
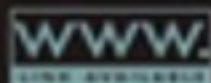


Igor Ushakov

PROBABILISTIC RELIABILITY MODELS

 WILEY

 www.linuxserver.org

Igor Ushakov

Probabilistic Reliability Models

Contents

1	Preface	6
2	Acronyms and notations	7
3	What is reliability?	9
3.1	Reliability as a property of technical objects	9
3.2	Other “ilities”	9
3.3	Hierarchical levels of analyzed objects.	12
3.4	How can reliability be measured?	12
3.5	Software reliability.....	13
3.5.1	Case Study: Software failures avalanche.....	14
4	Unrecoverable Objects	15
4.1	Unit	15
4.1.1	Probability of failure-free operation.....	15
4.1.2	Mean time to failure.....	16
4.2	Series systems.....	16
4.2.1	Probability of failure-free operation.....	16
4.2.2	Mean time to failure.....	18
4.3	Parallel System	19
4.3.1	Probability of failure-free operation.....	19
4.3.2	Mean time to failure.....	23
4.4	Structure of type « <i>k-out-of-n</i> »	25
4.5	Realistic models of loaded redundancy.....	27
4.5.1	Unreliable switching process.....	27
4.5.2	Non-instant switching.....	28
4.5.3	Unreliable switch.....	28
4.5.4	Switch serving as interface.....	30
4.5.5	Incomplete monitoring of operating unit.....	31
4.5.6	Periodical monitoring of the operating unit.....	32
4.6	Reducible structures.....	33
4.6.1	Parallel-series and series-parallel structures.....	33
4.6.2	General case of reducible structures.....	34
4.7	Standby Redundancy	35
4.7.1	Simple Redundant Group.....	35
4.7.2	Standby redundancy of type “ <i>k-out-of-n</i> ”	37
4.8	Realistic models of unloaded redundancy	38
4.8.1	Unreliable switching process.....	38
4.8.2	Non-instant switching.....	39
4.8.3	Unreliable switch.....	39
4.8.4	Switch serving as interface.....	41
4.8.5	Incomplete monitoring of the operating unit.....	42
5	Recoverable systems. Markov models	45
5.1	Unit	45

5.1.1	Markov Model	45
5.2	Series System	51
5.2.1	Turning-off system during recovery	52
5.2.2	System in poperating state during recovery. Unrestricted repair.	54
5.2.3	System in operating state during recovery. Restricted repair. 56	
5.3	Dubbed system	57
5.3.1	General description	58
5.3.2	Non-stationary availability coefficient	59
5.3.3	Stationary availability coefficient	62
5.3.4	Probability of failure-free operation	63
5.3.5	Stationary coefficient of interval availability	65
5.3.6	Mean time to failure	66
5.3.7	Mean time between failures	67
5.3.8	Mean recovery time	68
5.4	Parallel systems	69
5.5	Structures of type “ m out of n ”	70
6	Recoverable systems. Heuristic models	74
6.1	Preliminary notes	74
6.2	Poisson process	77
6.3	Procedures over Poisson processes	79
6.4	Asymptotic thinning procedure over stochastic point process	81
6.5	Asymptotic superposition of stochastic point processes	83
6.6	Intersection of flows of narrow impulses	85
6.7	Heuristic method for reliability analysis of series recoverable systems	87
6.8	Heuristic method for reliability analysis of parallel recoverable systems	88
6.8.1	Influence of unreliable switching procedure	89
6.8.2	Influence of switch’s unreliability	90
6.8.3	Periodical monitoring of operating unit	91
6.8.4	Partial monitoring of operating unit	92
6.9	Brief historical overview and related sourcews	93
7	Time redundancy	95
7.1	System with possibility of restarting operation	95
7.2	Systems with “admissibly short failures”	97
7.3	Systems with a Time Accumulation	98
7.4	Case study: Gas-pipeline with an underground storage	99
7.5	Brief historical overview and related sourcews	101
8	“Aging” units and systems of “aging” units	102
8.1	Chebyshev bound	102
8.2	“Aging” unit	102
8.3	Bounds for probability of failure-free operations	103
8.4	Series system consisting of “aging” units	106
8.4.1	Preliminary Lemma	106
8.5	Series system	107

8.5.1	Probability of failure-free operation	107
8.5.2	Mean time to failure of series system	110
8.6	Parallel system	111
8.6.1	Probability of failure-free operation	111
8.6.2	Mean time to failure	114
8.7	Bounds for the coefficient of operational availability	116
8.8	Brief historical overview and related sourcecs	117
9	Two-Pole Networks	119
9.1	General comments	119
9.1.1	Method of Direct Enumeration	121
9.2	Method of Boolean Function Decomposition	123
9.3	Method of Paths and Cuts	125
9.3.1	Esary-Proschan Bounds	125
9.3.2	"Improvements" of Esary-Proschan bounds	127
9.3.3	Litvak-Ushakov Bounds	129
9.3.4	Comparison of the Two Methods	133
9.4	Brief historical overview and related sourcecs	134
10	. Performance Effectiveness	136
10.1	Effectiveness Concepts	136
10.2	General idea of effectiveness evaluation	137
10.2.1	Conditional case study: Airport traffic control system. 139	
10.3	Additive Type of System's Units Outcome	141
10.4	Case study: ICBM control system	143
10.5	Systems with intersecting zones of action	144
10.6	Practical Recommendation	149
10.7	Brief historical overview and related sourcecs	150
11	On system survivability	152
11.1	Brief historical overview and related sources	156
12	Multistate systems	158
12.1	Preliminary notes	158
12.2	Generating Function	158
12.3	Universal Generating Function	161
12.4	Multistate series system	162
12.4.1	Series connection of piping runs	163
12.4.2	Series connection of resistors	164
12.4.3	Series connections of capacitors	166
12.5	Multistate parallel system	168
12.5.1	Parallel connection of piping runs	168
12.5.2	Parallel connection of resistors	169
12.5.3	Parallel connections of capacitors	169
12.6	Reducible systems	170
12.7	Conclusion	176
12.8	Brief historical overview and related sources	176
13	180

Appendicis	180
A. Main distributions related to Reliability Theory	180
A1. Discrete distributions	180
A.1.1. Degenerate distribution	180
A1.2. Bernoulli distriburion	181
A.1.3. Binomial distribution	181
A.1.4. Poisson distribution	183
A.1.5. Geometric distribution	185
A.2. Continuous distributions	186
A.2.1. Intensity function	186
A.2.2. Continuous uniform distribution	187
A.2.3. Exponential distribution	187
A.2.5. Erlang distribution	189
A.2.6. Hyper-exponential distribution	189
A.2.7. Normal distribution	191
A.2.8. Weibull-Gnedenko distribution	191
B. Laplace transformation	194
C. Markov processes	198
C.1. General Markov process	198
C.1.2. Non-stationary availability coefficient	199
C.1.3. Probability of failure-free operation	201
C.1.4. Stationary availability coefficient	203
C.1.5. Mean time to failure and mean time between failures	204
C.1.6. Mean recovery time	204
C.2. Birth-death process	205
D. General Bibliography	209

1 Preface

*Dedicated to Boris Gedenko
who was my mentor in my entire life.*

Why I have decided to write this book?

From the very beginning of my carrier I have been working at the junction of engineering and mathematics – I was a reliability engineer. As an engineer by education, I never had a proper mathematical background; however life forced me to submerge in depth of probability theory and mathematical statistics. And I was lucky meeting at the very beginning of my carrier the “Three Pillars on Which Rested the Reliability Theory” in Russia – they were Boris Gnedenko, Alexander Solovyev and Yuri Belyaev. They helped me to understand nuances and physical sense of many mathematical methods.

So, I decided to share with the readers with my own experience as well as with many real mathematical insights, which happened with my when I submerged into Reliability Theory.

Boris Gnedenko once told me: “Mathematical reliability models are engendered by practice, so they have to be adequate to reality and should not be too complex by its nature”.

To get understanding of “real reliability”, one comes through a series of painful mistakes in solving real problems. Engineering intuition arrives to mathematicians after years of working in reliability engineering, and, at the same time, proper mathematical knowledge comes to reliability engineers after multiple practical uses of mathematical methods and having “finger sensation” of formulas and numbers.

I remember my own thorny way in the very beginning of my professional carrier... So, I decided to write a reliability textbook that contains as much as possible “physical” explanations of mathematical methods applied for solving reliability problems and, as well, “physical” explanations of engineering objects laid in the basis of mathematical models.

At the end of the book, the reader can find a wide list of monographs on reliability. Nevertheless, I would like to mark out a few books that, in my opinion, are basical in this area. They are (in order of publication) the monographs by Igor Basovsky (1961), Lloyd and Miron Lipov (1962), Richard Barlow and Frank Proschan (1965), Boris Gnedenko, Yury Belyaev and Alexander Solovyev (1965). These books cover the entire area of probabilistic reliability modeling and contain many important theoretical and practical concepts.

Igor Ushakov
September 2011,
San Diego,
California.

2 Acronyms and notations

Acronyms

AC	Availability coefficient
d.f.	Distribution function
DFR	Decreasing failure rate
FR	Failure rate
GF	Generating function
i.i.d.	Independent identically distributed (about r.v.)
IFR	Increasing failure rate
LT	Laplace transform
MDT	Mean down time
MTBF	Mean time between failures
MTTF	Mean time to failure
OAC	Operational availability coefficient
PEI	Performance effectiveness index
PFFO	Probability of failure-free operation
r.v.	Random variable
RBD	Reliability block-diagram
TTF	Random time to failure
UGF	Universal generating function

Notations

$F(t)$	Distribution function
K	System stationary availability coefficient
$K(t)$	System non-stationary availability coefficient
$P(t)$	Probability of system's failure-free operation
$p_k(t)$	Probability of failure-free operation of unit k
$Q(t)$	Probability of system's failure
$q_k(t)$	Probability of failure of unit k
T	Mean time to/between failures
λ	Failure rate
τ	Down time
X	Random variable
ξ	Random time to/between failures
η	Random recovery time
$\psi(\cdot)$	System's structural function

$\sum_{1 \leq k \leq n}$	Sum by k from 1 to n .
$\prod_{1 \leq k \leq n}$	Product by k from 1 to n .
\cup	Logic sum (“or”)
\cap	Logic product (“and”)
$\bigcup_{1 \leq k \leq n}$	Logic sum by k from 1 to n .
$\bigcap_{1 \leq k \leq n}$	Logic product by k from 1 to n .
$\overline{i = 1, n}$	Set of natural numbers from 1 to n .

3 What is reliability?

3.1 Reliability as a property of technical objects

Reliability of a technical object is its ability of successful performing required operations. Usually it is assumed that an object is used in accordance with its technical requirements and is supported by appropriate maintenance.

One of the outstanding Russian specialists in cybernetics, academician Axel Berg told: "Reliability is quality expanded in time".

Reliability is a broad concept. Of course, its main characterization is the failure-free operation during performance of required tasks. However, it includes also such features as availability, longevity, recoverability, safety, survivability and other important properties of technical objects.

Speaking about reliability, one has to introduce a concept of failure. What does it mean: "successful operating"? Where is the limit of "successfulness"?

In reliability theory, one usually analyzes systems consisting of units, each of which has two states: operational and failure. If some "critical" set of units has failed, it leads to the system failure. However, not always a unit's failure leads to the "total" system failure, it can only decrease its ability but main system parameters still could be in appropriate limits.

However, such "instantaneous" failure is only one of possibilities. The system can fail due to monotonous drifting of some parameters that can bring the entire system to the unacceptable level of performance.

In both cases, one needs to formulate failure criteria.

3.2 Other "ilities"

Reliability itself is not the final target of engineering design. An object can be almost absolutely reliable in "greenhouse condition"; however, at the same time, it can be too sensitive to real environment. Another situation: an object is sufficiently reliable but during operation it produces unacceptable pollution that contaminates natural environment.

Below we discuss some properties closely connected to the concept of reliability.

Maintainability. Failure free operation is undoubtedly very important property. However, assume that a satisfactory reliable object needs long and expensive restoration after a failure. In other words, maintainability is another important

property of recoverable systems. Maintainability, in turn, depends on multiple factors itself.

The quality of restoration an object after failure as well as time spent on restoration significantly depends on repairmen qualification, availability of necessary tools and materials, etc.

Safety. Development of large scale industrial objects attracts attention to safety problem. It is clear that an object has not only to perform its main operating functions, but it is very important that the “successful operation” is not dangerous for personnel’s health and does not harm ecology.

One of the most tragic events of this kind happens in 1984. It was the Bhopal Gas Tragedy – one of the world’s worst industrial catastrophes. It occurred at the Union Carbide India Limited pesticide plant in India. The catastrophe led to almost immediate death of about 7 thousand people and then about 8 thousand died from gas-related diseases. In addition, over half a million people got serious injuries. Then, in 1986 explosion and fire had happened at the Chernobyl Nuclear Power Plant in the former Soviet Union. Large quantities of radioactive contamination were released into the atmosphere, which spread over much of Western USSR and Europe. It is considered the worst nuclear power plant accident in history. Thousands of workers were killed almost instantaneously, about one million cancer deaths occurred between 1986 and 2004 as a result of radioactive contamination.

Actually problem of safety appears not only in context of failures. A number of “reliable” industrial plants are extremely unsafe for the people who work there or live in the area.



Figure 1. Typical “industrial landscape” with terrible air pollution.

Survivability. The problem of survivability is very close to the reliability and safety problems. This is an object's property to survive under extreme natural impacts or intentional hostile actions.

In this case, nobody knows the moment of disaster, so an object has to have some "warranty level" of safety factor. In our time, the survivability problem is extremely important for large scale terrestrial energy systems.

The 1999 Southern Brazil blackout was the largest power outage ever. The blackout involved San Paulo, Rio de Janeiro, and other large Brazilian cities, affecting about 100 million people.

Then in 2003 there was a widespread power outage known as the North-East Blackout. It was the second most widespread blackout in history that affected 50 million people in Canada and the United States.

On March 11 of 2011 a ferocious tsunami spawned by one of the largest earthquakes ever recorded slammed Japan's Eastern coast. This earthquake officially named the Great East Japan Earthquake was 9 magnitudes (in Richter scale). Tsunami waves reach up to 40 meters that struck the country and in some cases traveling up to 10 kilometers inland Japan. States of emergency were declared for five nuclear reactors at two power plants. There were some severe damages, though consequences were much less than after Chernobyl.

Problem of survivability became essential in our days when unpredictable by location and strength terrorist acts are initiated by religious fanatics.

Stability. An object performance occurs in unstable conditions: environment can change, some simultaneously performing operations can conflict with each other, some disturbances can happen, etc. An object has to have an ability to return to normal operable state after such inner or outer influences.

Durability. Reliability as a concept includes such a property as durability. For instance, mechanical systems, having some fractioning parts, can be very reliable at the first several hundred hours, however after some period of time due to wearing-out processes it becomes to fail more and more frequently, and became unacceptable for further use.

Conservability. This is the property of the object to continuously maintain the required operational performance during (and after) the period of storage and transportation. This property is important for objects that are keeping as spares or are subjects of long transportation to the location of the use.

3.3 Hierarchical levels of analyzed objects.

Analyzing reliability, it is reasonable to introduce several hierarchical levels of technical objects. Below we will consider systems, subsystems and units. All these terms are obvious and understandable; nevertheless we will give some formal definitions for the further convenience.

A *unit* is an indivisible (“atomic”) object of the lowest hierarchical level in the frame of current reliability analysis.

A *system* is an object of the highest hierarchical level destined for performing required tasks.

Of course, concepts of unit and system are relative: a system in one type of analysis can be a unit in consideration of a larger scale object and vice versa. In addition, sometimes it is reasonable to introduce an intermediate substance – subsystem. It can be a part of a system that destined for performing a specific function or a separate constructive part.

System reliability indices can be expressed through corresponding indices of its units and subsystems.

3.4 How can reliability be measured?

Reliability can be and has to be measured. However, what measures should be used for reliability?

Distance can be measured in kilometers and miles; weight in kilograms and pounds; volume in liters and gallons. What kinds of index or indices are appropriate for reliability?

Of course, reliability index depends on the type of a technical object, its predestination, and regime of operating, as well as on some other factors that are usually rather individual.

Generally speaking, all technical objects can be divided into two main classes: unrecoverable and recoverable. All single-use technical objects are unrecoverable. For instance, anti-aircraft missile is used only once. It can be characterized by the probability that the required operation is completed.

A reconnaissance satellite is also a single-use object. However, for this object the best reliability index is an average time of operating without failure: the more time the satellite on the orbit, the more useful information will be collected.

Most of technical objects we are dealing with are recoverable ones: they can be restored after a failure and can continue their operations.

Let us consider a passenger jet. It is almost obvious that the most important reliability index is the probability that a jet successfully completed its flight. Of

course, one should think about longevity and convenience of technical maintenance, though these indices are undoubtedly secondary.

Let us notice that the same object may be considered as recoverable or not depending on the concrete situation. It is clear that for the same passenger jet some critical failure, having been occurred during the flight (for instance, engine failure), cannot be corrected. Thus in this case a jet should be considered as unrecoverable during a flight.

Anti-missile defense systems are working in regime “on duty”, i.e. they have to be in operable state at any arbitrary chosen moment of time. For an airport dispatcher system it is very important to be in an operational state at some required moment of time and successfully operate during an airplane landing. Thus for such systems the most important property is availability.

For a passenger bus, probably one of the main reliability characterization is the duration of failure free operation, because it means that the number of unexpected stops due failures is minimal. Same reliability index is convenient for trucks: it delivers the best economical efficiency during operations.

For most home appliances, cars, technical equipments, durability is very important, because it saves money of the user. At the same time, one does not need “immortal” personal computer, because in 2-3 years it will be anyway obsolete and should be replaced by a modern one. There are several common sense rules that one should keep in mind choosing reliability indices:

- (1) They have to reflect specificity of the object and its operating process;
- (2) They have to be simple enough and to have an understandable physical sense;
- (3) They have to be calculable analytically or numerically;
- (4) They have to be empirically confirmed by special tests or during real exploitation.

The number of indices chosen for characterization of reliability of a technical object should be as limited as possible, since multiple indices can only lead to confusion. Do not use “weighted” indices because they usually have no physical sense.

3.5 Software reliability

Software reliability requires a special topic. Frankly speaking, there are too much confusion and misunderstanding.

Nobody doubts that reliability in technical context is a concept associated with time and randomness. If there is an object (especially, immaterial) that exists in

sense “beyond the time” and its failure do not occur randomly, how we can talk about reliability?

Take a look: what is software? This is a set of commands arranged in a special order. It reminds a book “written” for a “hardware-reader” that can “read” it when and if needed.

Is it possible to say about “reliability of a book”, keeping in mind its contents? Of course, a book can contain errors (“failures”) but these errors are everlasting property of this specific book! These errors can be deleted in the next edition of the book but they are and they will remain forever in this particular edition.

The same picture we observe with software if some “inner program conflict” or “inconvenient “ set of input data appears again and again, that will lead to repeating failures. And it does not depend on current time, and it is not random at all.

For software, we should say about quality, which depends on programmer’s qualification and carefulness of testing. To say about “frequency of software failures” is hardly correct.

3.5.1 Case Study: Software failures avalanche.

In 1970s, the author, being an engineer at R&D Institute of the former Soviet Union, participated in a design of an automatic control system for missile defense. Reliability requirements for the system were extremely high.

After design completion, long and scrupulous tests began. Hardware and software have been multiply checked and rechecked: the system seemed “absolutely reliable”. But all of a sudden, a long series of software failures happened in a row!

Acceptance Commission was in panic...

After careful analysis it was found that a young lieutenant who was working as an operator mentioned that some sequence of specific commands led to computer fault. He made a corresponding note in a Test Protocol though, being too much curious, continued to try the same commands multiply.

Definitely, recording several tens faults was unreasonable. It was only one fault of software recorded. Afterwards, the software had been corrected...

However, there is a question: how you should characterize software reliability? The only fault has been recorded during 50 hours of testing. May you say that the software failure occurs once in 50 hours on average? Moreover, the program had been “repaired”. So, does it mean that after this the software became “absolutely reliable”?

Who knows when and how next time such “inconvenient” circumstances may occur in real operating regime?

4 Unrecoverable Objects

4.1 Unit

An indivisible (“atomic”) part of a system under current consideration is called a unit. Process of an unrecoverable unit operation is defined by its random time to failure (TTF). Let us denote this random variable (r.v.) by ζ and its distribution function (d.f.) by $F(t) = \Pr \{ \zeta < t \}$.

4.1.1 Probability of failure-free operation

The probability of failure-free operation (PFFO) of unrecoverable unit during time t equals, by definition,

$$P(t) = \Pr \{ \zeta \geq t \} = 1 - F(t). \quad (1)$$

If d.f. is exponential, then

$$P(t) = \exp(-\lambda t). \quad (2)$$

For highly reliable unit when $\lambda t \ll 1$, there is a good approximation:

$$P(t) \approx 1 - \lambda t. \quad (3)$$

(This estimation gives understated estimate with error of order $(\lambda t)^2$.)

Sometimes it is reasonable to introduce the so-called “indicator function” that is defined as follows:

$$x(t) = \begin{cases} 1, & \text{if unit is in operational state at moment } t, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

It is clear that x is a binary Boolean¹ r.v. The unit’s PFFO can be defined in new terms as:

$$P(t) = E\{x(t)\} = 1 \cdot P(t) + 0 \cdot [1 - P(t)], \quad (5)$$

where $E\{\cdot\}$ is operator of mathematical expectation.

¹ This type of variable is named after **George Boole** (1815 – 1864) who was an English mathematician and philosopher. He invented Boolean logic – the basis of computer logic.

This form of presentation will be sometimes useful below. (For compactness of formulas we will omit sometimes argument t .)

4.1.2 Mean time to failure

The MTTF of unrecoverable unit in general case is calculated as

$$T = E\{\xi\} = \int_0^{\infty} t \cdot f(t) dt = \int_0^{\infty} P(t) dt \quad T = E\{X\} = \int_0^{\infty} P(t) dt. \quad (6)$$

For exponential d.f., this integral gives:

$$T = \int e^{-\lambda t} dt = \frac{1}{\lambda} \quad (7)$$

$$T = \frac{1}{\lambda}$$

4.2 Series systems

A series system is such a system, for which failure of any its unit leads to inevitable failure of the entire system. Usually, these systems present a serial connection of its subsystems or units.

The series structure is one of the most common structures considered in engineering practice. In reliability engineering, for describing the logical connection of system units, one uses the so-called reliability block diagrams (RBD). For system of n units, RBD can be presented in the following form:

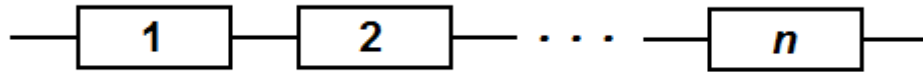


Figure 1. Reliability block-diagram for series system.

4.2.1 Probability of failure-free operation

Denote the system random TTF by ξ and units TTF's by ξ_k , where $1 \leq k \leq n$, then from the definition of a series system follows:

$$\xi = \min_{1 \leq k \leq n} \xi_k. \quad (8)$$

This statement is easily understood from the figure below:

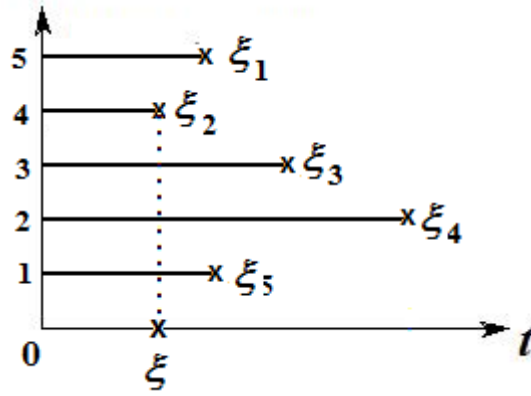


Figure 2. Illustration of a series system TTF.

Often one uses Boolean equations for description of reliability structures. For series system Boolean expression is:

$$\phi(\mathbf{X}) = \bigcap_{i=1}^n x_i . \quad (9)$$

where $\mathbf{X} = (x_1, x_2, \dots, x_n)$. For independent units, $P(t) = E\{\phi(X(t))\}$, and by the Theorem of multiplications, we can immediately write:

$$P(t) = \Pr\left\{\bigcap_{i=1}^n x_i = 1\right\} = \Pr\{\zeta_1 \geq t\} \cdot \Pr\{\zeta_2 \geq t\} \cdot \dots \cdot \Pr\{\zeta_n \geq t\} = \prod_{k=1}^n \Pr\{\zeta_k \geq t\} = \prod_{k=1}^n p_k(t). \quad (10)$$

where the probability of failure-free operation of unit k is denoted by $p_k(t) = \Pr\{\zeta_k \geq t\}$.

Let's introduce notation $q_k(t) = 1 - p_k(t)$. If system's units are highly reliable, i.e. $\max_{1 \leq k \leq n} q_k(t) \ll \frac{1}{n}$, then:

$$P(t) = \prod_{k=1}^n [1 - q_k(t)] \approx 1 - \sum_{i=1}^n q_k(t). \quad (11)$$

From formula (10), one can make the following conclusions:

- A series system's reliability decreases (increases) if the reliability of any unit decreases (increases).
- A series system's reliability decreases (increases) if the number of units is decreases (increases).
- A series system's reliability is worse than reliability of any of its units.

If each unit has exponential d.f. of TTF, then for a series system consisting of such units, one can write:

$$P(t) = \prod_{k=1}^n \exp(-\lambda_k t) = \exp\left(-t \sum_{k=1}^n \lambda_k\right) \quad (12)$$

For systems, consisting of highly reliable units, for which $t \cdot \max_{1 \leq k \leq n} \lambda_k \ll \frac{1}{n}$, one can write a convenient approximation

$$P(t) \approx 1 - t \sum_{k=1}^n \lambda_k. \quad (13)$$

If all system units were identical, then

$$P(t) = \exp(-\lambda n t). \quad (14)$$

For “feeling” the numbers, consider a system consisting of units with $p(t_0) = 0.999$. In the table below, one can see how reliability of the system decreases with the number of units increase.

Table 1. System reliability dependence on the system scale.

<i>n</i>	10	100	1000	10000
<i>P(t)</i>	0.99005	0.904837	0.367879	<i>Practically zero</i>

By the way, from this table, one can see that approximation formula is practically acceptable for values of $\lambda n t$ of order 0.1.

4.2.2 Mean time to failure

Now consider the MTTF of a series system. For any unrecoverable series system, the random TTF, ξ , can be expressed through a unit's random TTF's (ξ_k) in the following way:

$$X = \min_{1 \leq k \leq n} X_k \quad \xi = \min_{1 \leq k \leq n} \xi_k. \quad (15)$$

In general case, the MTTF can be found only in a standard way as

$$T = E\{\xi\} = \int_0^{\infty} P(t) dt \quad (16)$$

For exponential distributions, there is a closed expression:

$$T = \int_0^{\infty} \exp\left(-t \sum_{k=1}^n \lambda_k\right) dt = \frac{1}{\sum_{k=1}^n \lambda_k} \quad (17)$$

where λ_k is the parameter of the corresponding d.f.

For system with identical units with MTTF equals T^* for all $k = 1, 2, \dots, n$, one has:

$$T = \left(\frac{n}{T^*}\right)^{-1} = \frac{T^*}{n} \quad (18)$$

i.e. the system MTTF is inverse proportional to the number of units.

4.3 Parallel System

Another principal structure in reliability theory is a *parallel system*. This system is in operational state until at least one their unit is operable. This type of redundancy is called *loaded redundancy* or even in engineering jargon “*hot redundancy*”. Usually, in practice the operating and all redundant units are assumed to be identical. In addition, all units are assumed independent.

RBD for a parallel system of n units is presented in Figure 3 below.

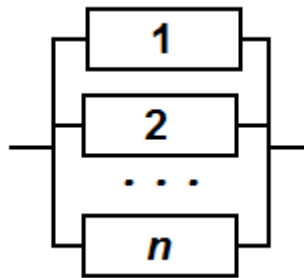


Figure 3. Reliability block diagram for parallel system.

4.3.1 Probability of failure-free operation

Let denote again the system random TTF by ζ and units TTF's by ζ_k , where $1 \leq k \leq n$, then from the definition of a parallel system follows:

$$\zeta = \max_{1 \leq k \leq n} \zeta_k. \quad (19)$$

This statement is easily to understand from the figure below:

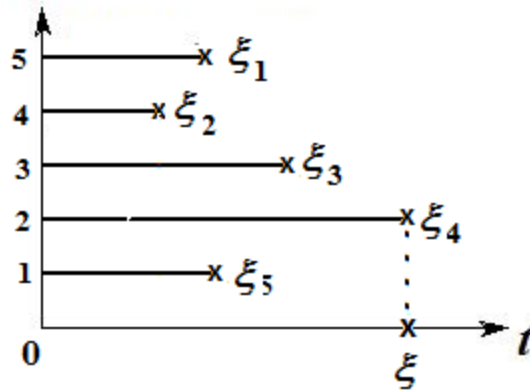


Figure 5. Illustration of parallel system TTF.

From the definition of a parallel system follows that it can be described by a Boolean function of the form:

$$\phi(x) = \bigcup_{k=1}^n x_k = x_1 \cup x_2 \cup \dots \cup x_n, \quad (20)$$

First, we transform expression (20) using de Morgan² Law of algebra of logics. This law states that

$$x_1 \cup x_2 = \overline{\bar{x}_1 \cap \bar{x}_2}, \quad (21)$$

where \bar{x} denote a complement to x . Actually, de Morgan Law can be easily proved by Ancinet Greeks rule: “Look at the drawing”. Indeed, look at the diagrams on Figure 6 that are called Venn³ diagrams:

² **Augustus De Morgan** (1806–1871) was a British mathematician and logician.

³ **John Venn** (1834-1923) was a British logician and philosopher.

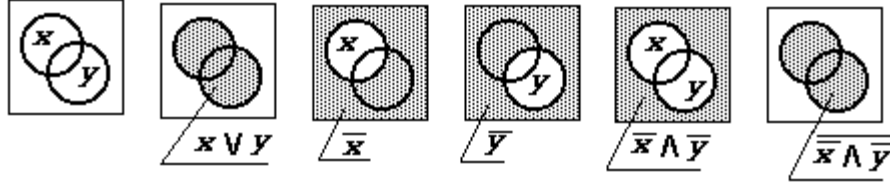


Figure 6. Venn diagrams for proving de Morgan Law.

You can see that the second and the sixth pictures present the same sets. This law can be easily expanded on arbitrary number of x 's. Let us demonstrate it on the adding x_3 :

$$x_3 \cup \overline{x_1 \cap x_2} = \overline{\overline{x_3} \cap \overline{x_1} \cap \overline{x_2}} = \overline{x_3} \cap \overline{x_1} \cap \overline{x_2}, \quad (22)$$

where we additionally use another de Morgan Rule that double complement to x is x itself. The last statement again is clear from the figure below.

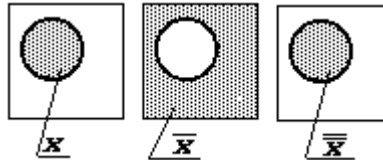


Figure 7. Venn diagrams for proving de Morgan Rule of double complement.

Thus in general case, one has the following expression:

$$\bigcup_{k=1}^n x_k = \overline{\bigcap_{k=1}^n \overline{x_k}}, \quad (23)$$

From (23) follows formula for PFFO:

$$P = E\{\phi(X)\} = 1 - \prod_{i=1}^n q_i. \quad (24)$$

where $\Pr\{x_k = 0\} = q_k$. The same result follows from the definition of a parallel system: using the Theorem of multiplications, one gets

$$Q(t) = \Pr\{(\xi_1 < t) \cap (\xi_2 < t) \cap \dots \cap (\xi_n < t)\} = \prod_{k=1}^n \Pr\{\xi_k < t\} = \prod_{k=1}^n q_k(t) \quad (25)$$

where $Q(t)$ is the probability of parallel system failure, $Q(t) = 1 - P(t)$, and $q_k(t)$ is the probability of unit k failure, $q_k(t) = 1 - p_k(t)$.

Thus, the PFFO of a parallel system is

$$P(t) = 1 - Q(t) = 1 - \prod_{k=1}^n q_k(t) \quad (26)$$

Sometimes a different form equivalent to (26) is used

$$\begin{aligned} P(t) &= p_1(t) + q_1(t) \cdot p_2(t) + q_1(t) \cdot q_2(t) \cdot p_3(t) + \dots + q_1(t) \cdot q_2(t) \cdot \dots \cdot q_{m-1}(t) \cdot p_m(t) \\ &= p_1(t) + q_1(t) \cdot [p_2(t) + q_2(t) \cdot [p_3(t) + \dots + q_{m-1}(t) \cdot p_m(t)]] \end{aligned} \quad (27)$$

This expression can be explained as following:

Pr { a parallel system successfully operates } =

Pr { the first unit is up during time t ;

or

if the first unit has failed, the second is up during time t ;

or

if both of these units have failed, then the third one;

and so on... }.

From formula (26), one can make the following conclusions:

- A parallel system's reliability decreases (increases) if the reliability of any unit decreases (increases).
- A parallel system's reliability decreases (increases) if the number of units decreases (increases).
- A parallel system's reliability is higher than reliability of any of its units.

If each of the system's unit has an exponential TTF distribution $p_k(t) = \exp(-\lambda_k t)$, for a highly reliable system where $\max_{1 \leq k \leq n} q_k(t) \ll \frac{1}{m} \ll \frac{1}{n}$, one can write $q_k(t) \approx \lambda_k t_0$, and, finally,

$$P(t) = 1 - \prod_{k=1}^n [1 - \exp(-\lambda_k t)] \approx 1 - t^n \prod_{k=2}^n \lambda_k \quad (28)$$

If all units of a parallel system are identical with exponentially distributed TTF then (28) turns to the following:

$$P(t) \approx 1 - (\lambda t)^n \quad (29)$$

In other words, the distribution function of parallel system TTF has a Weibull⁴-Gnedeno⁵ distribution with parameters $\alpha = \lambda^n$ and $\beta = n$ (see appendix A.2.8).

For “having a sensation” of the numbers, consider a parallel system of n identical units with $p = 0.9$. In the table below, one can see how reliability of the system significantly is increased with the number of units increased.

Table 2. Depending of parallel system’s PFFO on the number of parallel units.

n	2	3	4	5	...	10
$P=0.9$	0.991	0.9991	0.99992	0.999992	...	<i>Practically 1</i>

4.3.2 Mean time to failure

The MTTF of a parallel system in general case can be calculated only by integration of corresponding function $P(t)$. However, when each unit has exponential distribution of TTF, an analytic expression can be derived. For this purpose, write the PFFO in the form (27). Simple integration immediately gives us the result:

$$T = \sum_{1 \leq k \leq n} \frac{1}{\lambda_k} - \sum_{1 \leq k < j \leq n} \frac{1}{\lambda_k + \lambda_j} + \dots + (-1)^n \frac{1}{\sum_{1 \leq k \leq n} \lambda_k} \quad (30)$$

If, in addition, all units are identical, then (30) turns into the following:

:

$$T = \frac{1}{\lambda} \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \right) = \frac{1}{\lambda} \sum_{1 \leq k \leq n} \frac{1}{k} \quad (31)$$

where $\frac{1}{\lambda}$ is the MTTF of a single unit.

Explanation of this formula is understandable on the basis of the following simple and “physical” arguments. Consider a system, consisting of n identical and independent units each of which has exponential distribution with parameter λ . Assume that we have a series system of n units but after first failure the

⁴ **Ernst Hjalmar Waloddi Weibull** (1887-1979) was Swedish engineer, scientist, and mathematician. In the middle of 1930-s suggested a model of “weakest link” type. He introduced a two-parameter distribution of rather universal kind.

⁵ **Boris Vladimirovich Gnedenko** (1912-1995) was an outstanding Soviet mathematician who proved in a very beginning of 1940-s a cycle of limit theorems concerning extreme r.v.’s. The so-called Weibull distribution was a particular case of the entire class of limit distributions. This fact gives rise to call this distribution by two names.

system instantly transforms into a system of $(n-1)$ units and continue working with no failure. MTTF of that system equals $\frac{1}{n\lambda}$. Next system works on average $\frac{1}{(n-1)\lambda}$ until its failure and instantly transforms into a series system of $(n-2)$ units. Such transformation continues until the last survivor that is working on average time $\frac{1}{\lambda}$. This procedure is illustrated for a parallel system with $n = 5$ in Figure 8 below.

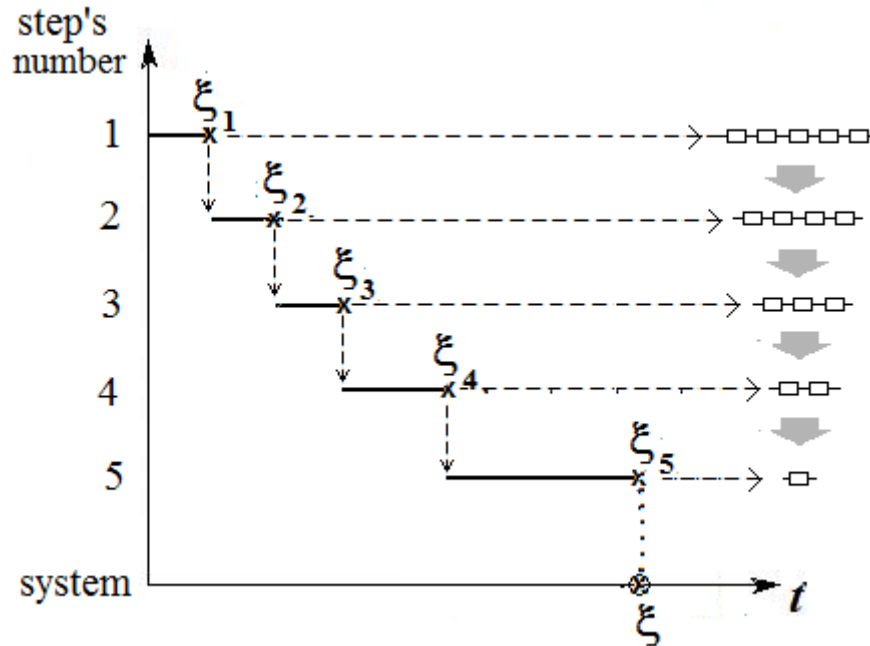


Figure 8. Explanation of deriving formula (35).

With unlimited n increase, MTTF of a parallel system approaches infinity though this increase is very slow. Some numbers are given in the table below.

Table 3. Increase of parallel system MTBF depending on number of parallel units.

Number of parallel units	MTTF increase
2	=1.50
3	≈1.83
5	≈2.28
10	≈2.93
15	≈3.32
20	≈3.60

For large n one can use Euler⁶ formula for a harmonic set

$$\sum_{k=1}^m \frac{1}{k} \approx \ln m + C \quad \sum_{k=1}^n \frac{1}{k} \approx \ln n + C, \quad (32)$$

where C is the Euler constant: $C \approx 0.5772$. However, hardly anybody will use multiple loaded redundancy for MTTF increase, because it is too ineffective.

4.4 Structure of type « k -out-of- n »

A system with such structure consists of n units and remains in operational state until $n - k + 1$ of units have failed. Structural function of such a system can be written as follow:

$$\psi(\mathbf{x}) \begin{cases} 1, & \text{if } \sum_{i=1}^n x_i \geq k \\ 0, & \text{otherwise.} \end{cases} \quad (33)$$

Factually, such a system can be considered as a series system of k units with $n-k$ redundant units, each of which can replace any one of failed operating units. Conditional RBD of this type of system is presented in Figure 9.

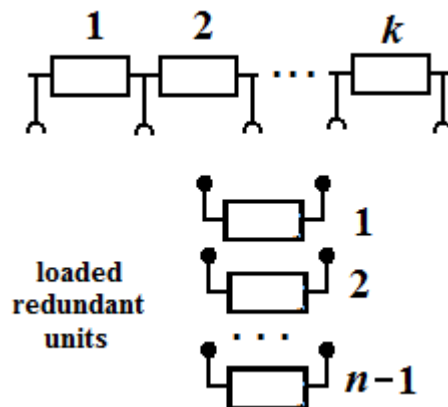


Figure 9. Conditional RBD of unrecoverable loaded “ k -out-of- n ” system.

⁶ **Leonard Euler** (1707-1783) was Swiss, German and Russian scientist who made significant contributions in mathematics, physics, mechanics and astronomy. The second half of his life worked in Sanct Petersburg Academy of Sciences, Russia.

In general case, the formula for PFFO can be written as:

$$P_{k\text{-out-of-}n}(t) = \sum_{j=k}^n \binom{n}{j} [p(t)]^j [q(t)]^{n-j} = 1 - \sum_{j=0}^{k-1} \binom{n}{j} [p(t)]^j [q(t)]^j. \quad (34)$$

where $\binom{n}{j} = \frac{n!}{j!(n-j)!} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-j+1)}{1 \cdot 2 \cdot \dots \cdot j}$ is a binomial coefficient.

For highly reliable units (when $q \ll \frac{1}{n}$), one can use an approximation:

$$P_{k\text{-out-of-}n}(t) \approx 1 - \binom{n}{k-1} [q(t)]^{n-k+1}. \quad (35)$$

If units have exponential distribution of TTF, then using arguments analogous to those used for (30), one gets:

$$T_{k\text{-out-of-}n} = \frac{1}{\lambda} \sum_{i=k}^n \frac{1}{i}. \quad (36)$$

Notice that when $k=n$, the structure transforms into an ordinary series system of n units, and when $k=1$, it transforms into an ordinary parallel system.

As a rule, structures «2-out-of-3» are found in engineering practice. For such system Boolean expression has the form:

$$\phi(\mathbf{x}) = (x_1 \cap x_2 \cap x_3) \cup (\bar{x}_1 \cap x_2 \cap x_3) \cup (x_1 \cap \bar{x}_2 \cap x_3) \cup (x_1 \cap x_2 \cap \bar{x}_3). \quad (37)$$

The RBD for such system is presented in Figure 10.

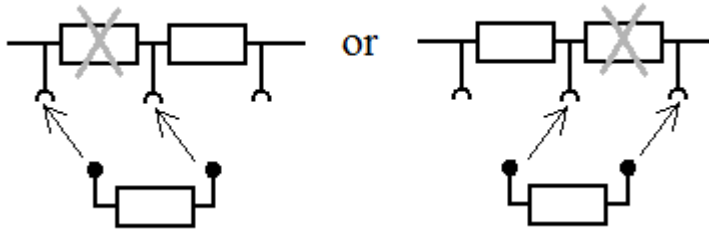


Figure 10. Connection a redundant unit instead of a failed operating unit.

From (34) for identical units follows

$$P_{2\text{-out-of-}3} = E\{\phi(\mathbf{x})\} = p^3 + 3p^2q. \quad (38)$$

4.5 Realistic models of loaded redundancy

On the paper redundant systems are sufficiently reliable. However, a real life is harder than a paper project...

Usually, one should use monitoring of operating unit for switching to redundant one after a failure. In addition, there should be a switching device, etc. Thus, the main problem is in constructing a realistic mathematical model. There is no universal solution: a reliability engineer has to take into account all specific features of analysed equipment and constructs an individual mathematical model for it.

We begin with dubbed system, one of the most often case of redundancy in engineering practice. We will consider this case in more details, taking into account some realistic assumptions.

4.5.1 Unreliable switching process

What happens if switching process is unreliable? How much it will affect on the dubbed system reliability?

Let us denote probability of successful switching by π . Then PFFO of unrecoverable dubbed system can be written as:

$$P(t) = (1 - \pi) p(t) + \pi \{1 - [q(t)]^2\}, \quad (39)$$

We would like to demonstrate numerical results using for this purpose a simple model realized in MS Excel. In this example, TTF distribution is taken exponential.

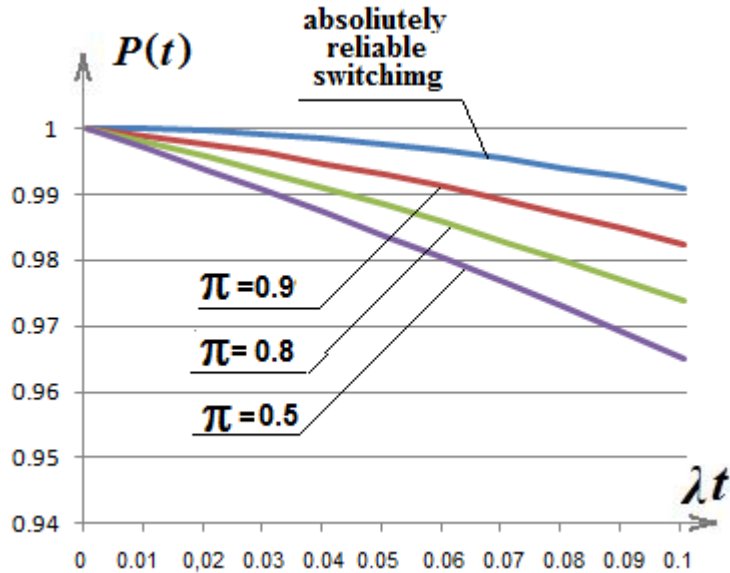


Figure 11. Dependence of a dubbed unrecoverable system on reliability of switching.

4.5.2 Non-instant switching

Usual assumption that switching from a failed operational unit to redundant one is instantaneous is erroneous. Actually, many systems have some “functional inertia”: it can stand short down times. In other words, there is some “acceptable” switching time, ε , which does not lead to the dubbed system failure.

Switching time itself can be random with some d.f., $F_{\text{switch}}(t)$. In this case, with probability $F_{\text{switch}}(\varepsilon)$ switching is considered as successful and with probability $1 - F_{\text{switch}}(\varepsilon)$ system fails.

It is clear that this case differs from the previous model only by terminology and notations.

4.5.3 Unreliable switch

Now assume that switching process is ideal; however a switch itself is unreliable and can fail with time. Thus, if switch has failed before an operating unit failure, then utilization of redundant unit will be impossible. Let us present a conditional RBD for this case in the following form:

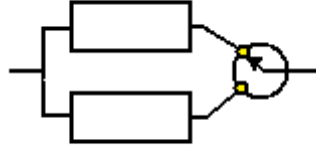


Figure 12. Conditional RBD of unrecoverable dubbed system with an unreliable switch.

Let us first find the probability, p^* , that the switch has failed after the operating unit failed, i.e. switching to redundant unit is successful.

$$p^* = P\{ \xi_{\text{switch}} > \xi / t \} = \int_0^t P_{\text{switch}}(t) \cdot f(t) dt, \quad (40)$$

where ξ is the operating unit TTF, ξ_{switch} is the switch TTF, $P_{\text{switch}}(t)$ is the switch PFFO, and $f(t)$ is density function of unit TTF.

If assume that all d.f. are exponential, it is easy to find:

$$p^* = \int_0^t \exp(-\lambda_{\text{switch}} t) \cdot \lambda \cdot \exp(-\lambda t) dt = \frac{\lambda}{\lambda_{\text{switch}} + \lambda} .$$

$$(41) \int_0^{t_0} \exp(-\lambda_{\text{nep}} t) \cdot \lambda \exp(-\lambda t) dt = \frac{\lambda_{\text{nep}}}{\lambda_{\text{nep}} + \lambda}$$

So, the PFFO of such system can be written as:

$$P(t) = p(t) + q(t) \cdot p^* \cdot p(t). \quad (42)$$

$$P(t_0) = p(t_0) + q(t_0) \cdot p^* \cdot p(t_0)$$

Below you can see numerical illustration of a switch unreliability influence.

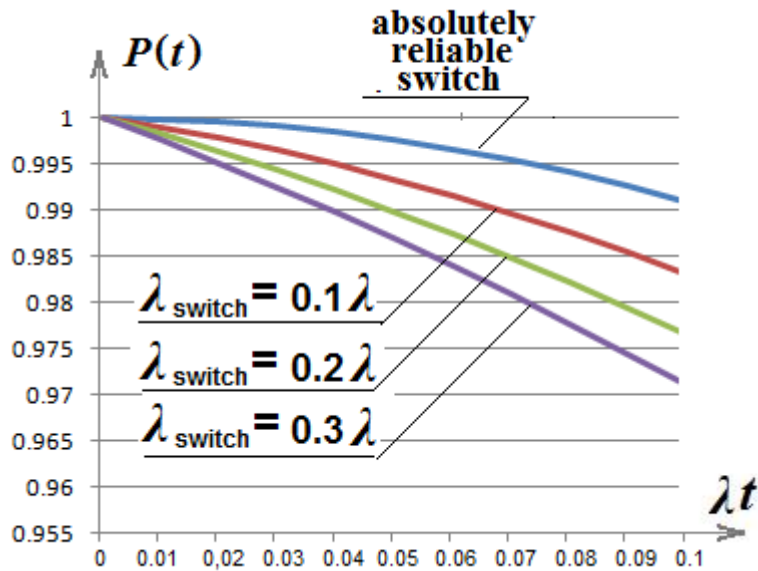


Figure 13. Dependence of unrecoverable dubbed system on switch reliability.

4.5.4 Switch serving as interface

Often (especially, in computer systems) the switch plays a role of a special kind of interface between the redundant group and a remaining part of the system. It means that such switch-interface is a necessary unit of the system and, actually, has to be considered as a series unit to the dubbed system. Conditional RBD for this case is presented below.

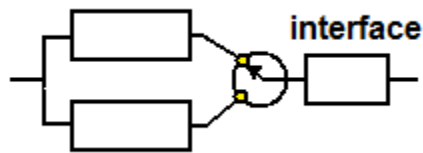


Figure 14. Conditional RBD of unrecoverable dubbed system with switch-interface.

Here we assume that switching is ideal. The switch failure becomes apparent only at the moment of required switch after a operating unit failure. In this case, the PFFO can be calculated by formula:

$$P(t) = P_{switch}(t) \cdot \{1 - [q(t)]^2\} \quad (43)$$

Again for simplicity of numerical calculations assume that all distributions are exponential.

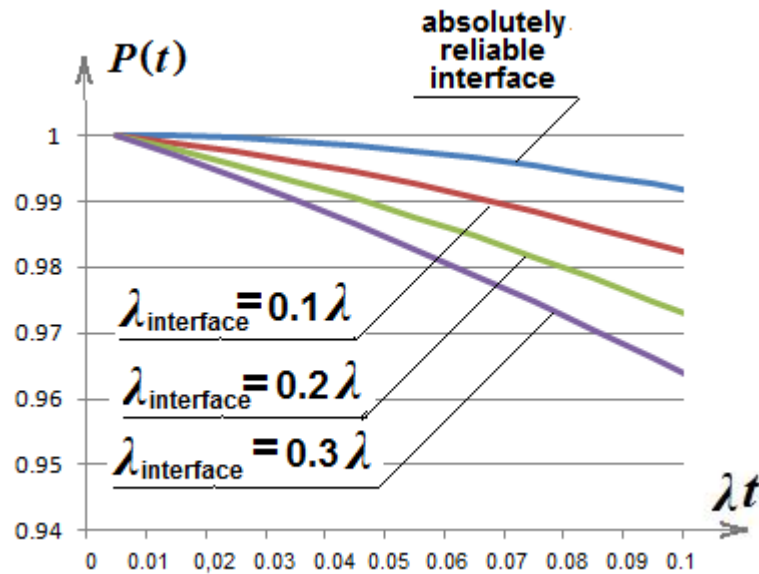


Figure 15. Dependence of unrecoverable dubbed system on switch-interface reliability.

From this example, one can see that the switch-interface reliability plays crucial role. Moreover, if switch-interface reliability is comparable with unit reliability, then duplication almost has no practical sense.

All these models are given to demonstrate how important can be “secondary” factors”, concerning switching from failed unit to redundant one.

4.5.5 Incomplete monitoring of operating unit

However, switching is not the only important factor when one deals with the redundancy group analysis. Also there is very important the monitoring procedure: without determination of operating unit failure, it is impossible to make a decision “to switch or not to switch”.

Assume that some part of the operating unit, say, a ($a < 100\%$), is not controlled at all, i.e. any failure of this part becomes a hidden failure, and switching to a redundant units does not occur. Conditional RBD for this case is presented below.

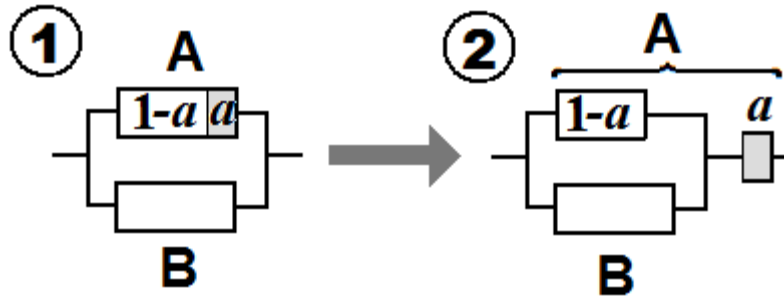


Figure 16. Initial RBD of loaded dubbed system with incomplete monitoring (1) and its equivalent presentation (2).

The PFFO of such system can be obtained with the use of the following formula:

$$P(t) = p_a(t) \cdot [1 - q_{1-a}(t) \cdot q(t)], \quad (44)$$

where $p_a(t)$ – the PFFO of a non-controlled part of the operating unit and $q_{1-a}(t)$ – failure probability of a controlled part of the operating unit. Below there are results of numerical calculations.

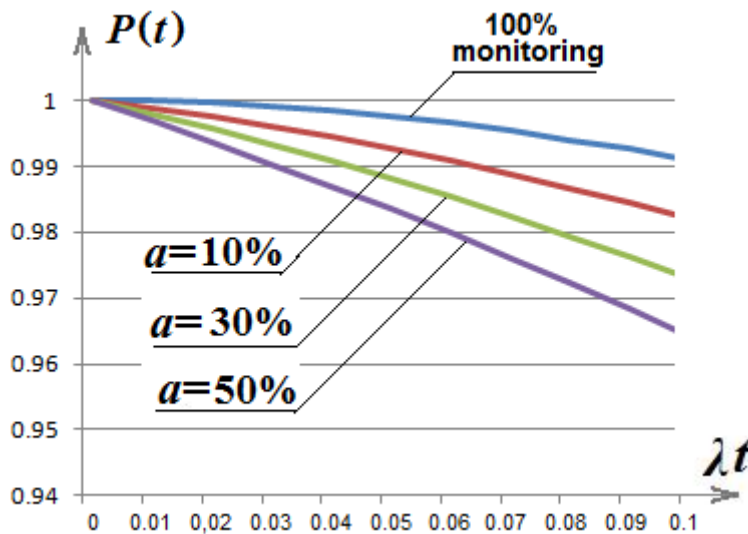


Figure 17. Dependence of unrecoverable dubbed system on the operating unit monitoring completeness.

