



IntechOpen

Concise Reliability for Engineers

Authored by Jaroslav Mencik



CONCISE RELIABILITY FOR ENGINEERS

Edited by **Jaroslav Menčík**

Concise Reliability for Engineers

<http://dx.doi.org/10.5772/62009>

Edited by Jaroslav Mencik

Contributors

Jaroslav Menčík

© The Editor(s) and the Author(s) 2016

The moral rights of the and the author(s) have been asserted.

All rights to the book as a whole are reserved by INTECH. The book as a whole (compilation) cannot be reproduced, distributed or used for commercial or non-commercial purposes without INTECH's written permission.

Enquiries concerning the use of the book should be directed to INTECH rights and permissions department (permissions@intechopen.com).

Violations are liable to prosecution under the governing Copyright Law.



Individual chapters of this publication are distributed under the terms of the Creative Commons Attribution 3.0 Unported License which permits commercial use, distribution and reproduction of the individual chapters, provided the original author(s) and source publication are appropriately acknowledged. If so indicated, certain images may not be included under the Creative Commons license. In such cases users will need to obtain permission from the license holder to reproduce the material. More details and guidelines concerning content reuse and adaptation can be found at <http://www.intechopen.com/copyright-policy.html>.

Notice

Statements and opinions expressed in the chapters are those of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in Croatia, 2016 by INTECH d.o.o.

eBook (PDF) Published by IN TECH d.o.o.

Place and year of publication of eBook (PDF): Rijeka, 2019.

IntechOpen is the global imprint of IN TECH d.o.o.

Printed in Croatia

Legal deposit, Croatia: National and University Library in Zagreb

Additional hard and PDF copies can be obtained from orders@intechopen.com

Concise Reliability for Engineers

Edited by Jaroslav Mencik

p. cm.

Print ISBN

Online ISBN 978-953-51-2278-4

eBook (PDF) ISBN 978-953-51-6653-5

We are IntechOpen, the world's largest scientific publisher of Open Access books.

3,250+

Open access books available

106,000+

International authors and editors

112M+

Downloads

151

Countries delivered to

Our authors are among the
Top 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Meet the editor



Prof. Jaroslav Menčík, originally mechanical engineer, has spent many years in technological and materials research at various institutions and universities in Czech Republic, Australia, Germany, UK, Japan, and other places. He is the author of two monographs on mechanical properties and many papers in scientific journals and conferences, with more than 1200 quotations. Since joining the University of Pardubice in 1996 he has been active in reliability engineering. He headed several grant projects and international scientific conferences on the topic. His lectures on reliability have been attended by students from various countries and branches: mechanical, civil and electrical engineering, transport technology, economy. With this extensive experience he decided to prepare a concise book for students and other people who want to get some insight into reliability, regardless their professional orientation. In writing, he put therefore emphasis on universal rules and methods. Jaroslav Menčík is married and has a daughter and a son.

Contents

Preface XIII

Section 1 Reliability Basics 1

Chapter 1 **Basic Terms of Reliability 3**
Jaroslav Menčík

Chapter 2 **Probability Basics 7**
Jaroslav Menčík

Chapter 3 **Characteristics of Reliability 19**
Jaroslav Menčík

Chapter 4 **Bathtub Curve 27**
Jaroslav Menčík

Chapter 5 **Reliability of Systems 33**
Jaroslav Menčík

Chapter 6 **Time to Failure of Deteriorating Objects 43**
Jaroslav Menčík

Chapter 7 **Maintenance 53**
Jaroslav Menčík

Chapter 8 **Diagnostics 57**
Jaroslav Menčík

Chapter 9 **Failure Analysis 61**
Jaroslav Menčík

- Chapter 10 **Approaches to Ensuring the Reliability and Safety** 71
Jaroslav Menčík
- Section 2 Advanced Reliability** 79
- Chapter 11 **Weibull Distribution** 81
Jaroslav Menčík
- Chapter 12 **Failure Modes and Effects Analysis** 89
Jaroslav Menčík
- Chapter 13 **Fault Tree Analysis and Reliability Block Diagrams** 97
Jaroslav Menčík
- Chapter 14 **Load – Resistance Interference Method** 101
Jaroslav Menčík
- Chapter 15 **Monte Carlo Simulation Method** 107
Jaroslav Menčík
- Chapter 16 **Latin Hypercube Sampling** 117
Jaroslav Menčík
- Chapter 17 **Economy of Reliability** 121
Jaroslav Menčík
- Chapter 18 **Reliability Assessment with a Small Amount of Data** 133
Jaroslav Menčík
- Chapter 19 **Robust Design, Sensitivity Analysis, and Tolerance Setting** 143
Jaroslav Menčík
- Chapter 20 **Reliability Testing and Verification** 155
Jaroslav Menčík
- Chapter 21 **Multicriterial Condition Evaluation and Fuzzy Methods** 167
Jaroslav Menčík
- Chapter 22 **Bayesian Methods** 173
Jaroslav Menčík

Section 3	Epilogue – Ways to Higher Reliability	181
Chapter 23	Epilogue – Ways to Higher Reliability Jaroslav Menčík	183
Chapter 24	Sources of Information on Reliability Jaroslav Menčík	189
Chapter 25	Standards Related to Reliability Jaroslav Menčík	195
Chapter 26	Software for Reliability Analysis Jaroslav Menčík	201

Preface

Reliability, in general, is the ability to fulfill the demanded tasks. Our life is often influenced by the reliability or unreliability of the things we use, such as home appliances, machines or cars, and by the reliability of processes and services, such as the supply of electricity, telephone services, or keeping the timetable of transport means. Even the people we meet can be considered as reliable or unreliable; everyone has such experience.

The opposite of reliability is unreliability and failures. Failures mean losses, such as costs of repairs and losses due to dropout of production, and also fatalities or injuries and damages to property or the environment as well as sometimes the loss of good reputation of the manufacturer, which can even contribute to his bankruptcy.

In this respect, reliability is related closely to safety and also to quality. The quality of a product or service is its ability to ensure customer satisfaction. And reliability is nothing else than the ability to keep the quality in time. It would be of little use to buy a nice and powerful car if it would fail the next day and have to stay for several weeks in a workshop. Nobody would say that a product is of high quality if it fails several times during short time.

These days, much is spoken about quality, but it is actually reliability that must be aimed at. During the last 100 years, attention has been devoted to reliability and its increase. The main factors influencing reliability have been revealed and the foundations of the reliability theory have been laid out. In addition, practical techniques have been developed for achieving high reliability. All this has brought its fruits. For example, the warranty time of cars has increased from 6 months, as common 60 years ago, to 6 years today. This could not be possible without the increase of reliability thanks to the systematic effort in this direction. And this trend should continue.

Today, the methods of reliability engineering are taught in numerous courses and textbooks. These are often limited only to certain branches (e.g. reliability of machines, electric appliances, software, civil engineering structures, or reliability of services). However, many rules and methods are universal and applicable in various areas of technology, as well as in our life. Yes, even human life can be considered from the reliability point of view, and life insurance companies use the mathematics common for reliability calculations. Moreover, as the occurrence of failures is accompanied by uncertainty, the methods for increasing reliability are, in principle, suitable for the reduction of any uncertainties.

The author of this book, originally a mechanical engineer, has spent many years in research. This is also an area with many uncertainties. Later, at the University of Pardubice, he has been teaching reliability engineering and solving related problems. His lectures are attended by students from various countries and branches: mechanical, civil, and electrical engineer-

ing, transport technology, and economy. With this extensive experience, he decided to prepare a concise book for students and others who want to get some insight into reliability, regardless of their professional orientation. Therefore, he put emphasis on reliability rules and methods useful for various branches. This book is no cookbook with recipes for the solution of a narrow group of problems. It explains the basic principles and universal methods of wide applicability, as it is well known that progress can also be achieved by the transfer of ideas and methods from one branch to another.

The book is divided into two parts. The first part (Chapters 1.1 to 1.10) explains the basic terms and simple methods for the determination of reliability characteristics, which form the base for any reliability evaluation. To understand the contents of this section, no special knowledge is necessary. In the second part (Chapters 2.1 to 2.12), more advanced methods are explained, such as failure modes and effects analysis, load-resistance interference method, the Monte Carlo simulation method, cost-based reliability optimization, basic approaches to reliability testing, and methods based on Bayesian approach or fuzzy logic suitable for the processing of rather vague information. The practical examples included help in understanding the individual topics. All examples can be solved without a special software; Excel or even a pocket calculator is sufficient.

The book is complemented with information on the standards for reliability evaluation, software for reliability, sources of information on reliability, list of references, and an index.

The author wanted to write a brief book that can serve as an introduction to reliability and to the study of a special literature on this topic. He wishes that this book brings the readers pleasure and helps them in increasing the reliability of anything.

Acknowledgments

During the preparation of this book, I made use of the experience gained during my work and teaching at various institutions. I am grateful to my colleagues for their advice as well as to my students for their questions. Special thanks belong to Prof. Hynek Šertler and Prof. Milan Lánský, who oriented me in reliability issues after I joined the University of Pardubice; Prof. Dietrich Munz of the Research Centre of Karlsruhe (KIT); and Prof. Milan Holický of the Klokner Institute of the Czech Technical University. I am also very grateful to the Group for Reliability at the Czech Society for Quality, Prague, and its head, Prof. Václav Legát of the Czech University of Life Sciences in Prague, for the permanent effort to cultivate a sense of reliability in the society.

Prof. Jaroslav Menčík
University of Pardubice
Dopravní fakulta Jana Pernera, KMMČS
Czech Republic

Reliability Basics

Basic Terms of Reliability

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62354>

Abstract

Basic terms are explained, such as reliability, failure, fault, limit state, quality, safety, repair, renewal, maintenance, availability and dependability, inherent and operational reliability.

Keywords: Reliability, safety, failure, limit state; repair, renewal, maintenance, availability, dependability

In matters of reliability, people from various branches — manufacturers, customers, technicians, and lawyers — must often communicate together, especially when trying to find the answers to the following questions: “What happened and why?”, “Who is guilty of this accident?”, “Who should pay the damages caused by the failure?”, or “How should the warranty be defined?”. Therefore, it is necessary that all participants understand certain technical terms in the same way. The most important expressions are explained in this section; more rigorous definitions can be found in the standards, such as ISO, IEC, and others, as listed in Appendix 2.

Reliability, in general, is the ability of an object (or process or service) to fulfil the demanded tasks and meet the specifications under given conditions. The specifications (i.e. technical parameters) must be written in the accompanying technical documentation. The conditions of use must also be specified, for example, the temperature range, in which the object will keep the assumed parameters.

Reliable operation is interrupted or terminated by failures. **Failure** is an event leading to the loss of the ability to fulfil the demanded tasks and meet the specifications. Examples are fracture of a component due to overloading or fatigue, collapse of a structure, loss of electric contact, unacceptable deformations or wear, or some parameters out of the allowable limits.

Fault is a defect in the component or product. Faults can also be present in software.

Limit state. From the reliability point of view, every object can be either in an operational state or in a failed state. The border between both states is the limit state. Some objects fail suddenly and the failure is complete. The technical condition of other objects becomes gradually worse. For some time, however, they are able to fulfil their purpose, though in limited extent (with worse technical parameters or lower safety); the failure is partial. However, if certain parameters exceed the corresponding limit values, the object gets either unfit for further use or destroyed.

Reliability is a component of quality. **Quality** is the ability of a product or service to ensure full customer satisfaction. The quality of an object (e.g. a car) can be judged with respect to several criteria, such as power, maximum speed, noise, fuel consumption, and color. Examples of quality in services are the possibility to provide quick connection to the Internet or to keep the timetable of trains or buses (and also the cleanliness in the vehicles, etc.). Reliability, in essence, is the ability to keep the quality with time. Nobody would say that a car is of good quality if it fails repeatedly during a short time.

A further important characteristic is safety. **Safety** is the ability of an object not to endanger the human health or life, the environment, and properties.

If an object fails, it must be repaired. The word **repair** denotes the works for restoring full operability after a failure. However, the situation is more complex. Only some items are repaired, and we call them **repairable**. Some objects are **unrepairable** (e.g. electric lamp bulbs). If the filament has burned, the bulb must be replaced by a good one. Some items are repairable (e.g. an electric motor), but in some cases they are not repaired after failure but **replaced** by good ones, just to reduce the downtime of the object in which they were mounted (e.g. a gear box in a locomotive). If necessary, the repair of the failed part can be done at a suitable time elsewhere. For some mass-produced items, their replacement by new ones is cheaper than repair. Therefore, one often speaks of **repaired** or **unrepaired** objects.

A more general expression than repair is **renewal**, which means putting the object back into the condition as if it “were new”. Renewal can be achieved by repair, replacement, or maintenance.

Maintenance comprises cleaning, exchange of oils and dirty filters, tightening of locked screws and other adjustments, and repair of minor faults or damaged paints – generally small works for restoring full operability.

Availability is the readiness for correct service. In complex structures with long life (e.g. a car or a locomotive), failures inevitably appear from time to time. In this case, the term availability is used, which considers the times of operation as well as of repairs. Later, it will be shown how availability can be calculated and used for reliability evaluation.

All these terms together form the so-called **dependability**. This is a general term used to characterize availability and factors that influence it: reliability, maintainability, and maintenance support. Briefly,

- availability = readiness for correct service,

- reliability = continuity of correct service (i.e. without failures),
- safety = absence of catastrophic consequences for the users and the environment,
- maintainability = ability of the object to be maintained (and repaired), and
- maintenance support = technical and organizational means enabling maintenance and repairs.

In this context, reliability has narrower meaning than in the aforementioned definition.

When speaking of reliability, two kinds should be distinguished: inherent and operational.

Inherent reliability is the reliability “built-in” into the object during the design stage by using an appropriate concept, materials and dimensions, and also by suitable conditions of manufacture or assembly. There are many pieces of evidence that it is the design that is most important for achieving high reliability. Some means will be shown later.

Operational reliability is the reliability achieved in operation. It depends on the way of operation and maintenance. Reckless operation and poor maintenance can significantly shorten the life of every object. However, neither very good maintenance can mitigate the faults and weaknesses caused by improper design.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

Probability Basics

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62355>

Abstract

The main concepts of probability theory are explained, such as probability, random quantity, population, sample, mean, average, standard deviation, coefficient of variation, probability density, distribution function, quantile, critical value, confidence interval and testing of hypotheses. Important probability distributions are also shown.

Keywords: Probability, sample, mean, standard deviation, distribution function, quantile, confidence interval, testing of hypotheses

The occurrence of failures is usually accompanied by some uncertainty. This is due to many factors that we cannot control, and call them therefore **random**. Similarly, we speak about **random events**, which can happen or not, depending on random influences. For their prediction, we use the concept of **probability** and the related methods. However, before these methods will be explained, several words are addressed here to those who have no or little knowledge of this topic. There are also methods that can improve reliability without probability tools, e.g. Failure Mode and Effect Analysis, which will be explained later. Nevertheless, such methods are suitable only in some cases, whereas the formulas based on probability can facilitate the solution of many reliability problems. Because computers can do all the necessary work, the only thing a user of probabilistic methods needs is some understanding of the basic terms and concepts. The following pages will try to help him or her.

Probability is a quantitative measure of the possibility that a random event occurs. The simplest definition of probability P is based on the occurrence of an event in a numerous repetition of a trial:

$$P = n / N, \tag{1}$$

where N is the total number of trials and n is the number of trials with a certain outcome (e.g. a tossed coin with the eagle on the top, the number of days with the maximum temperature higher than 20°C , or the number of defective components). Probability is a dimensionless quantity that can attain values between 0 and 1; zero denotes the impossible event and 1 denotes a certain event. A **random variable** is a variable that can attain various values with certain probabilities. Random quantities are **discrete** and **continuous**. Examples of **discrete random quantities** are the number of failures during a certain time, number of vehicle collisions, number of customers in a queue or number of their complaints, and number of faulty items in a batch. **Continuous random quantities** can attain any value (in some interval), such as strength of a material, wind velocity, temperature, diameter, length, weight, time to failure (expressed in hours, kilometers, loading cycles, or worked pieces), duration of a repair, and, also, probability of failure. Examples are depicted in Figure 1.

Random quantities can be described by **probability distribution** or by single numbers, called **parameters**, if they are related to the population (i.e. the set of all possible members or values of the investigated quantity), or **characteristics**, if they are calculated from a sample of a limited size. Parameters are usually denoted by Greek letters and characteristics are denoted by Latin letters.

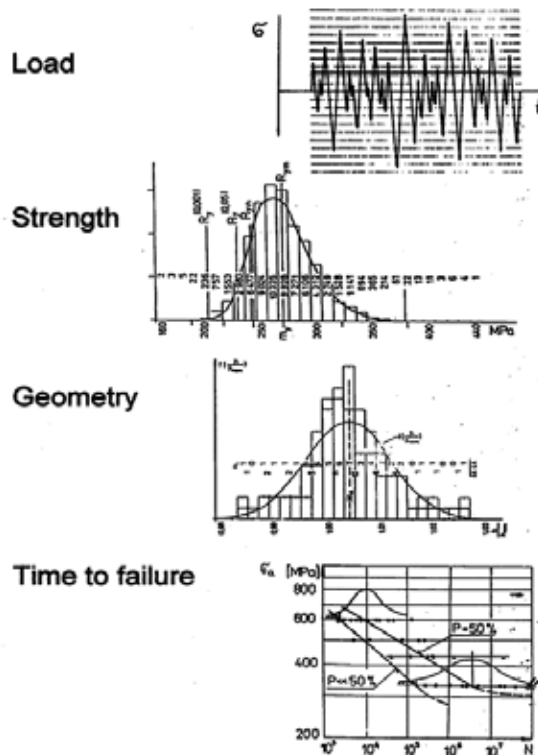


Figure 1. Examples of random quantities.

1. Description by parameters

The main parameters (or characteristics) of random quantities are given below, with the formulas for calculation from samples of limited size.

Mean μ (or **average value**) characterizes the position of the quantity on numerical axis; it corresponds to its centroid,

$$\bar{x} = \frac{\sum x_j}{n} \quad (2)$$

x_j is the j th value and n is the size of the sample. The summation is done over all n values.

Variance σ^2 (or s^2) characterises the dispersion of the quantity, and is calculated as

$$s^2 = \sqrt{\frac{\sum (x_j - \bar{x})^2}{n - 1}} \quad (3)$$

Standard deviation σ (or s) is defined as the square root of scatter,

$$s = \sqrt{\frac{\sum (x_j - \bar{x})^2}{n - 1}} \quad (4)$$

It has the same dimension as the investigated variable x , and therefore it is used more often than scatter.

Coefficient of variation ω (or v) characterizes the relative dispersion, compared to the mean value,

$$v = \frac{s}{\bar{x}} \quad (5)$$

It can thus be used for the comparison of random variability of various quantities.

A disadvantage of the average value is its sensitivity to the extreme values; the addition of a very high or very low value can cause its significant change. A less sensitive characteristic of the "mean" of a series of values is **median m** . This is the value in the middle of the series of data ordered from minimum to maximum (e.g. $m = 4$ for the series 2, 6, 1, 8, 10, 4, 3).

2. Description by probability distribution

A more comprehensive information is obtained from probability distribution, which informs how a random variable is distributed along the numerical axis. For discrete quantities,

probability function $p(x)$ is used (Fig. 2), which expresses the probabilities that the random variable x attains the individual values x^* ,

$$p(x^*) = P(x = x^*) . \quad (6)$$

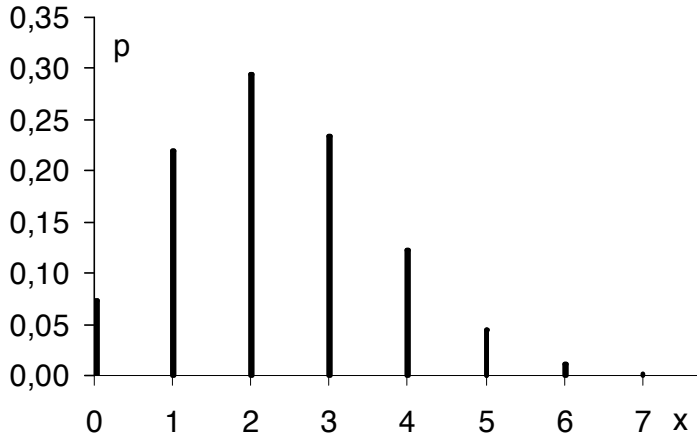


Figure 2. Binomial distribution. (An example; parameters: $p = 0.23$, $n = 10$.)

Probability density $f(x)$ is used for continuous quantities and shows where this quantity appears more or less often (Fig. 3a). Mathematically, it expresses the probability that the variable x will lie within an infinitesimally narrow interval between x^* and $x^* + dx$.

Distribution function $F(x)$ is used for discrete as well as continuous quantities (Fig. 3b) and expresses the probability that the random variable x attains values smaller or equal to x^* :

$$F(x^*) = P(x \leq x^*) . \quad (7)$$

These functions are related mutually as

$$f(x) = dF/dx, \quad F(x) = \int_{-\infty}^x f(x) dx, \quad \text{or } F(x) = \sum_{i=1}^n p(x_i) . \quad (8)$$

Figure 3 shows two possibilities for depicting these functions: by histograms or by analytical functions. Histograms are obtained by dividing the range of all possible values into several intervals, counting the number of values in each interval and plotting rectangles of heights proportional to these numbers. To make the results more general, the frequencies of occurrence in individual intervals are usually divided by the total number of all events or values. This

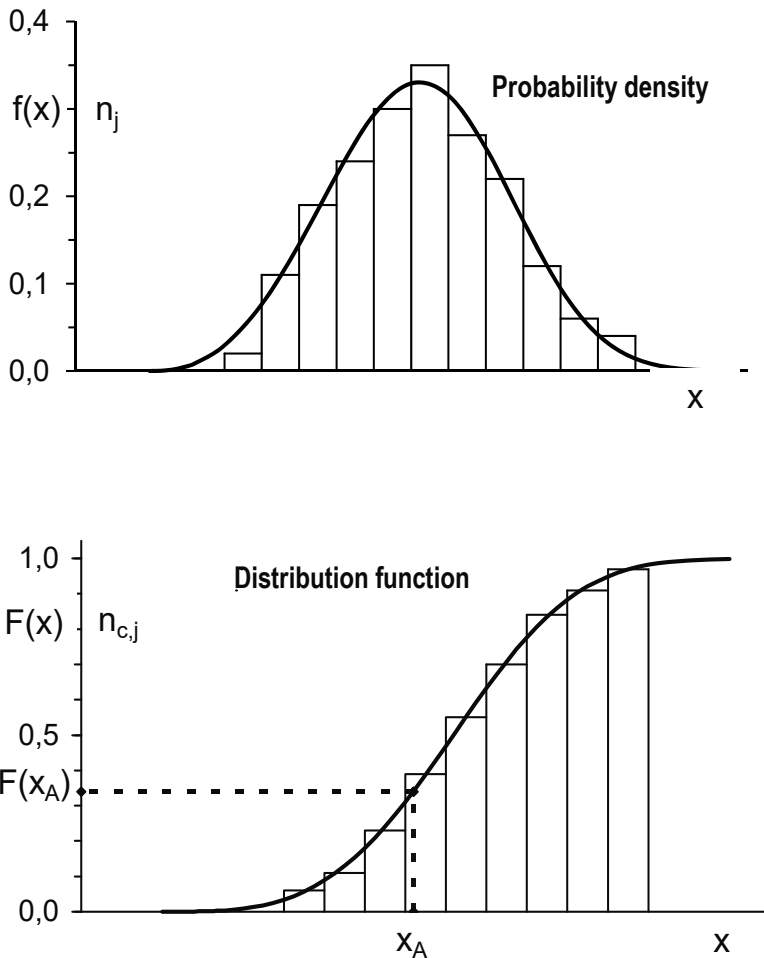


Figure 3. (a) Probability density $f(x)$ and (b) distribution function $F(x)$ of a continuous quantity.

gives relative frequencies (a) or relative cumulative frequencies (b), which approximately correspond to probabilities.

Fitting such histogram by a continuous analytical function gives the probability density or distribution function (solid curves in Fig. 3).

The probability of some event (e.g. snow height x lower than x_A) can be determined as the corresponding area below the curve $f(x)$ or, directly, as the value $F(x_A)$ of the distribution function.

Also very important are the following two quantities.

Quantile is such value of the random quantity x , that the probability of x being smaller (or equal) to is only α ,

$$P(x \leq x_\alpha) = \alpha. \quad (9)$$

Quantiles are inverse to the values of distribution function (Fig. 3b),

$$x_\alpha = F^{-1}(\alpha), \quad (10)$$

and are used for the determination of the “guaranteed” or “safe” minimum value of some quantity, such as the minimum expectable strength or time to failure.

Critical value (Fig. 3b) is such value of the random quantity x , that the probability of its exceeding is only β ,

$$P(x > x^\beta) = \beta. \quad (11)$$

The critical values are used for the determination of the expectable maximum value of some quantity, such as wind velocity or maximum height of snow in some area. They are also used for hypotheses testing, for example whether two samples come from the same population. Probability β is complementary to α ; $\beta = 1 - \alpha$,

$$x_\alpha = x^{1-\beta}, \quad x^\beta = x_{1-\alpha}. \quad (12)$$

More about the basic probability definitions and rules can be found, for example, in [1 – 5].

3. Probability distributions common in reliability

Several probability distributions exist, which are especially important for reliability evaluation. For discontinuous quantities, it is binomial and Poisson distribution. The main distributions for continuous quantities used in reliability are normal, lognormal, Weibull, and exponential. For some purposes also, uniform distribution, Student’s t -distribution, and chi-square (χ^2) distribution are used. The brief descriptions follow; more details can be found in the special literature [1 – 5].

Binomial distribution (Fig. 2) gives the probability of occurrence of x positive outcomes in n trials if this probability in each trial equals p . An example is the number of faulty items in a sample of size n if their proportion in the population is p . The probability function is

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad (13)$$

and the mean value is $\mu = np$. This distribution is discrete and has only one parameter p , which can be determined from the total number m of positive outcomes in n trials as $p = m/n$.

Poisson distribution is similar to binomial distribution but is better suitable for rare events with low probabilities p . The probability function giving the probability of occurrence of x positive outcomes in n trials is

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (14)$$

λ is the distribution parameter that corresponds to the average occurrence of x (and, in fact, to the product np of binomial distribution.)

Normal distribution, called also Gauss distribution, resembles a symmetrical bell-shaped curve (Fig. 4). It is used very often for continuous variables, especially if the variations are caused by many random factors and the scatter is not too big (cf. the central limit theorem). The probability density is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (15)$$

with the mean μ and standard deviation σ as parameters. There is no closed-form expression for the distribution function $F(x)$; it must be calculated as the integral of the probability density, cf. Equation (8). In practice, various approximate formulas are also used.

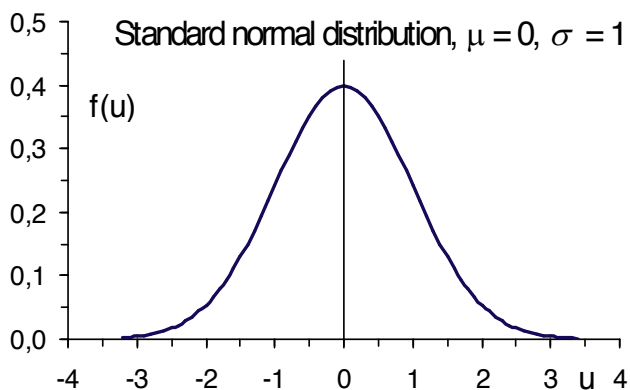


Figure 4. Normal distribution (probability density).

Standard normal distribution corresponds to normal distribution with parameters $\mu = 0$ and $\sigma = 1$ (Fig. 4). The expression for probability density is usually written as

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2), \quad (16)$$

u is the standardised variable, related to the variable x of the normal distribution as

$$u = (x - \mu) / \sigma . \quad (17)$$

It expresses the distance of x from the mean as the multiple of standard deviation. It is useful to remember that 68,27% of all values of normal distribution lie within the interval $(\mu \pm \sigma)$, 95,45% within $(\mu \pm 2\sigma)$, and 99,73% within $(\mu \pm 3\sigma)$.

Log-normal distribution is asymmetrical (elongated towards right, similar to Weibull distribution with $\beta = 2$ in Fig. 5) and appears if the logarithm of random variable has normal distribution.

Weibull distribution (Fig. 5) has the distribution function

$$F(t) = 1 - \exp \{ - [(t - t_0) / a]^b \} , \quad (18)$$

with three parameters: scale parameter a , shape parameter b , and threshold parameter t_0 that corresponds to the minimum possible value of x . The probability density $f(x)$ can be obtained easily as the derivative of distribution function. Weibull distribution is very flexible, thanks to the shape parameter b (Fig. 5). It is often used for the approximation of strength or time to failure. It belongs to the family of **extreme value distributions** [5, 6] and appears if the failure of the object starts in its weakest part. The determination of parameters of this very important distribution will be explained in Chapter 11.

Exponential distribution is a special case of Weibull distribution for shape parameter $b = 1$, cf. Fig. 5, with the distribution function

$$F(t) = 1 - \exp (t / T_0) , \quad (19)$$

which may be used, for example, for the times between failures caused by many various reasons and also in complex systems consisting of many parts. This distribution has only one parameter, T_0 , which corresponds to the mean μ and has the same value as the standard deviation σ .

The following three distributions are important especially for the determination of confidence intervals, for statistical tests, and for the Monte Carlo simulations, as it will be shown later.

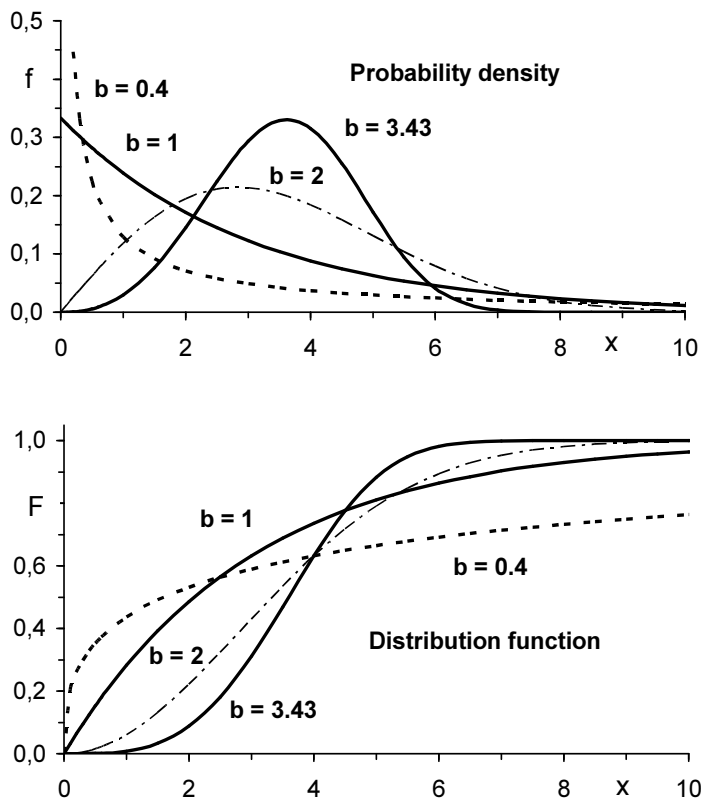


Figure 5. Weibull distribution for various values of shape parameter b .

Uniform distribution has constant probability density, $f = \text{const}$, in the interval $\langle a; b \rangle$, so that it looks like a rectangle. The mean value is the average of both boundaries, $\mu = (a + b)/2$, and the scatter equals $\sigma^2 = (b - a)^2/12$.

χ^2 distribution is a distribution of the sum of n quantities, each defined as the square of standard normal variable. An important parameter is the number of degrees of freedom. For more, see [1–5].

t -distribution (or Student's distribution) arises from a combination of χ^2 and standard normal distribution. It looks similar to normal distribution but also depends on the number of degrees of freedom; see [1–5].

The values of distribution functions and quantiles of the above distributions can be found via special tables or using statistical or universal computer programs, such as Excel.

Finally, two important probabilistic concepts should be explained.

Confidence interval. A consequence of random variability of many quantities is that every measurement or calculation gives a different result depending on the used specimen or input

value. Thus, the average $= \Sigma x_i/n$ is usually determined from several (n) values for obtaining a more definite information. This, however, does not say how far the actual mean μ can be from it. For this reason, confidence interval is often determined, which contains (with high probability) the actual value. For example, the confidence interval for the mean is

$$\bar{x} - t_{\alpha, n-1} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha, n-1} \frac{s}{\sqrt{n}}, \quad (20)$$

and s are the average and standard deviation of the sample of n values, and $t_{\alpha, n-1}$ is the α -critical value of two-sided t -distribution for $n - 1$ degrees of freedom. The probability that the true mean μ will lie inside the interval (20), is $1 - \alpha$. Confidence intervals can also be determined for other quantities.

A note. Also one-sided critical values exist. Such value (α') corresponds to the probability that the t -value will be either higher or lower than the pertinent critical value. α' is related to α as $\alpha' = \alpha/2$. When using statistical tables or computer tools one must be aware how was the pertinent quantity defined.

Testing of hypotheses. Often, one must decide which of the two products or technologies is better. The decision can be based on the value of the characteristic parameter (e.g. the mean). However, the values of individual candidates usually differ. If the differences are not big, one must consider that a part of the variability of individual values is due to random reasons. Statistical tests can reveal whether the differences between characteristic values of both compared samples are only random or if they reflect a real difference between both types of products. The value of the pertinent test criterion is calculated from basic statistical characteristics of each sample and compared with the critical value (of the probability distribution) of this criterion. If the calculated value is larger than the unlikely low critical value, we conclude that the difference is not random. If it is smaller, we usually conclude that there is no substantial difference between both populations. These tests are explained in the literature [1–5] and available in various statistical or universal programs. Also Excel offers several tests (e.g. for the difference between the mean values or scatters of two populations).

Example 1

The diameters of machined shafts, measured on 10 pieces, were $D = 16.02, 15.99, 16.03, 16.00, 15.98, 16.04, 16.00, 16.01, 16.01,$ and 15.99 mm, respectively. Calculate: (a) the average value and the standard deviation. Assume that the diameters have normal distribution, and calculate (b) the 95% confidence interval for the mean value μ_D and also (c) the interval, which will contain 95% of all diameters.

