficulties in the construction of the corresponding transition graph.

Let us consider the simple case represented in Figure 7.4a in more detail. Incidentally, this case shows the special behavior of the availability coefficient. The system of differential equations is constructed in the usual way:

The LST of (7.42) gives the following system of algebraic equations:

$$\begin{aligned} (A_0 + 5)^{-1} - p_2(p_2 = 1) \\ - A_0^{-1}(i) + (f l y + 5) \leq \text{Pi}(5) = 0 \end{aligned} \tag{7.4}$$

$$3) \quad s \leq p_0(s) + \quad + \quad = 1$$

and the solution for  $\leq p(s)$  is

$$\leq p(s) = \frac{s^2 + f i_2 s + \dots}{s^2 + (A_0 + I X, + i x_2) s + A_0 j u, + \dots} \quad \wedge \wedge$$

Denote the eigenvalues (roots) of the denominator by  $s_k$ :

$$\wedge = \dots \pm \mathbf{V T} \quad a \quad a^2 \quad - b \quad s_3 = 0$$

where

$$\begin{aligned} a &= A_0 + j u, + j t_2 \\ b &= A_0 / (t i, + M i) + M 1 M 2 \end{aligned}$$

Note that the discriminant of the denominator is negative for any  $A_0$ , and  $H_2$ , which leads to the complex roots  $j j$  and  $s_2$ . A representation of  $\leq p_0(s)$  is found in the form

$$\leq p_0(s) = \frac{A}{s} + \frac{B}{s - s_1} + \frac{C}{s - s_2} = \frac{(A + B + C)s^2 - \dots + B s_2 + C S / / s + A s_1 s_2}{(A + B + C)s^2 - \dots + 5_2) + B s_2 + C S / / s + A s_1 s_2} \tag{7.45}$$

322 REPAIRABLE DUPLICATED SYSTEM

The following system is obtained as a result:

$$B + c = i \frac{Ld}{b}$$

$$BS_2 + CSJ = A \frac{MI}{b}$$

Because of the complex conjugate roots  $s_1$  and  $s_2$ , one can write

$$\sim C_1 e^{s_1 t} + \bar{C}_1 e^{\bar{s}_1 t} = 0$$

Thus, for the real parts of these roots,

Now we can find  $K_0(t)$  in the form

For the complex root  $s = a + ip$ ,

$$e^{st} = e^{at} (\cos pt + i \sin pt)$$

Taking into account that the complex roots  $s_1$  and  $s_2$  are conjugate, we may

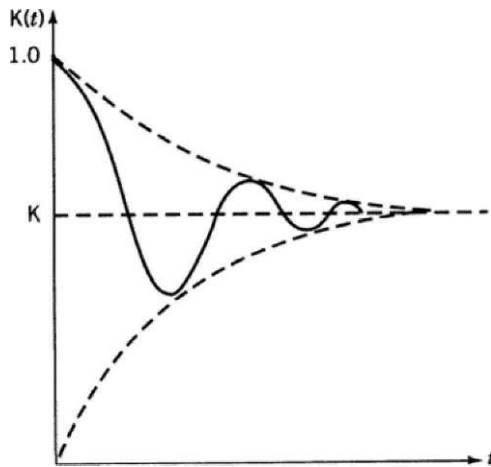


Figure 7,7, Time dependence of the availability coefficient for a unit with an IFR distributed repair time. Here  $K = \frac{V-1V2/P}{1 + I^2}$  and  $1 < I < *^2$



t

324

REPAIRA

BLE  
Write

DUPLICA

TED

SYSTEM

$$e^{-\lambda t} + e^{-\mu t} \leq 2$$

$$\cos \sqrt{b - \frac{a^2}{4} t}$$

The final result is

$$\begin{aligned}
 & \frac{M}{1} + \frac{M}{2} \\
 & \frac{M}{b} \\
 & *o(') \ll b - \frac{a^2}{4} t
 \end{aligned}$$

In this particular case the nonstationary availability coefficient is periodically oscillating with a decreasing amplitude (see Figure 7.7).

## 7.5 DUPLICATION WITH SWITCH AND MONITORING

Because of their relative simplicity, the mathematical models of a duplicated system with renewal allow one to consider some sophisticated cases close to real situations. Indeed, most "classical" mathematical models of redundant systems with repair are based on the assumption that the redundant group of units has an ideal switch which performs its functions reliably, without errors and delays. Moreover, the units are supposed to be totally and continuously monitored; that is, the occurrence of an operating or redundant unit failure becomes known immediately. It is clear that such assumptions are far from real. Sometimes, of course, these factors may be neglected. (But only sometimes!)

When a duplicated system is described by a Markov model, it is possible to provide an analysis of reliability by accounting for some additional factors. Obviously, it does not lead to especially clear and understandable models. At any rate, the solution can be derived.

Below we consider several examples which illustrate how one may construct appropriate mathematical models. We will not present the final results because they are inevitably bulky. A computer must be used for the numerical calculations. But, as is well known, no computer can substitute for the human mind, at least during the first stage of any research: one needs to be able to construct an appropriate mathematical model and only after this may one resort to computer calculation.

### 7.5.1 Periodic Partial Control of the Main Unit

We start with a simple example. A duplicated system consists of two independent identical units. One of them is in an operating position (the main unit) and the other is in a redundant position. The unit's failure rate depends on the current occupied position: operating or waiting. Let us assume that only

### 326 REPAIRABLE DUPLICATED SYSTEM

part of the main unit can be monitored continuously. The state of the remaining nonmonitored part of the main unit can be checked only periodically. In other words, if a failure has occurred in the nonmonitored part of the main unit, no switching to replace this failed unit is performed. A periodic test discovers that the main unit has failed and only then the switching might be performed. Thus, before this test, the duplicated system remains in a state of "hidden failure." The switching is assumed to be instantaneous. The redundant unit is continuously monitored, so a repair of the failed redundant unit begins instantly after a failure has occurred. Of course, the same thing happens if a failure occurs in the monitored part of the main unit.

During a repair of the main unit, all of its failures—both in the monitored and nonmonitored parts—are deleted. In other words, the repaired unit becomes as good as new. As soon as the failure of the main unit is detected (by any means—continuous or periodical monitoring), the redundant unit is switched into the main position. After repair, the unit becomes the redundant one. If one finds both units have failed, the total system repair is performed. After repair, the system, as a whole, becomes as good as new.

For the use of a Markov model, let us assume that monitoring is provided at randomly chosen moments of time. Moreover, assume that the distribution of the length of the periods between the tests is exponential. We mention that such an assumption is sometimes close to reality: in a computer, tests can be applied between runs of the main programs, and not by a previously set strict schedule.

The transition graph for this case is presented in Figure 7.8. The following notation is used in this example:

- $M$  is the operational state of the main unit.
- $M^*$  is the "hidden failure" of the main unit.
- $\bar{M}$  is the failure state of the main unit.
- $R$  is the operational state of the redundant unit.
- $\bar{R}$  is the failure state of the redundant unit.
- $\lambda_1$  is the failure rate of the nonmonitored part of the main unit.
- $\lambda_2$  is the failure rate of the monitored part of the main unit.
- $\lambda_3$  is the failure rate of the redundant unit.
- $\mu$  is the intensity of repair of a single unit.
- $\nu$  is the intensity of repair of the duplicated system as a whole.
- $\nu$  is the intensity of periodical tests.

The transition graph presented in Figure 7.8a is almost self-explanatory. Notice that for this case there are two states of system failure:  $[M^* \bar{R}]$  and  $[\bar{M} \bar{R}]$ . These failure states are denoted by bold frames in the figure.

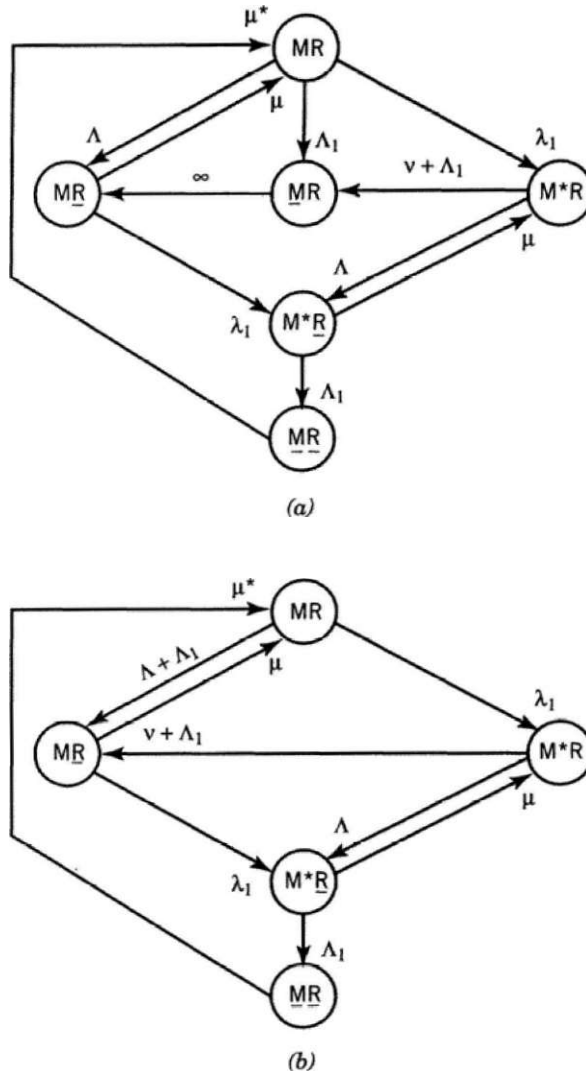


Figure 7.8. Transition graphs for a duplicated system with a partially monitored main unit: (a) graph including instantaneous "jumps" (intensity equal to  $\mu^*$ ) (b) equivalent graph excluding the state in which the system spends no time.

We will not describe the routine procedure of finding the reliability indexes. Our main goal is to build a mathematical model from the verbal description and to clarify all of the needed assumptions.

### 7.5.2 Periodic Partial Monitoring of Both Units

The duplicated system consists of two identical independent units. One of them is in an operating position (the main unit) and the other is in a

### 328 REPAIRABLE DUPLICATED SYSTEM

redundant position. The unit's failure rate depends on the occupied position: operating or waiting. Let us assume that only a part of each unit can be monitored continuously. (The monitored parts are identical in both units.) The state of the remaining nonmonitored parts of each of these units can be checked only periodically. Tests of the main and redundant units have different periods (intensity).

The switching system is analogous to that described in the previous example. If one knows that both units have failed, the repair is performed until complete renewal of the system. Let us consider two possible means of repair: (a) there are independent repair facilities for each unit, and (b) there is only one repair facility. After repair, the unit becomes as good as new.

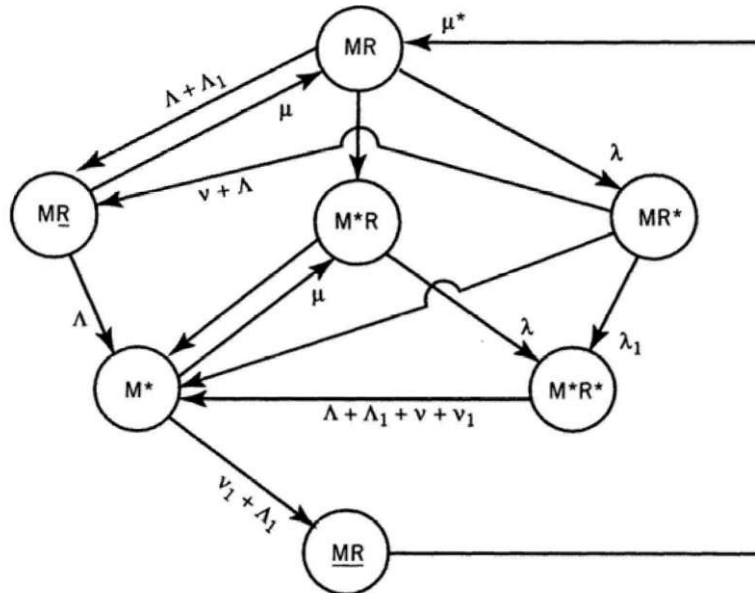
The transition graph for this case is presented in Figure 7.9. The following notation is used in this example:

- \*  $M$  is the operational state of the main unit.
- \*  $M^*$  is the "hidden failure" of the main unit.
- $M$  is the failure state of the main unit.
- \*  $R$  is the operational state of the redundant unit,
- \*  $R^*$  is the "hidden failure" of the redundant unit.
- $R$  is the failure state of the redundant unit,
- \*  $A_j$  is the failure rate of the nonmonitored part of the main unit.
- \*  $A_j$  is the failure rate of the monitored part of the main unit.
- \*  $A$  is the failure rate of the nonmonitored part of the redundant unit.
- \*  $A$  is the failure rate of the monitored part of the redundant unit.
- \*  $f_i$  is the intensity of repair of a single unit.
- $n^*$  is the intensity of repair of the duplicated system when there are two failed units,
- \*  $y$ , is the intensity of periodic tests of the main unit,
- \*  $v$  is the intensity of periodic tests of the redundant unit.

The transition graph presented in Figure 7.9 is almost self-explanatory. We only discuss the following two transitions:

1.  $[M R^*]$  to  $[M R]$ : This transition occurs if (a) an extra failure appears in the continuously monitoring part of the redundant unit or (b) the periodic test has found a "hidden failure."
2.  $[A_f / ?^*]$  to  $[M^* J^*]$ : This transition occurs if a failure appears in the continuously monitoring part of the main unit. Then the main unit is directed to repair and is substituted by the redundant one with a "hidden failure."

These failure states are again denoted by bold frames in the figure.



**Figure 7.9.** Transition graph for duplicated system with partially monitored main and redundant units.

We will again write no equations. This is obviously a routine procedure.