4.5.6 Periodical monitoring of the operating unit

For unrecoverable redundant system, periodical monitoring has no sense at all: any failure is detected “post mortem”, when a operating unit has already failed, however switching has not occurred.

4.6 Reducible structures

Pure series and pure parallel systems met in engineering practice not so often. In general case, systems have more complex structures. However most of such structures can be reduced to a conditional unit by some simple procedures. Such systems are called reducible.

4.6.1 Parallel-series and series-parallel structures

The simplest reducible systems are parallel-series and series-parallel structures presented in Figures 19(a) and 19(b), respectively.

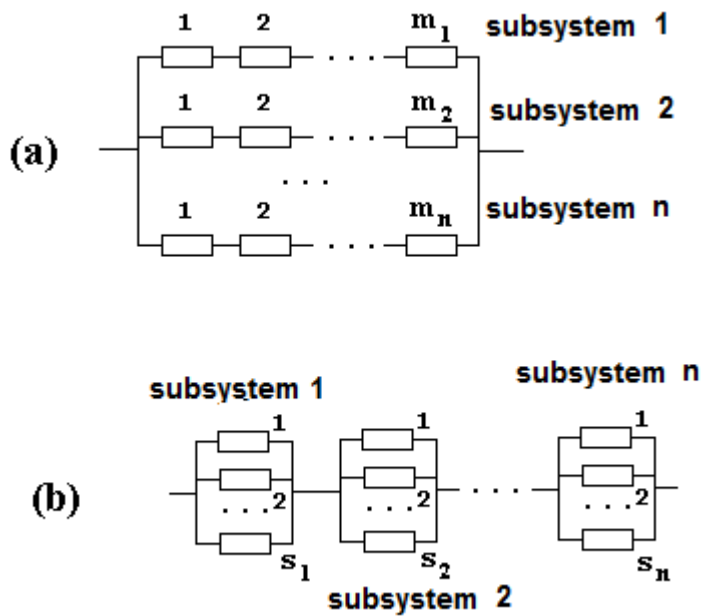


Figure 18. RBD for parallel-series and series-parallel systems.

We will write only expressions for PFFO without trivial explanations:

$$1. \quad P = 1 - \prod_{j=1}^n \left[1 - \prod_{k=1}^{m_j} p_k \right], \quad (45)$$

and

$$2. \quad P = \prod_{j=1}^n \left[1 - \prod_{k=1}^{s_j} (1 - p_k) \right]. \quad (46)$$

Of course, such idealized systems also seldom met in engineering practice.

4.6.2 General case of reducible structures

Avoiding general consideration, let us demonstrate the procedure of reducing on a simple particular RBD.

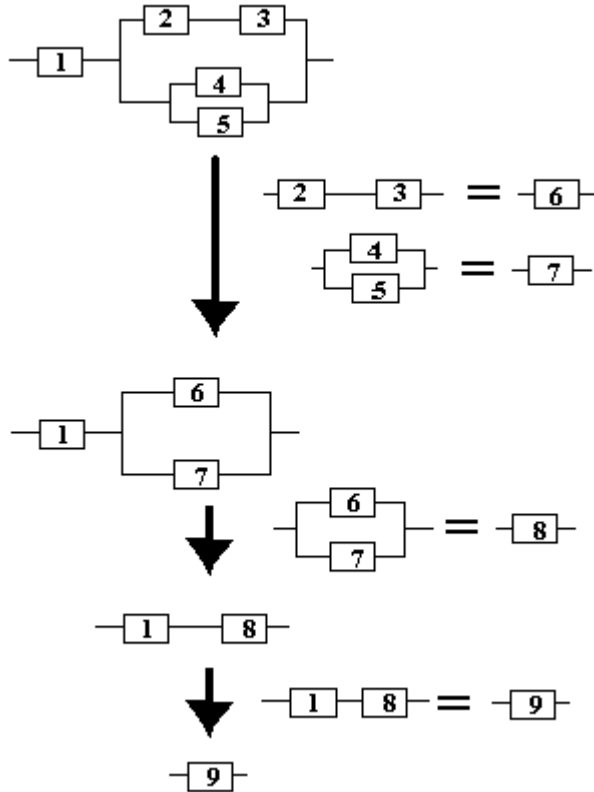


Figure 19. Example of reducing RBD of a system to a single unit.

At the beginning, we distinguish series structure (units «2» and «3»), and parallel structure (units «4» and «5»), and form new “equivalent” «6» и «7». Then units «6» and «7» are transformed into units «8». And finally, we get a single equivalent unit «9».

Construction of the expression for system PFFO is starting from the bottom of the scheme of transformation:

$$P = p_9 = p_1 \cdot p_8 = p_1 \cdot (1 - q_6 q_7) = p_1 \cdot \{1 - (1 - p_2 p_3) \cdot [1 - (1 - q_4 q_5)]\}. \quad (47)$$

Of course, not all structures are reducible; some of them will be considered later.

4.7 Standby Redundancy

Most technical systems have spare parts that can almost instantaneously replace failed operating unit. Of course, in practice such replacement takes a time, though most mathematical models assume that the time of replacement equals 0.

This type of redundancy is named standby redundancy. In this case redundant units are not included in an "active" system's structure. Moreover, these redundant units cannot fail until they occupy an active position. Of course, redundant units have to be identical to operating ones by all parameters, including reliability indices.

4.7.1 Simple Redundant Group

A system consisting of a single operating unit and $n-1$ standby units is called a redundant group. Conditional RBD of redundant group with standby redundancy can be presented in the form:

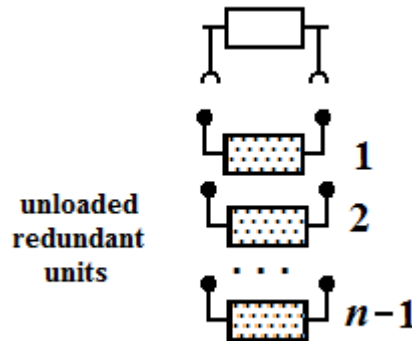


Figure 20. Conditional RBD for standby redundant group. (Standby units are dotted.)

In this case, the random time of the system's successful operation ξ equals

$$\xi = \sum_{1 \leq k \leq n} \xi_k \quad (50)$$

Thus a system's MTTF can be written immediately:

$$T = E\{\xi\} = E\left\{\sum_{1 \leq k \leq n} \xi_k\right\} = \sum_{1 \leq k \leq n} E\{\xi_k\} = \sum_{1 \leq k \leq n} T_k = nT^* \quad (51)$$

where T^* is the single unit's MTTF.

Remember the well-known property of the mean: formula (51) is valid even if the standby units are dependent.

The probability of a system's successful operation $P(t)$ can be written as

$$P(t) = \Pr\{\xi \geq t\} = \Pr\left\{\sum_{k=1}^n \xi_k \geq t\right\}. \quad (52)$$

It is known that the distribution of the sum of random variables is calculated as a convolution by the formula:

$$P^{(n)}(t) = 1 - F^{*n}(t) = \int_0^t P^{(n-1)}(t-x) dF(x). \quad (53)$$

where $P^{(k)}(t)$ is the PFFO of the system with $k-1$ standby units (k units in the entire redundant group).

Formula (53) factually gives only idea of calculation, since in most practical cases only numerical calculations are applicable.

However, in engineering practice, especially for electronic devices, the most frequently used distribution $F(t)$ is exponential. The standby group's random TTF has the Erlang⁷ d.f. of the n -th order (see Appendix A.2.5), and the probability of a failure-free operation is

$$P(t) = \Pr\left\{\sum_{1 \leq k \leq n} \xi_k \geq t\right\} = \sum_{0 \leq k \leq n} \frac{(\lambda t)^k}{k!} e^{-\lambda t} = 1 - \sum_{n+1 \leq k < \infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad (54)$$

For $\lambda t \ll 1$ the approximation can be written as

$$P(t) \approx 1 - \frac{(\lambda t)^{n+1}}{(n+1)!}. \quad (55)$$

In conclusion, notice that standby redundancy is more effective than loaded redundancy. This follows from a simple fact that

⁷ **Agner Krarup Erlang** (1878 – 1929) was a Danish mathematician, statistician and engineer, who invented the fields of traffic engineering and queuing theory. Erlang also created the field of telephone networks analysis.

$$\xi_{standby} = \sum_{1 \leq k \leq n} \xi_k \geq \max_{1 \leq k \leq n} \xi_k = \xi_{loaded} \quad (56)$$

The equality is never attained because of the strongly positive values of ξ 's.

Of course, the reader should keep in mind that standby redundancy, in practice, demands some time to switch a unit into an active regime. More reasonable is considering unloaded redundancy as spare parts.

4.7.2 Standby redundancy of type “k-out-of-n”

This type of redundancy can be presented by conditional RBD, depicted in Figure 21.

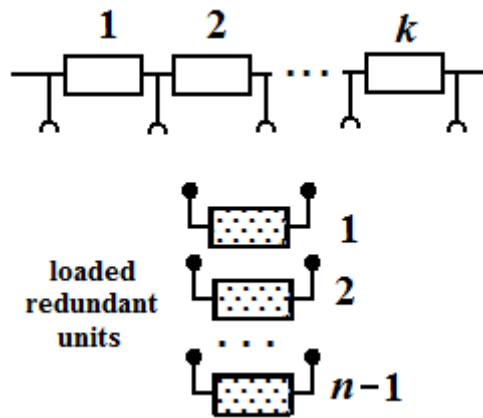


Figure 21. Conditional RBD for standby “k-out-of-n” redundant group.

It is clear that pure standby redundancy hardly can be implemented in real technical system. Mostly this type of model is used for evaluation of spare units sufficiency. In this case the role of series system plays all units of the same type within analyzed equipment.

Consider a series system of k operating units. The system is supported by $n-k$ standby units which can replace any failed unit of the group of k . In general case, formulae for PFFO and MTTF cannot be written in a simple closed form except for the case of an exponentially distributed random TTF of units. We may write the result basing our explanation on simple arguments.

Mean time to failure

Recall again that we assume that the units are i.i.d.

The system failure rate equals $k\lambda$. There are $n - k$ possible replacements, with average interval $\frac{1}{k\lambda}$ between them. So the system MTTF, i.e. average time until stock's depletion equals $\frac{n - k + 1}{k\lambda}$.

Probability of failure-free operation

The probability of a system's successful operation when its units have exponential TTF is described by Poisson⁸ distribution:

$$P(t) = \sum_{i=0}^{n-k} \frac{(k\lambda t)^i}{i!}. \quad (57)$$

4.8 Realistic models of unloaded redundancy

Ideal model of standby redundancy is constructed under assumption that a standby unit is immediately switched instead of failed operating unit. This is practically impossible: here should be some kind of functional inertia, i.e. the system has to assume a possibility of short down times needed for switching. Moreover, here (as well as in the case with loaded redundancy) there are many additional factors influencing on standby redundant group.

Just for simplicity of some expressions, let us assume that units have exponential distribution of TTF. In this case, simple and understandable formulas can be written.

4.8.1 Unreliable switching process

If switching is ideal, then for unrecoverable unloaded dubbed system PFFO is written as $P(t) = e^{-\lambda t} (1 + \lambda t)$. Let successful switching occurs with probability π . In this case, system PFFO, $P(t)$, can be calculated as:

⁸ **Siméon Denis Poisson** (1781–1840) was a French mathematician, geometer, and physicist. He made outstanding contribution in Probability Theory and Theory of Stochastic Processes.

$$P(t) = (1 - \pi) \cdot e^{-\lambda t} + \pi \cdot e^{-\lambda t} (1 + \lambda t) \approx e^{-\lambda t} (1 + \lambda \pi t). \quad (58)$$

How serious is a resulting error due to assumption concerning ideal switching? Let us give numerical illustration, using a simple model based on MS Excel.

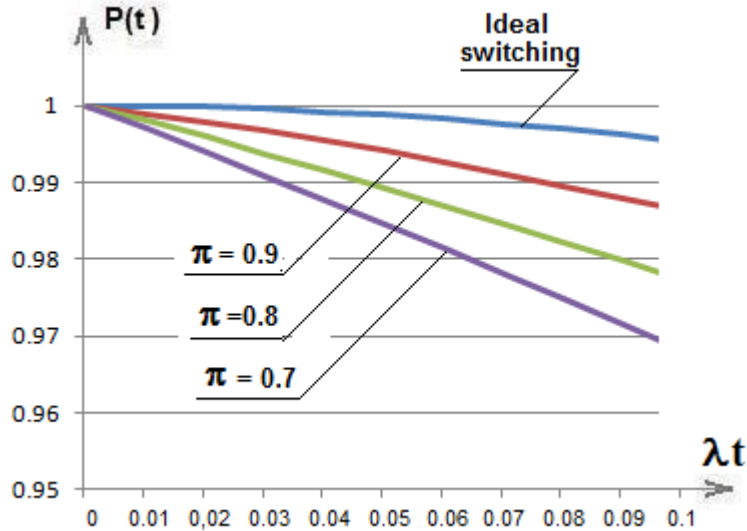


Figure 22. Dubbed system's PFFO depending of probability of switching failure.

4.8.2 Non-instant switching

Assume that acceptable switching time equals ε , i.e. if the switching duration, ω less than ε ($\omega < \varepsilon$) the system does not fail. If this switching time is a random value with a known distribution function, $F_{\text{switch}}(t)$, the probability of acceptably switching time, π , can be easily calculated: $\pi = F_{\text{switch}}(\varepsilon)$.

After this, one can use the results of the previous section.

4.8.3 Unreliable switch

Assume that switching procedure is ideal but a switch itself can fail with time. The RBD of such system is presented in Figure 23.

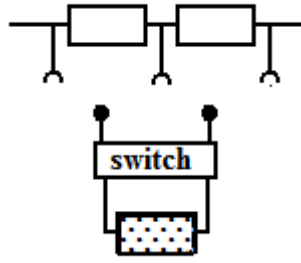


Figure 23. Unloaded redundancy with an unreliable switch.

If the switch has failed before the operating unit, use of spare unit becomes impossible.

Probability, p^* , that switching device has failed after an operating unit is

$$p^* = P\{\xi_{\text{switch}} > \xi/t\} = \int_0^t P_{\text{switch}}(t) f(t) dt \quad (59)$$

where ξ – unit's TTF, ξ_{switch} – switch's TTF, $P_{\text{switch}}(t)$ – switch's PFFO, and $f(t)$ – density function of unit's TTF.

If assume that all random variables have exponential distribution, one finds by simple integration:

$$p^* = \int_0^{\infty} \exp(-\lambda_{\text{switch}} t) \cdot \lambda \cdot \exp(-\lambda t) dt = \frac{\lambda}{\lambda + \lambda_{\text{switch}}} .$$

$$(60) \int_0^{t_0} \exp(-\lambda_{\text{nep}} t) \cdot \lambda \exp(-\lambda t) dt = \frac{\lambda_{\text{nep}}}{\lambda_{\text{nep}} + \lambda}$$

Again we can get expression for PFFO, substituting p^* instead of π in formula (58).

$$P(t_0) = p(t_0) + q(t_0) \cdot p^* \cdot p(t_0).$$

Under assumption of exponential distribution of all TTF's, the formula for PFFO can be easily written:

$$P(t) = \frac{\lambda_{\text{switch}}}{\lambda + \lambda_{\text{switch}}} \cdot e^{-\lambda t} + \frac{\lambda}{\lambda + \lambda_{\text{switch}}} \cdot e^{-\lambda t} \cdot (1 + \lambda t) = e^{-\lambda t} \cdot \left(1 + \frac{\lambda^2 t}{\lambda + \lambda_{\text{switch}}} \right) \quad (61)$$

This expression has a clear sense: with probability $\frac{\lambda_{\text{switch}}}{\lambda + \lambda_{\text{switch}}}$ a switch has failed before a operating unit failure, and with probability $\frac{\lambda}{\lambda + \lambda_{\text{switch}}}$ a switch

performs hooking-up a redundant unit. The graph below numerically illustrates the influence of the switch reliability on the system reliability.

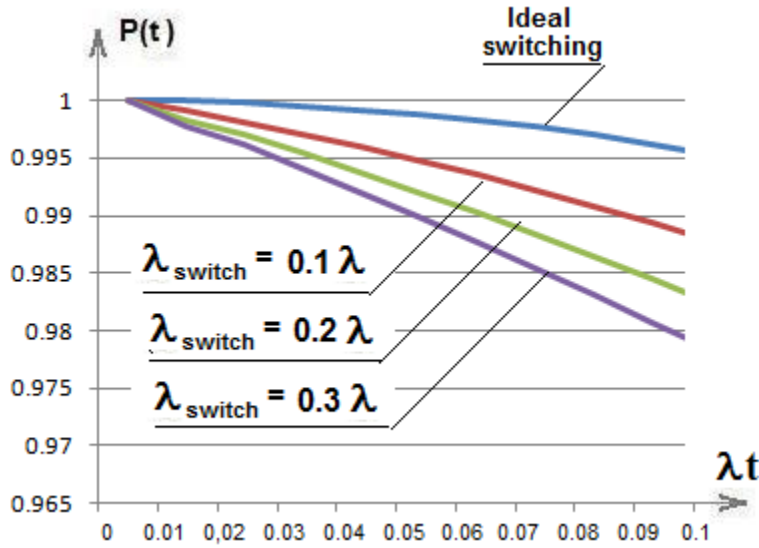


Figure 24. System's PFFO depending on switch reliability.

4.8.4 Switch serving as interface

In some situations, a switching device is a necessary part of the system operation. For instance, it can use as interface between redundant group and the remaining part of the system. In this case, a switching device has to be considered as a series unit.

Conditional RBD of such a system can be presented as follows.

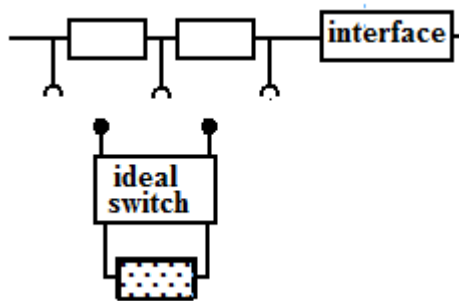


Figure 25. Conditional RBD for unloaded duplication with switch-interface.

The system fails if redundant group has failed or if the switching device has failed. In this case,

$$P(t) = \exp(-\lambda_{switch}t) \cdot (1 + \lambda t) \cdot e^{-\lambda t} = \exp[-(\lambda_{switch} + \lambda)t] \cdot (1 + \lambda t), \quad (63)$$

(We assume that switching procedure itself is ideal.)

Dependence of system PFFO on parameters of a unit and unreliability of switch -interface presented below.

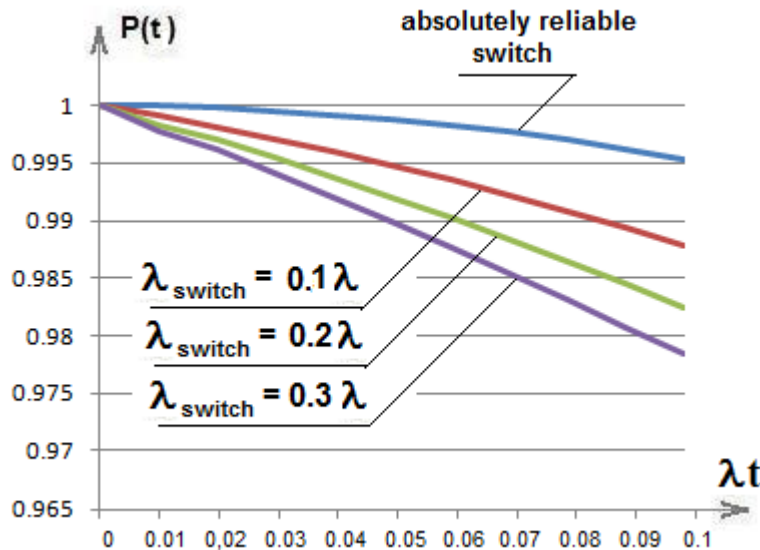


Figure 26. System's PFFO depending on switch reliability.

These graphs show how important is role of switching device reliability. Indeed, dubbed system cannot be more reliable than the switch-interface.

4.8.5 Incomplete monitoring of the operating unit

As we saw above, monitoring of operating system is a very significant factor for unrepairable redundant systems.

Assume that a part, say, a ($a < 100\%$), of an operating unit is not monitoring, i.e. its failure leads to the system failure, since there is no indication for switching to a standby unit.

Conditional RBD for this case is presented in Figure 27.

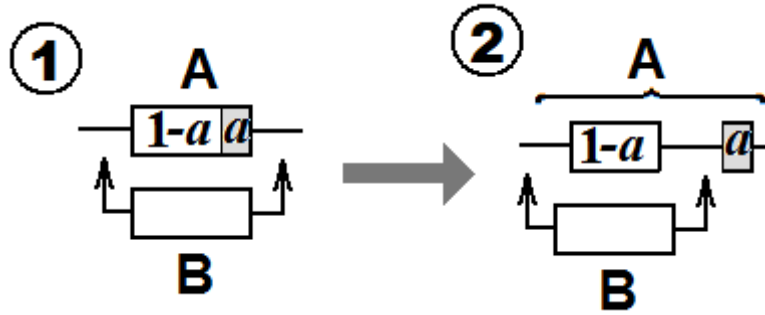


Figure 27. Initial RBD of standby dubbed system with incomplete monitoring of the operating unit and its equivalent transformation.

PFFO of this system is expressed as

$$P(t) = \Pr\{\xi_a < \xi_{1-a}\} \cdot \Pr\{\xi_a > t\} + \Pr\{\xi_{1-a} < \xi_a\} \cdot [\Pr\{\xi_{1-a} > t\} + \Pr\{\xi_{1-a} < t\} \cdot \Pr\{\xi > t - \xi_{1-a}\}] \quad (64)$$

In general case, this formula cannot be written in closed form and only numerical integration is possible. However if all distributions are exponential, one can derive expression for PFFO in close form:

$$P(t) \approx \frac{\lambda_a}{\lambda} \cdot \exp(-\lambda_a t) + \frac{\lambda_{1-a}}{\lambda} \cdot \exp(-\lambda t) \cdot (1 + \lambda t). \quad (65)$$

Comparison of different variants of monitoring completeness is given below.

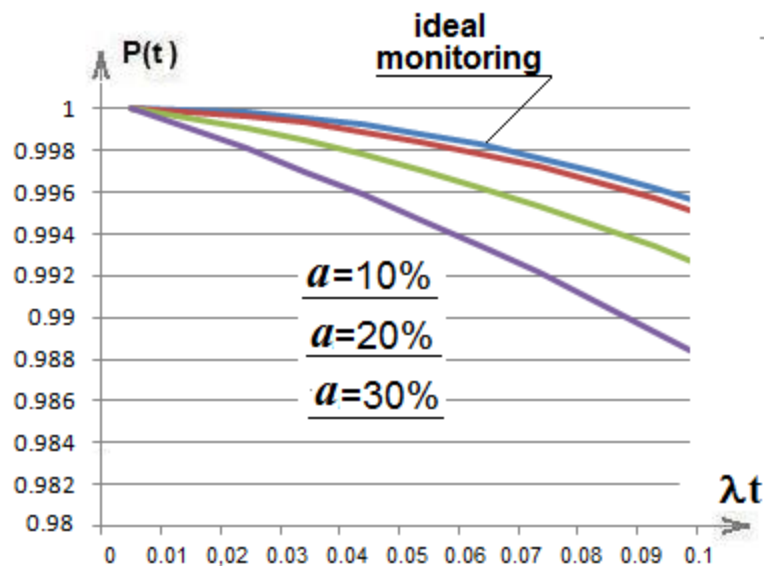


Figure 28. System's PFFO depending on completeness of operating unit monitoring.

Notice that periodical monitoring of operating unit in the case of standby redundancy also has no sense, since after “hidden” failure the system occurs in failure state.

5 *Recoverable systems. Markov models*

Reliability of recoverable systems with arbitrary distributions of units' time to failure practically cannot be described analytically in convenient and “workable” form. So, we restricted ourselves with Markov⁹ models. By the way, almost all stationary reliability indices for non-Markov models can be derived by substitution corresponding MTTF and MDT's in formulas obtained for Markov models.

5.1 Unit

A recoverable unit is defined by its two main parameters – MTTF and mean recovery time (MDT). Usually, one assumes that after failure a recovered unit is identical (in statistical sense) to the initial one, so in this case, MTTF and MTBF coincide. We are beginning with the simplest case when both distribution TTF and recovery time are exponential.

5.1.1 Markov Model

General description

In most academic approaches random TTF and random recovery time are assumed exponentially distributed that gives a possibility to use Markov model for reliability study. Let parameter of the TTF distribution is λ and parameter of the recovery time is μ . In other words, MTBF (MTTF) and mean recovery time are known from the beginning: $T = \frac{1}{\lambda}$ and $\tau = \frac{1}{\mu}$.

At any moment of time, the unit can be in one of two states: it is either operable or it is failed. A convenient form of presentation of such process is the transition graph presented in Figure 1. Let us denote the operable state with symbol “0”, and the failed state with symbol “1”.

⁹ **Andrey Andreyevich Markov** (1856-1922) was a Russian mathematician. He is best known for his work on theory of stochastic processes. His research later became known as Markov chains.

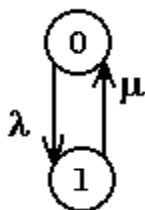


Figure 1. Transition graph for a recoverable unit.

The unit transition process can be described as an alternative renewal process. It is represented by a sequence of mutually independent r.v.'s ζ_k (unit's operational time) and η_k (unit's recovery time). An example of time diagram is presented in Figure 2.

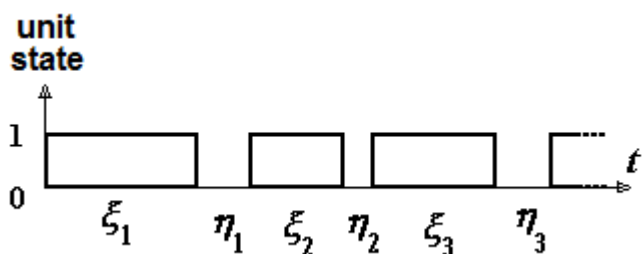


Figure 2. Example of time diagram for a unit.

Equations for finding non-stationary availability coefficient.

Let us find the probability, $p_0(t)$, that at moment $t + \Delta t$ a unit occurs in state "0". There are two possibilities:

- at moment t unit was in state "0" and did not leave it during infinitesimally small time interval Δt , that happens with probability $1 - \lambda \Delta t$, or
- at moment t it was in state "1" and moved to the state "0" during the time interval Δt , that happens with probability $\mu \Delta t$.

This immediately gives the equation:

$$p_0(t + \Delta t) = (1 - \lambda \Delta t) p_0(t) + \mu \Delta t p_1(t) \quad (3)$$

Form (3) we easily obtain

$$\frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} = -\lambda p_0(t) + \mu p_1(t). \quad (4)$$

and in the limit as $\Delta t \rightarrow 0$, we obtain the following differential equation:

$$\frac{d}{dt} p_0(t) = -\lambda p_0(t) + \mu p_1(t). \quad (5)$$

This represents the simplest example of Chapman¹⁰-Kolmogorov¹¹ Equation. To solve it with respect to any $p_k(t)$, we need to have one more equation. The second equation, which is called normalization equation, is:

$$p_0(t) + p_1(t) = 1 \quad (6)$$

which means that at any moment the unit must be in one of two possible states.

We need also to determine the initial condition for the solution of the system of differential equations. Usually, one assumes that at moment $t = 0$ the unit is in operational state, i.e. $p_0(0) = 1$.

It is clear that $p_0(t)$ is non-stationary availability coefficient that shows the probability that a unit has been found in state “0” at a given moment t under condition that at moment $t = 0$ the unit was in state “0”.

Finding this reliability index can be done with the help of different methods. We will use the Laplace¹² transform (LT) to make the presentations of other solutions in the book uniform. Brief information about Laplace transforms the reader can find in Appendix B.

Non-Stationary Availability Coefficient

The system of above differential equations (5) - (6) with the initial condition $p_0(t) = 1$ has the LST form:

$$\begin{aligned} -1 + s\varphi_0(s) &= -\lambda\varphi_0(s) + \mu\varphi_1(s) \\ \varphi_0(s) + \varphi_1(s) &= \frac{1}{s}. \end{aligned} \quad (7)$$

¹⁰ **Sydney Chapman** (1888–1970) was a British mathematician and geophysicist.

¹¹ **Andrey Nikolaevich Kolmogorov** (1903–1987) was a great Soviet Russian mathematician, of the 20th century, who advanced various scientific fields, among them probability theory, topology, intuitionistic logic, turbulence, classical mechanics and computational complexity.

¹² **Pierre-Simon, marquis de Laplace** (1749–1827) was a French mathematician and astronomer whose work was pivotal to the development of mathematical astronomy and statistics.

or, in the canonical form

$$\begin{aligned}(\lambda + s)\varphi_0(s) - \mu\varphi_1(s) &= 1 \\ s\varphi_0(s) + s\varphi_1(s) &= 1.\end{aligned}\tag{8}$$

To solve (8), we can use Cramer's¹³ Rule:

$$\varphi_0(s) = \frac{\begin{vmatrix} 1 & -\mu \\ 1 & s \end{vmatrix}}{\begin{vmatrix} \lambda + s & -\mu \\ s & s \end{vmatrix}} = \frac{s + \mu}{s^2 + (\lambda + \mu)s}.\tag{9}$$

To invert this LST, we have to present it in a form of a sum of terms of types a/s or $b/(s+a)$, inverse functions for which are a constant and an exponential function, respectively.

The denominator of fraction in (9) can be written as $s^2 + (\lambda + \mu)s = (s - s_1)(s - s_2)$, where s_1 and s_2 are polynomial roots that are, as can be easily found: $s_1 = 0$ and $s_2 = -(\lambda + \mu)$. Now we can write:

$$\varphi_0(s) = \frac{A}{s - s_1} + \frac{B}{s - s_2} = \frac{A}{s} + \frac{B}{s + \lambda + \mu}\tag{10}$$

where A and B are 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 two representations equal:

$$\frac{A}{s} + \frac{B}{\lambda + \mu + s} = \frac{s + \mu}{s(\lambda + \mu + s)},\tag{11}$$

And so we obtain a new system for A and B by equalizing the coefficients of the polynomials:

¹³ **Gabriel Cramer** (1704 –1752) was a Swiss mathematician. In linear algebra, **Cramer's rule** is a theorem, which gives an expression for the solution of a system of linear equations in terms of the determinants.

$$\begin{aligned} A + B &= 1 \\ A(\lambda + \mu) &= \mu. \end{aligned} \tag{12}$$

It is easy to find

$$\begin{aligned} A &= \frac{\mu}{\lambda + \mu} \\ B &= 1 - \frac{\mu}{\lambda + \mu} = \frac{\lambda}{\lambda + \mu} \end{aligned} \tag{13}$$

Thus, the LST of interest can be written as

$$\varphi_0(s) = \frac{\mu}{\lambda + \mu} \cdot \frac{1}{s} + \frac{\lambda}{\lambda + \mu} \cdot \frac{1}{\lambda + \mu + s}. \tag{14}$$

Finally, the non-stationary availability coefficient, i.e., the inverse LST of (14), is

$$K(t) = p_0(t) = \frac{\mu}{\mu + \lambda} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t} \tag{15}$$

The function $K(t)$ showing the time dependence of the system availability is presented in Figure 3.

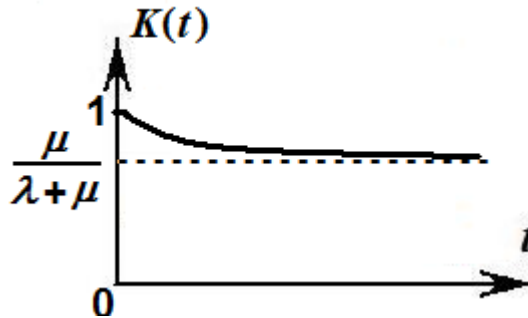


Figure 3. Graph of $K(t)$ with initial conditions $p_0(0) = 1$.

By the way, if initial condition is $p_1(0) = 1$, then the graph of $K(t)$ will be as follows:

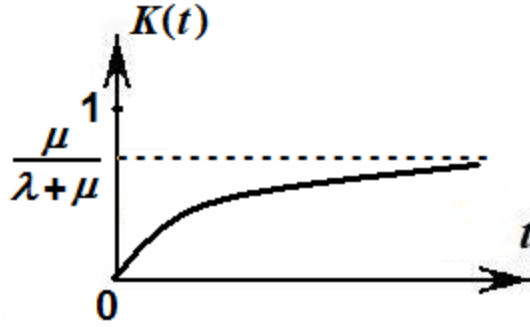


Figure 4. Graph of $K(t)$ with initial conditions $p_1(0) = 1$.

The graph shows that after awhile $K(t)$ approaches stationary value independently on the initial state. Since index $K(t)$ is practically almost never used in practice, we restrict ourselves by considering it for a recoverable unit.

Stationary Availability Coefficient

As we mentioned above, if $t \rightarrow \infty$, $K(t)$ approaches its limit value that is called the stationary availability coefficient, K :

$$K = \lim_{t \rightarrow \infty} K(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} \lim_{t \rightarrow \infty} \exp[-(\lambda + \mu)t] = \frac{\mu}{\lambda + \mu} = \frac{T}{T + \tau} \quad (16)$$

Actually, availability coefficient can be defined as an average portion of time, when a unit is in operating state. In turn, this is the average portion of time when a unit is in operating state during a single cycle “operating-recovering”. So, expression (16) for a recoverable unit can be written directly from the definition of availability coefficient.

It is time to repeat that we use all this rather sophisticated mathematics solely to demonstrate general methodology on simplest examples.

In practice, one usually considers highly reliable objects, for which condition $\frac{E\{\eta\}}{E\{\xi\}} \ll 1$ is satisfied. In this case, it is possible to use a very good approximation:

$$K = \frac{T}{T + \tau} = \frac{1}{1 + \frac{\tau}{T}} \approx 1 - \frac{\tau}{T} = 1 - \frac{\lambda}{\mu} \quad (17)$$

Error of this approximation does not exceed $\left(\frac{\lambda}{\mu}\right)^2$.

Probability of Failure-Free Operation

Since MTTF is given and Markov model is considered, one can immediately write:

$$P(t) = p_0(t) = e^{-\lambda t}. \quad (18)$$

Notice that for highly reliable unit there is a simple and accurate approximation:

$$P_o(t) \approx 1 - \lambda t. \quad (19)$$

This approximation has an error of order $(\lambda t)^2$.

Coefficient of interval availability

The easiest way to get this reliability index (that we denote by $R(t_0)$) is to use memoryless, or Markovian property. For this case, we can just multiply availability coefficient by PFFO, i.e.

$$R(t_0) = P(t, t_0) \cdot K = P(0, t_0) \cdot K = \frac{T}{T + \tau} \exp(-\lambda t_0). \quad (20)$$

For highly reliable systems, one can write an approximation:

$$R(t_0) \approx 1 - \lambda(t_0 + \tau). \quad (21)$$

Remark: We analyzed this simple case with such a scrupulosity only to demonstrate different possible ways of obtaining the needed result. We do this to avoid explanations below with unnecessary additional details for more complex models. The same purpose drives us to use a homogeneous mathematical technique for all routine approaches.

5.2 Series System

Recoverable series systems differ by their recovery processes. First of all, some systems have to be turned off during recovery after failure. In this case there is a single failed unit under restoration. Another case: system continues to stay in an operational state, so during recovering a currently failed unit there may appear new failures. In principle, in this case one can observe even a situation when all system's units are failed. It can happen if, for instance, a recovery process is very slow.

In addition, the number of repair facilities can be restricted, so failed units can form a queue for recovering.

5.2.1 Turning-off system during recovery

Assume that distributions of the TTF, $F_i(t)$, and of the recovery time, $G_i(t)$, are exponential for all units. Denote parameters of these d.f.'s by λ_i and μ_i , respectively.

Let after failure of any unit, the system is turned off during recovery of the failed unit, so other units cannot fail until recovery completion. The transition graph for such a system is presented in Figure 5.

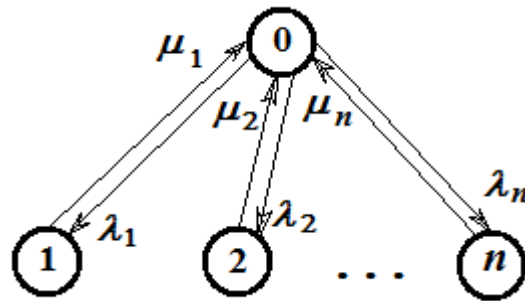


Figure 5. Transition graph for series system that is turning-off during recovery.

We won't write the equations to obtain results for this case. As much as possible, we will try to use simple verbal explanations.

Probability of a failure-free operation

Any exit from state "0" (see Figure 5) leads to failure. Hence,

$$P(t) = \exp\left(-\sum_{1 \leq i \leq n} \lambda_i t\right) . \quad (22)$$

Thus, by its PFFO the system is equivalent to a single unit with a failure rate equals Λ :

$$\Lambda = \sum_{1 \leq i \leq n} \lambda_i \quad (23)$$

Mean time to failure

From (23), one easily finds that

$$T = \frac{1}{\sum_{1 \leq i \leq n} \lambda_i}. \quad (24)$$

If all system's units have exponential distributions of TTF, then the system MTTF and MTBF coincide.

Mean recovery time.

Let us consider a general case, where all units differ by their repair time $\tau_i = \frac{1}{\mu_i}$. It is clear that a current system's failure due to unit i is:

$$p_i = \frac{\lambda_i}{\Lambda}, \quad (25)$$

where Λ is defined in (23).

Thus, mean recovery time can be found easily as weighed value:

$$\tau = \sum_{1 \leq i \leq n} \frac{p_i}{\mu_i} = \frac{1}{\Lambda} \sum_{1 \leq i \leq n} \frac{\lambda_i}{\mu_i}. \quad (26)$$

Stationary Availability Coefficient

Using (23) and (26), one easily writes

$$K = \frac{1}{1 + \sum_{1 \leq i \leq n} \lambda_i \tau_i} \approx 1 - \sum_{1 \leq i \leq n} \lambda_i \tau_i. \quad (27)$$

It is important to notice that if distributions $F_i(t)$ and $G_i(t)$ are not exponential, the expression:

$$K = \frac{1}{1 + \sum_{1 \leq i \leq n} \frac{\tau_i}{T_i}} \approx 1 - \sum_{1 \leq i \leq n} \frac{\tau_i}{T_i}. \quad (28)$$

remains valid. (Conditions of high reliability for the approximation correctness conserves.)

Stationary Interval Availability Coefficient

Since distribution of the system's TTF is exponential, we can use the expression $R(t_0)=KP(t_0)$ where $P(t_0)$ and K are defined in (22). and (27), respectively.

5.2.2 System in poperating state during recovery. Unrestricted repair.

Consider a recoverable series system of n independent units with n independent repair facilities for a case when the system continues to stay in up state, so any its unit may fail during recovery process of the previously faied unit.

In this case, the system's reliability indices can be obtained in a very siple way.

Probability of a failure-free operation and mean time to failure

The system PFFO and MTTF coincide with those considered above in (22) and (24). It is also clear that MTBF is equal to MTTF, since all units have exponential distribution of TTF.

Mean recovery time.

Let us consider a general case where each unit has its own mean repair time $\tau_i = \frac{1}{\mu_i}$. It is clear that a current system's failure due to unit i occurs with probability:

$$p_i = \frac{\lambda_i}{\Lambda}, \quad (29)$$

where Λ is the total system failure rate defined in (23).

Thus, mean recovery time can be found easily as weighed value:

$$\tau = \sum_{1 \leq i \leq n} \frac{p_i}{\mu_i} = \frac{1}{\Lambda} \sum_{1 \leq i \leq n} \frac{\lambda_i}{\mu_i}. \quad (30)$$

Notice that recovery time has a hyper-exponential distribution (see Appendix A.2.6):

$$\Pr\{\eta_{C_{ucm.}} \geq t\} = \frac{\lambda_i \exp(-\lambda_i t)}{\sum_{i=1}^n \lambda_i} \quad (31)$$

Stationary Availability Coefficient

For independent units, one immediately writes:

$$K = \prod_{1 \leq i \leq n} \frac{1}{1 + \lambda_i \tau_i} \approx 1 - \sum_{1 \leq i \leq n} \lambda_i \tau_i. \quad (32)$$

(Conditions of approximation correctness are the same as above in analogous cases.)

Stationary Operational Availability Coefficient

Since distribution of the system TTF is exponential, we can use the expression $R(t_0) = KP(t_0)$ where $P(t_0)$ and K are defined in (22). and (27), respectively.

Remark: If the number of independent repair facilities k is less than n , solution for different units becomes very clumsy and actually has pure “academic” interest.

Just for demonstration the mentioned above fact, let us consider relatively simple system of 2 different independent units and one repair facility. Assume that repair of failed units conforms to the rule “first-in-first-out”. (In Figure 6, system’s failure states are shadowed.)

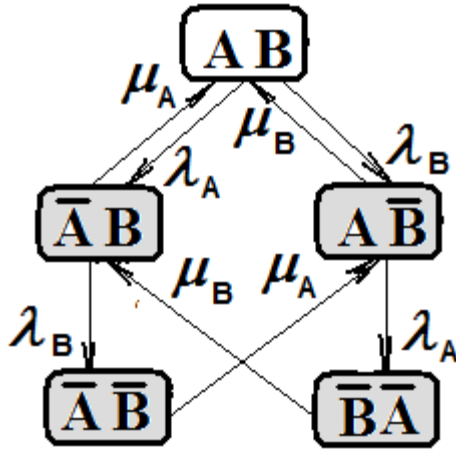


Figure 6. Example of a transition graph for a series system of two units with a single repair facility.

5.2.3 System in operating state during recovery. Restricted repair.

Sometimes when repair facilities are restricted in its ability of simultaneous recovery several failures, a queue of failed units can be formed. Naturally, it leads to the recovery time increase. Analysis of such systems in general case cannot lead to obtaining convenient formulas. However, if all units of series system are assumed identical (rather rear case in real engineering practice!), the problem of finding reliability indices becomes solvable. In this case one can use birth-and-death model described in Appendix C2. In this particular case, when there are only k repair facilities ($k < n$) for a series system of n units, transition graph has the form depicted in Figure 7:

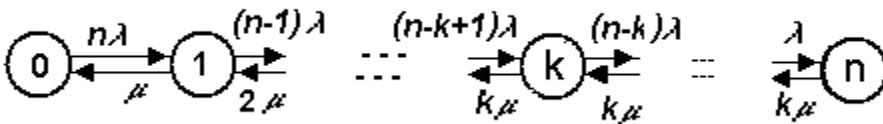


Figure 7. Transition graph for a system of n identical units with k repair facilities.

In the considered case, the system of equations is:

$$\begin{aligned}
 n\lambda p_0 &= \mu p_1 \\
 (n-1)\lambda p_1 &= 2\mu p_2 \\
 &\dots \\
 (n-k+1)\lambda p_{k-1} &= k\mu p_k \\
 (n-k)\lambda p_k &= k\mu p_{k+1}
 \end{aligned} \tag{33}$$

$$\dots$$

$$\lambda p_{n-1} = k\mu p_n .$$

Standard solution of the system (33), given in Appendix C.1.4, has the form:

$$p_1 = n\rho p_0 = \binom{n}{1} \rho p_0$$

$$p_2 = \frac{n-1}{2} \rho p_1 = \frac{n(n-1)}{2} \rho^2 p_0 = \binom{n}{2} \rho^2 p_0$$

$$\dots$$

$$p_k = \frac{n-k+1}{k} \rho p_{k-1} = \frac{n(n-1)\dots(n-k+1)}{1\cdot 2\cdot \dots\cdot k} \rho^k p_0 = \binom{n}{k} \rho^k p_0$$

$$p_{k+1} = \frac{n-k}{k} \rho p_k = \frac{n(n-1)\dots(n-k)}{1\cdot 2\cdot \dots\cdot k\cdot k} \rho^k p_0 = \binom{n}{k-1} \frac{n-k}{k} \rho^{k+1} p_0$$

$$\dots$$

$$p_n = \binom{n}{k} \frac{(n-k)!}{k^{n-k}} \rho^n p_0, \tag{34}$$

where $\rho = \frac{\lambda}{\mu}$.

Since sum of all these probabilities equals 1 (condition of the total probability), from (34), one easily finds the stationary availability coefficient:

$$K = p_0 = \left[1 + \sum_{i=1}^k \binom{n}{i} \rho^i + \binom{n}{k} \rho^k \sum_{j=0}^{n-k} \frac{(n-k)!}{j! k^j} \rho^j \right]^{-1} \tag{35}$$

If the considered system is highly reliable, i.e. $\rho \ll \frac{1}{n}$, one can write approximation:

$$K \approx 1 - n\rho. \tag{36}$$

Actually, it says that for highly reliable systems, it is enough to have a single repair facility, because the probability of occurrence another failure during recovery time is infinitesimally small.

5.3 Dubbed system

We begin with this particular case of parallel systems since because it allows demonstrate mathematical technique on clear and understandable level.

5.3.1 General description

A dubbed recoverable system with embedded loaded redundant unit is probably the most common case of redundancy in engineering practice. This simple structure allows perform a general analysis for all possible configurations: loaded and unloaded redundancy for restricted and unrestricted number of repair facilities. Transition graphs for all these cases are presented in Figure 7.

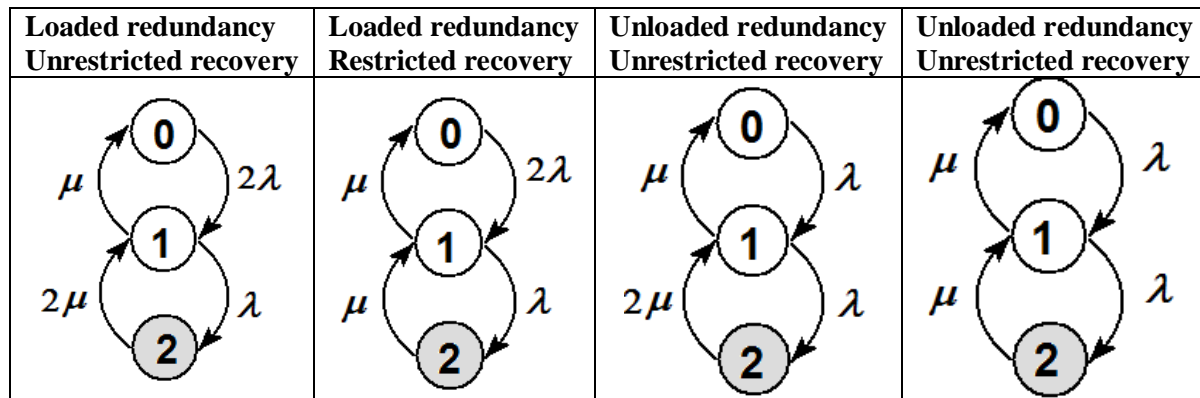


Figure 7. Transition graphs for four cases of recoverable dubbed system.

Usually, one makes the following assumptions:

1. In case of loaded redundancy, failures of both units occur independently;
2. After failure of an operating unit, switching to redundant unit is instantaneous and absolutely reliable;
3. Recovering of failed unit begins immediately if there are available repairing resources;
4. After recovering a unit becomes as well as initial.

We will find solution of the problem in general case, using the transition graph depicted in Figure 8.

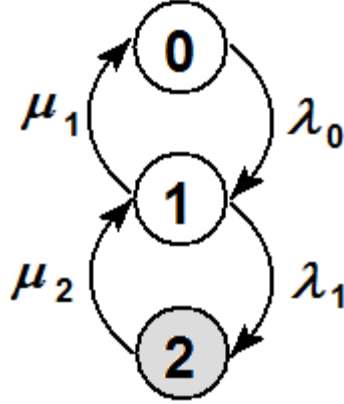


Figure 8. Transition graph for general Markov model of recoverable dubbed system.

5.3.2 Non-stationary availability coefficient

Let us write a system of differential equations in the same way as we deed it for a single recoverable unit.

$$\begin{aligned} \frac{d}{dt} p_0(t) &= -\lambda_0 p_0(t) + \mu_1 p_1(t) \\ \frac{d}{dt} p_1(t) &= \lambda_0 p_0(t) - (\lambda_1 + \mu_1) p_1(t) + \mu_2 p_2(t) \\ 1 &= p_0(t) + p_1(t) + p_2(t) \\ p_0(0) &= 1 . \end{aligned} \tag{37}$$

Laplace transform gives us the following system of algebraic equations:

$$\begin{aligned} -1 + s\varphi_0(s) &= -\lambda_0\varphi_0(s) + \mu_1\varphi_1(s) \\ s\varphi_1(s) &= \lambda_0\varphi_0(s) - (\lambda_1 + \mu_1)\varphi_1(s) + \mu_2\varphi_2(s) \\ \frac{1}{s} &= \varphi_0(s) + \varphi_1(s) + \varphi_2(s) \end{aligned} \tag{38}$$

Since for dubbed system both states «0» and «1» are operational, expression for availability coefficient can be written as:

$$K(t) = p_0(t) + p_1(t) = 1 - p_2(t) , \tag{39}$$

i.e. LST of function $K(t)$ is

$$\varphi_K(s) = 1/s - \varphi_2(s). \quad (40)$$

To find $\varphi_K(s) = 1/s - \varphi_2(s)$ from system of equation (38), we can apply the Cramer Rule:

$$\begin{aligned} \varphi_2(s) &= \frac{\begin{vmatrix} \lambda_0 + s & -\mu_1 & 1 \\ -\lambda_0 & \lambda_1 + \mu_1 + s & 0 \\ s & s & 1 \end{vmatrix}}{\begin{vmatrix} \lambda_0 + s & -\mu_1 & 0 \\ -\lambda_0 & \lambda_1 + \mu_1 + s & -\mu_2 \\ s & s & s \end{vmatrix}} = \\ &= \frac{\lambda_0 \lambda_1}{s[s^2 + s(\lambda_0 + \lambda_1 + \mu_1 + \mu_2) + \lambda_0 \lambda_1 + \lambda_0 \mu_2 + \mu_1 \mu_2]}. \end{aligned} \quad (41)$$

Finally, expression for needed LST is:

$$\varphi_K(s) = \frac{1}{s} - \varphi_2(s) = \frac{s^2 + s(\lambda_0 + \lambda_1 + \mu_1 + \mu_2) + \lambda_0 \mu_2 + \mu_1 \mu_2}{s[s^2 + s(\lambda_0 + \lambda_1 + \mu_1 + \mu_2) + \lambda_0 \lambda_1 + \lambda_0 \mu_2 + \mu_1 \mu_2]} \quad (42)$$

For preparing to inverse LST, let us present (3.7) in the form of simple fractions of the following form:

$$\varphi_K(s) = \frac{A}{s - s_1} + \frac{B}{s - s_2} + \frac{C}{s - s_3}, \quad (43)$$

where s_1, s_2 and s_3 are roots of denominator in (42). In this case, s_2 and s_3 are conjugate roots, and $s_1 = 0$:

$$s_{2,3} = -\frac{\alpha}{2} \pm \sqrt{\frac{\alpha^2}{4} - \beta}, \quad (44)$$

where, in turns, $\alpha = \lambda_0 + \lambda_1 + \mu_1 + \mu_2$ and $\beta = \lambda_0 \lambda_1 + \lambda_0 \mu_2 + \mu_1 \mu_2$. (Naturally, if $s_2 = s_3$, one uses L'Hopital¹⁴ rule.)

The next step is finding coefficients A, B and C in (3.8). First, make in (3.8) reduction to a common denominator

¹⁴ **Guillaume François Antoine, Marquis de l'Hôpital** (1661-1704) was a French mathematician. l'Hôpital's rule for calculating limits involving indeterminate forms $0/0$ and ∞/∞ did not originate with l'Hôpital, it appeared in print for the first time in his famous book, which was a first systematic exposition of differential calculus.

$$\varphi_K(s) = \frac{A(s-s_2) \cdot (s-s_3) + B(s-s_1) \cdot (s-s_3) + C(s-s_1) \cdot (s-s_2)}{(s-s_1) \cdot (s-s_2) \cdot (s-s_3)} \quad (45)$$

Two fractions (42) and (45) with equal denominators are equal if and only if polynomials in nominators are also equal. This condition permits us to write immediately:

$$s^2 + s(\lambda_0 + \lambda_1 + \mu_1 + \mu_1) + \lambda_0\mu_2 + \mu_1\mu_2 = A(s-s_2) \cdot (s-s_3) + B(s-s_1) \cdot (s-s_3) + C(s-s_1) \cdot (s-s_2) \quad (46)$$

The right side of the equality (46) can be rewritten as

$$s^2(A+B+C) - s[A(s_2+s_3) + B(s_1+s_3) + C(s_1+s_2)] + As_2s_3 + Bs_1s_3 + Cs_1s_2 \quad (47)$$

In result, (46) and (47) allow us to compile a new system of algebraic equations for finding coefficients A, B and C:

$$\begin{aligned} A+B+C &= 1, \\ A(\lambda+\mu) - 2B(\lambda+\mu) - C(\lambda+\mu) &= 3\lambda+3\mu \\ A(\lambda+\mu)^2 &= 2\lambda^2 + 4\lambda\mu + 2\mu^2 \end{aligned} \quad (48)$$

We omit simple however rather boring mathematical exercises and present the final result for all four different cases of recoverable dubbed system in the table below:

Table 1. Formulas of $K(t)$ for four cases of recoverable dubbed system.