Solution

- a. The average value is $= \Sigma D_i/n = 16.007$ mm and standard deviation $s = 0.01889$ mm.
- b. Confidence interval for the mean, calculated by Eq. (20), is (with two-sided critical value $t_{0.05; 10-1} = 2.2622$):

$$16.007 - 2.2622 \frac{0.01889}{\sqrt{10}} < \mu_D < 16.007 + 2.2622 \frac{0.01889}{\sqrt{10}}$$

$$15.993 < \mu_D < 16.020 \text{ mm} .$$

- c. The individual values can be expected (under assumption of normal distribution) within the interval $-u_{\alpha/2} \times s < d < +u_{\alpha/2} \times s$, where $u_{\alpha/2}$ is $\alpha/2$ - critical value of standard normal distribution (corresponding to probability $\alpha/2$ that the diameter will be larger than the upper limit of the confidence interval, and $\alpha/2$ that it will be smaller than the lower limit). In our case, $u_{0.025} \approx 1.96$, so that $16.007 - 1.96 \times 0.01889 < D < 16.007 + 1.96 \times 0.01889$; that is $D \in (15.970; 16.044)$. The reliability of prediction could be increased if tolerance interval were used instead of confidence interval; cf. Chapter 18.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
 University of Pardubice, Czech Republic

References

- [1] Freund J E. Modern elementary statistics. 6th ed. Englewood Cliffs, New Jersey: Prentice-Hall; 1981. 561 p.
- [2] Freund J E, Perles B E. Modern elementary statistics. 12th ed. New Jersey: Prentice-Hall; 2006. 576 p.
- [3] Suhir E. Applied Probability for Engineers and Scientists. New York: McGraw-Hill; 1997. 593 p.
- [4] Montgomery D C, Runger G C. Applied Statistics and Probability for Engineers. 4th ed. New York: John Wiley; 2006. 784 p.
- [5] Rao S S. Reliability-Based Design. New York: McGraw-Hill; 1992. 569 p.
- [6] Gumbel J E. Statistics of Extremes. New York: Columbia University Press; 1958. 375 p.

Characteristics of Reliability

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62356>

Abstract

The basic reliability characteristics are explained: time to failure, probability of failure and of failure-free operation, repairable and unrepairable objects. Mean time to repair and between repairs, coefficient of availability and unavailability, failure rate. Examples for better understanding are included.

Keywords: Time to failure, mean time to failure, mean time between failures, mean time to repair, availability, unavailability, failure rate

Reliability is usually characterized by the probability of failure or by the time to failure. If failure is considered as a single event (e.g. collapse of a bridge), regardless of the time, only its probability is of interest. If we want to know when the failures can occur, their time characteristics are also important. In this chapter, time-dependent failures will be dealt with. Here, one must distinguish between unrepaired and repaired objects depending on whether the failed object is discarded or repaired and again put into service.

1. Unrepaired objects

The basic quantity for unrepaired objects is the **time to failure** t_f . If a group of identical objects is put into operation, the individual pieces begin to fail after some time, and it is also possible to express the number of failed pieces as a function of time, $n_f(t)$. A more universal quantity is the relative proportion of the failed items, that is, the number of the failed items related to the number n of monitored objects, $n_f(t)/n$. This ratio approximately expresses the **probability of failure** $F(t)$ during the time interval $\langle 0; t \rangle$;

$$F(t) = \int_0^t f(t)dt \approx n_f(t)/n \tag{1}$$

Function $F(t)$ is the **distribution function of the time to failure**, also called **failure function** (Fig. 1a). An aid for easier remembering: the letter F is also the first letter of the word failure.

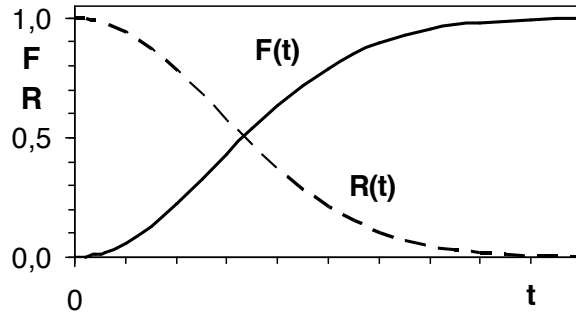


Figure 1. (a) Failure function $F(t)$ and (b) reliability function $R(t)$.

The **probability of failure-free operation $R(t)$** expresses the probability that no failure occurs before the time t ;

$$R(t) = \int_t^\infty f(t)dt \approx [n - n_f(t)]/n \tag{2}$$

$R(t)$ shows the gradual loss of serviceable objects (Fig. 1b) and is called **reliability function** (therefore the symbol R). R is complementary to F ,

$$R(t) + F(t) = 1. \quad R = 1 - F, \quad F = 1 - R. \tag{3}$$

The **probability density of the time to failure, $f(t)$** , expresses the probability of failure during a very short time interval Δt at time t , related to this interval:

$$f(t) = \frac{dF(t)}{dt} \approx \frac{n_f(t + \Delta t) - n_f t}{n \Delta t}; \tag{4}$$

the unit is s^{-1} or h^{-1} . The right-hand side of Equation (4) indicates how the probability density can be determined from empirical data, $n_f(t + \Delta t)$ expresses the number of failed parts from 0

to $t + \Delta t$, and $n_f(t)$ is the number of failures that occurred until the time t . In fact, probability density $f(t)$ shows the distribution of failures in time, similar to Fig. 3a in Chapter 2.

Useful information on reliability is obtained from a very simple characteristic, the average or **mean time to failure** or *MTTF*, which is generally defined as

$$MTTF = \bar{t} = t_{mean} = \int_0^{\infty} t f(t) dt = \int_0^{\infty} R(t) dt \quad (5)$$

The mean time to failure can be calculated from operational records as the average of the group of measured times to failure,

$$MTTF = (1/n) \sum t_{f,j} \quad (6)$$

Remark: Equation (6) is appropriate if all objects have failed. If components with very long life are tested, the tests are usually terminated after some predefined time or after failure of certain fraction of all components. In such cases, modified formulas for *MTTF* must be used; see Chapter 20 or [1].

2. Repaired objects

If a repairable item fails, it is repaired and again put into operation. After the next failure, it is again repaired and put into operation, etc. One can thus speak of a flow of operations and repairs (Fig. 2). If we denote each interval as “uptime” t_{up} or “downtime” t_{down} , we can calculate the **mean time between failures**, *MTBF*, and the **mean time to repair**, *MTTR*:

$$MTBF = (1/n) \sum t_{up,j} \quad (7)$$

$$MTTR = (1/n) \sum t_{down,j} \quad (8)$$

If data for a high number of values t_{up} and t_{down} are available, the distribution of these times can also be obtained and used.

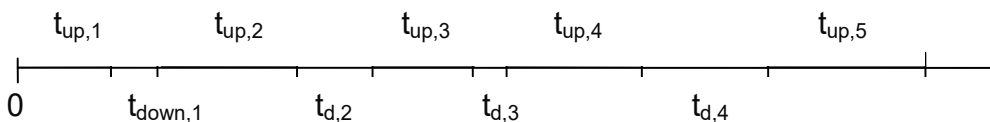


Figure 2. Flow of operations (uptimes, t_{up}) and repairs (downtimes, t_d).

The mean time between failures and mean time to repair can be used to characterize the probability that the object will be serviceable at a certain instant or not. The **coefficient of availability, COA**, is defined as [2, 3]:

$$COA = \sum t_{up} / t_{tot} = \sum t_{up} / (\sum t_{up} + \sum t_{down}), \quad (9)$$

where $\sum t_{up}$ is the sum of times of operation during the investigated interval (e.g. 1 month or year), $\sum t_{down}$ is the sum of down times in this interval, and t_{tot} is the total investigated time. The coefficient of availability can also be calculated as

$$COA = MTBF / (MTBF + MTTR); \quad (10)$$

MTBF is the mean time (of operation) between failures and *MTTR* is the mean time to repair (generally, the mean down time caused by failures).

The coefficient of availability simply says what part of the total time is available for useful work. It also expresses the average probability that the object will be able to fulfill the expected task at any instant.

The complementary quantity, **coefficient of unavailability**,

$$COU = \sum t_{down} / (\sum t_{up} + \sum t_{down}) = MTTR / (MTBF + MTTR) = 1 - COA, \quad (11)$$

says how many percent of the total time are downtimes. It also expresses the probability that the object will not be able to perform its function at a demanded instant. For example, $COA = 0.9$ means that, on average, the vehicle (or machine) is only 90% of all time in operation, and 10% of the total time it is idle due to failures. In other words, there is a 90% probability that the object will be available when needed and a 10% probability that it will not be available. Even the simple records from operation can give the basic values of probabilities and reliability.

It must be reminded here that the time of a repair is not always the same as the downtime when the object (e.g. a machine) does not work. In addition to the net time of the repair, some logistic times are often necessary, which sometimes last much longer than the repair.

3. Failure rate

A very important reliability characteristic is **failure rate $\lambda(t)$** . Basically, failure rate expresses the probability of failure during a time unit. Unlike probability, which is nondimensional, failure rate has a dimension. It is t^{-1} , for example, h^{-1} or % per hour for machines, components, or appliances, km^{-1} for vehicles, etc. Two cases must be distinguished, depending on whether the object after failure is repaired or not.

Unrepaired objects

The failed item is discarded. This is typical of simple unrepairable objects, such as lamp bulbs, screws, windows, integrated circuits, and many inexpensive parts. Also, a living being cannot be repaired, if it has died. Some objects could be repaired after failure but are not, because of economic reasons. Thus, the term nonrepaired objects can be used as more universal.

Failure rate expresses the probability of failure during a time unit but is related only to those objects that have remained in operation until the time t , that is, those that have not failed before the time t . Failure rate is defined as

$$\lambda(t) = f(t) / R(t) ; \tag{12}$$

$f(t)$ is the probability density of failure ($=dF/dt$), and $R(t)$ is the probability that the object was operated until the time t . An illustrative idea of failure rate can be gained from a simple formula for its calculation from the data from operation:

$$\lambda(t) = [n_f(t + \Delta t) - n_f(t)] / \{[(n - n_f(t))\Delta t]\} ; \tag{13}$$

n is the total number of the monitored objects, $n_f(t)$ is the number of the objects failed until the time t , $[n_f(t + \Delta t) - n_f(t)]$ is the number of objects failed during the time from t to $t + \Delta t$, and Δt is a short time interval. [Remark: Formula (13) is only approximate and often exhibits big scatter. A more accurate value of the instantaneous failure rate $\lambda(t)$ can be obtained from several n_f values occurring in a wider interval around the time t .]

The fraction of failed objects, $F(t)$, increases with time, and the fraction of objects that have not failed, $R(t)$, decreases.

Equation (1) relates mutually three variables: λ , f , and R . Fortunately, it can be transformed into simple relationships of two quantities. First, it can be rewritten as follows:

$$\lambda(t) = f(t) / R(t) = [dF(t) / dt] / R(t) = - [dR(t) / dt] / R(t) . \tag{14}$$

The separation of the variables leads to the differential equation of first order,

$$\lambda(t) dt = - dR(t) / R(t) . \tag{15}$$

The integration and transformation lead to the following expression for the probability of operation as a function of time:

$$R(t) = \exp\left(-\int_0^t \lambda(t) dt\right) . \tag{16}$$

The probability of failure is

$$F(t) = 1 - R(t). \quad (17)$$

With respect to Equations (12) to (17), any of the four quantities f , F , R , and λ is sufficient for the determination of any of the remaining three quantities.

The mean time to failure can be calculated using Equation (5).

Repaired objects

After a failure, the object is repaired and continues working. In complex systems, the failed part can also be replaced by a good one to reduce the downtime. The number of working objects remains constant, so that $R(t) = 1$. Failure rate (1) thus corresponds to the failure probability density, $\lambda(t) = f(t)$. In this case, the term **hazard rate** is used as more appropriate, but the expression failure rate is also very common.

Example 1

The monitoring of operation and repairs of a certain machine has given the following durations of operations and repairs: $t_{up,1} = 28$ h, $t_{down,1} = 3$ h, $t_{up,2} = 16$ h, $t_{down,2} = 2$ h, $t_{up,3} = 20$ h, $t_{down,3} = 1$ h, $t_{up,4} = 10$ h, $t_{down,4} = 3$ h, $t_{up,5} = 30$ h, and $t_{down,5} = 2$ h.

Tasks.

1. Determine the mean time between failures and mean time to repair.
2. Determine the coefficient of availability (COA) and unavailability (COU).
3. Express the average probability (in %) that the machine (a) will be able to work at any instant (R) and (b) will not be able to work (F).

Solution

1. Mean time between failures $MTBF = \sum t_{up,j}/n = (28 + 16 + 20 + 10 + 30)/5 = 104/5 = 20.8$ h.
Mean time to repair $MTTR = \sum t_{down,j}/n = (3 + 2 + 1 + 3 + 2)/5 = 11/5 = 2.2$ h.
2. Coefficient of availability $COA = \sum t_{up,j}/t_{tot} = \sum t_{up,j}/(\sum t_{up,j} + \sum t_{down,j}) = 104/(104 + 11) = 0.90435$.
Coefficient of unavailability: $COU = \sum t_{down,j}/t_{tot} = 11/(104 + 11) = 0.09565$. Another way of calculation: $COA = MTBF/(MTBF + MTTR) = 20.8/(20.8 + 2.2) = 0.90435$ and $COU = MTTR/(MTBF + MTTR) = 2.2/(20.8 + 2.2) = 0.09565$.
3. The probability that the machine will be able to work at any instant equals the coefficient of availability; $R = COA = 0.90435 \approx 90,4\%$. Similarly, $F = COU = 0.09565 \approx 9,6\%$.

Example 2

In a town, $N = 30$ buses are necessary for assuring reliable traffic on 15 routes. However, due to failures and maintenance, several buses are unavailable every day. As it follows from long-

term records, the mean availability of the buses is $COA = 0.85$. How many reserve buses (N_r) are necessary? What is the total necessary number of buses N_{tot} ?

Solution

The coefficient of availability can be calculated as the number of operable buses, N_{up} , divided by the total number of vehicles, $COA = N_{up}/N_{tot}$ from which $N_{tot} = N_{up}/COA$. With the above numbers, $N_{tot} = 30/0.85 = 35.29$. To reliably ensure the public traffic, 36 buses are thus necessary. The number of reserve vehicles is $36 - 30 = 6$ buses.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

References

- [1] Bednařík J et al. Reliability techniques in electronic practice (In Czech: Technika spolehlivosti v elektronické praxi). Praha: SNTL; 1990. 336 p.
- [2] Bentley J P. Introduction to Reliability and Quality Engineering. Harlow, England: Addison-Wesley; 1999. 202 p.
- [3] Ireson W G, Coombs C F Jr, Moss R Y. Handbook of reliability engineering and management. 2nd ed. New York: McGraw-Hill; 1996. 816 p.

Bathtub Curve

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62357>

Abstract

Typical time course of failure rate of unrepaired objects, called bathtub curve, is shown and its main stages are explained: period of early failures, useful life, and period of aging and deterioration. Attention is paid to the useful-life period, where the failure rate is constant and the distribution of times to failure (or between failures) is exponential. Illustrative examples are included.

Keywords: Failure rate, bathtub curve, early failures, steady-state operation, period of aging, exponential distribution

Failure rate, as defined in Chapter 3, can change with time. Figure 1 shows the time course of $\lambda(t)$ typical of nonrepairable objects, such as electrical bulbs, pumps, switches, or springs, and also living beings, including humans. Such course can be obtained if the operation of a high number of objects of the same kind is monitored. Due to its shape, resembling a longitudinal section of a bathtub, the curve has got the nickname **bathtub curve**. It can be divided into three stages with characteristic time courses related to different reasons of failures.

Stage I. Failure rate $\lambda(t)$ is high at the beginning and decreases with time. The failures occur due to errors in design, weak components or inferior materials, due to faults appearing during manufacture or building, or due to mistakes caused by an inexperienced personnel or user. A weak newborn baby more easily succumbs to an infectious disease. Software errors also belong to this category. The failed components are discarded and not used any more, the customer gradually becomes familiar with the use of a product, and the errors in software are corrected. This period is called the stage of **early failures** or **infant mortality**.

Stage II. Failure rate λ is low and approximately constant. In contrast to early failures, caused by the inherent weakness of the object, the failures during stage II occur mostly due to external reasons, such as overloading, collision with another object, weather or natural catastrophes,

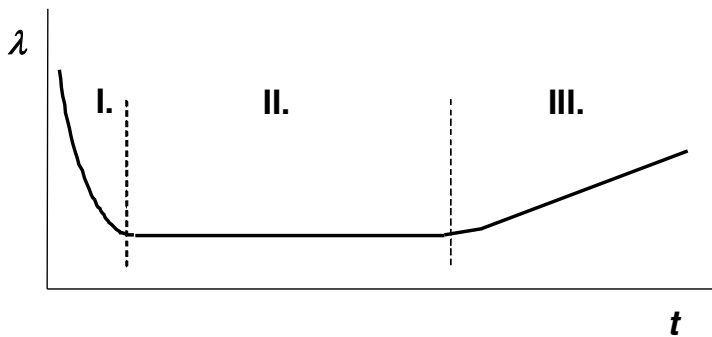


Figure 1. Bathtub curve (a schematic). $\lambda(t)$ – failure rate; t – time. I – stage of early failures; II – steady state, useful life, III – wear-out period.

hidden defects, and mistakes of the personnel. (In the case of people, the reasons for the “failures” during this stage are traffic accidents, diseases, wars, and murders.) Depending on the object and conditions, failure rates for various objects can be very different. Stage II represents the major part of the life and is called the **useful life** or the period of **steady-state operation**.

Stage III. Failure rate $\lambda(t)$ increases with time. The failures in this stage are caused by wear, fatigue, corrosion, or gradual deterioration of the material, for example due to UV radiation (plastics) or ozone (rubber). This period is called the **wear-out period** or **aging**.

Figure 1 shows the general shape of the time course of failure rate. In reality, various patterns of $\lambda(t)$ can occur. Today, many advanced products, when put into operation, have the failure rate constant from the beginning, without the period of early failures. This can be achieved by using high-quality materials and reliable components admitted only after entrance tests and by excluding potentially risky solutions as early as in the design stage, thanks to computer modeling and the simulation of various design solutions and conditions of operation. Also, thorough controls and checks during manufacture or building are an efficient tool for avoiding early failures or significant reduction of their number. Examples are cars, TV sets, washing machines, and other consumer goods. In the past, the so-called burning-in period was used for some products before putting them into operation. During this period, the objects were some time switched on, often under somewhat higher voltage or load, so that the weaker components failed during this period, before the object was sold to the customer and put into service. Today, thanks to special tests and the use of high-quality components, the burning-in period is not necessary. An evidence of the generally better situation today is the significantly longer warranty time provided by the manufacturers of many products.

Also, stage III, the wear-out period, can be avoided for more complex objects if their technical condition is monitored and the critical parts approaching stage III are replaced in time by new ones. This case belongs to repairable objects. The “bathtub curve” here consists only of periods I and II (early failures and useful life) or even only period II (steady-state operation).

Remark: The failures from external reasons can happen at any time; the instantaneous resultant failure rate equals the sum of failure rates from all reasons.

Special case: $\lambda = \text{const.}$

This is a very important case, as constant failure rate can often be assumed (approximately) for the prevailing period of useful life (stage II in Fig. 1). With $\lambda = \text{const.}$, the probability of failure during the interval $\langle 0; t \rangle$ follows from Equations (15) and (16) in Chapter 3:

$$F(t) = 1 - R(t) = 1 - \exp\left(-\int_0^t \lambda dt\right) = 1 - \exp(-\lambda t) \quad (1)$$

The reliability (i.e. the fraction of serviceable objects) decreases with time as

$$R(t) = \exp(-\lambda t), \quad (2)$$

The distribution of times to failure is exponential with the probability density

$$f(t) = dF(t) / dt = \lambda \exp(-\lambda t) \quad (3)$$

and the mean value

$$MTTF = \bar{t} = t_{\text{mean}} = \int_0^{\infty} t f(t) dt = 1 / \lambda. \quad (4)$$

Vice versa, the failure rate of some kind of components can be obtained from the mean time to failure,

$$\lambda = 1 / t_{\text{mean}}. \quad (5)$$

The time course of reliability may thus also be expressed as

$$R(t) = \exp(-t / t_{\text{mean}}) ; \quad (6)$$

note that the argument in the exponential function is nondimensional.

The mean time to failure (and also the mean time between failures) can be calculated by Equation (4). With $\lambda = \text{const.}$,

$$MTTF = t_{mean} = \int_0^{\infty} [\exp(-\lambda t) dt] = 1 / \lambda. \quad (7)$$

The empirical determination of the mean time to failure is based on the testing or monitoring of a group of components of the same kind and measuring their times to failure,

$$MTTF = t_{mean} = (1 / n) \sum t_j; \quad (8)$$

the summation is done for all n tested objects. The mean failure rate is obtained easily as

$$\lambda = 1 / MTTF. \quad (9)$$

In design, the knowledge of failure rate λ of a component, found from the manufacturer's catalog or by measurement, enables the determination of the mean time to failure, which is important for the determination of the overall reliability of more complex systems (cf. Chapter 5).

Exponential distribution is typical of systems consisting of many elements, where failures happen from various reasons, as usual in electric or electronic appliances. However, one should not forget that the period with constant failure rate often becomes dominant only after some time t_0 from putting the system into operation. In such cases, the time t in Equation (6) must be replaced by $t - t_0$.

Note: One must always keep in mind that the mean time between failures, calculated as the reciprocal value of failure rate, has nothing in common with the mean time to failures caused by aging or fatigue. Failure rate given in catalogs is determined from the period of steady-state operation. For example, a high-quality component has a failure rate $\lambda = 10^{-6} \text{ h}^{-1}$. However, this does not mean that these components will work until $t_f = 1/\lambda = 10^6 \text{ h}$. They fail after a much shorter time, for example after 10,000 h, when they enter stage III (wear out).

Example 1

A device should work 2 h without failure, and such operation should be 99% guaranteed. (There may be only 1% probability of failure during this time.) Assume that you can choose from various devices available in the market. What are the demanded failure rate and the mean time to failure of a suitable device? Assume exponential distribution of the time to failure.

Solution. The probability of failure-free operation is $R(t) = \exp(-\lambda t)$. Taking logarithms gives $\ln R = -\lambda t$, from which the demanded failure rate is $\lambda = -(1/t) \ln R$.

For the demanded $t = 2 \text{ h}$ and $R = 0.99$, the necessary failure rate is $\lambda = -(1/2) \ln 0.99 = 0.005025 \approx 0.005 \text{ h}^{-1}$. The demanded mean time to failure is $MTTF = 1/\lambda = 1/0.005 = 200 \text{ h}$ or more.

Example 2

A ventilator (air fan) has exponential distribution of times to failure with the mean time $MTTF = 10,000$ h. Calculate the probability that the ventilator does not fail during the first 800 h after being put into operation. What is the probability of failure during this time?

Solution.

Probability of not failing: $R(t) = \exp(-t/t_{\text{mean}}) = \exp(-800/10,000) = 0.923$ (=92.3%).

Probability of failure: $F(t) = 1 - R(t) = 1 - 0.923 = 0.077$ (=7.7%).

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

Reliability of Systems

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62358>

Abstract

Many objects consist of more components. The mutual arrangement of the individual elements influences the resultant reliability. The formulae are shown for the resultant reliability of series arrangement, as well as for parallel and combined arrangement. The possibility of reliability increasing by means of redundancy is explained, and also the principle of optimal allocation of reliabilities to individual elements. Everything is illustrated on examples.

Keywords: Reliability, systems, series system, parallel system, probability of failure, time to failure, failure rate, redundancy, reliability allocation

Many objects consist of more parts or elements. From reliability point of view, an element is any component or object that is considered in the investigated case as a whole and is not decomposed into simpler objects. An element can be a lamp bulb, the connecting point of two electric components, a screw, an oil hose, a piston in an engine, and even the complete engine in a diesel locomotive. Also, the individual operations or their groups in a complex manufacturing or building process can be considered as elements.

An example of a simple system is an electric lamp made by a light bulb, socket, switch, wires, plug, and the lamp body. An extremely complex system is an aircraft, containing tens of thousands of mechanical, hydraulic, or electric elements. Each of them can fail. This increases the probability that the whole system fails. The resultant reliability depends on the reliability of the individual elements and their number and mutual arrangement. A suitable arrangement can even increase the reliability of the system. In this chapter, important cases will be shown together with the formulas for the calculation of resultant reliability. Two basic systems are series and parallel, and their combinations are also possible.

1. Series system

From reliability point of view, a series system (Fig. 1a) is such, which fails if any of its elements fails. For example, a motorcycle cannot go if any of the following parts cannot serve: engine, tank with fuel, chain, frame, front or rear wheel, etc., and, of course, the driver. All these elements are thus arranged in series. Elements are also screws and many other things. If failure of any component does not depend on any other component, the reliability of the system is obtained simply as the product of the reliabilities of individual elements,

$$R = R_1 \times R_2 \times R_3 \times \dots \times R_n = \prod R_j. \quad (1)$$

A practical conclusion is that “the reliability of a series system is always lower than the reliability of any of its components”.

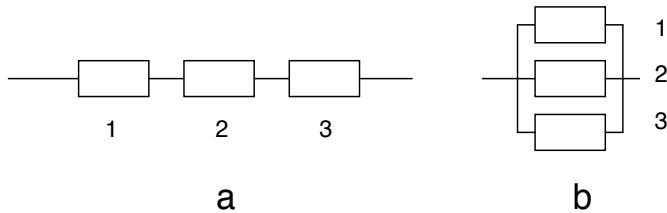


Figure 1. Examples of series system (a) and parallel system (b).

The probability of failure is complementary to reliability, i.e.

$$F = 1 - R. \quad (2)$$

The characteristic features of series arrangement will be shown on several examples.

Example 1

The resultant reliability of two components is $R = R_1 \times R_2$. For example, if $F_1 = 0.1$ and $F_2 = 0.2$, then $R_1 = 0.9$ and $R_2 = 0.8$ and $R = 0.9 \times 0.8 = 0.72$. This is less than the reliability of the weaker component no. 2. The probability of failure has increased to $1 - 0.72 = 0.28$, i.e. more than the failure probability F_2 .

Example 2

The reliability of a series system with three elements with $R_1 = 0.9$, $R_2 = 0.8$, and $R_3 = 0.5$ is $R = 0.9 \times 0.8 \times 0.5 = 0.36$, which is less than the reliability of the worst component ($R_3 = 0.5$). This reminds of the well-known saying “The chain is as weak as its weakest link” (which, however, does not consider that several components can fail simultaneously).

Example 3

The influence of the number of elements (and thus complexity of the system) can be illustrated on several systems where all components have the same probability of failure $F_1 = 0.02$; the corresponding reliability $R_1 = 0.98$. What will be the reliability of a system composed of (a) 2 components, (b) 10 components, (c) 50 components, and (d) 200 components?

Solution: (a) $R = R_1 \times R_1 = 0.98^2 = 0.960$; (b) $R = R_1^{10} = 0.98^{10} = 0.817$; (c) $R = R_1^{50} = 0.98^{50} = 0.364$; and (d) $R = R_1^{200} = 0.98^{200} = 0.0176$.

One can see that the drop of reliability is significant especially for high numbers of components. Although one component has relatively high reliability (98%), a system with 200 such parts is practically unable to work, as it has reliability lower than approximately 2% and probability of failure 98%! Complex large systems must therefore be assembled from very reliable elements.

Until now, we have assumed that the reliability of individual parts does not change with time. If it varies, Equation (1) changes to

$$R(t) = R_1(t) \times R_2(t) \times R_3(t) \times \dots \times R_n(t) = \prod R_j(t) ; \quad (3)$$

the resultant probability of failure is obtained as

$$F(t) = 1 - R(t) , \text{ or } F(t) = 1 - \prod [1 - F_j(t)] . \quad (4)$$

The reliability of components is often characterized by failure rate λ . If the failure rate may be assumed constant (especially in systems containing many elements), the decrease of reliability with time is exponential, $R(t) = \exp(-\lambda t)$, and Equation (3) changes to

$$\begin{aligned} R(t) &= \exp(-\lambda_1 t) \times \exp(-\lambda_2 t) \times \exp(-\lambda_3 t) \times \dots \times \exp(-\lambda_n t) = \\ &= \exp [-(\lambda_1 + \lambda_2 + \lambda_3 + \dots + \lambda_n)t] = \exp(-\lambda t) . \end{aligned} \quad (5)$$

The distribution of times to failure of such system is again exponential, with the resultant failure rate equal the sum of individual failure rates,

$$\lambda = \sum \lambda_i . \quad (6)$$

This means that "the failure rate of a series system is always higher (and the mean time between failures shorter) than that of individual components, and the reliability $R(t)$ decreases with time faster".

The mean time between failures is

$$MTBF = 1 / \lambda. \quad (7)$$

The decrease of reliability with time is illustrated in Figure 2 for several systems with different numbers of elements. One can see a very fast drop of reliability in systems with many components. This must be accounted for if guaranteed operation of a complex object during certain time is demanded. This issue will be treated in detail later.

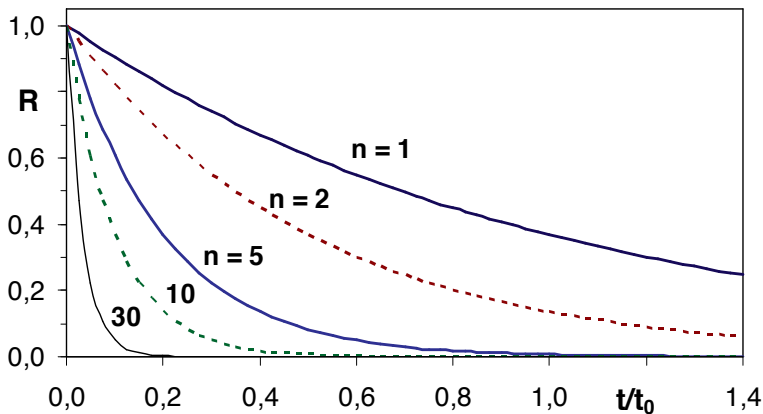


Figure 2. Series system. Time course of reliability for various number of elements n .

2. Parallel system

A parallel system (Fig. 1b) is such, which fails only if all its parts fail. An example is a four-cylinder engine. It will fail only if all four cylinders are unable to run. If one, two, or even three cylinders do not work, the fourth one is still able to put the car into motion (though with significantly reduced power).

The probability of a simultaneous occurrence of mutually independent events equals the product of individual probabilities. In parallel systems, the resultant probability of failure is thus calculated as

$$F(t) = F_1(t) \times F_2(t) \times F_3(t) \times \dots \times F_n(t) = \prod F_j(t). \quad (8)$$

Reliability is complementary to probability of failure, i.e.

$$R(t) = 1 - F(t), \text{ or } R(t) = 1 - \Pi[1 - R_i(t)]. \quad (9)$$

For example, if two components are arranged in parallel, each with reliability $R_1 = R_2 = 0.9$, that is, $F_1 = F_2 = 0.1$, the resultant probability of failure is $F = 0.1 \times 0.1 = 0.01$. The resultant reliability is $R = 1 - 0.01 = 0.99$. The probability of failure has thus dropped 10 times. This feature is sometimes used for reliability increasing by using redundant parts (see later).

If the reliability of elements is characterized by failure rates, the situation is more complex than in a series system, even if the failure rates of the individual elements are constant. For the simplest case of two components, with $R_1(t) = \exp(-\lambda_1 t)$ and $R_2(t) = \exp(-\lambda_2 t)$,

$$\begin{aligned} F(t) &= F_1(t) \times F_2(t) = [1 - R_1(t)] \times [1 - R_2(t)] = \\ &= [1 - \exp(-\lambda_1 t)] \times [1 - \exp(-\lambda_2 t)] = \\ &= 1 - \exp(-\lambda_1 t) - \exp(-\lambda_2 t) + \exp[-(\lambda_1 + \lambda_2)t], \end{aligned} \quad (10)$$

and

$$R(t) = 1 - F(t) = \exp(-\lambda_1 t) + \exp(-\lambda_2 t) - \exp[-(\lambda_1 + \lambda_2)t]. \quad (11)$$

The distribution is no more exponential and the failure rate is not constant. The mean time to failure is

$$MTTF = \int_0^{\infty} R(t) dt = \int_0^{\infty} [\exp(-\lambda_1 t) + \exp(-\lambda_2 t) - \exp[-(\lambda_1 + \lambda_2)t]] dt = \lambda_1^{-1} + \lambda_2^{-1} - (\lambda_1 + \lambda_2)^{-1} \quad (12)$$

For identical components, with $\lambda_1 = \lambda_2 = \lambda$,

$$MTTF = \lambda^{-1} + \lambda^{-1} - (\lambda + \lambda)^{-1} = (3/2)\lambda^{-1} = (3/2)MTTF, \quad (13)$$

i.e. by 50% longer than the mean time to failure of individual components.

The solution for parallel systems with more elements can be obtained in similar way. However, it is much more complicated. Analytical solutions exist only in very simple cases; more effective is the use of the Monte Carlo simulation method, explained in Chapter 15.

Generally, the reliability of parallel arrangement can be characterized as follows:

“The probability of failure-free operation of a system with several parallel elements is always higher than that of the best element in the system.” The situation is depicted in Figure 3. Also,

the mean time to failure of a parallel system is always longer than that of any of its parts. For this reason, parallel arrangement is sometimes used to increase reliability (see further).

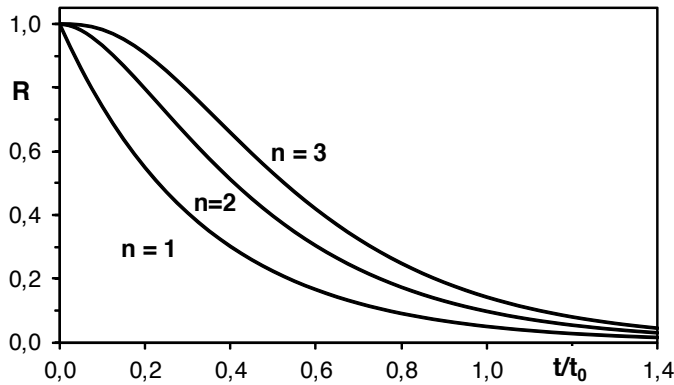


Figure 3. Parallel system. Time course of reliability for various number of elements n .

3. Combined arrangement

In some systems, series and parallel arrangements of elements appear together (Fig. 4). The resultant reliability can be found using step-by-step solution and gradual simplification. The group of elements arranged in series is replaced by one element with equivalent reliability parameters. Parallel elements can sometimes also be replaced by an equivalent element, and so on. The situation is easier if the time dependency of reliabilities does not need to be considered. Unfortunately, if reliability is characterized by failure rates, the failure rate for parallel arrangement is not constant and no simple and accurate analytical solutions exist, only approximate. Better results can be obtained using numerical simulation methods.

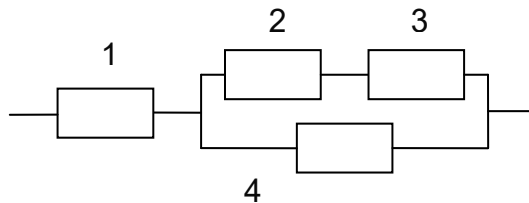


Figure 4. Combined system.