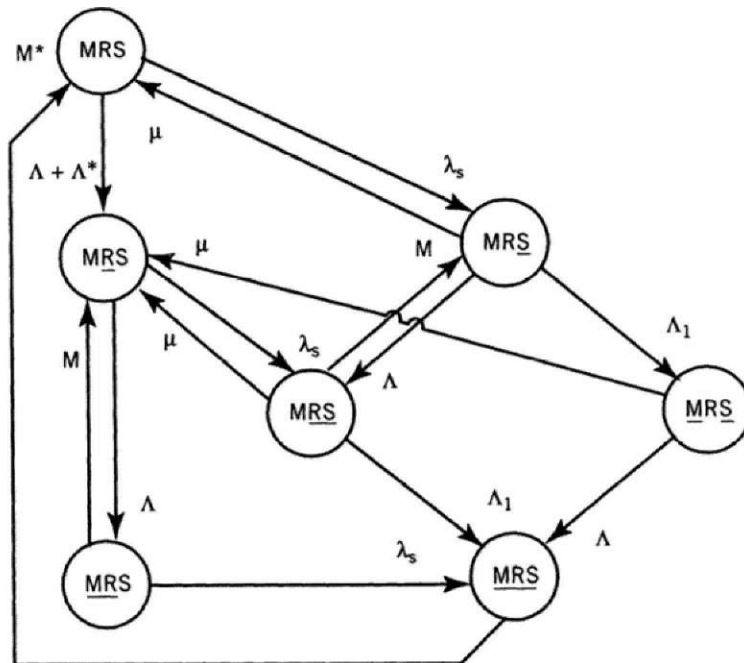
As an exercise, consider the system when, after the failure of both units,

We will again write no equations. This is obviously a routine procedure.

As an exercise, consider the system when, after the failure of both units, the system is subjected to a total renewal: the system is repaired as a whole until both units are as good as new (the transition from the system failure state to the failure-free state).

### 7.5.3 Unreliable Switch

Consider a duplicated system with an unreliable switching device. A switching failure becomes known immediately and its repair begins at once. There is only one repair facility. Repair is performed in accordance with a FIFO (first-in, first-out) rule. If both units have failed, the total repair is performed. The failure of the main unit, occurring during the repair of the switch, leads to a system failure even if the redundant unit is operational. But a switching failure itself does not interrupt the main unit's successful operation. The monitoring of both units is supposed to be continuous and ideal. Repairs of both units and the switch are supposed to be independent.



**Figure 7.10.** Transition graph for duplicated system with an unreliable switch.

The transition graph for this case is presented in Figure 7.10. The following notation is used in this example:



As in the previous examples, the transition graph presented in Figure 7.10 is almost self-explanatory. We only discuss the following transitions:

1.  $[M R 5]$  to  $[M R 5]$ : This transition occurs if the switch has been repaired: at the moment of the termination of repair, the redundant unit is instantly directed to the position of the failed main unit, and the latter begins to be considered as a redundant unit directed to repair.
2.  $[M R 5]$  to  $[M R 5]$  and  $[M R 5]$  to  $[M R 5]$ : These two transitions depend on which unit is repaired first: if the switching device is in repair, it is impossible to put the repaired redundant unit into the position of the main unit.

These failure states are again denoted by bold frames in the figure.

#### 7.5.4 Unreliable Switch and Monitoring of Main Unit

A duplicated system consists of two identical independent units: main and redundant. The unit failure rate depends on the occupied position. Let us assume that only a part of the main unit can be monitored continuously. The state of the remaining nonmonitored part of the unit can be checked only periodically. The switching device works as described above. Repairs of both the units and the switch are independent. After repair, a unit (or a switch) becomes as good as new.

The transition graph for this case is presented in Figure 7.10. The following notation is used in this example:

Note that the switching device may be one of the two following main types: (a) as we considered before, a switching failure does not interrupt the system's operation, or (b) a switching failure interrupts the system operation. In the latter case, the switch is a necessary part of the system. This may occur, for example, if the switch plays the role of an interface between the duplicated system's output and the input of another system or subsystem.

Of course, there are many other concrete examples of this type. We can only repeat that our main goal is to explain the methodology of modeling and not to give a list of the results or to make the reader exhausted with boring solutions of bulky equations.

The mathematical techniques used in this section are simple enough. But the results obtained are not always very clear or "transparent" for further analysis: What will happen if one changes some parameters? What will happen if the switching or monitoring methods are changed? Of course, in practical situations an engineer would like to have correct *and* simple formulas to perform a quick and understandable analysis of the designed system. Fortunately, for highly reliable systems (we emphasize again that this is the most important practical case!), it is possible to develop such simple and sufficiently accurate methods. The reader can find such methods in Chapter 13 dedicated to heuristic methods in reliability.

## CONCLUSION

It seems that the first paper on the analysis of a duplicated system with repair (renewal) was published by Epstein and Hosford (1960). They solved the problem for a purely Markov model when the distributions—both TTF and repair time—were exponential. They solved the problem with the help of birth and death processes. Their model is also described in Gnedenko and Kovalenko (1987). Here the solution of the same problem for the duplicated system was found for both active and underloaded redundancy.

A systematic investigation of renewal systems, in particular, the duplicated system, may be found in Ushakov (1985, 1994). Belyaev (1962) developed an elegant method of so-called "striping Markov processes" which has allowed one to solve the problem with no assumption of exponentiality on the repair time. Independently, Gaver (1963) obtained practically the same results with the help of traditional methods.

Gnedenko (1964a, 1964b) obtained solutions for the general case when both distributions are arbitrary. Theorems concerning the asymptotic behavior of renewal duplicated systems have been obtained by D. Gnedenko and Solovyev (1974, 1975). and, practically simultaneously, by Gnedenko, Belyaev, and Solovyev (1969). The method of fictitious states (stages) for "Markovization" of non-Markov models as applied to queuing systems takes its origin from Erlang's work. A comprehensive exposition of existent mathematical results related to the problem can be found in Ushakov (1985, 1994), where

results concerning the various models of duplicated renewal systems are presented. A detailed review of asymptotic methods is given by Gertsbakh (1984).

## REFERENCES

- Belyaev, Yu. K. (1962). Striping Markov processes and their applications to reliability problems (in Russian). *Proc. Fourth All-Union Meeting on Probab. Theory Math. Statist., Vilnius (Lithuania)*, pp. 309-323.
- Epstein, B., and T. Hosford (1960). Reliability of some two unit redundant systems. *Proc. Sixth Nat. Symp. on RQC*, pp. 466-476.
- Gavcr, D. P. (1963). Time to failure and availability of parallel systems with repair. *IEEE Trans. RQC*, vol. R-12, pp. 30-38.
- Gertsbakh, I. B. (1984). Asymptotic methods in reliability theory: a review. *Ada. in Appl. Probab.* vol. 16.
- Gnedenko, B. V. (1964a). On duplication with renewal, *Engrg. Cybernet.*, no. 5, pp. 111-118.
- Gnedenko, B. V. (1964b). On spare duplication. *Engrg. Cybernet.*, no. 4, pp. 3-12.
- Gnedenko, B. V., and I. N. Kovalenko (1987). *Introduction to Queuing Theory*, 2nd ed. (in Russian). Moscow: Nauka.
- Gnedenko, B. V., Yu. K. Belyaev, and A. D. Solovyev (1969). *Mathematical Methods of Reliability Theory*. New York: Academic.
- Gnedenko, D. B., and A. D. Solovyev (1974). A general model for standby with renewal. *Engrg. Cybernet.*, vol. 12, no. 6.
- Gnedenko, D. B., and A. D. Solovyev (1975). Estimation of the reliability of complex renewable systems. *Engrg. Cybernet.*, vol. 13, no. 3.
- Solovyev, A. D. (1970). Standby with rapid repair. *Engrg. Cybernet.*, vol. 4, no. 1.
- Solovyev, A. D. (1971). Asymptotic behavior of the time of first occurrence of a rare event. *Engrg. Cybernet.*, vol. 9, no. 3.
- Solovyev, A. D. (1972). Asymptotic distribution of the moment of first crossing of a high level by birth and death process. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.*, Issue 3.
- Ushakov, I. A., ed. (1985). *Reliability of Technical Systems: Handbook (in Russian)*. Moscow: Radio i Sviaz.
- Ushakov, I. A., ed. (1994). *Handbook of Reliability Engineering*. New York: Wiley,

## EXERCISES

- 7.1 There are two transition graphs (see Figures E7.1a and b). State 2 is a failed state in both cases.
- Give a verbal description of the two systems to which these graphs correspond.
  - Which system has a larger MTTF?

334 REPAIRABLE DUPLICATED SYSTEM

- (c) Which system has a larger MTBF?
- (d) What is the difference between the MTTF and MTBF of the first system?
- (e) Which system has a larger mean repair time?

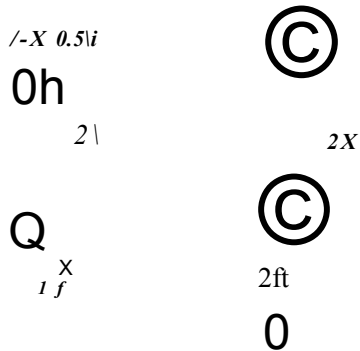


Figure E7.1.

- 7.2 Depict a transition graph for a unit with an exponentially distributed time to failure and a repair time having an Erlang distribution of the third order.
- 7.3 Depict a transition graph for a renewable unit with an exponentially distributed time to failure.  $P(t) = e^{-\lambda t}$ , and with a repair time distributed as

$$G(t) = p_1 e^{-\lambda t} + p_2 + p_3$$

- 7.4 Given an interpretation of the transition graph depicted in Figure E7.4.



Failure Figure E7.4.

**O**

Failure

**SOLUTIONS**

- 7.1 (a) The first system is a duplicate system with a loaded redundant unit where after the system has failed it is renewed as a whole. The second system is an ordinary duplicate system of two independent identical units operating in a loaded regime.
- (b) Both systems have the same MTTF.
- (c) The first system has a larger MTBF than the second one because after each failure this system starts from state 0. After a system failure the second system starts from state 1 where there is a possibility of entering a failed state immediately.
- (d) There is no difference at all.
- (e) The first system has twice as large a repair time.
- 7.2 The solution is depicted in Figure E7.2.

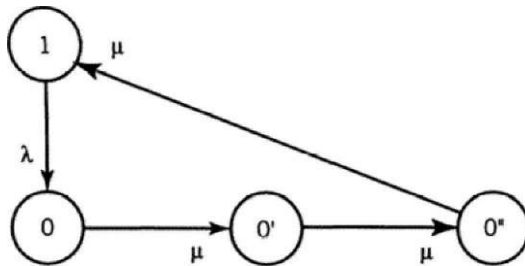


Figure E7.2.

- 7.3 The transition graph is depicted in Figure E7.3.

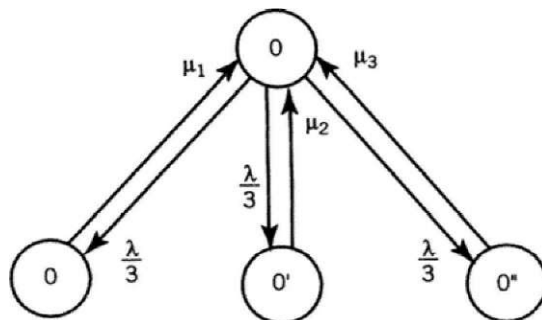


Figure E7.3.

- 7.4 The system is a series connection of an unrepairable unit with an exponentially distributed time to failure with parameter  $\lambda$ , and a repairable duplicated group with failure rate  $\lambda$  and intensity of repair  $\mu$ . The redundant group consists of units in a loaded regime; there are two repair facilities. The structure of the system is depicted in Figure E7.5.

X

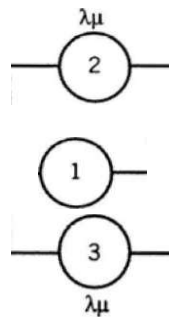


Figure E7.5.

## CHAPTER 8

# ANALYSIS OF PERFORMANCE EFFECTIVENESS

### 8.1 CLASSIFICATION OF SYSTEMS

#### 8.1.1 General Explanation of Effectiveness Concepts

Modern large-scale systems are distinguished by their structural complexity and their requirements for sophisticated algorithms to facilitate the functioning and interacting of their subsystems. On the one hand, this allows them to fulfill many different operations and functions, while, on the other hand, it leads to stable operations with a sufficient level of effectiveness even with some failed units and subsystems and/or under extreme influences of the external environment.

The adaptation of a complex system to external influences and to internal perturbations is possible only because of the redundancy of the system's structure and its ability to readjust its functions under various circumstances. In other words, the feature of modern technical systems is not only an extreme increase in the number of interacting units but also the appearance of entirely new qualitative properties. One of these properties is the stability of operation mentioned above.

It is also very important that modern large systems, such as information systems (computer and communications networks, control systems, etc.), energy systems (electric power networks, oil and gas pipelines, etc.), and transportation systems (railroads, highways, airlines, etc.) are multifunctional. Such systems, as a result of external and internal influences, can perform some functions perfectly and, at the same time, completely interrupt the performance of other functions. This means that, according to one criterion, a system as a whole could be considered successful and, by another criterion,





### 340 ANALYSIS OF PERFORMANCE EFFECTIVENESS

it could be considered failed. A researcher encounters the usual difficulties associated with multicriteria analysis.

But even for a complex system predestinated for one type of operation, there is generally no strict definition of failure. Often in such systems even a significant set of failed units could lead only to a decrease in performance and not to a complete system failure. This happens because of a partial "overlapping" of different subsystem (unit) functions, the presence of different feedbacks, the means of error correction, and so forth.

We consider several simple examples. In a regional power system, a failure of some subsystem (e.g., the failure of the fuel transportation system of an electric power plant) can be compensated for by using fuel from storage. In another case a deficiency of energy can be compensated for by a partial increase in the power of neighboring plants. Sometimes clients might use another type of energy supply. Under conditions of a severe energy deficit, clients with lower levels of priority might be temporarily "turned off" from an energy system in order to decrease the total damage.

Sometimes a completely operational system might be unable to perform some of its functions because of a harmful coincidence of external circumstances. For example, consider a communications network. All equipment in the system could be in a perfect operating state, but weather may spoil the opportunity to use certain radio channels. The same effect may be observed if there is some neighboring influence of other radio transmission systems. Even network users may create excessively heavy communication traffic which can lead to system performance failures.

Of course, from a client's viewpoint, he or she is quite indifferent to the reason for a breakdown in communication: either it happens because of a system failure or because of an overloading of the communications network.

For all such systems it is natural to speak about performance effectiveness. In each concrete case the index (or indexes) of performance effectiveness should be chosen with respect to the type of system under consideration, its destination, conditions of operation, and so forth. The physical characteristics of the *performance effectiveness index* (PEI) are usually completely defined by the nature of the system's outcome and can be evaluated by the same measures. In most practical cases we can measure a system's effectiveness in relative units. We might take into account the nominal (specified) value of a system's outcome as the normalizing factor. In other words, the PEI is a measure of the quality and/or volume of the system's performed functions or operations; that is, it is a measure of the system's expediency.