Loaded redundancy <i>Unrestricted recovery</i>	$1 - \frac{2\lambda^2}{\varepsilon_1\varepsilon_2} \left[1 - \frac{1}{\varepsilon_1 - \varepsilon_2} \right] (\varepsilon_1 e^{-\varepsilon_1 t} - \varepsilon_2 e^{-\varepsilon_2 t}) \approx 1 - \gamma^2 \left[1 - \left(2 - \exp\left(-\frac{\lambda t}{\gamma}\right) \right) \exp\left(-\frac{\lambda t}{\gamma}\right) \right],$ <i>where</i> $\varepsilon_{1,2} = \frac{\lambda}{2\gamma} (1 + \gamma)(3 \pm 1)$ <i>and</i> $\gamma = \frac{\lambda}{\mu}$
Loaded redundancy <i>Restricted recovery</i>	$1 - \frac{2\lambda^2}{\varepsilon_1\varepsilon_2} \left[1 - \frac{1}{\varepsilon_1 - \varepsilon_2} \right] (\varepsilon_1 e^{-\varepsilon_1 t} - \varepsilon_2 e^{-\varepsilon_2 t}) \approx 1 - 2\gamma^2 \left[1 - \left(1 + \frac{\lambda t}{\gamma} \right) \exp\left(-\frac{\lambda t}{\gamma}\right) \right],$ <i>where</i> $\varepsilon_{1,2} = \frac{\lambda}{2\gamma} \left(2 + 3\gamma \pm \sqrt{4\gamma + \gamma^2} \right)$ <i>and</i> $\gamma = \frac{\lambda}{\mu}$
Unloaded redundancy <i>Unrestricted recovery</i>	$1 - \frac{2\lambda^2}{\varepsilon_1\varepsilon_2} \left[1 - \frac{1}{\varepsilon_1 - \varepsilon_2} \right] (\varepsilon_1 e^{-\varepsilon_1 t} - \varepsilon_2 e^{-\varepsilon_2 t}) \approx 1 - \frac{\gamma^2}{2} \left[1 - \left(2 - \exp\left(-\frac{\lambda t}{\gamma}\right) \right) \exp\left(-\frac{\lambda t}{\gamma}\right) \right],$ <i>where</i> $\varepsilon_{1,2} = \frac{\lambda}{2\gamma} \left(3 + 2\gamma \pm \sqrt{1 + 4\gamma} \right)$ <i>and</i> $\gamma = \frac{\lambda}{\mu}$

Unloaded redundancy Restricted recovery	$1 - \frac{2\lambda^2}{\varepsilon_1 \varepsilon_2} \left[1 - \frac{1}{\varepsilon_1 - \varepsilon_2} \right] (\varepsilon_1 e^{-\varepsilon_1 t} - \varepsilon_2 e^{-\varepsilon_2 t}) \approx 1 - \gamma^2 \left[1 - \left(1 + \frac{\lambda t}{\gamma} \right) \exp\left(-\frac{\lambda t}{\gamma}\right) \right],$ <p style="text-align: center;">where $\varepsilon_{1,2} = \frac{\lambda}{\gamma} (1 + \gamma \pm \sqrt{\gamma})$ and $\gamma = \frac{\lambda}{\mu}$</p>
--	---

Let us notice that in this case when both units are mutually independent (case of loaded redundancy with unrestricted repair facilities), the solution for $K(t)$ can be obtained directly from the definition of dubbed system of mentioned type:

$$K(t) = 1 - [1 - K_{unit}(t)]^2 = 1 - \left[\frac{\lambda}{\lambda + \mu} \cdot e^{-(\lambda + \mu)t} \right]^2 = 1 - \frac{\lambda^2}{(\lambda + \mu)^2} \cdot e^{-2(\lambda + \mu)t} \quad (49)$$

REMARK: Notice again that non-stationary availability coefficient in practice is used extremely rare, so these deductions were done exclusively from methodological viewpoint, rather than for practical purposes.

We omit deriving availability coefficient for other particular cases, since it will be a boring use of the same standard methods. Nevertheless, it is interesting to compare all four cases.

5.3.3 Stationary availability coefficient

For finding this reliability index, one can use the equation system (37), replacing all derivatives by 0, all $p_k(t)$ by constant p_k , and omitting the initial condition. Then one gets the following system of algebraic equations:

$$\begin{cases} 0 = -\lambda_0 p_0 + \mu_1 p_1 \\ 0 = \lambda_0 p_0 - (\lambda_1 + \mu_1) p_1 + \mu_2 p_2 \\ 1 = p_0 + p_1 + p_2 \end{cases} \quad (50)$$

From the first equation, one gets

$$p_0 = \frac{\mu_1}{\lambda_0} p_1 \quad (51)$$

After substitution (51) into the second equation in (50), one gets:

$$0 = \mu_1 p_1 - (\lambda_1 + \mu_1) p_1 + \mu_2 p_2 = -\lambda_1 p_1 + \mu_2 p_2 \Rightarrow p_1 = \frac{\mu_2}{\lambda_1} p_2 \quad (52)$$

Finally, after substitution (51) and (52) into the third equation in (50):

$$p_2 + \frac{\mu_2}{\lambda_1} p_2 + \frac{\mu_1}{\lambda_0} \cdot \frac{\mu_2}{\lambda_1} p_2 = 1 \Rightarrow p_2 = \frac{1}{1 + \frac{\mu_2}{\lambda_1} + \frac{\mu_1}{\lambda_0} \cdot \frac{\mu_2}{\lambda_1}} = \frac{\lambda_0 \lambda_1}{\lambda_0 \lambda_1 + \lambda_0 \mu_2 + \mu_1 \mu_2} \quad (53)$$

or, finally:

$$K = \left(1 + \frac{\lambda_0 \lambda_1}{\lambda_0 \mu_2 + \mu_1 \mu_2} \right)^{-1}. \quad (54)$$

From (54), it easy to compile the following table.

Table 2. Formulas of K for four cases of recoverable dubbed system.

Loaded redundancy <i>Unrestricted recovery</i>	$\frac{1}{1 + \gamma^*} \approx 1 - \gamma^2$, where $\gamma^* = \frac{\gamma^2}{1 + 2\gamma}$ and $\gamma = \frac{\lambda}{\mu}$.
Loaded redundancy <i>Restricted recovery</i>	$\frac{1}{1 + \gamma^*} \approx 1 - 2\gamma^2$, where $\gamma^* = \frac{2\gamma^2}{1 + 2\gamma}$ and $\gamma = \frac{\lambda}{\mu}$.
Unloaded redundancy <i>Unrestricted recovery</i>	$\frac{1}{1 + \gamma^*} \approx 1 - \frac{\gamma^2}{2}$, where $\gamma^* = \frac{\gamma^2}{2(1 + \gamma)}$ and $\gamma = \frac{\lambda}{\mu}$.
Unloaded redundancy <i>Restricted recovery</i>	$\frac{1}{1 + \gamma^*} \approx 1 - \gamma^2$, where $\gamma^* = \frac{\gamma^2}{1 + \gamma}$ and $\gamma = \frac{\lambda}{\mu}$.

Naturally, availability coefficient for the first case can be obtained directly:

$$K = 1 - (1 - K)^2 = 1 - \left(\frac{\lambda}{(\lambda + \mu)} \right)^2 = 1 - \frac{1}{\left(1 + \frac{1}{\gamma} \right)^2} = 1 - \frac{\gamma^2}{(1 + \gamma)^2} \approx 1 - \gamma^2 \quad (55)$$

5.3.4 Probability of failure-free operation

Transition graph s for this case are presented in Figure 9. State “2” corresponding to system’s failure is absorbing one. Since there is no difference between restricted and unrestricted repair.

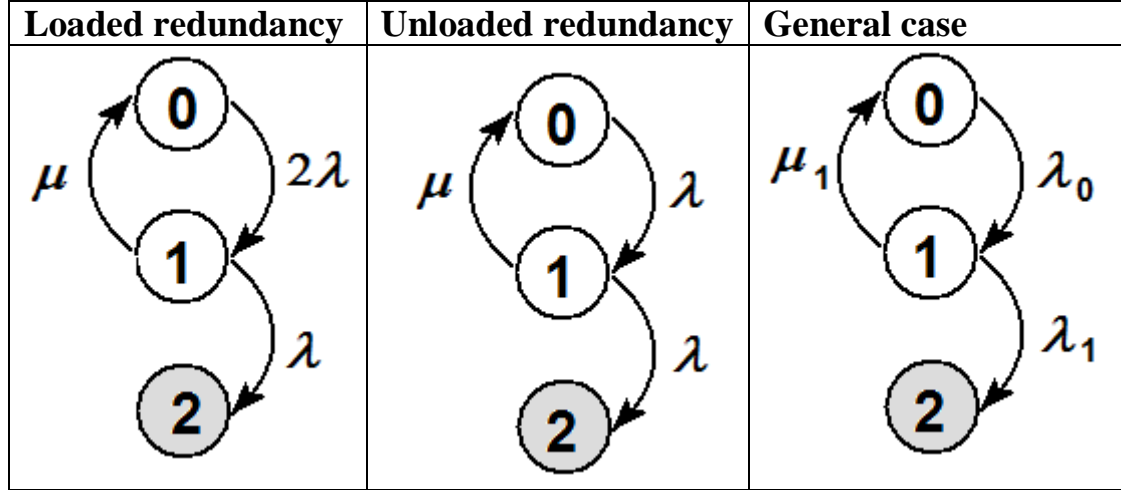


Figure 9. Transition graphs for calculations $P(t)$ of recoverable dubbed system.

The system of linear differential equations in this case is:

$$\begin{aligned} \frac{d}{dt} p_0(t) &= -\lambda_0 p_0(t) + \mu_1 p_1(t) \\ \frac{d}{dt} p_1(t) &= \lambda_0 p_0(t) - (\lambda_1 + \mu_1) p_1(t) \\ p_0(0) &= 1. \end{aligned} \quad (56)$$

After applying Laplace transform, one gets the following system of algebraic equations:

$$\begin{aligned} -1 + s\varphi_0(s) &= -\lambda_0\varphi_0(s) + \mu_1\varphi_1(s) \\ s\varphi_1(s) &= \lambda_0\varphi_0(s) - (\lambda_1 + \mu_1)\varphi_1(s) \end{aligned} \quad (57)$$

Since for a dubbed system $P^{(0)}(t) = p_0(t) + p_1(t)$, the solution in terms of Cramer Rule has the form:

$$\varphi^{(0)}(s) = \varphi_0(s) + \varphi_1(s) = \frac{\begin{vmatrix} 1 & -\mu_1 \\ 0 & \lambda_1 + \mu_1 + s \end{vmatrix} + \begin{vmatrix} \lambda_0 + s & 1 \\ -\lambda_0 & 0 \end{vmatrix}}{\begin{vmatrix} \lambda_0 + s & -\mu_1 \\ -\lambda_0 & \lambda_1 + \mu_1 + s \end{vmatrix}} = \frac{s + \lambda_0 + \lambda_1 + \mu_1}{s^2 + s(\lambda_0 + \lambda_1 + \mu_1) + \lambda_0\lambda_1}. \quad (58)$$

Here the superscript at $P^{(0)}(t)$ and $\varphi^{(0)}(s)$ denotes that solution has been got

under the mentioned above initial conditions.

Using the same procedure as above, one gets in this case solution:

$$P^{(0)}(t) = \frac{1}{s_1^* - s_2^*} (s_1^* \exp(s_2^* t) - s_2^* \exp(s_1^* t)), \quad (59)$$

where $s_{1,2}^* = -\frac{\alpha^*}{2} \pm \sqrt{\left(\frac{\alpha^*}{2}\right)^2 - \beta^*}$, $\alpha^* = \lambda_0 + \lambda_1 + \mu_1$ and $\beta^* = \lambda_0 \lambda_1$.

Below we will need a solution of the same system of linear differential equations with initial conditions $p_1(0) = 1$, i.e. when the starting moment is a moment of the system's recovery completion. The corresponding system of algebraic equations for Laplace transforms has the form:

$$\begin{aligned} s\varphi_0(s) &= -\lambda_0\varphi_0(s) + \mu_1\varphi_1(s) \\ -1 + s\varphi_1(s) &= \lambda_0\varphi_0(s) - (\lambda_1 + \mu_1)\varphi_1(s) \end{aligned} \quad (60)$$

Let us omit routine deductions absolutely similar those above, and write the result:

$$\varphi^{(1)}(s) = \varphi_0(s) + \varphi_1(s) = \frac{\begin{vmatrix} 0 & -\mu_1 \\ 1 & \lambda_1 + \mu_1 + s \end{vmatrix} + \begin{vmatrix} \lambda_0 + s & 0 \\ -\lambda_0 & 1 \end{vmatrix}}{\begin{vmatrix} \lambda_0 + s & -\mu_1 \\ -\lambda_0 & \lambda_1 + \mu_1 + s \end{vmatrix}} = \frac{\lambda_0 + \mu_1}{s^2 + s(\lambda_0 + \lambda_1 + \mu_1) + \lambda_0 \lambda_1}. \quad (61)$$

Inverse Laplace transform is:

$$P^{(1)}(t) = \frac{1}{s_1^* - s_2^*} \left[(s_1^* - \lambda_0 - \lambda_1) \exp(s_2^* t) - (s_2^* - \lambda_0 - \lambda_1) \exp(s_1^* t) \right] \quad (62)$$

where roots s_1^* and s_2^* are the same as for (59). Here again the superscript at $P^{(1)}(t)$ corresponds to initial condition $p_1(0) = 1$.

5.3.5 Stationary coefficient of interval availability

This is one of the most important reliability indices of recoverable systems. For dubbed system, this index can be written as a formula of total probability:

$$R(t_0) = p_0 P^{(0)}(t) + p_1 P^{(1)}(t), \quad (63)$$

where p_0 and p_1 are stationary probabilities of states “0” and “1”, respectively; probabilities $P^{(0)}(t)$ and $P^{(1)}(t)$ are conditional PFFO’s for successful completion an operation, starting at corresponding states. Probabilities $P^{(0)}(t)$ and $P^{(1)}(t)$ are taken from (59) and (62), correspondingly. Stationary probabilities p_0 and p_1 can be found by using (51) – (53).

$$p_1 = \frac{\mu_2}{\lambda_1} p_2 = \frac{\lambda_0 \mu_2}{\lambda_1 (\lambda_0 \lambda_1 + \lambda_0 \mu_2 + \mu_1 \mu_2)} \quad (64)$$

and

$$p_0 = \frac{\mu_1}{\lambda_0} p_1 = \frac{\mu_1 \mu_2}{\lambda_1 (\lambda_0 \lambda_1 + \lambda_0 \mu_2 + \mu_1 \mu_2)}. \quad (65)$$

We do not write the obvious final expression for $R(t_0)$ because it is too clumsy. However, approximations for highly reliable systems have rather simple and compact forms, though it is more convenient to write all of them for specific cases:

Table 3. Approximate formulas of $R(t_0)$ for four cases of highly reliable recoverable dubbed systems.

Loaded redundancy <i>Unrestricted recovery</i>	$\frac{1}{1 + \gamma^*} \exp\left(-\frac{2\gamma\lambda t_0}{1 + 2\gamma}\right)$, where $\gamma^* = \frac{\gamma^2}{1 + 2\gamma}$ and $\gamma = \frac{\lambda}{\mu}$.
Loaded redundancy <i>Restricted recovery</i>	$\frac{1}{1 + \gamma^*} \exp\left(-\frac{2\gamma\lambda t_0}{1 + 2\gamma}\right)$, where $\gamma^* = \frac{2\gamma^2}{1 + 2\gamma}$ and $\gamma = \frac{\lambda}{\mu}$.
Unloaded redundancy <i>Unrestricted recovery</i>	$\frac{1}{1 + \gamma^*} \exp\left(-\frac{\gamma\lambda t_0}{1 + 2\gamma}\right)$, where $\gamma^* = \frac{\gamma^2}{2(1 + \gamma)}$ and $\gamma = \frac{\lambda}{\mu}$.
Unloaded redundancy <i>Restricted recovery</i>	$\frac{1}{1 + \gamma^*} \exp\left(-\frac{\gamma\lambda t_0}{1 + 2\gamma}\right)$, where $\gamma^* = \frac{\gamma^2}{1 + \gamma}$ and $\gamma = \frac{\lambda}{\mu}$.

5.3.6 Mean time to failure

MTTF can be easily found from standard formula:

$$T^{(0)} = \int_0^{\infty} P^{(0)}(t) dt \quad (66)$$

Let us notice that Laplace transform of function $P_0(t)$ with substitution $s = 0$ also gives $T^{(0)}$. Indeed,

$$\varphi(s) \Big|_{s=0} = \int_0^{\infty} e^{-st} P^{(0)}(t) dt \Big|_{s=0} = \int_0^{\infty} P^{(0)}(t) dt = T^{(0)} \quad (67)$$

Thus $T^{(0)}$ can be found with the help of Laplace transform (58):

$$T^{(0)} = \frac{s + \lambda_0 + \lambda_1 + \mu_1}{s^2 + s(\lambda_0 + \lambda_1 + \mu_1) + \lambda_0 \lambda_1} \Big|_{s=0} = \frac{\lambda_0 + \lambda_1 + \mu_1}{\lambda_0 \lambda_1} = \frac{1}{\lambda_0} + \frac{1}{\lambda_1} + \frac{\mu_1}{\lambda_0 \lambda_1}, \quad (68)$$

5.3.7 Mean time between failures

For a dubbed system MTTF and MTBF are different. It is clear because functions $P^{(0)}(t)$ and $P^{(1)}(t)$ are different. MTBF can be found with the help of Laplace transform (61):

$$T^{(1)} = \frac{\lambda_0 + \mu_1}{s^2 + s(\lambda_0 + \lambda_1 + \mu_1) + \lambda_0 \lambda_1} \Big|_{s=0} = \frac{\lambda_0 + \mu_1}{\lambda_0 \lambda_1} = \frac{1}{\lambda_0} + \frac{\mu_1}{\lambda_0 \lambda_1}. \quad (67)$$

Notice that $T^{(0)}$ is larger than $T^{(1)}$ on value of $\frac{1}{\lambda_0}$. It is clear: this is the average “travel” time from state “0” to state “1”.

The same result can be obtained also in a different way. From the statement above follows:

$$T^{(0)} = \frac{1}{\lambda_0} + T^{(1)}. \quad (68)$$

Coming to state “1” the process stays there before going to state “0” or state “2” on average time $\frac{1}{\lambda_1 + \mu_1}$ and after it returns to state “0” with probability $\frac{\mu_1}{\lambda_1 + \mu_1}$ or moves to absorbing state “2” with probability $\frac{\lambda_1}{\lambda_1 + \mu_1}$. It gives a possibility to write the following recurrent equation:

$$T^{(1)} = \frac{1}{\lambda_1 + \mu_1} + \frac{\mu_1}{\lambda_1 + \mu_1} T^{(0)}. \quad (69)$$

Substituting (69) into (68) gives the final expression:

$$T^{(0)} = \frac{1}{1 - \frac{1}{\lambda_1 + \mu_1}} \left(\frac{1}{\lambda_0} + \frac{1}{\lambda_1 + \mu_1} \right) = \frac{\lambda_0 + \lambda_1 + \mu_1}{\lambda_0 \lambda_1} = \frac{1}{\lambda_0} + \frac{1}{\lambda_1} + \frac{\mu_1}{\lambda_0 \lambda_1}. \quad (70)$$

From (69) and (70) follows that

$$T^{(1)} = T^{(0)} - \frac{1}{\lambda_0} = \frac{1}{\lambda_1} + \frac{\mu_1}{\lambda_0 \lambda_1}. \quad (71)$$

We give sometimes several different deductions of the same results exclusively for a single purpose: to help the readers to develop “mathematical intuition”.

The final results for all types of recoverable dubbed systems are presented in Table 4.

Table 4. MTTF and MTBF for four types of recoverable dubbed systems

	$T^{(0)}$		$T^{(1)}$	
	exact	approximate	exact	approximate
Loaded redundancy	$\frac{1}{\lambda} \cdot \frac{1+3\gamma}{2\gamma}$	$\frac{1}{2\lambda\gamma}$	$\frac{1}{\lambda} \cdot \frac{1+2\gamma}{2\gamma}$	$\frac{1}{2\lambda\gamma}$
Unloaded redundancy	$\frac{1}{\lambda} \cdot \left(2 + \frac{1}{\gamma} \right)$	$\frac{1}{\lambda\gamma}$	$\frac{1}{\lambda} \cdot \left(1 + \frac{1}{\gamma} \right)$	$\frac{1}{\lambda\gamma}$

5.3.8 Mean recovery time

The mean recovery time, τ , for this simple Markov model coincides with average duration of staying the process in state “2”. As follows from transition graphs, this time is equal to $\frac{1}{\mu_2}$. So, this reliability index depends on the number

of repair facilities. For unrestricted recovery (two repair facilities) $\frac{1}{2\mu}$, and for restricted repair facilities (a single repair facility), $\frac{1}{\mu}$.

5.4 Parallel systems

By definition, a parallel system is factually a single operating unit with a group of identical redundant units, which are independent in sense of failing. Such an idealized scheme has few relations to a real engineering practice though is of theoretical interest. Speaking about parallel systems consisting of recoverable units, one has to keep in mind four main possible cases, presented in Table 5.

Table 5. Main cases of parallel systems.

		Regime of recovering	
		Unrestricted repair	Restricted repair
Regime of redundant units	Loaded	Factually, all n units in parallel are independent	As soon as the number of failed units exceeds the number of repair facilities, failed units form a waiting line.
	Unloaded	Failure of such a system occurs only if during repair of first failed units all other units have failed.	As in above case, when the number of failed units exceeds the number of repair facilities, failed units form a waiting line.

Probably, the mathematical description of such systems is given by “Birth-Death Process that is considered in details in Appendix C2.

Let a parallel system consists of n units, i.e. one operating unit and $n-1$ redundant ones. Assume that there are k repair facilities for recovering failed units, $k \leq n - 1$. Let us denote states by natural numbers $0, 1, 2, \dots$, where the number of a state corresponds to the number of failed units. Then all four cases can be described, actually, by very similar linear transition graphs.

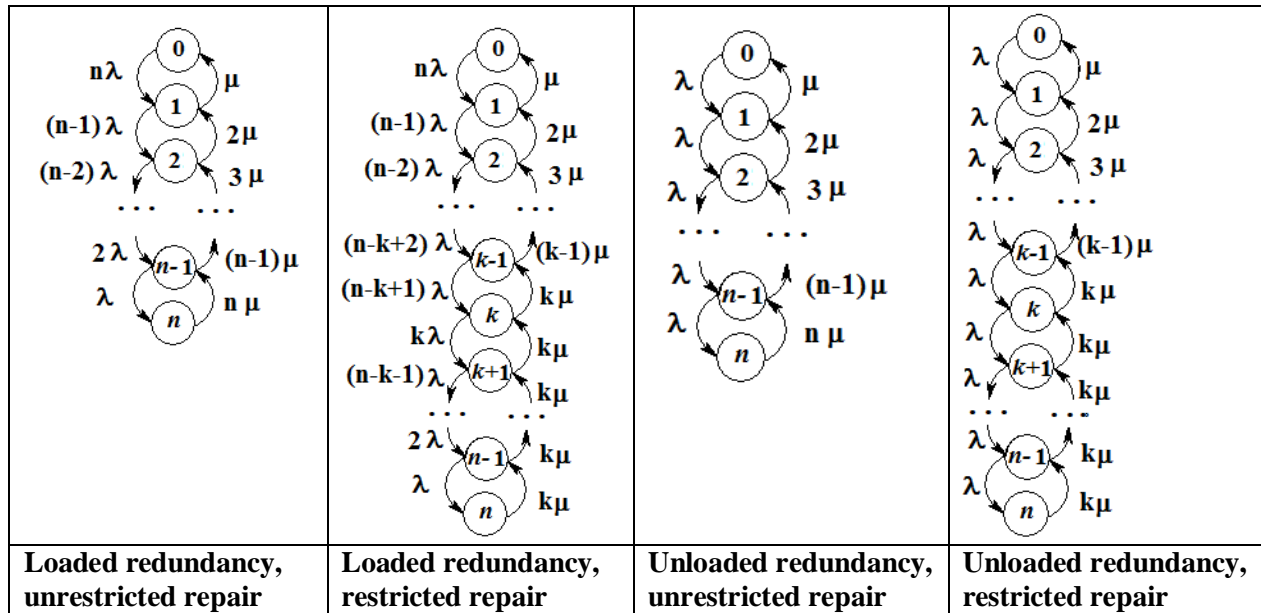


Figure 10. Transition graphs for corresponding “Birth-Death Processes”.

For each of these transition graphs, the system of differential equations, corresponding system of algebraic equations for Laplace transforms and algebraic equations for stationary probabilities can be easily found with the help of Appendix B. We will omit them, first of all, because of “pure parallel” systems with multiple loaded redundant units are rare in engineering practice. Much more interesting is the structure that is described in the next section.

5.5 Structures of type “ m out of n ”

Much more realistic is a series system of identical independent units with a common group of redundant units. Formally, such a structure appears if a system consists of units of several types. A set of units of the same type can be considered as a “series system”, for which there is a stock of spare units. It is reasonable to consider these spare units as unloaded; these units are waiting for being switched into operating position after one of operating units has failed. Failed units are directed to a repair shop, from where after recovery they again enter the system’s stock. Switching of spare unit into an operating position is usually assumed instantaneous. (Of course, this assumption is almost correct if switching time is relatively small.)

In this case, the system as a whole can be presented as a series connection of such “ m_k -out-of- n_k ” subsystems.

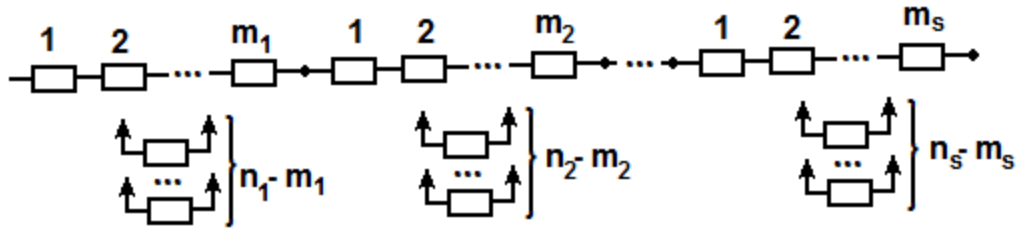


Figure 11. Conditional presentation a system as a series connection of “ m -out-of- n ” subsystems.

Transition graph for one of such series subsystems of m operating units and common group of $n-m$ unloaded spare units is presented below. States of system failure are shadowed.

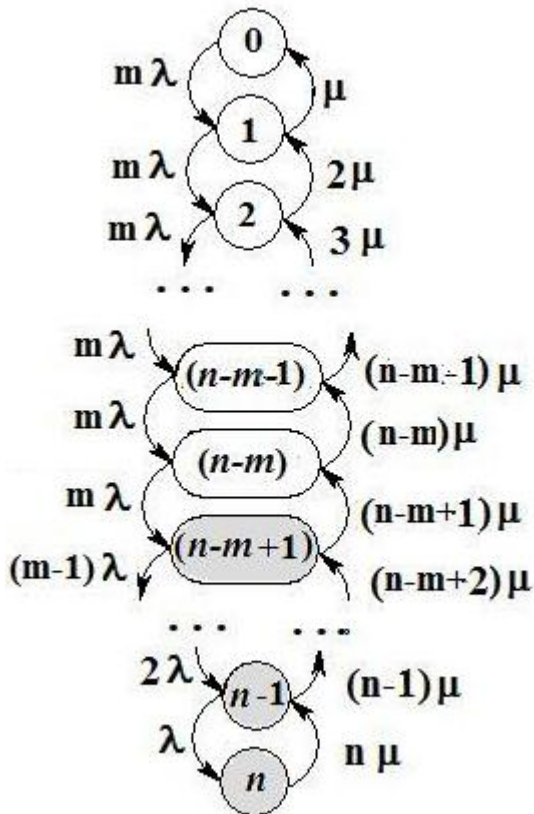


Figure 12. Transition graph of “ m -out-of- n ” system with unloaded spare units.

This model of redundancy is, probably, one of the most useful for practical purposes in reliability engineering. We will repeat most of deductions given in Appendix C2 for this particular, however very important case. We omit only technical details minutely described in the mentioned Appendix.

The system of differential equations in this case takes the form:

$$\begin{aligned}
\frac{d}{dt} p_0(t) &= -m\lambda p_0(t) + \mu\varphi_1(t) \\
\frac{d}{dt} p_1(t) &= m\lambda p_0(t) - p_1(t)(m\lambda + 2\mu) + 2\mu\varphi_2(t) \\
&\dots\dots\dots \\
\frac{d}{dt} p_{n-m-1}(t) &= m\lambda p_{n-m-2}(t) - [m\lambda + (n-m-1)\mu]p_{n-m-1}(t) + (n-m)\mu\varphi_{n-m}(t) \\
\frac{d}{dt} p_{n-m}(t) &= m\lambda p_{n-m-1}(t) - [m\lambda + (n-m)\mu]p_{n-m}(t) + (n-m+1)\mu\varphi_{n-m+1}(t) \\
\frac{d}{dt} p_{n-m+1}(t) &= m\lambda p_{n-m}(t) - [(m-1)\lambda + (n-m+1)\mu]p_{n-m+1}(t) + (n-m+2)\mu\varphi_{n-m+2}(t) \\
&\dots\dots\dots \\
\frac{d}{dt} p_{n-1}(t) &= 2\lambda p_{n-2}(t) - [\lambda + (n-1)\mu]p_{n-1}(t) + n\mu\varphi_n(t) \\
\frac{d}{dt} p_n(t) &= \lambda p_{n-1}(t) - n\mu\varphi_n(t).
\end{aligned} \tag{72}$$

The initial condition in most reliability application is taken in the form $p_0(0)=1$.

Of course, system of equations (72) can be solved with the help of methods described in Appendix B. The probability of failure-free operation can be found as:

$$P(t) = \sum_{0 \leq i \leq n-m} p_i(t) \tag{73}$$

We will not spend time for pure mathematical exercises because such systems are oriented on long run, and, consequently, stationary availability coefficient for such systems is more appropriate reliability index rather than PFFO. In this case the system (72) transforms into the system of algebraic equations:

$$\begin{aligned}
m\lambda p_0 &= \mu\varphi_1 \\
(m\lambda + 2\mu)p_1 &= m\lambda p_0 + 2\mu\varphi_2 = \\
&\dots\dots\dots \\
[m\lambda + (n-m-1)\mu]p_{n-m-1} &= m\lambda p_{n-m-2} + (n-m)\mu\varphi_{n-m} \\
[m\lambda + (n-m)\mu]p_{n-m} &= m\lambda p_{n-m-1} + (n-m+1)\mu\varphi_{n-m+1} \\
[(m-1)\lambda + (n-m+1)\mu]p_{n-m+1} &= m\lambda p_{n-m} + (n-m+2)\mu\varphi_{n-m+2} \\
&\dots\dots\dots \\
[\lambda + (n-1)\mu]p_{n-1} &= 2\lambda p_{n-2} + n\mu\varphi_n
\end{aligned} \tag{74}$$

$$\lambda p_{n-1} = n\mu p_n.$$

Since equations in (74) are mutually dependent, it is necessary additionally to use equation of total probability:

$$\sum_{k=0}^n p_k = 1. \quad (75)$$

As it has been stated in Appendix C2, actually it is more convenient to write the equations of balance for for “cuts” of transition graph rather than for states. It actually leads to the solution almost directly. Remind that the balance means that flows back and forth through a cut between neighbor states of the transition graph are equal. Thus, on the base of thransition graph depicted in Figure 12 we can write:

$$\begin{aligned} m\lambda p_0 &= \mu p_1 \\ m\lambda p_1 &= 2\mu p_2 \\ m\lambda p_2 &= 3\mu p_3 \\ &\dots\dots\dots \\ m\lambda p_{n-m-1} &= (n-m)\mu p_{n-m} \\ m\lambda p_{n-m} &= (n-m+1)\mu p_{n-m+1} \\ (m-1)\lambda p_{n-m+1} &= (n-m+2)\mu p_{n-m+2} \\ &\dots\dots\dots \\ (n-1)\mu p_{n-1} &= 2\lambda p_{n-2} \\ \lambda p_{n-1} &= n\mu p_n. \end{aligned} \quad (76)$$

Introducing for convenience notation $\rho = \frac{\lambda}{\mu}$, one gets system (76) in the form:

$$\begin{aligned} p_1 &= m\rho p_0 \\ p_2 &= \frac{m}{2}\rho p_1 = \frac{m^2}{1 \cdot 2}\rho^2 p_0 \\ p_3 &= \frac{m}{3}\rho p_2 = \frac{m^3}{1 \cdot 2 \cdot 3}\rho^3 p_0 \\ &\dots\dots\dots \\ p_{n-m} &= \frac{m}{n-m}\rho p_{n-m-1} = \frac{m^{n-m}}{(n-m)!}\rho^{n-m} p_0 \\ p_{n-m+1} &= \frac{m}{n-m+1}\rho p_{n-m} = \frac{m \cdot m^{n-m}}{(n-m+1)!}\rho^{n-m+1} p_0 \end{aligned} \quad (77)$$

$$\begin{aligned}
p_{n-m+2} &= \frac{m-1}{n-m+2} \rho p_{n-m+1} = \frac{m \cdot (m-1) \cdot m^{n-m}}{(n-m+2)!} \rho^{n-m+2} p_0 \\
p_{n-m+3} &= \frac{m-2}{n-m+3} \rho p_{n-m+2} = \frac{m \cdot (m-1) \cdot (m-2) \cdot m^{n-m}}{(n-m+3)!} \rho^{n-m+3} p_0 = \\
&= \frac{m!}{(m-3)!} \cdot \frac{m^{n-m}}{(n-m+3)!} \rho^{n-m+3} p_0 \\
&\quad \dots\dots\dots \\
p_{n-2} &= \frac{n-2}{3} \rho p_{n-1} = \frac{m!}{3!} \cdot \frac{m^{n-m}}{(n-2)!} \rho^{n-2} p_0 \\
p_{n-1} &= \frac{n-1}{2} \rho p_{n-1} = \frac{m!}{2!} \cdot \frac{m^{n-m}}{(n-1)!} \rho^{n-1} p_0 \\
p_n &= \frac{n}{1} \rho p_{n-1} = \frac{m!}{1!} \cdot \frac{m^{n-m}}{n!} \rho^n p_0
\end{aligned}$$

Using (75), one can write the solution:

$$p_0 = \left(1 + \sum_{k=1}^{n-m} \frac{m^k}{k!} \rho^k + \sum_{k=n-m+1}^n \frac{m^{n-m}}{k!} \rho^k \prod_{k=n-m+1}^{n-k} \frac{(m-1)!}{(k-n+m-1)!} \right)^{-1}. \quad (78)$$

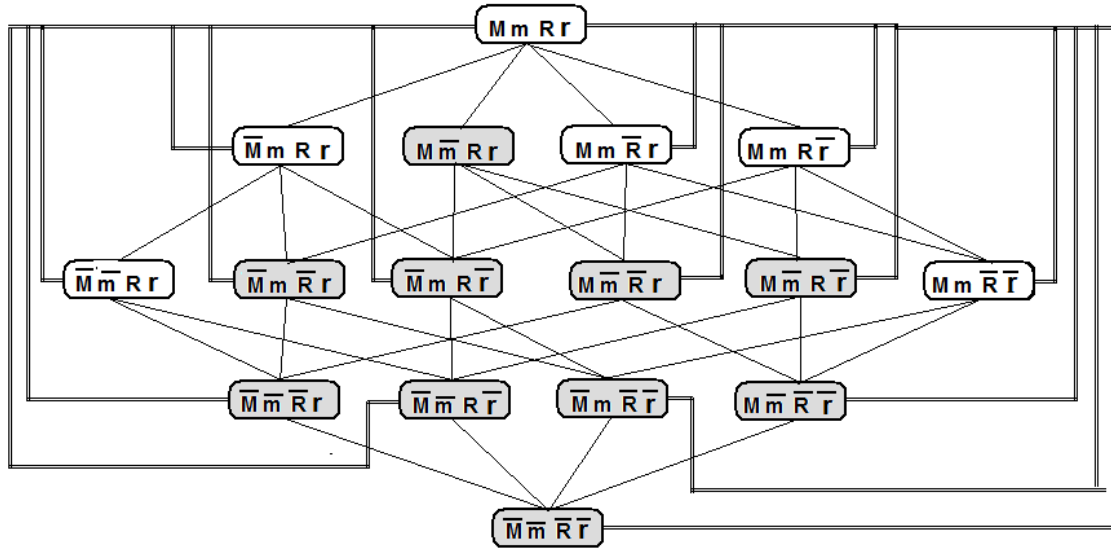
For any p_k solution can be easily obtained by substitution p_0 in corresponding equation of (77).

6 Recoverable systems. Heuristic models

6.1 Preliminary notes

In the previous chapter we presented analysis of recoverable parallel systems by the means of Markov models. The reader had a chance to receive evidence that even idealized simplest models needs rather sophisticated mathematical technique. What if the model will be a little bit more realistic? For instance, the dubbed system has a non-reliable switch? What if a operating unit (unit on operational position) has not complete monitoring of its state, so can leads to a hidden system failure? (Of course, such kinds of important practical features of redundant systems can be continued.) Below a transition graph for the system described above is

presented with additional condition: repair failed part is performing in accordance the rule FIFO (first-in-first-out). In this case, the transition graph is rather ugly!



- Legend:** **M** - monitored part of main unit; **m** - uncontrolled part of main unit;
R - monitored part of redundant unit; **r** - uncontrolled part of redundant unit ;
 - means failure of corresponding part ;
 — means recovering transition ;
 - - means transition due to failure ;
 ○ - operational system state ;
 ◐ - state of system failure (probably,hidden).

Figure 1. Transition graph for “simple” recoverable dubbed system with unreliable switching and incomplete control of operating and redundant units.

Hardly anyone will take courage to solve a corresponding system of differential or even algebraic equations!

At the same time, such a problem can be easily solved with the help of heuristic methods.

	<p>Do not think that heuristic is something taken “<i>from the ceiling</i>”. Heuristic we are talking about is an approximation method based on strong mathematical asymptotical results working for analysis of highly reliable systems.</p> <p>However, what to do if the system is not highly reliable? The answer is</p>
<p>simple: in this case, you should think about improvement of system reliability, not spend your time on senseless “reliability analysis” with deducting ugly and useless “five-store” formulas.</p>	

A researcher is often faced with the problem of finding a "solution" of a problem when the problem is practically unsolvable. However, in spite of the problem's "insolvability", a solution must be found! And even if an exact or “ideal”

solution cannot be found, a designer is forced to make a practical decision, since required problem has to be solved!

Remember that famous legend about “Columbus's egg”. It refers to a brilliant idea or discovery that seems simple and easy after the fact. That story of how Christopher Columbus, having been told that discovering the Americas was no great accomplishment, challenged his critics to make an egg stand on its tip. After his challengers gave up, Columbus did it himself : he tapped it gently on the table breaking it slightly and, with this, the egg stood on its end, because he flattened its tip. His sailors were buzzing: “We would do the same!” Columbus answered: “You wpcould, however I did it!”

When there is no exact analytic solution and the problem is still too hard for even a Monte Carlo simulation, the only possibility is to use a heuristic procedure (heuristics).

Sometimes heuristics are thought to lead to an arbitrary "solution", based only on an “I-personally-believe” type of argument. We oppose such "heuristics", as we understand the term "heuristic" to be an extension of analytical methods in areas where such methods cannot be exactly proven. Sometimes we omit some specified conditions, sometimes we make additional assumptions and are not insure that the method of solution is still correct. Sometimes we change an analyzed phenomenon description to allow the use of available mathematical tools.

In fact, the building of a mathematical model is always a heuristic procedure itself. No mathematical model completely reflects all of the properties of a real object. We always create "an ideal image" based on a real object, and then build a mathematical model for this idealized image.

Moreover, approximate calculations can be viewed as "good proven heuristic". Thus, heuristics is an inevitable part of mathematical modeling.

Below we will introduce several heuristic approaches. Generally, they concern the constructing of models.

Anyway, never use a cannon for hanting birds: a simple fowling-piece is quite enough for this purpose.



Figure 2. "Columbus Breaking the Egg"
by William Hogarth (1667-1764).

6.2 Poisson process

In Reliability Theory, the Poisson process occupies a special place. Models based on the Poisson process are rather simple and very convenient for numerical analysis. However, it is not like the drunk who is looking for his lost keys under the street lamp because that's the most illuminated place, not because that's the place where he dropped the keys. The utilization of the Poisson models in reliability has a very strong empirical background.

Most of complex systems consist of large number of relatively reliable units. Flow of system failures is generated by many "sub flows" of units' failures. These "sub flows" are mutually independent because they are formed by independent units, for highly reliable units probability of intersection of failures are negligible and, finally, after relatively short time the process of system's failure becomes stationary, i.e. does not change its probabilistic properties with time, and these properties become very close to the properties of the Poisson process: the Markov property, the ordinariness and stationarity. Let us explain these properties in more detail.

- (1) *Markov property*, means that the future development of the process does not depend neither on current state of the process, nor on its entire prehistory.
- (2) *Ordinarness* means that there are no simultaneously happening events or even the so-called "points of concentration". In other words, if time interval $\Delta \rightarrow 0$, the probability of occurrence of more than one event, $p_{k>1}(\Delta)$ within this interval becomes infinitesimally small in comparison with $p_1(\Delta)$:

$$\lim_{\Delta \rightarrow 0} \frac{p_{k>1}(\Delta)}{p_1(\Delta)} \rightarrow 0.$$

- (3) *Stationarity* means invariance to the shift operator, i.e. probabilistic characteristics of the process (for instance, mean number of failures in interval of given length t) depend on the length of the interval and do not depend on its location on time axis.

The Poisson process especially well describes the process of generating of electronic equipment failures. Not in vain, the Poisson process is often called "the process of rare events". In the theory of stochastic processes the Poisson process plays a role which is analogous to that of the normal distribution in probability theory.

The Poisson process is defined as a point stochastic process that is formed by sequence of independent random variables with the same exponential distribution. Probability that within fixed time interval $[x, x+t]$ exactly k events of the Poisson process occur has the Poisson distribution:

$$p_k(x, x+t) = \frac{(\lambda t)^k}{k!} \exp(-\lambda t). \quad (1)$$

Parameter λ is called *intensity of the Poisson process*. By definition it is equal to the mean number of events (in our case failures) in a unit of time, that is $\lambda=1/T$, where T is average distance between events (in our case T is MTBF).

Actually, formula (1) complies with the three properties that were formulated almost on qualitative level. Indeed, take an arbitrary fixed interval of length t and divide it into small subintervals Δ_i so that $\sum_{\forall i} \Delta_i = t$.

The mean number of failures within interval Δ_i is:

$$E \{ \Delta_i \} = \lambda \Delta_i. \quad (2)$$

From the other hand, for the same interval Δ_i one can write another expression for the mean number of failures:

$$E \{ \Delta_i \} = \sum_{0 \leq j < \infty} j p_j(\Delta_i), \quad (3)$$

where $p_k(\Delta_i)$ is the probability that exactly k failures will have occurred within interval Δ_i . Taking into account the property of ordinariness, the following equation can be written for total probability:

$$p_0(\Delta_k) + p_1(\Delta_k) + o(\Delta_k) = 1, \quad (4)$$

where $o(\Delta_k)$ means infinitesimally small value in comparison with $p_1(\Delta_k)$. Thus, property of ordinariness allows rewrite (3) as follows:

$$\lim_{\Delta_k \rightarrow 0} E(\Delta_k) = p_1(\Delta_k) \quad (5)$$

or, taking into account (2) and (4):

$$p_1(\Delta_i) = \lambda \Delta_i \quad (6)$$

and

$$p_0(\Delta_i) = 1 - \lambda\Delta_i . \quad (7)$$

Let us find now the probability of no failures within entire interval of length t . Due to the Markov property:

$$P(t) = \prod_{\forall i} p_0(\Delta_i) . \quad (8)$$

In limit, under condition of uniform tend all Δ_i to 0, and remembering that $\sum\Delta_i = t$, we get:

$$P(t) = \lim_{\Delta \rightarrow 0} (1 - \lambda\Delta)^{t/\Delta} = \exp(-\lambda t) . \quad (9)$$

This has completed the proof that the three characterization properties of the Poisson process leads to recurrent point process with intervals between events distributing exponentially.

6.3 Procedures over Poisson processes

Before begin with explanations of a suggested heuristic method, let us consider two simple procedures: *thinning* of the Poisson process and *superposition* of the Poisson processes.

Thinning procedure.

Procedure of thinning consists in the following. Let us exclude from an arbitrary Poisson process points with constant probability q and keep them with probability $p = 1 - q$. What kind of process we will get in this case?

Get the result without any special mathematical proofs basing only on characterization properties of the Poisson process.

Markov property of a new point process conserves because probability of exclusion of any point does not depend on entire prehistory of the process.

Ordinariness property conserves due to the fact that points are only excluding from the process.

Stationarity of the process conserves due to constant value of probability q on all time axis.

Thus, since all characterization properties have been preserved under such thinning procedure, the resulting process is the Poisson one.

Of course, not any “thinning procedure” with Poisson process leads to such results. Assume that probability of point exclusion, for instance, is decreasing in time. In this case, the flow of failures will remind a traditional aging process of a

system consisting of wearing-out units: before “complete death” a system will fail more and more often, i.e. no properties of the Poisson process are observed.

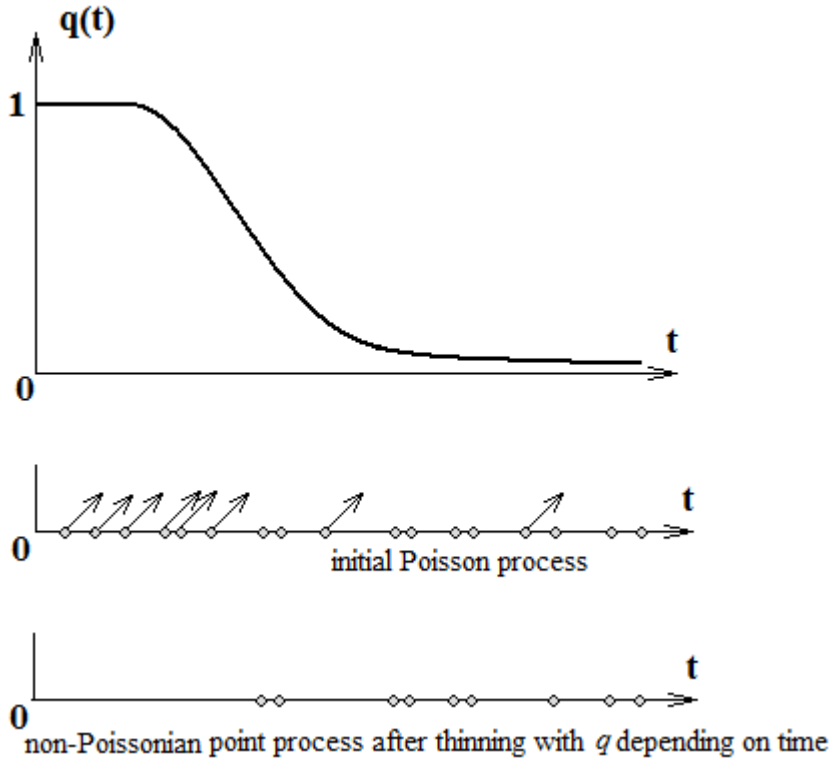


Figure 3. Example of thinning with probability q depending on time.

Another example of thinning that does not conserve Markov property is thinning with probabilities of exclusion points from the process depending on the ordinal number of point in initial process. For instance, let $q_k = 1$ for all $k \neq 2^{3x}$ and $k \neq 2^{3x+1}$ where x is a natural number, and $q_k = 0$, otherwise. In other words, we have a new point process: after interval presenting sum of three i.i.d. exponential random variables with probability 1 follows a single exponentially distributed time interval, In other words, the process becomes alternative process formed an alternative sequence of intervals with Erlang and exponential (see Figure 4).

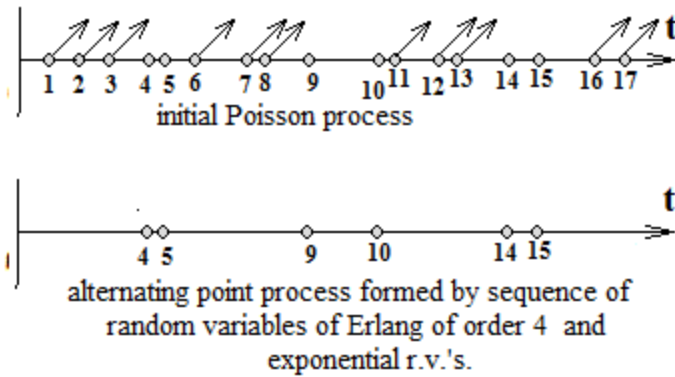


Figure 4. Example of thinning with probability q depending ordinal number of point in initial process.

Superposition procedure.

The second important procedure is the so-called superposition of independent Poisson processes.

Markov property of a new point process conserves due to independence of initial processes and the Markov property of each ingredient process.

Ordinariness property conserves due to the fact that exponential distribution is continuous, hence exact coincidence of two independent random variables is impossible. By the same reason, appearance of points of concentrations also becomes an impossible event.

Stationarity of the resulting process conserves due to initial stationarity of each incoming ingredient.

Thus, since all characterization properties of Poisson process have been preserved under such procedure of superposition, the resulting process is also Poisson one with parameter equals to sum of parameters of inputting sub processes.

These simple explanations are given only to give the reader more exact filling of this two procedures those are so important in Theory of stochastic point processes.

The approximate methods suggested below are based on two important limit theorems in the theory of point stochastic processes.

6.4 Asymptotic thinning procedure over stochastic point process

We begin with a renewal process. By definition, a counting process, for which the interarrival times are i.i.d. with an arbitrary so-called "forming distribution" $F(t)$ is said to be a renewal process. In this sense, the Poisson process is a particular case of a renewal one.

Let us apply the thinning procedure to this process multiply, excluding points step-by-step as it shown in Figure 5.

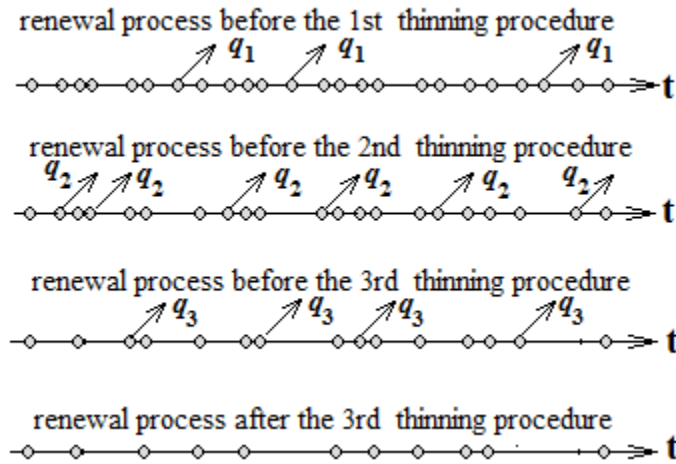


Figure 5. Three first steps of thinning a renewal process.

The strong definition of such asymptotic thinning procedure is known as the Limit Renyi¹⁵ Theorem:

If

$$Q_n = q_1 q_2 \dots q_n \rightarrow 0 \text{ as } n \rightarrow \infty,$$

then in limit the thinned renewal process approaches Poisson one. Of course, such procedure leads to infinite growth of the interarrival times. To keep the resulting process of the same intensity (i.e. with the same average length of intervals as in the initial renewal process), one needs simultaneously to “compress” the time axis: new time scale after n thinning procedures has to be $Q_n t$.

Some time later Yu. Belyaev¹⁶ generalized the result on stochastic point processes beyond renewal ones.

¹⁵ **Alfred Rényi** (1921 – 1970) was a Hungarian mathematician who known for his contributions in combinatorics, graph theory, number theory but mostly in probability theory.

¹⁶ **Yuri Konstantinovich Belyaev** (b. 1932) is Russian statistician, Professor of Moscow State University and Professor Emeritus of University of Umeå (Sweden). Pupil of A.N. Kolmogorov.

6.5 Asymptotic superposition of stochastic point processes

Another important asymptotic result is the Khinchin¹⁷-Ososkov¹⁸ Theorem of the superposition (union) of independent renewal processes. Later this result was independently expanded on more general stochastic point processes by B. Grigelionis¹⁹ and I. Pogozhev²⁰ in the theorem named after them,

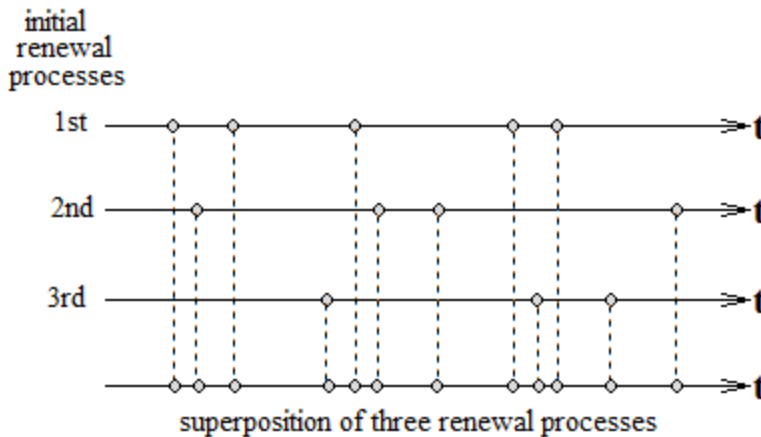


Figure 6. Example of superposition of three renewal processes,

The theorem states that superposition of n independent renewal processes forms in limit the Poisson process, when $n \rightarrow \infty$. Indeed, let us again check if such procedure leads to three properties of the Poisson process.

Each renewal process consisting the resulting one, possesses the property of “restricted aftereffect”, i.e. its future depends only on the moment of the last occurred event. If the number of superposed renewal processes is large, the process, in a sense, “forgets” its past: too many other independent events “intervene” between two neighbor arrivals of the same process. Thus, it is understandable, on intuitive level, that process asymptotically begins to possess Markov property.

Ordinariness property conserves due to the fact that “forming distributions” $F_k(t)$ of initial renewal processes are continuous, and due to it exact coincidence of two arrivals or appearance of points of concentrations are impossible events.

Stationarity of the resulting process is delivered by the fact that each of initial renewal processes has its own constant intensity.