4. Redundancy

Reliability can be increased if the same function is done by two or more elements arranged in parallel. This is called redundancy. Two kinds of redundancy can be distinguished: structural and algorithmic. **Structural redundancy** uses more components for the same purpose. Examples include dual-circuit brakes in modern cars, a reserve water pump in a power plant, joining of two load-carrying parts using more rivets than necessary for safe transfer of the load, a spare electric generator to ensure safe power supply in a hospital, or a reserve electric line. **Redundancy** can be **active** (the parallel elements work or are loaded simultaneously) or **standby**. In the latter case, only one element is loaded or works, whereas the second (third, etc.) redundant element is switched on just if the first one has failed. The advantage of standby redundancy is that only one component is loaded and exposed to wear or other kinds of deterioration. A disadvantage is that such arrangement usually needs a **switch** or similar item, which increases the costs and can also contribute to the unreliability of the system.

The second case is algorithmic redundancy. This means the repetition of some operations, for example measurement or check for defects in some kinds of nondestructive control, such as X-ray or ultrasonic revealing of internal defects in castings or fatigue cracks in airframes or wings, as well as the proofreading of a paper for finding errors. Algorithmic redundancy is commonly used in the transmission of signals and information, from the simple addition of parity bits (check digits) to complex systems for safe information coding.

5. Reliability allocation

Until now, we determined the resultant reliability of a system composed of more components. In the design of complex systems, an opposite problem appears: what should be the reliabilities of individual parts so that the reliability of the whole system is equal to some demanded value (or better)? Several methods of reliability allocation were proposed. The simplest one for series systems uses **equal apportionment**, which distributes the reliability uniformly among all members. If the resultant reliability should be R and the system consists of n components in a series, each of the reliability R_i , then it follows from Equation (1) that $R = R_i^n$, so that every single element should have the reliability

$$R_i = R^{1/n}. \quad (14)$$

If failure rates are considered, then the failure rate λ_i of every element should be

$$\lambda_i = \lambda / n, \quad (15)$$

where λ is the demanded failure rate of the system.

Also other apportionments are possible. Not always has each available component the reliability R_i or λ_i corresponding exactly to Equation (14) or (15). Such values can serve as a guide for finding the parameters so that the resultant reliability (1), (3), or (6) fulfills the requirements. In the reliability allocation, other criteria can also be considered, such as the importance of individual parts.

Example 4

A system consists of three parallel components (Fig. 1b) with probabilities of failure (during a certain, unspecified time): $F_1 = 0.08$, $F_2 = 0.20$, and $F_3 = 0.20$. Calculate the resultant probability of failure (F) and of failure-free operation (R). Assume that the components are independent.

Solution. In parallel systems, $F = F_1 \times F_2 \times F_3 = 0.08 \times 0.20 \times 0.20 = 0.0032$. $R = 1 - F = 1 - 0.0032 = 0.9968$. (Compare the results with the failure probabilities of individual components!)

Example 5

Calculate the mean time to failure and failure rate of a system consisting of four elements in a series (like in Fig. 1a). The individual elements have exponential distribution of the time to failure with failure rates $\lambda_1 = 8 \times 10^{-6} \text{ h}^{-1}$, $\lambda_2 = 6 \times 10^{-6} \text{ h}^{-1}$, $\lambda_3 = 9 \times 10^{-6} \text{ h}^{-1}$, and $\lambda_4 = 2 \times 10^{-5} \text{ h}^{-1}$. Calculate the probability of failure (in %) during the time $t = 500$ hours of operation.

Solution.

$$\lambda = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = (8 + 6 + 9 + 20) \times 10^{-6} = 43 \times 10^{-6} \text{ h}^{-1}.$$

$$MTTF = 1 / \lambda = 1 / 43 \times 10^{-6} = 23,256 \text{ h}.$$

$$F(t) = 1 - \exp(-\lambda t) = 1 - \exp(-43 \times 10^{-6} \times 500) = 0.9787 = 97.87\%. R = 1 - F = 2.13\%.$$

Example 6

Calculate the resultant probability of failure (F) and failure-free operation (R) for a combined series-parallel system (Fig. 4). Assume that the components are independent. The failure probabilities of individual elements are: $F_1 = 0.08$, $F_2 = 0.30$, $F_3 = 0.20$, and $F_4 = 0.10$.

Solution. The system must be solved step-by-step. First, the reliability of elements 2 and 3 in a series is calculated: $R_{2-3} = R_2 \times R_3 = (1 - F_2) \times (1 - F_3) = (1 - 0.3) \times (1 - 0.2) = 0.7 \times 0.8 = 0.56$. The probability of failure is complementary to reliability, so that $F_{2-3} = 1 - R_{2-3} = 1 - 0.56 = 0.44$. Then, the reliability of this F_{2-3} group arranged in parallel with element 4 is obtained as $F_{4,2-3} = F_4 \times F_{2-3} = 0.10 \times 0.56 = 0.056$. The resultant reliability of the whole system is obtained as the reliability of component 1 in a series with the subsystem 4,2-3. Here, the reliabilities must be multiplied. The resultant reliability thus is

$$\begin{aligned} R &= R_1 \times R_{4,2-3} = (1 - F_1) \times (1 - F_{4,2-3}) = \\ &= (1 - 0.08) \times (1 - 0.056) = 0.92 \times 0.944 = 0.86848. \end{aligned}$$

The resultant probability of failure is $F = 1 - R = 1 - 0.86848 = 0.13152 \approx 0.13$.

Example 7

The failure rate of a system of five components arranged in a series should be $\lambda = 2.0 \times 10^{-5} \text{ h}^{-1}$. Determine the failure rate of individual components provided that all can have the same λ_i .

Solution. The resultant failure rate of this series system is $\lambda = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5$. For identical components, it is $\lambda = 5\lambda_i$. The demanded failure rate of each part is $\lambda_i = \lambda/5 = 2.0 \times 10^{-5} / 5 = 4.0 \times 10^{-6} \text{ h}^{-1}$.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

Time to Failure of Deteriorating Objects

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62359>

Abstract

This chapter explains the prediction of the time to failure in the following cases: fatigue of metallic components under cyclic loading or in the presence of cracks, static fatigue, wear and creep, variable loading (damage accumulation). Prediction of the time to failure based on monitoring of the changing response. Probabilistic aspects of the lifetime prediction. The determination of the time to failure is illustrated on examples.

Keywords: Failure, time to failure, fatigue, static fatigue, wear, creep, damage accumulation, prediction, monitoring

As we have seen in Chapter 4, failure rate often changes with time (Fig. 1 there). This is because the main causes of failures change with time. The early failures (stage I) are mostly caused by errors in design, manufacture, assembly, or building process or due to hidden defects, and the instant of their occurrence cannot be predicted. The failures in stage II (useful life) arise from external reasons (random overloading, collision with another object, climatic events, and errors of personnel) and also cannot be predicted. Only if the pertinent failure rate is known one can predict approximately how often failures can be expected and take suitable measures to mitigate their effects.

The failures in stage III (aging and wear-out) arise due to the internal “weakness” of the object and appear after some time of operation even under appropriate conditions of use. Many objects fail due to wear, fatigue, creep, corrosion, or other processes of gradual deterioration. Fortunately, in such cases a possibility exists to predict (with higher or lower accuracy) the time when the object is about to fail, provided that the relationship between the load intensity and the rate of deterioration is known. Two principal ways exist:

1. If the basic mechanism of degradation is known, one can express the deterioration rate as a function of the characteristic load and then derive a formula for the calculation of the

time to failure. For the known load, the time to failure can thus be predicted in advance (e.g. during the design stage). Vice versa, it is possible to find such dimensions of loaded parts, which will guarantee an acceptably low rate of deterioration and thus the demanded life.

2. Due to deterioration, the load response of many objects also gradually changes. If a quantity characterizing the degradation is known (e.g. the magnitude of vibrations), it is possible to monitor its time development in the real object. These data can be extrapolated, and the time can be predicted at which the characteristic quantity reaches the critical value.

Both approaches may be combined. During design, the time to failure can be predicted, and during operation, this prediction is updated with respect to the actual time history of the operation and response. The object is never allowed in operation until the instant of expected failure. At a reasonable time before it, a check of its condition is made and a suitable time for repair or next inspection is proposed.

Both cases will be discussed here in detail.

1. Prediction of the time to fatigue failure from the law of material degradation

A typical example is the fatigue of metallic parts under cyclic or periodic loading. If the characteristic stress is higher than the so-called fatigue limit, a small crack arises in the component after some time of operation and grows slowly, and when it reaches the critical size, the component breaks. The number of cycles to failure N_f depends on the stress amplitude σ_a in the loading cycle [1, 2]. In the simplest case, it can be expressed by the so-called $S - N$ (or Wöhler) curve (Fig. 1):

$$N_f = A\sigma_a^{-m}; \quad (1)$$

A and m are constants obtained by testing the material or component. Several variants of fatigue equation exist, but, basically, their character is similar to Equation (1). Sometimes, the stress range in the loading cycle, $\Delta\sigma = \sigma_{\max} - \sigma_{\min}$ is worked with instead of the stress amplitude σ_a .

Equation (1) can be rearranged to obtain the stress amplitude or range corresponding to the demanded life:

$$\sigma_a = \left(A / N_f\right)^{(1/m)}. \quad (2)$$

This formula can be used in the dimensioning of a component for a certain prescribed life.

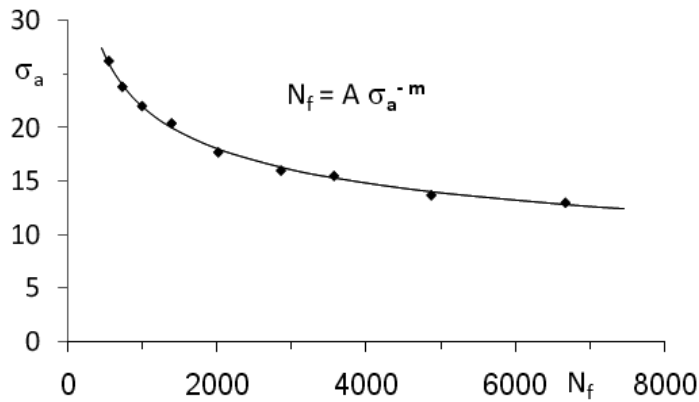


Figure 1. $S - N$ curve for fatigue. N_f – number of cycles till failure, σ – characteristic stress or load (amplitude σ_a or range $\Delta\sigma$).

2. Time to fatigue failure of objects with cracks

Equation (1) does not assume any previous damage to the component. Sometimes, however, one or more cracks or similar defects are present in the body from the beginning. The behavior of bodies with cracks is studied by fracture mechanics [1 - 4]. Crack growth is influenced not only by the stress, but also by the crack size (Fig. 2). Both quantities form together a very important parameter, called **stress intensity factor** K ,

$$K_I = \sigma Y(a) a^{1/2}; \tag{3}$$

σ is the nominal stress in the crack region, a is the characteristic length of the crack, and $Y(a)$ is a factor characterizing the crack shape and size and stress distribution. The subscript of K denotes the mode of crack opening; number I means simple opening, which is the most important case. If the stress intensity factor attains the critical value K_C for the given material, fast fracture follows. The corresponding critical crack length is

$$a_c = \left\{ K_{IC} / \left[\sigma_{\max} Y(a_c) \right] \right\}^2. \tag{4}$$

In components exposed to periodic loading, the crack can grow very slowly even if the stress intensity factor is lower than the critical value. The period of subcritical crack growth can last from minutes to years and can be predicted via the relationship between the crack velocity and the stress intensity factor. The crack velocity v during the period most important for delayed failure (region II in Fig. 3a) can often be approximated by the Paris-Erdogan equation [3]:

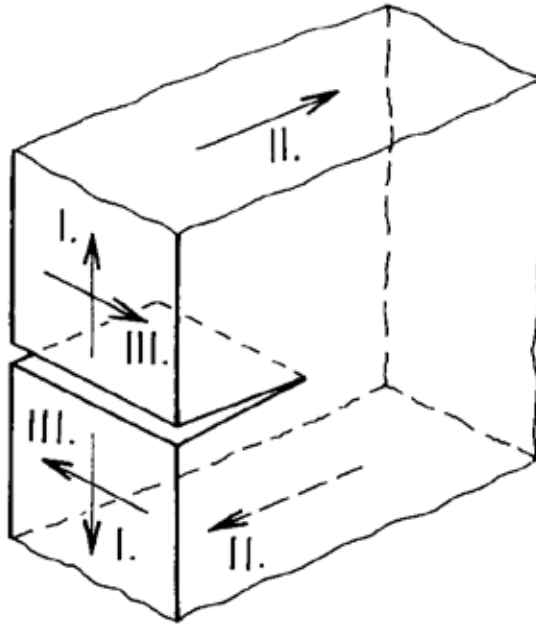


Figure 2. Body with a crack. Characteristic modes of crack opening.

$$v = da / dN = B\Delta K_I^n; \tag{5}$$

da/dN is the increment of crack length per loading cycle, ΔK_I is the range of stress intensity factor in the loading cycle, $\Delta K_I = K_{I,max} - K_{I,min}$, and B and n are the constants for the given material and environment. Inserting ΔK_I from the modified Equation (3) into Equation (5) and separating a and N , one can arrive at the following expression for the number of cycles for the crack growth from the initial length a_i to the length a :

$$N = \int_{a_0}^a \frac{da}{B\Delta\sigma^m [Y(a)]^m a^{m/2}} \tag{6}$$

Fast fracture occurs if the stress intensity factor attains critical value K_{IC} also called fracture toughness. The corresponding critical crack length a_c , used as the upper limit in the integral (6), is given by Equation (4), with σ_{max} denoting the maximum stress in the loading cycle. The resultant formula for N_f in the simplest case (constant stress range, small crack enlargement, and thus $Y \approx \text{const}$) is basically similar to Equation (1), with $\Delta\sigma$ instead of σ_a . The number of cycles to failure is roughly indirectly proportional to some power of the characteristic stress range or amplitude. It is thus possible in design to propose such dimensions of the cross-section that the stresses will be so low to guarantee the demanded lifetime.

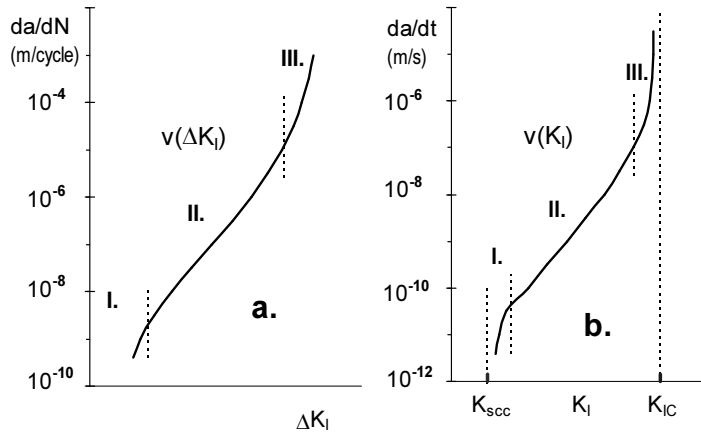


Figure 3. Crack growth velocity v as a function of the stress intensity factor (a schematic). a – metallic materials, periodic loading; b – brittle materials, static load.

3. Static fatigue, wear, and creep

In some materials, fatigue occurs even under constant load. Examples are glass and some ceramics as well as some metals in a corrosive environment. In this case, called static fatigue, it is possible to express the time to failure t_f as a function of acting stress. The velocity v of very slow crack growth depends not on the amplitude but on the value of the stress intensity factor K_I (Fig. 3b), and in the important part of the $v(K)$ diagram, the velocity can be approximated by a power-law function:

$$v = da / dt = AK_I^n. \quad (7)$$

The relationship for the time to failure is similar to Equation (1) or (6) with the stress amplitude σ_a replaced by characteristic stress σ and the number of cycles N_f by the time to failure t_f . The relationships similar to Equation (1) are also used for the prediction of the life of ball bearings and other components exposed to **wear** or of parts exposed to **creep** and other kinds of gradual deterioration. In all these cases, the time to failure is roughly indirectly proportional to some power of the characteristic load P ,

$$t_f = AP^{-n}; \quad (8)$$

the time to failure in rotating parts can be expressed by means of a number of revolutions. Equation (8) is the simplest formula; the relationship in some cases is more complex. For details, the reader is referred to a special literature, for example [1 - 4].

The consequences of random variability of load and uncertainties in the determination of parameters in fatigue equation will be dealt with in Chapter 19.

4. Variable loading

Until now, we have assumed a constant load amplitude. Often, it varies. Figure 4 depicts a regular operation regime of a machine, with four characteristic stages. Examples of irregular or random regime are bogie of a car or components of an engine. In all these cases, the concept of **damage accumulation** is used. Various hypotheses and models have been proposed [1, 2]. Here, only the simplest concept of linear damage accumulation, also named the Palmgren-Miner rule, will be explained.

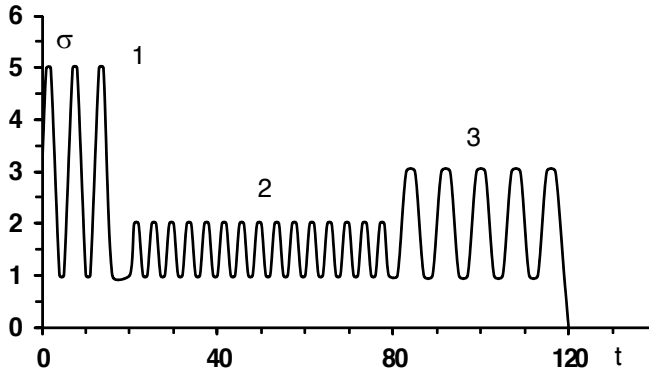


Figure 4. Variable loading (a schematic for damage accumulation).

The basic idea of the **Palmgren-Miner rule** is that every loading cycle contributes to damage and exhausts a minute part of the life. Damage (in fact relative damage) D is then defined as the ratio of the number of loading cycles, which the object has undergone, and the number of cycles (under the same kind of loading), which would cause failure,

$$D = N / N_f. \tag{9}$$

For example, if a component could sustain 1000 loading cycles until failure, then one loading cycle has exhausted 1/1000 of its fatigue life. Failure occurs if $N = N_f$, that is if $D = 1$ (in this case, $D = 1$ for $N = 1000$). Note the difference between N and N_f !

If the loading pattern is more complex, consisting of various loading blocks, each with different amplitude and different number of loading cycles (Fig. 4), the total damage is obtained as the sum of damages caused during the individual blocks,

$$D = \sum D_j = \sum N_j / N_{f,j} \quad (10)$$

Again, failure occurs if $D = 1$. In this way, it is possible to find the number of loading cycles or blocks until failure. If the damage, corresponding to one day of operation, is D_1 , then the object can sustain $1/D_1$ days before it fails.

Special procedures have been developed for cases where the load changes irregularly, in a random way, such as constructions of airplanes or cars. For more, see [1 - 3].

5. Prediction of the time to failure from the response

The gradual deterioration of many objects can be characterized by various quantities, such as amplitude of vibrations, noise level, deformation under load, or loss of material by wear or corrosion. The pertinent quantity (y) grows gradually to the critical value, corresponding to failure or just an unacceptable condition (Fig. 5). In some cases, it is possible to monitor this quantity and its development in time. These data can be fitted by a suitable function $y(t)$, and the time to failure is predicted as such, for which y reaches the critical value y_C (Fig. 5). The knowledge of y_C is thus very important. However, it cannot be predicted accurately due to uncertainties in the determination of material parameters, loads, influence of environment, and other factors. For these reasons, the object may never be let in operation till the instant of expected failure. At a reasonable time before it, a check of its condition must be done. This time is denoted as alert, t_A , with the appropriately chosen value y_A . The determination of t_A is also shown in Figure 5. At this time, the object is inspected, and the obtained value $y(t_A)$ serves as a base for the decision on the maintenance, renovation, termination of further operation, or allowing the object in operation until the next inspection. Practical examples are shown at the end of this chapter.

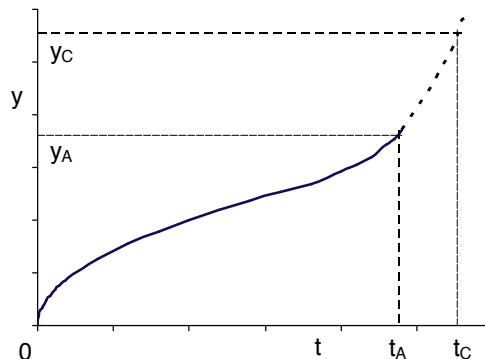


Figure 5. Time course of gradual deterioration (e.g. wear) – a schematic.

6. Probabilistic aspects of the deterioration curves and lifetime prediction

When dealing with the constants of a fatigue curve, taken from the material data sheets, or with the constants of a curve showing the development of a certain parameter with time, one must not forget that they are only approximate, obtained by testing several samples only. One should always be aware of the scatter of individual values. Actually, each curve corresponds to a particular specimen despite the same test conditions. The testing of many samples would thus give a group or distribution of $S - N$ or similar curves (Fig. 6). The data can be processed in several ways. It is possible to use all experimental data points and calculate the parameters corresponding to the "average" curve, giving the "mean" times to failure. In this case, however, the actual times to failure will be in 50% of all cases longer than the times calculated using these parameters and in 50% shorter, which could be dangerous. The reliability of the prediction can be increased using confidence band around the regression curve based on the scatter of the individual values around this curve (see also Chapters 18 and 19). Another approach fits only certain quantiles of the times to failure for various stress levels. In this way, for example, the 5% quantile curve can be obtained. This is such curve that the probability of failing earlier is only 5%. Similarly, the curves for other reliabilities can be constructed. This approach is possible if many data points (several tens or more) are available.

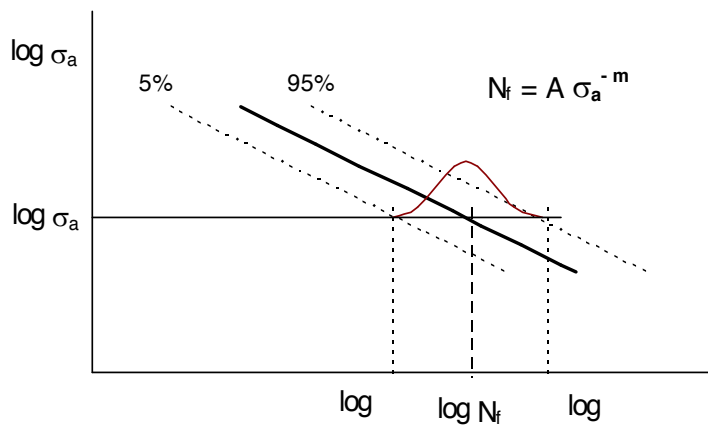


Figure 6. Scatter of fatigue values and curves.

When taking the constants for calculations from a material database or literature, one should therefore know how they were obtained and to what conditions they correspond. Scatter also influences the values of other material parameters, such as strength, fracture toughness K_{IC} , constants in the equation for subcritical crack growth or for creep, and so on. Also here, one must distinguish between the work with the mean values or with certain quantiles.

The uncertainty due to the scatter of experimental data must always be considered in the predictions of the time to failure or the alert time. The higher uncertainty, the longer time before reaching the critical state should an inspection be made. We return to this topic in Chapter 19.

Example 1

A component is loaded by the sinusoidal mechanical stress of the amplitude $\sigma_a = 120$ MPa. The fatigue ($S - N$) curve of the material (cf. Fig. 1) is $N_f = A\sigma_a^{-m}$, with the constants $m = 4.0$ and $A = 5.0 \times 10^{13}$ (for σ_a given in MPa).

Task. Determine the number of cycles to failure N_f . What part of the life will be exhausted after $N = 20,000$ loading cycles?

Solution. $N_f = A\sigma_a^{-m} = 5.0 \times 10^{13} \times 120^{-4.0} = 241,126$ cycles.

$D = N / N_f = 20,000 / 241,126 = 0.0829 = 8.29 \%$.

Example 2

Determine the allowable stress amplitude for the component from Example 1.6/1 so that the component can sustain $N = 600,000$ cycles.

Solution. The rearrangement of the $S - N$ curve gives $\sigma_a = (A/N_f)^{1/m}$. With $N = 600,000$ and the constants $m = 4.0$ and $A = 5.0 \times 10^{13}$, the allowable stress is $5.0 \times 10^{13}/600,000^{1/4.0} = 95.54 \approx 95$ MPa.

Note. In reality, some factor of safety would be used, either for increasing the number of cycles to failure or for reducing the allowable stress. The stress can be reduced by more ample dimensioning.

Example 3

The technical condition of a machine can be characterized by the amplitude y of vibrations. This amplitude was measured once a day. During 10 days, the following values were measured:

Day No.:	1	-	2	-	3	-	4	-	5	-	6	-	7	-	8	-	9	-	10
$y(\text{mm})$:	0.10	-	0.12	-	0.16	-	0.20	-	0.24	-	0.28	-	0.33	-	0.42	-	0.48	-	0.62

Task. Fit the measured data: (a) by a straight line, and (b) by exponential function and predict:

1. the amplitude on 14th day, and
2. when the amplitude will reach the critical value $y_c = 1.00$ mm.

Solution. The measured data were plotted using Excel. Then, the command Add Trendline was used (Fig. 7). Linear approximation (a) was: $y = a + bt = -0.002667 + 0.054121t$; the coefficient of determination, characterizing the quality of fit was $r^2 = 0.9445$. Exponential fit (b) was $y = 0.085250 \times \exp(0.19762t)$, with the coefficient of determination $r^2 = 0.9952$. (Remark: The higher the r^2 , the better fit; $r^2 = 1$ means perfect fit.)

The approximations have given the following amplitudes of vibrations on 14th day: $y_{\text{lin}} = 0.755$ mm according to the linear fit, and $y_{\text{exp}} = 1.356$ mm according to the exponential fit. The

amplitude $y = 1.00$ mm could be expected for $t = 18.52$ days according to the linear fit and for $t = 12.46$ days according to the exponential approximation.

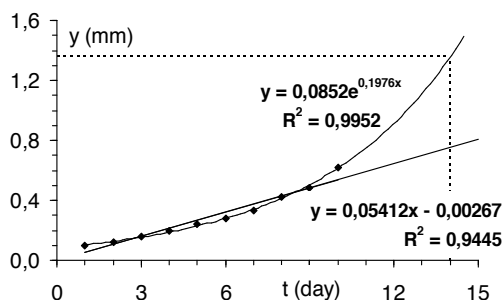


Figure 7. Vibration amplitude as a function of time (an example). Fitting by various regression functions and predictions.

One can see the big differences between the predictions done using different approximations. Caution is necessary especially for longer intervals of predictions and also with respect to the consequences of a wrong prediction.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty, University of Pardubice, Czech Republic

References

- [1] Fuchs H O, Stephens R I. Metal Fatigue in Engineering. New York: Wiley and Sons; 1980. 336 p.
- [2] Stephens R I, Fatemi A, Stephens R R, Fuchs H O. Metal fatigue in engineering. 2nd ed. New York: John Wiley & Sons; 2001. 473 p.
- [3] Broek D. Elementary engineering fracture mechanics. Boston: Martinus Nijhoff Publishers; 1982. 469 p.
- [4] Menčík J. Strength and fracture of glass and ceramics. Amsterdam: Elsevier; 1992. 357 p.

Maintenance

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62360>

Abstract

Various approaches to maintenance are explained: maintenance after failure (breakdown maintenance), preventive maintenance, on-condition maintenance, reliability centred maintenance (RCM), the use of technical diagnostics.

Keywords: Maintenance, preventive maintenance, on-condition maintenance, reliability centred maintenance (RCM), technical diagnostics

The operational reliability and service life of machines, vehicles, various appliances, bridges, and many other long-life objects are strongly influenced by **maintenance**. This term generally denotes small works for restoring full operability, such as cleaning, exchange of oils and filters, and tightening of the locked screws and other adjustments, as well as the repairs of paints or minor faults or the exchange of small damaged parts.

Maintenance is related to the kinds of objects and failures, and its techniques and strategies have been developed along with the development of technology. Historically first was the **maintenance after failure** or **breakdown maintenance**. It was the only strategy common when the machines were relatively simple, their number and productivity was low, and the knowledge about their operation and failures was limited. The losses caused by the corresponding downtimes had to be accepted, as no other solution was known. The strategy of repairs and maintenance after failure is used also today, namely in two cases. The first one are the sudden failures that cannot be predicted, and the second case are the failures of cheap objects, where a failure has none or only insignificant consequences.

During the first half of the 20th century, substantial changes have occurred. The machines, vehicles and other objects became more complex, with failures from various reasons. The number of failures grew, the more so that some products, such as cars, were produced in large

series in production lines. The consequence of a failure of a machine in such line was the stopping of the whole line, and the losses caused by this break became much bigger. Also a failure of a complex and expensive object (e.g. a locomotive, a big ship, or an aircraft) meant big losses. Many of these failures were due to wear or fatigue and occurred after a certain time of operation. Thanks to the understanding of these processes, it became possible to predict approximately the time to failure or to a significant degree of wear in some kinds of items (bearings, cylinders and piston rings, sealing, and valves), or the time to the critical loss of efficiency of oil or air filters, etc. All this has led to the change of the maintenance philosophy. Gradually, the so-called **preventive maintenance** was introduced. Its principle is simple: the endangered parts are repaired or replaced by new ones in fixed regular intervals based mostly on experience. The organization of preventive maintenance is simple and cheap. This approach is suitable for gradual failures (deterioration due to wear, fatigue, corrosion, etc.) and is effective especially if the scatter of the deterioration rate or the time to failure of individual items of a certain kind is low.

In complex systems consisting of many components, failures occur from various reasons, in irregular intervals, and the failures of the system usually occur due to other reasons than the fatigue of key components. During this useful period of the life, the failure rate of the system is approximately constant and the time between failures has exponential distribution (cf. Chapter 4). The maintenance according to a fixed plan then loses any substantiation.

Another drawback of preventive maintenance is that the replacement of components in fixed intervals can be too early for some of them, which is uneconomical, and too late for some others, which can mean worse efficiency of operation or a higher risk of failure. As the methods for ascertaining the technical condition become gradually better, an approach based on the use of **technical diagnostics** became more common (see also Chapter 8). This kind of maintenance, based on the actual condition of the object, is called **on-condition maintenance**. The monitoring of the technical condition and its time development are used for the determination of the optimum time for replacement. Modern cars, rail vehicles, aircrafts, and some other items, as well as the pertinent service stations are equipped with devices enabling the appropriate diagnostics. The diagnosed objects are used best, but additional costs for the diagnostics are necessary.

In machinery, chemical, and some other industries, maintenance has become a very important branch. However, the related costs are high. Ways are therefore sought to reduce them. Currently, the so-called **reliability-centred maintenance (RCM)** is being introduced [1]; see also IEC 60300-3-11. The idea is to eliminate all maintenance works that are not necessary. Often, the failure of a non-crucial component does not endanger the operation of the whole object. With the RCM approach, the system is sometimes allowed in operation also after the failure of unimportant components, and their repairs and maintenance are done only when the more important parts approach their limit state. If the RCM strategy should be introduced, a thorough analysis is done first to reveal the consequences of all possible failures, and the maintenance plan is adjusted appropriately.

Maintenance is a wide topic that goes much beyond this concise book. The reader is referred to the literature, such as [1 - 3].

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

References

- [1] Moubray J. RCM II - Reliability-centered Maintenance. London: Butterworth-Heinemann; 1997. 426 p.
- [2] Ireson W G, Coombs C F Jr, Moss R Y. Handbook of reliability engineering and management. 2nd ed. New York: McGraw-Hill; 1996. 816 p.
- [3] Legát V et al. Management and engineering of maintenance (in Czech: Management a inženýrství údržby). Kamil Mařík Professional Publishing, 2013. 570 p.

Diagnostics

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62361>

Abstract

This chapter explains the tasks of technical diagnostics and basic diagnostic quantities, such as the time of use, time of operation, structural parameters, operational parameters and cost indicators. The main kinds of diagnostics are described, as well as its advantages and drawbacks.

Keywords: Technical diagnostics, diagnostic quantities, time of operation, structural parameters, operational parameters, cost indicators, diagnostic systems

The reliability of many objects, machines, transport means, various appliances, or processes can be increased and operational costs can be reduced by the application of diagnostics. This is a branch of technology dealing with the methods and means for ascertaining the condition of investigated objects. Even the “reliability” of humans can be increased and the life can be prolonged in some cases if the health condition is diagnosed (e.g. by cancer screening of people achieving a certain age or by measuring specific body parameters of patients in hospitals). In this chapter, only the basic diagnostic terms will be explained briefly, with an emphasis on technical objects. However, the applicability is much more general.

The main tasks of **technical diagnostics** are as follows:

1. Determination of the technical condition of the diagnosed object: (a) now (current condition), (b) in the past, and (c) in the future.
 2. After a failure: determination of the place, extent, and cause of the failure (detection, localization, and identification of the failure).
 3. If the technical condition of the object is good: prediction of its future development (Figs. 5 and 7 in Chapter 6), determination of the further duration of operation, and determination of the time for the next inspection or maintenance.
-

The **technical condition** of an object is the condition characterizing its ability to fulfill the demanded functions (under assumed conditions). It can be described by diagnostic quantities.

The **diagnostic quantity** (or parameter) is a quantity that carries information on technical condition and can be used for diagnostics. Various quantities can serve for diagnosing. The basic kinds of diagnostic quantities are as follows:

1. Time of use

It is the total (calendar) time of operation and breaks. This quantity yields certain information on the corrosion or aging but not about wear. For example, if one wants to buy a second-hand car, the first information, which is looked at, is its age.

2. Operating time

- a. The time necessary for doing some work (the time for which the construction was loaded or working; also the number of loading cycles). This quantity, however, does not sufficiently reflect the variable operating regime and loads. This makes no problem for objects with monotonic conditions of operation.
- b. The amount of the performed work (driven distance, amount of consumed fuel, area of the harvested field, number of machined pieces, product of the mileage, weight of cargo transported by a truck, etc.). The buyer of a second-hand car is also interested in the total mileage the car has driven.

3. Structural parameters

Parameters that directly characterize the technical condition or extent of damage of individual components (wear of the cylinders in an engine or of other components; e.g. tires, size of cracks, magnitude of deformations, the clearance between the shaft and bearing, the value of electrical resistance, decrease of thickness of load-carrying parts due to corrosion or wear, and amount and condition of the lubricating oil).

These quantities have very high informative value. A drawback is that they often can be determined only on a disassembled object.

4. Operational parameters

They are derived from structural parameters and express the properties of the diagnosed object or illustrate its changed technical condition. Examples are vibrations, temperatures, fuel consumption, noise, and efficiency. They characterize the technical condition directly and can be monitored without disassembling. As such, they are sometimes measured by the user, but sometimes during special tests when the structure is exposed to specific loads (e.g. load tests of a bridge). A drawback is that operational parameters can be measured only on objects able of operation.