Of course, a system's efficiency is not an absolute measure. It depends on the type of functions and tasks being performed and the operating environment. A system which is very efficient under some circumstances and for some operations might be quite useless and ineffective under another set of circumstances and/or operations.

In general, a PEI is dimensional. The dimension of the PEI depends on the system's outcomes, as we mentioned above. When it is possible in the

following discussion, we shall consider the PEI as a ratio of the expected system outcome to its maximal outcome. In this case the PEI is nondimensional. Of course, we always assume a larger outcome is a better outcome. The use of a nondimensional PEI is very convenient in many practical evaluations.

Sometimes we encounter "pessimistic" measures which characterize a system's performance. For example, consider the acceptable error of a technological process or the permissible volume of pollution of a plant. Such indexes measure "ineffectiveness" rather than effectiveness. Usually, in such cases one can reformulate the desired outcome in "positive" terms.

If a system's outcome has an upper bound, the PEI can be expressed in a normalized form; that is, it may be considered as having a positive value lying between 0 and 1. Then we have  $PEI = 0$  if the system has completely failed and  $PEI = 1$  when it is completely operational. For intermediate states,  $0 < PEI < 1$ .

When considering a system's effectiveness, one should remember the property of monotonicity introduced earlier. In this context, an increase in the reliability of any unit leads to a simultaneous increase in the system's effectiveness. Also, a failure of any unit can only decrease (not increase) a system's effectiveness.

It is convenient for system design to determine a PEI in relative units, because in this case one does not need to measure an absolute value of a system's outcome for different states. The absolute values of a PEI are very convenient if we must compare several different competitive variants of a system. They allow us to compare variants of a system with different reliability and different efficiencies of performance. It is clear that reliability alone does not completely solve the problem of engineering design.

### 8.1.2 Classes of Systems

Consider a system consisting of  $n$  units. As before, we suppose that any system unit has two states: an operating state and a failed state. Let  $x_i$  be the indicator of the  $i$ th unit's state:  $x_i = 1$  when the unit is up and  $x_i = 0$  when the unit is down. The system then has  $2^n$  different states as determined by the states of its units. Denote a system state by  $X = (x_1, x_2, \dots, x_n)$ .

If we consider the process of a system's evolution in state space, then for each unit we should consider the process  $x_i(t)$ , and for the system as a whole, the process  $X(t)$ . The transformation of system states  $X(t)$  characterizes the system's behavior. On the basis of knowledge about this process, we can analyze a system's effectiveness.

Taking into account the length of a system's performance, it is reasonable, for effectiveness analysis, to distinguish two main classes of systems: instant and enduring.

Some systems are characterized by their instant outcome at a moment of time. The current effectiveness of an *instant system* is completely determined

by its state at the moment of performance. It is clear that no instant system exists in reality because any task has some duration. Strictly speaking, we consider a system whose duration of performance is negligibly short in comparison with time intervals between changing system states  $X(t)$ . This means that

$$P\{X(t) = X(t + t_0)\} = 1 - e^{-\lambda t_0} \quad (8.1)$$

where  $t_0$  is the system's task duration and  $e$  is a practically negligible value. (The size of  $e$  depends on the required accuracy of analysis.)

From (8.1) it follows that the current effectiveness of a system is completely determined by the current system state  $X = X(t)$ . For this state the effectiveness coefficient equals  $W_x$ , and the system's PEI can be determined as the expected value of  $If_x$ .

Examples of practical instant systems are missiles, production lines (during production of a single item), and a communications network during an individual call.

For an enduring system condition (8.1) is not valid. The effectiveness of an enduring system depends on a trajectory of the system's transition from one state to another. In this case the fact that some particular units have failed is very important, but the moments and the order of their failures are also equally important. In other words, for these systems the effectiveness is determined by a trajectory of states changing during the system's performance of a task.

Examples of enduring systems are different technological and chemical processes, information and computer systems, aircraft, and so on.

## 8.2 INSTANT SYSTEMS

Let  $h_{X_k}(t)$  denote the probability that an instant system at moment  $t$  is in state  $X_k(t)$ . We assume that the current effectiveness of the system being in any state can be evaluated. Let us denote this value for state  $X$  as  $If_x$ . It is natural to determine  $W$  as the expected value of  $W_x$ , that is,

$$W = \sum_{k=1}^N h_{X_k}(t) W_{X_k} \quad (8.2)$$

where  $N = 2^n$  is the total number of different system states.

It is clear that an absolutely accurate calculation of a system's effectiveness when  $n \geq 1$  is a difficult, if not unsolvable, computational problem. First of all, it is connected with the necessity of determining a large number of coefficients  $W_k$ . Fortunately, it is sometimes not too difficult to split all of the system's states into a relatively small number of classes with close values  $W_k$ .

If so, we need only to group appropriate states and calculate the corresponding probabilities.  $W^{\wedge}$ , can then be calculated as

$$K_{ysM} = E W, E M < \tag{8-3}$$

where  $M$  is the number of different levels of the values of  $W_x$  and  $G_j$  is the set of system states for which  $IV_x$  belongs to the  $j$ 'th level.

Later we shall consider special methods for the evaluation of the effectiveness of higher-dimensional systems.

Let us evaluate a system's effectiveness for a general case. For notational simplicity, we omit the time  $t$  in the expressions below. Let  $h_0$  denote the probability that all units of the system are successfully operating at moment

$$K - n P_i \tag{8-4}$$

Let  $h_i$  denote the probability that only the  $i$ 'th unit of the system is in a down state at moment  $t$  (repairable systems can be considered as well as unrepairable). Then

$$h_i - q_i \prod_{j=1}^{i-1} P_j = h_0 = S_i K \tag{8.5}$$

where, for brevity, we introduce  $g_i = q_i/p_i$  and  $h_{ij}$  denotes the probability that only the  $i$ th and  $j$ th units of the system are in down states at moment  $t$ :

$$h_{ij} = h_0 \prod_{k=1}^{i-1} P_k \prod_{l=1}^{j-1} P_l = g_i g_j h_0 \tag{8.6}$$

and so on.

We can write the general form of this probability as

$$- n P_i n * - n * \tag{8-7}$$

where  $G_x$  is the set of subscripts of the units which are considered operational in state  $X$  and  $G_x^c$  is the complementary set. Sometimes it is reasonable to write (8.7) for any  $X$  as

$$* \mathbf{x} = \mathbf{n} p_M \sim \tag{8-8}$$

It is clear that (8.7) and (8.8) are equivalent. Using (8.4) to (8.8), we can

344 ANALYSIS OF  
PERFORMANCE  
EFFECTIVENESS

rewrite (8.3)

$$\hat{W}_{\text{sys}} = \frac{1 + \sum_{i=1}^n w_i g_i + \sum_{j=1}^n w_j g_j}{\sum_{i=1}^n w_i c_i / s_i} \quad (8.9)$$

where  $W_0$  is the system effectiveness for state  $X_0$  and  $W_b, W_{tj}, \dots$  are normalized effectiveness coefficients for states  $X_b, X_{tj}, \dots$ . In other words,  $W_b = W_b/W_0, W_u = W_u/W_0, \dots$ .

For a system consisting of highly reliable units, that is,

$$\max_{\{s_i\}} q^* \ll - \frac{1}{r t} \quad (8.10)$$

expression (8.9) can be approximated as

$$W_{\text{sys}} \approx \frac{\sum_{i=1}^n W_i q_i \sim W_0 (1 - \sum_{i=1}^n w_i c_i / s_i)}{\sum_{i=1}^n w_i c_i / s_i} \quad (8.11)$$

Here  $w_i = 1 - W_i$  has the meaning of a "unit's significance."

**REMARK.** It is necessary to note that, strictly speaking, it is wrong to speak of a "unit's significance." The significance of a unit depends on the specific system state. For example, in a simple redundant system of two units, the significance of any unit equals 0 if both units are successfully operating, but if a single unit is operating at the time, then its significance equals 1. Other examples are considered below.

Consider some particular cases of (8.11). If  $p_i(t) = \exp(-A_i t)$  is close to 1 and, consequently,  $\lambda_i(r) = A_i r$ , then (8.11) can be approximated by

$$W_{\text{sys}} \approx \frac{W_0 (1 - \sum_{i=1}^n w_i A_i r)}{\sum_{i=1}^n w_i A_i r} \quad (8.12)$$

We can see that "the significance of unit" is reflected, in this case, in the factual failure rate. (See the previous remark.)

If  $p_i$  is a stationary availability coefficient, that is,  $p_i = \lambda_i / (\lambda_i + r_i)$  where  $T_i$  is the MTBF of the  $i$ 'th unit and  $r_i$  is its idle time and  $\lambda_i \gg r_i$ , then it is possible to write the approximation

$$\hat{W}_{\text{sys}} \sim W_0 \frac{\sum_{i=1}^n w_i T_i}{\sum_{i=1}^n w_i} \quad (8.13)$$

Again, we can consider "the significance of a unit" in a new form keeping in mind the same precautions as above.

We should once more emphasize that all approximate expressions (8.9) to (8.13) are valid only for highly reliable systems.

**Example 8.1** To demonstrate the main ideas, we first use a simple system consisting of two redundant units. Let the system's units have corresponding probabilities of successful operation equal to  $p_1$  and  $p_2$ . The problem is to find the probability of the system's successful operation.

*Solution.* By the definition of a duplicate system,  $H_{s,} = W_2 - 1$  and  $W_n = 0$ . Thus, for this particular case

$$\hat{W}_{syst} = 1 - (1 - p_1)(1 - p_2) = 1 - (1 - p_1 - p_2 + p_1 p_2) = p_1 + p_2 - p_1 p_2$$

which completely coincides with the corresponding expression for the probability of failure-free operation of a duplicate system.

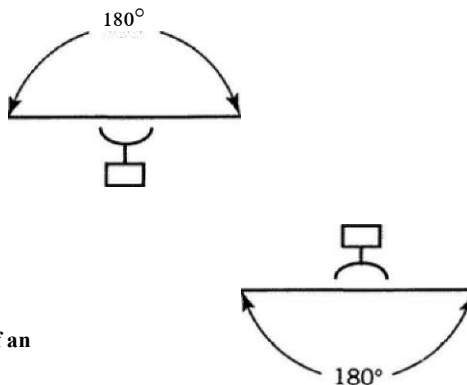
It is to be understood that  $W$  is a generalization of a common reliability index. Everything depends on the chosen coefficients  $W_x$ .

Now we consider more interesting cases which cannot be put into the framework of a standard reliability scheme.

**Example 8.2** An airport traffic control system consists of two stationary radars each with an effective zone of  $180^\circ$  (see the schematic plot of the system in Figure 8.1). For this example let us assume that the effectiveness of the system in a zone with active radar coverage equals 0.7. The availability coefficient for each radar is equal to 0.9, (Of course, nobody would use such an ineffective system in practice!) We will assume that if only one radar is operating, it means that the system PEI = 0.5. It is necessary to evaluate the PEI for the system.

*Solution.*

$$= (0.7) \cdot p_1 + (1/2)(0.7) \cdot p_2 + (1/2)(0.7) \cdot p_2 - (1/2)(0.7) \cdot p_1 + (1/2)(0.7) \cdot p_2 = (0.7)(0.9) = 0.63$$



**Figure 8.1.** Schematic representation of an airport radar system.



**Example 8.3** Consider the same airport traffic control system as in Example 8.2. To increase the effectiveness of the system, the operating zones of the radars overlap. In addition, we assume that within the overlapped zone, the effectiveness of service is higher. Let us say that the coefficient of effectiveness in an overlapping zone is practically equal to 1, while the same coefficient of effectiveness in an ordinary zone is 0.7.

The system's effectiveness is determined as the average probability of success weighted by the size of the zones with their corresponding effectiveness coefficients. There are two possibilities to design a system with overlapping zones. These two cases are depicted in Figure 8.2. The availability coefficient of each radar again equals 0.9. The problem is to compare the effectiveness of both variants and to choose the best one.

*Solution.* Consider the first variant, A, with two radars in the north zone and two radars in the south zone (see Figure 8.2a). It is clear that we can consider two independent subsystems, each delivering its own outcome to the control system as a whole. The outcome of one subsystem is equal to one-half of the system's total outcome. Denote the effectiveness indexes of these two subsystems and of the whole system by  $W_u$ ,  $W_z$ , and  $W_{\text{system}}$  respectively. Because of the identity  $W_x = W_z$ ,  $W^{\wedge} = 2W_x = 1W_z$ .

Each subsystem can be in one of two useful states:

- Both radars are operating, and the probability of this is  $(0.9 \times 0.9) = 0.81$ ; the coefficient of effectiveness in the zone is equal to 1.
- Only one radar is operating, and the probability of this is  $(0.9 \times 0.1) = 0.09$ ; the coefficient of effectiveness in the zone is equal to 0.7.

(Recall that each subsystem covers only one-half of the zone of the operating system.)

$$W_{\text{system}} = 2[(0.81)(1)(0.5) + (0.09)(0.7)(0.5)] = 0.873$$

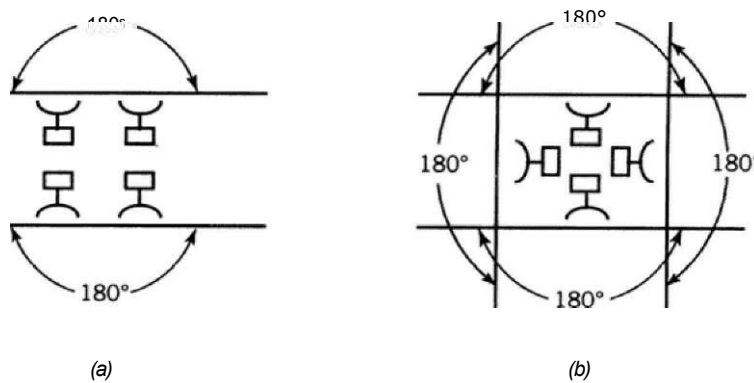


Figure 8.2. Two possible ways of using redundant radars for an airport radar system.