¹⁷ **Alexander Yakovlevich Khinchin** (1894 –1959) was a Soviet mathematician and one of the most significant people in the Soviet school of probability theory.

¹⁸ **Gennady Alexeevich Ososkov** (b. 1931) is Soviet and Russian mathematician, pupil of A.N. Kolmogorov and A.Ya. Khinchin.

¹⁹ **Bronyus Igno Grigelionis** (b. 1935) is Lithuanian mathematician. Pupil by B.V. Gnedenko.

²⁰ **Ivan Borisovich Pogozhev** (1923-2011) was Soviet and Russian mathematician.

Thus, since all characterization properties of Poisson process are presented, the resulting point process asymptotically approaches Poisson one with parameter equals to sum of parameters of inputting sub processes.

Of course, in this case interarrival intervals go to 0, so we again should change time scaling: we have to “stretch out” the time axes to keep length of interarrival intervals in a reasonable scale.

The main requirement for correct using the Theorem of superposition is the condition of “uniformity” of point processes that compose the resulting process. For instance, assume that we superpose a single regular point process with constant interarrival time, τ , and infinite number of Poisson processes of such type that their intensity, λ_k , decrease in such a way that

$$\lim_{n \rightarrow \infty} \sum_{1 \leq k \leq n} \lambda_k = \Lambda, \text{ where } \Lambda \ll \frac{1}{\tau} \text{ (see Figure 7).}$$

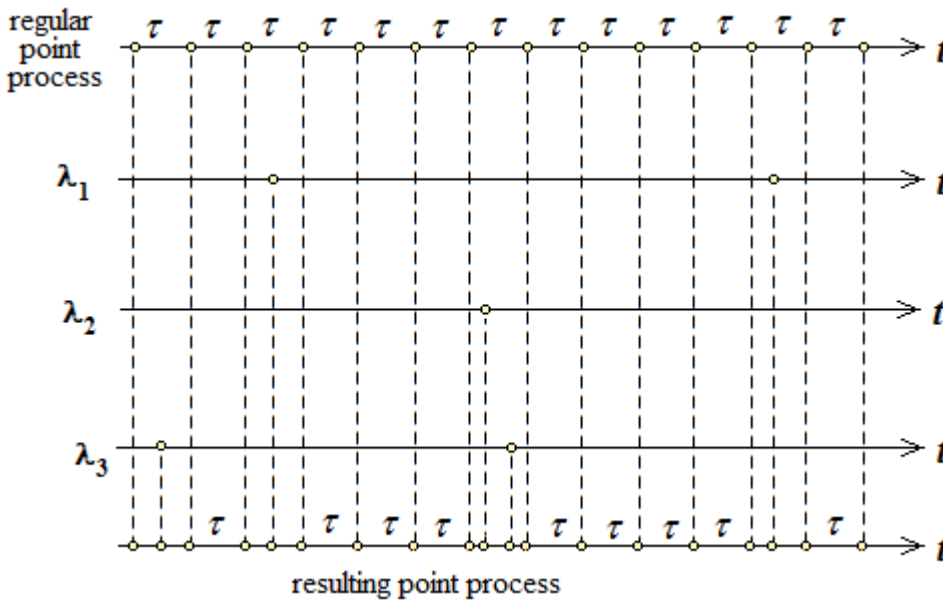


Figure 7. Superposition “dense” regular point process with a number of “weak” Poisson processes.

From this figure, one can see that points of regular process prevail over points of other processes.

* * *

Now we will show that both of these theorems are extremely constructive and effective for a heuristic analysis of highly reliable repairable systems. Of course, rigid asymptotic results used in a prelimit case gives only approximations, however what does it mean “correct model”? A model is always a model. We would like to underline once more that model constructing is “an art”, and preciseness of

probabilistic analysis of real objects and phenomena depends, in first turn, on their understanding, rather than on filigree mathematical exercises.

Everybody understands that the main imperfection of any heuristic method usually lies in the impossibility of defining the domain where the results obtained with its help are valid. But in this particular case, we can suggest a simple (and convenient!) rule: if you have obtained a high value of a reliability index, the application of the heuristic method was correct. It is not a bad rule because otherwise a system is improper for practical use!

The best way to explain a heuristic method in detail is to show examples of how it works.

6.6 Intersection of flows of narrow impulses

The main idea is in the following. Consider two alternating renewal processes (flows of impulses), one with interarrival time T_1 and impulse's width τ_1 , and another with corresponding parameters T_2 and τ_2 . In assumption of exponentiality of interarrival time, one can write $\lambda_1 = \frac{1}{T_1}$ and $\lambda_2 = \frac{1}{T_2}$.

Intersection of two impulses can happen if the end edge of an impulse of the first flow overlaps with the front edge of an impulse of the second flow, or if the end edge of an impulse of the first flow overlaps with the end edge of an impulse of the second flow. Notice that probability of this event is equivalent to the probability that the front edge of an impulse of the second flow appears within the impulse with the width equal to $\tau_1 + \tau_2$ (see Figure 8).

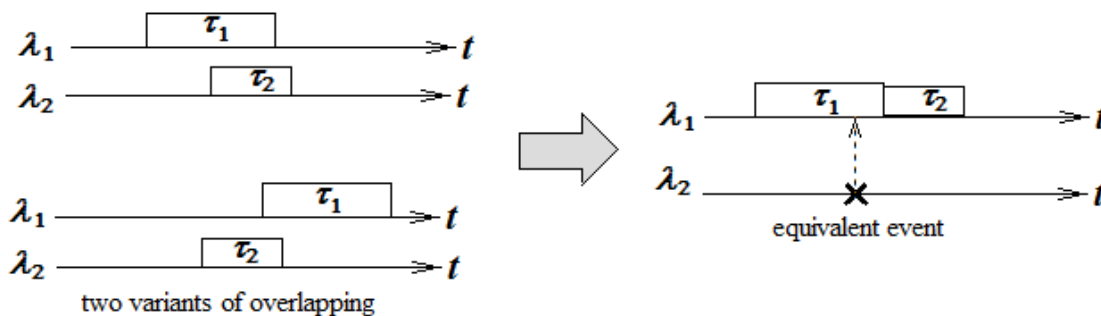


Figure 8. Demonstration of equivalency of two models of impulses overlapping.

In result, the probability of impulses of the first flow overlap with impulses of the second flow can be calculated approximately as:

$$\pi_1 \approx 1 - \exp(-\lambda_2(\tau_1 + \tau_2)) \approx \lambda_2(\tau_1 + \tau_2). \quad (10)$$

Now let us apply these results and mentioned above asymptotic theorems to reliability analysis of simplest series and parallel systems.

First, consider a series system of two units. What happens if two intervals of down time are overlapped? Consider the case of restricted recovery (the unit failed during recovery of previously failed one is waiting in line). The total recovery time equals $\tau_\Sigma = \tau_1 + \tau_2$.

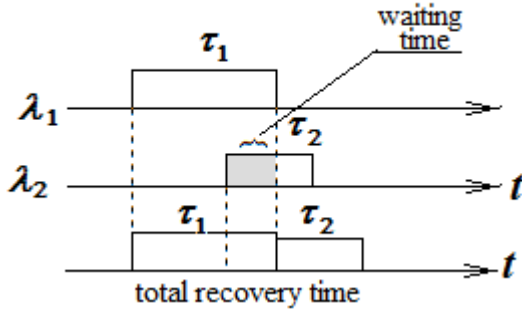


Figure 8. Explanation of forming the total recovery time in case of restricted recovery.

In result, there are three ingredients of the resulting flow:

- 1) Flow of impulses of mean width equals $\tau_\Sigma = \tau_1 + \tau_2$. Intensity of this flow defined in $\Lambda = \lambda_1 \pi_1 \approx \lambda_1 \lambda_2 (\tau_1 + \tau_2)$;
- 2) Flow of impulses of the first flow (width of impulse equals τ_1 and intensity of this flow is $\lambda_1^* \approx \lambda_1 - \Lambda$);
- 3) Flow of impulses of the second flow (width of impulse equals τ_2 and intensity of this flow is $\lambda_2^* \approx \lambda_2 - \Lambda$).

Being practical, let us evaluate possible values of the found parameters.

If a system is highly reliable, for instance, its availability coefficient is at least 0.95 and the number of units is about 50, then availability coefficient of each unit has to be of order $K = \sqrt[50]{0.95} \approx 0.999$. If assume that mean down time is about 0.5 hour, it leads to value of failure rate about 0.002 [1/hrs]. In such assumptions value of $\Lambda \approx 0.002^2 [1/hrs] = 0.000004 [1/hrs]$, i.e. in practice no sense in “catching the fleas”.

Thus, considering series systems with such minor corrections has no sense.

Now consider a dubbed system keeping the same notations for units' parameters. In this case, the system failure occurs if and only if two impulses of down time overlapped. It means that the system failure rate is approximately Λ . In this case, the total recovery time is equal to the duration of time interval when both units are failed. This event may occur by two ways, as it depicted in Figure If

recovering time would be exponentially distributed, this total system recovery intensity is equal to

$$\mu = \frac{1}{\tau_1} + \frac{1}{\tau_2} = \frac{\tau_1 + \tau_2}{\tau_1 \tau_2}, \quad (11)$$

or

$$\tau_{\Sigma} = \frac{\tau_1 \tau_2}{\tau_1 + \tau_2}. \quad (12)$$

In [Gnedenko, Ushakov, 1994] it is shown that (11) is valid, at least approximately, for wide class of distributions.

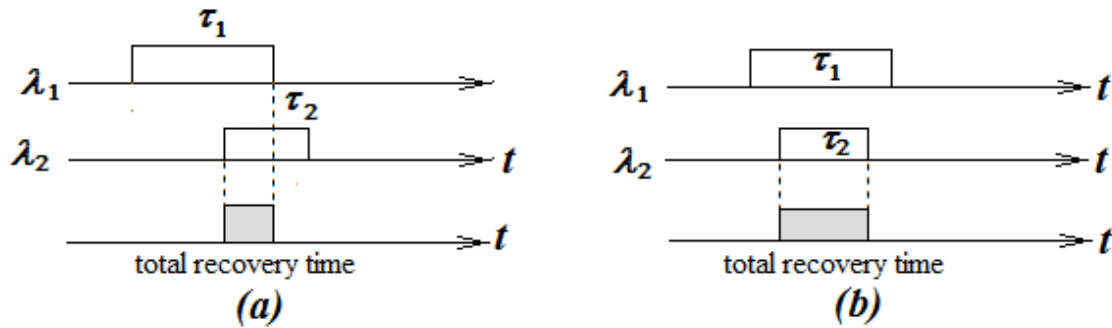


Figure 9. Two ways of forming the system recovery time: partial and complete overlapping.

6.7 Heuristic method for reliability analysis of series recoverable systems

Let a series system consist of n units. The i th unit has a MTTF T_i and a mean repair time τ_i . In accordance with theorem about asymptotic behavior of point processes superposition, failure rate of such series system is:

$$\lambda_{\text{Syst.}} = \left(\sum_{1 \leq i \leq n} \frac{1}{T_i} \right). \quad (13)$$

The system's mean down time can be calculated as a weighted average:

$$\tau_{\text{Syst.}} = \frac{1}{\lambda_{\text{Syst.}}} \sum_{1 \leq i \leq n} \lambda_i \tau_i \quad (14)$$

The system's MDT has hyperexponential distribution (see Appendix 2.2.6).

Probability of failure-free operation

$$P(t_0) \approx \exp\left(-t_0 \sum_{i=1}^n \frac{1}{T_i}\right). \quad (15)$$

Availability coefficient

$$K = \frac{1}{1 + \lambda_{\text{Syst.}} \tau_{\text{Syst.}}} \approx 1 - \sum_{i=1}^n \lambda_i \tau_i \quad (16)$$

Operational availability coefficient

$$R(t_0) \approx \left(1 - \sum_{i=1}^n \lambda_i \tau_i\right) \cdot \left(1 - t_0 \sum_{i=1}^n \lambda_i\right) \approx \left(1 - \sum_{i=1}^n \lambda_i (\tau_i + t_0)\right) \quad (17)$$

6.8 Heuristic method for reliability analysis of parallel recoverable systems

Presenting use of the heuristic method to recoverable parallel systems, we restricted ourselves by dubbed systems.

Operating process of dubbed recoverable systems can be illustrated by the following time diagram (see Figure 10).

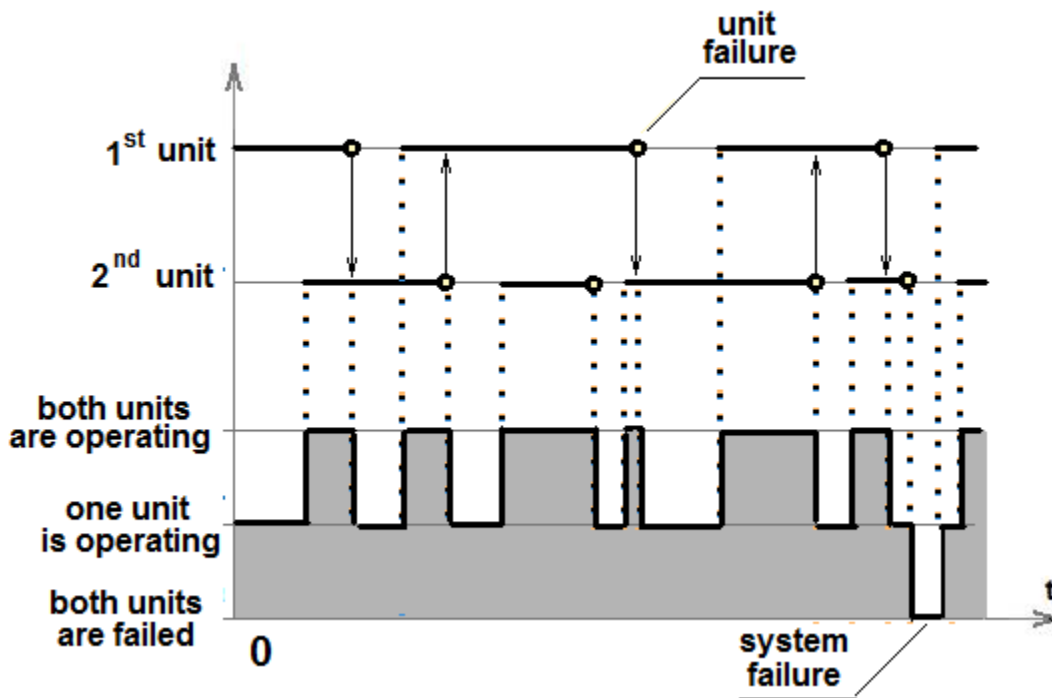


Figure 10. Operating process of a dubbed recoverable system.

Let us begin with a stationary availability coefficient. For a single recoverable unit this coefficient has the form $\frac{1}{1 + \frac{\tau}{T}}$ where τ is MDT and T is MTBF. If time between failures has exponential d.f., one can write the approximation $1 - \lambda\tau$.

Thus, for dubbed system availability coefficient approximately equals:

$$K \approx 1 - [1 - (1 - \lambda\tau)]^2 = 1 - (\lambda\tau)^2 . \quad (18)$$

Now we consider some important special cases.

6.8.1 Influence of unreliable switching procedure

In reality, dubbed system can fail during the process of switching from a failed operating unit to a redundant one,

If π is the probability of successful switching, then the system will fail, on average, after $\frac{1}{1-\pi}$ switching. In other words, dubbed system failure rate due to this particular cause is equal to $\lambda(1-\pi)$. Assume that MDT in case of switching process failure is τ_s . Another ingredient of system's failure flow is simultaneous failure of both units simultaneously. The failure rate due to this case is $\lambda^2\tau$ and MDT is 0.5τ . These simple considerations lead to the following RBD (see Figure 11).

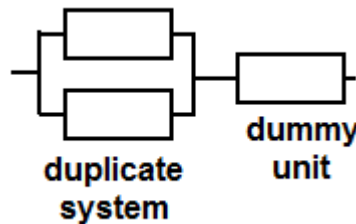


Figure 11. Conditional RBD of a dubbed system with unreliable switching procedure.

In this case, the resulting availability coefficient of dubbed system can be written as:

$$K \approx [1 - (\lambda\tau)^2] \cdot \frac{1}{1 + \lambda(1 - \pi)\tau_s} \approx 1 - (\lambda\tau)^2 - \lambda\tau_s(1 - \pi) \quad (21)$$

Dependence of availability coefficient on the level of switching reliability is presented in the table below.

Table 1. Dependence of availability coefficient on reliability of switching procedure.

		Unit's coefficient of availability		
		0.9	0.95	0.99
π	0.9	0.979	0.992	0.9989
	0.95	0.985	0.995	0.9994
	0.99	0.989	0.997	0.9998
	1	0.99	0.998	0.9999

6.8.2 Influence of switch's unreliability

Sometimes a switch plays a role of a distinctive interface. It means that the switch failure leads to the system failure. The RBD for this case is presented in Figure 12.

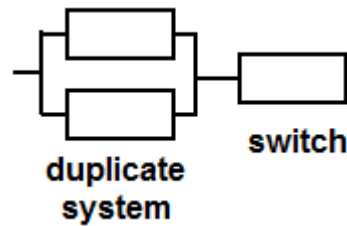


Figure 12. RBD for a dubbed system and a switch-interface.

In this case the system availability coefficient can be written as:

$$K \approx [1 - (\lambda\tau)^2] \cdot K_s \quad (22)$$

where K_s is availability coefficient of the switch-interface.

In the table below, one can find some numerical illustrations.

Table 2. Dependence of availability coefficient on availability of switch-interface.

		Unit's availability coefficient		
		0.9	0.95	0.99
K_s	0.9	0.891	0.95	0.99
	0.95	0.94	0.898	0.8999
	0.99	0.98	0.948	0.9499

	1	0.99	0.9975	0.9999
--	----------	------	--------	--------

6.8.3 Periodical monitoring of operating unit

Sometimes for checking if a operating unit of dubbed system is operational or not, one has to perform special periodical testing. In this situation, a system fails after operating unit failure until it will have been detected. Thus, the system is found in the state of undetected failure during time interval with average length 0.5θ , where θ is period of testing (see Figure 13).

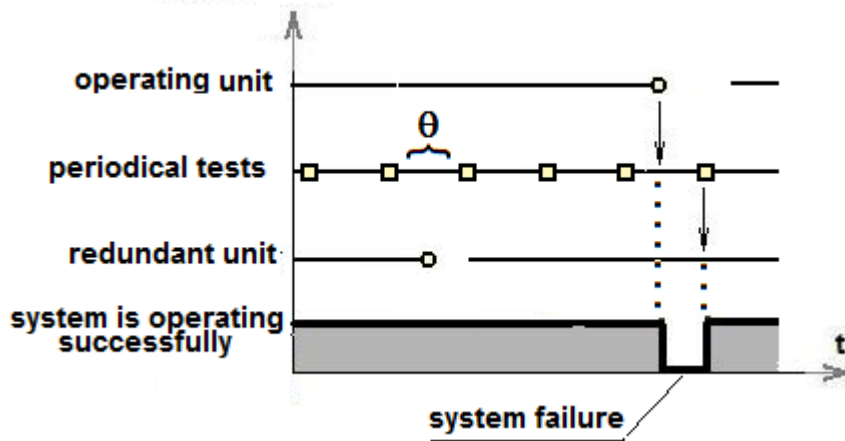


Figure 13. Time diagram illustrating system undetected failure forming.

In this case system's availability coefficient can be written as:

$$K \approx [1 - (\lambda\tau)^2] \cdot (1 - 0.5\lambda\theta) \approx 1 - (\lambda\tau)^2 - 0.5\lambda\theta \quad (23)$$

Formula (23) shows that the testing period has to be significantly less than a single unit MTBF, otherwise there will be null effect of redundancy. Numerical examples are given in the table below.

Table 3. Dependence of availability coefficient on frequency of operating unit testing.

		Unit's availability coefficient		
		0.9	0.95	0.99
θ	0.1	0.94	0.948	0.95
	0.02	0.98	0.988	0.99
\bar{T}	0.01	0.98	0.993	0.995
	0	0.99	0.9975	0.9999

6.8.4 Partial monitoring of operating unit

Efficiency of dubbed system depends on completeness of monitoring of an operating unit, which is currently performing needed operation. Usually testing of a redundant unit is made easier than a unit on operational position because testing itself can interfere with operating functions of the system. Consider a dubbed system with partially monitored operating unit. Time diagram of functioning process of such system is presented in Figure 13.

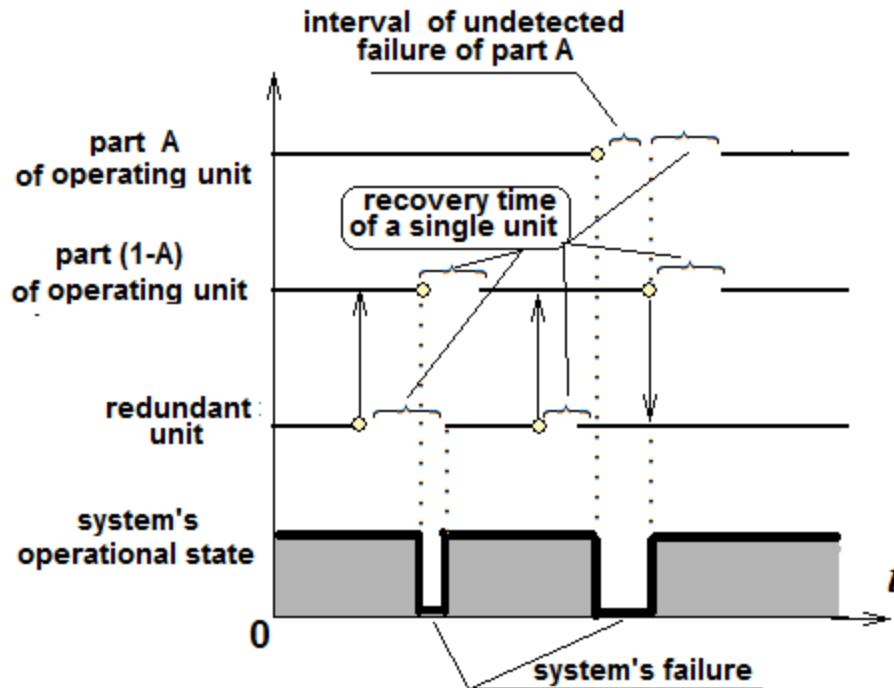


Figure 13. Time diagram of functioning process for dubbed system with partially monitored operating unit.

The RBD of such system can be presented in a conditional form as follows:

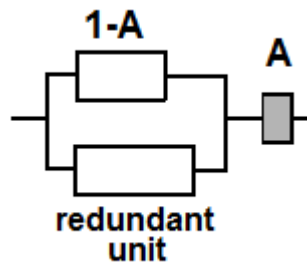


Figure 14. Conditional RBD for a dubbed system with partially monitored operating unit.

For such system, availability coefficient can be written easily as:

$$K_{System} = [1 - (1 - K) \cdot (1 - K_{1-A})] \cdot K_A \approx 1 - \lambda \lambda_{1-A} \tau^2 - \lambda_A \tau \quad (24)$$

From (24), one can see that even if an uncontrollable part is small enough, the entire effect of redundancy can be practically nullified, moreover, it can even worsen the system total reliability (in the table below corresponding values are shadowed).

Table 3. Dependence of recoverable dubbed system reliability on the portion of uncontrolled part of the operating unit.

		portion of uncontrolled part			
		0	0.01	0.05	0.1
unit's availability coefficient	0.8	0.96	0.95	0.91	0.86
	0.9	0.99	0.98	0.94	0.89
	0.95	0.9975	0.988	0.948	0.9
	0.99	0.9999	0.9899	0.9499	0.8999

6.9 Brief historical overview and related sourcecws

Here we offer only papers and highly related books to the subject of this chapter. List of general monographs and textbooks, which can include this topic, is given in main bibliography at the end of the book.

One of the first who suggested applying process of random impulses for reliability analysis of recoverable systems was N. Sedyakin²¹. His ideas were not based on mentioned above asymptotic theorems, therefore were not accepted at his time by mathematical community.

Theory of stochastic point processes and main asymptotic theorems, used for developing the heuristic method, are presented in the following publications. Bibliography is given in chronological-alphabetical ordering for better exposition of historical background of the subject.

Bibliography:

- Cox, D., W.Smith (1954) On the superposition of renewal processes. Biometrika, vol.41, No.(1-2).
- Ososkov, G.A. (1956). A limit theorem for flows of similar events. Theory Probab. Appl., V. 1, No. 2
- Renyi, A. (1956). Poisson-folyamat egy jemllemzese (Hungarian). Ann. Math. Statist, V. 1, No. 4.
- Smith, W.L. (1958) Renewal theory and its ramifications, in Journal of the Royal Statistical Society, series B, 20:243-302.
- Cox, D.R. (1962). Renewal Theory. John Wiley, New York.
- Grigelionis, B (1963). On convergence of sums of steps stochastic processes to a Poisson Process. Theory Probab. Appl. V.8, No.2
- Gnedenko, B.V. (1964). On duplication with renewal. Engrg Cybernet. No. 5
- Gnedenko, B.V. (1964). On spare duplication. Engrg Cybernet. No. 4
- Sedyakin, N. M..(1965) Elements of the Theory of Random Impulse Flows (in Russian). Moscow, Sovetskoe Radio.

²¹ Nikolai Mikhailovich Sedyakin (1922-1969) was Soviet applied mathematician.

- Cox, D.R, and V. Isham (1980). Point Processes. Chapman and Hall.

7 Time redundancy

Some systems possess ability to neglect failures even without redundant units due to some kind of “insensitivity” to short failures or possibility to restart a needed operation. In this case, one says about the so-called *time redundancy*.

Let us consider several main types of systems with time redundancy.

7.1 System with possibility of restarting operation

Assume that the needed operation continues time t_0 , though it can be completed within some interval $[0, t]$ where $t > t_0$. So, the problem is to find the probability that there is at least one interval between failures that is larger than t_0 .

There are two cases:

1. Failure durations are negligibly short, though after each failure has to begin its operation from the beginning;
2. Failure durations are not negligible, so they decrease remaining potentially useful time.

First case. Each failure destroys current results of system operation, and every time the system is forced to begin its operation from the beginning.

Let $R(t | t_0)$ denote the probability that during interval $[0, t]$ there will be at least one period between failures exceeding required value t_0 . The system performs its operation successfully during time t if two events occur:

- there is no failures in time interval $[0, t_0]$;
- a failure has occurred at $x < t_0$, however during remaining period $[x, t_0]$ time at least once the system successfully perform its operation.

The latter event, evidently, cannot occur if $t - x < t_0$. This verbal explanation leads us to the recurrent expression

$$R(t | t_0) = P(t_0) + \int_0^{t_0} R(t - x | t_0) dF(x), \quad (1)$$

where $F(x)$, as usual, is d.f of the system TTF.

This is an equation relating to Volterra²² type equations. Equations of such a recurrent type are usually solved numerically. We will not provide here a mathematical technique for this solution.

For exponentially distributed TTF, one can write a simple approximation based on the following arguments. Successful system operations can occur at the first attempt. If a failure happens in the first interval, then the system begins the second attempt, and so on. Under condition of highly reliable system $\lambda t_0 \ll 1$, the probability of appearance of more than one failure within interval of duration t_0 is negligibly small. At the same time, the conditional distribution of a single event of Poissonian process has a uniform distribution within a fixed interval.

Let us explain the latter statement. Consider a Poisson process. If $\lambda t_0 \ll 1$, then the conditional probability that there is a single failure within the considered interval under condition that there is at least one failure is very close to one. Conditional density of failure location within interval t is:

$$f(t | \text{there is exactly 1 failure}) = \frac{\lambda e^{-\lambda t}}{\lambda t e^{-\lambda t}} = \frac{1}{t}$$

Thus, for interval t_0 on average a failure occurs at moment $\frac{t_0}{2}$, and in this case there remain $t - \frac{t_0}{2}$ units of time for restarting a new attempt. In other words, during time t the system has a possibility to restart operation on average $\frac{t - t_0}{0.5t_0}$ times, i.e. it is equivalent to corresponding number of loaded redundant units (taking into account that the number of redundant units is integer.). Denote the integer part of $\frac{t - t_0}{0.5t_0}$ by Ω . Then we can write the following bounds:

$$1 - (q(t_0))^\Omega \leq P(t | t_0) \leq 1 - (q(t_0))^{\Omega+1}. \quad (2)$$

If time resource is small, the effect of such time redundancy is negligible. Using formula (1) and the same heuristic arguments, we can write for a case when $t < 2t_0$:

$$P(t | t_0) = p(t_0) + q(t - t_0) \cdot p(t_0) = p(t_0) \cdot [1 + q(t - t_0)]. \quad (3)$$

²² **Vito Volterra** (1860 – 1940) was an Italian mathematician and physicist, known for his contributions to mathematical biology and integral equations.

Below there are some numerical examples where the total operational time is larger in comparison with t_0 .

Table 1. Increase of PFFO depending on increase of the total operational time.

		Total time		
		110%	120%	130%
$p(t_0)$	0.99	0.991	0.992	0.993
	0.98	0.982	0.984	0.986
	0.97	0.973	0.976	0.979
	0.96	0.964	0.968	0.972
	0.95	0.955	0.959	0.964

Sesond case: If failures are non-instant, one has to take into account lengths of idle periods between failures. Let $G(t)$ denote a distribution of idle time during recovery. This case is very close to the previous one with the difference that restarting of the system occurs after time of recovery, not immediately after a failure.

This verbal description permits us to write the following recurrent expression:

$$R(t | t_0) = P(t_0) + \int_0^{t_0} \left[\int_0^{t-x} R(t-x-y | t_0) dG(y) \right] dF(x), \quad (4)$$

where again $R(t | t_0) = 0$, if $t < t_0$.

For this case, we cannot suggest any “pleasant results”; solution can be obtained only by numerical methods.

7.2 Systems with “admissibly short failures”.

Consider a system that has some kind of “functional inertia”: if recovery time, η , is less than ε , the system does not “feel” it and continue successfully perform its functions.

It is clear that for highly reliable systems when, $E\{\xi\} \gg E\{\eta\}$, one can apply the asymptotic theorem of point processes thinning. If $G(x)$ is d.f. of recovery time, then $G(\varepsilon)$ is the probability that a system failure has been excluded during thinning operation. In other words, such “short” failure has no influence on the system operation. In this case, for general distribution of TTF, ξ , we will have a new random variable, ξ^* , that has distribution, $F^*(t)$, which is defined as:

$$F^*(t) = \begin{cases} F(t) & \text{with probability } G(\varepsilon) \\ F^{*2}(t) = \int_0^t F(t-x)dF(x) & \text{with probability } G(\varepsilon) \cdot [1-G(\varepsilon)] \\ F^{*3}(t) = \int_0^t F^{*2}(t-x)dF(x) & \text{with probability } G(\varepsilon) \cdot [1-G(\varepsilon)]^2 \\ F^{*4}(t) = \int_0^t F^{*3}(t-x)dF(x) & \text{with probability } G(\varepsilon) \cdot [1-G(\varepsilon)]^3 \\ & \text{and so on} \end{cases} \quad (5)$$

where $F^{*n}(t) = \int_0^t F^{*(n-1)}(t-x)dF(x)$ is a convolution of order n , i.e. distribution of sum of n random variables. Thus, r.v. ξ^* is the sum of random number of random variables ξ and this number has geometric distribution. It is known from the probability theory that asymptotically such sum has an exponential distribution. So, for large n one can approximateli write:

$$P(t) = \exp\left(-\frac{t}{T} \cdot G(\varepsilon)\right). \quad (6)$$

7.3 Systems with a Time Accumulation

Some systems accumulate time of successful operation during a total period of performance, θ . The system operation is considered completed if during period θ the totalaccumulated operational time exceeds t_0 . In this case we consider an alternating process of operating and idle periods.

Denote the probability that the total accumulate operational time is larger than t_0 units during period θ as $P(t_0|\theta)$. For this probability one consider two events lead to success:

- a system works without failures during time t_0 from the beginning;
- a system has failed at moment $x < t_0$, was repaired during time y and at the remaining interval of $\theta - x - y$ tries to accumulate $t_0 - x$ units of time of a successful operation. This description leads us to the recurrent expression:

$$P(t_0|\theta) = P(t_0) + \int_0^{t_0} \left[\int_0^{\theta-x} P(t_0 - x / \theta - x - y) dG(y) \right] dF(x). \quad (7)$$

This expression is correct for the case where a system starts to perform at $t=0$. In general case such equation can be solved only numerically.

This subject as a whole requires much more room and details. There are many interesting detailed models concerning, for instance, computer systems.

7.4 Case study: Gas-pipeline with an underground storage

Consider a simplest gas-pipeline system with an underground storage, which allows a user to get gas supply during a pipeline break out.

Let η be the pipeline's random idle time with distribution $G(t)=\Pr\{\eta \leq t\}$ and ξ be its TTF with a distribution $P(t)=\Pr\{\xi \leq t\}$. The storage volume equals V . The speed of the storage expenditure (after a pipeline failure) equals α , and its speed of refilling equals β . A process of expenditure and refilling the storage is depicted in Figure 1. For simplicity, we assume that the storage begins refilling immediately after the pipeline's repair.

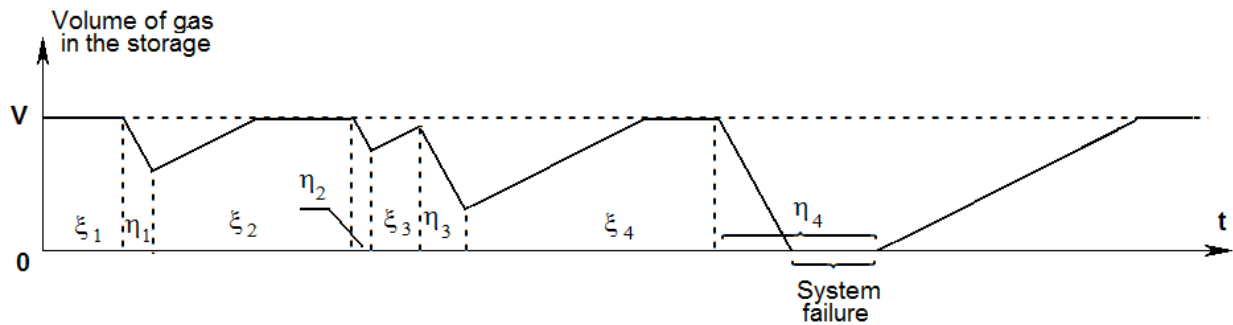


Figure 1. Example of a process of expenditure and refilling storage.

The system's failure occurs when a user does not obtain gas (the storage becomes empty). It is clear that due supplying from the storage, a user does not "feel" short failure times of the pipeline.

Assume that pipeline failures occur "not too often" and the probability of the storage's exhaustion during a pipeline's repair is "small enough". (The meaning of the expressions in inverse commas will be explained below.) Also assume that pipeline MTBF is much larger than the average duration of the pipeline's repair.

Under these assumptions, one can consider the process of the pipeline's disruption occurrences as a renewal stochastic process. An appearance of the system failures can be considered as a "thinning" procedure because a pipe-line's disruption rarely leads to the system's failure. In other words, the process of system failures might be approximated by a Poisson Process. Note that for an acceptance of this hypothesis, the probability of developing a pipeline's failure in a system's failure should be small enough (practically less than 0.05).

Let us denote $q = 1 - G\left(\frac{v}{\alpha}\right)$, then the system MTTF equals

$$T_{syst} \approx \frac{T + \tau}{1 - G\left(\frac{V}{\alpha}\right)}. \quad (8)$$

For the PFFO, one can write

$$P_{syst} \approx \exp\left(-\frac{t}{T_{syst}}\right). \quad (9)$$

This expression gives an upper bound because actually we assume that the storage is refilled instantaneously. The smaller the probability q , the better the bound.

One can easily obtain a lower bound. Assume that any pipeline failure, which appears during the refilling, leads to the system's failure. This bound is lower because, as a matter of fact, not each failure during refilling leads to the system's failure. The probability of a system's failure under this assumption equals

$$q^* = 1 - G\left(\frac{V}{\alpha}\right) + \int_0^x R\left(\frac{\alpha x}{\rho}\right) dG(x). \quad (10)$$

Obviously, the probability q^* will be larger if we write

$$q^* = 1 - G\left(\frac{V}{\alpha}\right) + G\left(\frac{V}{\alpha}\right) \cdot R\left(\frac{V}{\rho}\right). \quad (11)$$

i.e., consider that each pipeline failure has always a maximal duration equal to $\frac{V}{\rho}$.

Expression (11) can be rewritten as

$$q^* = 1 - G\left(\frac{V}{\alpha}\right) \left[1 - R\left(\frac{V}{\rho}\right)\right]. \quad (12)$$

Now we can write a lower bound for the system's MTTF

$$T_{syst}^* = \frac{T + \tau}{q^*} \quad (13)$$

and for the system's PFFO

$$P_{\text{sys}}^* \approx e^{-\frac{t}{T_{\text{sys}}^*}} \quad (14)$$

Note that in practical cases values of the MTTF and MTBF coincide. For more details, see [Rudenko and Ushakov, 1989].

7.5 Brief historical overview and related sources

Here we offer only papers and highly related books to the subject of this chapter. List of general monographs and textbooks, which can include this topic, is given in main bibliography at the end of the book.

The so-called time redundancy was first considered in the books by G. Cherkosov and B. Kredentser. Bibliography below is given in chronological-alphabetical ordering for better exposition of historical background of the subject.

Bibliography

Cherkosov, G.N. (1974). Reliability of Technical Systems with Time Redundancy. (Russian). Moscow, Sovetskoe Radio.

Kredentser, B.P. (1978). Prediction of Reliability of Systems with Time Redundancy. (Russian). KiNaukova Dumka.

Rudenko, Yu.N., and I.A. Ushakov (1989). Reliability of Energy Systems. (Russian). Novosibirsk, Nauka.

Obzherin, Yu.E., and A.I. Peschansky (1994) Reliability of unstructured systems with excess time. Cybernetics and Systems Analysis, Vol.30, No.6.

Obzherin, Yu.E., and A.I. Peschanskii. (2004). Reliability analysis of a system with combined time reserve. Cybernetics and Systems Analysis. Volume 40 Issue 5

Obzherin, Yu. E., and A. V. Skatkov (2010). On the time to failure of systems with large replenishable reserve time. Journal of Mathematical Sciences, Vol. 57, No. 5

8 “Aging” units and systems of “aging” units

8.1 Chebyshev bound

In practice very often we know only MTBF of a unit and sometimes, additionally, from some physical hypotheses, that a unit is “aging”, i.e. its failure rate $\lambda(t)$ is increasing in time. Even this scant information permits to get some reasonable boundary estimates of reliability indices.

If distribution of TTF is unknown, one can write the following upper bound based on the Chebyshev²³ inequality:

$$\Pr\{\xi - E\{\xi\} \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}, \quad (1)$$

where ξ is random variable (in our case, unit’s TTF), and $E\{\xi\}$ and σ^2 are its mean and variance; ε is an arbitrary positive constant. Let us demonstrate the proof of the statement. By definition:

$$\Pr\{\xi - E\{\xi\} \geq \varepsilon\} = \int_{|x - E\{\xi\}| \geq \varepsilon} dF(x). \quad (2)$$

Since the domain of integration is $(1/\varepsilon) |x - E\{X\}| \geq 1$, one can write:

$$\int_{|x - E\{X\}| \geq \varepsilon} dF(x) \leq \frac{1}{\varepsilon^2} \int_{|x - E\{X\}| \geq \varepsilon} (x - E\{X\})^2 dF(x) \leq \frac{1}{\varepsilon^2} \int_{-\infty}^{\infty} (x - E\{X\})^2 dF(x) = \frac{\sigma^2}{\varepsilon^2}, \quad (3)$$

That completes the proof.

From (3) one can see that the universal estimate gives rather rough estimates and only for arguments that are lying from the mean on a distance larger than σ . Nevertheless, it is possible to get better estimates if here is some additional information about type of d.f.

8.2 “Aging” unit

In Reliability Theory aging units were introduced in [Barlow-Proschan, 1965]. They called corresponding class of TTF distributions as IFR-distributions where IFR stands for Increasing Failure Rate.

²³ **Pafnuty Lvovich Chebyshev** (1821 – 1894) was a Russian mathematician. He is known for his work in the field of probability, statistics and number theory.

For further deductions, we will use the fact that exponential and degenerate distribution functions represent boundary ones for the entire class of IFR distributions. Indeed, the constant is the boundary function between class of decreasing and increasing functions; and the second one is “the most increasing”: it is a delta function.

Notice that degenerate distribution is the distribution function for non-random variable, i.e. constant. Its distribution function is the Heaviside²⁴ step function defined as:

$$D^*(x,t) = \begin{cases} 0 & \text{if } t < x \\ 1 & \text{otherwise.} \end{cases} \quad (4)$$

where x is a point of discontinuity.

On Figure 1 failure rates for both cases are depicted.

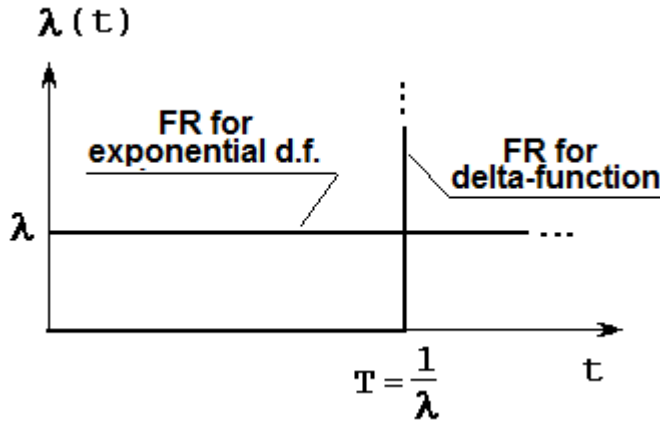


Figure 1. Failure rates for exponential and degenerate distributions with equal MTTF's.

8.3 Bounds for probability of failure-free operations

Lower bound. Let us consider three functions: IFR, exponential and degenerate distributions, all with the same mean equals T . Denote these functions by $P(t)$, $E(t)$ and $D(t)$, respectively.

²⁴ **Oliver Heaviside** (1850–1925) was a self-taught English electrical engineer, mathematician, and physicist who adapted complex numbers to the study of electrical circuits, invented mathematical techniques to the solution of differential equations (equivalent to Laplace transforms), and independently co-formulated vector analysis. Although at odds with the scientific establishment for most of his life, Heaviside changed the face of mathematics and science for years to come.

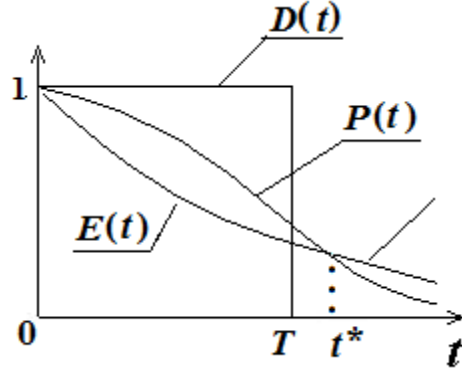


Figure 2. Degenerate , IFR and exponential distributions with equal MTTF's.

It is easy to show that t^* , the crossing point of $P(t)$ and $E(t)$, lays on the right of T . From the condition of equality of the means follows that areas restricted by each of curves $G(t)$, $E(t)$ and $P(t)$ and abscissa are equal. It follows that for $E(t)$ and $P(t)$, the equality stands:

$$\int_0^{t^*} [P(t) - E(t)] dt = \int_{t^*}^{\infty} [E(t) - P(t)] dt \quad (5)$$

Since $G(t) > P(t)$ for all $t < T$, then:

$$\int_0^T [P(t) - E(t)] dt < \int_0^T [G(t) - E(t)] dt = \int_{t^*}^{\infty} [E(t) - P(t)] dt \quad , \quad (6)$$

that confirms that inequality $t^* > T$ is satisfied.

From above follows that for any IFR distribution with the mean T the following lower bound exists:

$$P(t) \geq \begin{cases} \exp(-t/T) & \text{for } t < T \\ 0 & \text{for } t \geq T \end{cases} \quad . \quad (7)$$

Upper bound. If $t \leq T$ the upper bound for PFFO is trivial: $P(t) \leq 1$. The upper bound for $t > T$ needs some auxiliary arguments. Consider a family of exponential functions, $E_k^*(t)$, truncated from the right and such that their means are equal to T . It is obvious that for each truncated exponent $\lambda_k < \lambda = \frac{1}{T}$.. So, the ‘‘aging’’ function $P(t)$ cross any of truncated exponential functions $E_k^*(t)$ twice: first from above on the left of t_k , and then in t_k (see Figure 3). This fact follows from the condition of the means equality that means that the corresponding areas under curves are equal, i.e.

$$\int_{t: P(t) > E_k^*(t)} [P(t) - E_k^*(t)] dt = \int_{t: E_k^*(t) > P(t)} [E_k^*(t) - P(t)] dt. \quad (8)$$

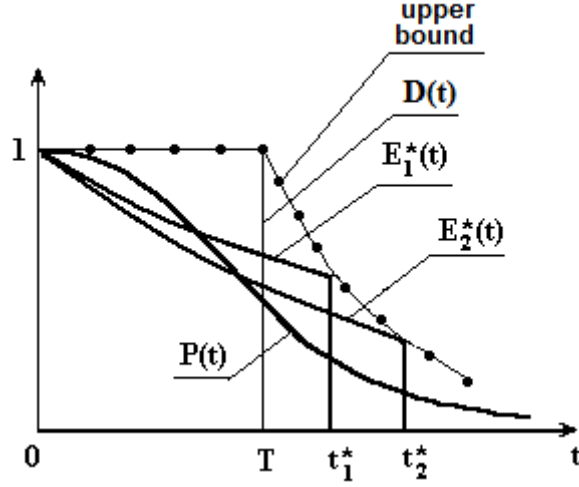


Figure 3. Explanation of the upper bound construction.

For constructing the upper bound, that is set of all $E_k^{\&}(t_k^*)$. In other words, $\int_0^{t_1^*} E_k^{\&}(t) dt = T$ or each exponential function with parameter λ_k , one needs to find such t_k^* that

$$\int_0^{t_1^*} E_k^{\&}(t) dt = \int_0^{t_1^*} \exp(-\lambda_k^{\&} t) dt = T. \quad (9)$$

Integration of (9) gives us

$$1 - \lambda_k^* T = \exp(-\lambda_k^* t_k^*). \quad (10)$$

Thus, we can construct the continuous upper bound for $P(t)$:

$$P(t) \leq \begin{cases} 1 & \text{for } t \leq T, \\ \exp(-\lambda^* t) & \text{for } t > T, \end{cases} \quad (11)$$

where λ^* depends on corresponding t_k and is found from equation of type (10).

Notice that for practical purposes, one is more interested in the lower bound, since it gives a warranty value for PFFO.

8.4 Series system consisting of “aging” units

8.4.1 Preliminary Lemma

For the further analysis, we will need the following lemma.

Lemma. Let for functions $f(x)$ and $g(x)$ the following conditions are satisfied:

- (1) $f(x)$ is a monotone bounded and non-negative function on positive semi-axis,
- (2) $g(x)$ is absolutely integrated function on positive semi-axis,
- (3) $g(x)$ possesses the following property: $g(x) \geq 0$ for $x < a$ and $g(x) \leq 0$ for $x \geq a$, and additionally:
- (4) $g(x)$ satisfies the following condition:

$$\int_0^{\infty} g(x) dx = 0. \quad (12)$$

Under these conditions, if function $f(x)$ increases (decreases) in x , the following inequality takes place:

$$\int_0^{\infty} f(x)g(x)dx \leq (\geq) 0. \quad (13)$$

Proof. We present the proof in the form of a chain of equalities.

$$\begin{aligned} \int_0^{\infty} f(x)g(x)dx &= \int_0^a f(x)g(x)dx + \int_a^{\infty} f(x)g(x)dx \leq (\geq) \\ &\leq (\geq) \int_0^a \left[\max_{0 \leq x \leq a} f(x) \right] g(x)dx + \int_a^{\infty} \left[\min_{0 \leq x \leq a} f(x) \right] g(x)dx = \\ &= f(a) \int_0^a g(x)dx + f(a) \int_a^{\infty} g(x)dx = f(a) \int_0^{\infty} g(x)dx = 0. \end{aligned} \quad (14)$$

The common sense of this lemma is easily understood from Figure 4: it is clear that in the final integral square s_2 is taken with smaller “weight” than square s_1 .

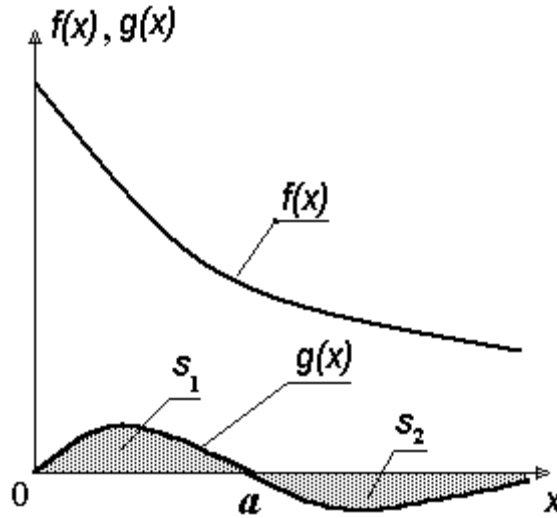


Figure 4. Graphical explanation of lemma for decreasing function $f(x)$.

8.5 Series system

8.5.1 Probability of failure-free operation

Consider a series system, elements of which are numerated in order of increasing their MTTF's: $T_1 \leq T_2 \leq \dots \leq T_n$.

Lower bound. The lower bound for the system, $\underline{P}(t)$, is defined as the product of the lower bounds of its units:

$$\underline{P}(t) = \begin{cases} E(t) = \prod_{i=1}^n E_i(t) \text{ for } t \leq \min_{1 \leq i \leq n} T_i \\ 0 \text{ for } t > \min_{1 \leq i \leq n} T_i \end{cases} \quad (15)$$

where $E_i(t) = \exp(-t/T_i)$, and T_i is the i -th unit MTTF.

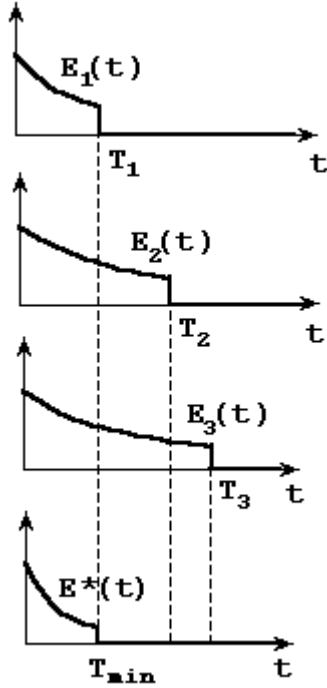


Figure 5. Explanation of the lower bound of PFFO construction on the example of a series system of three units.

Using (7), one can immediately write the lower bound for the series system PFFO in the:

$$\underline{P}(t) \geq \begin{cases} \exp\left(-t \sum_{i=1}^n (T_i)^{-1}\right) & \text{for } t < T_{\min} \\ 0 & \text{for } t \geq T_{\min} \end{cases} \quad (16)$$

Thus, the lower bound for series system PFFO has a non-trivial sense only vor time interval $[0, T_{\min}]$.

Upper bound. Using (2.8), one can write the upper bound, \bar{P} , for a series system consisting of n independent “aging” units. This time we avoid long formal explanations, referring to Figure 6.

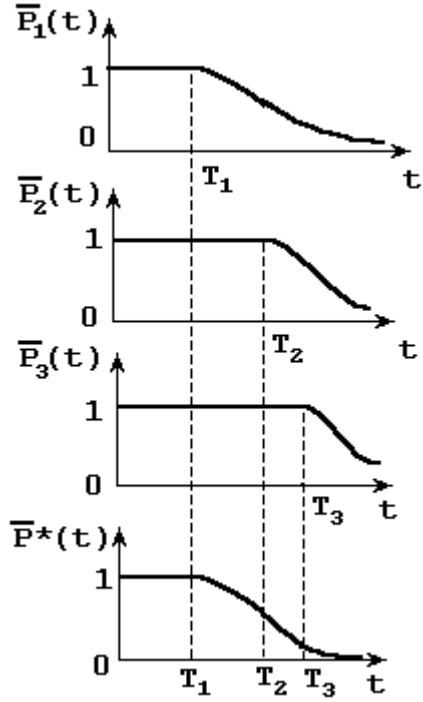


Figure 6. Explanation of the upper bound of PFFO construction on the example of a series system of three units.

The system upper bound for PFFO is defined as:

$$\bar{P}^*(t) = \prod_{i=1}^n P_i^*(t). \quad (17)$$

After substitution (11) in (17), the upper bound expression has the form:

$$\bar{P}(t) = \begin{cases} 1 & \text{for } t \leq T_1, \\ \exp(-\lambda_1 t) & \text{for } T_1 < t \leq T_2, \\ \exp[-(\lambda_1 t + \lambda_2 t)] & \text{for } T_2 < t \leq T_3, \\ \dots \\ \exp\left(-\sum_{i=1}^n \lambda_i t\right) & \text{for } t \geq T_n, \end{cases} \quad (18)$$

where λ_k depends on corresponding moment of truncation.

Unfortunately, this upper bound has almost no practical interests: we are usually interested in PFFO values for $t \ll T_{\min}$.

If each unit's d.f. has a small variation coefficients, then one can assume that

$$P(t_0) \approx p_{min}(t_0), \quad (19)$$

where index *min* corresponds to d.f. with the minimum MTTF, i.e.

$$T_{min} = \min_{1 \leq k \leq n} T_k. \quad (20)$$

To explain this in graphical way, consider a system of two units (Figure 7).