5. Cost indicators

The changes of technical condition lead to the changes of some components of operational costs (e.g. those for fuel consumption). Therefore, the following quantities are sometimes monitored: total cumulative costs $C(t)$, average unit costs $C_1(t) = C(t)/t$, and instantaneous unit

costs $dC_1(t)/dt$. They can be used for the determination of the cost-based optimum instant for renewal, as it will be described in Chapter 17.

The advantage, as well as disadvantage, of cost indicators is that they depend not only on the technical condition of the object, but also on the current level of prices.

Application of technical diagnostics

Diagnostics can be carried out in several ways. The simplest is **manual diagnostics**, either visual or with the use of universal devices (e.g. observation of the tires of a car, inspection of the condition of a certain object, such as a building, bridge, or dam; measurement of deformations, either permanent or under certain load, strain gauge measurement of stresses, measurement of vibrations or noise). Semiautomatic diagnostics needs special devices, but it works according to a fixed regime, which can also be controlled by a technician. Automatic diagnostics is performed by a special system, where a computer controls the testing of all important parameters and functions. The system can work in an adaptive manner: it selects (and stores) only the important quantities to be measured, with respect to the course and results of the ongoing diagnostic process.

Diagnostic systems can be on-board or station-type. Stationary systems are used for various objects and are either stable (immovable) or mobile. Stable systems are used in service stations for technical control (e.g. analysis of exhaust gases or adjustment of lights in cars). A mobile system is installed in a vehicle (e.g. in a van, a special wagon or another transportable appliance for measurements on railways). For example, aircrafts, locomotives or cars have today often installed sensors for the demanded quantities. Then, a measuring vehicle comes and its devices are connected with them via the plugs.

On-board diagnostic systems are installed in the diagnosed object and measure during its usual operation (e.g. the amount of fuel or temperature of bearings). Often they are only oriented at certain parameters (e.g. dynamic response of a wagon bogie during a ride).

Advantages of diagnostics

1. Diagnostics enables quick finding of the place and cause of a failure and thus the reduction of the time necessary for the repair.
2. The regular monitoring of technical condition makes possible the prevention of serious failures as well as the prevention of the fast degradation of properties or worsening the economy of operation by finding the optimum time for the maintenance and renewal.
3. Diagnostics enables the planning of more serious repairs and reconstructions in advance, so that they can be done in the most suitable time, with minimum time losses and unfavorable consequences for the operation, production, or traffic.

Drawbacks of diagnostics

The use of diagnostics increases the total costs. Diagnostics is useful, but not for any costs. One generally aims at the minimum sum of the costs for diagnostics and the losses caused by insufficient diagnosing, and a compromise is often necessary.

More about diagnostics can be found in the literature, for example [1, 2]. Some books deal with diagnostics from the point of view of data and information processing; others are devoted to the instruments for diagnostics or to the application of diagnostics in special branches, such as machines, electrical appliances or electricity supply, and vehicles, and to the use of diagnostics in medicine.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

References

- [1] Legát V et al. Management and engineering of maintenance (in Czech: Management a inženýrství údržby). Kamil Mařík Professional Publishing, 2013. 570 p.
- [2] Kreidl M et al. Diagnostic systems (In Czech: Diagnostické systémy). Praha: ČVUT; 2001. 352 p.

Failure Analysis

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62362>

Abstract

Two approaches to failure analysis are explained: analysis of individual failures and statistical analysis. Various criteria for failure sorting and classification are presented, as well as the main causes and mechanisms of failures. The text is accompanied by figures with characteristic fracture patterns. The chapter is complemented by an example of computer aided sorting of failures in railway driving vehicles.

Keywords: Analysis, failure, statistics, Pareto diagram, failure mechanism, computer aided analysis, railway vehicles

The analysis of failures is very important for revealing the cause of a particular failure, for taking appropriate measures to avoid similar failures in the future, and for improving similar products or processes. Historically, failure analysis has also contributed to the creation of new disciplines of mechanics and other branches, and to the better design and reliability of many products.

Two basic kinds of failure analysis can be distinguished: analysis of individual failures and statistical analysis. However, failures can also be sorted according to other criteria.

1. Analysis of individual failures

This kind of analysis aims at finding the cause of particular failures. Generally, it has three sources: appearance of the failed object, history of the accident, and information on the loads, material properties, and conditions of the service or operation.

An observation of the appearance of the failed item can often reveal the fracture origin (Figs. 1 and 2) and the internal cause of the failure (e.g. a material defect). Fracture mechanics can

help in the determination of the magnitude of stresses and forces acting at the critical place at the instant of fracture. The appearance of the fracture surface (Fig. 1) and the crack trajectory (Fig. 2) can inform about the time course of the fracture process and also about the characteristic failure mode (brittle fracture or fatigue fracture) and the kinds of acting stresses (e.g. shear stress leading to torsion fracture; Fig. 2g). Some details are visible with the naked eye, and some need an electron microscope. It is important to create a detailed photographic documentation or, at least, a thorough description of the failed object.

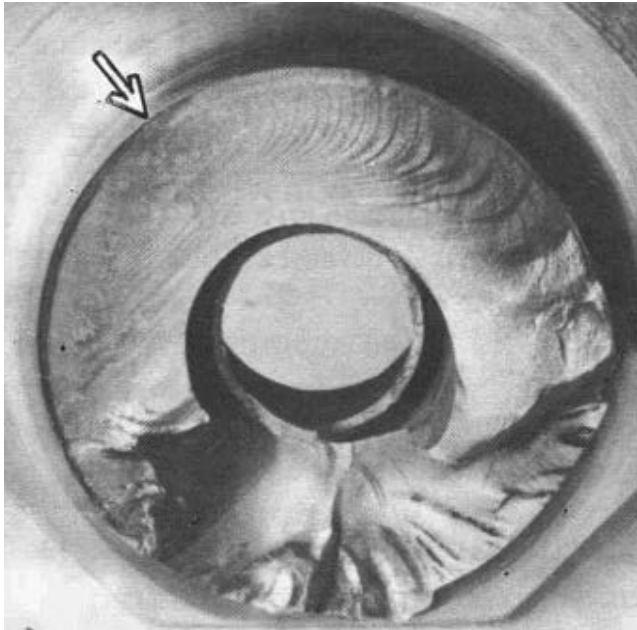


Figure 1. Fracture surface of a steel shaft broken due to fatigue. The arrow indicates the origin of the fatigue process and direction of crack propagation.

Failure analysis uses the time course of the accident and the situation before it, the detailed history of operation, and the conditions of use of the object. The analysis can lean on records from operation (time course of pressures, temperatures, other loads acting on the object, the conditions of environment and the personnel). The sources are records of measuring devices, logbooks, and protocols from inspections.

Often, a computer analysis of the stresses acting in the object is done or also the analysis of material properties, including the mechanical testing of specimens taken out from the critical parts (tensile test, fatigue tests, test of notch resistance or fracture toughness, etc.).

Generally, this kind of analysis can result in measures for the prevention of similar failures in similar components or structures in the future. Well-known are extensive (and expensive) analyses done after aircraft accidents, but thorough analyses became common after every accident with critical consequences.

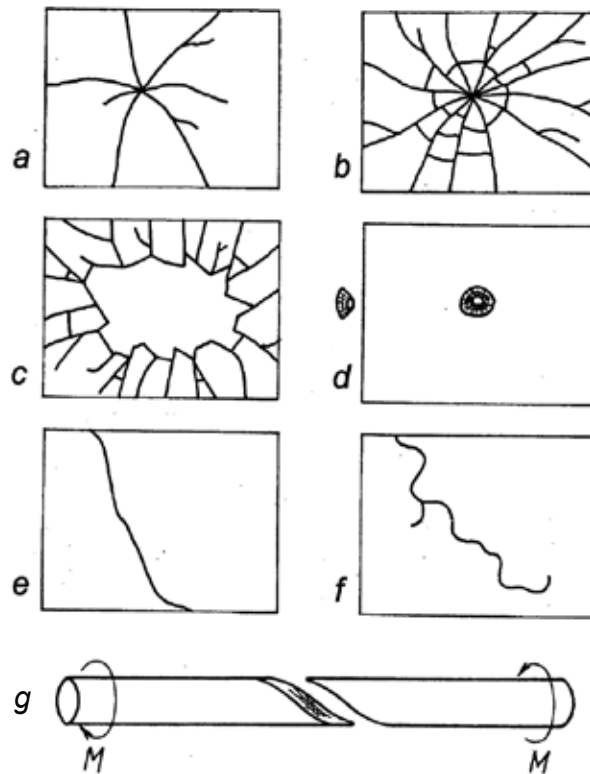


Figure 2. Examples of crack trajectories. a – f: cracks in glass panes (adapted from [1]). a – force acting at the left edge; b – pressure acting over all the area; c – like b, but more intensive load; d – thermal stress; e – fracture due to the detonation in a room; f – a shot through the plate, g – a shaft of brittle material, broken by twisting.

Some rules for failure analysis are also summarized in [1, 2]. Examples of many failures and their analysis can be found in the literature [2 - 6]. Interesting also is the TV series “Seconds from disaster” or the analyses of aircraft accidents observable via YouTube.

2. Statistical failure analysis

This kind of analysis works with a high number of failures of certain kind (e.g. failures of bridges, gear boxes or buildings, as well as vehicle collisions in the analysis of traffic accidents, interruptions of electricity supply, or death cases in a hospital). It uses the records on accidents and failures and records from service stations or repair workshops. The failures can be sorted according to various criteria, such as the kind of the failed object, the place or time of the occurrence, or the cause.

Statistical analysis needs a high number of values and the knowledge of statistical methods. Useful is the software able to sort the data according to various criteria and to perform statistical tests. It enables one to distinguish rare failures from systematic ones and helps in

finding the common cause of some failures. Compared to the analysis of individual failures, statistical analysis can reveal hidden relations and reasons for some kinds of failures. As a consequence, it can help to introduce system measures for reliability enhancement, such as 100% control of welds or bought components, introduction of certain regulations, such as the demand for special qualification for some kinds of work, fire regulations, standards for building of metallic structures, prohibition of building in the area endangered by flooding or avalanches, or prescription of preventive inspections in health care.

3. Sorting of failures

Failures can be sorted according to various criteria. Understanding the character and cause of a particular failure helps in deciding what means should be used to avoid such failures in the future or to reduce their consequences. In this paragraph, sorting according to the character of failures is shown first followed by sorting according to various criteria.

A failure can be:

- complete or partial,
- sudden or gradual,
- random or systematic,
- insignificant or significant,
- independent or dependent.

The inclusion of the investigated failure into the proper group facilitates the selection of an appropriate strategy for avoiding similar failures in the future.

Failures can further be sorted according to

- kind of the object,
- seriousness of consequences,
- frequency of occurrence,
- time (duration) of operation failure,
- time (date) of occurrence,
- period in the lifetime,
- location of occurrence,
- cause,
- failure mechanism,
- other criteria.

The individual criteria are discussed further in detail.

Kind of the object (component or appliance). In civil or mechanical engineering, it is usual to record, study, and analyze the failures of bridges, gear boxes, fans or pumps of certain kind; failures of traction rail vehicles, cars or airplanes of brand "xxx" or type "yyy", failures of brakes, failures of electric or hydraulic appliances, etc. In medicine, one can observe and investigate diseases (or deaths) of children or adults, some kind of disease, etc. In traffic, accidents of certain vehicles, such as buses, motorcycles, or trucks, can be studied. This topic will also be addressed at the end of this chapter.

Severity and frequency of occurrence. Failures can be sorted according to the consequences (e.g. "insignificant – minor – mean – serious – critical – fatal"). (Other classifications are possible as well.) The significance of a certain kind of failures can be evaluated similarly to the FMEA analysis, as explained in Chapter 12. In addition to the severity of a particular kind of failure, it is also possible to consider its frequency or probability of occurrence ("how often it happens?"), to assign weights to them, and to form a common criterion "consequences × frequency". Generally, and for civil engineering structures in particular, it is reasonable to distinguish the failures of serviceability and the failures of load-carrying capacity.

Time of operation until failure. This quantity can be measured in hours or seconds, kilometers, number of pieces processed until the cutting tool becomes blunt, etc. The distribution of a large amount of these values, plotted along the time axis, enables one to distinguish early failures and wear-out failures, etc. It is also important for the determination of basic reliability characteristics, such as MTTF and MTBF, or for planning maintenance and renewal.

Time of occurrence. The occurrence of failures can sometimes depend on time [e.g. on the daytime (hour), year season (influence of weather), or even the day in a week]. For example, many years ago, the ironic term "Monday car" was used in the United States because of a much higher failure rate of cars assembled on that day, perhaps the aftermath of the weekend.

Stage in the lifetime of the object. Generally, three stages exist, where failure occurs or can be initiated:

1. Concept proposal and preparation of manufacture or building.
2. Manufacture, assembly, or building.
3. Operation of the component, object, or appliance.

The individual stages can become sources of different causes and kinds of failures (e.g. "child diseases" appearing soon after putting the object into operation or failures due to wear and aging after a long time of service). The knowledge of the typical features of failures in the individual stages helps in the choice of a proper strategy for improvement.

Place of origin. Three examples can be given: the fifth mould in a multisectional machine for glass bottle-making, various parts of a road in the case of traffic accidents, or the operator who has the highest rate of failures. The knowledge of the place where the failures occur most often helps one to better identify the reason for the failures. Also, it reduces the time necessary for repair.

Failure cause. Basically, the cause of a failure can be internal or external. An **internal cause** means that the component was "weak" for the assumed load. Such failures can be avoided by better design, dimensioning, or manufacture. Failures due to **external cause** are those caused by overloading, collision with another object or due to another failure. The efficient way to mitigate them needs the knowledge of failure cause in the particular case.

Principal causes of failures

1. Natural influences or disasters (also collision with another object);
2. Unpredictable, random, or insufficiently known influences (e.g. load, environment);
3. Deficient project (errors in the concept, dimensioning, choice of materials, etc.);
4. Defective material or semifinished goods and bought components;
5. Unsuitable process of the manufacture, assembly, or building;
6. Disobeying the prescribed conditions for manufacturing (the proper material) or operation (e.g. overloading);
7. Insufficient control and verification tests;
8. Wear, fatigue, change of dimensions or properties of the object in time;
9. Insufficient maintenance;
10. Human factor:
 - Ignorance, insufficient knowledge;
 - Negligence, disorderliness, laziness, carelessness;
 - Errors, inattention, absent-mindedness, bad psychological condition;
 - Unsubstantiated reliance in other people;
 - Excessive thriftiness, greed;
 - Malicious intention.

Mechanisms and causes of mechanical failures

1. Fracture:
 - ductile (with well observable permanent deformations; it occurs due to the overloading of components from tough materials);
 - brittle (without observable permanent deformations; it occurs in brittle materials, at low temperatures, dynamic load, impact, notches, cracks);
 - fatigue [with typical appearance (Fig. 1); it occurs under harmonic or periodic loading or even under constant load];
2. Strength reduction due to cracks;

3. Wear (adhesive, abrasive, erosive, fatigue, and cavitation);
4. Corrosion (chemical, electrochemical, oxidation, and erosion of electric contacts);
5. Permanent deformations (by overloading);
6. Permanent deformations in the surface layer due to high contact stresses;
7. Gradual changes in shape and dimensions due to creep (in some materials even under normal temperatures);
8. Temporal property changes due to changes in the environment (temperature, humidity...);
9. Material changes due to UV radiation or ozone (aging of polymers and rubber);
10. Chemical changes promoted by diffusion (e.g. carbonation of concrete by CO₂);
11. Changes of dimensions and dynamic properties due to the loss of material (corrosion or abrasion) or its gain (ice accretion, water-logging, deposition of products of chemical reactions);
12. Changes caused by living organisms (rodents, insects, fungi, mould, rotting).

Also other criteria for sorting can be used. The high number of available failure data enables their sorting and analysis according to several criteria simultaneously. Such analysis can reveal relationships and influences unknown as yet.

4. Classification of failures

Especially two aspects of failures are important: severity and frequency or probability of occurrence. In the case of a high number of various kinds of failures, a **Pareto analysis** is very informative. In this analysis, kinds of failures are rank-ordered according to the frequency of occurrence. The pertinent histogram shows at first sight the failures that occur most often and the rare ones (Fig. 3). The typical shape of a Pareto diagram has led to the saying "20% of all causes are responsible for 80% of all troubles, and 80% of causes are responsible for 20% of problems". However, this is only a saying and not a law of nature.

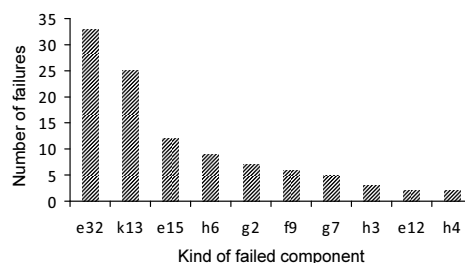


Figure 3. Pareto diagram (an example).

Also, the knowledge of relative frequencies of occurrence, corresponding to probabilities, is important. One should keep in mind that a product with 100 failures per 1million pieces is much more reliable than another product with "only" 10 failures, but per 1000 pieces.

The significance of failures can be evaluated according to the consequences and to the frequency of occurrence. The overall importance is evaluated with respect to both criteria, as described in the "Severity and frequency of occurrence" paragraph and in Chapter 12.

5. Computer-aided failure analysis and record keeping

Many items today are very complex and can fail from various reasons. Manufacturers or users of products such as cars, locomotives, pumps, etc., often produce or operate many pieces, so that the number of various failures can be very high. A consequent reliability analysis needs a system for the classification of failures (and also a system for the evidence of times between failures and times to repair, a system for data collection, and tools for statistical data analysis). Here, a simple system for the classification of failures in railway driving vehicles [7] will be shown as an example. This system classifies the failures with respect to the (1) kind of vehicle, (2) structural group in the vehicle, (3) subassemblies, and (4) specification of the failure. Each of these four categories is again divided into several subcategories. The code for each failure thus represents a four-digit number of the form ABCD, whose digits specify the situation in the individual subcategories, for example:

A. Kinds of railway vehicles

1 - Electric locomotives for DC

2 - Electric locomotives for AC

...

7 - Diesel locomotives

...

B. Structural groups in railway vehicles

1 - Bogie

2 - Body

3 - Mechanical equipment

...

8 - Protective paints, signs, labeling

...

C. Subassemblies

The number here characterizes the specific properties of individual groups, typical for some kinds of vehicles, purpose, and kind of use (feeding, power transmission, shifting, etc.). The number of subassemblies can be different in individual groups.

D. Specification of failures with respect to the consequences

0 - Component without failure

1 - Repair of faulty or damaged component

2 - Exchange of faulty or damaged component

For example, the code 2312 in the above "ABCD" system means electric locomotive for alternating current (2), failure of mechanical equipment (3) at traction part (1), and the pertinent component had to be replaced by a new one (2).

Various systems exist (also as a part of maintenance management systems) or can be created according to the specific requirements of the user (e.g. a system for evidence and the classification of failures combined with the tools for cost analysis).

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

References

- [1] Menčík J. Strength and fracture of glass and ceramics. Amsterdam: Elsevier; 1992. 357 p.
- [2] Ireson W G, Coombs C F Jr, Moss R Y. Handbook of reliability engineering and management. 2nd ed. New York: McGraw-Hill; 1996. 816 p.
- [3] Neale B S, editor. Forensic engineering. The investigation of failures. London: Thomas Telford Publishing; 2001. 219 p.
- [4] Ratay R T, editor. Forensic Structural Engineering Handbook. New York: McGraw-Hill; 2009. 688 p.
- [5] Wearne P. Collapse. When Buildings Fall Down. New York: TW Books; 2000. 255 p.

- [6] Lancaster J. Engineering catastrophes. 3rd ed. Abbingdon: CRC Press, Woodhead Publishing Ltd; 2005. 269 p.
- [7] Sojka, J., Menčík, J. Analysis of failures in railway driving vehicles. In: Proc. Int. Conf.: Reliability and Diagnostics of Transport Structures and Means 2002 (J. Menčík, editor). Pardubice: University of Pardubice; 2002; p. 318 – 322

Approaches to Ensuring the Reliability and Safety

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62363>

Abstract

This chapter presents various ways to reliability ensuring. The philosophies fail-safe, safe-life and damage-tolerant are explained briefly, as well as deterministic and probabilistic approach. Then, the important methods are explained, such as allowable stress, use of standards, load and resistance factor design, probabilistic approach and proof testing.

Keywords: Reliability, safety, design, fail-safe, safe-life, damage-tolerant design, allowable stress, codes, load and resistance factor design, proof testing, probability

From a reliability point of view, every technical object can be either in a serviceable state or in a failed state. The boundary between both is the **limit state**. Some objects fail suddenly. The condition of other objects changes gradually (e.g. due to wear or corrosion). They are able to fulfill their purpose for a long time, though in a limited extent (worse technical parameters or lower safety); the failure is partial. However, if certain parameter exceeds a specified **limit value**, the object either becomes destroyed or unfit for further use; the failure is complete.

In civil and mechanical engineering, two kinds of limit states are distinguished: limit state of load-carrying capacity and that of serviceability (usability). Exceeding the limit state of load-carrying capacity leads to the destruction of the object, often with fatal consequences. If the limit state of usability is exceeded (e.g. large deformations), the object cannot fulfill its function properly, but the consequences are not fatal. Correspondingly, the demanded degrees of safety of an object can differ depending on the kind of the limit state and consequences of its exceeding.

The boundary between serviceable and failed state can be described by one number (e.g. the stress value, whose exceeding means fracture of the component) or by some analytical expression (e.g. the relationship between the critical load for buckling of a compressed column and its slenderness). Also, the condition of the fracture of a shaft loaded simultaneously by twisting and bending depends on the ratio of both load components; see also Figure 1.

The **basic condition of reliability and safety** says: “The resistance to load effects must be higher than these effects”.

1. Basic philosophies for ensuring safety in the design stage

The two most common approaches are (1) Fail-safe and (2) Safe-life. A special case of the safe-life approach is the (3) Damage-tolerant design.

The **Fail-safe approach** understands that the important parts of the object can fail and tries to do everything that such failure will not be fatal for the whole object. This is mostly achieved using redundant components or circuits. **Redundancy** can be **active**, with all parts loaded or working simultaneously, or **standby**, where the redundant component is switched-on only if the principal component has failed; see also Chapter 5.

The **Safe-life approach** tries to do everything to ensure that the component or object can sustain all expectable loads during the assumed life. Basically, it means sufficient dimensioning (the knowledge of all possible loads is important) and the use of materials that do not deteriorate or whose rate of degradation is acceptably low. The dimensioning for “infinite” life of the items exposed to fatigue or ample dimensioning of parts exposed to creep also belongs here.

In contrast to the previous case, which assumed a “perfect” object at the beginning of service, the **damage-tolerant approach** assumes that the component contains some defects (e.g. cracks), which will gradually grow during the operation. The components are dimensioned so that these defects cannot attain critical size during the expected lifetime. The knowledge of the defect growth velocity as a function of load is necessary. This approach uses fracture mechanics, but it is similar to the safe-life approach.

The determination of the time to failure, or dimensioning of a component for the demanded life was explained in Chapter 6.

2. Deterministic versus probabilistic approach

The procedures for the design and check of reliability depend on whether random influences are considered. Some quantities can be considered as deterministic (e.g. the number of teeth in gears or the distance of bearings in a gearbox). Other quantities, such as the strength of material, loads, or action of environment (e.g. wind velocity), have random character, with values varying in some intervals. There are also other sources of uncertainties (e.g. the

computational methods and models characterizing the limit state). Historically, various approaches for ensuring reliability have been developed. They can be divided into two groups: deterministic and probabilistic.

In the deterministic approach, every quantity (load, strength, etc.) is described by one number of constant value. The design in this case is simple, as well as the check of safety, as it will be shown in this section.

The probabilistic approach is based on the fact that some quantities (e.g. load or strength) vary due to random reasons and cannot be sufficiently described by one number only. This approach works with the probability distributions of the pertinent quantities and determines the probability of failure or probability of exceeding the allowable values. The component or structure is considered safe (or reliable) if the probability of failure is lower than certain allowable value.

The probabilistic approach can give more accurate answers but is more demanding than the deterministic approach. It needs a basic knowledge of the probability theory, some computer tools for the work with random quantities (even Excel is sufficient in simple cases), and, of course, the knowledge of probability distributions of the pertinent random variables. If their types and parameters or histograms are not known, this kind of analysis cannot be made. One must also be sure that the statistical characteristics (of the materials or parts) used in the design will correspond to reality. The probabilistic methods for reliability assessment are described in Part 2 of this book.

3. Design using allowable stress

This is a traditional approach in mechanical engineering. The safety condition is

“The maximum operating stress must not exceed the allowable stress”.

The allowable stress is obtained by dividing the nominal strength of the material $\sigma_{n,s}$ (ultimate strength or yield stress; do not confuse it with standard deviation) by the so-called **factor of safety** k_s :

$$\sigma_{\text{allow}} = \sigma_{n,s} / k_s. \quad (1)$$

The meaning of this factor could be interpreted roughly as “failure would occur at k_s -times higher stress”. However, the situation is more complex. The value of safety factor is chosen to “cover” all uncertainties related to the material, the component, and the conditions of operation. Therefore, this approach is also somehow related to probability, but only very loosely. For example, the allowable stress is such value that practically all pieces of this material will be stronger. For metallic parts, σ_{allow} is calculated so that the “minimum” strength, given in material data sheets as 5% quantile (i.e. corresponding to 5% probability that weaker pieces can appear), is divided by a factor of safety, which is chosen, in accordance with the years of

experience, so that the probability that σ_{allow} will be lower than the maximum acting stress is negligibly low.

The factor of safety k_s is a number usually between 1 and 3 and, in some cases even more. Generally, the higher the uncertainties, the higher the k_s . Its values are based on experience; they can be found in the literature; manufacturers use their own well-proven values. The safety factor is often given as a single number (e.g. 2.5), but sometimes it is calculated as the product of several partial coefficients; for example,

$$K_S = s_1 \times s_2 \times s_3 \times s_4 \times s_5 \times s_6, \quad (2)$$

where s_1 characterizes the importance of the component (low or high), s_2 – the technology of manufacturing, s_3 – the material testing, s_4 – the way of strength calculations, s_5 – the quality of manufacturing (e.g. casting or machining by turning or grinding), and s_6 characterizes the possible overloading; s_j is closer to 1 for smaller uncertainty in the j -th factor. This approach enables the consideration of the variability and level of knowledge about the individual factors.

Two values of safety factor should be distinguished. The first is the demanded or target value, which is used in the design stage. The other is the value corresponding to the actual situation. Sometimes, the dimensions of the cross-section (e.g. the wall thickness) should satisfy various criteria, and another criterion than strength can be decisive (e.g. thermal resistance). In such case, the wall will be thicker, so that the actual safety against overloading will also be higher than that originally demanded in design.

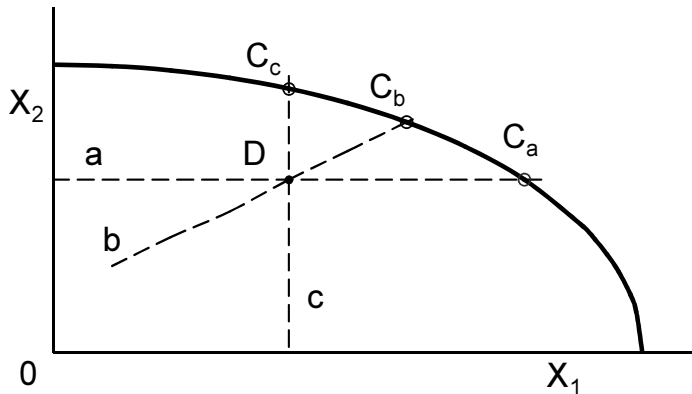


Figure 1. Limit curve, separating safe and failure regions. The failure depends on two quantities: X_1 and X_2 . Curves a,b,c show the various paths of overloading.

When the safety against overload is to be determined, one should consider the actual “path” of overloading leading to the collapse (Fig. 1). Sometimes, several loads act simultaneously, but some of them are constant, such as the dead weight of a bridge, and only some can increase,

for example the traffic load. The actual safety against overloading by traffic is here obtained as the ratio of the actual traffic load at the instant of collapse to the nominal traffic load. These issues are important in dimensioning. A misunderstanding of the term “safety” can lead to unnecessarily high costs.

4. Design according to standards

Standards are often used for the design and dimensioning of bridges, cranes, pressure vessels, and many metal or other important constructions. Also, vehicles, aircrafts, and electric appliances are often designed using various standards, such as ISO, ASME, or Euro-codes; see Appendix 2. In some cases, their use is compulsory, but sometimes it is only a matter of agreement between the manufacturer and the customer. The advantage of standards is that they are usually created as a result of cooperation of many specialists, often from various countries, and are based on a thorough analysis and experimental verification. They are updated from time to time to reflect the progress in the state of knowledge. Generally, standards represent an efficient way for obtaining safe items and constructions. The advantage for the design engineer, manufacturer or builder is that the design and calculations according to proven formulas and procedures in standards are straightforward and simple. Moreover, if the structure fails and the designer or builder can prove that he has done everything according to the standards, he cannot be prosecuted. On the contrary, the design according to codes is somewhat conservative, and the standards do not solve all eventualities (e.g. certain combinations of loads). In such cases, other approaches can be more appropriate.

5. Load and Resistance Factor Design (LRFD)

This approach is common in the design of civil engineering structures, where it is used in some standards, e.g. for steel constructions [1]. The safety condition, in general, is

“The design value of load effect must not exceed the design value of the resistance.”

The term “design value” means the value assumed in design (e.g. recommended or prescribed in a standard), because the actual values are not known exactly yet.

For a component or structure, this condition can be written as

$$\gamma_n S_d \leq R_d; \tag{3}$$

S_d is the effect of maximum load, R_d is the resistance (e.g. the load-carrying capacity or the allowable deformation), and γ_n is the factor characterizing the purpose of the object. The subscript **d** means “design” and denotes the value considered in design; the pertinent standards usually show how the design value is related to the mean or nominal value.

In this approach, the uncertainties are divided into two groups: those related to the load and those related to the material or components. The design value of the load effect F_d is obtained as the product of characteristic load F_k and partial safety factor γ_F for the load,

$$F_d = \gamma_F F_k. \quad (4)$$

The design value of the resistance (or strength) f_d is obtained as the characteristic value of strength f_k divided by the partial factor of reliability of the material γ_m ; for example

$$f_d = f_k / \gamma_m; \quad (5)$$

f_k can be either the characteristic value of the yield strength or the ultimate strength (e.g. 5% quantile). The characteristic values and the partial safety factors are given in pertinent standards (e.g. [1]).

As we can see, the actual values of loads and properties were replaced in the LRFD approach by the design values given in codes. This approach is reasonably conservative and the procedures are arranged so that they enable fast control in standard cases.

Note: Load and resistance are also used in the determination of failure probability in the so-called load-resistance interference method, as described in Chapter 14.

6. Probabilistic approach

If probabilistic approach to reliability assessment is used, only general recommendations for allowable probabilities can usually be found instead of definite obligatory values. Generally, the allowable probability of a failure should be related closely to its consequences. Some idea about these probabilities can be obtained from two examples. The first one is from aviation technology. Usually, 1:10,000 is the acceptable probability of critical failure for an aircraft at the end of its service life, just before decommissioning. The acceptable probability of failure at the time of its putting into service must be several orders lower.

The second example is the Eurocode for metal constructions [1], which gives the following design probabilities of failure for the limit states of load-carrying capacity and usability. They are differentiated according to the assumed level of reliability or safety, as given in Table 1.

The above numbers correspond to the reliability of the whole object. If it consists of many components, the failure probabilities of single elements must be appropriately lower (see Chapter 6), of the order 10^{-5} to 10^{-10} . Similarly, if reliability is assessed via failure rate, the allowable failure rates of elements must be very low. In such cases, specific problems arise in design. First, it can be difficult to prove very high reliability of the pertinent item, because the number of tested samples must be high and the duration of tests are very long. For example, the failure rate of 10^{-6} h^{-1} could be (roughly) verified in a test with one component tested for 10^6 h or with 1,000,000 components tested for 1 h. None of these cases is practicable and a

Reliability level	Limit states of:	
	Load-carrying capacity	Usability
Reduced	5×10^{-4}	16×10^{-2}
Usual	7×10^{-5}	7×10^{-2}
Increased	8×10^{-6}	23×10^{-3}

Note: The unrounded numbers in Table 1 look rather strange. The reason is that reliability index β (see Chapter 14) was used originally instead of probabilities, and the relationship between β and P_i is nonlinear (e.g. $P_i = 23 \times 10^{-3}$ corresponds to reliability index $\beta = 2.00$; see the distribution function of standard normal distribution).

Table 1. Recommended design probabilities of failure P_i [1]

compromise must be found. Another example is related to the guaranteed strength. The value of 0.001% quantile of strength, determined from three tests only, does not make great confidence. Many more tests would be better. However, this would also cost much more money. The big manufacturers of standard electric and electronic components can make extensive tests and use sophisticated techniques for testing and processing the results, as it will be mentioned in Chapter 20, so that their data are trustworthy. Often, however, the means for testing are much more limited, and the predictions are less safe. Usually, the allowable reliability is a compromise between the demands for high reliability and the money available for reliability ensuring. In some cases, the optimum reliability can be found from the condition of minimum total costs consisting of the purchase costs and the costs caused by failure. This topic is treated in Chapter 17. However, this approach cannot be used if very high reliability or safety is demanded. In some cases, the so-called **ALARP** philosophy is used, demanding that the risk should be “as low as reasonably practicable”.

Generally, the demands for increasing reliability should not be unrealistically high. Useful information on the actual situation can be obtained from the statistics of failures. If the current probability of failure of a certain item is $1:10^2$, it will be easier to reduce it to $1:10^4$ than $1:10^8$.

In some cases, quite different approaches are used for guaranteeing very high reliability. Sometimes, certain technologies or activities are prescribed or, vice versa, forbidden by law (e.g. building family houses in the areas endangered by flooding or avalanches). Another means is proof testing, as explained in the following paragraph.

7. Proof testing

In these tests, all components are exposed to certain overload, so high that the “weak” parts are destroyed during the test. Destruction is complete with components of brittle materials, whereas the ductile parts are often only permanently deformed. (For example, overpressure tests common for pressure vessels belong also to proof tests.) Basically, it is sufficient if the proof-test stress or load is equal to a certain value higher than the maximum load expectable in service. In some cases, proof tests are also used for ensuring a sufficient life of components from brittle materials exposed to static fatigue due to the corrosive action of environment. This

fatigue causes very slow growth of preexisting minute cracks until the critical size. The methods of fracture mechanics together with the knowledge of the velocity of subcritical crack growth under stress enable the calculation of the necessary proof-test stress guaranteeing a sufficient life of the parts that have passed the test. Such approach was used, for example, in the design of glass windows in the American orbital laboratory Skylab [2, 3]. Theoretical foundations of proof testing are explained in detail in [2, 3] and [4].