**TABLE 8.1 Analysis of Variant B of Example 8.3**

Type of State	Number	Probability	Effectiveness Coefficient	Product
NSEW	1	$0.9^4$	1	0.6561
N'SEW	4	$4(0.9)^3(0.1)$	$(1/2X1 + 0.7)$	0.3060
N'S'E'W	2	$2(0.9)^2(0.1)^2$	0.7	0.01134
N'SE'W	4	$4(0.9)^2(0.1)^2$	$(1/4)1 + (1/2X0.7)$	0.01944
N'S'E'W	4	$4(0.9X0.1)^*$	$(1/2X0.7)$	0.00126

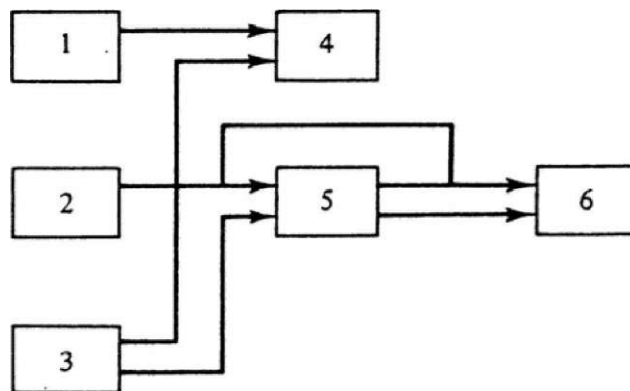
Now let us consider the second variant, B (see Figure 8.2b). In this case we have to analyze  $2^3 - 1 = 7$  different states. The results of this analysis are presented in Table 8.1. Here we denote the corresponding radars by N, S, E, and W and use the symbols N', S', E', and W' to denote their idle states. The final result can be found by summing all values in the last column of Table 8.1:

$$w = 0.99314$$

Thus, variant B is the preferable one.

Example 8.4 As Russian nonmilitary authors, we never had access to information about former Soviet military systems, even the out-of-date systems. So for illustration we are forced to use an illustrative narrative from the proceedings of one of the early IEEE Reliability Conferences.

Just after World War II there were anti-aircraft missile systems of the following simple type. There was a radar searching for a target with information displayed on a screen. After locating the target, a conveying radar was switched in and information about the target was processed by a computer and displayed by the same monitor. If the searching radar failed, the conveying radar was used for searching (with a lower efficiency). The last step



**Figure 8.3. Simplified block diagram of an aircraft radar system. 1 = searching radar; 2 = optical device; 3 = conveying radar; 4 = display; 5 = computer; 6 = control equipment.**

**TABLE 8.2 Effectiveness of Different Modes of the System in Example 8.4**

Number	P <sup>utje</sup>		Stage of Operation		W <sub>x</sub>
	Searching	Finding	Guiding		
1	Searching radar	Display	Watching radar and computer	1.0	
2	Watching radar	Display	Watching radar and computer	0.6	
3	Searching radar	Display	Optical equipment and computer	0.3	
4	Optical equipment	Optical equipment	Optical equipment and computer	0.1	
5	Optical equipment	Optical equipment	Optical equipment	0.1	

connected with the destruction of the target was fulfilled by means of control equipment and a controlled missile. In case of a failure of the electronic equipment (which was so unreliable at that time!), a pilot could use an optical system for pursuing the target (see Figure 8.3).

Thus, the system could be in different states because of the failures of the equipment. Different modes of the system under consideration are presented in Table 8.2. The probabilities of a successful operation at some given moment of time are

For the searching radar,  $p_1 = 0.80$ .

For the optical equipment,  $p_2 = 0.99$ .

For the watching radar,  $p_3 = 0.80$ .

For the display,  $p_4 = 0.95$ .

For the computer,  $p_5 = 0.90$ .

For the control system,  $p_b = 0.95$ .

*Solution.* Let the probability of the  $k$ -th mode be denoted by  $h_k$ . Then

$$h_1 = p_1 p_2 p_3 p_4 p_5 p_6 = 0.52$$

$$= 0.8 \cdot 0.99 \cdot 0.8 \cdot 0.95 \cdot 0.9 \cdot 0.95$$

$$h_3 = p_3 p_4 p_5 p_6 = 0.13$$

$$K = p_1 p_2 p_3 p_4 p_5 p_6 + p_3 + \dots + p_5 p_6$$

(in this case one should take into account the impossibility of performing the operation by means of previous modes)

350 ANALYSIS OF PERFORMANCE EFFECTIVENESS

$$a_5 = p_2 Q_s P_i (i_4 + p^* Q 1^3)^{11} \circ^{-12}$$

The final result is that the probability of success is equal to

$$W_{\text{sys}} = (0.51) \blacksquare 1 + (0.13)(0.6) 4 - (0.13)(0.3) + (0.08)(0.15) + (0.01)(0.1) - 0.66$$

---


$$\frac{\max \langle 7, (r, r + f_0) \rangle}{\text{lsin}} \ll - \frac{12}{n} \quad (8.15)$$

8.3 ENDURING SYSTEMS

If the period of time it takes to perform a system task is sufficiently long, that is, during this period a number of different state changes can occur, then one needs to investigate an *enduring system*. In this case a probabilistic measure is distributed over a continuous space of trajectories of the changing system states. Let  $Z(t, t + f_0)$  denote some fixed trajectory. In the continuous trajectory space we can determine a density  $f_z$  for each such trajectory. At the same time, if a system moves from one state to another in correspondence to such a trajectory, one can characterize it by some predetermined outcome (effectiveness), say  $W_z$ .

Now we can write an expression similar to (8.2)

$$\int_{G_z} W_z dF(Z) \tag{8.14}$$

where  $G_z$  is the space of all possible system state trajectories in the interval  $(t, t + t_0)$ . The simplicity of (8.14) is deceptive. In general, it is very difficult to compute the densities of trajectories and to find analytical expressions for the outcomes of a system for each particular case. (We will return to this topic later.) To illustrate this statement, consider a simple duplicated system consisting of two unrepairable units. Initially, both units are in an operating state. The expression for this case is

$$\begin{aligned} & \int_{P_1}^{t_1+t} W_1 dF(t_1) + \int_{P_2}^{t_2+t} W_2 dF(t_2) \\ & + \int_{P_1 \setminus P_2} W_{12} dF(t_1) dF(t_2) \end{aligned}$$

Here  $W$  again is a normalized effectiveness coefficient relative to the nominal trajectory with no failures.

Thus, even for a very simple enduring system, the expression for the evaluation of effectiveness is quite complex. But the complexity of the expression is not all that makes this problem difficult. One also needs very detailed information about the reliability of the system's units as well as some knowledge about the effectiveness coefficients for different trajectories. In this case one is interested in finding an approximate solution.

Let us denote  $\#_i(t) = 1 - p_i(t)$ . If the following condition is valid:

it is possible to write an

$$E_{i < n} L(t, t + t_0) \sim r'' W_i(x_t) dF_i(x,)$$

approximate formula for

unrepairable systems:

$$1_{\text{syst}} \sim 0$$

For repairable systems such an analysis becomes extremely difficult and boring. For numerical calculations one can introduce a discrete lattice to describe the system's trajectory in the space of system states. But in this case one encounters a complex factorial problem. Of course, the largest practical difficulties arise in the determination of the effectiveness coefficients for different state trajectories in both cases: continuous and discrete.

For enduring systems  $B^{\wedge}$ , is also a generalized index in comparison with the standard reliability indexes. As usual, a generalized (or more or less universal) method permits one to obtain any different particular solutions but with more effort. So, for simple reliability problems, one need not use this general approach. At the same time we should mention (hat, in general, a system performance effectiveness analysis cannot be done via common reliability methods.

We first consider two simple examples which can be solved by the use of standard reliability methods.

**Example 8.5** Consider a unit operating in the time interval  $[0, Z]$ . An outcome of the unit is proportional to the operating time; that is, if a random TTF is more than  $Z$ , then the outcome of the unit is proportional to  $Z$ . Let  $p(t)$  be the probability of a failure-free operation during time  $t$  and let  $q(t) = 1 - p(t)$ . Find an effectiveness index  $H_{\text{syst}}$ .

*Solution.* Simply reformulating the verbal description gives the result

$$W^{\wedge} = Zp(Z) + \int_0^Z t dq(t)$$

and, after integrating by parts,

$$- Zp(Z) + Zq(Z) - \int_0^Z q(t) dt - Z - \int_0^Z q(t) dt = \int_0^Z p(t) dt$$

As one can see (and as would be expected), the result coincides with the conditional MTTF inside the interval  $[0, Z]$ .

Example 8.6 Consider a system consisting of  $n$  identical and independent units. The system's behavior can be described with the help of the birth and death process (BDP). The effectiveness of the system during the time interval  $[(t_0 + t_D)]$  is completely determined by the lowest state which the system attains. If the system's lowest state is  $k$ , denote the effectiveness coefficient by  $W_k$  (see Figure 8.4). The problem is to determine B'



(a)

k + 1

(b)

UL

$\frac{i}{1.j} >$

(c)

**Figure 8.4.** Sample of a stochastic track: (a) an observed track without absorbing states; (b) a track absorbed at state  $k + 1$ ; (c) a track absorbed at state  $k$ .

*Solution.* To solve the problem, one should write the BDP equations (see Chapter 1). At moment  $t = 0$ , assume all system units are operating; that is, the system is in state 0. Let  $a_k$  be the transition intensity from state  $k$  to state  $k + 1$  and let  $f_k$  be the transition intensity from state  $k$  to state  $k - 1$ . If we consider a process without an absorbing state, then the system of linear



differential equations is

$$\frac{d}{dt} (a_k + P_k)R_k(t) + \dots = 0 + \sum_{i=k+1}^n R_{k+i}(t)$$

for  $0 < k < n, a_{n+1} = 0$  (8.16)

and the initial condition is  $R_0(0) = 1$ .

To solve the problem, we should solve (8.16)  $n$  times for different absorbing states. Namely, we should solve  $n$  subproblems of type (8.16) for absorbing states  $n, n - 1, \dots, 1$ . Let  $R_k^*(t)$  denote the probability that the process is in the absorbing state when the process is "cut" up to the absorbing state  $k$  (see Figure 8.4). From (8.16) it is clear that  $R^*(t)$  differs from  $R_k(t)$ . Moreover, the sum of the  $R_k^*(0)$ 's over all  $k$  does not equal 1.

We can use the methods described in Chapter 1. But our purpose is not to actually find the above-mentioned probabilities. For further consideration of this particular example, let us assume that we know the probabilities  $R_k(t)$  which are the probabilities of reaching an absorbing state  $k$  in the corresponding subproblem (8.16).

It is clear that if state  $k$  is absorbing,  $R_k^*(t)$  is the probability that in the original state space the process would also reach states with larger subscripts. This means that

$$R^*(t) = \sum_{k=j}^n S_k(t)$$

where  $S_k(t)$  is the probability that the worst state that the initial process reached in  $[0, t]$  is  $k$ . Hence,  $S_k(t) = R_k(t) - R_{k+1}^*(t)$ . The final result is the following:

$$\sum_{k=0}^n w_k S_k(t)$$

**Example 8.7** Consider a system that involves the collection and transmission of information. The system consists of two identical and independent communication channels. If a channel fails, the system capacity decreases to 0.3 times a nominal value. For simplicity, assume that each channel is characterized by an exponentially distributed TTF with parameter  $\lambda$ . Let the duration of a given information collection equal  $(0.1)/\lambda$ . The volume of the collected

information is proportional to the operating time, that is,

$$\begin{aligned}
 W_0 &= t \\
 W_j(x) &= W_2(x) = * + (0.3)(f \sim x) = (0.3)f + (0.7) \text{ or} \\
 W_{j2}(x_1, x_2) &= \min(x_{1t}, x_{2t}) + 0.3[\max(x_{1t}, x_{2t}) - \min(x_{1t}, x_{2t})] \\
 &= 0.3 \max(x_{1t}, x_{2t}) + (0.7) \min(x_{1t}, x_{2t})
 \end{aligned}$$

The task is to calculate expressed via the absolute amount of collected information.

*Solution.* In this case (8.14) can be written in the form

$$\begin{aligned}
 W_{mtCO} - P^2 * + \int_0^t p f'(0.3) t + (0.7)x | e^{-\lambda x} dx \\
 + \int_0^t \lambda e^{-\lambda x} \int_0^x [(0.3)x + (0.7)y] dx dy
 \end{aligned}$$

where we have denoted  $p = p(it) = e^{-\lambda t}$ , After substituting the input data  $p = 0.905$  and  $t = 1/\lambda$ , one obtains the final result

$$W_{\text{sys}} = (0.819 + 0.109 + 0.002)(1/\lambda) \text{ m } 0.94(1/\lambda)$$

This value is the amount of information collected by the system as a whole during a time equal to the MTTF of a single channel.

### 8.4 PARTICULAR CASES

Below we consider several particular cases for which one can obtain simple results. Such kinds of structures are often encountered in practice.

#### 8.4.1 Additive Type of a System Unit's Outcome

We first consider an instantaneous system containing  $n$  independent units. Each of them performs its own task, which implements a determined portion  $W_i$  of the total system outcome. Therefore, the system's outcome  $H'_{\text{sys}}$  can be represented as the sum of the  $W$ -s. Each unit  $i$  can be in one of two states: successfully operating or failure, with probabilities  $p_i$  and  $q_i$ , respectively. For such a system  $W_i$  can be written as

$$\begin{aligned}
 &= \mathbf{E} W_i \\
 & \quad t < i < n
 \end{aligned} \tag{8.17}$$

Expression (8.17) can also be written for dependent units. This follows from

the fact that the expected value of a sum of random values equals the sum of its expected values, regardless of their dependence.

For concreteness, let us consider a system with two types of units. Let us call a unit an *executive unit* if it produces a portion of the system's outcome. All of the remaining units will be called *administrative units*. The system's outcome again consists of the sum of the individual outcomes of its executive units. The coefficient of effectiveness of the  $i$ th executive unit,  $i = 1, \dots, N$ , depends on  $X$ , the state of both the structural and executive units of the system, that is,  $W_i(X)$ ,  $1 < i \leq n$ ,  $N < n$ .