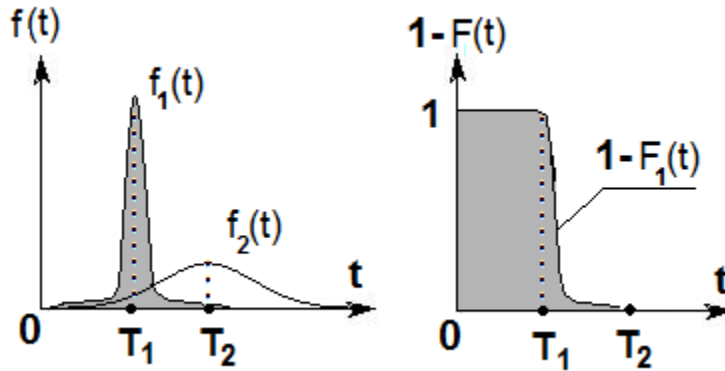


Figure 7. Influence of the weakest unit TTF distribution on PFFO of a series system of two units.

From this figure, one can see that in this case one observes “the rule of the weakest link”: reliability of the system depends practically only on reliability of the less reliable unit.

8.5.2 Mean time to failure of series system

Upper bound. Since

$$\underline{P}(t) \leq \min_{1 \leq i \leq n} P_i(t), \quad (21)$$

one can write:

$$T \leq \min_{1 \leq i \leq n} \int_0^{\infty} P_i(t) dt = \min_{1 \leq i \leq n} T_i. \quad (22)$$

Lower bound. For getting this bound, use results of lemma above. Let us show that the substitution exponential distribution instead of IFR one (if MTTF's of these units are equal) leads to the system's MTTF decrease.

Assume that the n -th "aging" unit is replaced by a unit with exponential distribution of TTF. Calculate the increment of MTTF, Δ :

$$\Delta = \int_0^{\infty} \prod_{i=1}^n P_i(t) dt - \int_0^{\infty} \exp\left(-\frac{t}{T_n}\right) \cdot \prod_{i=1}^{n-1} P_i(t) dt = \int_0^{\infty} [P_n(t) - \exp(-t/T_n)] \prod_{i=1}^{n-1} P_i(t) dt \quad (23)$$

Since function $P_n(t)$ crosses function $\exp\left(-\frac{t}{T_n}\right)$ only once and from above and their means are equal, then function

$$g(t) = P_n(t) - \exp\left(-\frac{t}{T_n}\right) \quad (24)$$

corresponds to function $g(x)$ of the lemma above. At the same time, function

$\prod_{i=1}^{n-1} P_i(t)$ corresponds to decreasing function $f(x)$ of the same lemma. Thus,

replacement an IFR distribution by an exponential one leads to decrease of the series system MTTF: $\Delta < 0$. Performing such substitutions systematically, one gets the following lower bound for the system MTBF:

$$\underline{T}_{\text{сис.т.}} \geq \int_0^{\infty} \prod_{i=1}^n \exp\left(-\frac{t}{T_i}\right) dt = \int_0^{\infty} \exp\left(-t \sum_{i=1}^n \frac{1}{T_i}\right) dt = \left(\sum_{i=1}^n \frac{1}{T_i}\right)^{-1}. \quad (25)$$

So, the final result can be written as:

$$\left(\sum_{i=1}^n \frac{1}{T_i}\right)^{-1} \leq T \leq \min_{1 \leq i \leq n} T_i. \quad (26)$$

8.6 Parallel system

8.6.1 Probability of failure-free operation

Upper bound.

Let us write formula for the probability of failure of the parallel system:

$$Q(t) = \prod_{i=1}^n q_i(t), \quad (27)$$

where $q_i(t)$ is the probability of the i -th unit failure. Again let us numbering units in order of their MTTF decreasing: $T_1 < T_2 < \dots < T_n$. Notice that for the n -th unit for all $t < T_n$ the upper bound equals $\bar{p}_n(t)=1$, or, equivalently, $q_{n,n}(t) = \underline{Q}(t) = 0$. For $t \geq T_n$, all $q_i(t) = 1 - \exp(-\lambda_i^* t)$, where parameters λ_i^* were found in (10). Thus, for the PFFO of the parallel system consisting of n independent “aging” units, the lower bound, $\underline{Q}(t)$, can be written as:

$$\underline{Q}(t) = \begin{cases} 0 & \text{for } t \leq T_n \\ \prod_{i=1}^n [1 - \exp(-\lambda_i^* t)] & \text{for } t > T_n. \end{cases}, \quad (28)$$

Naturally, that from (28), one gets the upper bound for the system PFFO:

$$\bar{P}(t) = \begin{cases} 1 & \text{for } t \leq T_n \\ 1 - \prod_{i=1}^n [1 - \exp(-\lambda_i^* t)] & \text{for } t > T_n. \end{cases} \quad (29)$$

Notice again that for small time, this bound is trivial and non-informative. If these bound is meaningless, why we pay attention to it? This is given just for information and for protection those who would like to find such bounds. Remember Ancient Greeks who said: “Well-competent means armed”.

Lower bound. Again begin with the upper bound for the system failure probability. Use formula (27) and pay attention to Figure 8, that is, in a sense, a mirror to Figure 6.

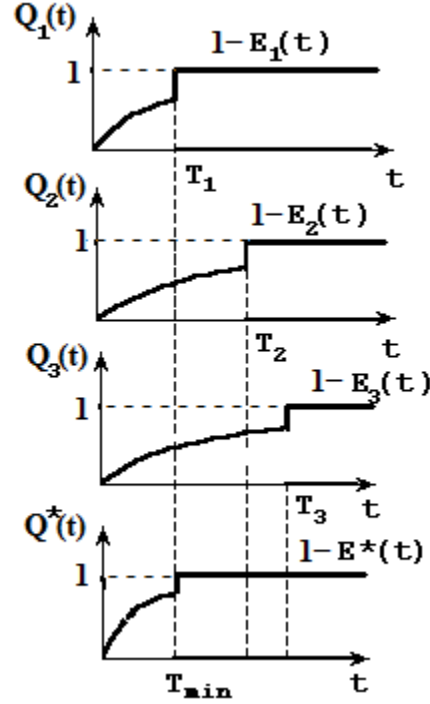


Figure 8. Explanation of the lower bound of failure probability construction on the example of a parallel system of three units.

From Figure above, one can see that for $t > T_3 = T_{max}$ the upper bound is $\bar{Q}(t) = 1$, so the lower bound for PFFO on the same time interval is $\underline{P}(t) = 0$. From Figure 8, one can also see that for $t \leq T_1$, the upper bound for the system's failure probability is

$$\bar{Q}(t) = \begin{cases} \left[1 - \exp\left(-\frac{t}{T_1}\right) \right] \cdot \left[1 - \exp\left(-\frac{t}{T_2}\right) \right] \cdot \left[1 - \exp\left(-\frac{t}{T_3}\right) \right] & \text{for } 0 < t \leq T_1 \\ \left[1 - \exp\left(-\frac{t}{T_2}\right) \right] \cdot \left[1 - \exp\left(-\frac{t}{T_3}\right) \right] & \text{for } T_1 < t \leq T_2 \\ 1 - \exp\left(-\frac{t}{T_3}\right) & \text{for } T_2 < t \leq T_3 \\ 1 & \text{for } t > T_3 \end{cases} . \quad (30)$$

In general case, using the same arguments, one gets:

$$\bar{Q}(t) = \begin{cases} \prod_{i=1}^n [1 - \exp(-t/T_i)] & \text{for } 0 < t \leq T_1, \\ \prod_{i=2}^n [1 - \exp(-t/T_i)] & \text{for } T_1 < t \leq T_2, \\ \dots \\ 1 & \text{for } t > T_n \end{cases} \quad (31)$$

Naturally, the lower bound for the system PFFO can be obtained as complement:

$$P(t) = \begin{cases} 1 - \prod_{i=1}^n [1 - \exp(-t/T_i)] & \text{for } t \leq T_1, \\ 1 - \prod_{i=2}^n [1 - \exp(-t/T_i)] & \text{for } T_1 < t \leq T_2, \\ \dots \\ 0 & \text{for } t > T_n \end{cases} \quad (32)$$

In conclusion notice that in reliability engineering practice the most important bounds, doubtlessly, are the lower ones, since they gives a warranty reliability index value. Fortunately, for “aging” units and systems of “aging” units these bounds are sufficiently informative.

8.6.2 Mean time to failure

Lower bound. This bound can be obtained in usual way:

$$T = \int_0^{\infty} \left[1 - \prod_{i=1}^n q_i(t) \right] dt \geq \max_{1 \leq i \leq n} \int_0^{\infty} [1 - q_i(t)] dt = \max_{1 \leq i \leq n} \int_0^{\infty} P_i(t) dt = T_{\max} \quad (33)$$

Upper bound. Again let us use the result of the above lemma. Let units are numbered in accordance with increasing their MTTF's. The parallel system MTTFis:

$$T = \int_0^{\infty} \left[1 - \prod_{i=1}^n q_i(t) \right] dt \quad (34)$$

As above, let us replace the n -th “aging” unit for a unit with exponential distribution of TTF and the same MTTF:

$$\begin{aligned}
\Delta &= \int_0^{\infty} \left[1 - \prod_{i=1}^n q_i(t) \right] dt - \int_0^{\infty} 1 - \left[\left(1 - \exp\left(-\frac{t}{T_n}\right) \right) \cdot \prod_{i=1}^{n-1} q_i(t) \right] dt = 1 - \int_0^{\infty} \prod_{i=1}^{n-1} q_i(t) \cdot \left[\left(1 - \exp\left(-\frac{t}{T_n}\right) \right) - q_n(t) \right] dt \\
&= 1 - \int_0^{\infty} \prod_{i=1}^{n-1} q_i(t) \cdot \left[p_i(t) - \exp\left(-\frac{t}{T_n}\right) \right] dt.
\end{aligned} \tag{35}$$

The integrand in the last term of (35) is a product of two functions. The first one is product of probabilities of unit failures that corresponds to decreasing function $f(x)$ in the above lemma, and the second one is a difference that corresponds to function $g(x)$ in the same lemma. In accordance with the lemma $\Delta < 0$ that means that replacement “aging” unit for a unit with exponential TTF and the same MTTF increases the parallel system MTTF.

Applying such replacement systematically, we obtain finally:

$$\begin{aligned}
T &\leq \int_0^{\infty} \left[1 - \prod_{i=1}^n \left(1 - \exp\left(-\frac{t}{T_i}\right) \right) \right] dt = \int_0^{\infty} \sum_{i=1}^n \exp\left(-\frac{t}{T_i}\right) dt - \int_0^{\infty} \sum_{1 \leq i < j \leq n} \exp\left(-\frac{t}{T_i}\right) \exp\left(-\frac{t}{T_j}\right) dt + \dots \\
&\dots + (-1)^{n+1} \int_0^{\infty} \prod_{i=1}^n \exp\left(-\frac{t}{T_i}\right) dt,
\end{aligned} \tag{36}$$

that gives, in result, deserved bound.

Thus MTTF of a parallel system consisting of n independent “aging” units has the following lower and upper bounds:

$$\max_{1 \leq i \leq m} T_i \leq T_{\text{сист.}} \leq \sum_{i=1}^m T_i - \sum_{1 \leq i < j \leq m} (T_i^{-1} + T_j^{-1})^{-1} + \dots + (-1)^{m+1} \left(\sum_{i=1}^m T_i^{-1} \right)^{-1}. \tag{37}$$

Remark: If system’s units have IFR-distributions of TTF with almost equal MTTF’s and very small variation coefficient, then TTF of units could be grouped densely (see Figure 9).

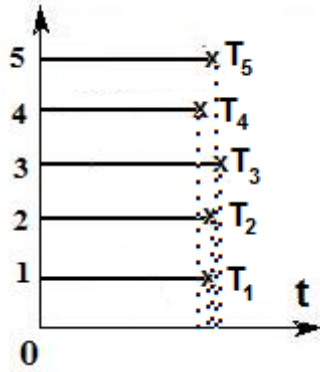


Figure 9. Example of uniy TTF's with close values of MTTF and small variation coefficients.

In this case $T_{\min} \approx T_{\max}$ and it means that reliability indices for series and parallel systems are very close! Indeed, in limit case, when all distributions are degenerate and have the same MTTF, we have $\max_{1 \leq i \leq n} T_i = \max_{1 \leq i \leq n} T$ for any n . In this case, addition an extra unit to a series system does not decrease its reliability as well as addition an extra unit to a parallel system does not increase its reliability.

8.7 Bounds for the coefficient of operational availability

We know that the PFFO during interval t_0 for a recoverable object that starts its operation at arbitrary random moment t equals

$$R(t_0) = \frac{1}{T} \int_t^{\infty} P(t) dt, \quad (38)$$

where $P(t)$ is unit's PFFO and T is its MTTF.

Use again the above lemma. Consider three functions: $P(t)$, $G(t)$ and $E(t)$ introduced above. First take function $g_1(t) = P(t) - G(t)$ that satisfies to conditions for function $g(t)$ in the above lemma. Then:

$$\int_{t_0}^{\infty} [P(t) - G(t)] dx < 0 \Rightarrow \int_{t_0}^{\infty} P(t) dx < \int_{t_0}^{\infty} G(t) dx. \quad (39)$$

Next take function $g_2(x) = P(t) - G(t)$, for which we have:

$$\int_{t_0}^{\infty} [E(t) - P(t)] dx < 0 \Rightarrow \int_{t_0}^{\infty} E(t) dx < \int_{t_0}^{\infty} P(t) dx. \quad (40)$$

From (39) and (40) follows:

$$\int_{t_0}^{\infty} G(t)dx < \int_{t_0}^{\infty} P(t)dx < \int_{t_0}^{\infty} E(t)dx \quad (41)$$

Since $\int_{t_0}^{\infty} G(t)dt = T - t_0$ and $\int_{t_0}^{\infty} E(t)dt = \int_{t_0}^{\infty} \exp\left(-\frac{t}{T}\right)dt$, after substitutions into (41), one gets:

$$T - t_0 < \int_{t_0}^{\infty} P(t)dt < T \exp(t_0) , \quad (42)$$

and, finally:

$$1 - \frac{t_0}{T} < R(t_0) < \exp\left(-\frac{t_0}{T}\right). \quad (43)$$

Obviously, for $t_0 > T$, the left side of inequality turns 0.

Notice that bounds (43) are extremely good for highly reliable objects. Moreover, for small t_0 , one can write a simple and very precise approximation:

$$R(t_0) \approx 1 - \frac{t_0}{T} \quad (45)$$

Stationary coefficient of interval availability is equal to $R^*(t_0) = KR(t_0)$. It means that for highly reliable objects, one can write:

$$R^*(t_0) \approx \left(1 - \frac{t_0}{T}\right) \left(1 - \frac{\tau}{T}\right) \approx 1 - \frac{t_0 + \tau}{T} = 1 - \lambda(t_0 + \tau). \quad (46)$$

Factually, approximation (46) is universal for reliability engineering: one should not know the type of unit's TTF distribution. The only condition has to come true: an object (unit or system) has to be "aging", and this condition comes true practically in all real cases.

8.8 Brief historical overview and related sourcecs

Here we offer only papers and highly related books to the subject of this chapter. List of general monographs and textbooks, which can include this topic, is given in main bibliography at the end of the book.

Swedish engineer, scientist, and mathematician E. Weibull, who paid attention to the wearing-out processes and introduced a two-parameter distribution of rather universal kind. A few years later B. Gnedenko proved a cycle of limit theorems concerning extreme r.v.'s. The so-called Weibull distribution appeared a particular case of the entire class of limit distributions.

Then in 20 years R.Barlow and F. Proschan developed the modern theory of IFT distrivutions that appears very constructive in Reliability Theory.

Bibliography below is given in chronological-alphabetical ordering for better ecposition of historical background of the subject.

Bibliography

- Weibull, W. (1939). A statistical theory of the strength of materials. *Ing.Vetenskaps Akad. Handl.*, No. 151.
- Gnedenko, B.V. (1941) Limit theorems for maximum order statistic (in Russian). *Reports of the Academy of Sciences the USSR*, vol.32, No.1
- Gnedenko, B.V. (1943) Sur la Distribution Limite du Terme Maximum d'Une Serie Aleatorie. *Ann. Math.*, vol.44, No.3
- Weibull, W. (1951) A statistical distribution function of wide applicability. *J. Appl. Mech.* Vol.18, No.3.
- Weibull, W. (1951) A statistical distribution function of wide applicability. *J. Appl. Mech.* Vol.18, No.3.
- Barlow, R.E., A.W.Marshall, and F. Proschan (1963). Properties of probability distributions with monotone hazard rate. *Ann. Math. Stat.*, vol.34.
- Barlow, R.E., and A.W.Marshall. (1964). Bounds for distributions with monotone hazard rate. I and II, *Ann. Math. Stat.*, 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.
- Solovyev, A.D., and I.A. Ushakov (1967) . Some bound on a system with aging elements. (Russian)Automst. Comput. Sci., vol. 6.
- Raizer, V. (2009). Reliability assessment due to wear. RTA Journal, Vol.4, No.1. Gnedenko Forum.

9 Two-Pole Networks

9.1 General comments

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 of existing networks, e.g. communication and computer networks, transportation railroads, gas and oil pipelines, electric power systems and others, have a structure which cannot be described in terms of reducible structures.

In general, networks reliability can be analyzed from different viewpoints. If network is designated for transportation of some material flows then it can be characterized by an ability to deliver required amount of product from sender to receiver. In telecommunication systems the system has to permit its customers to be able to contact each other. We will consider two-pole networks where successful operation is characterized by connectivity between points A and B (see Figure 1).

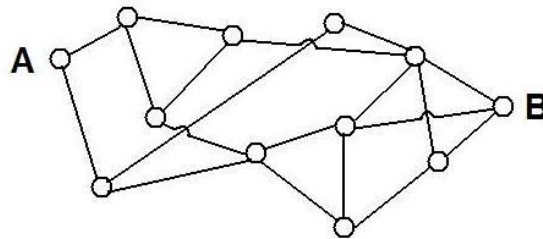


Figure 1. Example of a two-pole network.

It is clear that for two-pole network connectivity necessary and sufficient condition is:

- There is at least one *path* from A to B .
- or
- There are no *cuts* between A and B .

Formally, a path is defined as a set of edges that provides connectivity between points A and B . It is also useful to introduce a *minimum path*, i.e. such set of edges that deletion any of its edges violates network connectivity. A cut is such a set of edges that deletion of all its edges leads to the loss of the network connectivity. For

cuts, a concept of *minimum cut* is also very important. This is such a set of deleted edges that recovering any of them returns connectivity to the network. Examples of paths and cuts are given in Figure 2.

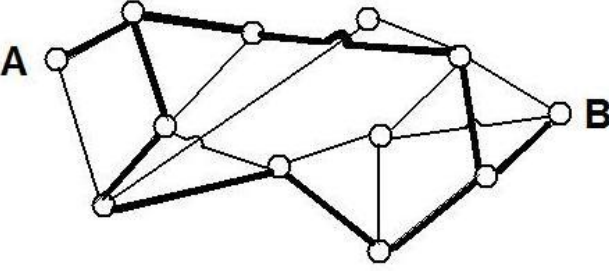
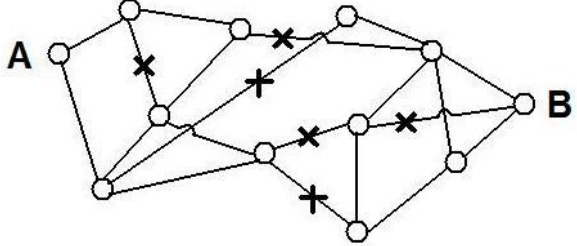
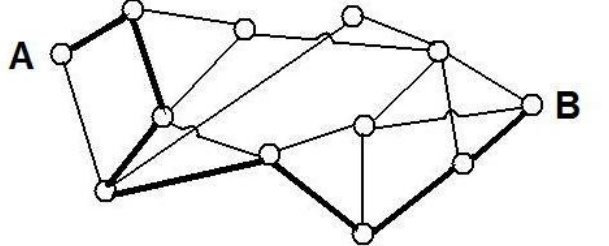
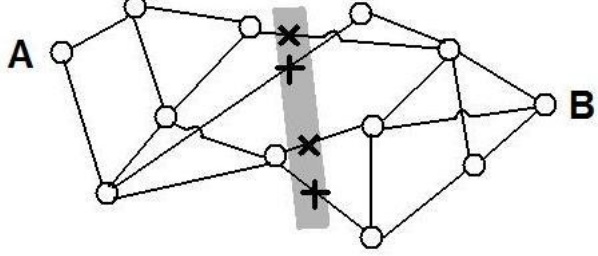
Sample of paths	Sample of cuts
	
Set of edges forming one of paths between A and B	Set of edges forming one of cuts between A and B
	
Set of edges forming one of minimum paths between A and B	Set of edges forming one of minimum cuts between A and B

Figure 2. Examples of paths and cuts of a two-pole network.

For minimum paths and minimum cuts of two-pole network, one can introduce the so-called structural functions that, in a sense, repeat concept of series and parallel connections. For minimum path $X^{(\pi)}$ the structural function has the following Boolean form:

$$\phi^{(\pi)}(X^{(\pi)}) = \bigcap_{i \in X^{(\pi)}} x_i \quad (1)$$

or in verbal explanations:

$$\phi^{(\pi)}(X^{(\pi)}) = \begin{cases} 1, & \text{if and only if all } x_i = 1, x_i \in X^{(\pi)}, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

For minimum cut $X^{(\kappa)}$, the structural function can be written as:

$$\phi^{(\kappa)}(X^{(\kappa)}) = \bigcup_{i \in X^{(\kappa)}} x_i \quad (3)$$

In verbal form (3) is explained as:

$$\phi^{(\kappa)}(X^{(\kappa)}) = \begin{cases} 0, & \text{if and only if all } x_i = 0, x_i \in X^{(\kappa)}, \\ 1, & \text{otherwise.} \end{cases} \quad (4)$$

For further convenience, let us introduce symbol $A \Leftrightarrow B$ to denote connectivity between vertices A and B .

9.1.1 Method of Direct Enumeration

The simplest example of a system with a non-reducible structure is the so-called "bridge structure" (Figure 3).

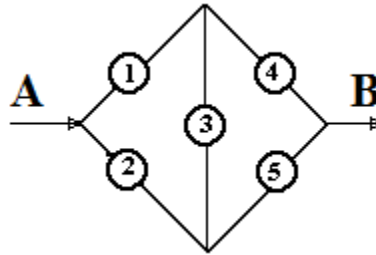


Figure 3. Bridge structure.

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.

For this simple system, it is possible to enumerate all states and check each of

Table 1. System states defined by units' states.

States of units					Vector	Value
x_1	x_2	x_3	x_4	x_5	X_k	$\phi(X_k)$
1	1	1	1	1	X_1	1
0	1	1	1	1	X_2	1
1	0	1	1	1	X_3	1
1	1	0	1	1	X_4	1
1	1	1	0	1	X_5	1
1	1	1	1	0	X_6	1
0	0	1	1	1	X_7	0
0	1	0	1	1	X_8	1
0	1	1	0	1	X_9	1
0	1	1	1	0	X_{10}	1
1	0	0	1	1	X_{11}	1

1	0	1	0	1	X_{12}	1
1	0	1	1	0	X_{13}	1
1	1	0	0	1	X_{14}	1
1	1	0	1	0	X_{15}	1
1	1	1	0	0	X_{16}	0
0	0	0	1	1	X_{17}	0
0	0	1	0	1	X_{18}	0
0	0	1	1	0	X_{19}	0
1	0	0	0	1	X_{20}	1
1	0	0	1	0	X_{21}	0
1	0	1	0	0	X_{22}	0
1	0	0	0	1	X_{23}	0
1	0	0	1	0	X_{24}	1
1	1	0	0	0	X_{25}	0
0	0	0	0	1	X_{26}	0
0	0	0	1	0	X_{27}	0
0	0	1	0	0	X_{28}	0
0	1	0	0	0	X_{29}	0
1	0	0	0	0	X_{30}	0
0	0	0	0	0	X_{32}	0

If edges of the bridge structure are mutually independent, one can easily find the probability of each state. For instance, probability that state X_8 will be realized is: $\Pr\{X=X_8\} = \Pr\{x_1=0\} \Pr\{x_2=1\} \Pr\{x_3=0\} \Pr\{x_4=1\} \Pr\{x_5=1\} = q_1 p_2 q_1 p_2 p_2$, where $p_k = \Pr\{x_k = 1\}$ and $q_k = 1 - p_k$.

Finally, the probability of bridge connectivity, P , can be calculated by the formula:

$$P = E\{\phi(X)\} = \left\{ \sum_{1 \leq k \leq 32} \phi(X_k) \cdot P(X_k) \right\} \quad (5)$$

Omitting intermediate results, we give the final formula for the connectedness probability (in the case of identical units) in two forms

$$P = p^5 - 5p^4 + 2p^3 + 2p^2 \quad (6)$$

$$P = 1 - 2q^2 - 2q^3 + 5q^4 - 2q^5 \quad (7)$$

Expression (7) is useful in case of highly reliable systems where $q \ll 1$, because it permits to get a simple approximation

$$P \approx 1 - 2q^2. \quad (8)$$

Of course, such a method of direct enumeration allows one to compute the probability of the connectedness of a non-reducible two-pole network only in principle. This method is nonapplicable for problems one meets in practice.

9.2 Method of Boolean Function Decomposition

Sometimes the method of decomposition of Boolean functions $\phi(X)$ is very effective. Any Boolean function can be decomposed in respect to its argument:

$$\phi(x_1, \dots, x_k, \dots, x_n) = x_k \phi(x_1, \dots, 1_k, \dots, x_n) \cup \bar{x}_k \phi(x_1, \dots, 0_k, \dots, x_n), \quad (9)$$

or even a set of its arguments (for instance, two of them):

$$\begin{aligned} \phi(x_1, \dots, x_k, \dots, x_n) = & x_k x_j \phi(x_1, \dots, 1_k, \dots, 1_j, \dots, x_n) \cup \bar{x}_k x_j \phi(x_1, \dots, 0_k, \dots, 1_j, \dots, x_n) \cup \\ & x_k \bar{x}_j \phi(x_1, \dots, 1_k, \dots, 0_j, \dots, x_n) \cup \bar{x}_k \bar{x}_j \phi(x_1, \dots, 0_k, \dots, 0_j, \dots, x_n) \end{aligned} \quad (10)$$

where we use 1_k (or 0_k) to show that 1 (or 0) is placed on the k th position. If we interpret the terms of the Boolean function as events, we can say that these events in (9) and (10) are mutually exclusive. For instance, the first term in (9) includes x_k , and the second one includes \bar{x}_k . In this case, we can write for (9)

$$\mathbf{E}\{\phi(x_1, \dots, x_k, \dots, x_n)\} = \mathbf{E}\{x_k \phi(x_1, \dots, 1_k, \dots, x_n)\} + \mathbf{E}\{\bar{x}_k \phi(x_1, \dots, 0_k, \dots, x_n)\}. \quad (11)$$

Since x_k and $\phi(x_1, \dots, 1_k, \dots, x_n)$ are independent as well as \bar{x}_k and $\phi(x_1, \dots, 0_k, \dots, x_n)$, (11) can be finally rewritten in the form

$$\mathbf{E}\{\phi(x_1, \dots, x_k, \dots, x_n)\} = \mathbf{E}\{x_k\} \mathbf{E}\{\phi(x_1, \dots, 1_k, \dots, x_n)\} + \mathbf{E}\{\bar{x}_k\} \mathbf{E}\{\phi(x_1, \dots, 0_k, \dots, x_n)\}. \quad (12)$$

Let us apply this rule to the bridge structure. Choose unit x_3 for decomposition. Then (12) can be rewritten in the concrete form as

$$\begin{aligned} \mathbf{E}\{\phi(x_1, x_2, x_3, x_4, x_5)\} &= \mathbf{E}\{x_3\} \mathbf{E}\{\phi(x_1, x_2, 1, x_4, x_5)\} + \mathbf{E}\{\bar{x}_3\} \mathbf{E}\{\phi(x_1, x_2, 0, x_4, x_5)\} = \\ &= p_3 \mathbf{E}\{(x_1 \cup x_2) \cap (x_4 \cup x_5)\} + q_3 \mathbf{E}\{(x_1 \cap x_4) \cup (x_2 \cap x_5)\}. \end{aligned} \quad (13)$$

So, we came to a ‘‘mixture’’ of series-parallel and parallel-series systems. It becomes clear with the following explanations. What does it mean that $x_3=1$? It means that in the initial bridge structure, unit x_3 is absolutely reliable (always in the operational state). Thus, the bridge structure becomes a simple series-parallel structure. Similarly, $x_3=0$ means that unit x_3 is eliminated from the structure (always in a failed state). This means that the structure becomes a parallel-series structure. (See Figure 4.)

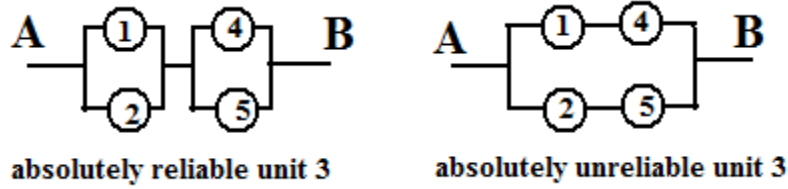


Figure 4. Transformation initial bridge structure into series-parallel and parallel-series systems depending on state of unit 3.

Finally, we may write

$$P = p_3[(1-q_1q_2)(1-q_4q_5)] + q_3[1 - (1-p_1p_4)(1-p_2p_5)]. \quad (14)$$

We should mention that a Boolean function can be decomposed by any variable. In this particular example, such decomposition can be done with respect to any x_k . For example let us decompose the same structure in respect to unit 1. Corresponding structures are presented in Figure 5,

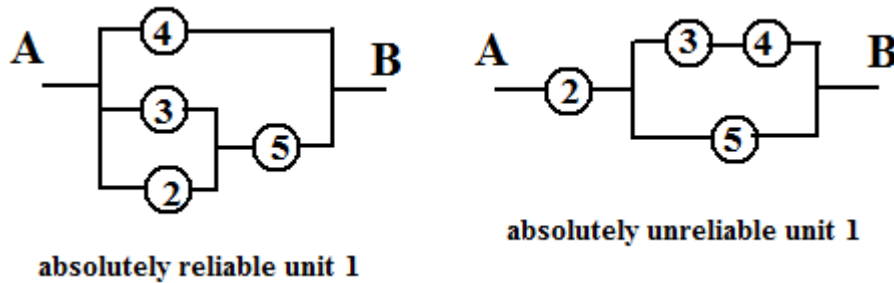


Figure 5. Transformation initial bridge structure into series-parallel and parallel-series systems depending on state of unit 1.

In this case the expression for the bridge structure connectivity has the form:

$$P = p_1\{1 - (1 - q_4) \cdot [1 - (1 - q_3q_2)p_5]\} + q_3p_2[1 - q_5(1 - p_3p_4)] \quad (15)$$

Of course, (14) and (15) are equivalent, though (14) is much more elegant!

In general, the decomposition method is not practically effective, even if one applies it using decomposition in respect to several Boolean variables. In short, this idea of network decomposition in fact nearly always represents only a nice illustrative example, not an effective tool for engineers.

All of the difficulties connected with the numerical analysis of non-reducible structures lead to a need to find other methods. One effective analytical method is obtaining lower and upper bounds of the unknown value of the probability of connectivity.

9.3 Method of Paths and Cuts

9.3.1 Esary-Proschan Bounds

Consider an arbitrary two-pole network. Assume that all network vertices (or nodes) are absolutely reliable. As we stated above, for two-pole network connectivity at least one minimum path has to exist. Structural function $\phi(\mathbf{X})$ for arbitrary two-pole network can be written as:

$$\phi(\mathbf{X}) = \bigcup_{1 \leq k \leq N} \phi_k^{(\pi)}(\mathbf{X}) \quad (16)$$

Using de Morgan Rule, (16) can be rewritten in the form:

$$\phi(\mathbf{X}) = \bigcup_{1 \leq k \leq N} \phi_k^{(\pi)}(\mathbf{X}) = \overline{\bigcap_{1 \leq k \leq N} \overline{\phi_k^{(\pi)}(\mathbf{X})}} \quad (17)$$

where N is the number of all minimum paths of considering two-pole network.

Factors in (17) are mutually dependent, since different paths may have the same links. From the Probability Theory we know that if events X_1 and X_2 are dependent

$$\Pr\{X_1 \cap X_2\} \neq \Pr\{X_1\}\Pr\{X_2\}. \quad (18)$$

In [Barlow, Proschan, 1965] associated random variables were introduced. Two r.v.'s X and Y are called associated if their covariance is positive:

$$\text{Cov}(X, Y) = E\{X - E\{X\}\} \cdot E\{Y - E\{Y\}\} \geq 0, \quad (19)$$

They naturally generalized this concept on multivariate case that appeared very useful and productive for reliability analysis of multicomponent complex systems. For associated r.v.'s,

$$\Pr\{X_1 \cap X_2 \cap \dots \cap X_n\} = \Pr\left\{\bigcap_{1 \leq k \leq n} X_k\right\} \geq \prod_{1 \leq k \leq n} \Pr\{X_k\} \quad (20)$$

Notice that paths of two-pole network are positively correlated, i.e. increase reliability of one of them cannot lead to decrease of reliability of another one. Indeed, if reliability of a common unit for both of these two paths is improved, it improves reliability both those paths simultaneously. (Same is observed with a common unit's reliability decrease: both paths decrease their reliability.)

So, since for parallel connection of paths, the probability of disconnection $\prod_{1 \leq k \leq n} \Pr\{X_k = 0\}$ delivers the lower bound, the upper bound for the connectiveness of the two-pole network has the form:

$$\bar{P} = 1 - \prod_{1 \leq k \leq N} (1 - \Pr\{\phi_k^{(\pi)}(\mathbf{X}) = 0\}) \quad (21)$$

So, (21) is the upper bound for probability of the two-pole network connectivity.

To feel a real sense of all these abstract deduction, let us again demonstrate the method on a bridge structure, presented in Figure 3. All possible minimum paths of this structure are given in Figure 6.

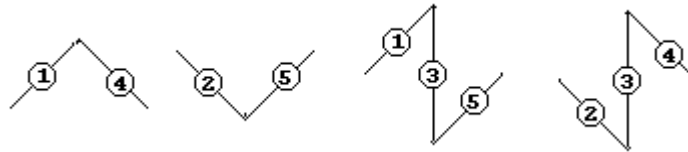


Figure 6. Minimum paths of a bridge structure.

Expression (21) for this particular case takes form

$$P = 1 - (1 - p^2)^2 \cdot (1 - p^3)^2 \quad (22)$$

We can formulate a similar natural condition of a connectedness violation by the following equivalent statement expressed via a network's minimum cuts. As we noticed above, for violation of the two-pole network connectivity links of at least one minimum cut have to be failed. Structural function $\phi(\mathbf{X})$ for arbitrary two-pole network can be expressed through minimum cuts as follows:

$$\phi(\mathbf{X}) = \bigcap_{1 \leq k \leq M} \phi_k^{(\kappa)}(\mathbf{X}) \quad (23)$$

where M is the number of all minimum cuts of considering two-pole network.

Notice again that network's minimum cuts may be interdependent because they may contain the same units, and again we observe the positive correlation. Keeping this fact in mind, one can write:

$$\underline{P} = \prod_{1 \leq k \leq M} \left(1 - \prod_{i \in \beta_k} q_i \right) \quad (24)$$

Formula (24) gives the lower bound of value P .

Let us illustrate this method on the bridge structure. Minimum cuts of the bridge structure are presented in Figure 7.

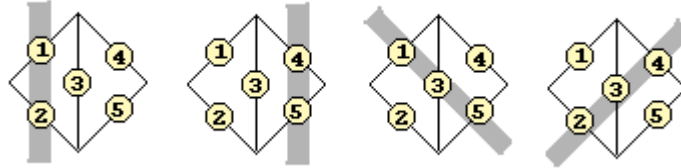


Figure 7. Minimum cuts of a bridge structure.

Below in Figure 8, numerical comparison of the exact values and the upper and lower bounds is given.

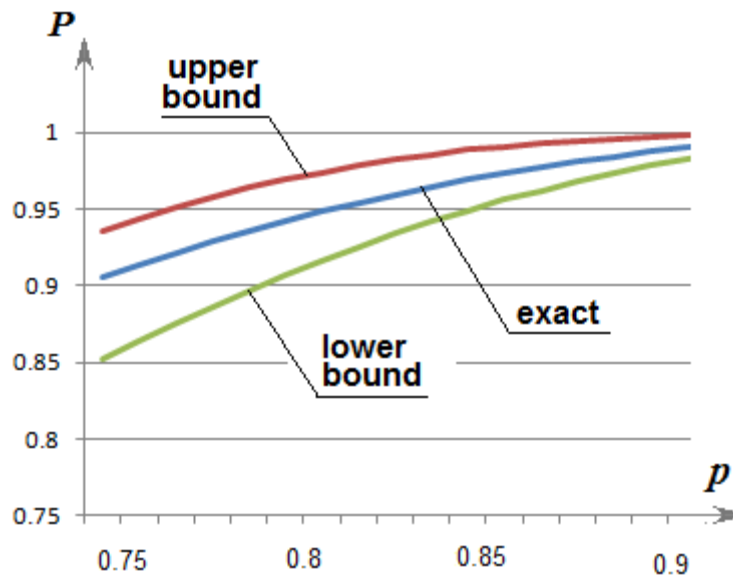


Figure 8. Comparison of exact values of bridge structure connectivity and its Esary-Proschan bounds.

9.3.2 “Improvements” of Esary-Proschan bounds

These bounds are based on sets of non-intersecting simple paths and cuts.

We first point out that, for complex networks, the enumeration of all of the different simple paths and cuts is a very difficult problem demanding a huge

computer memory and an enormous computational time. For systems of any practical dimension, this enumeration problem is essentially impossible.

For this reason, one sometimes attempts to make the computations shorter. On a heuristic level, an explanation of the main idea of such an attempt follows. If we consider a very complex multi-unit network, we often can find that a lower bound includes some very "thick" cuts, i.e., cuts with a large number of links. It is clear that such a parallel connection is characterized by an extremely high reliability. There is a temptation to exclude such "thick" cuts from consideration: they are very reliable in comparison with the remaining cuts. In other words, the value $1-q_1q_2\dots q_k$ is so close to 1 that it seems reasonable to replace it with 1. This leads to the increase of the reliability index. Thus after this, the new lower bound should be even higher than the initial strong lower bound. However, the higher the lower bound the better!

Analogously, the strong upper bound includes some "very long" series connections which may be very unreliable. Again the question arises: why should one take into account such practically absolutely unreliable series connection of units for a computation a reliability index of a parallel connection? Indeed, for very large m , value of $p_1p_2\dots p_m \rightarrow 0$. If one neglects such "very long" paths, the new upper bound becomes lower. This again produces a better upper bound than we have initially!

We must emphasize that such a "heuristic heuristic" leads to the very rough mistakes. Indeed, the higher the lower bound, the better, but only if the lower bound remains a lower bound!

We may obtain strange results using these "simplifications" and "improvements" of the bounds: the obtained "improved" bounds may not be even bound the unknown value at all! In fact, an "improved" lower bound, obtained in such manner, may be even larger than an "improved" upper bound! Thus, we may obtain new "bounds" which lost all mathematical meaning at all!

Once more we would like to emphasize that a real heuristic is not an arbitrary guess on an "intuitive level". In our opinion, a heuristic usually must be an "almost proven" simplification of an existing strong solution. Sometimes, instead of a proof one may deliver a set of numerical examples, covering the parametrical area of domain, as a confirmation of the heuristic's validity. Such "experimental mathematics" occupies more and more room in computational methods and very often replaces exact proofs.

Let us illustrate some possible mistakes of using the above-mentioned "simplification" on an example of a bridge structure consisting of identical units. Represent approximations of an upper bound $U = 1-(1-p^2)^2(1-p^3)^2$ in the form:

$$U^* \approx 1-(1-p^2)^2 \quad , \quad (25)$$

where we keep only the “shortest” minimal paths, and of a lower bound

$L = (1-q^2)^2(1-q^3)^2$ in the form:

$$L^* \approx (1-q^2)^2 \tag{26}$$

where we keep only the “thinnest” minimal cuts.

It is easy to check that $L^* \geq U^*$. We prefer not to use straight dull transformations to prove this fact. We show this on numerical examples for $p=0.9$, $p=0.1$ and $p=0.5$:

Table 2. Comparison of U^* and L^* for various value of p .

p	U^*	L^*
0.9	0.9639	0.9801
0.5	0.4375	0.5625
0.1	0.0199	0.0361

We see that “improved” lower bound became larger than “improved” upper bound!

9.3.3 Litvak-Ushakov Bounds

First of all, remind a very important property of reliability of systems – property of monotonicity. This property absolutely natural and means, in common terms, that if reliability of any system unit is increased, it cannot lead to decreases of the system reliability as a whole. Of course, inverse statement is also true: decreasing unit’s reliability cannot result in the system reliability increase. Construction of Litvak-Ushakov bounds is based on this property. The idea is to find a set of independent minimum paths, i.e. paths that that have no the same link in different paths.

For expansion of the procedure of constructing such set, let us begin with an example. In Figure 9, the first minimum path (links 1, 4, 8 and 11) is marked with black lines. Remember this path and exclude its entire links (the second figure). In remaining part of two-pole network find another minimum path (links 3, 5, 9 and 12). After deleting the second path from the initial network, we see that points A and B are disjoint.

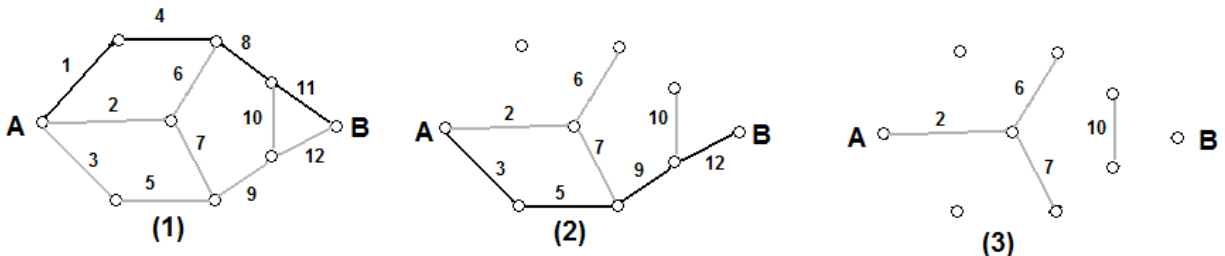


Figure 9. Procedure of constructing non-intersected minimum paths.

This is the end of procedure of finding set of independent minimum paths: initial network is reduced to parallel connection of two minimum paths. Now look at Figure 10: one gets the same two paths by excluding the links that are depicted in Figure 8 (3).

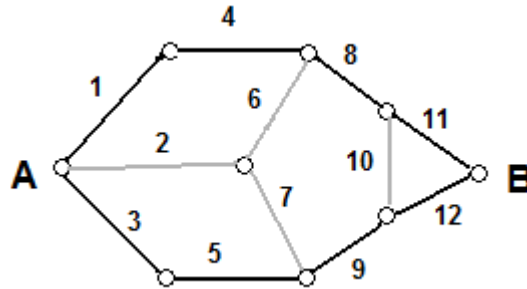


Figure 10. Construction of set of independent minimum paths by exclusion links 2, 6, 7 and 10.

Excluding any link is equivalent to decrease its reliability to zero. So, the probability of connectivity between nodes A and B for two parallel minimum paths is lower than the same probability for the initial network.

By the way, the number of independent minimum paths cannot be larger than the number of links in the “thinnest” minimum cut. Moreover, if, for instance, the path depicted in Figure 11 is chosen, there are no other independent paths in the same two-pole network.

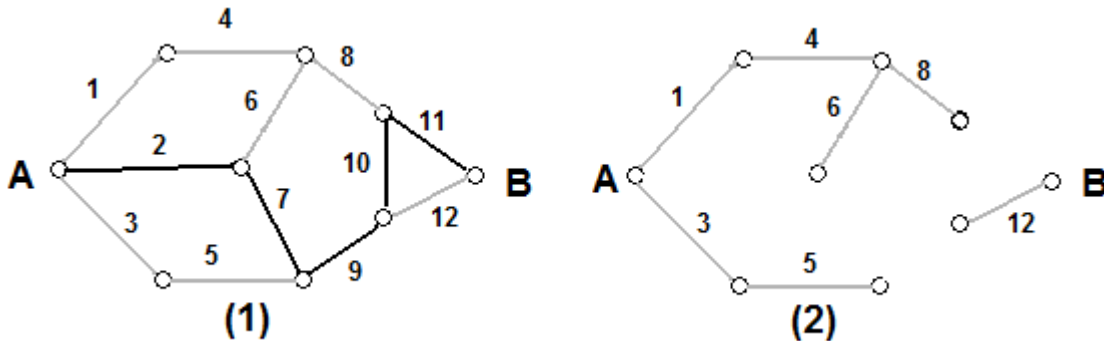


Figure 11. An example of a case with a single minimum path.

In general case, one can find several sets of minimum paths in the same two-pole network. Each such set forms a parallel connection of different minimum paths with its own probability of connectivity that lower than analogous probability for the initial two-pole network. It is clear that the maximum of lower bounds is the best one. Denote by $G_k^{(\pi)}$ k -th subset of minimum paths, $k = \overline{1, N}$, and by $g_{ki}^{(\pi)}$ subset of

links composing the i -th path of the k -th subset. Then the lower Litvak-Ushakov bound is:

7 Let us again give more detailed explanation on a bridge structure with identical links. All possible minimum paths are depicted in Figure 12.

Set $G_1^{(\pi)}$	Set $G_2^{(\pi)}$	Set $G_3^{(\pi)}$
$g_{11}^{(\pi)} = (x_1, x_4), g_{12}^{(\pi)} = (x_2, x_5)$	$g_{21}^{(\pi)} = (x_1, x_3, x_5)$	$g_{31}^{(\pi)} = (x_2, x_3, x_4)$

Figure 12. All minimum paths of bridge structure.

The lower bound of Litvak-Ushakov type for bridge structure is:

$$\underline{P} = \max\{[1 - (1 - p^2)^2], p^3, p^3\} = 1 - (1 - p^2)^2 \quad (28)$$

The upper bound can be obtained from series connection of independent cuts. For expansion of the procedure of constructing the upper bound, let us again refer to the two-pole network presented in Figure 9. In Figure 13 we demonstrate procedure of sequential construction of a set of independent cuts (each cut here marked with grey strip). After choosing the first cut, we gather all right ends of corresponding links into a single node. This procedure is equivalent to putting into the network absolutely reliable links that (by definition of structural monotomity) can only increase reliability atire network.

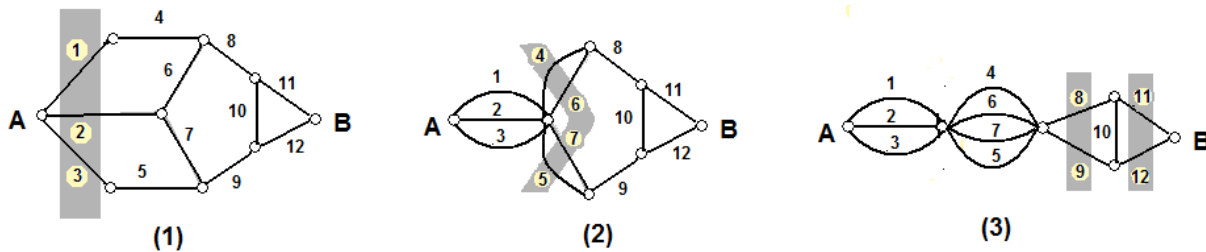


Figure 13. Procedure of constructing non-intersected minimum paths.

At the last step in this particular case, we can assume link 10 absolutely reliable that opleted the construction of this set of independent cuts. The final result can be presented in Figure 14.

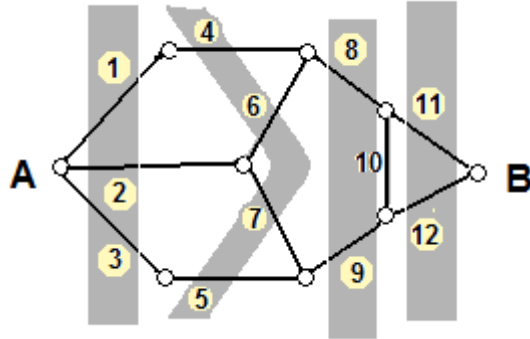


Figure 14. Initial network with marked minimum cuts.

Notice that the number of independent minimum cuts cannot be larger than the number of links in the “longest” minimum path. Moreover, if in Figure 15 one chooses cut formed by links 3, 7, 10 and 11, it will be the only cut.

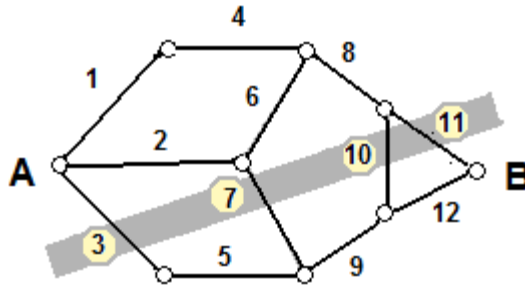


Figure 15. An example of a case with a single minimum path.

In general case, there are several sets of minimum cuts in the same two-pole network. Each such set forms a series connection of different minimum cuts with its own probability of connectivity that higher than analogous probability for the initial two-pole network. It is clear that the maximum of lower bounds is the best one. Denote by $G_k^{(\kappa)}$ k -th subset of minimum cuts, $k = \overline{1, M}$, and by $g_{ki}^{(\kappa)}$ subset of links consisting the i -th cut of the k -th subset. Then the upper Litvak-Ushakov bound can be written as:

$$\bar{P} = \min_{1 \leq k \leq M} \left\{ \prod_{i \in G_k^{(\kappa)}} \left(1 - \prod_{j \in g_{ki}^{(\kappa)}} (1 - p_j) \right) \right\} \quad (29)$$

Let us again give more detailed explanation on a bridge structure with identical links. All possible minimum cuts are depicted in Figure 12.

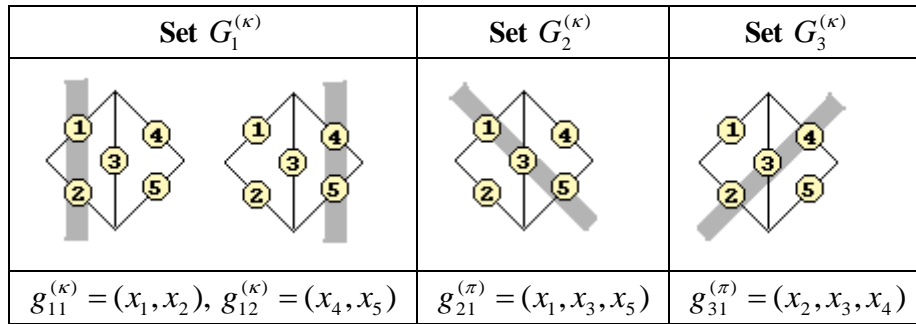


Figure 12. All minimum cuts of bridge structure.

The upper bound of Litvak-Ushakov type for bridge structure is:

$$\bar{P} = \min\{[1-(1-p)^2] \cdot [1-(1-p)^2], 1-(1-p)^3, 1-(1-p)^3\} = [1-(1-p)^2]^2. \quad (27)$$

9.3.4 Comparison of the Two Methods

Again we will not misuse mathematical deductions and refer to results of numerical calculations (see Figure 13). Here we use subscripts EP for Esary-Proschan bounds and subscript LU for Litvak-Ushakov bounds. What we can see from the table?

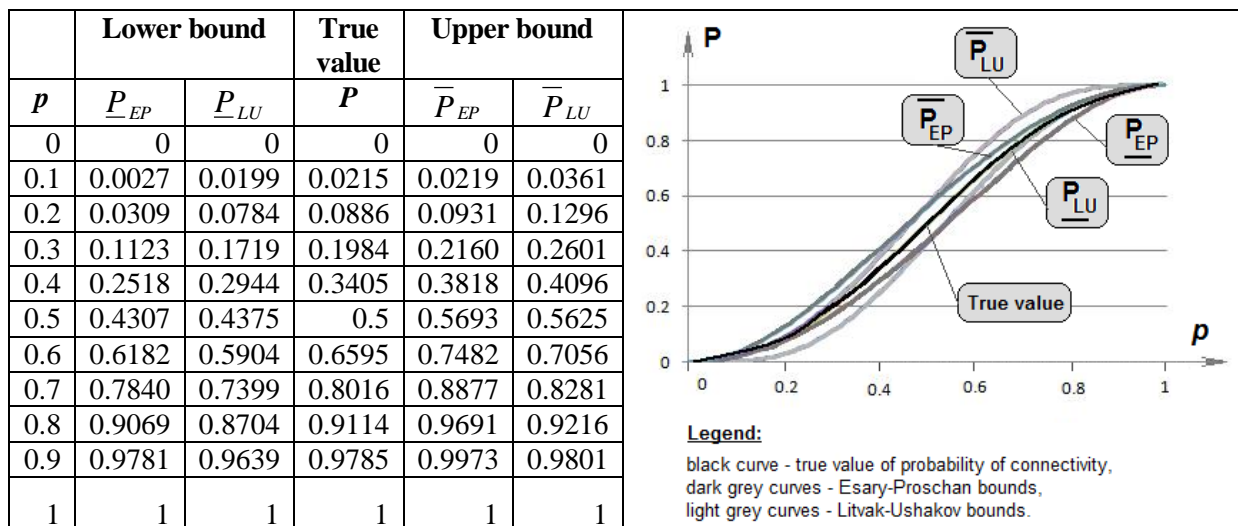


Figure 13. Comparison of Esary-Proschan and Litvak-Ushakov bounds.

For higher values of unit's reliability index, p , the best lower bound is delivered by P_{EP} . It is important to mention that this bound is one of the most important: for highly reliable two-pole networks: we need to know the warranty level of reliability. (Let us repeat again that if the network is not reliable enough, one should think about reliability improvement, not about reliability evaluation.) Around $p = 0.5$, Esary-Proschan and Litvak-Ushakov bounds cross each other. Then at the area of small values of p , Litvak-Ushakov method delivers the better lower bound. One of the main advantages of the method of obtaining bounds by this method is simplicity of obtaining sets of disjoint paths and cuts.

9.4 Brief historical overview and related sources

Here we offer only papers and highly related books to the subject of this chapter. List of general monographs and textbooks, which can include this topic, is given in main bibliography at the end of the book.