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

References

- [1] ENV 1993-1-1 Eurocode 3. Design of steel structures. (also Czech standard ČSN 731401, 1998.)
- [2] Jacobs D F, Ritter J E jr. Uncertainty in minimum lifetime predictions. *J. Am. Ceram. Soc.* 1976; 59: No. 11/12, 481 – 487
- [3] Wiederhorn S M, Fuller E R jr, Mandel J, Evans A G. An error analysis of failure prediction techniques derived from fracture mechanics. *J. Am. Ceram. Soc.* 1976; 59: No. 9/10, 403 – 411
- [4] Menčík J. *Strength and fracture of glass and ceramics*. Amsterdam: Elsevier; 1992. 357 p.

Advanced Reliability

Weibull Distribution

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62375>

Abstract

Weibull distribution is very flexible in fitting empirical data, such as strength or time to failure. Several methods for the determination of parameters are described, including direct fitting using solvers available in universal programs. Also finding of parameters of exponential distribution is described. The use of Weibull distribution is illustrated on examples.

Keywords: Probability, reliability, Weibull distribution, exponential distribution, determination of parameters, least squares method, solver

A special position in reliability assessment pertains to Weibull distribution, which offers great flexibility in fitting empirical data. The distribution function (Fig. 1a) is

$$F(t) = 1 - \exp\left\{-\left[(t-t_0)/a\right]^b\right\}, \quad (1)$$

with parameters a , b , and t_0 . The scale parameter a is related to the values of t and ensures that the distribution is independent of the units of t (e.g. minutes or hours). The constant b is shape parameter. Depending on its value, Weibull distribution can approximate various, even very different shapes (Fig. 5 in Chapter 2). It is suitable for the characterization of time to failure as well as strength or load; therefore, it became popular in reliability assessment. The constant t_0 is the threshold value that corresponds to the minimum possible value and characterizes the position of the distribution on the t -axis. (t is the usual symbol for time; for other quantities, other symbols may be used.)

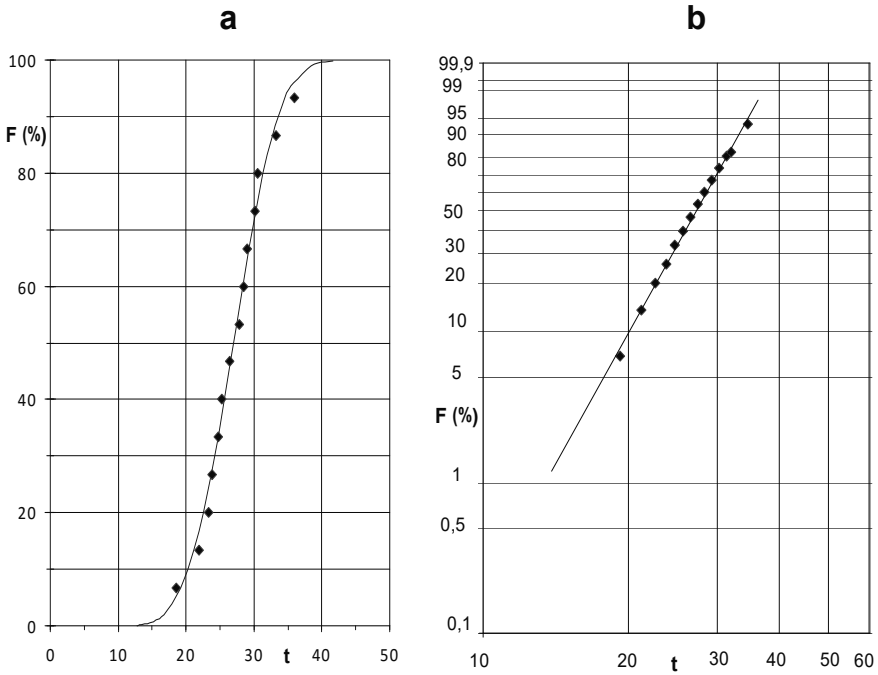


Figure 1. Weibull distribution function $F(t)$: (a) original coordinate system, (b) transformed coordinates (Weibull probabilistic paper).

1. Determination of parameters in a two-parameter distribution

The strength or time to failure cannot attain negative values, so that the threshold parameter is often assumed zero, $t_0 = 0$. The distribution function (1) will thus have only two parameters:

$$F(t) = 1 - \exp\left[-(t/a)^b\right]. \tag{2}$$

Parameters a and b can be found easily, as the transformed data can be fitted by a straight line. Double logarithmic transformation and rearrangement change Equation (2) to

$$\ln t = \ln a + (1/b) \ln\left\{\ln\left[1/(1 - F)\right]\right\}, \tag{3}$$

which corresponds to the equation of straight line (Fig. 1b)

$$Y = A + BX, \quad (4)$$

where $Y = \ln t$, $X = \ln\{\ln[1/(1 - F)]\}$, $A = \ln a$, $B = 1/b$.

The method of linearization was very popular in the past, and it is still often used for the determination of parameters from the operation data via a special diagram, called Weibull paper (Fig. 1b). For its construction, the individual measured values t_j and the corresponding values F_j of the empirical distribution function are needed. The t_j values are obtained by rankordering the n data from operation (e.g. times to failure) from the minimal value ($j = 1$) to maximal ($j = n$). The corresponding values of distribution function are calculated as

$$F_j = j / (n + 1); \quad (5)$$

j is the rank number and n is the total number of measured values. The explanation of formula (5), common for order statistics, is simple. If we have, say, 100 values and order them from the minimal to maximal, then the probability F that t will be smaller or equal to the lowest of 100 values, t_1 , is 1:100. The probability of $t \leq t_2$ is 2/100, etc.; generally, $F_j = j/n$. In Equation (5), 1 was added to the denominator because of mathematical correctness; the probability F that t will be smaller or equal t_n must be smaller than 1, simply because if more measurements would be done, values higher than t_n could appear. Also other formulas exist for the calculation of empirical F_j values [e.g. $F_j = (j - 1/2)/n$], but none can be recommended unequivocally, especially when considering the fact that bigger errors in the determination of distribution parameters can arise due to the small amount of data than due to the formula used for F_j .

The regression constants A , B can be obtained by fitting the empirical data by a straight line (using Weibull paper or a program for curve fitting, such as "Insert Trendline" in Excel). Then, the constants in the distribution function (2) are obtained from A and B by inverse transformation:

$$b = 1/B, \quad a = \exp(A). \quad (6)$$

Plotting the empirical data into the coordinate system $X = \ln\{\ln[1/(1 - F)]\}$, $Y = \ln t$, enables a good visual check. In the ideal case, if Equation (2) is valid, the data lie on a straight line.

2. Determination of parameters in a three-parameter distribution

A two-parameter distribution is not always suitable. Sometimes, the transformed data do not lie on a straight line, or it is obvious that the distribution should have a threshold value t_0 higher than zero. In such case, the use of a two-parameter distribution as a base for dimensioning could lead to uneconomical design, and a three-parameter function (1) would be better.

The parameters in this distribution can be found by the procedure for a two-parameter function if t in Equation (2) is replaced by the expression $t - t_0$; the constant t_0 must be chosen in advance. For various t_0 values, the shape of empirical distribution varies. The best t_0 value is such for which the transformed data best resemble a straight line. However, a more straightforward procedure exists.

Direct determination of parameters

The constants a , b , and t_0 can also be obtained in a simpler way without any transformation. The solution of Equation (1) for t gives the formula for quantiles:

$$t = t_0 + a \left\{ \ln \left[1 / (1 - F) \right] \right\}^{1/b} . \quad (7)$$

This equation and the least-squares method are used in search for such values of a , b , and t_0 , which minimize the sum of squared differences between the measured and the calculated values of t ,

$$(t_{j,\text{meas}} - t_{j,\text{calc}})^2 = \min ! \quad (8)$$

If a suitable solver is available for such minimization (one is present also in Excel), it is then sufficient to prepare one series of measured data, $t_{j,\text{meas}}$, and another series of the $t_{j,\text{calc}}$ values, calculated via Equation (7) for the same values of F_j using the parameters a , b , and t_0 . Solver's command to minimize the expression (8) by changing a , b , and t_0 will do the job. An example is shown at the end of this chapter.

Remark: Formula (7) is also suitable for the determination of a "minimum guaranteed value" (e.g. strength or time to failure) for acceptably low probability F .

In addition to flexibility, Weibull distribution has one more advantage. The shape parameter b in Equation (1) or (2) is related to the character of failures. This is well visible at the bathtub curve (Fig. 1 in Chapter 4). The values $b < 1$ are typical of decreasing failure rate λ and may thus indicate the period of early failures. On the contrary, $b > 1$ corresponds to increasing failure rate λ and is typical of the period of aging or wear out. The value $b = 1$ corresponds to the constant failure rate $\lambda = \text{const}$, with failures from many various reasons (see Chapter 4). The exponent b thus can inform generally about the possible kind of failures and about the period in the life of an object even if the amount of data is not large. However, caution is necessary. If the data from a long period are fitted by Weibull distribution, failures from various reasons and stages can be mixed, and the relation of b to the kind of failures is not unambiguous.

Remark: Weibull distribution was proposed in 1939 by the Swedish engineer Waloddi Weibull, who studied the strength of materials, life endurance of ball bearings, and fatigue life of mechanical components and other quantities. Later, it appeared that this very useful distribution belongs to the family of extreme value distributions [1, 2]. More on Weibull distribution and its applications can be found, for example, in [3 - 5].

3. Exponential distribution

Let us now look at a special and very important case. With the shape parameter $b = 1$, Weibull distribution simplifies to exponential distribution

$$F(t) = 1 - \exp[-(t/a)], \text{ or } F(t) = 1 - \exp[-(t - t_0)/a]. \quad (9)$$

The probability density and distribution function are depicted in Fig. 5. The parameters a and t_0 can be determined similarly as described above. If $t_0 = 0$, the remaining parameter a is usually calculated from the mean time to failure, as it will be shown in Chapter 20. Typical of exponential distribution is that the standard deviation has the same or similar value as the mean.

The determination of parameters and use of Weibull and exponential distribution will be demonstrated in the following examples.

Example 1

The strength (S) of a new alloy was measured on seven specimens, with the following results: 203, 223, 248, 265, 290, 313, and 342 MPa. Solve the following three problems:

A. Determine the parameters of Weibull distribution for this alloy using:

- a. Two-parameter distribution and linearized data;
- b. Two-parameter distribution, applying Solver on the original data without transformation); and
- c. Three-parameter distribution, applying Solver on the nontransformed data.

B. Calculate (for each case) the probability that the strength will be lower than 120 MPa.

C. Calculate (for each distribution) the “minimum guaranteed” strength such that the probability of the actual strength being lower equals: 0.05 – 0.01 – 0.001.

Solution.

Task A. Determination of distribution parameters

- a. Linearized two-parameter Weibull distribution. The strength values, ordered from minimum to maximum, are given in Table 1 together with the values of distribution function, calculated as $F_j = j/(n + 1)$, with $n = 7$; see also Fig. 2. The distribution function $F(t) = 1 - \exp[-(t/a)^b]$ was transformed to linear form; see Equation (4) and the following formulas. The transformed values are in the columns X_j and Y_j . Note: The values of distribution function are fixed (deterministic), as they correspond to the number of measured values, whereas the strengths exhibit random variations. Therefore, F is the independent variable and t is the dependent variable.

j	S _j	F _j	X _j	Y _j	S _{j,c,lin2}	S _{j,c,sol2}	S _{j,c,sol3}
1	203	0.1250	-2.0134	5.3132	196.6331	194.501	201.13921
2	223	0.2500	-1.2459	5.4072	228.3515	227.021	226.32602
3	248	0.3750	-0.7550	5.5134	251.2710	250.615	247.17612
4	265	0.5000	-0.3665	5.5797	271.0299	271.015	266.95925
5	290	0.6250	-0.0194	5.6699	289.9970	290.645	287.54013
6	313	0.7500	0.3266	5.7462	310.2209	311.624	311.21631
7	342	0.8750	0.7321	5.8348	335.7233	338.143	343.64209

Subscript c means calculated; lin2 – linearized, two parameters; sol2 – nonlinearized, Solver, two parameters; sol3 – nonlinearized, Solver, three parameters

Table 1. Measured values $S(F_j)$ and those calculated using three methods.

The transformed values were fitted by linear function (4); see columns X_j and Y_j in Table 2. The regression constants were $A = 5.673642$ and $B = 0.194844$. The inverse transformation has given $a = \exp A = 291.0928$ and $b = 1/B = 5.132311$, so that the two-parameter distribution function is $F(t) = 1 - \exp[-(S/291.0928)^{5.132311}]$. The corresponding calculated values S_j are in column $S_{j,c,lin2}$ and depicted by a curve in Fig. 2.

- b. Two-parameter distribution, application of Solver on untransformed data. In this case, the strength values $t_{j,calc}$ were calculated for the individual F_j values using Equation (7), with $t_0 = 0$. (See column $t_{j,calc,2p}$.) Now, a quantity for characterization of the quality of the fit was defined: sum of the squared differences of the measured and calculated strengths, $\sum(t_{meas} - t_{calc})^2$. The Solver then changes the constants a and b of the distribution function (7) automatically until the sum of squared differences attains a minimum. In the investigated example, the “optimum” constants were $a = 291.7807$ and $b = 4.964505$, near to the results of the linearized problem. The calculated values are in column $S_{j,c,sol2}$.
- c. Three-parameter distribution, application of Solver on untransformed data. The difference from the previous case is the full form of distribution function (7). Also here the sum of the squared differences of the measured and calculated strength, $\sum(t_{meas} - t_{calc})^2$, was minimized. The resultant constants were $a = 154.9796$, $b = 2.4156$, and $t_0 = 133.7975$. The calculated values are in column $S_{j,c,sol2}$. The calculated distribution function is plotted by the thick curve in Fig. 2. Also the distribution function of a two-parameter distribution function is shown (thin curve). The curves for cases (a) and (b) were very close to each other.

Task B. Determination of probability $S \leq 120$ MPa

The probabilities are as follows:

- a. 0.010533, (b) 0.012069, and (c) 0; the minimum possible value is $t_0 = 133.8$ MPa.

Task C. Determination of guaranteed strength

The results are in the following table.

Probability of lower strength	Guaranteed strength (MPa) with the constants from the method:		
	(a)	(b)	(c)
0.05	163.19	160.41	179.12
0.01	118.79	115.51	156.88
0.001	75.78	72.58	142.68

Note the big difference between the two- and three-parameter distributions for very low failure probabilities (cf. also Fig. 2). According to the three-parameter model, the minimum (threshold) strength is 133.8 MPa.

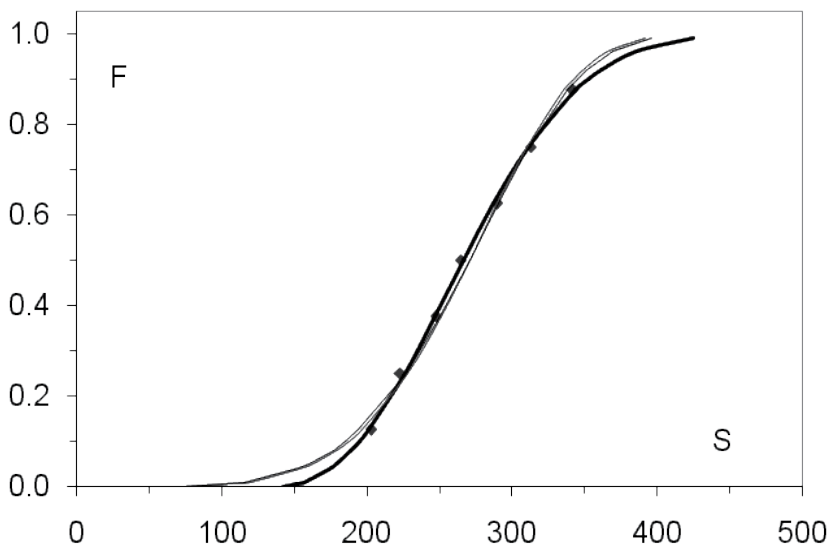


Figure 2. Measured values of strength (S) and approximate distribution functions (F) for various approximations in Example 1. Thick curve – case c, three-parameter function; thin curves – cases a, b, two-parameter curves.

Example 2

Eight components ($n = 8$) were tested until failure. The failures occurred at the following times t_j : 65, 75, 90, 120, 250, 510, 520, and 760 h. Calculate the mean time to failure and failure rate. Calculate also the standard deviation, so that you can assess whether exponential distribution may be used for the time to failure.

Solution.

$$MTTF = \sum t_i/n = (65+75+90+120+250+510+520+760)/8 = 298.750 \text{ h.}$$

The sample standard deviation [Equation (4) in Chapter 2] is $\sigma_{MTTF} = 264.288 \text{ h}$. This is reasonably close to the sample mean, and an exponential distribution may be assumed. For this case, failure rate $\lambda = 1/MTTF = 1/298.75 = 0.003347 \text{ h}^{-1}$. The determination of confidence interval for λ will be demonstrated in a similar case in Chapter 20.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty, University of Pardubice, Czech Republic

References

- [1] Rao S S. Reliability-Based Design. New York: McGraw-Hill; 1992. 569 p.
- [2] Gumbel J E. Statistics of Extremes. New York: Columbia University Press; 1958. 375 p.
- [3] Bentley J P. Introduction to Reliability and Quality Engineering. Harlow, England: Addison-Wesley; 1999. 202 p.
- [4] Abernethy R B. The New Weibull Handbook. 4th ed. North Palm Beach, Florida: Robert B Abernethy; 2001. ISBN 0-9653062-1-6.
- [5] VDA 3. Zuverlässigkeitssicherung bei Automobilherstellern und Lieferanten. Frankfurt am Main: Verband der Automobilindustrie e.V; 1984.

Failure Modes and Effects Analysis

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62364>

Abstract

Failure Mode and Effect Analysis (FMEA) is a simple procedure for systematic revealing of possible failures of structures or processes as early as in the design stage. The main steps of this procedure are explained. Classification of severity, frequency and possibility of early detection of the individual failure modes is shown, as well as the calculation of the risk priority number, which serves for finding the most dangerous causes of failures. The application of FMEA is shown on an example.

Keywords: Failure, failure mode, severity, frequency of occurrence, risk, FMEA

Until now, probabilistic methods were described. In this chapter, a nonprobabilistic method will be explained, which can increase reliability in a very effective way.

Failure modes and effects analysis (FMEA) is a simple procedure for systematic revealing possible failures of a structure or process as early as in the design or project stage and avoiding or mitigating them. The basic idea is that the prevention of failures is better and cheaper than their later detection and repairs. The term failure means here any loss of the ability of the object to perform its functions properly.

FMEA was used for the first time in the Apollo project. Today, it is compulsory in the design of aircrafts; very often it is used in the automotive industry and gradually spreads into other branches. Its use is recommended by quality standards such as ISO 9000. In the past, good designers and builders used a similar approach intuitively. The advantage of FMEA is the fact that it is a systematic procedure guaranteeing that everything will be done to prevent expectable failures of a component, structure, or process. A very important thing is that FMEA is not a matter of one expert only, but uses the knowledge and experience of people from various branches. Their cooperation can have synergic effects and bring further improvements into the design.

Failure modes and effects analysis can be done in 10 steps.

1. Formulation of the problem and establishing a FMEA team

FMEA can be done for a product (component or structure) or a process. A special team is usually formed for the pertinent task. The team should consist of designers, technologists, somebody responsible for the manufacture or building, and somebody representing the future user. His practical experience with the operation and maintenance of similar objects is invaluable.

Every FMEA team has its leader, either appointed by the management or selected by the team itself. The role of the leader is to organize and facilitate the FMEA sessions, to ensure the resources for the work, and to help the team to reach the consensus and to progress toward the completion of FMEA.

Before starting the analysis, it is necessary to define well its scope, the relation of the team to the management, and its competences and responsibility. It is also necessary to set the budget for the analysis as well as the deadline. All this, including the names of the team leader and members and the way of communication with the management should be written down in a document.

2. Review of the construction or process

The purpose for a product FMEA is to reveal problems that could result in safety hazards, product malfunctions, or a shortened life. The key question is "How can the product fail?"

The process FMEA should uncover the problems related to the manufacture, building, or assembly of the product. It is helpful to consider the five elements of a process: **people, materials, equipment, methods, and environment**. With these elements in mind, the key question is "How can the process failure affect the product, processing efficiency, or safety?"

During the first session, the members of the team should be sure that they understand all necessary details of the construction or process and their interrelations. To ensure it, every member should get in advance engineering drawings and documents of the product or a detailed flowchart of the process or operation. It is helpful to have an expert on the construction or process, who will be able to answer any questions the team might have.

3. Revealing of all potential failure modes

Once the team members understand the product (or process), they can begin thinking about the potential failure modes that can affect the product quality, reliability and safety during its

useful life. This should be done during one or more sessions organized according to the rules of brainstorming; information on previous failures is also useful.

In such meetings, no idea or comments should be rejected. However, some people personally involved in the design might feel offended by somebody’s finding the faults and mistakes. The role of the team leader is to facilitate the process, enhance the people to bring ideas and comments, and mitigate some negative psychological effects.

4. Listing of potential effects of each failure mode

Once the possible failure modes have been identified, they are written down into a special form (Fig. 1). Then, the FMEA team reviews each failure mode and identifies the potential effects of the failure should it occur. For every failure mode, there may be one or more effects. Again, everything is written into the FMEA form. This is very important, as this information is the base for assigning risk ratings to each failure mode. It is recommended to use the if-then thinking: “If this occurs, what are the consequences?” The form (Fig. 1) helps in taking measures for the elimination of some failures or reduction of their severity.

Product: _____
 Fmea Team: _____
 Team leader: _____ (part 1 - left side)

F M E A Process									
Ident. No.	Item, Function	Failure Mode	Effect of Failure	Severity	Cause of Failure	Occurrence	Current Controls	Detection	RPN
Total Risk Priority Number :									

FMEA Number: _____
 FMEA Date (original) _____
 (revised) _____
 Page: _____ of _____

(part 2 - right side)

Action Results							
Recommended Action	Responsibility and Target Completion Date	Action Taken	Severity	Occurrence	Detection	RPN	
Resulting Risk Priority Number :							

Figure 1. Failure modes and effects analysis worksheet. (In real worksheets, both parts are printed together.)

5. Assigning severity, occurrence, and detection ratings for each effect

Each effect is assigned three numbers characterizing its severity, frequency, and probability of early detection, and these numbers are written into the left part of the form (upper part of Fig. 1). Often, each of the ratings is based on a 10-point scale, with 1 being the best case and 10 the worst case; for example,

Severity rating scale:

10 – consequences dangerously high (failure could injure or kill); 8 – consequences very serious (failure renders the object unfit for use); 6 – moderate (failure results in partial malfunction); 4 – very low (there is minor performance loss); 3 – minor (the effects could be overcome without performance loss); 1 – none (failure would not be noticeable).

Occurrence rating scale:

10 – very high probability of occurrence [failure is (almost) inevitable]; 8 – high probability (repeated failures); 6 – moderate probability (occasional failures); 3 – low (relatively few failures); 1 – negligible (failure is unlikely).

Detectability rating scale:

10 – probability of detection (POD) is zero (the object is not inspected or the defect is not detectable); 8 – POD is low (the signs of failure are not easily detectable); 3 – POD is high (the signs of failure are easily detectable, the objects are 100% controlled); 1 – detection of approaching failure is certain [the emerging defect is obvious or there is 100% automatic control (regular inspections, if necessary)].

There are no fixed scales; the classification depends on the character of the object. However, it is important to establish a clear description of the points on each scale so that all team members have the same understanding and consensus of the ratings.

When assigning a severity rating, one must be aware that a single failure of a component can have several effects, and each effect can have a different severity.

The best method for determining the occurrence rating is to use actual data from the same or similar product or process. When actual failure data are not available, the team must estimate how often the pertinent failure mode can occur.

The detection rating tells how likely a failure can be revealed before it happens. If there are no controls, the probability of detection is low and the rating high (9 or 10).

6. Calculation of the Risk Priority Number (RPN) for each failure mode

Now, the RPN is calculated by multiplying the severity rating by the occurrence rating and the detection rating for each item (see the special column in Fig. 1):

$$\text{RPN} = \text{Severity} \times \text{Occurrence} \times \text{Detection} . \quad (1)$$

This number for a single item can be between 1 and 1000.

Then, the total RPN can be calculated by summing up the risk priority numbers for all failure modes (Fig. 1, at the bottom of the table). This number alone is meaningless, because each FMEA has a different number of failure modes and effects. However, it can serve for comparison with the revised total RPN once the improving measures have been proposed (see further).

7. Prioritizing the failure modes for action

The failure modes can now be ranked from the highest RPN to the lowest RPN. This can easily be accomplished by common spreadsheet programs (e.g. Excel).

The team must now decide which failure modes will be worked on to reduce their RPN. Usually, a limit value of RPN is chosen, and only those items are dealt with, whose RPN was higher. However, special attention must also be paid to all cases with the highest severity rating, such as 8 – 10.

8. Taking action for eliminating or reducing the high-risk failure modes

Each of the high-risk failure modes is discussed, and the team members propose measures to reduce its RPN. This number is a product of three terms (severity, occurrence, and detectability), and the reduction of each of them will reduce the RPN. However, the best way is to eliminate the reason for particular failure. For example, if a steel component can fail due to corrosion, the use of a stainless steel can fully avoid this danger. If there is no failure, there is no need to reduce its severity or frequency, nor improve its detectability.

Then, measures follow for the reduction of severity of a failure or their frequency. (Some of the failure modes have similar reasons.) Improvement can be reached by new design, by using other components or materials, by the improvement of input control for components or raw materials, and discarding the unsuitable ones. The third way to reduce RPN strives at the improvement of detection of failures in early stages (e.g. by building-in special elements or sensors or by periodic inspections). However, this does not mean an actual improvement of the structure.

9. Calculation of the resulting RPN as the failure modes are reduced

For each item corrected, new ratings are determined (severity, occurrence, and detectability) as well as the risk priority number (see part 2 of Fig. 1). Then, the total RPN is calculated for

the whole structure. This number can often be several tens of percent lower than the original RPN, partly thanks to the elimination of reasons for some failures. The comparison of both RPN shows how effective the FMEA was. It can also help in deciding what measures should be taken in cases of several possible ways of improvement, with different RPNs.

10. Taking action for improvements

The recommended measures for improvement are written into the FMEA form, including their ratings and RPN. However, the most important thing is to ensure that these measures will be realized. Thus, it must also be proposed who will be responsible for the corrective action, the date to which this action should be carried out, and the person who will check it (with respect to the competences of the FMEA team). The final FMEA forms are then submitted to the management.

Concluding remarks

Failure Mode and Effect Analysis, although it is very simple and does not work explicitly with probabilities, can significantly reduce the number of mistakes happening during the design, manufacture, and assembly or building of an object, as well as the number of failures occurring during its life. Thus, FMEA reduces the total costs and increases the safety, reliability, lifetime, and quality of the object. Very often, the design is improved.

Further details on FMEA can be found in the literature, e.g. [1–3]. FMEA has been incorporated into reliability standards, such as IEC 60812, and also commercial computer programs for FMEA are available, although the creation of own, purpose-tailored programs is easy.

A variant of FMEA exists, called FMECA (failure mode, effects, and criticality analysis), which puts more emphasis on the assessment of consequences of possible failures [3]. The principle, however, is the same as above.

Example 1

In a Failure Modes and Effects Analysis, done during the design of a home appliance, five possible failure modes were revealed. Their severity (S), probability of occurrence (O), and possibility of early detection (D) were classified as shown in the table below. Calculate the RPN for each failure mode and the resultant RPN for the whole appliance.

Solution

The individual values of RPN ($=S \times O \times D$) and the resultant value ($=\sum RPN_i$) are written in italics.

Failure mode no.	Severity	Occurrence	Detectability	RPN
1	8	6	2	96
2	4.5	7	2	63
3	6	3	4	72
4	2.5	4	7	70
5	5	6	3	90
Total RPN				391

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
 University of Pardubice, Czech Republic

References

- [1] McDermott R E, Mikulak R J, Beauregard M R. The Basics of FMEA. Portland: Resource Engineering; 1996. 90 p.
- [2] Menčík J. Failure mode and effect analysis - a tool for increasing the reliability and quality of constructions. In: Proc. Int. Conf. „Quality and Reliability in Building Industry“, Levoča, 24. - 26. 10. 2001. Košice: Technical University of Košice; p. 346 – 351
- [3] O’Connor Patrick D T. Practical Reliability Engineering. 4th ed. Chichester: John Wiley & Sons; 2002. 513 p.

Fault Tree Analysis and Reliability Block Diagrams

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62374>

Abstract

Fault tree analysis (FTA) strives to reveal all possible sources of critical failures. It starts from the most critical event (“top event”) and looks at its reasons, and continues in this way backwards to the initial events leading finally to the failure. So-called fault tree, plotted using the symbols of Boolean algebra can then be used for the construction of a reliability block diagram, which serves for finding the critical way and probability of failure. The principle of Markov analysis is explained as well.

Keywords: Failure, fault tree, fault tree analysis, FTA, top event, reliability block diagram, probability of failure, Boolean algebra

The failure modes and effects analysis (FMEA), explained in the previous chapter, strives for finding all possible sources of future failures. It starts with failures of single elements, with mistakes of personnel, etc., and looks for their consequences for the structure or process. It is very efficient but has two drawbacks. First, it reveals perhaps all sources of many possible failures, but only few of them are really serious and have fatal consequences, such as the collapse of the structure. Moreover, complex objects can fail in various ways. Second, FMEA is a rather qualitative analysis and does not give information on the probabilities of failure.

For these reasons, **fault tree analysis (FTA)** is also often used (IEC 61025). In contrast to the “bottom-up” inductive approach of FMEA, the Fault Tree Analysis is a deductive method and goes “top-down”. It starts with the so-called **top event** (critical event; e.g. the aircraft is falling down) and searches for all possible causes (e.g. failure of all engines, a broken wing, or an explosion in the aircraft). Then, the reasons for each of these causes are looked for, and so on, until basic events. If all these events are depicted, showing how the “upper” event follows the “lower” event, and so on, the so-called **fault tree** is obtained, which shows the straightest ways to critical failures. Special symbols are used for creating these diagrams (Fig. 1).

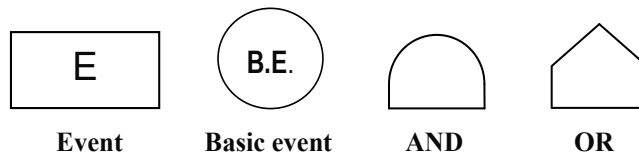


Figure 1. Symbols for Fault Tree Analysis.

A simple example with electric lighting in a room with two lamps is shown in Fig. 2. The top event is “there is darkness in the room”. This can happen if none of the two lamps lights, and four possible reasons exist for this (either both the lamps have failed, there is no voltage in the network, the switch is off or failed, or the fuse has burnt).

A single fault tree is used to analyze one and only one top event (or undesired event). FTA involves five principal steps:

1. Definition of the undesired event to be studied. A system engineer with a deep knowledge of the system can best help to define the undesired events.
2. Obtaining an understanding of the system. Analysts and system designers can help here.
3. Construction of the fault tree.
4. Evaluation of the fault tree.
5. Control of all identified hazards, with the effort to reduce the probability of their occurrence.

In contrast to FMEA, fault tree analysis is able to consider also events caused by external reasons.

Fault tree analysis is often used in the aviation industry, as well as chemical, petrochemical, nuclear power, and other high-hazard industries.

A fault tree can be converted into a **reliability block diagram (RBD)**. This is a scheme similar to Fig. 4 in Chapter 5, with series and parallel arrangement of blocks representing the individual elements or groups of them. Each element is characterized by a failure rate. A series arrangement fails if any of its elements fails. Parallel paths are redundant, that is, all elements must fail for the parallel network to fail. If the probabilities of individual events are known, one can calculate the failure probability of the system, as shown in Chapter 5.

A reliability block diagram RBD may be drawn using switches instead of blocks, where a closed switch represents a working component and an open switch represents a failed component. If a path may be found through the network of switches from the beginning to the end, the system is still working. The system can also be solved using the rules of Boolean algebra. Series paths can be replaced by AND gates and parallel paths with OR gates, etc.

In complex systems consisting of many blocks, various blocks can fail simultaneously. If connections exist between certain elements, the failure of one or even more blocks does not

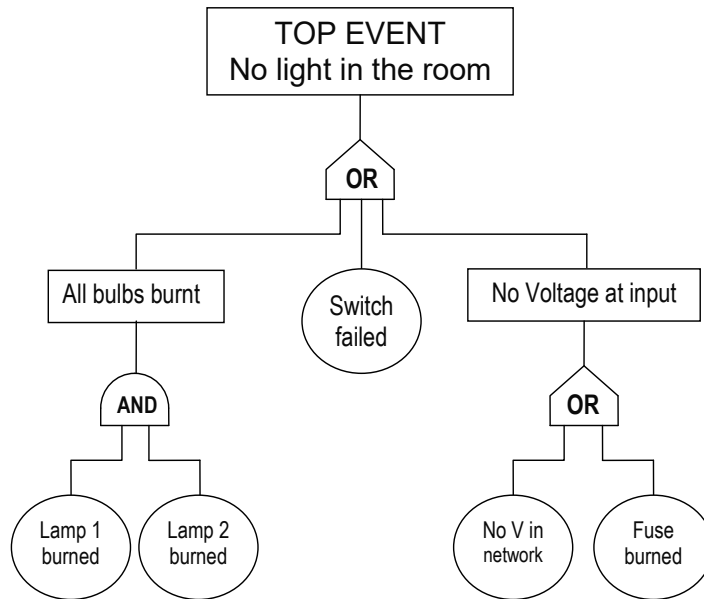


Figure 2. Fault tree for two lights in a room.

necessarily mean the failure of the whole system. Reliability in such systems is studied by the **cut set** or **tie set** methods. A cut set is obtained by drawing a line through the blocks, whose failures would cause the failure of the system. Tie sets are obtained by drawing lines through such blocks, which, if working, would ensure the operation of the system. This analysis helps in revealing the possible conditions for failure or in finding an arrangement with high resistance to failure.