In this case a unit's outcome depends on two factors: the operating state of the unit itself and the state of the system. Finally, we can write

$$\mathbf{E} \sum_{i=1}^N W_i(X) \tag{8.18}$$

where  $E\{W_i(X)\}$  is the unit's average coefficient of effectiveness. (In other words, it is a  $W_i$  of the separate  $i$ th unit.)

$$E\{W_i(X)\} = \mathbf{E} \Pr\{X\} f_i^*(X) \tag{8.19}$$

A clear practical example of such a system can be represented by the so-called nonsymmetrical branching system with a simple treelike hierarchical structure. This system consists of  $N$  executive units controlled by "structural" units at higher hierarchy levels (see Figure 8.5). The total number of hierarchy levels is  $M$ . Each executive unit of the system can produce its outcome if it is operating itself, and if all of its controlling units are also operating. Each executive unit  $i$  has its own outcome  $W_i$  being a portion of

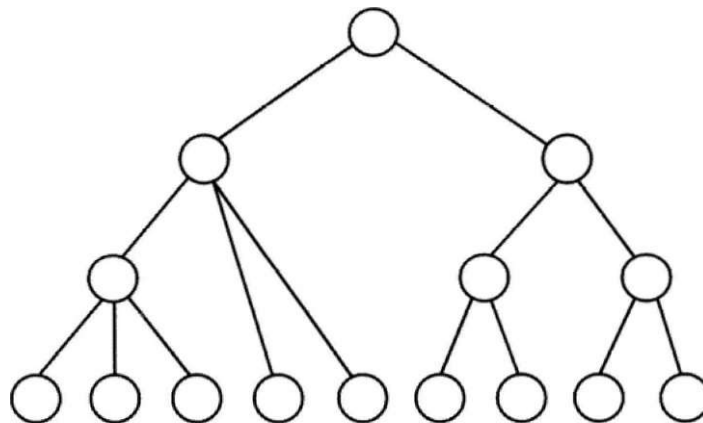


Figure 8.5. Sample of a nonsymmetrical branching system.

the total system outcome, that is,

$$E\{W\}$$

Denote the probability of a successful operation of the highest unit in the system hierarchy by  $p_0$ ; the corresponding probability of the units of the second level controlling the  $i$ th executive unit by  $p_{2i}$ ; the same for the third level,  $p_{3i}$ ; and so on. Thus, the successful operation of the  $i$ th executive unit can occur with probability

$$p_{i,j} \quad (8-20)$$

Now it is easy to calculate the system's effectiveness

$$E \quad (8.21)$$

Again, we use the fact that the mean for the sum of dependent random variables equals the sum of their means.

**Example 8.8** Consider a power supply system whose structure is presented in Figure 8.6. Units 0, 1, and 2 are structural and units 3 to 10 are executive. The outcome of each of them equals the power distributed to consumers (in conditional units). All absolute outcomes of the system units and their availability coefficients are presented in Table 8.3. Find  $W^{\wedge}$  with the condition of independence of the system's units.

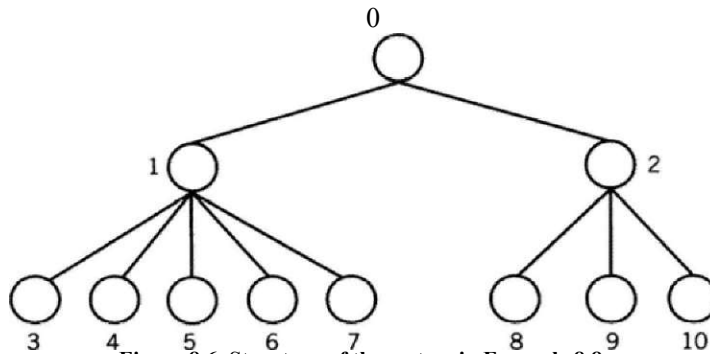


Figure 8.6. Structure of the system in Example 8.8.

TABLE 8.3 Parameters of the System's Units for Example 8.8

Unit	$P_i$	$w_i$	$P/W$
0	0.99	-	
1	0.98	-	
2	0.97	-	
3	0.9	5	4.5
4	0.95	10	9.5
5	0.9	10	9.0
6	0.9	5	4.5
7	0.95	15	14.25
8	0.98	10	9.8
9	0.9	5	4.5
10	0.95	15	14.25

*Solution.* Any executive unit performs its function if it is successfully operating itself, along with the common units and the corresponding units of the second level. Thus, using (8.21), one obtains

$$E = PM + \sum P_i L_i P W_i$$

$$= (0.99) [(0.98)(41.75) + (0.97)(28.55)] = 67.92$$

For an enduring system operating in a time interval  $[t, t + \Delta t]$ , the coefficient of effectiveness for the  $i$ th unit will depend

on the moment of its failure:  $W_i(x | t < x < t + \Delta t)$ . In this case an expression for  $W$  can also be written in a very simple form

$$W_i(x | t < x < t + \Delta t) = \int_0^{\Delta t} P_i (M + t w_i u + \lambda_i u) + \lambda_i V_i(j_c) dF(x)$$

where  $F(x)$  is the distribution of a random time to failure of the  $i$ th unit.

**Example 8.9** Let us consider a spy satellite designed for the collection and transmission of information. This system is unrepairable and can be considered as enduring. The system consists of three communication channels. Their capacities and failure rates correspondingly are:  $V_1 = 100$  Mbps,  $K_1 = 200$  Mbps,  $V_3 = 250$  Mbps, and  $\lambda_1 = 0.0001$  1/hr,  $\lambda_2 = 0.0003$  1/hr, and

**360 ANALYSIS OF PERFORMANCE EFFECTIVENESS**

$A_3 = 0.0004$  1/hr. Find  $W_{\text{sys}}$  (in absolute value) in two forms: (a) the mean capacity of the system as a whole at moment  $t = 1000$ , and (b) the mean volume of transmitted information during 5000 hours.

Solution, (a) For an arbitrary moment  $t$  one can write

$$= E$$

After the substitution of numerical input data

$$1000) - 100e^{-0.1} + 200e^{-0.3} + 250e^{-0.4} = 405.5 \text{ Mbps}$$

(b) If, during the period  $[0, t_a]$  in this example, there was no failure, a channel has collected  $H$  b i t s of information. If a failure has occurred at the moment  $t < t_a$ , then a channel has collected  $Wt$  bits of information. Taking this into account, one can write

$$\gg U(0, t_a) - E \quad - \frac{W}{A} \int_0^{t_a} e^{-\lambda t} dt$$

Note that, since the amount of transmitted information is proportional to  $t_0$ , the total operating time is

Substituting the input data (in the same time dimension), one obtains

$$I(t), \quad (0, 5000)$$

$$0.00001 \frac{100}{200} \quad 0.00004 \quad 3600 = 5.2 \cdot 10^3 \text{ Mbits}$$

$$\frac{250}{0.00003} - 0.39 + \frac{0.78}{0.00003} - 0.86$$

### 8.4.2 Systems with a Symmetrical Branching Structure

Now we consider a system whose structure presents a particular case of the system structure discussed in Example 8.8. The branching structure has a symmetry, which means that each controlling unit controls the same number of units in the lower level. Also, all units of the same hierarchical level have the same reliability characteristics; that is, the system is homogeneous.

### 362 ANALYSIS OF PERFORMANCE EFFECTIVENESS

Now we will consider a more complex measure of system effectiveness: one that depends in a nonlinear way on the number of executive units performing their functions successfully.

The successful performance of a unit of any hierarchy level means that the unit is in an operating state and all of its controlling units are also in



operating states. It is clear that an executive unit will not operate successfully if at least one "structural" unit which controls it has failed.

Note that the executive units of the system are dependent through their common controlling units. Indeed, a failure of any controlling unit leads to the failure of all controlled units at lower levels, including the corresponding executive units. Therefore, a failure of some "structural" unit leads to the stopping of successful operations of the corresponding branch as a whole; that is, the corresponding set of executive units does not produce its outcome. A failure of the highest-level controlling unit leads to an interruption of successful operations at all executive units.

It is understandable that the problem of effectiveness evaluation for dependent executive units is not trivial. We introduce the following notation:

$p_j$  is the probability of a successful operation of a unit in the  $j$ 'th hierarchy level,  $0 \leq j < n$ .

$a_j$  is the "branching power" of the  $(j-1)$ th-level unit which shows how many units of the  $j$ 'th level are controlled by this unit.

$x_j$  is the random number of successfully performing units in the  $j$ 'th hierarchy level.

$N_j$  is the total number of units in the  $j$ th level.

$P_{fixj}$  is the distribution of  $X_j$ .

$W(x_n)$  is the coefficient of the system's effectiveness if  $x_n$  executive units are successfully performing.

(See the explanations in Figure 8.7.)

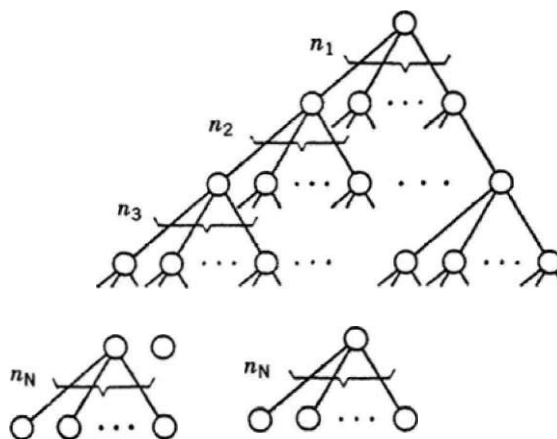


Figure 8.7. General scheme of a system with symmetrical branching structure.

**REMARK.** Here we use the two expressions: "successfully operating" and "successfully performing." Really these expressions have a very slight difference. In this text we will understand that "operating" means that a unit is in an up state itself, independent of the states of any other unit in the system, and "performing" means that the unit is operating itself and, at the same time, all of its controlling units are successfully operating. (See the explanations in Figure 8.7.)

For the system under consideration,

$$WV = Z Pn \{x_n\} W \{x_n\} = E \{W(x_n)\} \tag{8.22}$$

where  $V$ , is the total number of the executive units:

$$K = 11 \tag{8-23}$$

In general, the function  $W(x_n)$  can be arbitrary. For simplicity, let us suppose that  $W(x)$  is a continuous differentiable function of  $x$ . It is known that any such function can be represented in the form of a Taylor series. In the case under consideration,

$$W(x) = Z \sum_{k=0}^{\infty} \frac{d^k W(x_n)}{k!} x^k \tag{8.24}$$

For practical purposes, one can use an approximation taking (8.24) with relatively small  $k$ . Using (8.24), we can easily write

$$WV = E \{ Z \sum_{k=0}^k \frac{d^k W(x_n)}{k!} x^k \} = Z \sum_{k=0}^k \frac{d^k W(x_n)}{k!} M_k \tag{8.25}$$

where  $M_k$  is the moment of the distribution of the number of successfully performing executive units.

To find  $M_k$ , we write the moment generating function. First, consider a group of executive units depending on a single unit in the  $(n - 1)$ th level. We have  $N_{n-1}$  such groups. A random number of successfully operating executive units in a group,  $x$ , has a binomial distribution  $B(a_n, p_n)$ . The moment generating function for the distribution of successfully operating executive units of the above-mentioned group is denoted by

$$g(e^{-x}) = [p_n e^{-x} + q_n]^a \tag{8-26}$$

Now consider all executive units which depend on  $N_{n,x}$  controlling units at the  $(n - 1)$ th level. (At this stage of consideration we are not interested in all

of the remaining units in the system.) The random number of successfully operating units at this level,  $x_n$ , also has a binomial distribution

$$P_{M, X}$$

Note that if no units at the  $(n - 1)$ th level are successfully operating, no units at the  $n$ th level are successfully performing, even though all executive units are operating. This event occurs with probability:

$$P_{n-1}(0) = (1-p)^{n-1}$$

If only one unit at the  $(n - 1)$ th level is operating successfully, then not more than  $a_n$  executive units can perform successfully. The random number of successfully performing executive units will have a binomial distribution with moment generating function (8.26). This event occurs with probability

$$P_{n-1}(1) = \binom{n-1}{1} p^1 (1-p)^{n-2}$$

If two units at the  $(n - 1)$ th level are operating successfully then not more than  $2a_n$  executive units can perform successfully. The probability of this event equals  $P_{n-1}(2)$  where

Arguing in the same manner, we obtain the moment generating function of the distribution of the random number of all successfully performing executive units,  $x_n$ , taking into account the random number of successfully operating units at the  $(n - 1)$ th level as a whole:

$$G_n(e^x) = \sum_{k=0}^{n-1} P_{n-1}(k) G_k(e^x) \quad (8.27)$$

If we let

$$e(Y) = G(t) = [p e^t + q]^n \quad (8-28)$$

then (8.27) can be rewritten as

$$G_n(e^x) = \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} G_k(e^x) \quad (8.29)$$

From (8.28) it follows that

$$(8.30)$$

$$Y^a a_n \ln(p, e^{x''} + q_n)$$

We can continue similar arguments for the units at the remaining upper levels of the system's hierarchy.

Thus, we have the recurrent expression

$$G_{n-1}(e^Y) = G_{n-2}[(p_{n-1} + q_{n-1})e^Y] \quad (8.31)$$

Using the chain rule, we obtain recurrent expressions for the desired initial moments  $M_k$ .

Indeed, the first moment can be found in the following way:

$$M_1 = \int_0^\infty x G_{n-1}(e^x) dx = \int_0^\infty x G_{n-2}[(p_{n-1} + q_{n-1})e^x] dx = \int_0^\infty x G_{n-2}(e^x) dx + \int_0^\infty x G_{n-2}(e^x) dx \quad (8.32)$$

Finally,

$$M_1 = \sum_{i=1}^n p_i M_{i-1} + \sum_{i=1}^n q_i M_i \quad (8.33)$$

Continuing this recurrent procedure, we obtain the result in closed form

$$M_1 = \sum_{i=1}^n p_i M_{i-1} + \sum_{i=1}^n q_i M_i \quad (8.34)$$

We mention that (8.34) could be obtained in a simpler way. Indeed, it directly follows from (8.21) that

$$K \mathbf{ft} P_i = P_0 \mathbf{T1} P_i < *i$$

The second moment of the distribution of the random value  $x_n$  can be found in a similar way.

$$M^2 = \int_0^\infty x^2 G_{n-1}(e^x) dx = \int_0^\infty x^2 G_{n-2}[(p_{n-1} + q_{n-1})e^x] dx = \int_0^\infty x^2 G_{n-2}(e^x) dx + \int_0^\infty x^2 G_{n-2}(e^x) dx \quad (8.35)$$

The recurrent equation for  $M^2$  is

and the final result in closed form is

$$M_{n,2} = \sum_{i=1}^n a_i P_i (n - i + 1) P_i + n q, n \ll * / > * \quad (8.36)$$

Closed-form expressions for higher-order moments are enormously complicated. One is advised to use the above-obtained recurrent expressions for computer calculations.