One of the first work dedicated to the problem of two-pole networks connectedness [Moore and Shannon, 1956] related to connectedness analysis of so-called "hammock-like" networks. Then series of works by Esary and Proschan (1962, 1963, 1970) were dedicated to obtaining network bounds based on the enumerating of all simple paths and cuts. A decomposition method with the use of this same idea was proposed in [Bodin, 1970].

The bounds based on disjoint simple paths and cuts were obtained in [Ushakov and Litvak, 1977] and later were generalized in [Litvak and Ushakov, 1984]. Multiterminal (multipole) networks were investigated in [Lomonosov and Polessky, 1971].

Bibliography below is given in chronological-alphabetical ordering for better exposition of historical background of the subject.

Bibliography

- Moore, E., C. Shannon (1956) Reliable circuits using less reliable relays. J. Franklin Inst. No.9.
- Esary, J.D., and F. Proschan (1962). The reliability of coherent systems. In: Redundancy Techniques for Computing Systems. Spartan Books, Washington, D.C.
- Ushakov, I., Yu. Konenkov (1964). Evaluation of branching systems operational effectiveness. (Russian) In: "Cybernetics for Service to Communism". Energiya.
- Kelmans, A.K. (1970) On estimation of probabilistic characteristics of random graphs. Automat. Remote Control., Vol. 32, No.11.
- Stepanov, V. E. (1970) On the probability of graph connectivity. Probab. Theory and Appl., Vol. 15, No.1.
- Lomonosov, M.V., and V. Polessky (1971). Upper bound for the reliability of information networks. Problems Inform. Transmission. Vol. 7
- Lomonosov, M.V., and V. Polessky (1972). Lower bound of network reliability. Problems Inform. Transmission. Vol. 8
- Gadasin, V.A., and I.A. Ushakov (1975). Reliability of Complex Information and Control Systems (Russian). Sovetskoe Radio.
- Ushakov, I.A., and E.I. Litvak (1977). An upper and lower estimate of the parameters of two-terminal networks. Engrg. Cybernet. vol. 15
- Litvak, E., I. Ushakov. (1984) Estimation of parameters of structurally complex systems. Engrg, Cybernet. vol. 22, no.4.
- Kaustov V.A., E.I. Litvak, and I.A. Ushakov. (1986). The Computational Effectiveness of Reliability Estimates by the Method of Nonedge-Intersecting Chains and Cuts. Soviet Journal of Computer and System Sciences, Vol. 24, No. 4.

- Reinshke, K., I. Ushakov. (1987), Application of Graph Theory for Reliability Analysis, (in German), Verlag Technik.
- Reinshke, K., I. Ushakov. (1988), Application of Graph Theory for Reliability Analysis, (in Russian), Radio i Sviaz.
- Reinshke, K., I. Ushakov. (1988), Application of Graph Theory for Reliability Analysis, (in German), Springer.Elperin, T, I. Gertsbakx, M. Lomonosov (1991) Estimation of network reliability using graphs evolution models. IEEE Trans. Reliab., No.40.
- Ushakov, I. (1989). Reliability Analysis of Computer Systems and Networks (Russian). Mashinostroenie.
- Lomonosov, M. (1994) On Monte Carlo estimates in Network Reliability. Probab. Engng. Inf. Sci., No.8.
- Ushakov, I. (1994) Methods of Research in Telecommunications Reliability (An Overview of Research in the Former Soviet Union). RTA
- Rauzy, A. (2001) Mathematical foundations of minimal cut sets. IEEE Trans. on Reliability. Vol. 50, No.4.
- Lin, Y. (2002) Using minimal cuts to evaluate the system reliability of a stochastic-flow network with failures at nodes and arcs, Reliability Engineering & System Safety, 75.
- Fotuhi-Firuzabad, M., R. Billinton, T.S. Munian, B. Vinayagam. (2003) A novel approach to determine minimal tie-sets of complex network. IEEE Trans. on Reliability. Vol. 52, No.4.
- Wei-Chang Yen (2007) A Simple Heuristic Algorithm for Generating All Minimal Paths. IEEE Trans. on Reliability. Vol. 56, No.3.

10 . *Performance Effectiveness*

10.1 Effectiveness Concepts

Modern large scale systems are characterized by structural complexity and sophisticated algorithms to facilitate the functioning and interacting of its subsystems. One of main properties of such systems is that they have a significant “safety factor”: even a set of failures may not lead to the system failure, though can somehow decrease effectiveness of its performance. Indeed, telecommunication networks have highly redundant structures, transportation systems have a number of bypasses, oil and gas supply systems can change their regime to adjust to new conditions of operation. It allows these systems perform their operations with lower though still acceptable level of effectiveness even with some failed units and subsystems. For such systems, “white-and-black” approach is not appropriate: it is practically impossible to give a definition of system failure, it is more reasonable to say about some degradation of system’s abilities; it is natural to speak about performance effectiveness of such systems. In each concrete case, the feature of an index (or indices) of performance effectiveness should be chosen with respect to the type of system under consideration, its destination, conditions of its operation, etc. The physical sense of the *performance effectiveness index (PEI)* is 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. In other words, we might take into account the nominal (specified) value of a system's outcome as the normalizing factor. In other words, a PEI is a measure of quality and/or volume of its performed functions or operations, i.e., it is a measure of the system's expediency.

Of course, a system's efficiency depends on the type of currently performed functions and operating environments. A system which is very efficient under some circumstances might be quite useless and ineffective under another set of circumstances and/or operations.

If a system's outcome has an upper bound, the PEI can be expressed in a normalized form, i.e., 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 $0 \leq PEI \leq 1$.

Consider a system consisting of n units, each of which can be in two states: operating and failure. Let x_i be the indicator of the i th unit's state: $x_i=1$ when the unit

is operable and $x_i = 0$ when the unit has failed. The system has $N = 2^n$ different states determined by the states of its units. Denote a system state by $\mathbf{X} = (x_1, x_2, \dots, x_n)$.

With time the system change its state due to changes of states of its units. Transformation of system states $\mathbf{X}(t)$ in time characterizes the system's behavior. For this state, the performance effectiveness coefficient equals $W_{\mathbf{X}}$, and the system's PEI can be determined as the expected values of $W_{\mathbf{X}}$:

$$PEI = E\{W_{\mathbf{X}}\} = \sum_{1 \leq k \leq N} W_{\mathbf{X}_k} P(\mathbf{X}_k). \quad (1)$$

Nevertheless, there are systems, for which (1) is not valid. For these systems, effectiveness depends on an entire trajectory of the system's transition from one state to another during some predetermined time interval. In other words, for these systems the effectiveness is determined by a trajectory of states changing during the system's performance of task.

Examples of such systems are different technological processes, information and computer systems, aircrafts in flight, etc.

10.2 General idea of effectiveness evaluation

Let $h_{\mathbf{X}_k}(t)$ denote the probability that the system at the moment t is in the state $\mathbf{X}_k(t)$. We assume that the current any state can be evaluated. Let us denote effectiveness of the system being in state \mathbf{X} is $W_{\mathbf{X}}$. It is naturally to determine the mathematical expectation of $W_{\mathbf{X}}$ as:

$$W_{\text{sys}}(t) = \sum_{1 \leq k \leq N} h_{\mathbf{X}_k}(t) W_{\mathbf{X}_k} \quad (2)$$

It is clear that an absolutely accurate calculation of the system's effectiveness when $n \gg 1$ is a difficult, if not unsolvable, computational problem. First of all, one needs to determine a huge 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.

The W_{sys} can then be calculated as

$$W_{\text{sys}}(t) = \sum_{1 \leq j \leq M} W_j \sum_{\mathbf{X}_k \in G_j} h_{\mathbf{X}_k}(t) \quad (3)$$

where M is the number of different levels of the values of $W_{\mathbf{X}}$ and G_j is the set of system's states, for which $W_{\mathbf{X}}$ belongs to the j th level.

Let us evaluate system's effectiveness for a general case. For simplicity of notation, 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 the moment t :

$$h_0 = \prod_{1 \leq i \leq n} p_i. \quad (4)$$

Let h_i denote the probability that only the i th unit of the system is in a down state at the moment t . Then

$$h_i = q_i \prod_{1 \leq j \leq n; j \neq i} p_j = \frac{q_i}{p_i} h_0 = g_i h_0 \quad (5)$$

where, for brevity, we introduce $g_i = \frac{q_i}{p_i}$; and h_{ij} denotes the probability that only the i th and j th units of the system are in down states at the moment t :

$$h_{ij} = q_i q_j \prod_{1 \leq k \leq n; k \neq (i,j)} p_k = \frac{q_i q_j}{p_i p_j} h_0 = g_i g_j h_0 \quad (6)$$

and so on.

We can write the general form of this probability as

$$h_X = \prod_{i \in G_X} p_i \prod_{i \in \bar{G}_X} q_i = h_0 \prod_{i \in \bar{G}_X} g_i \quad (7)$$

where G_X is the set of subscripts of the units, which are considered operational in state \mathbf{X} ; and \bar{G}_X is the complimentary set.

Sometimes it is reasonable to write (7) for any \mathbf{X} as:

$$h_X = \prod_{\substack{1 \leq i \leq n \\ x_i \in X}} p_i^{x_i} q_i^{(1-x_i)}. \quad (8)$$

It is clear that (7) and (8) are equivalent. Using (4)-(8), we can rewrite (3) as:

$$W_{\text{sys}} = W_0 h_0 [1 + \sum_{1 \leq i \leq n} \tilde{W}_i g_i + \sum_{1 \leq i < j \leq n} \tilde{W}_{ij} g_i g_j + \dots] \quad (9)$$

where W_0 is the system's effectiveness for state X_0 , and $\tilde{W}_i, \tilde{W}_{ij}, \dots$, are normalized effectiveness coefficients for states X_i, X_{ij}, \dots . In other words, $\tilde{W}_i = \frac{W_i}{W_0}$, $\tilde{W}_{ij} = \frac{W_{ij}}{W_0}$,

For a system consisting of highly reliable units, i.e. when $\max_{1 \leq i \leq n} q_i \ll \frac{1}{n}$

expression (9) can be approximated as:

$$W_{\text{sys}} \approx W_0 \left(1 - \sum_{1 \leq i \leq n} q_i \right) \cdot \left(1 + \sum_{1 \leq i \leq n} \tilde{W}_i q_i \right) = W_0 (1 - \sum_{1 \leq i \leq n} q_i \tilde{w}_i) \quad (10)$$

Here $\tilde{w}_i = 1 - \tilde{W}_i$ has the meaning of a "unit's insignificance".

Remark: It is necessary to note that, strictly speaking, it is wrong to say about "unit's significance". The significance of a unit depends on the specific state of system. For example, in a simple redundant system of two units, the significance of any unit equals 0 if both units are successfully operating, but one unit is failed, then its significance of remaining one becomes equal to 1.

10.2.1 Conditional case study: Airport traffic control system.

An airport traffic control system consists of two stationary radars each with an effective zone of 180° (schematic plot of the system is presented in Figure 1). The availability coefficient for each radar is equal to $K = 0.95$.

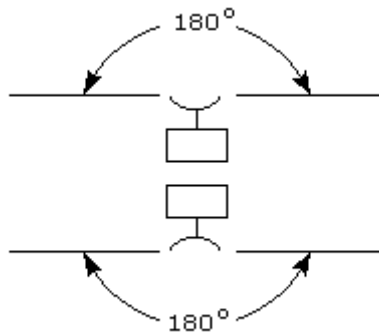


Figure 1. Structure of airport radar system.

Assume that if only one radar is operating, the system effectiveness drops to 50% of the nominal value.

It is easy to write the expression for PEI calculation:

$$PEI = K^2 \cdot 1 + 2K(1 - K) \cdot 0.5 = K^2 + K(1 - K) = K = 0.95. \quad (11)$$

There are two variants of the system effectiveness improvement (they are depicted in Figure 2).

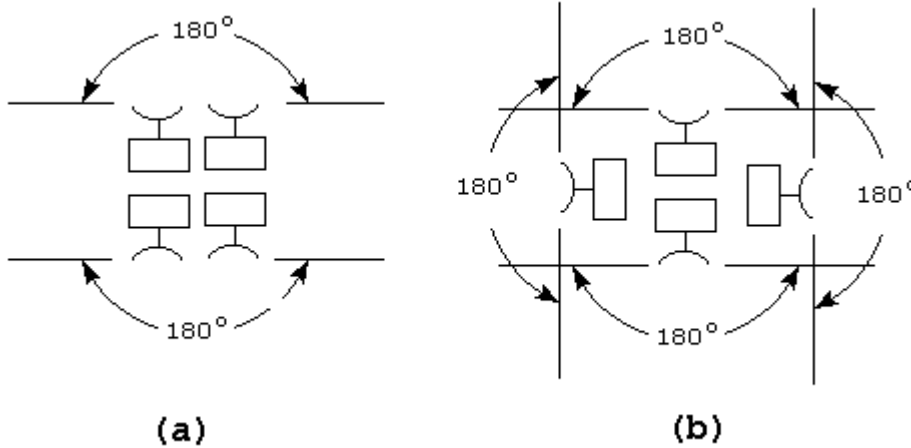


Figure 2. Two variants of airport radar system improvement.

First variant (a) is a usual redundancy.

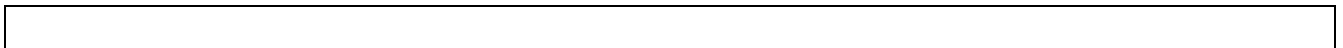
All radars operate	One radar has failed (4 different variants)	Two “opposite” radars have failed (4 various variants)	Two neighbor radars have failed (2 different variants)	Three radars have failed (4 different variants)	All radars have failed.

Figure 3. Possible locations of redundant radars failures.

In Figure 3 arrow shows the direction of radar radiation, and grey color denote the area covered by radiation of all currently operating radars. With the help of Figure 4, one easily can write the expression of PEI for this type of configuration:

$$PEI^{(1)} = [K^4 + 4K^3(1 - K) + 4K^2(1 - K)^2] \cdot 1 + [2K^2(1 - K)^2 + 4K(1 - K)^3] \cdot 0.5 \quad (12)$$

Now consider the second variant where radars form a “ring”.



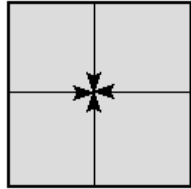
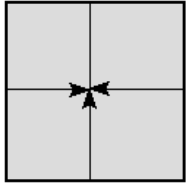
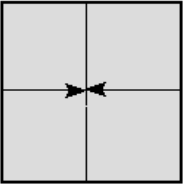
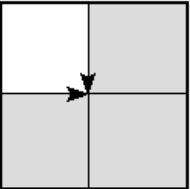
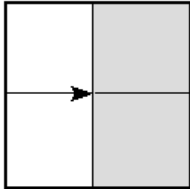
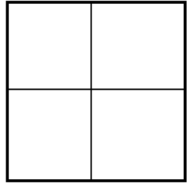
					
All radars operate	One radar has failed (4 different variants)	Two “back-to-back” radars have failed (2 various variants)	Two neighbor radars have failed (4 different variants)	Three radars have failed (4 different variants)	All radars have failed.

Figure 4. Possible locations of radars failures in case “ring structure”.

Basing on Figure 4, one can write the following expression:

$$PEI^{(2)} = [K^4 + 4K^3(1-K) + 2K^2(1-K)^2] \cdot 1 + 4K^2(1-K)^2 \cdot 0.75 + 4K(1-K)^3 \cdot 0.5. \quad (13)$$

These simple analysis shows that the second variant is better, though the difference is not too significant:

$$PEI^{(2)} - PEI^{(1)} = K^2(1-K)^2 \quad (14)$$

For “numerical feeling”, in Figure 5 comparison of both variants is given.

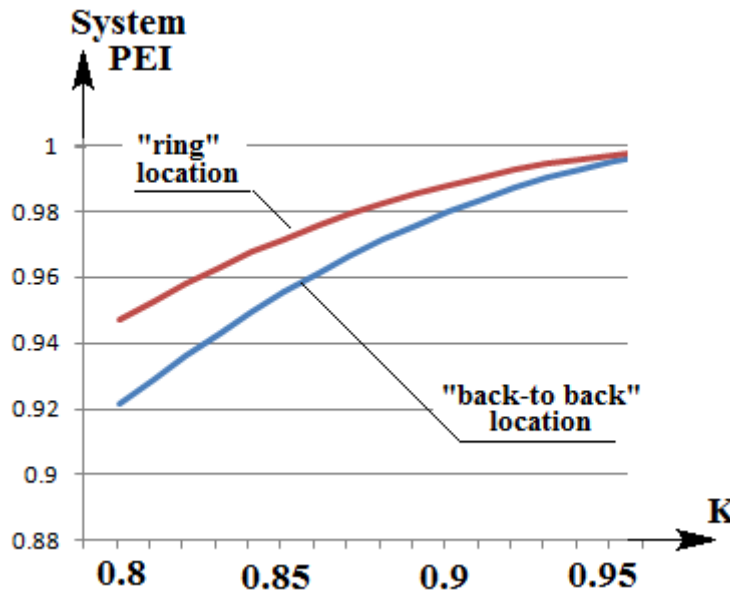


Figure 5. Comparison of two variants of radars location.

10.3 Additive Type of System’s Units Outcome

Let system consist of N “executive” units, each of which is characterized by its own outcome:

$$w_i = \begin{cases} W_i, & \text{if unit is operable} \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

All other units of the system are considered as “administrative” that provide required functioning of executive units. So the system PEI can be represented as the sum of the units’ outcomes:

$$PEI = E\left\{ \sum_{1 \leq i \leq N} w_i \right\} = \sum_{1 \leq i \leq N} E\{w_i\}. \quad (16)$$

Expression (11) is true even if system’s units are dependent. This follows from well known in mathematical statistics fact that the expected value of a sum of random variables equals the sum of its expected values, regardless of their dependence.

Let a system has the following structure (see Figure 1): Control center and a set of N executive units monitored from the center. Assume that for successful operating of an executive unit it is necessary both the center and the unit were operable.

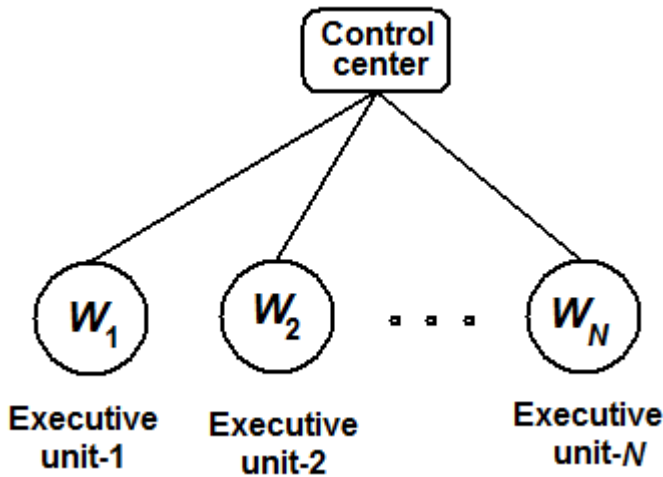


Figure 1. Structure of a system with several executive units.

The i th executive unit characterized by its own outcome in operable state, $W_i(X)$, and probability of successful operation, p_i , $1 \leq i \leq N$. The control center is characterized by its PFFO equals p_0 . In this case, the average outcome of the i th executive unit is equal to:

$$E\{w_i\} = p_0 p_i W_i \quad (17)$$

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:

$$PEI = p_0 \sum_{1 \leq i \leq N} p_i W_i \quad (18)$$

10.4 Case study: ICBM control system

A clear practical example of such a system can be represented by a so-called non-symmetrical branching system with a simple tree-like hierarchical structure. This system consists of N executive units controlled by "structural" units at higher hierarchy levels. This example is a reminiscent of the author's participation in design the system for ICBM controlling. Of course, this example presents a schematic and very much conditional structure of the system (see Figure 2).

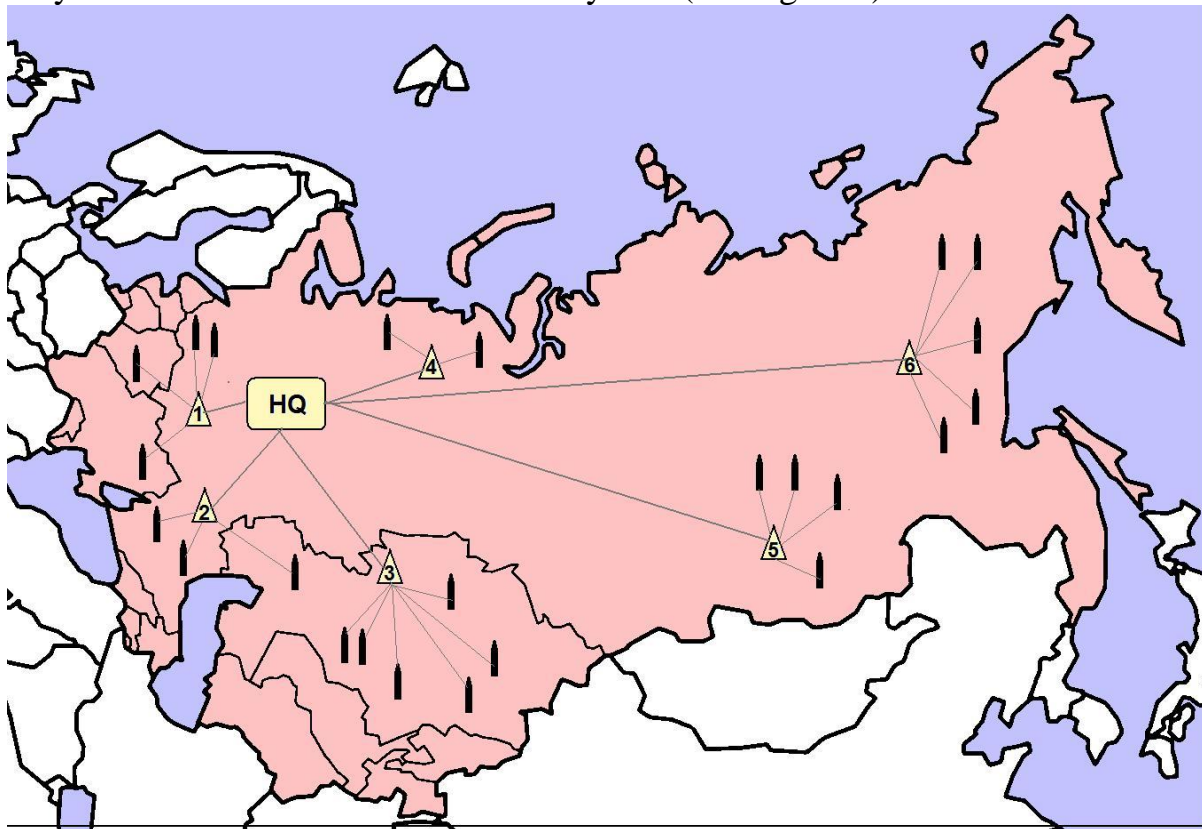


Figure 2. Tree-like hierarchical structure of ICBM system.

On this figure, Headquarter (HQ) controls regional centers (1, 2, ..., 6), and they, in turn, control underground silos with ICBMs. Assume that all silos have PFFO equal 0.99, Hadquarter PFFO is 0.995 and regional centers have various PFFO

due to the communication systems, natural environments, etc. Let the corresponding probabilities for regional centers are: $p_1 = 0.995$, $p_2 = 0.993$, $p_3 = 0.992$, $p_4 = 0.99$, $p_5 = 0.89$, $p_6 = 0.87$. The same system in more formalized form is presented in Figure 3.

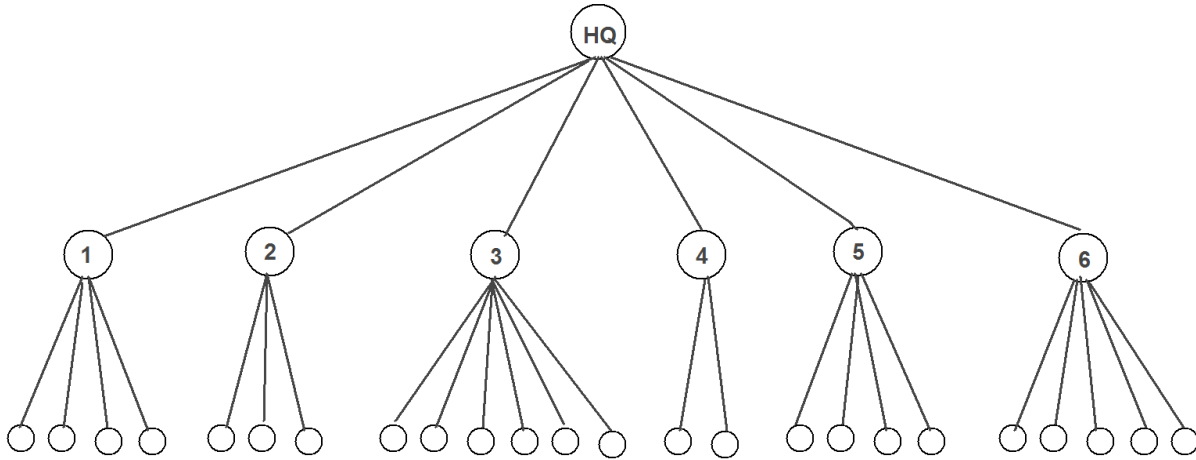


Figure 3. Formalized scheme of the system oriented in Figure 2.

For such type of a system, a natural measure of importance is the average number of available ISBMs. It is easy to calculate that for given parameters the average of available ICBMs equals ≈ 22.45 .

The question could arise: what is the way to increase the system PEI? For instance what is more effective: to increase the Headquarter PFFO from 0.995 to 0.999 or to improve silos and increase the PFFO level up to 0.995?

Simple calculations show that in the first case the system PEI will be ≈ 22.67 , and in the second case will be ≈ 22.57 . So, the difference in two variants is insignificant, however improving 24 silos, probably, leads to larger expenses.

10.5 Systems with intersecting zones of action

Suppose that a system consists of n executive units. Unit i has its own zone Z_i of action. Each unit is characterized by its own effectiveness of action W_i in the zone Z_i . These zones can be overlapping. Let us denote subzones, z , obtained in result of overlapping by z with subscripts corresponding to the number of zones that form these subzones (see Figure 4). Actually, in practice the number of different subzones is restricted enough because overlapping is observed only among neighboring zones.

Such mathematical models appear in the analysis of satellite intelligence systems, in radio communication networks, in power systems, and in anti-aircraft and anti-missile systems (overlapping zones of destruction). Effectiveness of

performance within a particular subzone depends on type of the systems and performed operations.

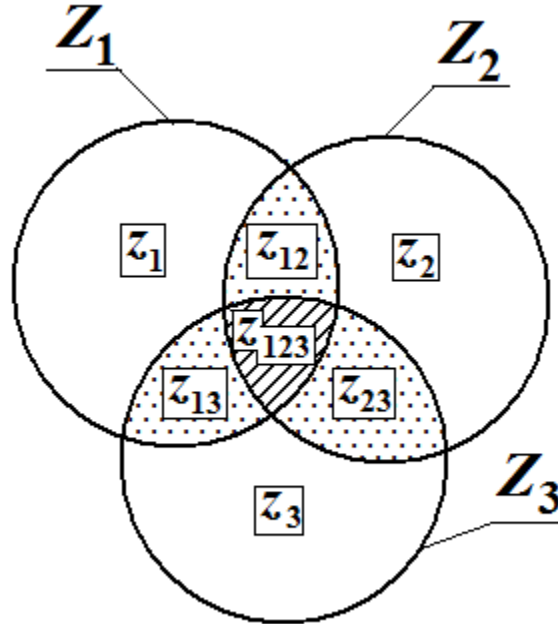


Figure 4. Example of a system consisting of three overlapping zones. Z_1 , Z_2 and Z_3 are zones and z_{12} , z_{13} , z_{23} and z_{123} are corresponding subzones.

Consider several particular cases.

1. Maximum “intensity index” within subzone.

Assume that each zone individually is characterized by its own “intensity index”, w_k , per a unit of square. It means that zone k with square Z_k and with no overlapping with other zones, has input $w_k Z_k$ in the total system’s PEI. If zones Z_k and Z_j overlap, then within the subzone z_{jk} “intensity index” is equal to $\max\{w_j, w_k\}$. Subzones with double overlapping are characterized by $\max\{w_i, w_j, w_k\}$, and so on.

This mathematical model can be used for describing a reconnaissance satellite (or spy satellite) system. In this case, if a territory is covered by several satellites, the best image is used for further processing.



Calculation algorithm in this case is simple enough, though enumeration remains enumeration. Let us demonstrate the methodology of calculation on a simple example of three zones. Let zone k is operable with probability p_k . The system can be in $2^3 = 8$ different states.

First, order all systems in order of value of “intensity indices”. Let, for concreteness, $w_1 > w_2 > w_3$.

All input data can be tabulated into the following table.

Table 1. Input data for calculation of the system PEI.

	z_1	z_2	z_3	z_{12}	z_{13}	z_{23}	z_{123}
1, 1, 1	w_1	w_2	w_3	w_1	w_1	w_2	w_1
0, 1, 1	0	w_2	w_3	w_2	w_3	w_2	w_2
1, 0, 1	w_1	0	w_3	w_1	w_1	w_3	w_1
1, 1, 0	w_1	w_2	0	w_1	w_1	w_2	w_1
1, 0, 0	w_1	0	0	w_1	w_1	0	w_1
0, 1, 0	0	w_2	0	0	0	w_2	w_2
0, 0, 1	0	0	w_3	0	w_3	w_3	w_3
0, 0, 0	0	0	0	0	0	0	0

The system’s performance effectiveness index is calculated by formula:

$$\begin{aligned}
 PEI = & p_1 p_2 p_3 [w_1 (z_1 + z_{12} + z_{13} + z_{123}) + w_2 (z_2 + z_{23}) + w_3 z_3] + \\
 & q_1 p_2 p_3 [w_2 (z_2 + z_{12} + z_{23} + z_{123}) + w_3 (z_3 + z_{13})] + p_1 q_2 p_3 [w_1 (z_1 + z_{12} + z_{13} + z_{123}) + w_3 (z_3 + z_{23})] + \\
 & p_1 p_2 q_3 [w_1 (z_1 + z_{12} + z_{13} + z_{123}) + w_2 (z_2 + z_{23})] + p_1 q_2 q_3 w_1 (z_1 + z_{12} + z_{13} + z_{123}) + \\
 & q_1 p_2 q_3 w_2 (z_2 + z_{12} + z_{23} + z_{123}) + p_1 p_2 q_3 w_3 (z_3 + z_{13} + z_{23} + z_{123})
 \end{aligned} \tag{19}$$

It is clear that even such a simplified example led to clumsy calculations. In practice, such systems have more or less homogeneous nature: subzones without overlapping have the same values of w_i , $= w^{(0)}$ subzones with a single overlapping have also the same values $w_{ij} = w^{(1)}$ and so on. In addition, zones are the same size and structure itself is “recurrent”, usually honeycomb type (see Figure 4).

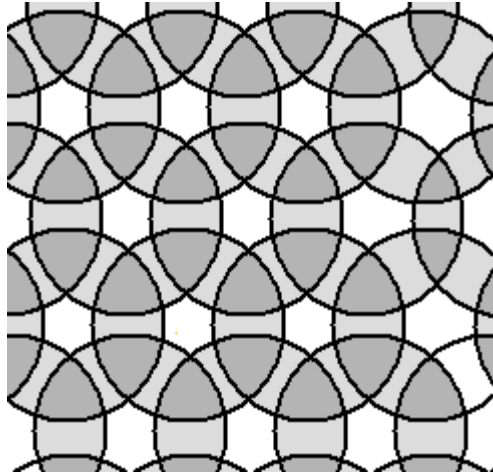


Figure 5. Honeycomb type of zones overlapping.

In this case, it is enough to analyze a single zone Z_1 with all its subzones formed with by 6 neighboring zones.

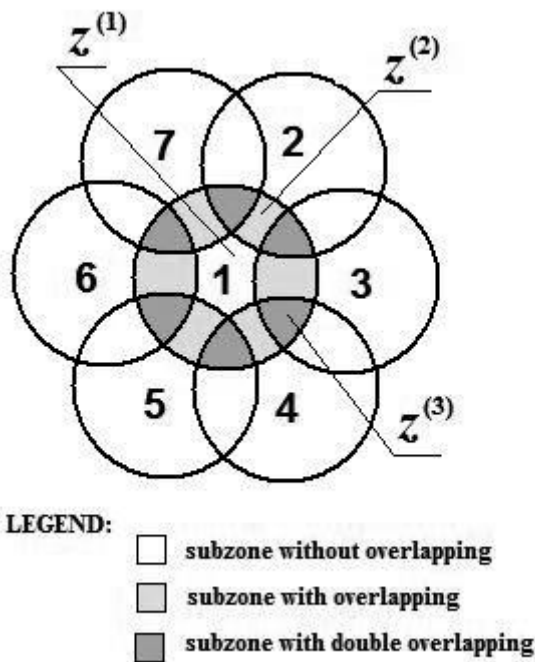


Figure 6. Zone Z_1 with its neighboring zones. Here $z^{(k)}$ denotes three types of subzone configuration.

However, even in this artificially simplified case, calculation of performance effectiveness index is simple due to enumerating nature of the problem. Let us consider a highly reliable system when probability of occurring more than two failures is insignificant. In this case, there are two types of a single failure and four types of two failures (see Figure 5).

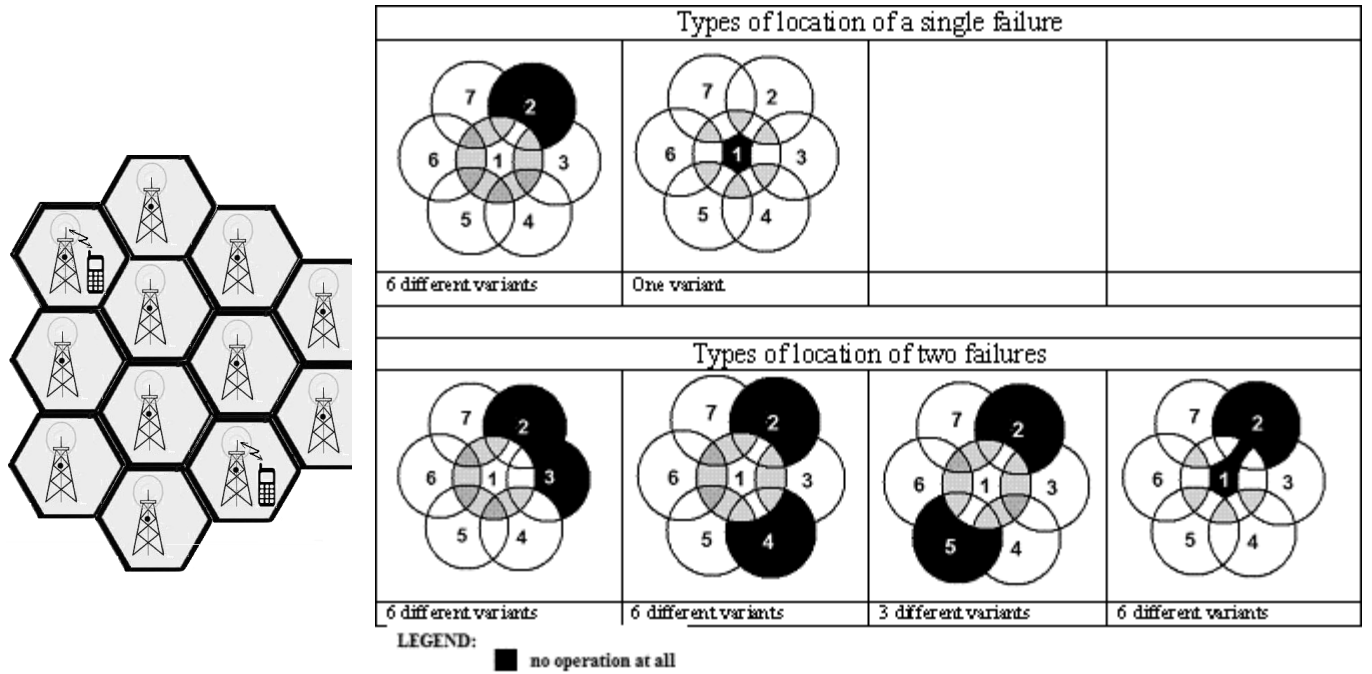


Figure 7. Possible locations of one and two failures.

Basing on Figure 6, one can write:

$$\begin{aligned}
 PEI \approx & p^7 (w^{(1)} z^{(1)} + 6w^{(2)} z^{(2)} + 6w^{(3)} z^{(3)}) + p^6 q \{ 6[w^{(1)} (z^{(1)} + z^{(2)}) + w^{(2)} (5z^{(2)} + 2z^{(3)}) + \\
 & + 3w^{(3)} z^{(3)}] + 6w^{(2)} z^{(3)} + 6w^{(1)} z^{(2)} \} + \\
 & + p^5 q^2 \{ 6[w^{(1)} (z^{(1)} + 2z^{(2)} + z^{(3)}) + w^{(2)} (4z^{(2)} + 2z^{(3)}) + 3w^{(3)} z^{(3)}] \} + \\
 & + p^5 q^2 \{ 6[w^{(1)} (z^{(1)} + 2z^{(2)}) + w^{(2)} (4z^{(2)} + 4z^{(3)}) + 2w^{(3)} z^{(3)}] \} + \\
 & + p^5 q^2 \{ 3[w^{(1)} (z^{(1)} + 2z^{(2)}) + w^{(2)} (4z^{(2)} + 4z^{(3)}) + 2w^{(3)} z^{(3)}] \} + \\
 & + p^5 q^2 \{ 6[w^{(1)} (5z^{(2)} + 2z^{(3)}) + 4w^{(2)} z^{(3)}] \}.
 \end{aligned} \tag{20}$$

Of course, (6) can be simplified by collecting terms but we omit primitive though rather boring transformations, keeping in mind that the main target in this case is explanation, not final result.

2. Boolean type of “intensity index” within subzone.

Such mathematical model can describe an area with a set of ground base stations serve the cell phone customers: a customer is served or not depending on the zone where he or she is currently in. If a customer is in a zone of intersection of several base stations, one of them is chosen for operation.

This case is similar to the previous one, if we put $w^{(1)}= w^{(2)}= w^{(3)} = w$. Using (5), one can immediately write an approximate expression:

$$PEI \approx w\{p^7(z^{(1)} + 6z^{(2)} + 6z^{(3)}) + p^6q[6(z^{(1)} + 6z^{(2)} + 2z^{(3)}) + 3z^{(3)} + 6z^{(3)}]\} \quad (21)$$

3. Redundancy type of “intensity index” within subzone.



Imagine that zones represent anti-aircraft or anti-missile areas of defense. In this case, a target in a subzone without overlapping can be shut down with probability p , within subzone with overlapping it happens with probability $1-q^2$, within subzone with double overlapping it happens with probability $1-q^3$, and so on.

To get approximate expression for PEI, one can use again (6), keeping in mind that in this case $w^{(1)} = p$, $w^{(2)} = 1-(1-p)^2$ and $w^{(3)} = 1-(1-p)^3$.

10.6 Practical Recommendation

Analysis of the system performance effectiveness 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 the successful analysis. Of course, the system analyst should also be acquainted with apply operations research methods. As any operations research problem, the task is concrete and its solution is more of an art than a science.

For the 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:

- formulation of an understandable and clear goal of the system;
- determination of all possible system's tasks (operations, functions);
- choice of the most appropriate measure of system effectiveness;
- division of a complex system into subsystems;
- compilation of a structural-functional scheme of the system which reflects the interaction of the system's subsystems;
- collection of reliability data;

- computation of the probabilities of the different states in the system and its subsystems;
- estimation of the effectiveness coefficients of different states;
- performance of the final computations of the system's effectiveness.

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.

Only in rare cases it is possible to find these coefficients by the means of analytical approaches. The most common method is simulation of the system behavior with the help of computerized models or physical analogues 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 the 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 an analysis of a system's effectiveness performance is not routine. To design a mathematical model of a complex system, in some sense, is a problem similar to the problem of the design of a system itself. Of course, there are no technological difficulties, no time or expense for engineering design and production.

10.7 Brief historical overview and related sources

Here we offer only papers and highly related books to the subject of this chapter. List of general monographs and textbooks, which can include this topic, is given in main bibliography at the end of the book.

Actually, the main idea of the system's performance effectiveness contains (with the accuracy of terminology) in [Kolmogorov, 1945]. The first application to Reliability Theory is described in [Ushakov, 1960].

Bibliography below is given in chronological-alphabetical ordering for better exposition of historical background of the subject.

Bibliography

1. Kolmogorov, A.N. (1945) The number of hits in a few shots and general principles for evaluating the effectiveness of the fire. Proc. Moscow Steklov Math. ин-та им. Inst.В. А. Стеклова, 12 , Изд-во АН СССР, М.–Л., 1945, 7–25, Vol. 12. Moscow-Leningrad, USSR Ac,Sci.
2. Ushakov, I.A. (1960) Effectiveness estimation of complex systems. In *Reliability of Electronic Equipment*. Moscow, Sovetskoe Radio.
3. Ushakov, I.A. (1966). Performance effectiveness of complex systems. In *On Reliability of Complex Technical Systems*. (Russian). Moscow, Sovetskoe Radio.

4. Ushakov, I. A.(1967). On reliability of performance of hierarchical branching systems with different executive units. Engrg. Cybernet.Vol.5, No.5.
5. Ushakov, I.A., and Yu. K. Konenkov (1964). Evaluation of branching systems operational effectiveness. (Russian) In: "Cybernetics for Service to Communism". Moscow, Energiya.
6. Ushakov, I.A., and S. Chakravarty (2002) Reliability influence on communication network capability. Methods of Quality Management, No.7.
7. Ushakov, I.A., and S. Chakravarty (2002) Reliability measure based on average loss of capacity. International Transaction in Operational Research, No.9.

11 On system survivability

With development extremely complex systems, especially worldwide terrestrial systems, a new problem arose – survivability of systems. Our life is full of unexpected extreme situations: various natural disasters like earthquakes, floods, forest fires, etc. Last decades are marked by evil terrorist actions performing by political terrorists or just mentally sick maniacs.

Dealing with natural disasters, we are protecting ourselves against the nature. Here we deal with unpredicted events, and as Albert Einstein²⁵ told: “God is subtle but he is not malicious”²⁶. Quite different situation with “Homo Sapiens”: he directs his evil will against the most vulnerable and painful points. In a sense, such actions are “more predictable”, though it is almost impossible to guess the next step of a crazy maniac or evil suicider.

Taking a historical prospective, we should remember Russian academician Alexei Krylov²⁷ who was one of the first who formulated the problem of survivability of sea vessels against enemy attack. He suggested a witty method of keeping a vessel afloat: in response on destruction of some vessel’s module he suggested to fill intentionally a symmetrical module with water and so keep the vessel in vertical position (see Figure below). This method of floodability is widely used since then and all military vessels now have module construction of their holds.

²⁵ **Albert Einstein** (1879 –1955) was a German-born American theoretical physicist who developed the theory of general relativity,

²⁶ Inscription in Fine Hall of the Princeton University.

²⁷ **Alexei Nikolaevich Krylov** (1863-1945) was Russian and Soviet academician, naval architect, mathematician and mechanician.

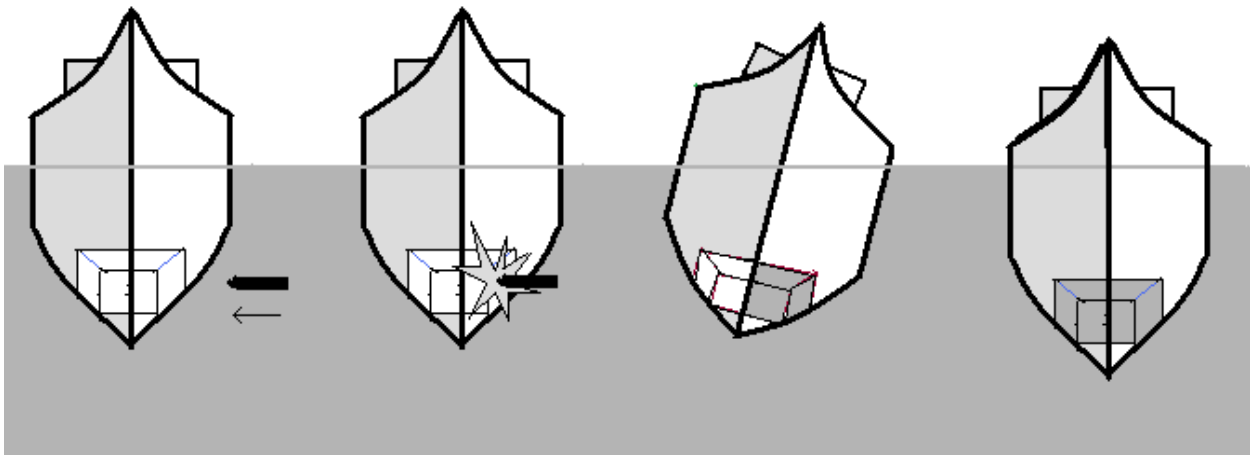


Figure 1. Explanation of Krylov's idea to keep a vessel afloat after enemy torpedo attack.

The nature of the survivability problem hints that there is no place for pure reliability methods: the destructive events are not predictable in probabilistic sense and there is no such factor as calendar time. Methodology of survivability (or, in inverse terms, vulnerability) is not developed sufficiently by now. However, it is impossible to pass by this problem. Below we will give some simple ideas with no pretensions.

It seems that in this case the most appropriate is minimax approach. In reliability analysis of networks we already met with the concepts of “minimal cut”, i.e. with such minimum set of system units which failures lead to the failure of the entire network.

Actually, the only known method of survivability investigation is the so called “What –If Analysis”. It is a structured brainstorming method of determining what consequences can happen with the considered system if one or another situation will have occurred. Usually, situations under consideration cannot be modeled physically, so the judgments are done by a group of experts on the basis of their intuition and previously known precedents.

One of natural measurement of survivability is the system effectiveness after a given set of impacts. Of course this residual effect depends on the number of impacts, their nature, intensity and location of exposure.

Due to numerical indefinites of expert estimates, comparison of survivability of various systems under some certain impacts is more reasonable than absolute measurements. Consider N various systems designated for performing the same actions (operations). Let each system consists of n units, each of them characterized its own “level of protection”, specified to each system. Denote π_{ij} the level of protection of unit i against impact of type j . Assume that there are n possible different impacts of different “destructive level” $\omega_1, \omega_2, \dots, \omega_n$. Ascribe

to each system's unit a loss in case of impact influence, u_{is} , where $i = \overline{1, n}$ and $s = \overline{1, N}$. These losses can be expressed in absolute (of the same dimension) or relative values. Let $U_s(x_{i_1}, x_{i_2}, \dots, x_{i_m})$ denote the system loss when set of its units $x_{i_1}, x_{i_2}, \dots, x_{i_m}$ is destructed. In a sense, the problem is to make such impacts assignment that delivers the maximum possible loss for a system. (It will be worst scenario for "defending side".)

Introduce additional indicator function of type;

$$\delta_{ip} = \begin{cases} 1, & \text{if } \omega_p > \pi_i, \text{ i.e. destructive level larger than protection,} \\ 0, & \text{otherwise} \end{cases}$$

For each system, on the basis of input data presented above, one compiles the following table:

Table 1. Loss value if impact j is applied to unit i of system s .

		Type of impact			
		1	2	...	n
Type of unit	1	$\delta_{11}u_{1s}$	$\delta_{12}u_{1s}$...	$\delta_{1n}u_{1s}$
	2	$\delta_{21}u_{2s}$	$\delta_{22}u_{2s}$...	$\delta_{2n}u_{2s}$

	n	$\delta_{n1}u_{ns}$	$\delta_{n2}u_{ns}$...	$\delta_{nn}u_{ns}$

Thus, in this table some of values are equal to 0, which means that some impacts are harmless for definite units. Assume that to any unit a single impact can be assigned, i.e. $\sum_{1 \leq i \leq n} \delta_{ip} = 1$ and $\sum_{1 \leq p \leq n} \delta_{ip} = 1$.

Notice that there are three general types of systems in sense of their "response" on a set of impacts:

- Systems with linear loss function, i.e. such functions that the total loss is equal of the sum of units' losses: $U_s(x_{i_1}, x_{i_2}, \dots, x_{i_m}) = \sum_{1 \leq r \leq m} u_{i_r, s}$;
- Systems with convex loss function, i.e. such function that the total loss is less than the sum of units' losses: $U_s(x_{i_1}, x_{i_2}, \dots, x_{i_m}) < \sum_{1 \leq r \leq m} u_{i_r, s}$;
- Systems with concave loss function, i.e. such function that the total loss is larger than the sum of units' losses: $U_s(x_{i_1}, x_{i_2}, \dots, x_{i_m}) > \sum_{1 \leq r \leq m} u_{i_r, s}$.

Consideration of two last cases in general form has no sense since functions $U_s(x_{i_1}, x_{i_2}, \dots, x_{i_m})$ should be defined for each specific case. Thus, let us focus on the

linear loss function. In this case, the problem of system survivability estimation reduced to finding

$$U_s = \max_{\delta} \left\{ \sum_{1 \leq i \leq n} \delta_{ip} u_{is} \mid \sum_{1 \leq i \leq n} \delta_{ip} = 1 \text{ and } \sum_{1 \leq p \leq n} \delta_{ip} = 1 \right\}$$

After such calculations for each s , $s = \overline{1, N}$, the final solution is found as $\min_{1 \leq s \leq N} U_s$.

To make the idea of survivability evaluation more transparent, let us consider a simple numerical example for two simple systems.

Illustrative example.

Consider two systems with the following parameters: u_{11}, u_{21}, u_{31} and u_{41} for the first system and u_{12}, u_{22}, u_{32} and u_{42} for the second system. System have the same “total importance”, i.e.: $u_{11} + u_{21} + u_{31} + u_{41} = u_{12} + u_{22} + u_{32} + u_{42}$, and the “hostile impacts are of the same intensity. Compile the table for comparison of survivability of these systems subjected the same impacts.

Table 1. Matrices of solution for hostile impacts on two differently protected systems.

The first system					The second system						
Level of protection		Intensity of impacts				Level of protection		Intensity of impacts			
		0.5	0.65	0.8	0.95			0.5	0.65	0.8	0.95
0.3	5	5	5	5	0.6	0	7	7	7		
0.6	0	7	7	7	0.6	0	8	8	8		
0.7	0	0	11	11	0.7	0	0	11	11		
0.9	0	0	0	15	0.6	0	12	12	12		

In the table the cells with chosen units and corresponding impacts are highlighted with grey. On the basis of this table, one can make some qualitative conclusions. In spite of “more reasonable” location of protection resources in the first system (the more important unit, the better protection), the total loss after a hostile attack is 38 conditional units. At the same time, the second system with the same “total importance” and with even allocation of protection resources has the total loss only 31 conditional units.

Let us consider the second situation: the same total hostile intensity of impacts is distributed more or less evenly. (It can occur, for instance, if terrorists do not know real importance of the system’s units or if they do not know the level of their protection.) This situation is reflected in Table 2.

Table 2. Matrices of solution for hostile impacts on two differently protected systems for even intensities.

The first system					The second system						
Level of protection	Intensity of impacts	Intensity of impacts				Level of protection	Intensity of impacts	Intensity of impacts			
		0.75	0.75	0.75	0.75			0.75	0.75	0.75	0.75
0.3	5	5	5	5	0.6	7	7	7	7		
0.6	7	7	7	7	0.6	8	8	8	8		
0.7	11	11	11	11	0.7	11	11	11	11		
0.9	0	0	0	0	0.6	12	12	12	12		

In this case, the first system is better protected against hostile strike. By the way, disinformation about importance of units and/or levels of protection can help for defending site.

11.1 Brief historical overview and related sources

One of the first work dedicated the problem of survivability was [Krylov, 1942]. Almost 40years later the problem of survivability arose in connection of volnurability of large scale energy systems [Rudenko and Ushakov, 1979, 1989]. This problem broke out last years in connection with organized terrorist activity [Ushakov, 2005, 2006; Lefvitin, et al. 2007-2010].

Bibliography below is given in chronological-alphabetical ordering for better ecposition of historical background of the subject.

Bibliography

- Krylov, A.N. (1942) The Ship Theory. Part 1: Ship Stability. Voenmorizdat.
- Rudenko, Yu.N., and I.A. Ushakov (1979) On Survivability of Complex Energy Systems. *Energy Systems and Transportation*", № 1. Moscow, USSR Ac.Sci.
- Kozlov, M., Yu. Malashenko, V. Rogozhin, I. Ushakov, and T. Ushakova. (1986) Survivability modeling energy systems: Methodology, model, implementation. Moscow, Computer Center Russian Ac. Sci.
- Rudenko, Yu.N., and I.A. Ushakov (1989). Reliability of Power Systems (Russian). Novisibirsk, Nauka.
- Levitin, G. (2004) Protection survivability importance in systems with multilevel protection. *Quality and Reliability Engineering International*, No. 20.
- Ushakov, I., (2005). Cost-effective approach to counter-terrorism. *Communication in Dependability and Quality Management*. Vol.8, No.3.
- Ushakov, I. (2006) Counterterrorism: Protection Resources Allocation. Reliability: Theory & Applications (No 2, 3, 4), Gnedenko Forum.
- Bochkov, A., and I. Ushakov (2007). Sensitivity analysis of optimal counter-terrorism resources allocation under subjective expert estimates. Reliability: Theory & Applications, Vol.2, No. 2. Gnedenko Forum.
- Korczak, E., and G. Levitin . (2007) Survivability of systems under multiple factor impact, *Reliability Engineering & System Safety*, 92(2).
- Levitin, G., H. Ben Haim. (2008). Importance of protections against intentional attacks. *Reliability Engineering & System Safety*, 93 (4).

- Van Gelder, P. (2008) Methods for risk analysis in disaster reduction. RTA Journal, Vol.3, No.2. Gnedenko Forum.
- Levitin, G. (2009) System survivability and defense against external impacts. Guest editorial. G. *International Journal of Performability Engineering*, 5(1).
- Levitin, G., H. Ben Haim, (2009). Minmax defense strategy for complex multi-state systems.. *Reliability Engineering & System Safety*.
- Levitin, G., K. Hausken. (2010). Defense and attack of systems with variable attacker system structure detection probability. *Journal of the Operational Research Society* 61.