Another approach to reliability analysis of complex systems uses the so-called Markov chains or **Markov analysis**. This analysis is suitable for systems whose components can be in two states, failed or not failed, and transitions from one state to another can happen from time to time. The analysis can be applied in cases where the response (or change of state) at a certain instant does not depend on previous events (so-called memoryless system) and the probabilities are known for the transition from operable state to failed state and vice versa; these probabilities are assumed constant. Markov analysis enables one to trace how the system evolves in time from certain initial conditions, and to see how quickly (and whether) it approaches to a steady state after a disturbing event. For example, if the probability of transition from the available state to a failed one is $P_{A \rightarrow F}$ and from the failed state to an available state is $P_{F \rightarrow A}$, and if the component was initially available ($P_0 = 1$), then the probability that it will be in a failed state after one step equals $P_1(F) = P_{A \rightarrow F}$. The probability of a failed state after the second step is $P_2(F) = P_{A \rightarrow F} \times P_{F \rightarrow F} + P_{A \rightarrow A} \times P_{A \rightarrow F}$; here the probability of the transition from the failed state into the failed state is $P_{F \rightarrow F} = 1 - P_{F \rightarrow A}$, whereas the probability of transition from the available state into an available state is $P_{A \rightarrow A} = 1 - P_{A \rightarrow F}$. The evolution can be depicted using Markov state transition diagrams and tree diagrams, which become

more and more complex with each step. Computer support is thus necessary. Markov analysis is used, for example, for the simulation and analysis of reliability of systems for electricity supply or reliability of software.

More details to fault tree analysis and reliability block diagram can be found in the literature [1 - 3]. These methods have also been incorporated into reliability standards, e.g. IEC 61025, and commercial computer programs for FTA are also available. More about cut set and tie sets can be found in [2, 3], more about Markov analysis is in [3 - 5].

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

References

- [1] Bentley J P. Introduction to Reliability and Quality Engineering. Harlow, England: Addison-Wesley; 1999. 202 p.
- [2] Ireson W G, Coombs C F Jr, Moss R Y. Handbook of reliability engineering and management. 2nd ed. New York: McGraw-Hill; 1996. 816 p.
- [3] O'Connor Patrick D T. Practical Reliability Engineering. 4th ed. Chichester: John Wiley & Sons; 2002. 513 p.
- [4] Freedman D. Markov Chains, Berlin: Springer; 1983. 382 p.
- [5] Bednařík J et al. Reliability techniques in electronic practice (In Czech: Technika spolehlivosti v elektronické praxi). Praha: SNTL; 1990. 336 p.

Load – Resistance Interference Method

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62368>

Abstract

Reliability and safety of a load carrying structure needs that its resistance R must be higher than the load effect S . So-called reliability margin $G = R - S$ and reliability index are used for reliability assessment and determination of failure probability if R and S are random quantities. This chapter explains the determination of parameters of the reliability margin and shows its use on examples, including the finding of suitable dimensions for achieving the demanded reliability.

Keywords: Reliability, safety, load, resistance, reliability margin, reliability index, failure, interference, probability

Many situations exist, which can be characterized as the conflict “load-resistance” or “action-barrier”. The reliability is ensured if the load effect is smaller than the resistance against it. An example is a load-carrying structure, such as a road bridge or a mast of a TV transmitter exposed to wind. If the instantaneous load acting on the structure is higher than its load-carrying capacity, the structure can collapse or its deformations will be larger than allowed. Several examples follow. If the voltage at the input of a device is higher than its electric strength, a breakdown of insulation will follow. If the amount of water, flowing into a reservoir during rain period, is higher than its capacity at that time, the water overflows the upper edge. The strength of a shrink-fitted connection depends on the overlap of the bolt in the hole (i.e. on the difference between the diameter of the bolt and the diameter of the hole). If this overlap is too small, the strength of the joint is insufficient. If the bus arrives at the train station later than at the time of the train departure, the passengers miss the journey. The consequences of these failures can range from negligible to fatal.

In all these cases, a tool is needed that can quantify the reliability. The object is reliable if its **resistance R** to a certain “load” is higher than the **load effect** or **stress S** . (The meaning of the

terms “load” and “resistance” can be very broad depending on the context.) For the quantification of reliability, the so-called **reliability margin** G is introduced, defined as

$$G = R - S. \quad (1)$$

This reliability margin shows, for example, how much the load-carrying capacity is higher than the load or how many minutes remain between the arrival and departure. The reliability condition can be written as

$$G = R - S > 0. \quad (2)$$

The case $G < 0$ corresponds to the resistance lower than load, which means failure.

Often, only the load varies (the wind force, for example), whereas the resistance R of a structure is constant. In such case, only the stress S is a random quantity, and the probability of failure is determined as the probability that S exceeds the value of R ,

$$P_f = P(S \geq R). \quad (3)$$

If the distribution function of the wind-caused stress in the structure, $F(S)$, is known, then its value corresponding to the known value R gives the probability that the stress will be lower than the strength, and the structure is safe. The probability of failure is the complement,

$$P_f = 1 - F(S = R). \quad (4)$$

Note that here the letter F denotes distribution function and the probability of failure is P_f . Vice versa, it is possible to determine the necessary strength R of the structure such that the probability of failure will not exceed the allowable value $P_{f,a}$.

Often, also the random variability of the resistance R must be considered. It is especially during the design stage that the actual parameters of the structure are not known yet: for example, strength or Young’s modulus of the material, characteristic stiffness of rubber bearings in a bridge, thickness of the flanges of rolled steel beams, etc. Only their nominal values are known in advance. The actual values vary randomly less or more around them and can be determined accurately only after the components have been purchased or manufactured.

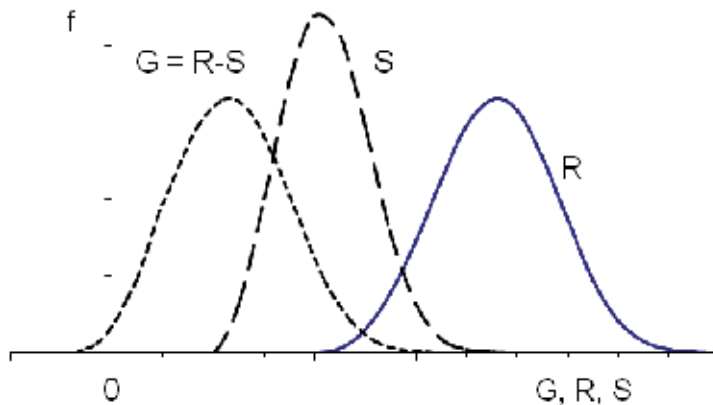
In such cases, both quantities R and S must be considered as random. They can be characterized by probability distribution or simply by the average value and standard deviation. The situation is depicted in Figure 1. If the distributions of both quantities do not overlap at all, no failure can occur. This can be achieved if the average resistance is sufficiently higher than the average load. However, the effort to ensure that the distributions R and S never overlap can be uneconomical, especially if the consequences of failure are not critical. Sometimes, it is

reasonable to admit a reasonably low probability of failure (e.g. small and short-time exceeding that of the allowable deformation). In this case, both distributions overlap (Fig. 1) and the probability of failure is proportional to the area below the overlapping part of both curves (see further). It is thus useful to know the failure probability for certain combinations of S and R or, vice versa, to determine in advance what cross-section dimensions should be used if the failure probability must not exceed some allowable value. The pertinent procedure, called the “load – resistance” or “stress – strength interference method” [1, 2], is explained further.

If the stress S and resistance R are random quantities, the reliability margin G is also a random quantity (Fig. 1), with its own probability distribution. Its mean μ and standard deviation σ can be calculated as

$$\begin{aligned} \mu_G &= \mu_R - \mu_S, & (a) \\ \sigma_G &= (\sigma_R^2 + \sigma_S^2)^{1/2}; & (b) \end{aligned} \tag{5}$$

the subscripts denote the corresponding quantities. When working with empirical data, μ and σ are replaced by the sample mean m and sample standard deviation s .



S – stress; R – resistance (strength); G – reliability margin.

Figure 1. Stress – strength interference method (a schematic).

Failure occurs if the load effect is higher than the resistance, i.e. if the reliability margin G in Eq. (1) is negative (Fig. 1). The corresponding critical value of G is 0, and the probability of failure P_f can be determined as

$$P_f = P(R \leq S) = P(G = 0) . \tag{6}$$

Reliability margin can be standardized to a nondimensional form in the following way:

$$u = (G - m_G) / s_G. \quad (7)$$

Its value for the transition between reliable state and failure ($G = 0$) equals

$$u = (0 - m_G) / s_G = -m_G / s_G. \quad (8)$$

The ratio

$$m_G / s_G = (m_R - m_S) / (s_R^2 + s_S^2)^{1/2} = \beta \quad (9)$$

is called the **reliability index** β and corresponds to the distance of the mean value of reliability margin G from 0, which is expressed as a multiple of standard deviation of G . The reliability index gives a simple measure of reliability, as it shows how far is the average value of reliability margin from the critical point. The situation is simple if the reliability margin G has normal distribution: in this case, an unambiguous relationship exists between β and the probability of failure: P_f equals the value of distribution function of standard normal distribution for $u = -\beta$. For example, $P_f = 0.02275$ for $\beta = 2$, $P_f = 0.00135$ for $\beta = 3$, and $P_f = 0.0000317$ for $\beta = 4$. Some standards for civil engineering constructions admit the reliability evaluation using the reliability index and give the characteristic values of β for various degrees of safety [3].

The advantage of the reliability index is that it is simple and its values are of the order of units, which is near to our way of thinking. A normal distribution of G may be assumed if the coefficient of asymmetry $\alpha_G < 0.3$. Otherwise, no simple relation between β and P_f exists, and other methods for the determination of failure probability are more appropriate (see further).

As the S and R curves can overlap or interfere (Fig. 1), the term **interference method** is used for this way of reliability assessment. Its use will be illustrated on two simple examples.

Example 1

Determine the probability of failure of a pull rod loaded by tensile force. The force magnitude is normally distributed with the mean $m_F = 140,000$ N and standard deviation $s_F = 14,000$ N. The diameter D of the rod is 20 mm. The stress is determined as $\sigma = F/A$, where the cross-section $A = \pi D^2/4 = 314.16$ mm². The mean and standard deviation of the stress are: $m_S = 140,000/314.16 = 445.6$ MPa and $s_S = 14,000/314.16 = 44.6$ MPa. The strength parameters of the used steel are: $m_R = 500$ MPa and $s_R = 50$ MPa. One can assume a normal distribution of R and S as well as of G .

Solution. The mean and standard deviation of reliability margin are $m_G = m_R - m_S = 500.0 - 445.6 = 54.4$ MPa and $s_G = (s_R^2 + s_S^2)^{1/2} = (50.0^2 + 44.6^2)^{1/2} = 67.0$ MPa. The reliability index is $\beta = m_G/s_G = 54.4/67 = 0.8117$. The probability of failure is $P_f = F(-\beta) = F(-0.8117) = 0.20848 = 20.85\%$. (Various programs can be used for finding the values of standard normal distribution function F ; the

appropriate command in Excel is "norm.s.dist" or "normsdist". The statistical tables for the distribution function of standard normal distribution give the same result.)

Example 2

The failure probability from Example 1 is too high and must be reduced to $P_f = 0.0001$. Find the appropriate diameter of the rod.

Solution. The reliability index for $P_f = 0.0001$ is $\beta = 3.719$. The material parameters m_R and s_R are the same as above, so that it is necessary to determine only the stress parameters. In fact, there are two unknown parameters, m_s and s_s . However, we can assume that the coefficient of variation v of the slightly larger cross-section will be the same as in the first variant (i.e. 10%; cf. the m_s and s_s values above), so that $s_s = vm_s = 0.1m_s$. Thus, only the mean stress m_s (and the corresponding rod diameter) are to be determined. Several possible methods for finding m_s exist. The first one, exact, is based on the solution of Equation (5) for given β , m_R , s_R , and unknown m_s . This approach leads to a quadratic equation and could be preferred by those who like mathematical analysis. The second approach uses the formula for the calculation of β , varies step-by-step the value m_s or rod diameter, and calculates repeatedly β or the failure probability until the target value of β or P_f is found. This solution can be facilitated using a suitable solver: if the formula for the calculation of failure probability as a function of shaft diameter was created [using relationships $P_f = F(-\beta)$, Eq.(7), and $A = \pi D^2/4$, $\sigma = F/A$], it is possible to "ask" the solver to change the diameter D until P_f attains the demanded value. In this way, the solver in Excel has given the (accurate) value $m_s = 285.81$ MPa (for $\beta = 3.719$). The corresponding cross-section area (for the load 140,000 N) is $A = F/m_s = 140,000/285.81 = 489.84$ mm², and the rod diameter is $D = 24.97 = 25.0$ mm. (The reader is encouraged to make the check by calculating the cross-section area A , mean stress m_s , reliability index β , and failure probability P_f for this diameter.) Note how dramatically the failure probability has decreased (from 0.2048 to 0.0001) by increasing the rod diameter from 20 to 25 mm.

These computations can be done even for the values of D or m_s chosen "by hand", without a special algorithm. This "primitive" approach, which does not need analytical abilities or solver, also leads quickly to an acceptable solution, the more so that some quantities (e.g. dimensions of standard rolled steel profiles) are not continuous, but graded.

Other distributions and approaches

If the stress and resistance have asymmetrical distributions that can be approximated by log-normal functions, the above approach may be used if the reliability condition is defined not as the difference of the strength and stress, Equation (1), but as their ratio:

$$G = R / S. \tag{10}$$

Taking the logarithms of Equation (8) gives an expression similar to Equation (1):

$$\log G = \log R - \log S. \tag{11}$$

Both transformed quantities, $\log R$ and $\log S$, have normal distribution and Equation (9) resembles Equation (1) in transformed coordinates ($\log T$ corresponding to G , etc.). Thus, Equation (9) can be treated by the procedures of interference method described above.

If the distributions of the resistance and stress are only known in the form of histograms, the probability of failure can be determined by numerical integration:

$$P_f = \int_{-\infty}^{\infty} F_R(S) f_S(S) dS, \quad (12)$$

and the probability that failure does not occur:

$$P_f = \int_{-\infty}^{\infty} F_S(R) f_R(R) dR \quad (13)$$

depending on what functions are available. The differentials dS and dR are replaced by finite intervals ΔS and ΔR . For more, see [2].

The probability of failure can also be determined by numerical simulation methods, such as Monte Carlo, which will be explained in the following chapter.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty, University of Pardubice, Czech Republic

References

- [1] Sundararajan C, editor. Probabilistic Structural Mechanics Handbook. New York: Chapman & Hall; 1995. 745 p.
- [2] Mrázik A. Theory of reliability of steel structures (In Slovak: Teória spoľahlivosti oceľových konštrukcií). Bratislava: VEDA; 1987. 360 p.
- [3] ENV 1993-1-1 Eurocode 3. Design of steel structures. (Czech standard ČSN 731401, 1998.)

Monte Carlo Simulation Method

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62369>

Abstract

The Monte Carlo method studies random phenomena using numerous fictitious experiments with computer-generated random numbers. Its principle is explained and also the principle of generation of random numbers with various probability distributions. Also more complex cases, such as the response surface method and generation of correlated random quantities are explained. The use of the Monte Carlo method is illustrated on several examples.

Keywords: Probability, random numbers, Monte Carlo method, correlation, response surface method, probabilistic transformation

Random phenomena or processes can be successfully studied by the Monte Carlo method [1 - 4]. This is a probabilistic method based on performing numerous fictive experiments using random numbers. It is used in various branches of science and technology. For example, in reliability it serves for the analysis of load-carrying capacity or deformations of a construction, for the determination of time to failure, resonant frequency of a mechanical structure, or an electric circuit, or for the study of behavior of a complex transport or production system.

The Monte Carlo method is close to the engineering way of thinking. It is universal and does not need a special knowledge of probability theory. The only information it needs is the relationship between the output and input quantities,

$$y = f(x) , \text{ or } y = f(x_1, x_2, x_3, \dots) , \quad (1)$$

and the knowledge of probability distributions of the input variables. The method uses numerous repetitions of trials with computer-generated random numbers and the relevant

mathematical operations. In each "trial", the input variables x_1, x_2, \dots, x_n are assigned random values, but such that their distributions correspond to the probability distribution of each variable. With these input values, the output quantity y is calculated via Equation (1). From the results, a histogram can be constructed (Fig. 1), which corresponds to the distribution of y .

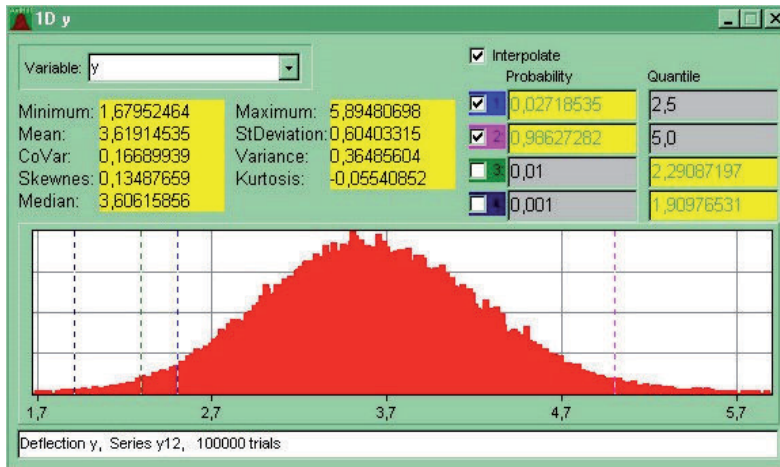


Figure 1. Histogram obtained by the Monte Carlo simulation program Ant-Hill [2].

The generated y values can be used for the determination of the average value and standard deviation, but also the probability that y will be lower or higher than a chosen value or lie within some interval. Also, the characteristic values that will be exceeded with higher probability than some allowable value (e.g. the guaranteed strength or time to failure). In a similar way, critical values can also be obtained, whose exceeding can be expected with only small probability (e.g. maximum load or deformation).

Today, various commercial computer programs exist for Monte Carlo simulations, but they can also be created. The base of such program is a generator of random numbers. Actually, they are not truly random but computer generated using a suitable deterministic algorithm. However, such algorithms are used, which generate numbers behaving nearly as if they were random.

The principle of these generators is simple. For example, the so-called congruential generator gives random numbers, distributed uniformly in the interval $(0; 1)$, in the following way. One number is chosen as the base for the series of random numbers u (e.g. $u_0 = 0.5284163$). Now, in the first step, this number is multiplied by some suitable number Q , for example 997. The product is $997 \times 0.5284163 = 526.8310511$. The first random number u_1 is then created as the part of the result, lying behind the decimal point; in our case, $u_1 = 0.8310511$. In the second step, u_1 is again multiplied by the same number Q , $997 \times 0.8310511 = 828.5579467$, and the second random number is created as the decimal part of the result (i.e. $u_2 = 0.5579467$). This procedure is repeated again and again. The formula for the random number in the j -th step is

$$u_j = \text{MOD}(Q \times u_{j-1}) ; \tag{2}$$

u_{j-1} is the random number from the previous step, and the symbol MOD (read "modulo") means the decimal part of the expression in brackets. The reader is encouraged to make several steps in this way; for a check, $u_3 = 0.2728599$. A long series of these numbers has approximately uniform distribution. Also other algorithms exist. For example, one generator of pseudorandom numbers with normal distribution uses the central limit theorem, etc. In any case, the use of "commercial" generators is strongly recommended, as they have undergone thorough statistical testing to prove that they behave as really "nearly random", and the period, after which the series of generated numbers is repeated, is very long, hundreds of millions of numbers. Generators of random numbers are usually a part of universal computer programs or languages. Even Excel can produce such numbers. Better tools are included in special packages for probabilistic analysis of reliability, such as FREET (www.freet.cz) or Ant-Hill (www.sbra-anthill.com), as well as in Matlab and other advanced software.

1. Creation of random numbers with nonstandard distributions

The commercial programs offer often-used distributions, such as uniform or normal. The random numbers, corresponding to other analytically defined distributions, can be generated via uniform distribution. The basic idea is that the distribution function F for any continuous random quantity is also random variable, distributed uniformly in the interval (0; 1). Thus, if the distribution function of random quantity x is described by the expression $z = F(x)$, then the random numbers of x can be obtained from the random numbers z with uniform distribution in (0; 1) using the inverse formula:

$$x = F^{-1}(z) . \tag{3}$$

Here, F^{-1} means inverse probabilistic transformation (Fig. 2). For example, the distribution function for exponential distribution is $z = F(x) = 1 - \exp(-x/x_0)$, with the parameter x_0 . The inverse transformation for this distribution is $x = -x_0 \ln(1 - z)$;

In some cases, the distribution of a random input quantity has a more complex shape and can be described by a histogram, obtained from experiments or monitoring. This histogram is then used for the construction of distribution function $F(x)$. This function can be approximated either by constant values of F in the individual subintervals of x (if the number of classes is high) or by interpolation within each class, such as

$$F(x) = F_i + \frac{F_{i+1} - F_i}{x_{i+1} - x_i} (x - x_i), \quad x = x_i + \frac{F(x) - F_i}{F_{i+1} - F_i} (x_{i+1} - x_i) \tag{4}$$

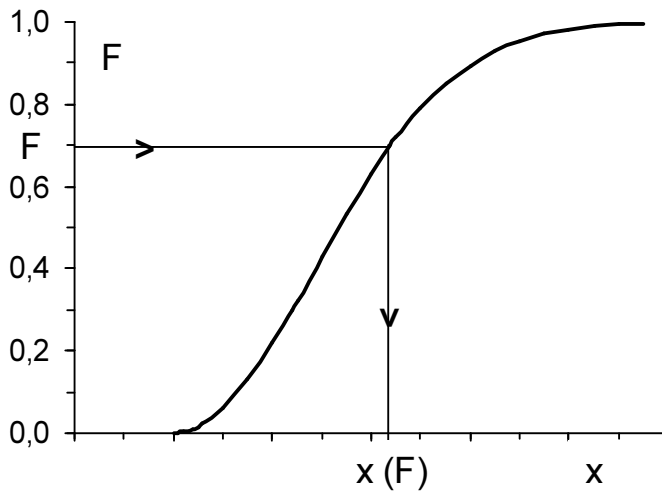


Figure 2. Generation of random numbers x by inverse probabilistic transformation.

$i = 1, 2, \dots, n$ denotes the interval. The formula on the right gives x corresponding to the probability F . The F values are then generated as random numbers with uniform distribution.

A typical feature of the Monte Carlo method is that the characteristic values (average, quantiles, probabilities corresponding to certain values of y , etc.), obtained as a result of n trials, are never the same as the results obtained in any other set of simulations. The results are thus only approximate and are closer to the actual values for a higher number of trials. The number of simulation steps n , needed for attaining some accuracy of results, is given approximately by the formula:

$$n = u_{\alpha/2}^2 (1 - P) / (P\delta^2) ; \tag{5}$$

P is the expected (estimated) probability of the investigated phenomenon, δ is the allowed relative error in the determination of P , $u_{\alpha/2}$ is the $\alpha/2$ -critical value of standard normal variable, and α is the probability that the actual value of P will lie outside the interval $P \pm \delta$. The necessary number of simulations thus grows significantly with decreasing probability. For example, if the assumed probability P is 0.01 and the allowed relative error δ is 10 % and confidence level $\alpha = 5\%$ (with $u_{\alpha/2} = 1.96$), then $\approx 40,000$ simulation trials are necessary. For $P = 0.0001$, it is as much as 4,000,000 trials, etc. [Note: Equation (5) is based on the fact that the number of outcomes of an event of probability P in n repetitions has binomial distribution, and this distribution can be approximated by normal distribution for high n .]

The characteristic features of the Monte Carlo method are illustrated on several examples at the end of this chapter. The reader is encouraged to work them out on a PC.

2. More complex cases — Response Surface Method

The direct use of the Monte Carlo method is suitable for simple relationships $y = f(x_1, x_2, \dots)$. Often, the response must be obtained by numerical solution (e.g. the finite-element method). If one trial lasts minutes or more, the thousands of simulations would consume too much time. In these cases, more effective is the combination of Monte Carlo with the response surface method (RSM). The principle is that the response is first calculated only for selected values of input variables, the results are fitted by a simple regression function (response surface; Fig. 3), and the Monte Carlo trials are done with this function.

The relationship between the output quantity y (deformation, load-carrying capacity, amplitude of vibrations, etc.) and the input variables can often be fitted by a polynomial function:

$$y = a_0 + \sum a_i x_i + \sum b_i x_i^2 + \dots + \sum c_{ij} x_i x_j + \dots \quad (6)$$

This approximation is suitable if the actual relationship between input and output has a similar character (e.g. $y \sim x^3$) or if the output quantity changes in the considered interval only little. If it differs significantly from a polynomial (e.g. $y \sim 1/x^3$ or $y \sim x^{1/2}$), expression (6) cannot give a good approximation in a wider interval. There are several ways for improvement, the starting point being a visual judgment of the character of the response. Linear or polynomial function may be used for the approximation of other relationships if suitable transformations are made. For example, the relationship $y = a/x^3$ can be expressed as $y = az$ by introducing a new variable $z = 1/x^3$; the relationship $y = ax_1/x_2^2$ can be converted to multiple linear regression $Y = A_0 + A_1 X_1 + A_2 X_2$ using logarithmic transformations, etc. Solvers in universal software (Excel, Mathcad, or Matlab) can find regression coefficients directly, without transformations.

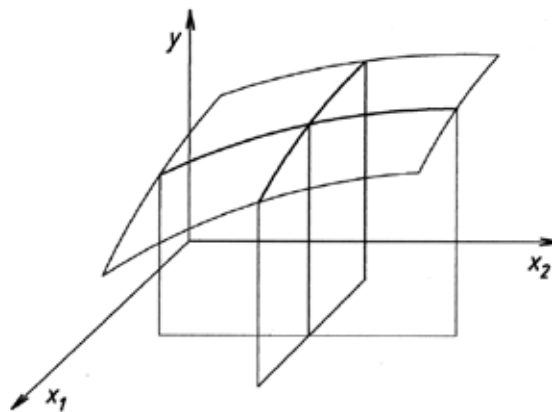


Figure 3. Response surface for two independent variables (a schematic, with cuts $x_1 = \text{const}$, $x_2 = \text{const}$).

The fit of response function can sometimes be improved by dividing the definition interval of some input quantities into subintervals and using different regression functions for each. This may be substantiated by the physical character of the problem. For example, the elastoplastic deflection of a beam obeys another law than purely elastic deformations.

The quality of the fit can be evaluated by means of residual standard deviation s_{res} , which characterizes the scatter of the measured values around the regression function. Also the maximum difference between the individual "accurate" values and those on the response surface can serve as a criterion. With good response surface, the individual differences are randomly positive and negative. Larger regions with differences of the same sign indicate that the chosen function does not correspond well to the character of the actual response.

3. Application of the Monte Carlo method for correlated quantities

The application of Monte Carlo simulations in problems with several input variables is simple if the individual input quantities are mutually independent (e.g. Young's modulus and the cross-section area of a beam). Sometimes, however, a relation between them exists; (e.g. between mass density and Young's modulus of concrete). In such case, one speaks about statistical dependence or correlation. A special case is the so-called autocorrelation, when the value of a random quantity at some point is related partly to the values at neighboring points or in preceding times. Examples are the properties of concrete or of soil at foundations or the temperature of a building structure: it varies during a day or from a day to day, but depends partly also on the season in the year.

The omission of correlations in the simulations can lead to errors. For example, a very low value of elastic modulus of concrete could be generated simultaneously with a very high value of strength, but this does not correspond to reality. If correlations are respected, the calculations reflect the reality better and the conclusions or predictions are more accurate, with smaller scatter. Sometimes, also, a quantity needed for the analysis is unavailable, but can be replaced by a correlated quantity. For example, if the direct measurement of the tensile strength of components in an existing massive steel structure is impossible, the information from hardness tests can sometimes be adapted.

The tightness of the relationship of two quantities is characterized by the correlation coefficient r , defined as

$$r_{ij} = cov(x_i, x_j) / (s_i s_j), \quad (7)$$

where $cov(x_i, x_j)$ is the covariance of x_i and x_j , and s_i and s_j are their standard deviations. The correlation coefficient r is nondimensional, ranging from -1 to $+1$. For $r = 0$, no mutual relationship exists, whereas $r = +1$ or -1 mean deterministic (functional) relationship. For $r > 0$, the x_j values grow with the growth of x_i and decrease for $r < 0$. (Note: The correlation coefficient is also equal to the square root of the coefficient of determination r^2 , given by programs for

curve fitting, available also in Excel.) Three examples with various values of r are shown in Figure 4.

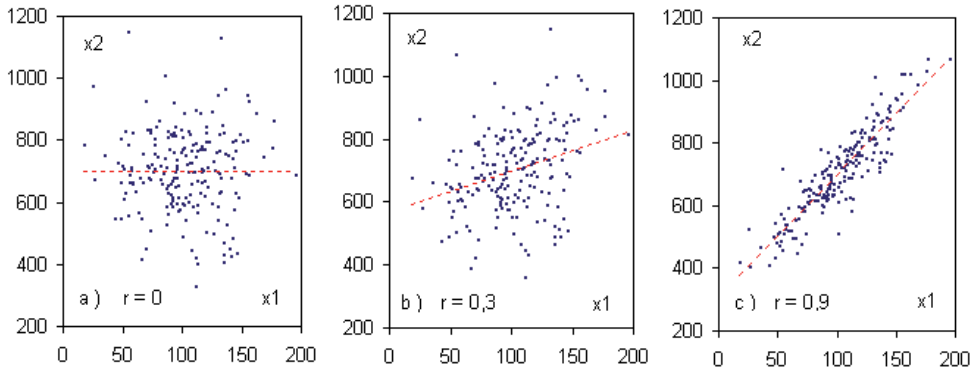


Figure 4. Two correlated quantities x_1 and x_2 with the same mean values ($\mu_1 = 100$ and $\mu_2 = 700$) and standard deviations ($\sigma_1 = 30$ and $\sigma_2 = 150$) and various correlation coefficients r [5].

If two correlated random quantities x_1 and x_2 should be generated, and if the regression function $x_{2,reg} = f(x_1)$ is known, as well as the coefficient of determination r^2 of this approximation, the following procedure may be used. First, the random value of x_1 is generated. Then, the corresponding value of x_2 is generated as [5, 6]

$$x_2 = f(x_1) + \Delta x_2 = f(x_1) + u s_{2,res} = f(x_1) + u s_2 (1 - r^2), \quad (8)$$

where $s_{2,res}$ is the residual standard deviation of quantity x_2 around the regression function f , and u is the quantile of standard normal distribution (provided that the distribution of individual values x_2 around f is normal). The right-hand part of Equation (8) uses the fact that the residual deviation $s_{2,res}$ of x_2 can be expressed by means of the standard deviation s_2 of x_2 and the coefficient of determination r^2 pertaining to the regression function $x_2 = f(x_1)$.

This approach may be used for linear, as well as nonlinear relationships between x_1 and x_2 . With some modification, it may also be used for multiple regression [6].

More information on the Monte Carlo method, especially on its use in the assessment of reliability of structures, including load-carrying capacity and lifetime, can be found in books [1 – 3] and proceedings of conferences [4], which contain many practical examples.

Example 1

Generate (e.g. using Excel) 500 random numbers with normal distribution with the mean $\mu = 5$ and standard deviation $\sigma = 1$. Calculate the sample average m and standard deviation s and compare them with μ and σ . Determine also the 5% quantile and the probability that x will be

larger than 8.0 and compare them with the exact values $x_{0,05} = 3.35515$, and $P(x > 8.0) = 0.99865$. Repeat the procedure and look at the new results. You can also make similar simulations with a lower number of generated values (e.g. $n = 50$) and also with a higher number ($n = 5000$).

Remark: Excel was mentioned here because it is ubiquitous and its use is easy. Everybody can thus try to solve such problems. The necessary routines Descriptive statistics, Histograms, Generation of random numbers, and Solver are installed in every Excel. However, they are not always directly available. If the command Data analysis does not appear on your screen after the command "Data", it must be activated. The procedure is as follows. Click on the button File, then on Possibilities (in this menu), then Add-Ins, then Analytical Tools (and Solver), and, finally, OK. After next pressing the command Data, the buttons Data Analysis and Solver appear in the upper part of the screen. It would be a pity not to use such powerful tools !

Example 2

Generate 10,000 random numbers (x) with uniform distribution in the interval (0; 1). Calculate the mean value and standard deviation (accurate values are $\mu = 0.5$ and $\sigma = 12^{-1/2}$), plot the histogram, and check if it corresponds to uniform distribution. Now, generate the second series of 10,000 numbers (Y) with the same parameters. Make the sums of two numbers with the same subscript j . Calculate now the parameters and create the histogram of the resultant distribution. (The accurate value of the mean is $\mu = \mu_1 + \mu_2 = 0.5 + 0.5 = 1$, and the distribution is triangular.) If you would make (in similar way) the sum of three and more random quantities with uniform distribution, you will see that the resultant distribution resembles more and more normal distribution, in agreement with the central limit theorem.

Example 3

Determine (by the Monte Carlo method) the mean time to failure from Example 5 in Chapter 5 (four elements in a series, each of exponential distribution with $\lambda_1 = 8 \times 10^{-6} \text{ h}^{-1}$, $\lambda_2 = 6 \times 10^{-6} \text{ h}^{-1}$, $\lambda_3 = 9 \times 10^{-6} \text{ h}^{-1}$, and $\lambda_4 = 2 \times 10^{-5} \text{ h}^{-1}$).

Solution.

The problem was solved using Excel. First, the mean times to failure of individual components were calculated (in standard way) as $m_{t,i} = 1/\lambda_i$ (i.e. $m_{t,1} = 125,000 \text{ h}$, $m_{t,2} = 166,667 \text{ h}$, $m_{t,3} = 111,111 \text{ h}$, and $m_{t,4} = 50,000 \text{ h}$). In the second step, four series of random values of time to failure of individual components were generated via inverse formula $t = -m_i \ln(1 - F)$ from random numbers F with uniform distribution in the interval (0; 1) and for the above parameters $m_{t,1}, \dots, m_{t,4}$. Each series had 1000 numbers. Then, in each simulation trial, the minimum of the four random numbers, corresponding to the times to failure of individual elements, was found, because the failure of the system of several components in a series occurs as the first element fails. In this way, the mean time to failure of the series system was $m_t = 23,376.3 \text{ h}$, which is very near the theoretical value 23,256 h. The standard deviation was $s_t = 24,137.7 \text{ h}$, again near the mean value, as typical for exponential distribution. (A histogram will confirm it.)

The mean values of the simulation series for the individual components are as follows (the numbers in brackets express the standard deviation): No. 1: 120,073 (126,158), No. 2: 161,217

(162,127), No. 3: 104,902 (103,727), and No. 4: 52,001 (53,723), everything in hours. One can see that the parameters of the generated variables are all near the parent parameters. With higher numbers of simulation trials, the differences would be even smaller.