**Example 8.10** Consider different variants of a branching system (see Figure 8.8). Each system has six executive units. The problem of interest is to choose the best structure for two cases: (a) the system outcome is a given linear function of the number of operating executive units; and (b) the system outcome is a given quadratic function of the number of operating executive units.

*Solution.* For the first case  $W(x_n) = Ax_n$ . Then

$$\hat{M}_{sys} = AM_r = Ap^{p_x a_x} \{p_2 a_2\} - MPoPxPi$$

that is, according to the chosen effectiveness measure, all four variants are equivalent.

If  $W(x_n) = Bx^2$  the effectiveness of any state of the system is proportional to the square of the number of successfully performing executive units. Then

$$\hat{M}_{sys} = E\{1F(*,*)\} = JW_{,,}^2 = 6Bp_0p_1p_2(6p_1p_2 + a_2q_xq_2p_2)$$

In this case the value of the system effectiveness increases as  $a_2$  increases. This means that the variant  $a$  is best for the second type of system effectiveness. For example, such situations appear when one considers Lanchester's models of the second order when the effectiveness of an army division is proportional to the square of the number of its combat personnel.



Figure 8.8. Variants of a system structure with six executive units.

The system with the largest  $a_z$  is the most effective. Thus, the higher the level of centralized control from the center, the better is the result.

**8.4.3 Systems with Redundant Executive Units**

Many instant systems are used for fulfilling a given task. For example, an anti-aircraft or anti-missile defense system is designed to destruct a target. To improve the system's effectiveness,  $N$  redundant executive units can be used. If the system is in state  $X$ , each executive unit fulfills its task with probability  $W_i(X)$ . For example, the efficiency of an anti-aircraft missile system depends on the state of its subsystems which are used for searching, controlling, and so on. There are two main cases: (1) when all executive units perform their common task simultaneously (2) when units perform the same task sequentially.

**Case 1** All units are dependent through the system state  $X$ . Then

$$W = \prod_{i=1}^N W_i(X) \tag{8.37}$$

In particular, for the branching system considered in the previous section, the problem can be solved in the following elegant way. Let  $D$  be the probability of success of a single executive unit (e.g., the kill probability of an enemy's aircraft). Then, if  $x_n$  executive units are acting simultaneously, the total probability of success is

$$W(x_n) = 1 - (1 - D)^{x_n} \tag{8.38}$$

An interesting particular case arises if we consider a symmetrical branching system. As discussed above, the effectiveness of the branching system is completely determined by the number of successfully performing executive units,  $x_n$ . Thus, using (8.2), we can write

$$W(x_n) = \sum_{i=0}^{x_n} C(x_n, i) (1 - D)^{x_n - i} D^i = 1 - G_{x_n}(1 - D) \tag{8.39}$$

The second term in (8.39) is a moment generating function with the substitution of  $1 - D$  as a variable. Thus, (8.39) can be rewritten as

$$W = 1 - G_{x_n}(1 - D) = 1 - G_{x_n}(p + qz) \tag{8.40}$$





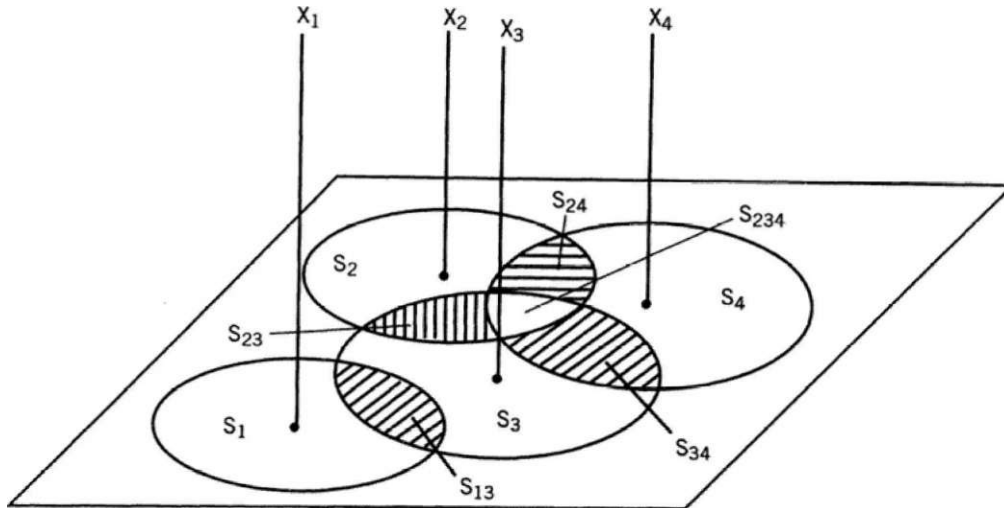


Figure 8.9. Sample of overlapped zones.

Fortunately, in practice, if one considers a territorial system, the number of overlapping zones is usually small enough. On the other hand, if there is a strong overlapping of different zones, the case can be significantly reduced: if all zones are totally overlapping, the system becomes a common redundant system, and its analysis involves no special analytic difficulties.

If we consider a territorial system, the number of units acting in the same zone is not usually large. A zone of the whole system action can be represented as

$$z = \bigcup_{i \in S_n} Z_i$$

For further purposes, let us introduce zones  $Z_a$  which are disjoint. Within each zone  $Z_a$  the same set of serving units is acting. The subscript represents the set of subscripts of the executive units acting in zone  $Z_a$ . Thus,  $i \in c_{ij}$  means that the  $i$ th unit serves in zone  $Z_a$ . Let the number of different zones  $Z_a$  be  $M$ , that is,  $1 < i < M$ . It is clear that zones  $Z_a$  are disjoint and we can write

$$U z_i, \quad 1 < i < M$$

As we mentioned above,  $M \ll 2^M$  in practice.

Because of failures the actual set of units operating in zone  $Z_u$  is random. In general, if includes  $m_y$  subscripts, zone  $Z_0$  can be characterized by  $2^{m_y}$  different possible levels of effectiveness. For each possible set of acting units,

say  $a_{jk}$ , where  $k \geq 2$ , we observe some specified coefficient of effectiveness  $W_{a_i}$ . As a result, for such a system we can write

$$\hat{\Lambda}_{\text{SYST}} = \mathbf{E} \sum_{i=1}^n \mathbf{Z} P_k W_{a_{ik}}. \quad (8.43)$$

Another, more compact representation of (8.43) is

$$\hat{\Lambda}_{\text{SYST}} = \mathbf{E} \sum_{i=1}^M Z_a E \{ W_{a_j} \} \quad (8.44)$$

Such a simple and obvious modification of the general expression (8.2) sometimes allows us to obtain constructive results for some important and interesting practical cases.

### 8.5.2 Additive Coefficient of Effectiveness

In this case for any set of acting units in the zone we have

$$- \mathbf{E} \quad (8.45)$$

I Gay

As an example, we can consider pollution in some region when each of several polluting companies makes its own "investment" in the total level of pollution. Pollution is assumed to be additive. (Of course, in this case it would be more reasonable to speak of loss rather than effectiveness.) It is clear that in this case for the system as a whole

$$\mathbf{E} W_{a_j} = \mathbf{Z} P.ZW \quad (8.46)$$

Let us illustrate this by a simple example.

**Example 8.11** Consider a system consisting of two units and three acting zones (see Figure 8.10). Let us denote

$Z_1 = Z$ ,  $Z_2 =$  the acting zone of the first unit.

$Z_3 = Z_2 + Z_3 =$  the acting zone of the second unit.

$Z_3 =$  the acting zone of both units.

The effectiveness coefficient of the first unit is  $W_1$ , and the effectiveness coefficient of the second unit is  $W_2$ . By assumption of the additive character of the joint effect of the units,  $W_3 = W_1 + W_2$  for zone  $Z_3$ .

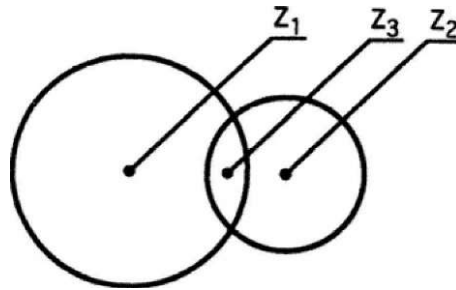


Figure 8.10. Two overlapped zones.

Because all zones  $Z_1$ ,  $Z_2$ , and  $Z_3$  are independent, we can write for the whole system

$$\begin{aligned}
 & P(Z_1 \cap Z_2) = p_1 p_2 (v_1 + w_2) + a_1 p_2 w_2 + p_1 g_2 w_1 \\
 & = P_1 Z_1 + P_2 Z_2 + Z_3 \\
 & = P_1 W_1 Z_1 + P_2 W_2 Z_2
 \end{aligned}$$

### 8.5.3 Multiplicative Coefficient of Effectiveness

In this case for any set of acting units in the zone we have

$$W_{\text{sys}} = W_j$$

It is more natural to consider a loss rather than a "positive" outcome. For example,  $1 - W_j$  is the kill probability of a target in the acting zone of the  $i$ th unit of the system. Thus, if the  $i$ th unit does not act, this probability is 0. In other words, the probability of the enemy's survival (or, more accurately, the loss of the attacker) is  $1 - W_j = 1$ . If a unit acts successfully, the enemy's damage is larger: the probability of the enemy's survival equals some  $W_i < 1$ . If in zone  $Z_a$  units  $i_1, i_2, \dots, i_k$  act together successfully, then the probability of the enemy's survival is

$$W_{\text{sys}} = W_{i_1} * W_{i_2} * \dots * W_{i_k}$$

Taking into account the probability of success of the units, we can write

$$W_{\text{sys}} = \prod_{j=1}^n (p_j W_j) \tag{8.47}$$

If we take into account that  $p_i W_i + q_i = 1 - p_i w_i$ , (8.47) can be rewritten as

$$E Z_{a/nd -PVi} \tag{8-48}$$

The final results (8.47) and (8.48) are illustrated by a simple example.

**Example 8.12** The system under consideration is the same as in Example 8.11 (see Figure 8.10 for an explanation). We use the same notation. To facilitate understanding, keep in mind the case of a target's destruction in the zone of defense. Thus,  $P_i$  is the probability of failure of the  $i$ th executive unit,  $P_i = 1 - q_i$ , and  $W_i$  is the probability of a target's destruction by the  $i$ th executive unit. Assume that the probability of a target's appearance in a zone is proportional to its size. Then the probability of the target passing through the defense zone equals

$$\begin{aligned} W^{\wedge} &= Z_x(W_x p_x q_x) + Z_2(W_2 p_2 + q_2) \\ &\quad + Z_3(p_1 p_2 W_1 W_2 + P_x q_2 w_x + p_2 q_x W_2 + q_x q_2) \\ &= Z_x(W_x p_x + q_x) + Z_2(W_2 p_2 + q_2) + Z_i(p_x W_x + q_x)(p_2 W_2 + q_2) \end{aligned}$$

or, equivalently

$$W^{\wedge} = 1 - w_1 p_1 + Z_2(1 - w_2 p_2) + Z_3(1 - p_1 w_1)(1 - p_2 w_2)$$

**8.5.4 Redundant Coefficient of Effectiveness**

Now we consider a "positive" outcome in a zone. This type of effectiveness coefficient is, in effect, complementary to the one considered in the previous section. For any set of acting units in the zone,

$$W^{\wedge} = 1 - n \cdot v \tag{8.49}$$

where  $v_i = 1 - w_i$  and  $w_i$  has the previous meaning, that is, the probability of success. For this system we obtain

$$E Z_{fl, i - n} \tag{8.50}$$

$$(p_i w_i + q_i)$$

If we again take into account that  $p_i w_i + q_i = 1 - p_i W_i$ , (8.50) can be rewritten as

$$E Z_{fl, i - n} = 1 - p_i W_i \tag{8.51}$$

Again, let us illustrate the final result by a simple example.

**Example 8.13** Consider again the system represented in Figure 8,10. Using the previous arguments, we have

$$\begin{aligned}
 & W^{\wedge} Z_x p_x W_i + Z_2 p_2 W_2 \\
 & + Z_3 [p_1 p_2 (1 - K_{> i} w_z) + p_1 q_2 W_1 + p_2 q_x w_2] \\
 & - Z^{\wedge} - O - p \cdot w O] + z_2 [1 - (1 - p_2 w_2)] \\
 & + Z_3 [1 - (q_x + q_z) (p_2 w_2 + q_2)]
 \end{aligned}$$

**8.5.5 Boolean Coefficient of Effectiveness**

This case is very close to ordinary redundancy applied to each zone. In other words, at least one executive unit must act in a zone to fulfill the operation within that zone. Thus, if  $a_j$  is a set of units acting in the  $j$ th zone and unit  $i$  delivers the outcome  $W_j$  in this zone, then any subset— $a^*$ ,  $a^* \subset a_j$  and  $a^* \neq \emptyset$ —delivers the same effectiveness  $W_r$ . For example, if we consider a communication with a zone, it is sufficient to have at least one path of connection with this zone. For the system as a whole, we have

(8.52)

$$\begin{matrix}
 & i - \mathbf{ru} \\
 \& j s M & i s a j
 \end{matrix}$$

We assume that (8.52) does not demand any additional comments.

**8.5.6 Preferable Maximal Coefficient of Effectiveness**

If a set of units  $a^*$  might act in the  $j$ th zone, then for the actual operation the unit with the maximal possible effectiveness coefficient is chosen

$$W^* = \max W_i \tag{8.53}$$

Enumerate all of the system units in decreasing order of their effectiveness indexes  $IV, : W_1 > W_2 > \dots > W_n$ . Then in the  $j$ th zone the  $k$ th one uses the unit which is characterized by the effectiveness coefficient  $W_k$  (of course,  $k \in a_j$ ) if and only if the  $i$ th unit itself is operational and there are no other operational units with  $i < k$ . This means that all units belonging to the set and having smaller numbers have failed at the moment of use. After this argument, it is simple to write the following expression:

$$- E Z_0 / E W_{k P k} \prod c_h \tag{8.54}$$

$\langle j \rangle M \text{ Area.} \quad \langle k \rangle$

376 ANALYSIS OF PERFORMANCE EFFECTIVENESS

\*  $i^a$

This follows directly from the formulation of the problem. We again think that there is no special need to explain it in more detail.