12 Multistate systems

12.1 Preliminary notes

Reliability analysis of multistate systems has a long enough history. First papers dedicated to this subject appeared as early as in 1978 [R. Barlow, et al., 1978; E. El-Newehi, et al. 1978]. Later appears several papers with introduction of a new technique for multi-system analysis [I. Ushakov, 1986, 1988, 1998]. And finally it was a real burst of research papers on the subject [A. Lisnianski, et al., 2003; G. Levitin, et al., 2003; G. Levitin, 2004; G. Levitin, 2005].

We begin analysis of explanation of a new technique introduced in [I. Ushakov, 1986].

12.2 Generating Function

In spite of constant attempting to make presentation “transparent” and very much “physical”, this time we begin with rather abstract statement of fundamental principles, on which Universal Generating Function is based.

Everybody knows about the generating function (GF) that is also called the discrete Laplace transform, or z -transform. Generating function is widely used in probability theory, especially, for finding convolutions of discrete distributions.

Generating Function, $\varphi(z)$, for positive random variable X is defined as polynomial

$$\varphi(z) = \sum_{1 \leq k < \infty} p_k z^{x_k} . \quad (1)$$

Power of z denotes the value of r.v., and coefficient at each term equals to the probability of realization random variable x_k of the random variable X . For instance, for binomial distribution

$$\varphi(z) = \sum_{1 \leq k < n} \binom{n}{k} q^k p^{n-k} z^k = (p + qz)^n \quad (2)$$

As one knows, binomial distribution, corresponding to (2), characterizes, in particular, a number of failures during testing of n independent and identical items. If one takes two samples of different sizes, say, n_1 and n_2 , from the same general set of events, the GF for such composition will be obtained as product

$$\varphi_{\Sigma}(z) = \left[\sum_{1 \leq k < n_1} \binom{n_1}{k} q^k p^{n_1-k} z^k \right] \cdot \left[\sum_{1 \leq j < n_2} \binom{n_2}{j} q^j p^{n_2-j} z^j \right] \quad (3)$$

Let us find the probability that in combined sample with $n_1 = 3$ and $n_2 = 5$ there will have occurred exactly $s = k + j = 3$ failures. For this purpose, compile auxiliary table of possible outcomes.

Table 1. Terms which multiplying leads to $s = 3$.

k	p_k	j	p_j	s	product
0	$\binom{3}{0} p^3$	3	$\binom{5}{3} q^3 p^2$	5	$\binom{3}{0} \cdot \binom{5}{3} q^3 p^5$
1	$\binom{3}{1} q p^2$	2	$\binom{5}{2} q^2 p^3$	5	$\binom{3}{1} \cdot \binom{5}{2} q^3 p^5$
2	$\binom{3}{2} q^2 p$	1	$\binom{5}{1} q p^4$	5	$\binom{3}{2} \cdot \binom{5}{1} q^3 p^5$
3	$\binom{3}{3} q^3$	0	$\binom{5}{0} p^5$	5	$\binom{3}{3} \cdot \binom{5}{0} q^3 p^5$
Total probability					$q^3 p^5 \sum_{0 \leq k \leq 3} \binom{3}{k} \cdot \binom{5}{3-k}$

Table 1. Terms which multiplying leads to $s = 3$.

k	p_k	j	p_j	s	product
0	$\binom{3}{0} p^3$	3	$\binom{5}{3} q^3 p^2$	5	$\binom{3}{0} \cdot \binom{5}{3} q^3 p^5$
1	$\binom{3}{1} q p^2$	2	$\binom{5}{2} q^2 p^3$	5	$\binom{3}{1} \cdot \binom{5}{2} q^3 p^5$
2	$\binom{3}{2} q^2 p$	1	$\binom{5}{1} q p^4$	5	$\binom{3}{2} \cdot \binom{5}{1} q^3 p^5$
3	$\binom{3}{3} q^3$	0	$\binom{5}{0} p^5$	5	$\binom{3}{3} \cdot \binom{5}{0} q^3 p^5$
Total probability					$q^3 p^5 \sum_{0 \leq k \leq 3} \binom{3}{k} \cdot \binom{5}{3-k}$

Theory of various transforms over binomial coefficients is very well developed. In particular, there is known the so-called Vandermond²⁸ convolution that in our case has the form:

²⁸ **Alexandre-Théophile Vandermonde** (1735 – 1796) was a French musician, mathematician and chemist. His name is now principally associated with determinant theory in mathematics.

$$\sum_{0 \leq k \leq 3} \binom{3}{k} \cdot \binom{5}{5-k} = \binom{8}{3} \quad (4)$$

So, the term $\binom{8}{3} q^3 p^5$ is actually the sixth one in polynomial expansion

$$\begin{aligned} \varphi_{\Sigma}(z) &= (p + qz)^3 (p + qz)^5 = (p + qz)^8 = \\ &= p^8 + \binom{8}{1} p^7 qz + \binom{8}{2} p^6 q^2 z^2 + \binom{8}{3} p^5 q^3 z^3 + \dots + \binom{8}{7} p q^7 z^7 + q^8 z^8. \end{aligned} \quad (5)$$

In a sense, this result was obvious from the beginning: performing of series of two binomial tests of volumes a and b is equivalent to a single test of volume $a + b$.

Of course similar deductions can be done with other discrete distributions. However, not this is the final target of our discussion.

Working with z-transforms “by hand” one uses polynomials because it is convenient to multiply coefficients (probabilities) and add powers at z . However, assume that we decided to write a program for computer. What we will do in this case?

We present the first polynomial $\varphi_1(z)$ as a set of pairs $\{(p_{11}, a_{11}), (p_{12}, a_{12}), \dots, (p_{15}, a_{15})\}$ and the second polynomial $\varphi_2(z)$ as a set of pairs $\{(p_{21}, a_{21}), (p_{22}, a_{22}), \dots, (p_{25}, a_{25})\}$ where p_{jk} 's are corresponding coefficients and a_{jk} 's are corresponding powers of polynomials in unfolded form. Then we arrange Descartes product of these two sets:

Table 2. Descartes product of two sets.

	(p_{11}, a_{11})	(p_{12}, a_{12})	...	(p_{15}, a_{15})
(p_{21}, a_{21})	(P_{11}, A_{11})	(P_{21}, A_{21})	...	(P_{51}, A_{51})
(p_{22}, a_{22})	(P_{12}, A_{12})	(P_{22}, A_{22})	...	(P_{25}, A_{25})
...
(p_{27}, a_{27})	(P_{17}, A_{17})	(P_{27}, A_{27})	...	(P_{27}, A_{27})

Thus, we have some kind of interaction. Here P_{jk} is found as $P_{jk} = p_{j1} \times p_{2k}$ and A_{jk} is found as $A_{jk} = a_{j1} + a_{2k}$.

However, for a computer there is no difference what kind of operations to perform over the first and the second terms of the considered pairs. This idea has

been put in the basis of introducing the so-called Universal Generating Function (UGF).

12.3 Universal Generating Function

We will present UGF only for reliability problems, so we restricted ourselves by units characterized by two parameters: probability of unit's particular state and by value of operational parameter associated with this state. Associated parameter can be a value of any system's outcome: productivity rate, capacity, resistance, inductivity, etc.

In this case we can keep (just for convenience of using habitual presentation) a polynomial form of specific type: powers of products of two terms, say, z_a and z_b will be presented by some transforms over powers of individual terms, namely:

$$p_a z_a^a \otimes_f p_b z_b^b = p_a p_b z^{f(a,b)} \quad (6)$$

where f is an arbitrary given function.

For further discussion, it will be more convenient to use the following form of presentation of (6):

$$\{p_a, a\} \otimes_f \{p_b, b\} = \{p_a p_b, f(a, b)\}. \quad (7)$$

Naturally, composition operator \otimes_f possesses commutativity property, i.e.

$$\otimes_f (a, b) = \otimes_f (b, a) \quad (8)$$

and associativity property, i.e.

$$\otimes_f (a, b, c) = \otimes_f (a \otimes_f (b, c)) = \otimes_f ((a \otimes_f b), c). \quad (9)$$

if the function $f(a,b)$ possesses these properties. In the most applications this is the case, though numerous exceptions exist (see, for example, G. Levitin, 2005).

To avoid terminological confusion, let us call $\otimes_f (a, b, c)$ interaction of variables a, b and c .

Let us now return back to multi-state systems.

Assume that unit k is characterized by the following discrete distribution of its operational parameter X_k : $P\{X_k = x_{kj}\} = p_{kj}$. Then we can characterize the distribution of the operational parameter of unit k with the following vector of pairs:

$$\mathbf{Q}_k = \{(p_{k1}, x_{k1}), (p_{k2}, x_{k2}), \dots, (p_{ks(k)}, x_{ks(k)})\} = \{(p_{kj}, x_{kj}), 1 \leq j \leq s(k)\}$$

where $s(k)$ is the number of different values of r.v. X_k .

Interaction of operational parameters of two units X_k and X_i can be written as

$$\begin{aligned} \mathbf{Q}_k \otimes_f \mathbf{Q}_i &= \{(p_{kj}, x_{kj}), 1 \leq j \leq s(k)\} \otimes_f \{(p_{il}, x_{il}), 1 \leq l \leq s(i)\} = \\ &= (p_{kj} \times p_{il}; f(x_{kj}, x_{il}), j = \overline{1, s(k)}, l = \overline{1, s(i)}) \end{aligned}$$

Interaction of operational parameters of N units can be written as

$$\otimes_f (\mathbf{Q}_1, \dots, \mathbf{Q}_k, \dots, \mathbf{Q}_N) = \left(\prod_{i=1}^N p_{i, m(i)}, f(x_{1, m(1)}, \dots, x_{N, m(N)}) \right) \quad (10)$$

for all combinations of $m(i)$, when $1 \leq m(i) \leq s(i)$.

In expression (10) there could be pairs with the same values of operational parameters, for instance, (P_1, A) , (P_2, A) , .. , (P_n, A) , then the resulting expression (10) has to be changed by collecting terms:

$$(P_1, A), (P_2, A), \dots, (P_n, A) = \left(\sum_{1 \leq k \leq n} P_k, A \right)$$

Let us consider several simple examples for demonstration of the use of UGF.

12.4 Multistate series system

For demonstrsation how UGF works, let us consider several simple numerical examples possessing a transparent physical sense.

12.4.1 Series connection of piping runs

Consider a simple oil pipe line (Figure 1) consisting of four runs of piping that we will call below as units (just for convenience).

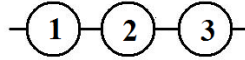


Figure 12.1. Conditional structure of oil pipe line with four series piping runs.

Piping run throughput, v , changes randomly due to various external and internal causes. Let pipeline units are characterized by the following distributions of pipe throughputs: throughput

First unit:

$$p_{11} = \Pr\{v = 100\} = 0.8; \quad p_{12} = \Pr\{v = 90\} = 0.15; \quad p_{13} = \Pr\{v = 80\} = 0.05$$

$$p_{14} = \Pr\{\text{complete failure}\} \approx 0.$$

Second unit:

$$p_{21} = \Pr\{v = 100\} = 0.9; \quad p_{22} = \Pr\{v = 90\} = 0.07; \quad p_{23} = \Pr\{v = 80\} = 0.03$$

$$p_{24} = \Pr\{\text{complete failure}\} \approx 0..$$

Third unit's has the same distribution as the second one.

Remark. Pipe run throughput is taken in some conditional units.

The entire pipe line is characterized by minimum current value of its units' throughputs, i.e. we have to use the operator \otimes_{\min} because

$$f_{SERIES}^{(v)}(v_k, v_j) = \min(v_k, v_j),$$

Let us consider the following recurrent procedure.

Step 1. First, consider interaction of parameters of units 1 and 2. Take a Descartes product presenting it in the form of the following table.

			Unit 1			
			State 1	State 2	State 3	State 4
			(0.8; 100)	(0.15; 90)	(0.05; 80)	(0; 0)
Unit 2	State 1	(0.9; 100)	$0.8 \cdot 0.9 = 0.72$ $\min(100, 100) = 100$	$0.15 \cdot 0.9 = 0.135$ $\min(90, 100) = 90$	$0.05 \cdot 0.9 = 0.045$ $\min(80, 100) = 80$	(0; 0)
	State 2	(0.07; 90)	$0.8 \cdot 0.07 = 0.056$ $\min(100, 90) = 90$	$0.15 \cdot 0.07 \approx 0.011$ $\min(90, 90) = 90$	$0.05 \cdot 0.07 \approx 0.004$ $\min(80, 90) = 80$	(0; 0)
	State 3	(0.03; 80)	$0.8 \cdot 0.03 = 0.024$ $\min(100, 80) = 80$	$0.15 \cdot 0.03 \approx 0.005$ $\min(90, 80) = 80$	$0.05 \cdot 0.03 \approx 0.002$ $\min(80, 80) = 80$	(0; 0)
	State 4	(0; 0)	(0; 0)	(0; 0)	(0; 0)	(0; 0)

In result, one obtains a new equivalent unit with the following distribution of the throughputs:

$$P_1 = \Pr\{v=100\} = 0.72;$$

$$P_2 = \Pr\{v=90\} = 0.056 + 0.135 + 0.011 = 0.202;$$

$$P_3 = \Pr\{v=80\} = 0.024 + 0.005 + 0.002 + 0.004 + 0.045 = 0.082;$$

$$P_4 = \Pr\{v=0\} = 0.$$

Remark: Here and below the sum of all probabilities is not equal exactly to 1 due to rounding of results of multiplications of corresponding probabilities.

Step 2. This new equivalent unit has to be combined with the third unit (see the table below).

			Equivalent unit			
			State 1	State 2	State 3	State 4
			(0.72; 100)	(0.202; 90)	(0.082; 80)	(0; 0)
Unit 3	State 1	(0.9; 100)	$0.72 \cdot 0.9 = 0.648$ $\min(100, 100) = 100$	$0.202 \cdot 0.9 = 0.182$ $\min(90, 100) = 90$	$0.082 \cdot 0.9 = 0.074$ $\min(80, 100) = 80$	(0; 0)
	State 2	(0.07; 90)	$0.72 \cdot 0.07 = 0.050$ $\min(100, 90) = 90$	$0.202 \cdot 0.07 \approx 0.014$ $\min(90, 90) = 90$	$0.082 \cdot 0.07 \approx 0.006$ $\min(80, 90) = 80$	(0; 0)
	State 3	(0.03; 80)	$0.72 \cdot 0.03 = 0.022$ $\min(100, 80) = 80$	$0.202 \cdot 0.03 \approx 0.006$ $\min(90, 80) = 80$	$0.082 \cdot 0.03 \approx 0.002$ $\min(80, 90) = 80$	(0; 0)
	State 4	(0; 0)	(0; 0)	(0; 0)	(0; 0)	(0; 0)

These results allow calculating the expected throughput of the pipe line, $E\{V\}$:

$$E\{V\} = 0.648 \cdot 100 + (0.050 + 0.014 + 0.182) \cdot 90 + (0.022 + 0.006 + 0.074 + 0.006 + 0.002) \cdot 80 = 0.648 \cdot 100 + 0.246 \cdot 90 + 0.111 \cdot 80 \approx 87.8.$$

One also can find the PFFO of this system by some chosen criterion of failure. For instance, if a failure criterion is $V < 90$, then PFFO is equal to:

$$P = \Pr\{V \geq 90\} = 0.648 + 0.246 = 0.894.$$

12.4.2 Series connection of resistors

Consider a simple chain of ohmic resistors with the structure analogous that presented in Figure 1. Resistance of each unit, ρ , can changes randomly due to

various environmental conditions or internal causes. Let resistors are characterized by the following distributions of resistance:

First unit:

$$p_{11} = \Pr\{\rho = 10 \text{ ohms}\} = 0.8; \quad p_{12} = \Pr\{\rho = 9 \text{ ohms}\} = 0.15; \quad p_{13} = \Pr\{\rho = 8 \text{ ohms}\} = 0.05$$

$$p_{14} = \Pr\{\text{complete failure}\} = 0.$$

Second unit:

$$p_{21} = \Pr\{\rho = 10 \text{ ohms}\} = 0.9; \quad p_{22} = \Pr\{\rho = 9 \text{ ohms}\} = 0.07; \quad p_{23} = \Pr\{\rho = 8 \text{ ohms}\} = 0.3$$

$$p_{24} = \Pr\{\text{complete failure}\} = 0.$$

Third unit has the same distribution as the second one.

The entire series connection of resistors is characterized by sum of its units' resistances, i.e. we have to use the operator \otimes because in this case

$$f_{SERIES}^{(\rho)}(\rho_k, \rho_j) = \rho_k + \rho_j.$$

Let us consider the following recurrent procedure.

Step 1. First, consider interaction of parameters of units 1 and 2. Take again a Descartes product presenting it in the form of the following table.

			Unit 1		
			State 1	State 2	State 3
			(0.8; 10 ohms)	(0.15; 9 ohms)	(0.05; 8 ohms)
Unit 2	State 1	(0.9; 10 ohms)	0.8·0.9 = 0.72 10+10=20	0.15·0.9=0.135 9+10=19	0.05·0.9=0.045 8+10=18
	State 2	(0.07; 9 ohms)	0.8·0.07=0.056 10+9=19	0.15·0.07≈0.011 9+9=18	0.05·0.07≈0.004 8+9=17
	State 3	(0.03; 8 ohms)	0.8·0.03=0.024 10+8=18	0.15·0.03≈0.005 9+8=17	0.05·0.03≈0.002 8+8=16

In result, one obtains a new equivalent unit with the following distribution of resistancy:

$$P_1 = \Pr\{\rho = 20 \text{ ohms}\} = 0.72;$$

$$P_2 = \Pr\{\rho = 19 \text{ ohms}\} = 0.056 + 0.135 = 0.191;$$

$$P_3 = \Pr\{\rho = 18 \text{ ohms}\} = 0.024 + 0.011 + 0.045 = 0.09;$$

$$P_4 = \Pr\{\rho = 17 \text{ ohms}\} = 0.005 + 0.004 = 0.009;$$

$$P_5 = \Pr\{\rho = 16 \text{ ohms}\} = 0.002.$$

Step 2. This new equivalent unit has to be “converged” with the third unit (see the table below).

		Unit 3			
		State 1	State 2	State 3	
		(0.9; 10 ohms)	(0.07; 9 ohms)	(0.03; 8 ohms)	
Equivalent unit	State 1	(0.72; 20 ohms)	$0.9 \cdot 0.72 = 0.648$ $10+20=30$	$0.07 \cdot 0.72 \approx 0.005$ $9+20=29$	$0.03 \cdot 0.72 = 0.022$ $8+20=28$
	State 2	(0.191; 19 ohms)	$0.9 \cdot 0.191 = 0.172$ $10+19=29$	$0.07 \cdot 0.191 = 0.013$ $9+19=28$	$0.03 \cdot 0.191 = 0.006$ $8+19=27$
	State 3	(0.09; 18 ohms)	$0.9 \cdot 0.09 = 0.081$ $10+18=28$	$0.07 \cdot 0.09 \approx 0.006$ $9+18=27$	$0.03 \cdot 0.09 = 0.003$ $8+18=26$
	State 4	(0.009; 17 ohms)	$0.9 \cdot 0.009 \approx 0.008$ $10+17=27$	$0.07 \cdot 0.009 \approx 0$ $9+17=26$	$0.03 \cdot 0.009 \approx 0$ $8+17=25$
	State 5	(0.002; 16 ohms)	$0.9 \cdot 0.002 \approx 0.002$ $10+16=26$	$0.07 \cdot 0.002 \approx 0$ $9+16=25$	$0.03 \cdot 0.002 \approx 0$ $8+16=24$

Remark. By the way, this case can be analyzed with a standard GF. We demonstrate it here just for some logical completeness.

These results allow calculating the expected resistance of the series connection of resistors:

$$E\{\rho\} = 0.648 \cdot 30 + (0.172 + 0.005) \cdot 29 + (0.081 + 0.013 + 0.022) \cdot 28 + (0.008 + 0.006 + 0.006) \cdot 27 + (0.002 + 0.003) \cdot 26 = 0.648 \cdot 30 + 0.177 \cdot 29 + 0.116 \cdot 28 + 0.02 \cdot 27 + 0.005 \cdot 26 \approx 28.49 \text{ ohms}.$$

One also can find the PFFO of this system by some chosen criterion of failure. For instance, if a failure criterion is $\rho < 27 \text{ ohms}$, then PFFO is equal to:

$$P = \Pr\{\rho < 28 \text{ ohms}\} = 0.648 + 0.177 + 0.116 = 0.941.$$

12.4.3 Series connections of capacitors

Consider a simple chain of ohmic resistors with the structure analogous that presented in Figure 1. Resistance of each unit, c , can changes randomly due to various environmental conditions or internal causes. Let resistors are characterized by the following distributions of resistance:

First unit:

$$p_{11} = \Pr\{c = 10 \mu F\} = 0.8;$$

$$p_{12} = \Pr\{c = 9 \mu F\} = 0.15;$$

$$p_{13} = \Pr\{c = 8 \mu F\} = 0.05$$

$$p_{14} = \Pr\{\text{complete failure}\} = 0.$$

Second unit:

$$p_{21} = \Pr\{c = 10 \mu F\} = 0.9;$$

$$p_{22} = \Pr\{c = 9 \mu F\} = 0.07;$$

$$p_{23} = \Pr\{c = 8 \mu F\} = 0.3$$

$$p_{24} = \Pr\{\text{complete failure}\} = 0..$$

Third unit's has the same distrigution as the second one.

The entire series connection of capacitors is characterized by sum of inverse values of its units' capacities, i.e. we have to use the operator \otimes_f , where

$$f_{SERIES}^{(c)}(c_k, c_j) = (c_k^{-1} + c_j^{-1})^{-1} = \left(\frac{1}{c_k} + \frac{1}{c_j} \right)^{-1} = \frac{c_k \cdot c_j}{c_k + c_j}.$$

Let us consider the following recurrent procedure.

Step 1. First, consider interaction of parameters of units 1 and 2. Take again a Descartes product presenting it in the form of the following table.

			Unit 1		
			State 1	State 2	State 3
			(0.8; 10 μF)	(0.15; 9 μF)	(0.05; 8 μF)
Unit 2	State 1	(0.9; 10 μF)	$0.8 \cdot 0.9 = 0.72$ $(10^{-1} + 10^{-1})^{-1} = 5$	$0.15 \cdot 0.9 = 0.135$ $(9^{-1} + 10^{-1})^{-1} \approx 4.74$	$0.05 \cdot 0.9 = 0.045$ $(8^{-1} + 10^{-1})^{-1} \approx 4.44$
	State 2	(0.07; 9 μF)	$0.8 \cdot 0.07 = 0.056$ $(10^{-1} + 9^{-1})^{-1} \approx 4.74$	$0.15 \cdot 0.07 \approx 0.011$ $(9^{-1} + 9^{-1})^{-1} = 4.5$	$0.05 \cdot 0.07 \approx 0.004$ $(8^{-1} + 9^{-1})^{-1} \approx 4.24$
	State 3	(0.03; 8 μF)	$0.8 \cdot 0.03 = 0.024$ $(10^{-1} + 8^{-1})^{-1} \approx 4.44$	$0.15 \cdot 0.03 \approx 0.005$ $(9^{-1} + 8^{-1})^{-1} \approx 4.24$	$0.05 \cdot 0.03 \approx 0.002$ $(8^{-1} + 8^{-1})^{-1} = 4$

In result, one obtains a new equivalent unit with the following distribution of capacity:

$$P_1 = \Pr\{c = 5 \mu F\} = 0.72;$$

$$P_2 = \Pr\{c = 4.74 \mu F\} = 0.056 + 0.135 = 0.191;$$

$$P_3 = \Pr\{c = 4.5 \mu F\} = 0.011;$$

$$P_4 = \Pr\{c = 4.44 \mu F\} = 0.024 + 0.045 = 0.069;$$

$$P_5 = \Pr\{c = 4.24 \mu F\} = 0.005 + 0.004 = 0.009;$$

$$P_6 = \Pr\{c = 4 \mu F\} = 0.002.$$

Step 2. This new equivalent unit has to be “converged” with the third unit (see the table below).

			Unit 3		
			State 1	State 2	State 3
			(0.9; 10 μF)	(0.07; 9 μF)	(0.03; 8 μF)
Equivalent unit	State 1	(0.72; 5 μF)	$0.9 \cdot 0.72 = 0.648$ $(10^{-1} + 5^{-1})^{-1} \approx 3.33$	$0.07 \cdot 0.72 \approx 0.005$ $(9^{-1} + 5^{-1})^{-1} \approx 3.21$	$0.03 \cdot 0.72 \approx 0.022$ $(8^{-1} + 5^{-1})^{-1} \approx 3.08$
	State 2	(0.191; 4.74 μF)	$0.9 \cdot 0.191 = 0.172$ $(10^{-1} + 4.74^{-1})^{-1} \approx 3.22$	$0.07 \cdot 0.191 \approx 0.013$ $(9^{-1} + 4.74^{-1})^{-1} \approx 3.10$	$0.03 \cdot 0.191 \approx 0.006$ $(8^{-1} + 4.74^{-1})^{-1} \approx 2.98$
	State 3	(0.011; 4.5 μF)	$0.9 \cdot 0.011 \approx 0.01$ $(10^{-1} + 4.5^{-1})^{-1} \approx 3.10$	$0.07 \cdot 0.011 \approx 0.006$ $(9^{-1} + 4.5^{-1})^{-1} = 3$	$0.03 \cdot 0.011 \approx 0.003$ $(8^{-1} + 4.5^{-1})^{-1} \approx 2.88$
	State 4	(0.069; 4.44 μF)	$0.9 \cdot 0.069 \approx 0.062$ $(10^{-1} + 4.44^{-1})^{-1} \approx 3.07$	$0.07 \cdot 0.069 \approx 0$ $(9^{-1} + 4.44^{-1})^{-1} \approx 2.97$	$0.03 \cdot 0.069 \approx 0.002$ $(8^{-1} + 4.44^{-1})^{-1} \approx 2.86$
	State 5	(0.009; 4.24 μF)	$0.9 \cdot 0.009 \approx 0.008$ $(10^{-1} + 4.24^{-1})^{-1} \approx 2.98$	$0.07 \cdot 0.009 \approx 0$ $(9^{-1} + 4.24^{-1})^{-1} \approx 2.88$	$0.03 \cdot 0.009 \approx 0$ $(8^{-1} + 4.24^{-1})^{-1} \approx 2.77$
	State 6	(0.002; 4 μF)	$0.9 \cdot 0.002 \approx 0.002$ $(10^{-1} + 4^{-1})^{-1} \approx 2.86$	$0.07 \cdot 0.002 \approx 0$ $(9^{-1} + 4^{-1})^{-1} \approx 2.77$	$0.03 \cdot 0.002 \approx 0$ $(8^{-1} + 4^{-1})^{-1} \approx 2.67$

These results allow calculating the expected capacity of the series connection of capacitors:

$$E\{c\} \approx 0.648 \cdot 3.33 + 0.172 \cdot 3.22 + 0.005 \cdot 3.21 + (0.01 + 0.013) \cdot 3.1 + 0.022 \cdot 3.08 + 0.062 \cdot 3.07 + 0.006 \cdot 3 + (0.008 + 0.006) \cdot 2.98 + 0.003 \cdot 2.88 + (0.002 + 0.002) \cdot 2.86 \approx 3.14.$$

One also can find the PFFO of this system by some chosen criterion of failure. For instance, if a failure criterion is $c < 3 \mu\text{F}$, then PFFO is equal to:

$$P = \Pr\{c < 3 \mu\text{F}\} = 0.648 + 0.172 + 0.005 + 0.023 + 0.022 + 0.062 + 0.006 = 0.938.$$

12.5 Multistate parallel system

12.5.1 Parallel connection of piping runs

Consider a section of oil pipeline with four parallel piping runs (Figure 12.2).

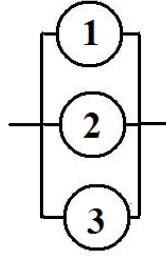


Figure 12.2. Conditional structure of oil pipeline with three parallel piping runs.

The entire pipe line is characterized by sum of current value of its units' throughputs, v , i.e. we have to use the operator \otimes because

$$f_{PARALLEL}^{(v)}(v_1, v_2) = v_1 + v_2.$$

It means that from methodological point of view calculations are coinciding with those in Section 12.4.2.

12.5.2 Parallel connection of resistors

Consider a parallel connection of ohmic resistors (see Figure 14.2). For parallel connections of two resistors, one should use the following function that determines the resistance of the pair of resistors:

$$f_{PARALLEL}^{(\rho)}(\rho_1, \rho_2) = \frac{\rho_1 + \rho_2}{\rho_1 \cdot \rho_2} \quad (13)$$

and the corresponding operator \otimes .

Formally this mathematical model coincides with that described in Section 12.4.3. One can use all numerical solutions from there with corresponding change of dimension.

12.5.3 Parallel connections of capacitors

Consider a parallel connection of electrical capacitors (see Figure 2). For parallel connections of two capacitors, one can write:

$$f_{PARALLEL}^{(c)}(c_1, c_2) = c_1 + c_2. \quad (14)$$

and use the operator \otimes .

Formally this type of manipulative interaction corresponds to that in Section 12.4.1, so one can use all numerical solutions from that section with corresponding change of dimension.

12.6 Reducible systems

Above there were considered only simple series and parallel structures. Naturally, UGF method can be with the same success applied to reducible systems in general. Since routine transforms in this case are very much similar to those described above, we will present only principal new ideas demonstrating them on a simple particular example.

Let a considered system has the structure presented in Figure 12.3.

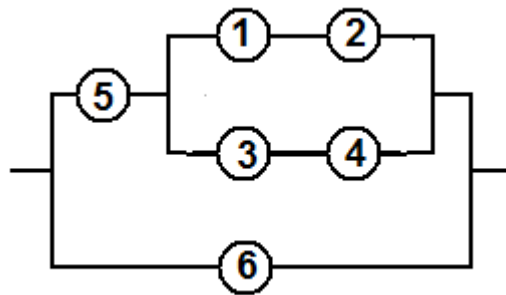


Figure 12.3. Example of reducible structure.

Compiling UGF for such structure is performed in accordance with the step-by-step reducing the initial structure by aggregating pairs of its modules (elements) into single equivalent modules. In other words, the structure is reducing sequentially as it is presented in Figure 12.4.

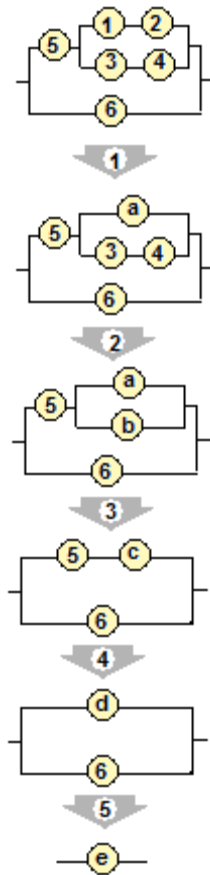


Figure 12.4. Step-by-step reduction of the structure depicted in Figure 12.43.

Notice that in this case one has to use two types of \otimes_f operators with functions f corresponding to series and to parallel connection of structural modules.. We will illustrate the methodology on two simple illustrative examples of a pipeline.

Example 12.1.

Consider a pipe line with a simple reducible structure presented in Figure 12.5.

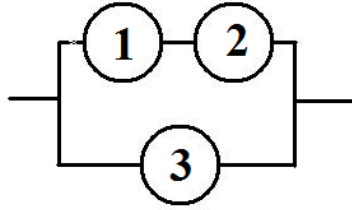


Figure 12.5. Structure of pipeline considered in Example 12.1.

Let the system consists of identical units with pipe troughput distribution of each units equals to:

$$\begin{aligned}
 p_{k1} &= \Pr\{v = 100\} = 0.8; \\
 p_{k2} &= \Pr\{v = 90\} = 0.15; \\
 p_{k3} &= \Pr\{v = 80\} = 0.05 \\
 p_{k4} &= \Pr\{\text{complete failure}\} = 0.
 \end{aligned}$$

Remark. Pipe run throughout is taken in some conditional units.

Step 1. Consider first the upper brunch of the pipeline consisting of series connection of units 1 and 2.

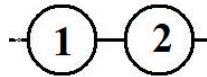


Figure 12.6. Structure for Step 1.

For this step we have to use the operator \otimes_{\min} , since

$$f_{SERIES}^{(v)}(v_k, v_j) = \min(v_k, v_j).$$

Results of calculations are presented in the table below.

			Unit 1			
			State 1 (0.8; 100)	State 2 (0.15; 90)	State 3 (0.05; 80)	State 4 (0; 0)
Unit 2	State 1 (0.8; 100)	0.8 · 0.8 = 0.64 min(100,100)=100	0.15 · 0.8=0.12 min(90,100)=90	0.05 · 0.8=0.04 min(80,100)=80	(0; 0)	
	State 2 (0.15; 90)	0.8 · 0.15 = 0.12 min(100,90)=90	0.15 · 0.15=0.0225 min(90,90)=90	0.05 · 0.15=0.0075 min(80,90)=80	(0; 0)	
	State 3 (0.05; 80)	0.8 · 0.05 = 0.04 min(100,80)=80	0.15 · 0.05=0.0075 min(90,80)=80	0.05 · 0.05=0.0025 min(80,80)=80	(0; 0)	
	State 4 (0; 0)	(0; 0)	(0; 0)	(0; 0)	(0; 0)	

In result, one obtains a new equivalent unit with the following distribution of troughputs:

$$P_1 = \Pr\{v=100\} = 0.64;$$

$$P_2 = \Pr\{v=90\} = 0.12 + 0.12 + 0.0225 = 0.2625;$$

$$P_3 = \Pr\{v=80\} = 0.04 + 0.0075 + 0.04 + 0.0075 + 0.0025 = 0.0975;$$

$$P_4 = \Pr\{v=0\} = 0.$$

Step 2. This new equivalent unit has to be combined in parallel with the third unit.

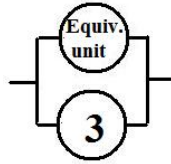


Figure 12.7. Structure for Step 2.

In this case we have to use the operator \otimes because

$$f_{PARALLEL}^{(v)}(v_k, v_j) = v_k + v_j.$$

Results of calculations are presented in the table below.

			Equivalent unit			
			State 1	State 2	State 3	State 4
			(0.64; 100)	(0.2625; 90)	(0.0975; 80)	(0; 0)
Unit 3	State 1	(0.8; 100)	0.64 · 0.8 = 0.512 100+100=200	0.2625 · 0.8=0.21 90+100=190	0.0975 · 0.8=0.078 80+100=180	0 100
	State 2	(0.15; 90)	0.64 · 0.15=0.096 100+90=190	0.2625 · 0.15≈0.039 90+90=180	0.0975 · 0.15≈0.015 80+90=170	0 90
	State 3	(0.05; 80)	0.64 · 0.05=0.032 100+80=180	0.2625 · 0.05≈0.013 90+80=170	0.0975 · 0.05≈0.005 80+80=160	0 80
	State 4	(0; 0)	0 100	0 90	0 80	0 0

Result in pipeline troughput distribution is:

$$P_1 = \Pr\{v=200\} = 0.512;$$

$$P_2 = \Pr\{v=190\} = 0.096 + 0.21 = 0.306;$$

$$P_3 = \Pr\{v=180\} = 0.032 + 0.039 + 0.078 = 0.149;$$

$$P_4 = \Pr\{v=170\} = 0.013 + 0.015 = 0.028;$$

$$P_5 = \Pr\{v=160\} = 0.005$$

These results allow calculating the average power of the pipe line:

$$V_{\text{average}} = 0.512 \cdot 200 + 0.306 \cdot 190 + 0.149 \cdot 180 + 0.028 \cdot 170 + 0.005 \cdot 160 \approx 192.9.$$

One also can find the PFFO of this system by some chosen criterion of failure. For instance, if a failure criterion is $V < 190$, then PFFO is equal to:

$$P = \Pr\{V \geq 190\} = 0.512 + 0.306 = 0.818.$$

Example 14.2.

Consider a pipeline with a structure depicted in Figure 12.8.

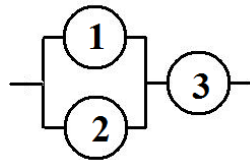


Figure 12.8. System structure for Example 14.2.

Let us use the following throughput distributions for the system units:

$$p_{11} = p_{21} = \Pr\{v = 100\} = 0.8;$$

$$p_{12} = p_{22} = \Pr\{v = 90\} = 0.15;$$

$$p_{13} = p_{23} = \Pr\{v = 80\} = 0.05;$$

$$p_{14} = p_{24} = \Pr\{\text{complete failure}\} \approx 0;$$

and

$$p_{31} = \Pr\{v = 200\} = 0.8;$$

$$p_{32} = \Pr\{v = 180\} = 0.15;$$

$$p_{33} = \Pr\{v = 160\} = 0.05;$$

$$p_{34} = \Pr\{\text{complete failure}\} \approx 0.$$

Step 1. At the beginning consider two units 1 and 2 connected in parallel.

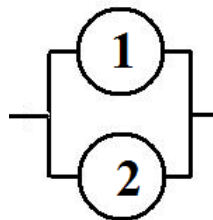


Figure 12.9. Structure for Step 1.

In this case one has to use operator \otimes since

$$f_{PARALLEL}^{(v)}(v_k, v_j) = v_k + v_j.$$

Calculation results are presented in the table below.

		Unit 1				
		State 1	State 2	State 3	State 4	
		(0.8; 100)	(0.15; 90)	(0.05; 80)		
Unit 2	State 1	(0.8; 100)	0.8·0.8 = 0.64 100+100=200	0.15·0.8=0.12 90+100=190	0.05·0.8=0.04 80+100=180	0 100
	State 2	(0.15; 90)	0.8·0.15 = 0.12 100+90=190	0.15·0.15=0.0225 90+90=180	0.05·0.15=0.0075 80+90=170	0 90
	State 3	(0.05; 80)	0.8·0.05 = 0.04 100+80=180	0.15·0.05=0.0075 90+80=170	0.05·0.05=0.0025 80+80=160	9 80
	State 4	(0; 0)	0 100	0 90	0 80	

In result, one obtains a new equivalent unit with the following distribution of troughputs:

$$P_1 = \Pr\{v=200\} = 0.64;$$

$$P_2 = \Pr\{v=190\} = 0.12 + 0.12 = 0.24;$$

$$P_3 = \Pr\{v=180\} = 0.04 + 0.0225 + 0.04 = 0.1025;$$

$$P_4 = \Pr\{v=170\} = 0.0075 + 0.0075 = 0.015;$$

$$P_5 = \Pr\{v=160\} = 0.0025;$$

$$P_6 = \Pr\{v=0\} = 0.$$

Step 2. At this step the obtained above equivalent unit has to be combined with the third unit using the operator \otimes_{\min} .

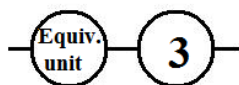


Figure 12.10. Structure at Step 2.

Calculations for the entire system are presented in the table below.

			Unit 3			
			State 1	State 2	State 3	State 4
			(0.8; 200)	(0.15; 180)	(0.05; 160)	(0; 0)
Equivalent unit	State 1	(0.64; 200)	$0.8 \cdot 0.64 = 0.512$ $\min(200, 200)$	$0.15 \cdot 0.64 = 0.096$ $\min(180, 200) = 180$	$0.05 \cdot 0.64 = 0.032$ $\min(160, 200) = 160$	0 $\min(0, 200) = 0$
	State 2	(0.24; 190)	$0.8 \cdot 0.24 = 0.192$ $\min(200, 190) = 190$	$0.15 \cdot 0.24 = 0.036$ $\min(180, 190) = 180$	$0.05 \cdot 0.24 = 0.012$ $\min(160, 190) = 160$	0 $\min(0, 190) = 0$
	State 3	(0.103; 180)	$0.8 \cdot 0.103 \approx 0.082$ $\min(200, 180) = 180$	$0.15 \cdot 0.103 \approx 0.015$ $\min(180, 180) = 180$	$0.05 \cdot 0.103 \approx 0.005$ $\min(160, 180) = 160$	0 $\min(0, 180) = 0$
	State 4	(0.015; 170)	$0.8 \cdot 0.015 = 0.012$ $\min(200, 170) = 170$	$0.15 \cdot 0.015 \approx 0.002$ $\min(180, 170) = 170$	$0.05 \cdot 0.015 \approx 0$ $\min(160, 170) = 160$	0 $\min(0, 170) = 0$
	State 5	(0.003; 160)	$0.8 \cdot 0.003 \approx 0.002$ $\min(200, 160) = 160$	$0.15 \cdot 0.003 \approx 0$ $\min(180, 160) = 160$	$0.05 \cdot 0.003 \approx 0$ $\min(160, 160) = 160$	$\min(0, 160) = 0$
	State 6	(0; 0)	0 $\min(200, 0) = 0$	0 $\min(180, 0) = 0$	0 $\min(160, 0) = 0$	$\min(0, 0) = 0$

These results allow calculating the expected throughput of the pipeline:

$$E\{V\} = 0.512 \cdot 200 + 0.192 \cdot 190 + (0.082 + 0.096 + 0.036 + 0.015) \cdot 180 + (0.012 + 0.002) \cdot 170 + (0.002 + 0.032 + 0.012 + 0.005) \cdot 160 = 0.512 \cdot 200 + 0.192 \cdot 190 + 0.229 \cdot 180 + 0.014 \cdot 170 + 0.051 \cdot 160 \approx 190.6$$

One also can find the PFFO of this system by some chosen criterion of failure. For instance, if a failure criterion is $V < 180$, then PFFO is equal to:

$$P = \Pr\{V \geq 180\} = 0.512 + 0.192 + 0.229 = 0.933.$$

* * *

We restrict ourselves with these two numerical examples, since from methodological viewpoint considering additional structures or different physical objects brings nothing new.

12.7 Conclusion

We limit ourselves with few examples that demonstrate main ideas of using UGF in reliability analysis. Of course, the reader can find objects of different physical nature and compile for them corresponding composition operators. Notice that UGF methodology is not limited by reliability problems. It can be successfully used for other multidimensional “generalized convolutions”.

12.8 Brief historical overview and related sources

Here we offer only papers and highly related books to the subject of this chapter. List of general monographs and textbooks, which can include this topic, is

given in main bibliography at the end of the book. All references are ordered chronologically, since it gives to the reader additional information about the history of the considered subject.

First publications on multistate systems appeared in the end of 1970s and beginning of 1980s. The theory of multistate systems reliability is now developed in a powerful branch of the modern Reliability Theory. Here we made an attempt to demonstrate the place and role of UGF, which first appeared in [Ushakov, 1986].

Now this new direction is flourishing mostly due to intensive and quite productive research by G. Levitin and A. Lisniansky.

Bibliography below is given in chronological-alphabetical order for better exposition of historical background of the subject.

Bibliography

- Barlow, R., and Wu, A. (1978). Coherent systems with multi-state components, *Mathematics of Operations Research*, vol. 3, 1978.
- El-Newehi, E., F. Proschan, and J. Seturaman, (1978) Multistate coherent systems, *J. Applied Probability*, vol. 15, 1978.
- Griffith, W. (1980). Multistate reliability models, *J. Applied Probability*, vol. 17.
- Fardis, M. N., C. A. Cornell, (1981) Analysis of coherent multistate systems, *IEEE Transactions on Reliability*, vol 30.
- Hudson, J.C., and K. Kapur. (1982). Reliability Theory for Multistate Systems with Multistate Components, *Microelectronics and Reliability*, vol. 22, No. 1.
- Fardis, M., and C. Cornell, (1982), Multistate reliability analysis. *Nuclear Engineering and Design*, Vol.60.
- Hudson . J. C. and K. C. Kapur. (1983), Reliability analysis of multistate systems with multistate components, *Transactions of Institute of Industrial Engineers*, vol. 15.
- Ebrahimi N. (1984), Multistate reliability models, *Naval Res. Logistics*, vol. 31.
- Ohio, F., and T. Nishida, (1984), On multistate coherent systems, *IEEE Transactions on Reliability*, vol. 33.
- Xue J. (1985) On Multistate System Analysis, *IEEE Transactions on Reliability*, vol. R-34.
- Xue, J. (1985) New approach for multistate system analysis. *Reliability Engineering* , Vol.10.
- Aven, T. (1985). Reliability evaluation of multi-state systems with multi-state components. *IEEE Transactions on Reliability*, Vol. 34, No.5.

- Ushakov, I. (1986). A Universal Generating Function. Soviet Journal of Computer and System Sciences, Vol. 24, No. 5.
- Natvig B., S. Sormo, A. Holen, and G. Hogasen. (1986). Multistate reliability theory - a case study. *Advances in Applied Probability*, Vol.18.
- Ushakov, I. (1988) Reliability analysis of multi-state systems by means of modified generating function. *Journal Information Process. Cybernet.* , vol. 24, no. 3.
- Aven, T. (1993) On performance measures for multistate monotone systems, *Reliability Engineering and System Safety*, vol. 41.
- Gnedenko, B.V., and I.A. Ushakov. Probabilistic Reliability Engineering. John Wiley & Sons, New York.1995.
- Ushakov, I.A. An object oriented approach to generalized generating function. Proc. of the ECCO-XI Conference (European Chapter on Combinatorial Optimization), Copenhagen, 1998.
- Levitin, G. (1999) A universal generating function approach for analysis of multi-state systems with dependent elements. *Reliability Engineering & System Safety*, vol. 84, No.3.
- Levitin, G., and A. Lisnianski. (1999) Importance and sensitivity analysis of multi-state systems using universal generating functions method, *Reliability Engineering & System Safety*, vol. 65.
- Chakravarty, S., and Ushakov, I. Effectiveness Analysis of Globalstar™ Gateways. Proceedings of Second International Conference on Mathematical Methods in Reliability (MMR'2000), Vol. 1, Bordeaux, France, 2000.
- Ushakov, I. (2000) The method of generalized generating sequences, *European Journal of Operational Research*, 125 (2).
- Chakravarty, S., and Ushakov, I. Reliability measure based on average loss of capacity. *International Transactions in Operational Research*, Vol. 9, No. 2, 2002.
- Rykov, V., and B. Dimitrov. (2002) On multi-state reliability systems, *Information Processes*, Vol.2, No.2.
- Lisnianski , A., and G. Levitin, (2003) Multi-State System Reliability. Assessment, Optimization, Applications. World Scientific.
- Levitin, G., A. Lisnianski, and I. Ushakov. (2003) Reliability of Multi-State Systems: A Historical Overview in *Mathematical and statistical methods in reliability*. World Scientific.
- Levitin, G. (2004) A universal generating function approach for the analysis of multi-state systems with dependent elements, *Reliability Engineering and System Safety*, 84/3.
- Levitin, G. (2005). The Universal Generating Function in Reliability Analysis and Optimization. Springer.

- Ding, Y., and A. Lisnianski. (2008). Fuzzy universal generating functions for multi-state system reliability assessment. *Fuzzy Sets and Systems*, 159 (3).

Appendicis

A. Main distributions related to Reliability Theory

Since each of reliability indices represents one or another characteristic of probability distribution, we give brief information about main distributions related to the subject.

A1. Discrete distributions.

A.1.1. Degenerate distribution

In a sense, it is a distribution of “non-random random variable”: this is the distribution of a constant value. In Reliability Theory a constant is used, for instance, for description of switching time to redundant unit or to duration of monitoring tests.

The degenerate distribution is localized at a point T on the real line. The cumulative distribution function of the degenerate distribution concentrated in point T is:

$$F(t) = \begin{cases} 0, & \text{if } t < T, \\ 1, & \text{otherwise.} \end{cases} \quad (1)$$

Its probability mass function and distribution function are depicted in Figure 1.

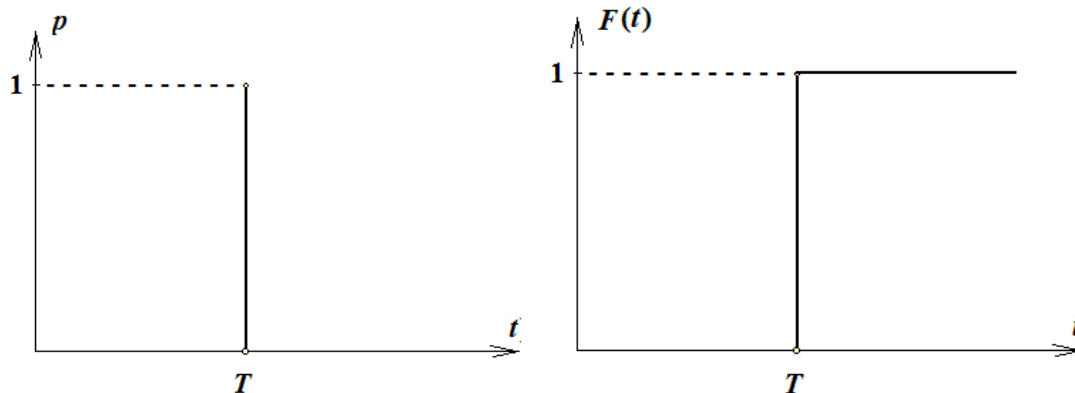


Figure 1. Probability mass function and distribution function of degenerate distribution.

A1.2. Bernoulli distribution

Let two mutually exclusive events are possible: success, which occurs with probability p , and failure, which occurs with probability $q=1-p$. Prescribe conditionally value 1 to success and value 0 to failure and introduce a random variable X that is called ***Bernoulli²⁹ random variable***:

$$X = \begin{cases} 1, & \text{if success has occurred,} \\ 0, & \text{otherwise.} \end{cases}$$

Distribution of the Bernoulli r.v. is called the Bernoulli distribution:

$$\Pr\{x = X\} = \begin{cases} p, & \text{if } X = 1, \\ q, & \text{if } X = 0, \end{cases} \quad (2)$$

or in more compact form:

$$\Pr\{x=X\} = p^X q^{X-1}. \quad (3)$$

Thus, the Bernoulli r.v. is a special case of degenerate r.v. when $T=1$.

The mathematical expectation (the mean) of a Bernoulli random variable X is

$$E\{X\} = 1 \cdot p + 0 \cdot q = p, \quad (4)$$

and its variance is:

$$\sigma^2\{X\} = pq. \quad (5)$$

A.1.3. Binomial distribution

If one observes a series of n Bernoulli r.v.'s, the number of successes (and, respectively, failures) is random. The distribution of this r.v. is named binomial distribution. Binomial r.v. is the sum of Bernoulli r.v.'s, i.e.

$$X = X_1 + X_2 + \dots + X_n = \sum_{k=1}^n X_k, \quad (6)$$

where X_k is Bernoulli r.v.

²⁹ **Jacob Bernoulli** (1654–1705) was a Swiss mathematician, one of the many prominent mathematicians in the Bernoulli family.

The probability of occurrence of exactly k successes is:

$$P(n; k) = \binom{n}{k} p^k q^{n-k}. \quad (7)$$

Distribution function in this case is written as:

$$F(n; m) = \sum_{k=0}^m \binom{n}{k} p^k q^{n-k}. \quad (8)$$

An example of the binomial distribution is given in Figure 2.

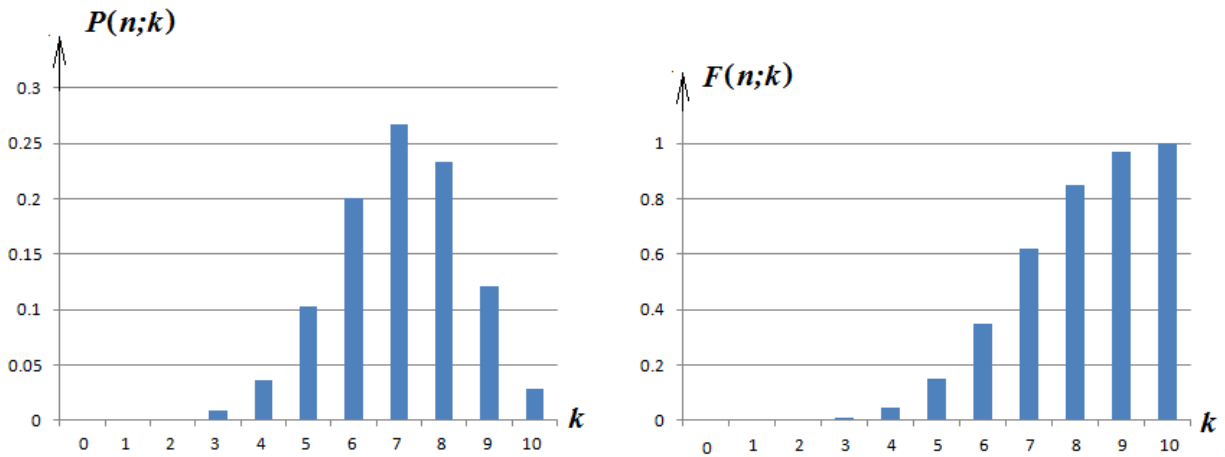


Figure 2. Mass function, $P(n;k)$, and cumulative function, $F(n;k)$, for binomial distribution with parameters $p = 0.7$ and $n = 10$.

Using (4) and applying the theorem about expectation of the sum of r.v.'s, one can immediately write:

$$E \{X_1 + X_2 + \dots + X_n\} = nE \{X_1\} = np. \quad (9)$$

Using (5) and the formula for the sum of variances, one gets:

$$\sigma^2 = npq. \quad (10)$$

Let us underline an important property of binomial distribution: if one performs two series of Bernoulli tests of n_1 and n_2 trials and probability of success in both cases are the same p , then it is equivalent to a single test with the total number of trials equals $n=n_1+n_2$. In other words, the convolution of two Binomial distribution functions produces the binomial distribution with a new parameter n .