Remark: Systems with parallel arrangement of components can be solved in a similar way, but instead of searching for the minimum of the times to failure in each trial, now the maximum will be sought, because the parallel system fails only after the failure of the component with the longest time to failure.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

References

- [1] Hammersley J M, Handscomb D. Monte Carlo Methods. New York: John Wiley; 1964. 184 p.
- [2] Marek P, Brozetti J, Guštar M, Tikalsky P, editors. Probabilistic Assessment of Structures using Monte Carlo Simulation. Prague: ITAM CAS CR; 2003. 471 p.
- [3] Marek P, Guštar M, Anagnos T: Simulation-based reliability assessment for structural engineers. Boca Raton: CRC Press; 1996. 384 p.
- [4] Marek P (editor): Reliability of constructions (in Czech: Spolehlivost konstrukcí). Proceedings of conferences. Ostrava: Dům techniky Ostrava; 2000–2012. ISBN 80-0201708-0, 800201489-8, 80-0202007-3, 80-0286246-33-8, ISBN 80-0201551-7.
- [5] Menčík J. Simulation assessment of reliability with correlated quantities. (In Czech: Simulační posuzování spolehlivosti při korelovaných veličinách.) In: Spolehlivost konstrukcí (Reliability of structures), 23.-24. 4. 2003; Ostrava: Dům techniky; 2003. 151 – 156.
- [6] Čačko J, Bílý M, Bukovec J. Measurement, evaluation and simulation of random processes. (in Slovak: Meranie, vyhodnocovanie a simulácia náhodných procesov.) Bratislava: VEDA; 1984. 210 p.

Latin Hypercube Sampling

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62370>

Abstract

The simultaneous influence of several random quantities can be studied by the Latin hypercube sampling method (LHS). The values of distribution functions of each quantity are distributed uniformly in the interval (0; 1) and these values of all variables are randomly combined. This method yields statistical characteristics with less simulation experiments than the Monte Carlo method. In this chapter, the creation of the randomized input values is explained.

Keywords: Probability, Monte Carlo method, Latin Hypercube Sampling, probabilistic transformation, randomization

The Monte Carlo method has two disadvantages. First, it usually needs a very high number of simulations. If the output quantity must be obtained by time-consuming numerical computations, the simulations can last a very long time, and the response surface method is not always usable. Second, it can happen that the generated random numbers of distribution function F (which serves for the creation of random numbers with nonstandard distributions) are not distributed sufficiently and regularly in the definition interval (0; 1). Sometimes, more numbers are generated in one region than in others, and the generated quantity has thus somewhat different distribution than demanded. This problem can appear especially if the output function depends on many input variables.

A method called Latin Hypercube Sampling (LHS) removes this drawback. The basic idea of LHS is similar to the generation of random numbers via the inverse probabilistic transformation (3) and Figure 2 shown in Chapter 15 [1, 2]. The difference is that LHS creates the values of F not by generating random numbers dispersed in chaotic way in the interval (0; 1), but by assigning them certain fix values. The interval (0; 1) is divided into several layers of the same width, and the x values are calculated via the inverse transformation (F^{-1}) for the F values

corresponding to the center of each layer. With reasonably high number of layers (tens or hundreds), the created quantity x will have the proper probability distribution. This approach is called stratified sampling. If the output variable y depends on several input quantities, x_1, x_2, \dots, x_m , it is necessary that each quantity is assigned values of all layers and that the quantities and layers of individual variables are randomly combined. This is done by random assigning the order numbers of layers to the individual input quantities.

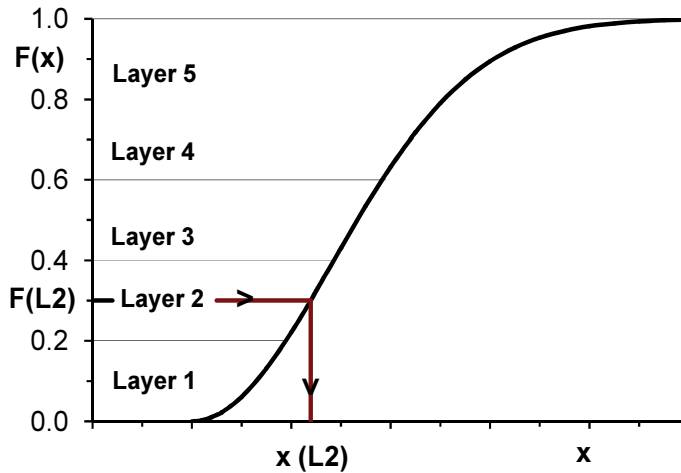


Figure 1. Principle of the LHS method.

The procedure is as follows. The definition interval of the distribution function F of each of m variables is divided into N layers. N , the same for all variables, also corresponds to the number of trials (= simulation experiments). In each trial, the order numbers of layers are assigned randomly to the individual variables (X_1, X_2, \dots, X_m). In this way, various layers of the individual variables are always randomly combined. In practice, this is achieved by means of random numbers and their rank-ordering. Then, each input variable is assigned the value corresponding to the center of the pertinent layer of its distribution function.

The application is illustrated on a case with four random quantities (X_1, X_2, X_3 , and X_4) and the definition interval of F divided into five layers (Fig. 1). Only five layers are used here for simplicity; usually, several tens of layers are used. In our case, Y will be calculated for five combinations of the four input quantities. Thus, $5 \times 4 = 20$ random numbers with uniform distribution in interval $(0; 1)$ are generated (see the table in the left part of Fig. 2). Then, the layer numbers for variable X_1 (for example) for individual trials are assigned with respect to the order of random values (for X_1) ranked by size from the maximum to minimum. Here, layer no. 3 (with the highest number 0.83) for the first trial, layer no. 1 for the second, no. 5 for the third, no. 2 for the fourth, and no. 4 for the fifth, corresponding to the numbers 0.56 - 0.25 - 0.83 - 0.17 and 0.30 in the column for X_1 . Similar operations are done for each variable. Thus, in the first trial, variables X_1, X_2, X_3 , and X_4 are assigned the values corresponding to the second,

fourth, second, and fifth layers of their distribution functions, respectively. Inverse probabilistic transformation F^{-1} is then used for the determination X_1 from $F_{1,1}$, etc.; see the table on the right. Now, the investigated quantity $Y = Y(X_1, X_2, X_3, X_4)$ is calculated five times. The obtained values $Y_1, Y_2, Y_3, Y_4,$ and Y_5 can be used for the determination of statistical characteristics (mean, standard deviation,...).

Random numbers (RN)					Layer numbers for individual trials				
Variable ----- Trial	X_1	X_2	X_3	X_4	Variable ----- Trial	X_1	X_2	X_3	X_4
	RN	RN	RN	RN		layer	layer	layer	layer
1	0.56	0.12	0.72	0.13	1	2	4	2	5
2	0.25	0.40	0.84	0.60	2	4	3	1	3
3	0.83	0.05	0.21	0.55	3	1	5	4	4
4	0.17	0.62	0.03	0.99	4	5	2	5	1
5	0.30	0.70	0.61	0.73	5	3	1	3	2

Figure 2. LHS method: assignment of layers to individual variables and trials.

Usually, several tens or hundreds of trials are made, which enable the construction of distribution function $F(Y)$ and determination of the mean value, standard deviation, various quantiles, and other characteristics.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty, University of Pardubice, Czech Republic

References

- [1] Florian A. An efficient sampling scheme: Updated Latin Hypercube Sampling. Probabilistic Engineering Mechanics, 7 (1992), issue 2, 123 – 130.
- [2] Olsson A, Sandberg G, Dahlblom G. On Latin hypercube sampling for structural reliability analysis. Probabilistic Engineering Mechanics. 2002; 25: issue 1, 47 – 68.

Economy of Reliability

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62371>

Abstract

The simultaneous influence of several random quantities can be studied by the Latin hypercube sampling method (LHS). The values of distribution functions of each quantity are distributed uniformly in the interval (0; 1) and these values of all variables are randomly combined. This method yields statistical characteristics with less simulation experiments than the Monte Carlo method. In this chapter, the creation of the randomized input values is explained.

Keywords: Probability, Monte Carlo method, Latin Hypercube Sampling, probabilistic transformation, randomization

Permanent effort exists to make the components and constructions more reliable and with longer life. Higher reliability can be achieved by better design and more ample dimensioning of load-carrying parts and by better maintenance. However, all these cost money, and one can ask: How is reliability related to the costs? Does an optimum reliability exist from the costs' point of view? How can it be found? This chapter is devoted to the following issues: (1) optimum time for the renewal of objects with gradual deterioration, (2) optimum dimensions of the cross-section of load-carrying components that can fail suddenly (e.g. due to overloading) or due to fatigue or similar processes, (3) optimum probability of failure, and (4) cost-based optimum strategy of the maintenance and renewal of a group of objects.

1. Optimum time for the renewal of deteriorating objects

Examples of deteriorating objects are machines, cars, bridges, cutting tools, pumps, or aircrafts. Basically, there are three kinds of costs related to these objects: (1) purchase costs C_0 , (2) costs for operation C_{op} , and (3) costs caused by a failure $C_{i,p}$. Their sum creates the total costs C_{tot}

$$C_{\text{tot}}(t) = C_0 + C_{\text{op}}(t) + C_{f,p}(t); \quad (1)$$

t denotes the time of operation. The general time course of the costs is depicted in Figure 1. The purchase costs C_0 is the money spent for buying or manufacturing or building the object. C_0 is spent at the instant of its purchase and remains constant until the failure or replacement by another object. The operation costs $C_{\text{op}}(t)$ have several components, such as costs of energies, fuels, processed material, common maintenance, and small repairs. Basically, these costs should grow approximately linearly with the time. For example, fuel is consumed continuously during any car ride. However, as the technical condition of the object gets gradually worse, the operation costs after some time start growing faster (the fuel consumption of a worn engine becomes higher, some small parts must be exchanged more often, etc.). The failure-related costs $C_{f,p}(t)$ mean here the probable costs, which can be determined as

$$C_{f,p}(t) = C_{f,\text{tot}} \times F(t); \quad (2)$$

$C_{f,\text{tot}}$ is the total costs caused by the failure, which consist of the price for the replacement of failed components or parts and additional costs, such as damages of other objects caused by this failure, costs due to the fall-out of production, and costs related to injuries or casualties. $F(t)$ means the probability of failure caused by gradual deterioration (see Figure 1 and Chapter 16). When a new object is put into service, this probability equals zero and remains very low for a relatively long part of its life. Later, however, it gradually grows faster and faster due to the worsening condition of the object.

Note: The concept of probable failure-related costs $C_{f,p}$ and economic optimization of the allowable probability of failure makes sense in cases with many objects involved, where failures can occur more often, but the total managed property is so high that the administrator can include the average losses easily into his expenses. For example, a collapse of a bridge is a very costly event. However, the administrator of a bridge network comprising hundreds of bridges knows that several million Euros will be needed every year for repairs and thus prepares funds for them. On the contrary, a small manufacturer, who owns only one workshop, can become bankrupt if the workshop collapses due to overloading by snow, just because he has not enough money to build a new one. The total actual loss C_0 (e.g. 10^5 €) is incomparably higher than the probable loss $C_{f,\text{tot}} \times F$ included into the economic optimization, which would make only 1 € for failure probability $F = 1:10^5$. All insurance systems are based on the idea of distributing the rarely occurring high losses over many subjects. (This also illustrates the fact that the concept of probable costs $C_{f,p}$ is sometimes inadequate, and the allowable failure probability must be based on the other criteria.)

The optimum lifetime from the economic point of view is that the object attains during its life the minimum costs per unit of the demanded production or service [e.g. per 1 km for a vehicle, per one machined part in the case of a cutting tool, per time unit of service (e.g. year) of a

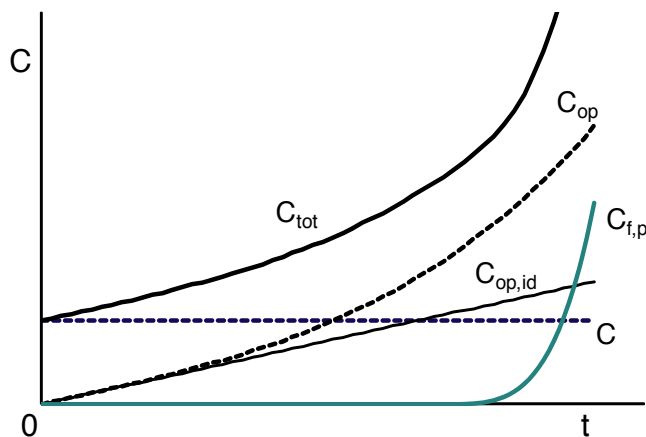


Figure 1. Development of costs with time (a schematic). C_0 , purchase costs; C_{op} , operation costs ($C_{op,id}$, ideal case); $C_{f,p}$ failure-related probable costs; C_{tot} , total costs; t , time.

bridge, etc.]. The lifetime means the time until the demolition, complete overhaul, or replacement by a new one. Equation (1) can be rewritten as

$$C_{tot,1}(t) = C_{tot}(t)/t = C_0/t + C_{op}(t)/t + C_{f,p}(t)/t. \quad (3)$$

The time course of individual components and of the total unit costs is depicted schematically in Figure 2. One can see that the total unit costs $C_{tot,1}$ attain a minimum at certain time t_{opt} . This is the optimum time for replacement. For all other times, the economy of operation is worse. This can be illustrated on the example of a car. Everybody understands that it will not be economical to buy a new car every year. Similarly, it will not be reasonable to use one car for 50 years or more because of the increasing fuel consumption and the necessity to buy spare parts more often (provided they would be available for so long time).

The optimum condition for renewal must be understood in a wider sense. From a mathematical point of view, the optimum is exactly the time, for which $dC_{tot,1}/dt = 0$, and nowhere else. However, only seldom can a repair be done at just this instant. On the other hand, the curve $C_{tot,1}(t)$ changes very slowly near the optimum (Fig. 2), and often it does not matter if the reconstruction of a bridge, for example, is done 1 month earlier or later. This makes the planning of maintenance and repairs easier.

2. Optimum dimensioning of load-carrying components – sudden failure

Sudden failures occur due to a “weak” component or overloading from external cause (e.g. snow, wind, flood, or error in operation). The failure can be a fracture, permanent deformation,

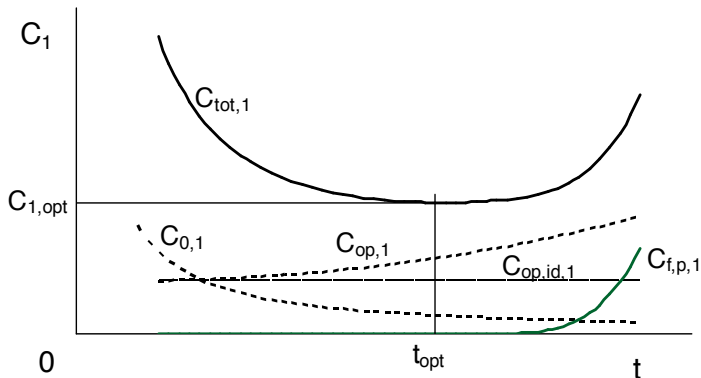


Figure 2. Development of unit costs with time. (The symbols are explained in Figure 1.)

or collapse of a structure. Failure by overloading can occur at any time. If it happens before the end of the assumed service life, the system must be put back into its original state either by repairing the failed component or by replacing it by a new one. More ample dimensioning of the load-carrying parts means higher safety and lower probability of failure due to overloading, but also higher price of the component. One can thus seek such dimensions, which will guarantee the best combination of high reliability and low costs.

The influence of the size of the cross-section on the probability of failure and total costs was studied by Menčík [1]. The main results for a bar loaded by tensile force are shown in Figures 3 to 5. The nondimensional scales express the ratio of individual values (C_{tot} and A) to the reference values C^* or A^* as a function of the cross-section area. Also, the failure probability P_f is shown. The individual curves correspond to various magnitudes of the failure-related costs (Fig. 3a) and to the various rates of the cost increase with the cross-section area (Fig. 3b).

Typical is the fast increase of costs with the decreasing cross-section at the left part of the diagram. For small cross-sections, the probability of failure increases and the failure-related costs are no more negligible. The cost growth after the minimum has been reached is usually much slower. The analysis for various cases has also revealed that the economically optimal failure probability is sometimes relatively high (e.g. 10^{-3} or even more). This is acceptable if the costs caused by a failure are low. For higher failure losses, the optimum failure probability will be lower. It is thus important that all possible losses be included into $C_{f,a}$. The relatively high number of failures of some product can even cause the loss of customer reliance, which can lead to the bankruptcy of the manufacturer. Another example is the disaster of the space shuttle Challenger in 1986 due to the failure of a sealing ring. The price of such sealing ring is minute. If, however, the price of a destroyed space shuttle should be added to it, the total costs and $P_{f,opt}$ would be quite different. The differences in optimum dimensions can exist if big differences in the probable failure-related losses exist (e.g. due to the different financial compensation for the accident victims in various countries).

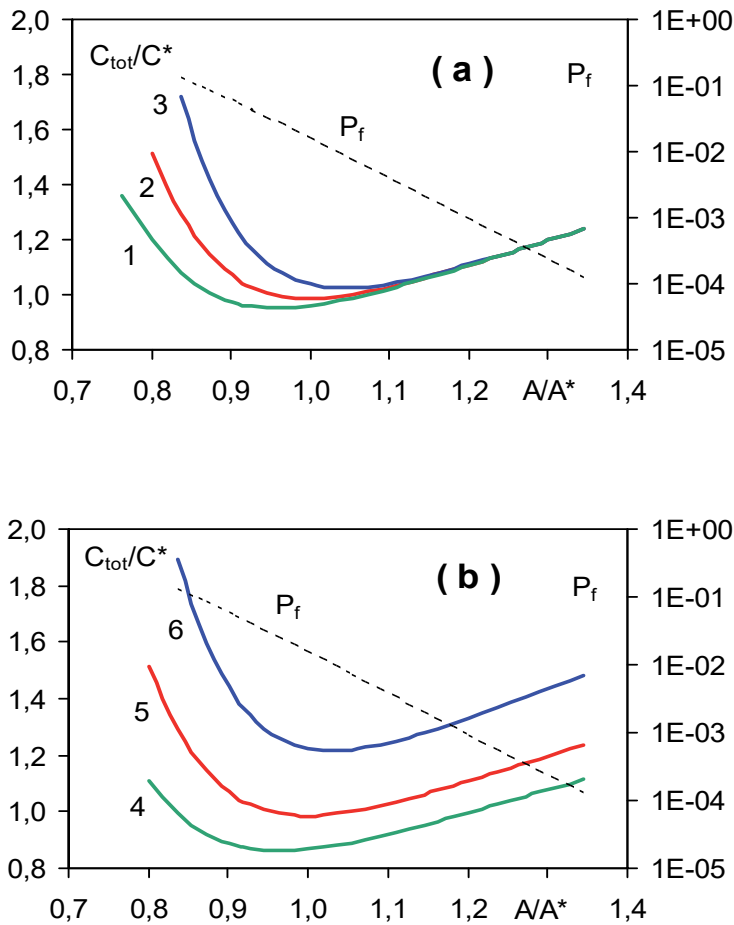


Figure 3. Total costs C_{tot} and failure probability P_f as functions of the cross-section area A (a schematic, after [1]). Curves 1 to 3 in (a) correspond to increasing failure costs and curves 4 to 6 in (b) correspond to increasing costs per unit of the cross-section area. The scales for costs and the cross-section area are standardized, and the scale for P_f is logarithmic.

Based on the same data as above, Figure 4 shows the total costs C_{tot} as a function of the failure probability P_f . Obvious is the fast increase of the costs for higher values of P_f due to the growing probable costs $P_f(C_0 + C_{f,a})$. The probable failure-related costs are usually negligible for $P_f < 10^{-6}$, but significant for $P_f > 10^{-1}$. It is also obvious that the probability of failure, corresponding to the minimum total costs, is relatively high. Sometimes, other criteria are thus decisive rather than the costs.

These examples assumed that the cross-section area can change continuously. The reality is often more complex: the cross-section of standardized components changes stepwise, and so also the price. The computations should thus be done for all possible nominal cross-sections and arrangements, and the variant with minimum total costs can be chosen.

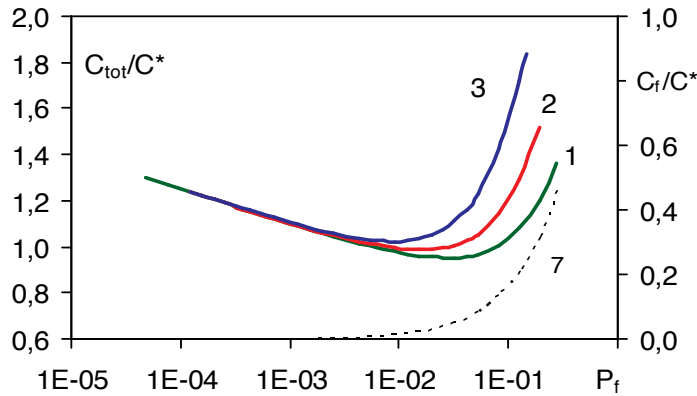


Figure 4. Total costs C_{tot} as functions of failure probability P_f [1]. Curves 1 to 3 correspond to increasing failure costs and curve 7 shows the probable failure costs P_f . The curves correspond to the same input values as in Figure 3a.

3. Optimum dimensioning of components with stress-enhanced deterioration

Examples of gradual deterioration are fatigue or wear of metallic parts as well as corrosion, creep at increased temperature, or carbonation of concrete. In these cases, the failure occurs after some time of operation depending on the load. Thus, the costs must be related to the time to failure or replacement of the component T_f , and the cost-effectiveness is evaluated according to the unit costs (i.e. the costs per unit time of operation),

$$C_{tot,1}(t) = C_{tot} / T_f. \tag{4}$$

The time to failure depends on the load effect S , which is the stress amplitude in cyclical loading, the force acting on a ball bearing, or the intensity of creep or corrosive environment. In the simplest case, the lifetime T_f decreases with increasing S according to the formula:

$$T_f = BS^{-m}; \tag{5}$$

B and m are material constants. Equation (5) is known as $S-N$ curve for fatigue, with S denoting the stress amplitude. However, as shown in Chapter 6, Equation (5) can also be used for the prediction of the time to failure of a body with a slow crack growth or the endurance of components exposed to creep at high temperatures, the lifetime of bearings, as well as several other cases.

The load effect S depends on the load L and the size of the cross-section under load as

$$S = L / Z, \tag{6}$$

where L is the amplitude of characteristic load (tensile force or bending moment) and Z is the characteristic parameter of the cross-section (the area for tension or section modulus for bending). The expected lifetime T_f may thus be expressed as

$$T_f = B(Z / L)^m, \tag{7}$$

and can be ensured by proper dimensioning (Z).

The relationships between cross-section size, costs, and time to failure may be used for studying the influence of the cross-section size on the cost-effectiveness. Figures 5 and 6 show the development of unit costs and time to failure as the function of the diameter of a shaft loaded in tension (Fig. 5) and bending (Fig. 6a,b). Also, these diagrams can be plotted in non-dimensional coordinates.

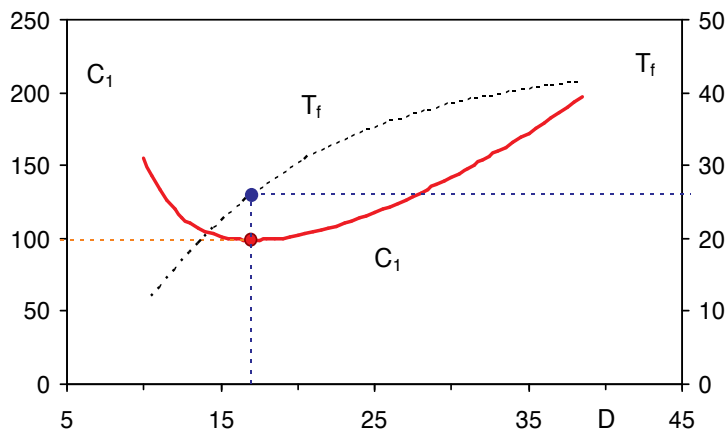


Figure 5. Unit costs C_1 and time to failure T_f as functions of shaft diameter D (a schematic, [1]). Component loaded in tension, fatigue exponent $m = 3.0$.

Several important conclusions can be drawn from this analysis. An optimum with minimum unit costs exists only in some cases. Sometimes, it lies outside the suitable interval of service times or acceptable dimensions of the cross-section. An important role is played by the kind of loading and fatigue exponent m . The situation for tensile loading is better: for the common values of fatigue exponent in metals ($2 < m < 6$), an optimum often exists (Fig. 5). For bending, the increase of the characteristic cross-section dimension (e.g. diameter D) causes the increase of the cross-section area, $A \sim D^2$, and also a faster increase of the section modulus ($Z \sim D^3$) and thus much faster growth of useful life ($T_f \sim Z^m$). As a consequence, any enlargement of the

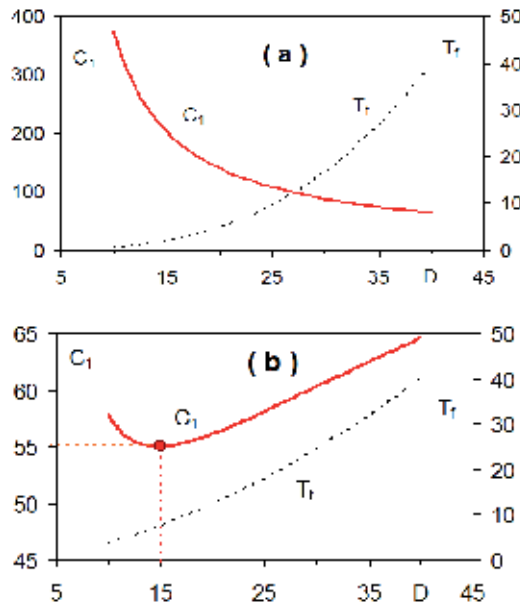


Figure 6. Unit costs C_1 and time to failure T_r as functions of shaft diameter D [1]. Bending load, fatigue exponent $m = 3.0$ (a) and $m = 1.7$ (b).

cross-section usually leads to the reduction of unit costs C_1 (Fig. 6a); the exception is low values of fatigue exponent, $m < 1.7$, where also a cost minimum exists (Fig. 6b).

Because of the many factors involved, it is impossible to formulate simple universal rules for finding the optimum size. A more practical way is to model the situation for admissible ranges of the input quantities. A graphical representation is very useful. Sometimes, it becomes obvious that the unit costs vary only monotonously or insignificantly in the possible range of input variables, which can make the choice of optimum dimensions easier.

4. Cost-based maintenance optimization of long-life objects

Examples of such objects are bridges, cooling towers, locomotives, or heavy machines in power plants. Two cases can be distinguished: maintenance optimization of a single object, and of a group of similar objects, such as bridges in a railway or road network.

Single object

The optimization tries to minimize the total costs spent from putting the object into operation until its replacement or reconstruction. The total costs C in the period T are [1 – 4]:

$$C = C_i + C_m + C_r + C_{f,p} + C_u + C_a - V_s; \tag{8}$$

C_i = inspection costs, C_m = maintenance costs, C_r = repair costs, C_u = user costs, C_a = additional costs, V_s = salvage value of the object at the end of the considered period, and $C_{t,p}$ = probable failure costs, as defined by Equation (2). These include the costs due to a collapse of the object (e.g. a bridge) or its closing if the collapse is imminent. Also, the expenses of users C_u should be included (delays and the necessity to use longer alternative routes).

Various strategies can be used for organizing the maintenance and repairs (e.g. no repair until the replacement, or regular maintenance plus small repairs), which would enable a longer time of operation until replacement. The optimization consists of comparing the costs C for several variants and choosing that with the lowest possible costs. For each variant, the total costs C , Equation (8), are calculated for the time interval considered (e.g. 5 or 20 years), and the costs for individual variants are rank-ordered and compared (Fig. 7). As some of the input data are only estimated, it is recommended to make several estimates for each maintenance strategy: with optimistic, probable, and pessimistic input values.

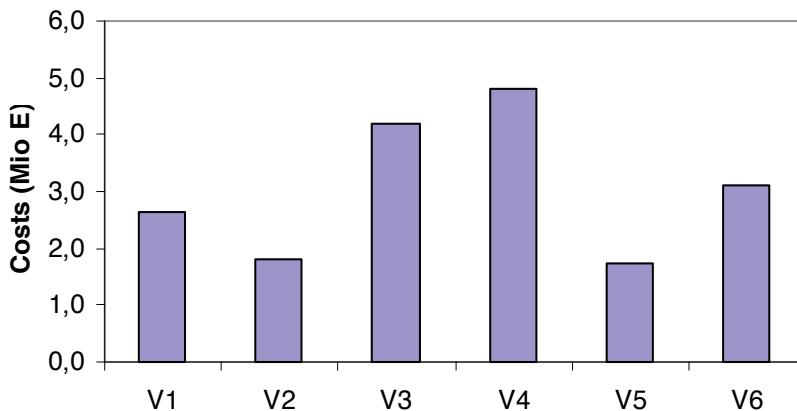


Figure 7. Comparison of total costs for six variants (an example).

Maintenance optimization for a group of objects

In an ideal case, the optimum variant for maintenance of each object in the network is found as described above. In real life, however, various constraints must often be respected. Very important is the amount of money available for maintenance and repairs in the individual years. It could happen that more bridges should be reconstructed simultaneously than the budget would permit, whereas, in other years, the working capacities for repairs would not be fully used. Thus, it is necessary to calculate the cost components for every bridge and maintenance variant for every year in the considered time interval, in which the renewal should be optimized (e.g. 20 years). These costs can be written into a table with the columns corresponding to years and the rows to individual bridges and maintenance variants (Fig. 8) and compared with the money available. A comparison of all variants reveals the best strategy.

In addition to the demand of uniform flow of money for maintenance and repairs, several other factors must be considered when prioritizing the objects (e.g. bridges) for repairs:

- The condition of the stock as a whole (the worst objects are of special interest);
- The rate of deterioration, as some objects can be in a better condition but deteriorate faster;

The importance of the object in the whole network or in some region.

	2015	2020	2025	2030	2035	SUM
Br. 1, V2						
Br. 2, V1						
Br. 3, V2						
Br. 4, V1						
Br. 5, V3						
Br. 6, V2						
Br. 7, V1						
SUM						

Figure 8. Example of cost matrix for seven bridges during the period 2015 to 2035. Br – bridge; Vj – maintenance variant.

For these reasons, maintenance optimization is sometimes divided into three steps [1, 3, 4]:

Step 1. Condition rating prioritization. All objects are ranked according to their condition as revealed by inspections. Only those with condition worse than a certain value will be considered for the maintenance in the near future. This preselection significantly reduces the number of candidates for further steps.

Step 2. Object importance prioritization. The role of individual objects in the network is considered. One bridge can be in a worse condition than another bridge. However, if it lies on an unimportant road, whereas the other is on a main road, the latter one will be repaired preferably.

Step 3. Optimization of money allocation. The possible maintenance strategies are compared with respect to the available money and the work capacities. Then, the strategy is chosen, which is economically most favorable for the stock in the longer period.

5. Time value of money

In long-term planning, one must be aware of the difference between spending money today and in the future. Due to interests, the value of (suitably deposited) money gradually increases,

so that the today's value V will, in n years, correspond to $V(1+r)^n$, where r is the interest rate. Thus, when the total costs during a long period are calculated as a sum of expenses arising in various times, the values of individual components should be converted to the same time base, usually to the time T_0 when the study is made. The common formula for the conversion of the j -th component C_{j,T_i} paid at time T_i is

$$C_{j,T,0} = C_{j,T_i} \frac{1}{(1+r)^{T_i-T_0}}. \quad (9)$$

However, due to inflation, the prices of material, components, and work gradually grow, and the gain from postponing an investment is smaller. If the inflation rate is close to the interest rate, the profit can be negligible, and the standardization (9) is not necessary for the comparison of individual variants.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
 University of Pardubice, Czech Republic

References

- [1] Menčík J. Optimum-Reliability Dimensioning: Cost-Based Approach. In: 3rd International Colloquium ASRANet 2006. Glasgow, 10 – 12th July 2006. Universities of Glasgow and Strathclyde, ASRANet Book of abstracts, p. 29, full text: CD-ROM. ISBN: 0-9553550-0-4.
- [2] Bień J, Elfgren L, Olofsson J, editors. Sustainable bridges. Wrocław: Dolnośląskie Wydawnictwo Edukacyjne; 2007. 490 p.
- [3] Menčík J, Rudolf P. Optimisation of maintenance strategy in a bridge network. In: Proc. Int. Conf.: Reliability and Diagnostics of Transport Structures and Means 2005 (J. Menčík, editor). Pardubice: University of Pardubice; 2005. p. 224 – 231.
- [4] Menčík J. Cost-based design and renewal strategy for long-life structures. In: Life-Cycle Civil Engineering (F. Biondini and D.M. Frangopol, editors), Varenna, 10-14 June 2008, Leiden: CRC Press (Taylor and Francis); 2008, p. 385 – 390. ISBN 978-0-415-46857-2.

Reliability Assessment with a Small Amount of Data

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62372>

Abstract

This chapter describes various methods for reduction of uncertainties in the determination of characteristic values of random quantities (quantiles of normal and Weibull distribution, tolerance limits, linearly correlated data, interference method, Monte Carlo method, bootstrap method).

Keywords: Random quantity, uncertainty, normal distribution, Weibull distribution, tolerance limits, correlation, interference method, Monte Carlo method, bootstrap method

Many quantities necessary for reliability assessment are obtained by observation or measurement (load and material characteristics). Often, the number of tests n is low. As a consequence, the distribution parameters are only estimated. Their true values can be different and thus also the other characteristics, such as quantiles. This could be dangerous especially if n is very low (tens or less). Uncertainty arising from a small amount of data should be taken into consideration in any reliability assessment. This chapter presents four methods that can increase the safety of these assessments. The first method is related to the determination of the guaranteed lowest or highest values (i.e. low probability quantiles of normal or Weibull distribution). The second method tries to mitigate the uncertainties related to the use of the S – R interference method. The third method is devoted to the uncertainties of the Monte Carlo simulations if the input distributions were obtained from a small amount of measurements. The last method explains briefly the principle of the so-called bootstrap technique.

1. Guaranteed values of quantiles

Guaranteed or safe value of a random quantity x is such value that will be exceeded (e.g. load) or not-attained (strength) only with a very low probability α . This value corresponds

to α -quantile in the latter case and to α -critical value in the former case; note that the α -critical value corresponds to $(1 - \alpha)$ -quantile.

A quantile can be calculated as the inverse value of the distribution function. However, because usually only the estimates of the parameters of a population are known, only an estimate of the quantile is obtained in this way. Its true value can be different. As a consequence, in some cases, the actual reliability will be lower than assumed, which is dangerous. The difference can be high especially if the statistical parameters of the population were obtained from a very small number of samples. For this reason, confidence interval should always be determined for the quantiles. Two very important cases are normal and Weibull distribution.