**8.5.7 Preferable Minimal Coefficient of Effectiveness**

In this case, if a set of units  $a^*$  might act in the  $j$ th zone, then for the actual operation the unit with the minimal possible effectiveness coefficient is chosen

$$W_a = \min_{i \in A_j} W_i \tag{8.55}$$

This kind of effectiveness coefficient can be chosen if one investigates damage rather than a "positive" outcome.

Actually, this case is not distinguished from the previous one. One can even keep formula (8.54) with only one very essential difference: the enumeration of the system's executive units must be done in an increasing order of effectiveness indexes  $W_1 < W_2 < \dots < W_n$ .

**8.6 ASPECTS OF COMPLEX SYSTEMS DECOMPOSITION**

As mentioned above, the problem of effectiveness analysis arises in connection with the analysis of complex systems. Thus, the more complex a system is, the more important and, at the same time, the more difficult is the evaluation of its effectiveness. Thus, the problem of simplifying the evaluation of effectiveness, in particular, the methods of decomposition, seems very important.

Above we considered systems consisting of units with two states: an operating state and an idle state. But one sometimes deals with complex systems consisting of many such subsystems which themselves can be considered as complex systems. This is equivalent to the consideration of a system consisting of units with more than two states.

Let  $n$  be the total number of system units. Suppose the system is divided into  $M$  subsystems by some rule (it can be a functional principle or a constructive one). Each  $j$ th subsystem includes  $n_j$  units and, consequently, has

$$n_j = 2^{n_j}$$

different states. Now the system consists of  $M$  new units, each with  $n_j > 2$  states. Of course, such a system representation does not lead to a decrease in the total number of system states  $n$  that is, it does not follow that

378 ANALYSIS OF PERFORMANCE EFFECTIVENESS

$$m = Y_i 2'' \text{ } \text{ } 2''$$

I sis *M*

But such a new system representation may still help to generate new ideas.



First, it may be possible to characterize subsystems via some simpler description. For example, we can find one main characterization parameter for the entire system. In this case the dimension of the problem could be essentially decreased.

Second, it may be possible to use a simpler description of the states of the subsystems in comparison with complete enumeration. In this case the number of subsystems  $M$  is usually not very large. The second case leads to the construction of upper and lower bounds on the system effectiveness index  $H_{ysl}$ . We will consider it in the next section.

### 8.6.1 Simplest Cases of Decomposition

It would be very constructive to represent a system's effectiveness index as a function of the  $W$ 's of its subsystems. Is this ever possible? If so, when? The problem is to present the system's effectiveness as a function of the subsystem's effectiveness:

Assume that for any system state  $a^*$ , which is expressed as a composition of subsystem states  $a^*$ , that is,  $a^* = (a^*_1, \dots, a^*_M)$ , the condition

$$\frac{EW_{AI}}{ISJSW} \tag{8.57}$$

is true. Then, for such an additive system, (8.56) can be written as

$$W_{SYST} = \sum_1 E_{Jf} \tag{8.58}$$

The statement is clear if one remembers that the mean of a linear function equals the function of the mean values of its variables. Thus, if it is possible to choose subsystems in such a way that (8.58) holds, we can use the simple expression (8.57).

The next results can be formulated for multiplicative systems. Note that for multiplicative systems (8.56) can be written as

$$\prod_{1 \leq j \leq M} n_{ty} \tag{8.59}$$

if and only if for any system state  $a^*$  which is expressed as a composition of subsystem states  $a^*$ , that is,  $a^* = (a^*_1, \dots, a^*_M)$ , the following condition is valid:

**(8.60)**

$n_K^*$

1 ZJZM

Expression (8.60) means that the  $W$  of any subsystem does not depend on the states of other subsystems.

Statement (8.59) becomes clear if one remembers that the mean of the product of independent random variables equals the product of the means of its variables. Thus, if subsystems are chosen in such a way that (8.60) holds, we can use the simple expression (8.59).

Unfortunately, the number of practical examples where we may obtain such a fantastic bargain in the evaluation of performance effectiveness is exhausted by these two trivial cases. Also, unfortunately, such systems are quite rare in engineering practice.

Fortunately, however, a similar approach can be used to obtain bounds of a system's effectiveness index in the case of regional systems.

### 8.6.2 Bounds for Regional Systems

1. Consider a regional system with a multiplicative effectiveness coefficient in a zone. Let us consider a zone with a set of executive units  $A$ . Assume that the system is divided into  $M$  subsystems. In this case the units of the set  $A$  can belong to several different subsystems. This means that the set  $A$  can be divided into several nonintersecting subsets  $A_j$ ,  $1 \leq j \leq M$ . (Some of the  $A_j$  can be empty.) If the  $H^i$ 's are normalized effectiveness coefficients, that is, if  $0 < (W^i q) < 1$ , then, for any  $A$ ,

$$n(W_{iP_i} + q_i) < Z \prod_{j=1}^M (W_{iP_i} + q_i) \quad (8.61)$$

From (8.61) it immediately follows that for these systems

$$\hat{W}_{\text{sys}} < Z \prod_{i=1}^M W_i \quad (8-62)$$

2. For systems with a redundancy type of effectiveness coefficient, we have

$$\lambda_{\text{sys}} < Z W, \quad (8.63)$$

To confirm (8.63), we show that this is correct for a zone with two acting units belonging to different subsystems. Keeping in mind that  $0 < W_i < 1, i = 1, 2$ , we can easily write

$$1 - (p_x w_x + q_1)(p_2 w_2 + q_2) < p_x W_y + p_2 W_2 \quad (8.64)$$

3. For systems with a Boolean type of effectiveness coefficient, one can obviously write

$$K^{**} E^{W_j} \quad (8-65)$$

$i s j s M$

4. For systems in which one chooses for operation the unit with the maximal effectiveness coefficient in a zone, we have

$$E_{P_k W_k} \mathbf{n} L E_{P_k W_k} \quad (8.66)$$

$k e a \quad K k, i <= a \quad k \& d j$

Expression (8.66) is clear as any product of  $\wedge$ 's is always less than 1.

5. For systems in which one chooses for operation the unit with the minimal effectiveness coefficient in a zone, we have absolutely the same result (8.66).

Unfortunately, we have obtained only one-sided bounds. From a practical point of view, a lower bound of any "positive" effectiveness index (the larger, the better) is reasonable: one has a guaranteed result. But all bounds considered here yield a restriction on the upper side.

### 8.6.3 Hierarchical Decomposition and Bounds

**Using Subsystem  $W_i$ 's** Let the system be represented as a composition of  $M$  subsystems. For each subsystem one can calculate its own effectiveness  $W_i$ . Let the  $j$ th subsystem include  $r_{ij}$  units. This subsystem has, in general,

$$m_j = 2^{r_{ij}}$$

different states. Thus, we must analyze  $m_j$  different states for each subsystem.

If it is possible to express a system's effectiveness index  $W_{\text{sys}}$  as a function of the  $W_i$ 's of the subsystems in the form

$$W_{\text{sys}} = f_i(W_i) \quad (8.67)$$

then it is enough to calculate  $W_{\text{sys}}$  as we did before,

$$V = E r\{X_{ji}\} W_{it} \quad (8.68)$$

68)  $1 \wedge i \wedge m j$

and, after this, use (8.67).

The total number of computations to obtain the desired result is proportional to

$$\sum_{j=1}^M 2^j \ll 2^M \quad (8.69)$$

Unfortunately, such a procedure cannot be used too frequently as functions such as (8.67) are seldom known. The two simplest examples were shown above. At any rate, this method allows one to obtain at least some rough estimates of the unknown value of  $W^*$ .

Let us give several examples of the effectiveness of such a decomposition. For a system consisting of  $n = 400$  units, a strict evaluation of the system's effectiveness is practically impossible because the number of all possible system states exceeds the so-called googol ( $10^{100}$ ). As the reader may know, the googol is sometimes jokingly called "the greatest number in the universe." Indeed, everything in the universe—its maximal diameter, its time of existence since the Big Bang, its total number of smallest elementary particles—measured by the smallest physical units (length or time, respectively) is smaller than this definitely restricted number. Thus, any attempt just to enumerate all the states of the above-mentioned system is unrealistic.

But if the system is divided into 20 subsystems, each consisting of 20 units, the number of calculations will still be large— $20 \cdot 20^{20} = 2 \cdot 10^7$ —but at least it is a realistic number. If it is possible to divide the system into 40 subsystems, the corresponding number equals  $40 \cdot 2^{40} \sim 4 \cdot 10^4$ , which is unconditionally acceptable.

**REMARK.** We mention that very complex systems are usually considered in engineering practice in a hierarchical way with more than two levels. This permits one to independently analyze first the system as a whole; then each subsystem as a part of the system, but performing its own functions; then some more-or-less autonomous parts of these subsystems; and so on. Such a mode is very effective in the evaluation of a system's effectiveness.

Let us again consider a system consisting of  $n = 400$  units. Suppose the system is divided into 5 subsystems, each subsystem is divided into 5 autonomous parts, where each part consists of 16 units. Thus, the total number of calculations can be evaluated as the total number of parts in the system, multiplied by the number of calculations for one such part:  $5 \cdot 5 \cdot (2^{16}) = 200,000$ . It is significantly less than in the initial case.

It is interesting to note that, for a system consisting of  $n = 1000$  units, one can obtain an even smaller number of calculations if the system is repre-

### 386 ANALYSIS OF PERFORMANCE EFFECTIVENESS

sented by a three-level hierarchy: 5 subsystems, each of 5 parts, each of 5 complex units, each of 8 units of the lowest level. The number of calculations required is equal to  $5 \cdot 5 \cdot 5 \cdot (2^8) = 32,000$ .



**Distributions of Subsystem Levels of  $W$**  A more accurate method than the previous one is described next. From the viewpoint of a system's user, all subsystem states can be divided into a very restricted number of groups. The states of each such group are characterized by a value close to the value of the subsystem's effectiveness coefficient  $W_j$ . It is clear that the number of such groups could be very small, say 10. This number does not depend on the initial number of subsystem states. (One has also to take into account that there is no necessity to consider groups with levels of effectiveness which appear with an infinitesimally small probability and/or with levels of one essentially negligible effectiveness.)

At any rate, the first step in the analysis of a system's effectiveness consists in a detailed analysis of each subsystem. For each subsystem  $j$ , we need to analyze all possible states  $XI < i < 2^m$ . Also, for each such subsystem we need to choose a reasonable lattice of the effectiveness coefficient values. Assume this lattice has  $K_j$  different cells:

- The first cell includes those states whose effectiveness coefficients  $W_{ij}$  satisfy the condition  $1 = W^{\wedge} < W_j < B_1$ , where  $B_1$  is the first threshold of the lattice; for all states belonging to this cell of the lattice, one computes the total probability  $R_j$  as the sum of the probabilities of all states whose effectiveness values are included in this cell.
- The second cell includes those states whose effectiveness coefficients  $W_{ij}$  satisfy the condition  $B_1 < W_j < B_2$ , where  $B_2$  is the second threshold of the lattice; the corresponding probability computed is  $R_2$ .
- The  $K_j$ th cell includes those states whose effectiveness coefficients  $W_{ij}$  satisfy the condition  $B_{(K_j - 1)} < W_j < 0$ , where  $B_K$  is the last threshold of the lattice; the corresponding probability computed is  $R_{K_j}$ .

We may now analyze the system as a whole. In each cell of the lattice, we choose a "middle" state which corresponds to the average value of  $W_j$ . For future analysis, this state now becomes a "representative" of all of the remaining states related to this cell. Thus, we choose  $K_j$  representatives for each subsystem. We should choose an appropriate number of representatives, say  $X^*$ . Each of them appears with probability  $R_{j,r}$ . The number of representatives is determined with respect to the required accuracy of the analysis and the available computer capacity.

After these preliminary steps we consider

$$K - N \quad (8.70)$$

different system states and, for each of them, evaluate the effectiveness coefficient. We then consider all  $K$  system states

388 ANALYSIS OF PERFORMANCE EFFECTIVENESS

$$X = (J^*; 1 < j < M) = (X^*_r, X^*_{r+1}, \dots, X^*_J)$$

and write the expression for  $W_{\text{sys}}$  with the use of (8.62)

$$PK1 = \sum_{\text{alt } X} W(X_{t_h}, \dots, X_{t_j}) \prod_{i=1}^n (1 - \alpha_i)^{m_i} \quad (8.71)$$

This expression is not too pleasant in visual form because of the notation used. Neither is it easy to compute. But its nature is simple and completely coincides with (8.61).

Of course, if we decide to distinguish several levels of a system's hierarchy, the methodology would be the same but the corresponding description, in a general form, would be even longer than (8.71). We would like to emphasize that a hierarchical model needs less computation.

This method of representative selection can be successfully used for obtaining lower and upper bounds,

1. Let us choose from among the states of the lattice cell a state with a minimal effectiveness coefficient and consider this state as a representative of this cell. Denote this aggregate state by  $Xp^{\text{TM}}$ . If we substitute  $Ay^{\text{TM}}$  instead of  $X^*$  in (8.71), we will obtain a lower bound for the system index,
2. If a state  $XJ^{\text{max}}$  with a maximal effectiveness coefficient is chosen as the representative of the cells, then the same procedure gives us an upper bound for  $W^{\text{max}}$ .

Thus, we obtain two-sided bounds for  $W^{\text{max}}$ :

$$W^{\text{TM}} < \leq W^{\text{max}} \quad (8.72)$$

In general, for practical purposes, it is enough to have an approximate expression (8.71). We should emphasize that reliability (and also effectiveness performance) computations are usually provided not for a precise evaluation of different indexes, but usually for a comparison of competitive variants at some design stage. For such purposes, we may use an approximate solution as a direction for design.

### 8.7 PRACTICAL RECOMMENDATION

An analysis of the performance effectiveness of a system must be carried out by a researcher who deeply comprehends the system as a whole, knows its operation, and understands all demands on the system. It is a necessary condition of successful analysis. Of course, the systems analyst should also be acquainted with operations research methods. As with any operations re-

### 390 ANALYSIS OF PERFORMANCE EFFECTIVENESS

search problem, the task is concrete and its solution is more of an art than a science.