A.1.4. Poisson distribution

This distribution is often used in various mathematical reliability models. For instance, it describes distribution of events of Poisson process in a fixed time interval. The “physical” sense of this distribution can be demonstrated on the following example. Consider the binomial distribution for the case when the number of Bernoulli trials, n is extremely large and, at the same time, value of p is very close to 1. In this case, numerical calculation of probability (7) presents a definite complexity: one needs to sum up a huge number of summands each of which is a product of large value of binomial coefficient by very small values of probabilities. In this case, it is reasonable to use a limit passage for $n \rightarrow \infty$, $q \rightarrow 0$, and $nq = \text{Const}$:

$$\lim_{n \rightarrow \infty} \binom{n}{k} q^k p^{n-k} = \frac{q^k}{k!} \lim_{n \rightarrow \infty} [n \cdot (n-1) \cdot \dots \cdot (n-k+1)] (1-q)^{n-k}. \quad (11)$$

Since $n \gg k$, then $\lim_{n \rightarrow \infty} [n \cdot (n-1) \cdot \dots \cdot (n-k+1)] = n^k$ and $\lim_{n \rightarrow \infty} (1-q)^{n-k} = \exp(-nq)$. In result one gets:

$$\lim_{n \rightarrow \infty} \binom{n}{k} q^k p^{n-k} = \frac{(nq)^k}{k!} \exp(-nq) \quad (12)$$

Formula (12) is the corollary of fundamental Poisson theorem.

Formula (20) can be used as approximation even for relatively small n . In Figure 4 four particular cases for $n = 20$ are depicted: B-1 is binomial distribution with $q = 0.1$, P-1 is corresponding Poisson distribution with $\lambda = 0.1 \cdot 20 = 0.2$; B-2 is binomial distribution with $q = 0.25$, P-2 is corresponding Poisson distribution with $\lambda = 0.25 \cdot 20 = 5$.

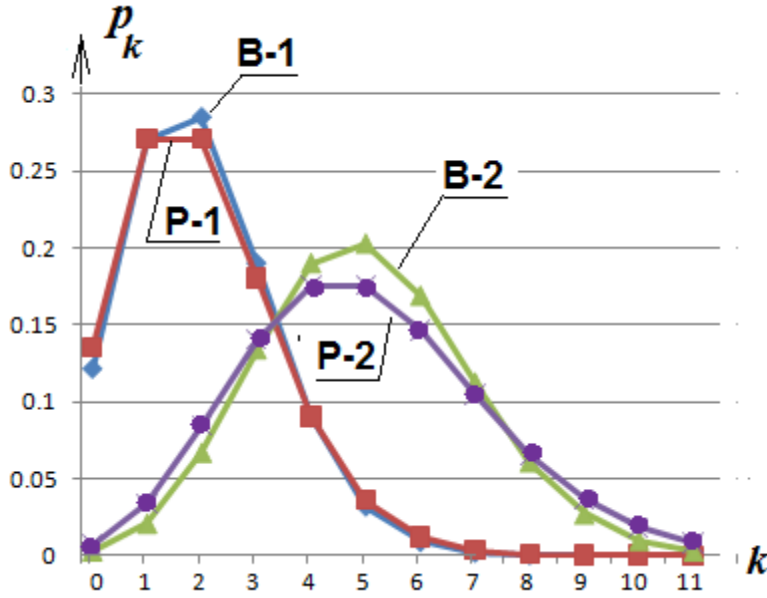


Figure 3. Comparison binomial and Poisson mass functions for two particular cases.

Now consider that during time interval $[0, t]$ random failures can occur with constant rate λ . Divide interval $[0, t]$ by small subintervals Δ . Then in each interval may occur $\lambda\Delta + o(\Delta)$ failures, where $o(\Delta)$ is a value of higher order of smallness. In other words, we consider $n = \frac{t}{\Delta}$ Bernoulli trials. So, we can value nq replace in this case for $\Lambda = \lambda t$. It means that (12) can be rewritten as:

$$p_k(\Lambda) = \frac{\Lambda^k}{k!} e^{-\Lambda} . \quad (13)$$

Cumulative function in this case is:

$$P_{j \leq k}(\Lambda) = \sum_{j=0}^k \frac{\Lambda^j}{j!} e^{-\Lambda} . \quad (14)$$

Examples of mass function and cumulative function of the Poisson distribution are given in Figures 5 and 6, respectively.

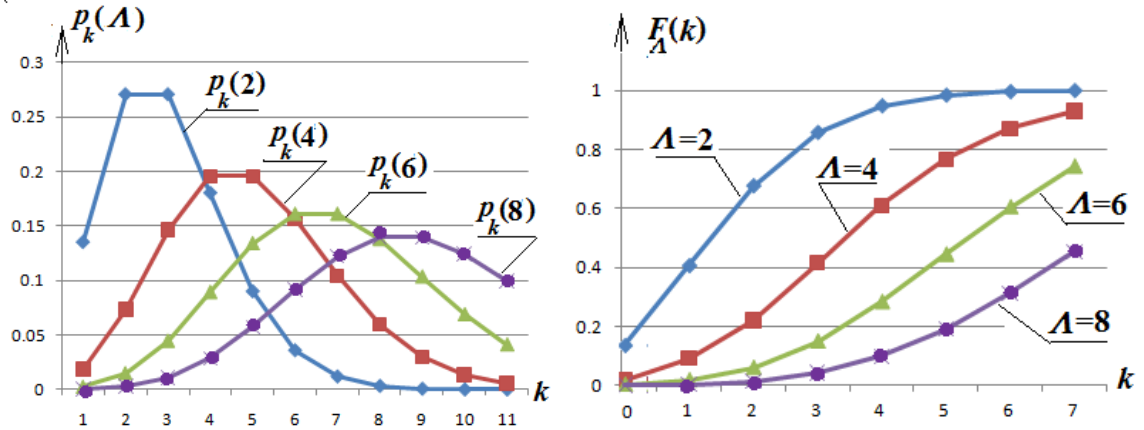


Figure 4. Mass function and cumulative function of Poisson distributions for four meanings of parameter λ equal to 2, 4, 6 and 8, respectively.

The mathematical expectation for Poisson distribution is defined in usual way:

$$E\{k\} = \sum_{k=0}^{\infty} k \frac{(\lambda t)^k}{k!} e^{-\lambda t} = \lambda t \sum_{k=1}^{\infty} \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} = \lambda t e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} = \lambda t . \quad (15)$$

The variance of this distribution equals its mean:

$$\sigma^2 = \lambda t. \quad (16)$$

A.1.5. Geometric distribution

Consider a series of Bernoulli trials. Denote the number of consecutive successes until first failure occurrence by X . Distribution of such r.v. is called geometric. Its mass function is expressed as:

$$\Pr\{X=k\} = p^x q . \quad (17)$$

Cumulative function of geometric distribution has the form”

$$\Pr\{X \leq x\} = q \sum_{k=0}^x p^k = 1 - q \sum_{k=x+1}^{\infty} p^k = 1 - qp^{x+1} \sum_{k=0}^{\infty} p^k = 1 - \frac{qp^{x+1}}{1-p} = 1 - p^{x+1}. \quad (18)$$

Notice that (18) can be obtained in different way on the basis of simple arguments. Probability $P\{X > x\} = p^{x+1}$ is the probability that $x+1$ successes occur in row. So, complementary probability is $P\{X \leq x\} = 1 - p^{x+1}$.

The mathematical expectation of r.v. X is found as usual:

$$E\{x\} = \sum_{x=0}^{\infty} xp^x q = pq \cdot \frac{d}{dp} \sum_{x=0}^{\infty} p^x = pq \cdot \frac{d}{dp} \left(\frac{1}{1-p} \right) = \frac{p}{q}. \quad (19)$$

By the way, sum $\sum_{x=0}^{\infty} p^x$ can be found in another way. Denote $y = 1 + p + p^2 + \dots$. For this infinite sum, the following equality is true: $1 + p + p^2 + \dots = 1 + py$. From here $y = \frac{1}{1-p} = \frac{1}{q}$ if geometric series converges.

A.2. Continuous distributions

A.2.1. Intensity function

For continuous distribution there is an important additional characteristic that is often used in reliability theory. This is intensity function, $\lambda(t)$, defined as conditional density function at moment t under condition that the considered r.v. is larger than t , i.e.:

$$\lambda(t) = \frac{1}{P(t)} \cdot \frac{dF(t)}{dt}. \quad (20)$$

One can rewrite (20) as:

$$\lambda(t) = -\frac{1}{P(t)} \cdot \frac{dP(t)}{dt} = \frac{d}{dt} \ln P(t). \quad (21)$$

For (21) follows that for any distribution, the probability of failure-free operation can be written in the form:

$$P(t) = \exp \left(- \int_0^t \lambda(t) dt \right). \quad (22)$$

Expression (22) leads beginners in reliability engineering to confusion: they call arbitrary d.f. presented in such a form “exponential distribution”. Indeed, for an exponential distribution always $\lambda(t) = const$.

For the probability that residual time to failure larger than t under condition that an object has already worked time x , one can write:

$$P(t | x) = \exp \left(- \int_x^{t+x} \lambda(t) dt \right). \quad (23)$$

A.2.2. Continuous uniform distribution

Continuous uniform distribution (or rectangular distribution) is defined on restricted interval $[x, y]$ with density function:

$$f(t) = \frac{1}{y-x}. \quad (24)$$

and distribution function:

$$F(t) = \begin{cases} 0 & \text{if } t \leq x \\ \frac{t-x}{y-x} & \text{if } x < t \leq y. \\ 1 & \text{if } t > y \end{cases} \quad (25)$$

Functions (37) and (38) are depicted I Figure 5.

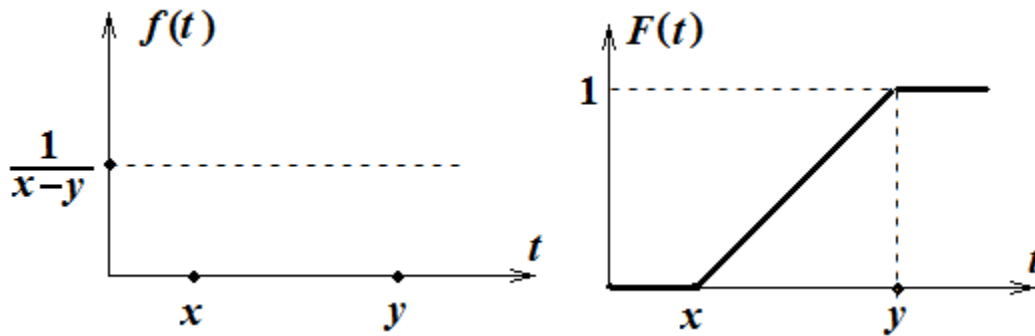


Figure 5. Density and distribution function of uniform distribution.

The mean of this distribution equals to:

$$E\{X\} = \frac{y-x}{2}, \quad (26)$$

and variance equals to:

$$\sigma^2 = \frac{(y-x)^2}{12}. \quad (27)$$

This distribution is widely used for Monte Carlo modeling.

A.2.3. Exponential distribution

Consider geometric distribution with the probability of success that is very close to 1, or, what is the same, the probability of failure is extremely small:

$q = 1-p \ll 1$. The probability that r.v. X , that equal the number of consecutive successes in such series of Bernoulli trials is more than some fixed n is equal to:

$$\Pr\{X \geq n\} = p^n = (1-q)^n. \quad (28)$$

Let each Bernoulli trial lasts time Δ . Denote the failure probability as $q = \lambda\Delta$, where λ is some constant. Assume that the number of trials, n , is sufficiently large. Let the total time of trials is $t = n\Delta$. Then (28) can be rewritten as:

$$P(t) = P\{X \geq n\} = (1-\lambda\Delta)^{t/\Delta}, \quad (29)$$

And after $\Delta \rightarrow 0$, one gets:

$$P(t) = \exp(-\lambda t). \quad (30)$$

Formula (30) gives a function complementary to the distribution function, so:

$$F(t) = 1 - \exp(-\lambda t). \quad (31)$$

This is exponential distribution function with parameter λ .

Thus, in a sense, exponential d.f. is a limit distribution for geometric one. Density function for this distribution is found as follows:

$$f(t) = \frac{d}{dt}[1 - \exp(-\lambda t)] = \lambda \exp(-\lambda t). \quad (32)$$

The mathematical expectation is:

$$E\{X\} = \int_0^{\infty} t \lambda \exp(-\lambda t) dt = \frac{1}{\lambda}. \quad (33)$$

and variance is:

$$\sigma = \frac{1}{\lambda^2}. \quad (34)$$

It is clear that exponential distribution, as well as geometric one, possesses the Markovian property, that is:

$$P(x, t+x | x > x) = P(0, t) = \exp(-\lambda t). \quad (35)$$

In reliability terms it means that an object that is in operational state at some moment t is in operational state by its reliability properties is equivalent to an absolutely new one. Of course, this assumption should be always taken into account before using exponential models for practical purposes: not always such assumption is adequate to a real technical object.

Intensity function, $\lambda(t)$, for exponential function is constant:

$$\frac{1}{P(t)} \cdot \frac{dF(t)}{dt} = \frac{1}{e^{-\lambda t}} \cdot \lambda e^{-\lambda t} = \lambda \quad (36)$$

A.2.5. Erlang distribution

Erlang distribution is a convolution of n identical exponential distributions, i.e. Erlang r.v. is a sum of n i.i.d. exponential r.v.'s. Erlang d.f. represents a particular case of Gamma-distributions with an integer shape parameter.

Erlang density function is:

$$f(t) = \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} \cdot \exp(-\lambda t), \quad t \geq 0. \quad (37)$$

The mean and variance of this distribution equal, correspondingly, to $\frac{n}{\lambda}$ and $\frac{n}{\lambda^2}$.

A.2.6. Hyper-exponential distribution

This distribution appears in some reliability models. Hyper-exponential distribution is weighed sum of exponential d.f.'s and defined as:

$$F(t) = 1 - \sum_{k=1}^n p_k \exp(-\lambda_k t). \quad (38)$$

where $\sum_{k=1}^n p_k = 1$ and all $p_k > 0$.

An example of hyper-exponential distributions is presented in Figure 8.

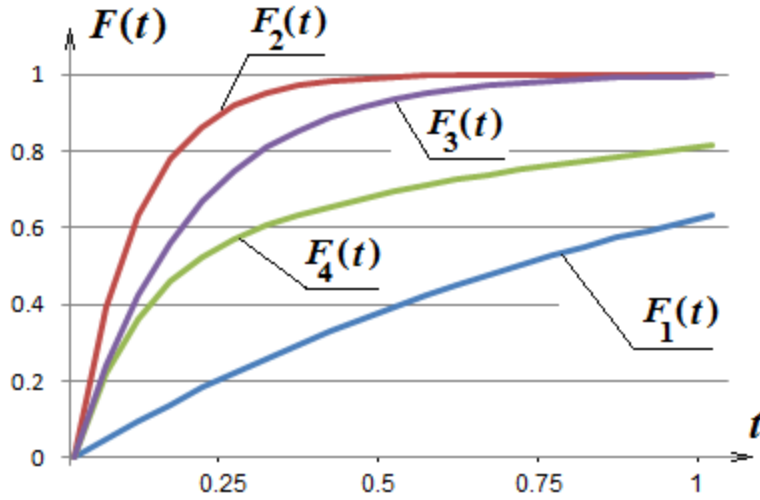


Figure 6. Exponential functions $F_1(t) = e^{-t}$ and $F_1(t) = e^{-10t}$, and hyper-exponential functions $F_3(t) = 0.5F_1(t) + 0.5F_2(t)$. For comparison, function $F_4(t) = e^{-5.5t}$ with parameter $\lambda = 0.5(1 + 10)$ is also presented.

Obviously that the mean of this distribution is equal to:

$$E\{X\} = \sum_{k=1}^n \frac{p_k}{\lambda_k}. \quad (39)$$

This distribution has decreasing intensity function:

$$\lambda(t) = \frac{\sum_{k=1}^n p_k \lambda_k \exp(-\lambda_k t)}{\sum_{k=1}^n p_k \exp(-\lambda_k t)}. \quad (40)$$

Indeed, function $\lambda(t)$ is monotone and for $t=0$ is equal to:

$$\lambda(0) = \frac{\sum_{k=1}^n p_k \lambda_k}{\sum_{k=1}^n p_k} = \sum_{k=1}^n p_k \lambda_k, \quad (41)$$

and for $t \rightarrow \infty$, as one can see directly from (41), $\lim_{t \rightarrow \infty} \lambda(t) = \min_{1 \leq k \leq n} \lambda_k$. It is clear that $\min_{1 \leq k \leq n} \lambda_k < \lambda(0)$. In this case, intensity function has the form depicted in Figure 7.

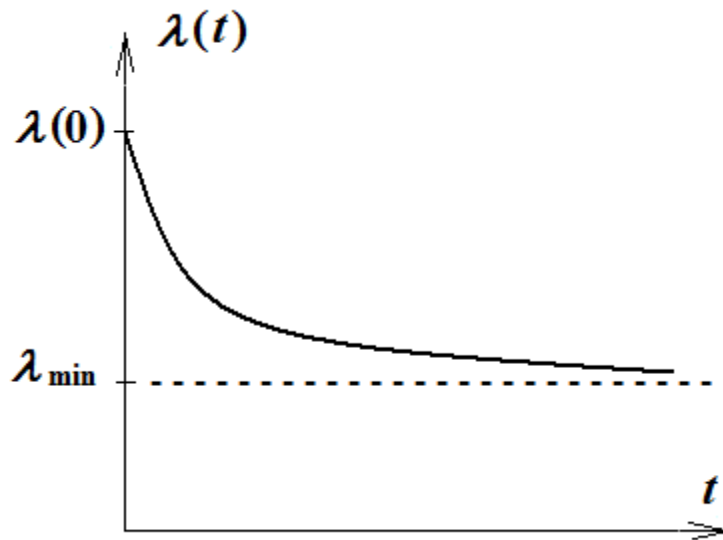


Figure 7. Function $\lambda(t)$ for hyper-exponential distribution.

A.2.7. Normal distribution

A.2.8. Weibull-Gnedenko distribution

In conclusion, let us consider Weibull-Gnedenko distribution that is widely used in applied reliability analysis. Distribution function in this case has the form:

$$F(t) = \begin{cases} 1 - \exp(-(\lambda t)^\beta) & \text{для } t \geq 0 \\ 0 & \text{для } t < 0 \end{cases}, \quad (42)$$

and density function is:

$$f(t) = \begin{cases} \lambda^\beta \beta t^{\beta-1} \exp(-(\lambda t)^\beta) & \text{для } t \geq 0 \\ 0 & \text{для } t < 0 \end{cases}. \quad (43)$$

Parameters λ and β are called scale and shape parameters, respectively. Examples of distribution and density functions depicted in figure

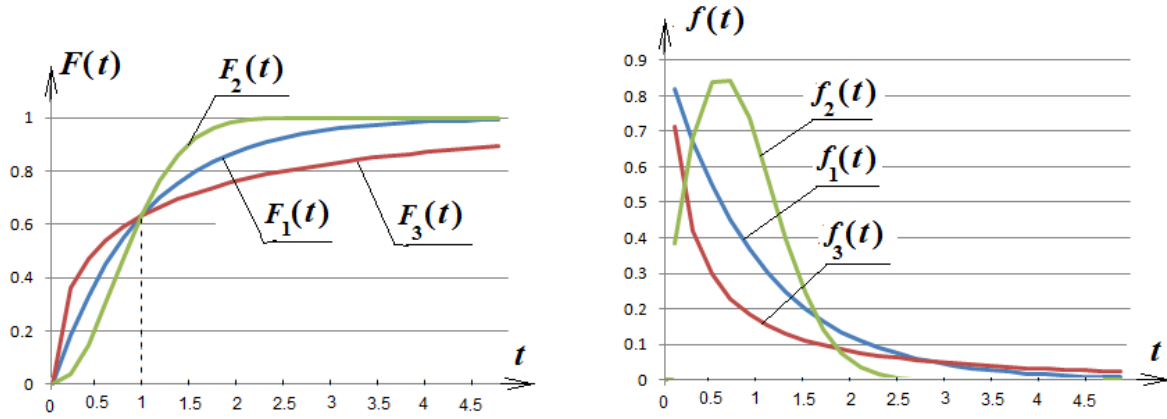


Figure 7. Distribution, $F(t)$, and density, $f(t)$, functions. All distributions with $\lambda=1$. Subscript “2” relates to the Weibull-Gnedenko distribution with parameter $\beta=2$ and subscript “3” relates to the distribution with parameter $\beta=0.5$.

Weibull-Gnedenko distribution has intensity function

$$\lambda(t) = \lambda^\beta \beta t^{\beta-1} \tag{44}$$

That is increasing for $\beta > 1$ and decreasing for $\beta < 1$. Obviously, for $\beta = 1$ this distribution coincides with exponential one. For $\beta=1$ intensity functions is linear.

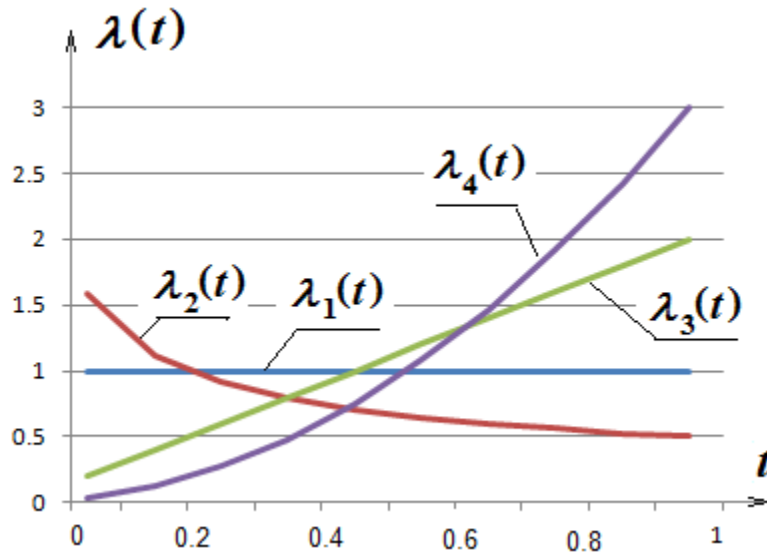


Figure 8. Examples of intensity functions for Weibull-Gnedenko distribution. All distributions with $\lambda=1$. Subscripts correspond: “1” to $\beta=1$ (exponential distribution), “2” to $\beta=0.5$, “3” to $\beta=2$ and “4” to $\beta=3$.

The mean of this distribution equals:

$$E\{X\} = \frac{1}{\lambda} \Gamma\left(1 + \frac{1}{\beta}\right), \quad (45)$$

and variance:

$$\sigma^2 = \frac{1}{\lambda^2} \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \left(\Gamma\left(1 + \frac{1}{\beta}\right) \right)^2 \right], \quad (46)$$

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

B. Laplace transformation

Recall that the Laplace³⁰ transform (LT), $\varphi(s)$, of function $y(t)$ is defined over the positive axis as:

$$\varphi(s) = \int_0^{\infty} y(t) e^{-st} dt. \quad (1)$$

For derivative of $y(t)$, that is for $y'(t)$, one can write, using integration by parts:

$$\int_0^{\infty} y'(t) e^{-st} dt = \int_0^{\infty} e^{-st} dy(t) = y(t) e^{-st} \Big|_0^{\infty} - \int_0^{\infty} f(t) d(e^{-st}) = -y(0) + s \int_0^{\infty} f(t) \cdot e^{-st} dt = -y(0) + s\varphi(s) \quad (2)$$

where $y(0)$ is the initial condition of the process, i.e. $y(t)$ at $t = 0$.

For integral of function $y(t)$, the expression for LT can be derived with using integration by parts:

$$\int_0^{\infty} \left[\int_0^t y(x) dx \right] \cdot e^{-st} dt = -\frac{1}{s} \int_0^{\infty} \left[\int_0^t y(x) dx \right] \cdot d(e^{-st}) = \frac{1}{s} \left[e^{-st} \int_0^t \mathbf{b}l(x) dx \Big|_0^{\infty} - \int_0^{\infty} \mathbf{b}l(t) \cdot e^{-st} dt \right] = \frac{1}{s} \cdot \varphi(s). \quad (3)$$

For system of differential equations, describing system's transit from state to state, one usually considers the following type of the linear differential equations:

$$\frac{d}{dt} p_k(t) = -p_k(t) \sum_{i \in e(k)} \alpha_{ki} + \sum_{i \in E(k)} \alpha_{ik} p_i(t) = -p_k(t) A_k + \sum_{i \in E(k)} \alpha_{ik} p_i(t) \quad (4)$$

where α_{ki} is the passage intensity from state “ k ” to state “ i ”; $e(k)$ is subset of the total set of states where the process can move at one step from state “ k ”, and $E(k)$ is subset of states from where the process can move at one step to state “ k ”; $A_k = \sum_{i \in e(k)} \alpha_{ki}$. If

there is no absorbing states, one needs to use the initial conditions of the type $p_k(0)$, i.e. $p_k(t)$ at moment $t = 0$.

Laplace transforms for system (4) has the form:

$$-p_k(0) + s\varphi_k(s) = -\phi_k(s) \sum_{i \in e(k)} \alpha_{ki} + \sum_{i \in E(k)} a_{ik} \varphi_i(s), \quad (5)$$

³⁰ **Pierre-Simon, marquis de Laplace** (1749–1827) was a French mathematician and astronomer whose work was pivotal to the development of mathematical astronomy and statistics.

That can be rewritten in open form as

$$\begin{aligned}
 -(s + A_1) \cdot \varphi_1(s) + \alpha_{21}\varphi_2(s) + \dots + \alpha_{n1}\varphi_n(s) &= p_1(0) \\
 \alpha_{12}\varphi_1(s) - (s + A_2) \cdot \varphi_2(s) + \dots + \alpha_{n2}\varphi_n(s) &= p_2(0) \\
 &\dots\dots\dots \\
 \alpha_{1n}\varphi_1(s) + \alpha_{2n}\varphi_2(s) + \dots - (s + A_n) \cdot \varphi_n(s) &= p_n(0)
 \end{aligned} \tag{6}$$

If considered Markov process has no absorbing states, equations in (6) are mutually dependent and in this case one has to replace any of them by the normalization equation:

$$s\varphi_1(s) + s\varphi_2(s) + \dots + s\varphi_n(s) = 1. \tag{7}$$

The same system of equations can be written in matrix form as:

$$\begin{vmatrix}
 -(s + A_1) & \alpha_{21} & \dots & \alpha_{n1} \\
 \alpha_{12} & -(s + A_2) & \dots & \alpha_{n2} \\
 \dots & \dots & \dots & \dots \\
 \alpha_{1n} & \alpha_{2n} & \dots & -(s + A_n)
 \end{vmatrix} \times \begin{vmatrix} \varphi_1(s) \\ \varphi_2(s) \\ \dots \\ \varphi_n(s) \end{vmatrix} = \begin{vmatrix} p_1(0) \\ p_2(0) \\ \dots \\ p_n(0) \end{vmatrix} \tag{7}$$

where

$$D = \begin{vmatrix}
 -(s + A_1) & \alpha_{21} & \dots & \alpha_{n1} \\
 \alpha_{12} & -(s + A_2) & \dots & \alpha_{n2} \\
 \dots & \dots & \dots & \dots \\
 \alpha_{1n} & \alpha_{2n} & \dots & -(s + A_n)
 \end{vmatrix} \tag{8}$$

is determinant of equation system (6). Solution of this system can be obtained with the Cramer Rule:

$$\varphi_k(s) = \frac{D_k(s)}{D(s)}. \tag{9}$$

where $D_k(s)$ is determinant in which the k-th column is substituted by the right column of absolute terms.

For inverse Laplace transforms, one uses the following procedure.

(a). Open the numerator and denominator of fraction (9) and write $\varphi_k(s)$ in the form:

$$\varphi_k(s) = \frac{A_0 + A_1s + \dots + A_ns^n}{B_0 + B_1s + \dots + B_{n+1}s^{n+1}}, \tag{10}$$

where A_k and B_k are unknown coefficients to be found.

(b). Find roots of the polynomial in the denominator of fraction (10):

$$B_0 + B_1s + \dots + B_{n+1}s^{n+1} = 0. \quad (11)$$

Let these roots are s_1, s_2, \dots, s_n . It means that:

$$B_0 + B_1s + \dots + B_{n+1}s^{n+1} = \prod_{j=1}^{n+1} (s - s_j). \quad (12)$$

(c). Write $\varphi_k(s)$ as the sum of simple fractions:

$$\varphi_A(s) = \frac{\beta_1}{s - s_1} + \frac{\beta_2}{s - s_2} + \dots + \frac{\beta_{n+1}}{s - s_{n+1}}, \quad (13)$$

where β_k 's are unknown coefficients.

(d). After reduction of fraction to a common denominator, write $\varphi_k(s)$ in the form:

$$\varphi_k(s) = \frac{\sum_{1 \leq j \leq n, i \neq j} \beta_j \prod (s - s_i)}{\prod_{1 \leq i \leq n+1} (s - s_i)} \quad (14)$$

(e). Open the numerator of the fraction and perform collecting terms:

$$\varphi_{E(k)}(s) = \frac{\gamma_0 + \gamma_1s + \gamma_2s^2 + \dots + \gamma_n s^n}{(s - s_1)(s - s_2) \cdot \dots \cdot (s - s_{n+1})}, \quad (15)$$

where γ_k 's are presented through known β_j 's and s_j 's.

(f). Polynomials (10) and (15) are equal if and only if $A_k = \gamma_k$. From these conditions, one finds unknown coefficients β_k .

(f). After finding β_k , one applies inverse Laplace transforms to (13).

For practical solutions, one can use existing tables of inverse Laplace transforms. A sample of the most important Laplace transforms, frequently used in reliability analysis, is given in the table below.

Table 1. Most important Laplace transforms.

Origin	Transform	Origin	Transform
$\alpha p_1(t) + \beta p_2(t)$	$\alpha \varphi_1(s) + \beta \varphi_2(s)$	$\int_0^t p_1(x) p_2(t-x) dx$	$\varphi_1(s) \cdot \varphi_2(s)$
$\frac{dp(t)}{dt}$	$-p(0) + s\varphi(s)$	$(-t)^k p(t)$	$\varphi^{(k)}(s)$

$\int_0^t p(x)dx + C$	$\frac{\varphi(s)}{s} + \frac{C}{s}$	$e^{\alpha} p(t)$	$\varphi(s - a)$
$p(bt)$	$\frac{1}{b} \varphi\left(\frac{s}{b}\right)$, for $b > 0$	$\frac{t^{n-1}}{(n-1)!}$	$\frac{1}{s^n}$, for $n = 1, 2, \dots$
$p(t-b)$	$e^{-bs} \varphi(s)$, for $b > 0$	$\frac{t^{n-1}}{(n-1)!} \cdot e^{-\beta t}$	$\frac{1}{(s - \beta)^n}$, for $n = 1, 2, \dots$

Usage of this table is almost obvious. For instance, if your solution in Laplace transforms is $\frac{\alpha}{\alpha + s}$, it means that in the “space of normal functions” the solution is $e^{-\alpha t}$.

Thus, the main idea of LT consists in replacing integro-differential equations by equivalent algebraic ones, which can be easier solved, and then make inverse transform for the obtained algebraic solution.

Why one needs to use LT? Explanation is simple: it makes solution of integro-differential equations simpler. The main idea of using LT exposed in Figure 2.

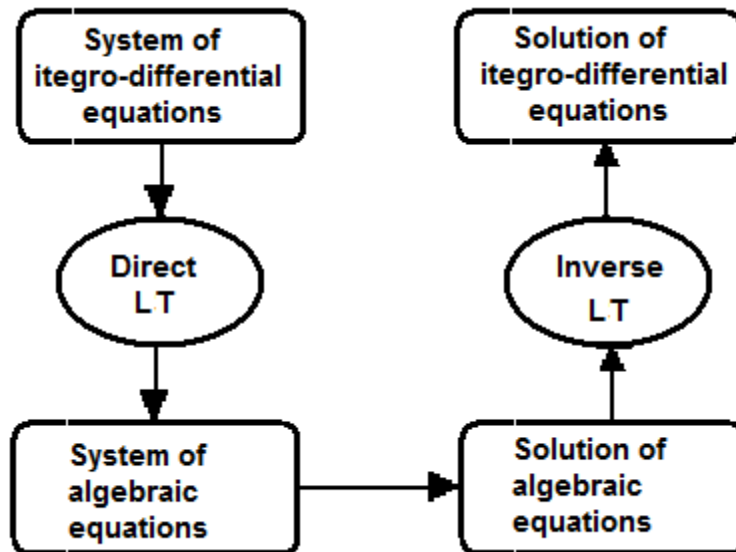


Figure 2. Graphical explanations of the idea of LT use.

C. Markov processes

C.1. General Markov process

From the very beginning, we would like to emphasize that a Markov model is an idealization of real processes. The 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 mathematical model are an appropriate fit to each other. If, in fact, they fit, then a Markov model is very convenient.

Assume that we can construct the transition graph which sufficiently describes a system's operation. (We use below reliability terminology only for the readers convenience.) 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 complimentary disjoint subsets, *up states* and *down states*. Necessary condition that transition from the subset of up states to the set of down states occurs is a failure of one of operating units. Of course, if a unit is redundant, the system's failure does not occur. An inverse transition may occur only if a failed unit is recovered 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, and n is the number units in the system. We set $s_i=1$ if the unit is operational and $s_i=0$, otherwise. If each system unit has two states, say, operational and failure, and the system consists of n units, the system, in principle, can have $N = 2n$ different states. System state $(s_1 = 1, \dots, s_i = 1, \dots, s_n = 1)$ will play a special role in further deductions, so let us assign to this state subscript "1".

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 change 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 being in a failed state has been recovered. The transition rate for this case equals the i -th unit's repair rate. These kinds of transitions are most common. For Markovian models, no more than a single failure can occur during infinitesimally small period of time.

For some reasons, sometimes one introduces absorbing states, i.e. such states that process once entering the state will never leave it. The sense of these states will be explained later.

We denote transitions from state to state on transition graphs with arrows. The rates (intensities) are denoted as weights on the arrows. After the transition graph has

been constructed, it can be used as a visual aid to determine different reliability indexes.

Consider a continuous Markov process with discrete states (Figure 1).

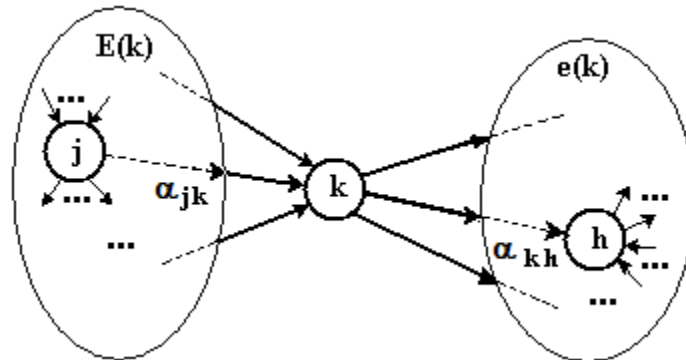


Figure 1. Conditional presentation of state k and its immediate neighbors.

On this graph, the process can move during infinitesimally small time interval from state k to one of states of subset $e(k)$, and at the same time the process can occur in state from some subset $E(k)$. Notice that a state can belong simultaneously to both subsets, i.e. the process can go from one state to another back and forth.

C.1.2. Non-stationary availability coefficient

This reliability index can be compiled with the help of transition graph above. In this case, one should choose the initial state, i.e. state in which system is at moment $t = 0$. For arbitrary state k , one can write the following formula of total probability

$$p_k(t + \Delta) = p_k(t) \cdot \left[1 - \Delta \sum_{i \in e(k)} \alpha_{ki} \right] + p_j(t) \cdot \left[\Delta \sum_{j \in E(k)} \alpha_{jk} \right]. \quad (1)$$

where α_{ki} is the transition intensity from state k to state i ; $e(k)$ is subset of the total set of states where the process can move at one step from state k ; and $E(k)$ is subset of states from where the process can move at one step to state k .

Indeed, the process occurs at moment $t + \Delta$ in state k by two ways:

- (a) At moment t , it is in state k and does not leave this state during infinitesimally small interval Δ ;
- (b) At moment t , it is in any state belonging to subset $e(k)$ and moves namely to state k during infinitesimally small interval Δ .

Equation (1) can be transformed into the form:

$$\frac{p_k(t+\Delta) - p_k(t)}{\Delta} = -p_k(t) \sum_{i \in e(k)} \alpha_{ki} + p_j(t) \sum_{j \in E(k)} \alpha_{jk} \quad (2)$$

Limiting passage for $\Delta \rightarrow 0$ leads to the differential equation for state k :

$$\frac{dp_k(t)}{dt} = -p_k(t) \sum_{i \in e(k)} \alpha_{ki} + \sum_{j \in E(k)} \alpha_{jk} p_j(t), \text{ for } k = \overline{1, N}. \quad (3)$$

with the initial condition $p_1(0) = 1$.

This system of differential equations in open form is:

$$\begin{aligned} \frac{dp_1(t)}{dt} &= -p_1(t) \sum_{i \in e(1)} \alpha_{1i} + \alpha_{21} p_2(t) + \dots + \alpha_{N-1;1} p_{N-1}(t) \\ \frac{dp_2(t)}{dt} &= \alpha_{12} p_1(t) - p_2(t) \sum_{i \in e(2)} \alpha_{2i} + \dots + \alpha_{N-1;2} p_{N-1}(t) \\ &\dots\dots\dots \\ \frac{dp_{N-1}(t)}{dt} &= \alpha_{1;N-1} p_1(t) + \alpha_{2;N-1} p_2(t) + \dots - p_{N-1}(t) \sum_{i \in e(N-1)} \alpha_{(N-1)i} + \alpha_{N;N-1} p_N(t) \\ p_1(t) + p_2(t) + \dots + p_N(t) &= 1 \end{aligned} \quad (4)$$

On the basis of (4) and using Appendix B, one can write the following algebraic system of equations in terms of Laplace transforms:

$$\begin{aligned} -1 - (s + \sum_{i \in e(1)} \lambda_{1j}) \cdot \varphi_1(s) + \lambda_{21} \varphi_2(s) + \dots + \lambda_{k1} \varphi_k(s) &= 0 \\ \lambda_{12} \varphi_1(s) - (s + \sum_{i \in e(2)} \lambda_{2j}) \cdot \varphi_2(s) + \dots + \lambda_{k2} \varphi_k(s) &= 0 \\ &\dots\dots\dots \\ \lambda_{1;N-1} \varphi_1(s) + \lambda_{2;N-1} \varphi_2(s) + \dots - (s + \sum_{i \in e(N-1)} \lambda_{k;j}) \cdot \varphi_k(s) &= q_m^* \\ \sum_{k=1}^N \varphi_k(s) &= \frac{1}{s} \end{aligned} \quad (5)$$

General form of solution of the considered equations system with the help of Laplace transforms is given in Appendix B.

C.1.3. Probability of failure-free operation

For finding the probability of a failure-free operation, absorbing states are introduced into the transition graph. They are the system's failure states.

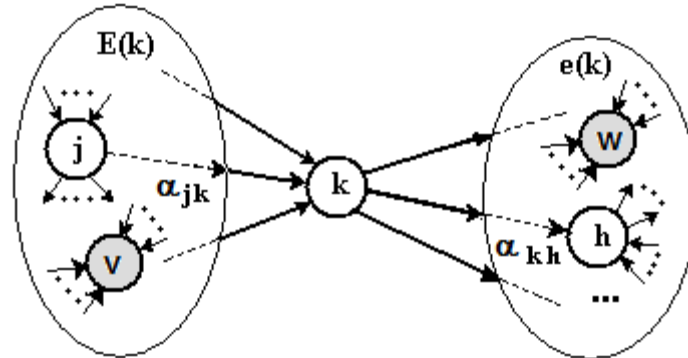


Figure 2. Absorbing states v and w .

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 immediately write for an operational state k formally the same equation as (3).

If the transition graph has M operational states (denote this set of states by Ω), we can construct M differential equations. (In this case, the equations are not linearly dependent. So, there is no need to use the normalization condition as one of equations.) Equations (5) and the initial conditions $p_k(0)$, $k = \overline{1, N}$, are used to find the probability of a failure-free operation of the system.

Actually, there are two special cases that are considered in reliability theory.

(1) At moment $t = 0$ the system is in state $s = (1, 1, \dots, 1)$, i.e. in the state where all units are operational, with probability 1.

(2) In a stationary process the system in arbitrary moment t can be in one of its M operable states with stationary probabilities p_k , $k = \overline{1, M}$.

Thus, systems of differential equations are the same:

$$\frac{dp_1(t)}{dt} = -p_1(t) \sum_{i \in e(1)} \alpha_{1i} + \alpha_{21}p_2(t) + \dots + \alpha_{M,1}p_M(t)$$

where D determinant of the system (13) is, D_k is the same determinant in which the k -th column substituted for the column of the absolute terms. Availability coefficient, K , is:

$$K = \sum_{k \in \Omega} p_k = \frac{1}{D} \sum_{k \in \Omega} D_k, \quad (15)$$

where Ω is the subset of all operational states.

C.1.5. Mean time to failure and mean time between failures

Recall that the mean time is defined as:

$$T = \int_0^{\infty} P(t) dt. \quad (16)$$

Now notice that if $\varphi(s)$ is Laplace transform for $P(t)$, then T can be obtained as:

$$T = \left[\int_0^{\infty} P(t) e^{-st} dt \right]_{s=0} = \varphi(s)|_{s=0} \quad (17)$$

To find MTTF, one has to take solution of equations system (6) with the initial condition (7). For MTBF, one has to take solution of the same equations system with the initial condition (8).

C.1.6. Mean recovery time

Finding this index is analogous to the finding of MTBF, however in this case the sets of transitive and absorbing states have to be redefined. In this case all operable states become absorbing and all failure states become transitive. If the total number of system states equals n , and among them m “up” states, then for remaining $k=n-m$ states (denote the set of these states by g) one has to compile system of k equations:

$$\begin{aligned} -(s + \sum_{\forall j} \lambda_{1j})\varphi_1(s) + \lambda_{21}\varphi_2(s) + \dots + \lambda_{k1}\varphi_k(s) &= q_1^* \\ \lambda_{12}\varphi_1(s) - (s + \sum_{\forall j} \lambda_{2j})\varphi_2(s) + \dots + \lambda_{k2}\varphi_k(s) &= q_2^* \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots & \end{aligned} \quad (18)$$

$$\lambda_{1,m}\varphi_1(s) + \lambda_{2,m}\varphi_2(s) + \dots - (s + \sum_{\forall j} \lambda_{k;j})\varphi_k(s) = q_m^*,$$

where each q_k^* , $k \in g$, is defined as:

$$q_k^* = \frac{p_j}{\sum_{j \in g} p_j}. \quad (19)$$

Solutin of system (18) can be found as

$$\tau = \sum_{j \in g} \varphi_j(0). \quad (20)$$

C.2. Birth-death process

Birth-death process (BDP) is one of the most important special cases of continuous-time homogenous Markov process where the states represent the current size of a population. This process has many applications in demography, biology, queuing theory, reliability engineering, and other areas.

Let us denote states by natural numbers 0, 1, 2, If the process at moment t is in state k , then during infinitesimally small time interval Δ it can with probability $\lambda_k \Delta + o(\Delta)$ proceed to state $(k+1)$, or with probability $\mu_k \Delta + o(\Delta)$ it can proceed to state $(k-1)$, or with probability $1 - (\lambda_k + \mu_k) \Delta + o(\Delta)$ will stay in state k . Notice that states “0” and “n” are so-called reflecting, i.e. $\lambda_0 = 0$ and $\mu_n = 0$. Corresponding transition graph is presented in Figure 3.

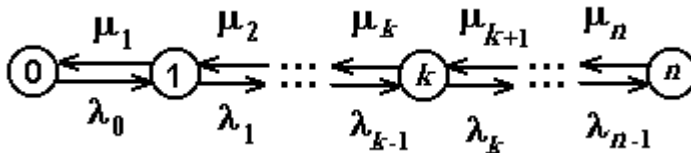


Figure 3. Transition graph for birth-death process.

For state k , one can write the following equation of dynamic balance:

$$p_k(t + \Delta) = p_{k-1}(t)[\lambda_{k-1}\Delta + o(\Delta)] + p_k(t)[1 - (\lambda_k + \mu_k)\Delta + o(\Delta)] + p_{k+1}(t)[\mu_k\Delta + o(\Delta)].$$

(21)

From (21) follows:

$$\frac{p_k(t + \Delta) - p_k(t)}{\Delta} = \lambda_{k-1}p_{k-1}(t) - p_k(t)(\lambda_k + \mu_k) + \mu_k p_{k+1}(t), \quad (22)$$

That gives after limiting passage for $\Delta \rightarrow 0$ the following differential equation:

$$\frac{d}{dt} p_k(t) = \lambda_{k-1}p_{k-1}(t) - p_k(t)(\lambda_k + \mu_k) + \mu_k p_{k+1}(t). \quad (23)$$

In analogous way, one can write equations for states “0” and “n”:

$$\frac{d}{dt} p_0(t) = -\lambda_0 p_0(t) + \mu_1 p_1(t). \quad (24)$$

and

$$\frac{d}{dt} p_n(t) = \lambda_{n-1}p_{n-1}(t) - \mu_n p_n(t). \quad (25)$$

In result, we have the following system of differential equations:

$$\begin{aligned} \frac{d}{dt} p_0(t) &= -\lambda_0 p_0(t) + \mu_1 p_1(t) \\ \frac{d}{dt} p_1(t) &= \lambda_0 p_0(t) - p_1(t)(\lambda_1 + \mu_1) + \mu_2 p_2(t) \\ &\dots\dots\dots \\ \frac{d}{dt} p_k(t) &= \lambda_{k-1}p_{k-1}(t) - p_k(t)(\lambda_k + \mu_k) + \mu_k p_{k+1}(t) \\ &\dots\dots\dots \\ \frac{d}{dt} p_n(t) &= \lambda_{n-1}p_{n-1}(t) - \mu_n p_n(t). \end{aligned} \quad (26)$$

The initial condition in most reliability application is taken in the form $p_0(0) = 1$.

Usually, one is interested in stationary probabilities, when $p_k(\infty) \rightarrow p_k$. It means that all $\frac{d}{dt} p_k(t) \rightarrow 0$ with $t \rightarrow \infty$. In this case the system (26) transforms into the system of algebraic equations:

$$\begin{aligned} 0 &= -\lambda_0 p_0 + \mu_1 p_1 \\ 0 &= \lambda_0 p_0 - (\lambda_1 + \mu_1) p_1 + \mu_2 p_2 \\ &\dots\dots\dots \end{aligned} \quad (27)$$

$$0 = \lambda_{k-1} p_{k-1} - (\lambda_k + \mu_k) p_k + \mu_k p_{k+1}$$

.....

$$0 = \lambda_{n-1} p_{n-1} - \mu_n p_n.$$

In addition, one has to use equation of total probability:

$$\sum_{k=0}^n p_k = 1. \tag{28}$$

Actually, the equations of balance for the birth-death process, written for “cuts” of transition graph, are more convenient.

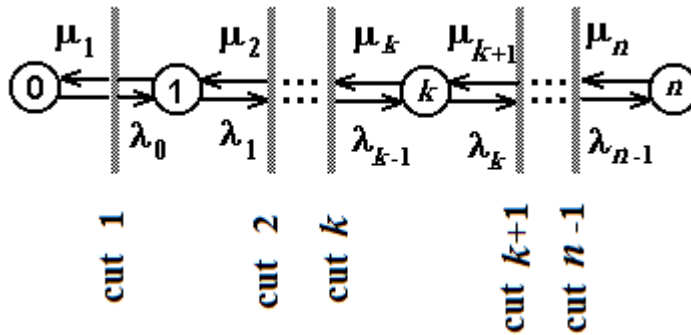


Figure 4. “Cuts” of transition graph for BDP.

Indeed, balance means that flows back and forth through a cut are equal. In this case, the system of balance equations has a very convenient form:

$$\begin{aligned} \lambda_0 p_0 &= \mu_1 p_1 \\ \lambda_1 p_1 &= \mu_2 p_2 \\ \dots\dots\dots & \\ \lambda_{k-1} p_{k-1} &= \mu_k p_k \\ \dots\dots\dots & \\ \lambda_{n-1} p_{n-1} &= \mu_n p_n. \end{aligned} \tag{29}$$

Introducing $\rho_k = \frac{\lambda_k}{\mu_k}$, one gets the same system (29) in the form:

$$\begin{aligned} p_1 &= \rho_1 p_0 \\ p_2 &= \rho_2 p_1 = \rho_1 \rho_2 p_0 \\ \dots\dots\dots \end{aligned}$$

$$p_k = p_0 \prod_{j=1}^k \rho_j \quad (30)$$

.....

$$p_n = p_0 \prod_{j=1}^n \rho_j .$$

Using (28), one can write the solution:

$$p_0 = \left(1 + \sum_{k=1}^n \prod_{j=1}^k \rho_j \right)^{-1} . \quad (31)$$

Thus, for any p_k solution is:

$$p_k = \frac{\prod_{j=1}^k \rho_j}{1 + \sum_{k=1}^n \prod_{j=1}^k \rho_j} . \quad (32)$$

D. General Bibliography

Handbooks

- Kozlov, B., and I. Ushakov. (1966) Brief Handbook of Reliability Calculations for Electronics Equipment, (in *Russian*). Moscow: Sovetskoe Radio.
- Kozlov, B., and I. Ushakov (1970). *Reliability Handbook*. New York: Holt, Rinehart and Winston.
- Kozlov, B., and I. Ushakov. (1975) Handbook of Reliability Calculations for Electronic and Automatic Equipment (in *Russian*). Moscow: Sovetskoe Radio
- Kozlov, B., and I. Ushakov. (1978) Reliability Handbook for Electronic and Automatic Systems (in *German*). Berlin: Verlag Technik.
- Kozlov, B.A., and I.A. Ushakov. (1979) Reliability Handbook for Engineers (in *German*). Springer.
- Ushakov, I. (ed.) (1985). Reliability of Technical Systems: Handbook. (Russian). Radio i Sviyaz.
- Ushakov, I. (ed.) (1989). Prirucka Spolehlivosti v Radioelektronice a Automatizacni technice. (*Check.*). SNTL.
- Rook, P. (1990). Software reliability handbook. Cluver.
- Kececioglu, D. (1991) Reliability engineering handbook. Prentice Hall.
- Ushakov, I.A., ed. (1994) *Handbook of Reliability Engineering*. Wiley.
- Ireson, W., C. Coombs, R. Moss (1995) Handbook of Reliability Engineering and Management. McGraw Hill.
- Kececioglu, D. (1995) Maintainability, availability, and operational readiness engineering handbook. Prentice Hall.
- Pecht, M., ed. (1995), *Product Reliability, Maintainability, and Supportability Handbook* CRC Press.
- Pahl, H. (2003) Handbook on Reliability Engineering. Springer.
- Misra, K. (2008) *Handbook of Performability Engineering* . Springer.
- Stapelberg, R. (2009) Handbook of Reliability, Availability, Maintainability and Safety in Engineering Design

Textbooks & Monographs

- Bazovsky, I. (1961). *Reliability Theory and Practice*. Prentice Hall.
- Lloyd, D.K., and M. Lipow (1962). *Reliability: Management, Methods, and Mathematics*. Prentice-Hall.
- Polovko, A. M. (1964). Fundamentals of Reliability Theory.(Russian). Nauka.
- Barlow, R.E., and F. Proschan (1965) *Mathematical Theory of Reliability*. Wiley.
- Gnedenko, B.V., Yu.K. Belyaev and A.D. Solov'yev (1965). *Mathematical Methods in Reliability Theory*. (Russian). Nauka
- Polovko, A. M. (1968). Fundamentals of Reliability Theory. Academic Press.
- Gnedenko, B.V., Yu.K. Belyaev and A.D. Solov'yev (1969). *Mathematical Methods in Reliability Theory*. Academic Press.
- Amstatter, B.L. (1971). *Reliability mathematics: Fundamentals ;practices; procedures*. McGraw-Hill.
- Barlow, R.E., and F. Proschan (1975) *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston

- Ryabinin I. (1976) Reliability of Engineering Systems. Principles and Analysis. (Russian)
- Kaufmann, A., Grouchko, D., Cruon, R. (1977). *Mathematical models for the study of the reliability of systems*. Academic Press.
- Kapur, K.C., and L.R. Lamberson (1977). *Reliability in Engineering Design*. Wiley.
- Barlow, R.E., and F. Proschan (1981) *Statistical Theory of Reliability and Life Testing*, 2nd ed. To Begin With.
- Gnedenko, B.V., ed. (1983) *Aspects of Mathematical Theory of Reliability* (Russian). Moscow, Radio i Svyaz.
- Osaki, S., Y.Hatoyama. (eds). (1984), *Stochastic Models in Reliability Theory*. Springer.
- Birolini, A. (1985). *On the use of stochastic processes in modeling reliability problems*. Springer.
- Melchers, R.E. (1987). *Structural reliability: Analysis and prediction*. Wiley.
- Rudenko, Yu.N., and I.A. Ushakov (1989). *Reliability of Power Systems* (Russian). Novosibirsk, Nauka.
- Hoyland, A., M. Rausand (1994). *System Reliability Theory: Models and Statistical Methods*. Wiley
- Gnedenko, B.V., and I.A. Ushakov (1995) *Probabilistic Reliability Engineering*. Wiley.
- Barlow, R.E. (1998) *Engineering Reliability*, SIAM.
- Modarres M., M. Kaminskiy and V. Krivtsov (1999). *Engineering and Risk Analysis: A Practical Guide* (2nd ed.), Marcel Dekker.
- Aven, T., and U. Jensen, (1999) *Stochastic models in reliability*, Springer.
- Kolowrocki, K. (2004). *Reliability of Large Systems*. Elsevier.
- Neubeck, K. (2004) *Practical Reliability Analysis*. Prentice Hall.
- Xie Min, Poh Kim-Leng, Dai Yuan-Shun (2004) *Computing System Reliability: Models and Analysis*, Kluwer.
- Bazovsky, I. (2004). *Reliability Theory and Practice*. Dover.
- Zio, E. (2007). *An introduction to the basics of reliability and risk analysis*. World Scientific.
- Epstein, B., and I. Weissman (2008). *Mathematical Models for Systems Reliability*. CRC Press.
- Ushakov, I.A. (2009). *Theory of System Reliability*.(Russian).Moscow, Drofa.
- Tobias, P.A., and D.C. Trindade (2010). *Applied Reliability*, 3rd edition. CRC Press