For simplicity of discussion, we demonstrate the effectiveness analysis methodology referring to an instant system. The procedure of a system's effectiveness evaluation, roughly speaking, consists of the following tasks:

- A formulation of an understandable and clear goal of the system.
- A determination of all possible system's tasks (operations, functions).
- A choice of the most appropriate measure of system effectiveness.
- A division of a complex system into subsystems.
- \* A compilation of a structural-functional scheme of the system which reflects the interaction of the system's subsystems.
- A collection of reliability data.
- \* A computation of the probabilities of the different states in the system and its subsystems.
- \* An estimation of the effectiveness coefficients of different states.
- \* A performance of the final computations of the system's effectiveness.

Of course, the effectiveness analysis methodology of enduring systems is quite similar, with the exception of some terms.

We need to remark that the most difficult part of an effectiveness analysis is the evaluation of the coefficients of effectiveness for different system states, in only extremely rare cases is it possible to find these coefficients by means of analytical approaches. At any rate, in the initial stages of a system's design there is no other way. The most common method is to simulate the system with the help of a computerized model or a physical analogue of the system. In the latter case, the analyst introduces different failures at appropriate moments into the system and analyzes the consequences. The last and the most reliable method is to perform experiments with the real system or, at least, with a prototype of the system.

Of course, one has to realize that usually all of these experiments set up to evaluate effectiveness coefficients are very difficult and they demand much time, money, and other resources. Consequently, one has to consider how to perform only really necessary experiments. This means that a prior evaluation of different state probabilities is essential: there is no need to analyze extremely rare events.

One can see that the analysis of a system's effectiveness performance is not routine. Designing a mathematical model of a complex system is, in some sense, a problem similar to the problem of designing a system itself. Of

**392 ANALYSIS OF PERFORMANCE EFFECTIVENESS**

course, there are no technological difficulties—no time or expense for engineering design and production.

It seems that the first paper devoted to the problem discussed in this chapter was the paper by Kolmogorov (1945). This work focused on an effectiveness measure of anti-aircraft fire. The total kill probability of an enemy's aircraft was investigated. The random nature of the destruction of different parts of an aircraft and the importance of these parts was assumed. It is clear that from a methodological viewpoint the problem of system effectiveness analysis is quite similar: one has only to change slightly the terminology.

The first papers concerning a system's effectiveness evaluation appeared in the early 1960s [see, e.g., Ushakov (1960, 1966, 1967)]. Some special cases of system effectiveness evaluation were considered in Ushakov (1985, 1994) and Netes (1980, 1984).

One can find an analysis of the effectiveness of symmetrical branching systems in Ushakov (1985, 1994) and Ushakov and Konyonkov (1964). Territorial (regional) systems with intersecting zones of action were studied in Ushakov (1985, 1994). Here one can also find an analysis of decomposition methods. The general methodology and methods of system effectiveness analysis are described in Ushakov (1985, 1994).

## REFERENCES

- Kolmogorov, A. N. (1945). A number of target hits by several shots and general principles of effectiveness of gun-fire (in Russian). *Proc. Moscow Inst. Math.*, Issue 12.
- Netes, V. A. (1980). Expected value of effectiveness of discrete system (in Russian). *Automat. Comput. Sci.*, no. 11.
- Netes, V. A. (1984). Decomposition of complex systems for effectiveness evaluation. *Engrg. Cybernet. (USA)*, vol. 22, no. 4.
- Ushakov, I. A. (1960). An estimate of effectiveness of complex systems. In *Reliability of Radioelectronic Equipment* (in Russian). Moscow: Sovetskoye Radio.
- Ushakov, I. A. (1966). Performance effectiveness of complex systems. In *On Reliability of Complex Technical Systems* (in Russian). Moscow: Sovetskoye Radio.
- Ushakov, I. A. (1967). On reliability of performance of hierarchical branching systems with different executive units, *Engrg. Cybernet. (USA)*, vol. 5, no. 5.
- Ushakov, I. A., ed. (1985). *Reliability of Technical Systems: Handbook* (in Russian). Moscow: Radio i Sviaz.
- Ushakov, I. A., ed. (1994). *Handbook of Reliability Engineering*. New York: Wiley.
- Ushakov, I. A., and Yu. K. Konyonkov (1964). Evaluation of effectiveness of complex branching systems with respect to their reliability. In *Cybernetics in Service for Communism* (in Russian), A. Berg, N. Bruevich, and B. Gnedenko, eds. Moscow: Nauka.

## EXERCISES

- 8.1 A conveyor system consists of two lines, each producing  $N$  items per hour. Each of the lines has an availability coefficient  $\kappa = 0.8$ . When one of the lines has failed, the other decreases its productivity to  $0.7N$  because of some technological demands. There is a suggestion to replace this system with a new one consisting of one line with a productivity of  $1.1N$  items per hour and an availability coefficient  $K_t = 0.9$ . Is this replacement reasonable from an effective productivity viewpoint or not?
- 8.2 A branching system has one main unit and three executive ones. There are two possibilities: (1) to use a main unit with PFFO  $p_0 = 0.9$  and an executive unit with PFFO  $p_1 = 0.8$  or (2) to use a main unit with PFFO  $p_0 = 0.8$  and an executive unit with PFFO  $p_1 = 0.9$ . Is there a difference between these two variants if the system's effectiveness depends on (a) the average number of successfully operating executive units, (b) a successful operation at least one executive unit, and (c) a successful operation of all executive units?

## SOLUTIONS

8.1 The old system of two lines has the following states:

- Both lines operate successfully. In this case the effective productivity of the system is  $2N$ . This state occurs with probability  $P = (0.8 \times 0.8) = 0.64$ .
- One line has failed and the other is operating. This state occurs with probability  $P = 2(0.8 \times 0.2) = 0.32$ . During these periods the system productivity equals  $0.7N$ .
- \* Both lines have failed. This state occurs with probability  $P = (0.2 \times 0.2) = 0.04$ . The system productivity obviously equals 0. Thus, the total average productivity of the old system can be evaluated as

$$W_M = (0.64)(2.0) + (0.32)(0.7) = 1.5$$

The new system has an average effective productivity equal to

$$H^w = (1.7)(0.9) = 1.53$$



**SOLUTIONS 395**

Thus, the average productivity of both systems is very close. The increase in productivity is about 1.5%. One has to solve this problem

taking into account expenses for installation of the new system, on the one hand, and the potential decrease in the cost of repair, on the other hand (the new system will fail less often).

- (a) The average number of successfully operating executive units depends only on the product  $p_n p_r$ , so both systems are equivalent.
- (b) For (1) one has  $W_{\text{sys}} = (0.9 \times 1 - 0.8)^3 \ll 0.898$  and for (2)  $W_{\text{sys}} = (0.8 \times 1 - 0.9)^3 = 0.799$ . Thus, the first variant is more effective.
- (c) For (1) one has  $W_{\text{sys}} = (0.9 \times 0.8)^3 * 0.460$  and for (2)  $W_{\text{sys}} = (0.8 \times 0.9)^3 \gg 0.584$ . In this case the second variant is more effective.

## CHAPTER 9

### TWO-POLE NETWORKS

Above we considered systems with a so-called "reducible structure." These are series, parallel, and various kinds of mixtures of series and parallel connections. As mentioned, they are two-pole structures which can be reduced, with the help of a simple routine, into a single equivalent unit. However, not all systems can be described in such a simple way.

We would like to emphasize that most existing networks, for example, communication and computer networks, transportation systems, gas and oil pipelines, electric power systems, and others, have a structure which cannot be described in terms of reducible structures, even if they are considered as two-pole networks.

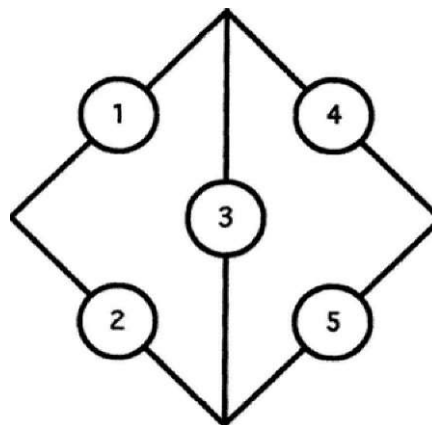


Figure 9.1. Bridge structure.

The simplest example of a system with a nonreducible structure is the so-called *bridge structure* (see Figure 9.1). This particular structure is probably not of great practical importance, but it is reasonable to consider it in order to demonstrate the main methods of analysis of such kinds of structures.

### 9.1 RIGID COMPUTATIONAL METHODS

#### 9.1.1 Method of Direct Enumeration

The bridge structure cannot be represented as a connection of parallel-series or series-parallel subsystems of independent units (links). For this system the structure function  $tp(X)$ , where  $X = (x_1, x_2, x_3, x_4, x_5)$ , can be written in tabular form (see Table 9.1) where all possible system states and corresponding structure function values are presented.

Because each Boolean variable has two possible different values, 0 or 1, the system can be characterized by  $2^5 = 32$  different states. In Table 9.1 we enumerate all possible values of the variables  $x_1, x_2, \dots, x_5$  and denote them as  $X_1, X_2, \dots, X_{32}$ . Some  $X_k$ 's are states of successful operation of the bridge system (the set G) and some of them are not (the set  $\bar{G}$ ). In this notation the structure function of the bridge system can be written as

$$p(x_1, \dots, x_5) = p(X_1) \cup p(X_2) \cup \dots \cup p(X_{32}) = \bigcup p(X_k) \quad (9.1)$$

The probability of a system's successful operation is

$$P\{p(x_1, \dots, x_5) = 1\} = E\{\bigcup_{X_k \in G} p(X_k)\} = \sum_{X_k \in G} P\{X_k\} \quad (9.2)$$

Each vector  $X_k$  can be expressed through its component  $x$ 's and  $\bar{x}$ 's. For example (see Table 9.1),

$$X_5 = (\bar{x}_1, x_2, \bar{x}_3, x_4, x_5)$$

From Table 9.1 it follows that the vector  $X_5$  belongs to G, and so it will be taken into account in (9.2). Then

$$p(X_5) = x_2 x_4 x_5$$

and

$$E\{p(X_5)\} = E\{x_2 x_4 x_5\} = E\{x_2\} E\{x_4\} E\{x_5\} = p_2 p_4 p_5$$

We do not write the detailed expression for  $p(X)$  here. This can be easily obtained from Table 9.1 by taking into account that the corresponding term

**TABLE 9.1 Description of the Structure Function of the Bridge Structure**

States of units		Vector			Value
$x$		*4	*5	**	$fW$
1	1	1	1	1	1
0	1	1	1	1	1
1	0	1	1	1	1
1	1	0	1	1	1
1	1	1	0	1	1
1	1	1	1	0	1
0	0	t	1	1	0
0	1	0	1	1	1
0	1	1	0	1	1
0	1	1	1	0	1
1	0	0	1	1	1
1	0	1	0	1	1
1	0	1	1	0	1
1	1	0	0	1	1
1	1	0	1	0	1
1	1	1	0	0	0
0	0	0	1	1	0
0	0	1	0	1	0
0	0	1	1	0	0
0	1	0	0	1	$X_{19}$
0	1	0	0	1	$\wedge 20$
0	1	0	1	0	0
0	1	1	0	0	$X_{22}$
1	0	0	0	1	$\wedge 23$
1	0	0	1	0	$\wedge 24$
1	0	1	0	0	$\wedge 25$
1	1	0	0	0	0
0	0	0	0	1	$X_{21}$
0	0	0	1	0	$\wedge 28$
0	0	1	0	0	$X_{29}$
0	1	0	0	0	$X_{30}$
1	0	0	0	0	0
0	0	0	0	0	$x_{32}$

1. Both the redundant group and the SD have not failed during a specified interval of time  $t$ .
2. The first unit chosen at random fails at some moment  $x < f$ , the SD performs a successful switch to one of the operating units of the remaining redundant group of  $m - 1$  units, and the new system performs successfully up to time  $f$ .
  1. The first unit operates successfully.
  2. After its failure there is a group of randomly chosen redundant units with operating SDs; this new system operates successfully during the remaining time.
2. After its failure at some moment  $x$ , there is a group of redundant units. The size of this group is random because some of them might have failed before the moment  $f$ . Let the number of operating redundant units at the moment  $x$  equal  $j$ . In some order we try to switch each of

these  $j$  operating units to the main position until a first successful switching occurs. The number of attempts before a success is distributed geometrically with parameter  $R$ . After  $k$  SDs have failed during switching, a successful attempt occurs ( $k$  is random). This means

$$A_0M_2 + M_1M_2 + \dots + A_0^i A_i$$

- $M$  is the operational state of the main unit.
- $\bar{M}$  is the failure state of the main unit.
- $R$  is the operational state of the redundant unit.
- $\bar{R}$  is the failure state of the redundant unit.
- $S$  is the operational state of the switch.
- $\bar{S}$  is the failure state of the switch.
- $\lambda$  is the failure rate of the main unit.
- $\lambda_s$  is the failure rate of the switch.
- $\lambda$  is the failure rate of the redundant unit.
- $\mu$  is the intensity of repair of a single unit.
- $\mu_s$  is the intensity of repair of the switch.
- $\mu^*$  is the intensity of repair of the system as a whole.
- $M$  is the operational state of the main unit.
- $M^*$  is the "hidden failure" state of the main unit.
- $\bar{M}$  is the failure state of the main unit.
- $R$  is the operational state of the redundant unit.
- $\bar{R}$  is the failure state of the redundant unit,
- $\lambda$  is the failure rate of the nonmonitored part of the main unit.
- $\lambda_1$  is the failure rate of the monitored part of the main unit.
- $\lambda$  is the failure rate of the redundant unit.
- $\lambda_s$  is the failure rate of the switch.
- $\mu$  is the intensity of repair of a single unit.
- $\mu_s$  is the intensity of repair of the switch.
- $\mu^* > \mu$  is the intensity of repair of the system as a whole.
- $\nu$  is the intensity of periodical tests of the main unit.

of type  $E\{p(J^{\wedge}A)\}$  has  $p_i$  for  $x_i = 1$  and  $q_i$  for  $x_i = 0$ . Based on Table 9.1, the following equation can be written:

$$E\{p^*(*)\} = E\{p^*(*)_1\} + E\{p^*(*)_2\} + \dots + E\{p^*(*)_{32}\}$$

Omitting intermediate results, we give the final formula for the connectivity probability (in the case of identical units) in two equivalent forms

$$E\{p^*(*)\} = p^5 - 5p^4 + 2p^3 + 2p^2 \tag{9.3}$$

$$E\{p(A')\} = 1 - 2q^2 - 2q^3 + 5q^4 - 2q^5 \tag{9.4}$$