Quantiles of normal distribution

The α -quantile of a quantity x is calculated as

$$x_\alpha = \mu + u_\alpha \sigma, \quad (1)$$

μ and σ are the mean and standard deviation and u_α is the α -quantile of the standard normal distribution. The uncertainty in the determination of the distribution parameters μ and σ can be reduced by means of confidence limits for the quantile. For quantities with normal distribution, usually tolerance limits $x_{\alpha, \text{tol}}$ are used instead. The lower (upper) tolerance limit demarcates the fraction α of the population, which will be lower (higher) than $x_{\alpha, \text{tol}}$ only with probability α , the risk of a wrong estimate being γ . The tolerance limits are determined using the formula [1, 2]

$$x_{\alpha, \text{tol}} = \bar{x} \pm k_\alpha s, \quad (2)$$

and s are the sample average and standard deviation, and k_α is the one-sided tolerance factor, which depends on α , the number of measurements n , and the risk γ that the prediction will be wrong. The plus sign pertains to the upper limit, whereas the minus sign pertains to the lower limit.

The difference between the values obtained using Eq. (1) or (2) is large especially for low n , α and γ . For example, the 0.1%-quantile according to (1) is $\mu - 3.09\sigma$, while the 0.1% lower tolerance limit (for $\gamma = 10\%$) is $-3.44s$ for $n = 100$, $-3.79s$ for $n = 30$, and $-4.63s$ for $n = 10$. If, for example, strength tests were made only on 10 specimens, and the standard deviation was obtained as $s = 0.1$, then the strength guaranteed with 99.9% is by $(4.63 - 3.09)s = 0.15$, i.e. by 15% of the average strength lower than the less conservative value according to Eq. (1)

The tolerance factors k can be found in statistical tables, e.g. [2]. Their exact determination is difficult. An approximate formula was proposed by Wallis [1, 3]:

$$k_\alpha = \left(u_\alpha + u_\gamma \sqrt{\frac{1}{n} + \frac{u_\alpha^2}{2(n-1)} - \frac{u_\gamma^2}{2n(n-1)}} \right) / \left(1 - \frac{u_\gamma^2}{2(n-1)} \right), \quad (3)$$

where u_α and u_γ denote the α - and γ -critical values of the standard normal distribution. The approximation (3) coincides with the exact solution for $n \geq 500$, and the difference slightly increases with decreasing sample size. For the confidence level $1 - \gamma = 0.9$, the difference exceeds 1% for $n < 20$ and 2.5% for $n < 10$ regardless the value of α .

This method can also be used for quantities with log-normal distribution. They can be transformed (by logarithms) into variables with normal distribution. Then, tolerance limits for quantiles of this new quantity can be calculated using the above approach. Finally, the tolerance limits for the original quantity are found by inverse transformation.

Quantiles of three-parameter Weibull distribution

Weibull distribution has a very flexible character and is used very often to characterize the strength or time to failure. The α -quantile x_α of a three-parameter Weibull distribution is determined as

$$x_\alpha = x_0 + a \left[-\ln(1 - \alpha) \right]^{1/b}, \tag{4}$$

where a and b are the scale and shape parameters of the distribution, respectively, and x_0 is the parameter of its position (threshold value). Again, only estimates of the true parameters can be obtained from a sample of limited size and thus also only an estimate of the quantile x_α . Especially for small sample sizes (n equal several tens or less), the sample parameters can differ significantly from those of the population [4]. Sometimes, the threshold value x_0 is assumed zero for safety reasons. However, this can yield unreasonably low values of low probability quantiles. The three-parameter distribution can be better, but confidence limits should always be given with quantiles, especially for small samples and low probability quantiles. These limits (L - lower; U - upper) can be computed as

$$x_{\alpha,L,U} = x_\alpha \pm \Delta_{\alpha,\gamma,n}, \tag{5}$$

where $\Delta_{\alpha,\gamma,n}$ is a certain function of the distribution shape, scatter of the measured values, probability α , confidence level $1 - \gamma$, and number of measurements n . Mann et al. [5] proposed a method for the determination of the confidence limits. Unfortunately, the procedure is complicated and tabulated values must be used. Menčík [4] has proposed a simple approximate expression for Δ based on the variation of the position and slope of the distribution function:

$$\Delta = \sqrt{t_{\gamma,n-1}^2 \frac{s^2}{n} + \left[1 - \left(\frac{2n}{\chi_{\gamma,2n}^2} \right)^{1/b} \right]^2 (x_\alpha - \bar{x})^2}, \tag{6}$$

where \bar{x} and s_x are the sample average and standard deviation, respectively; n is the number of measurements from which the distribution parameters were estimated; $t_{\gamma,n-1}$ is the one-sided

γ -critical value of the Student's distribution for $n - 1$ degrees of freedom; b is the shape parameter; and $\chi^2_{\gamma,2n}$ is the γ -quantile of the χ^2 -distribution for $2n$ degrees of freedom. The probability that the true value of x_α will be lower than the lower confidence limit is γ . The use of confidence intervals for quantiles may strongly be recommended if $n < 100$.

Linearly correlated data

If fatigue or wear processes occur, the relationship between characteristic load S and cycles to failure N can usually be described by a power-law function:

$$N = AS^{-m}, \quad (7)$$

where A and m are constants. This equation corresponds to a straight line

$$y = a + bx, \quad (8)$$

in coordinates $x = \log S$ and $y = \log N$. The constants $a = \log A$ and $b = -m$ are obtained by testing several specimens under various stress levels S , measuring the cycles to failure N , and fitting the transformed $y(x)$ data by linear regression function. Typical of fatigue is the large scatter of times to failure. This must be taken into account when determining the guaranteed time to failure for a given service stress or the allowable stress for the required lifetime.

The situation is easier if 10 or more specimens were used for each stress level: the pertinent N -values in each level can be ordered so that they (approximately) correspond to quantiles. For example, if $N=10$, then the lowest value corresponds to the 10% quantile, the second lowest corresponds to 20% quantile, etc. Then, the S - N curves may be constructed for various probabilities of survival by fitting only the pertinent quantiles. This is the best solution. Unfortunately, fatigue tests are time and cost demanding, so that often only several specimens are tested and Equation (8) is fitted to all data, thus representing the mean line. In this case, 50% probability exists that the true times to failure will be lower than those predicted via this line. Therefore, confidence intervals are also needed. The confidence interval for the points on the regression line (Fig. 1) is

$$y = a + bx \pm t_{\alpha, \nu} s_{res} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{(n-1)s_x^2}}, \quad (9)$$

where \bar{x} and s_x are the average value and standard deviation of x , calculated from all n values x_j used for the determination of regression constants a and b ; $t_{\alpha, \nu}$ is one-sided α -critical value of t -distribution for $\nu = n - 2$ degrees of freedom, and

$$s_{res} = \sqrt{\frac{\sum (y_j - a - bx_j)^2}{n - 2}} \quad (10)$$

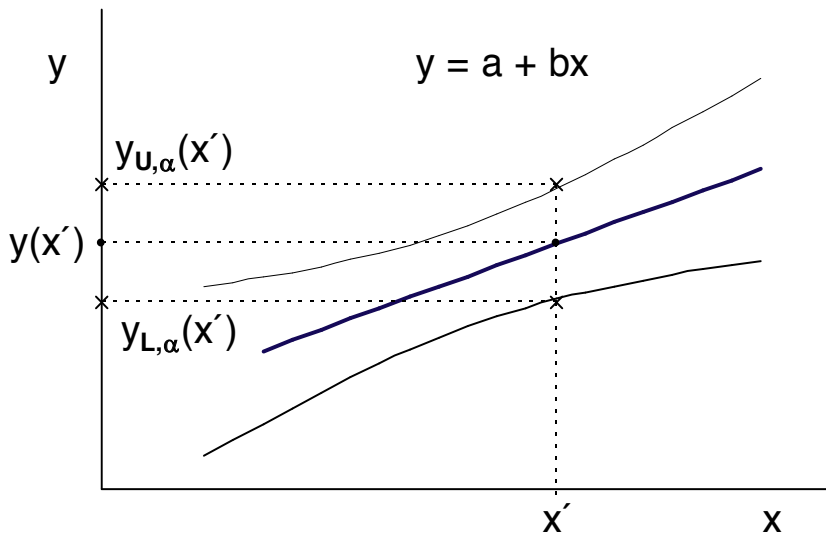


Figure 1. Confidence interval around regression curve (a schematic).

is the residual standard deviation, characterizing the scatter around the regression line; the summation is done over all measured values of y_j .

The modification of Equation (9), as proposed in [4], gives the approximate expression for tolerance limits for single values of y ,

$$y = a + bx \pm t_{\alpha, \nu} s_{res} \sqrt{\left(\frac{k_{\alpha}}{t_{\alpha, \nu}}\right)^2 + \frac{(x - \bar{x})^2}{(n - 1) s_x^2}}; \tag{11}$$

the minus (plus) sign pertains to the lower (upper) limit. In Equation (11), k_{α} is a one-sided tolerance factor, which can be calculated using Equation (3). The probability that $y(x)$ will be lower or higher than the tolerance limit (11) equals α , whereas the probability that this estimate is wrong is γ .

The intervals (11) for all $y(x)$ form a tolerance band around the regression line (9). The tolerance limits for the actual number of cycles (or time) to failure can be obtained using the inverse transformations $S = 10^x$, $N = 10^y$, $A = 10^a$, and $m = -b$.

2. Interference method for normally distributed stress and strength

The interference method, suitable if random “load” acts on an “object” whose “resistance” also exhibits random scatter, was explained in Chapter 14. Failure occurs if the load effect S is higher

than the resistance R . If the distributions of R and S interfere, the distribution of reliability margin $G = R - S$ can be found, and the probability of failure is determined as the value of the distribution function for $G = 0$. The solution is easy if R and S are normally distributed and their parameters are known, because here the distribution of G is also normal, with parameters

$$\mu_G = \mu_R - \mu_S, \quad \sigma_G = \sqrt{\sigma_R^2 + \sigma_S^2}; \quad (12)$$

μ and σ are the mean value and standard deviation; the subscript denotes the pertinent quantity. G can be transformed to standard normal variable u using the relationship

$$G = \mu_G + u\sigma_G. \quad (13)$$

Using the failure condition, $G = 0$, the probability of failure P_f can be found as the value of the distribution function for $u = -\mu_G/\sigma_G$.

However, instead of the parameters μ and σ of the distributions R and S , only their estimates \bar{x} and s are usually inserted into Eq. (13), which were obtained from samples of limited size n_R and n_S . As a consequence, one obtains only the estimates \bar{x}_G and s_G of the reliability margin. In such case it is necessary to use one-sided tolerance factor k instead of quantile u in Eq. (13). Otherwise the actual probability of failure can differ from the forecasted one by more than one order [6,7]. The tolerance factor k should be determined for the confidence level γ of the forecast and for the equivalent size n_G of the sample G , which must be calculated from the sample sizes and standard deviations of the samples R and S via the relationship

$$G = \mu_G + u\sigma_G. \quad (14)$$

The probability of failure is found as that corresponding to the lower tolerance limit

$$k = \bar{x}_G/s_G. \quad (15)$$

When dimensioning for given probability of failure, k is determined first via Eq. (3). For this value, \bar{x}_G is calculated from Eq. (15). Finally, the nominal size \bar{x} of cross-section is determined from \bar{x}_G and the known mean value of the load using Eq. (12). Diagrams for this purpose are given in [6, 7].

3. Monte Carlo method

If the investigated quantity x is a function of random input variables, its distribution can be obtained easily by the simulation Monte Carlo method (Chapter 15). In the simulation trials,

random values are assigned to input quantities so that their distribution corresponds to the probability distribution of the pertinent variable. However, the distribution parameters used in the simulations were obtained from samples of limited size. This means that the actual distributions can differ more or less from those used in the simulation. The corresponding uncertainties and errors persist in the results regardless of the number of Monte Carlo trials but can be reduced in two ways.

The first approach [8] uses random variation of distribution parameters. If the distributions of parameters of a random quantity x are known, this quantity can be generated so that, in each trial, random values are also assigned to its parameters, so that they vary randomly in their probable range. For example, the random quantity x of normal distribution with the known mean value μ and standard deviation σ can be generated using formula (1) with u_α replaced by the random value u of the standard normal distribution. If, instead of true parameters μ and σ , only the sample estimates m and s are known, the probable values of μ and σ can be generated in individual trials using modified expressions for their confidence limits. The corresponding formula for random values of x is

$$x = m + t_{n-1} \frac{s}{\sqrt{n}} + us \sqrt{\frac{n-1}{\chi^2_{n-1}}}; \quad (16)$$

u , t , and χ^2 are random values of normal, t , and χ^2 distribution, respectively (t , χ^2 for $n - 1$ degrees of freedom). One value of x thus needs three random numbers to be generated. The quantiles of t and χ^2 distributions can be expressed approximately by means of quantiles of standard normal distribution; a review of various approximations can be found in [1].

The second method [4] adds a random component to each generated number. The basic random values x_0 of a quantity x are created (via the inverse probabilistic transformation F^{-1}) from random values F uniformly distributed in interval (0; 1). Then, a random component Δ is added to each value x_0 created so that the result

$$x(F) = x_0(F) + \Delta(F) \quad (17)$$

has the same probability distribution around x_0 like the quantiles of the genuine variable x . The obtained x values create a blurred confidence band around their distribution function.

4. Bootstrap method

This method, which also uses the Monte Carlo simulations, was originally used for finding the statistical characteristics of random quantities from a relatively small number of data n [9]. It creates its own population, consisting only of the experimental values. From this population, n values are chosen randomly, and the characteristic X of interest is calculated (e.g. mean or a

quantile). This procedure is repeated many times. Then, the α -confidence interval for X is determined by one of the following ways. In the first approach, the average value m_x and standard deviation s_x of the pertinent quantity are calculated, and the lower (L) and upper (U) confidence limits are found using the expression

$$X_{L,U} = \bar{x} \pm u_\alpha s_x, \quad (18)$$

where u_α is the α -quantile of standard normal distribution; the probability of X being outside the limits (18) is 2α . This approach assumes that X has normal distribution. This needs not always be true, and various improvements were later proposed. (For more details, see [10].)

Another approach generates a large number of simulated data sets (at least 100). Then, the values of the characteristic of interest, calculated for each data set, are ranked into ascending order. The confidence bounds, corresponding, for example, to the 90% confidence interval, are obtained using the 5th and 95th values of the ordered values. However, when determining the confidence bounds for quantiles, this approach may only be used for quantiles sufficiently far from the tails of the distribution.

The bootstrap method can also be used in reliability assessment by the Monte Carlo technique, the more so that each simulation set gives a different value of X (e.g. the probability of failure P_f). Thus, the Monte Carlo simulation sets are repeated many times. In each set, P_f is determined. Then, its probable highest value is found by one of the above approaches.

5. Concluding remarks

Reliability assessment based on a small number of experimental values always means risk. A very important condition for the use of any probabilistic method is that the experimental sample adequately represents the whole population. The situation can be very dangerous if the population is not homogeneous, for example if several kinds of flaws and other defects can be responsible for the strength of a brittle material [11]. All characteristic kinds of defects must be present in the experimental sample (including the largest but rare ones), otherwise the predicted values of low probability strength can be wrong despite the determination of their confidence limits. The probability that a defect of probability of occurrence (e.g. 1:1000) will be found in a small sample consisting of only 10 pieces is really very low (only 1% compared to the probability 99% that such flaw will not be revealed). A similar situation exists, for example, when the maximum height of water waves at the sea coast should be predicted. Statistical characteristics can be obtained from long-term measurements, but if the waves at extremely rare tsunami events have not been included into the evaluation, the new coast structures will not be sufficiently protected against them.

When dealing with reliability assessment of some kind of structures, it is thus reasonable to gradually gather the data of all measurements (for the pertinent material or load) and to combine the newer data with older ones. For this purpose, Bayesian methods may be suitable, explained briefly in Chapter 22.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

References

- [1] Jílek M. Statistical tolerance limits (in Czech: Statistické toleranční meze). Praha: SNTL; 1988. 280 p.
- [2] Likeš J, Laga J. Statistical tables (in Czech: Základní statistické tabulky). Praha: SNTL; 1978. 564 p.
- [3] Wallis W A. Use of variables in acceptance inspection for percent defective. In: Selected Techniques of Statistical Analysis (C. Eisenhart, M. W. Hastay, W. A. Wallis, eds.). New York: McGraw-Hill Book Comp; 1947. p. 7 – 93
- [4] Menčík J. The Use of Weibull Distribution for Reliability Assessment. Sci. Pap. University of Pardubice, Ser. B: 1997, 3, p. 5 - 24.
- [5] Mann N, Schafer R E, Singpurwalla N D. Methods for Statistical Analysis of Reliability and Life Data. New York: John Wiley and Sons; 1974. 576 p.
- [6] Kececioglu D, Lamarre G. Designing mechanical components to a specified reliability and confidence level. Nuclear Engineering and Design. 1978; 50: 149 – 162.
- [7] Menčík J. Reliability assessment for normally distributed load and structure resistance with small amount of input data. Building Research Journal. 1997; 45: No. 4, 229 – 238.
- [8] Menčík J. Reliability assessment by the Monte Carlo method with small amount of data. Building Research Journal. 1999; 47: No. 1, 27 – 39.
- [9] Efron B, Tibshirani R J. An introduction to the bootstrap. Boca Raton: Chapman & Hall; 1993. 456 p.
- [10] DiCiccio T J, Efron B. Bootstrap confidence intervals. Statistical Science. 1996; 11: 189 – 211.
- [11] Cuccio J et al. Probabilistic Methods for ceramic component design and implications for standards. In: C. R. Brinkman, S. F. Duffy (eds.): Life Prediction Methodologies and Data for Ceramic Materials, ASTM STP 1201, ASTM, Philadelphia, 1994.

Robust Design, Sensitivity Analysis, and Tolerance Setting

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62373>

Abstract

This chapter describes various methods for reduction of uncertainties in the determination of characteristic values of random quantities (quantiles of normal and Weibull distribution, tolerance limits, linearly correlated data, interference method, Monte Carlo method, bootstrap method).

Keywords: Random quantity, uncertainty, normal distribution, Weibull distribution, tolerance limits, correlation, interference method, Monte Carlo method, bootstrap method

The reliability and safety of engineering objects are mostly formed during the design. Every design process has three stages:

1. Proposal of conception,
2. Determination of parameters,
3. Setting the tolerances.

Here, stages 2 and 3 will be explained in more detail, as they are very important for reliability.

1. Determination of optimum parameters — Robust design

After the concept of the construction (an engine, a bridge, a transmitter, etc.) has been proposed, it is necessary to determine all important parameters. However, input quantities often vary or can attain values different from those assumed in design. Good design ensures that the important output quantities will always lie within the allowable limits. This can be

achieved by a suitable choice of nominal values of input quantities and by setting their tolerances.

The nominal values of input quantities form together the **design point**. Its position should ensure the low sensitivity of the output parameters to the deviations of input quantities from nominal values. This is called **robust design** [1]. Figure 1 illustrates its principle on an example with one input variable x : the design point 1 is with high sensitivity, whereas point 2 is with low sensitivity. One can see that the changes of the output quantity y around point 2 are much smaller than around point 1, in both cases for the same changes of x . This also means that acceptable scatter of y can sometimes be achieved with lower demands on the accuracy of input parameters. The reliability is influenced not only by the scatter of input quantities, but also by the position of design point. The ideal position, with the lowest sensitivity to the parameter variations, corresponds to an extreme of the response function $y = f(x_1, x_2, \dots, x_n)$. Various optimization methods exist for finding this position, analytical or with computer modeling. Universal is the "simplex method", where the input variables approach the optimum step-by-step according to a simple algorithm [2, 3]. The graphical representation of the response is very informative. Also, the procedures of **design of experiments** (DOE) are suitable; see books by G. Taguchi and other authors [4 - 7]. The determination of optimum parameters should go hand in hand with the sensitivity analysis.

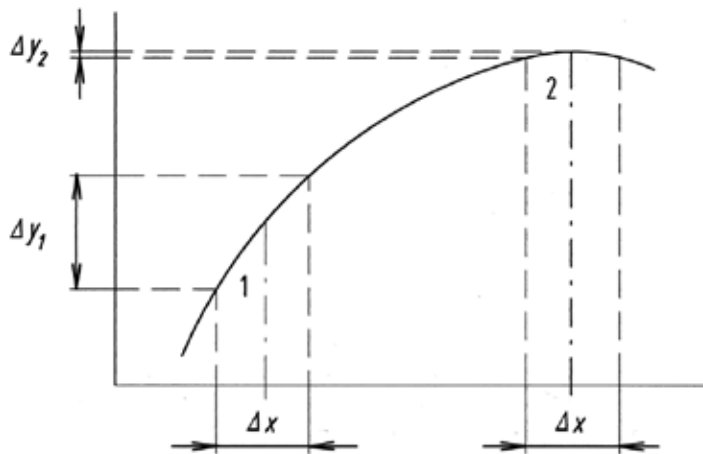


Figure 1. Principle of robust design. Note the influence of the position of the design point on the sensitivity of the output (Δy) to variations of input variable x .

2. Sensitivity analysis

After the design point has been found, the sensitivity analysis could be made to show the influence of the variations of input variables on the variability of the output [8]. The results

may be used for setting the tolerances of input quantities to keep the output in the allowable range. The sensitivity analysis can be done using analytical expressions or simulation methods. The analytical expression for the output variable y ,

$$y = f(x_1, x_2, \dots, x_n), \quad (1)$$

is known exactly only in simple cases (e.g. resonant frequency of an oscillator or deflection of a beam). Often, the **response function** must be found by numerical solution (e.g. using the finite element method). Then, an approximate expression is obtained by regression fitting the response computed for several combinations of input variables (Fig. 3 in Chapter 15).

The sensitivity analysis is usually done in two steps. First, the influence of individual variables is investigated. Several groups of computations are carried out, and in each group, only one variable (x_i) is changed, whereas the others keep their nominal values $x_{1,0}, x_{2,0}, \dots, x_{n,0}$ corresponding to the design point. Then, the y values for the individual groups are fitted by a suitable regression function (e.g. a polynomial),

$$y_i = a_0 + a_{i1}x_i + a_{i2}x_i^2 + a_{i3}x_i^3 + \dots, \quad (2)$$

or

$$y_i = y_0 + a_i(x_i - x_{i,0}) + b_i(x_i - x_{i,0})^2 + \dots, \quad (3)$$

the latter expression characterizes the changes of y as a function of deviations of the i -th input variable from the design point. These regression functions correspond to the cuts through the response surface (Fig. 3 in Chapter 15). The sensitivity analysis will depend on whether the deviations are small or large.

Small changes of the input and output quantities

In this case, linear approximation of the response function may be used, which yields simple expressions. The **sensitivity of the response** to the variations of individual variables is obtained from partial derivatives at the pertinent point,

$$c_i = \partial y / \partial x_i \approx \Delta y / \Delta x_i. \quad (4)$$

For linear approximation, the sensitivity coefficients c_i correspond to the constants a_{i1} in (2) and a_i in (3). Further information is obtained from **relative sensitivities**,

$$c_{ri} = \frac{\partial y}{\partial x_i} \frac{x_{i,0}}{y_0} \approx \frac{\Delta y}{y_0} \frac{\Delta x_i}{x_{i,0}}, \quad (5)$$

where y_0 and $x_{i,0}$ are the values corresponding to the design point. Coefficient c_{ri} expresses the change of y (in %, for example) caused by 1% deviation of x_i from the nominal value $x_{i,0}$. For linear approximation, $c_{ri} = a_i(x_{i,0}/y_0)$.

Generally, two kinds of sensitivity analysis can be made: (1) deterministic, which assumes that the deviations of individual quantities from nominal values have constant magnitude, and (2) stochastic, which assumes the random scatter of individual input quantities around their nominal values.

Both approaches will be illustrated on an example [9]. A cantilever flat spring of rectangular cross-section (Fig. 2) should be used in a precise measuring device. It is necessary to get an idea how the deviations of its individual dimensions and material properties from the nominal values will influence its compliance. The spring compliance C is given by the formula:

$$C = y / F = 4L^3 / (Ewt^3) ; \quad (6)$$

y is deflection, F is load, L is length, E is elastic modulus, w is spring width, and t is spring thickness.

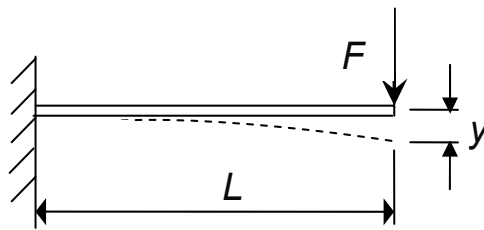


Figure 2. Spring for a measuring device (a schematic).

Deterministic analysis for small deviations

The increments of y are calculated via the first derivatives. The response surface is replaced by a tangent plane at the investigated point. For $y = f(x_1, x_2, \dots, x_n)$, the infinitesimal increment of y can be expressed generally as

$$dy = (\partial y / \partial x_1)dx_1 + (\partial y / \partial x_2)dx_2 + \dots + (\partial y / \partial x_n)dx_n, \quad (7)$$

where $\partial y / \partial x_1$ expresses partial derivatives. For practical reasons, the differentials are replaced by small finite increments Δ ,

$$\Delta y = (\partial y / \partial x_1)\Delta x_1 + (\partial y / \partial x_2)\Delta x_2 + \dots + (\partial y / \partial x_n)\Delta x_n. \quad (8)$$

In our example with the spring, the partial derivative of Equation (6) with respect to the first variable ($x_1 = L$) is

$$\partial C / \partial L = 3L^2 \times 4 / (Ewt^3) = [4L^3 / (Ewt^3)] \times 3 / L = (3 / L) \times C, \quad (9)$$

and the increment of compliance due to a small increment of the beam length ΔL is thus

$$\Delta C = 3C(\Delta L / L). \quad (10)$$

The formulas for other variables are obtained in a similar way. The resultant expression, involving the changes of all variables, is

$$\Delta C = C(3\Delta L / L - \Delta E / E - \Delta w / w - 3\Delta t / t), \quad (11)$$

and the relative sensitivity of the stiffness is

$$\Delta C / C = 3\Delta L / L - \Delta E / E - \Delta w / w - 3\Delta t / t. \quad (12)$$

This formula shows the influence of individual quantities. If the spring will be longer by 1% than the nominal value, the compliance will be higher by 3%; if the elastic modulus E will be higher by 1%, the compliance will be lower by 1%, etc. The constants at individual terms correspond to their exponents in Equation (6), and the signs depend on whether the quantity was in the numerator or denominator.

This preliminary analysis reveals which input quantities have very small influence on the variability of the output quantity y and may thus be considered as constants in the following analysis of simultaneous random variance of the input quantities. However, one must always keep in mind that the variance of the output depends on both the sensitivity c_i and the variance of the pertinent input quantity x_i .

Deterministic analysis for large deviations

The above approach is acceptable if the response function is linear or if the errors due to approximation by linear function are small. If the response function is nonlinear and the investigated ranges of input quantities are not small, the errors will not be negligible (Fig. 3). In such case, it is better to study the influence of deviations of input quantities by modeling the response without simplifications. For example, the influence of j -th variable can be studied from Equation (1), in which only x_j varies, whereas the others keep their values corresponding to the design point.

Influence of random variability – small scatter

The influence of random variability of input quantities can be investigated using the formula for the scatter of a function of several random variables. For small scatter,

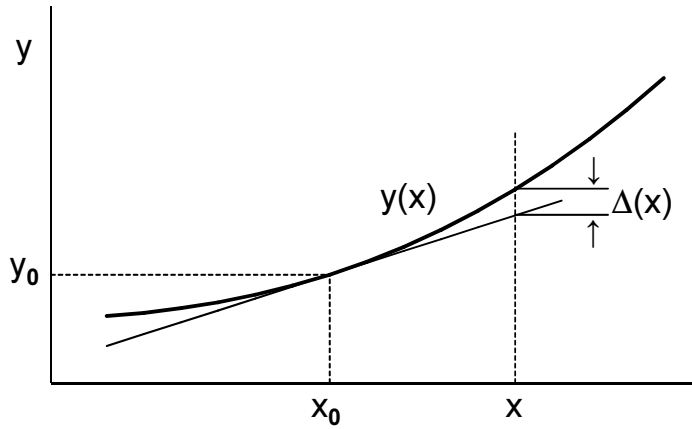


Figure 3. Error caused by linear approximation of response function $y = f(x)$.

$$s_y^2 = \left(\frac{\partial y}{\partial x_1}\right)^2 s_{x_1}^2 + \left(\frac{\partial y}{\partial x_2}\right)^2 s_{x_2}^2 + \dots + 2\left(\frac{\partial y}{\partial x_1}\right)\left(\frac{\partial y}{\partial x_2}\right) \text{cov}(x_1, x_2) + \dots, \tag{13}$$

where $s_{x_j}^2$ is the scatter of the j -th variable (quadrate of standard deviation). The far right-hand term is nonzero if the variables are correlated; often, it can be omitted. For linear approximation of y ,

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n, \tag{14}$$

the scatter is

$$s_y^2 = a_1^2s_{x_1}^2 + a_2^2s_{x_2}^2 + \dots + a_n^2s_{x_n}^2 + \dots \tag{15}$$

The individual components, $s_{y_j}^2 = a_j^2s_{x_j}^2$, give the scatter of y caused by random variations of j -th variable. Similarly to deterministic analysis, the contribution of a certain variable x_j to the total scatter is larger for large scatter of this variable ($s_{x_j}^2$) and for large sensitivity (a_j) of the output y to its changes.

The expression obtained by dividing Equation (10) or (12) by the total scatter s_y^2 gives the relative proportions of individual factors in the total scatter,

$$1 = a_1^2 \frac{s_{x_1}^2}{s_y^2} + a_2^2 \frac{s_{x_2}^2}{s_y^2} + \dots + a_n^2 \frac{s_{x_n}^2}{s_y^2} + \dots \tag{16}$$

The square root of scatter (10) is the standard deviation s_y . If the input quantities have normal distribution, the confidence interval for the output quantity y can be calculated as

$$y_{\text{lower,upper}} = y_0 \pm u_\alpha s_y; \quad (17)$$

the + or – sign corresponds to the upper (or lower) confidence limit and u_α is the α -critical value of standard normal distribution. The probability that y will lie out of these limits is 2α .

If Formula (12) is applied on the above example with a spring, one obtains the following expression for the standard deviation of the compliance caused, for example, by random variability of the length L :

$$s_{CL} = 3C(s_L / L); \quad (18)$$

cf. Equation (11). Similar expressions can be written for other variables. The random variability of all input quantities causes the following variability of the spring compliance:

$$s_c = [(3C / L)^2 s_L^2 + (C / E)^2 s_E^2 + (C / w)^2 s_w^2 + (3C / t)^2 s_t^2]^{1/2}. \quad (19)$$

The ratio of the standard deviation of a quantity and its mean is the variation coefficient,

$$v = \sigma / \mu, \quad (20)$$

so that the combination of Equations (16) and (17) gives the variation coefficient of the compliance,

$$v_c = s_c / C = [9v_L^2 + v_E^2 + v_w^2 + 9v_t^2]^{1/2}. \quad (21)$$

Stochastic analysis for large scatter

The above approach, based on the linearization of the response function, is suitable for small values of variance coefficients of input quantities, say $v_j \leq 10\%$. If their scatter is large, it is better to study the influence of variability or deviations of input quantities by the Monte Carlo simulation method. A preliminary assessment consists of making m simulation experiments with random variable only x_j , for $j = 1, 2, \dots, n$, and then calculating partial scatter $s_{y_j}^2$ of the obtained values y . Using the characteristics s_{x_j} , $x_{j,0}$ and y_{0j} one can determine the variation coefficients v_j or the sensitivity coefficients $a_j (= s_y / s_{x_j})$.

The approximate value of the total scatter is obtained by summing up the partial scatters,

$$s_y^2 = s_{y1}^2 + s_{y2}^2 + \dots + s_{ym}^2 + \dots \quad (22)$$

More accurate value is obtained if all input variables, x_1, x_2, \dots, x_n are considered as random in the Monte Carlo simulations, and the scatter is calculated from all values y_j . Dividing Equation (19) by the total scatter s_y^2 gives the relative influence of individual factors, like in Equation (13).

3. Determination of tolerances of input quantities

If the variability or deviation of the output quantity y from the nominal value is larger than allowed, it must be reduced. The procedure depends on whether the variability is random or deterministic.

Deterministic deviations

If the deviation of y is caused by the deviation of one or more input quantities, Equation (12) or (12), showing the contribution of individual factors to the total deviation of y , can be used to decide which factor should be aimed at. Let us assume that the deviation of y in Equation (12) is caused only by the deviation of x_j . The allowable magnitude of Δx_j , ensuring that the deviation of y does not exceed Δy , is

$$\Delta x_j \leq \Delta y / \left(\partial y / \partial x_j \right) . \quad (23)$$

For example, the allowable length tolerance of the above spring, ensuring the compliance tolerance ΔC , is

$$\Delta L = \Delta C (L / 3C) , \text{ or } \Delta L / L = (1/3) (\Delta C / C) . \quad (24)$$

The tolerances of other quantities can be determined in similar way. One must respect that the deviations of some input quantities influence the output in one direction, whereas the deviations of other quantities can have the opposite influence. Generally, the deviations of y depend on the deviations of input quantities and also on the sensitivity of y to the changes of x_j . The reduction of the tolerance of y can thus be accomplished by tightening the tolerances of individual input quantities or by changing the position of the design point towards lower sensitivity. The decision will also depend on the costs related to the individual adjustments.

Random variability of input quantities

The following analysis assumes that the range of probable occurrence of y (i.e. the half-width Δy_α of the α -confidence interval for y) is directly proportional to the standard deviation s_y , equal to the square root of the scatter. In production, the allowable limits of a quantity x are

also often determined as $x_{nom} \pm ks_y$, where k is a constant (e.g. a suitable quantile of standard normal distribution). With this assumption, the tolerance of y can be reduced from Δy_α to $\Delta y'_\alpha$ by reducing the standard deviation of y from the original value s_y to s'_y . This may be accomplished by the reduction of the variance or influence of input factors.

Often, the influence of one factor prevails (e.g. x_k). In such case, most of the scatter of y can be reduced by reducing its component due to this factor. As it follows from Equation (12), the scatter of y can be reduced by reducing the standard deviation $s_{x,k}$ or the sensitivity of y to the changes of x_k (coefficient a_k). The reduction of variance of y_k can be attained by more accurate manufacturing or by better control and sorting out the components that are out of the tolerance limits. The reduction of sensitivity of y to the changes of x_k can be accomplished by changing the parameters of the design point (Fig. 1). An example is a prestressed flange connection in steam turbines: the use of long bolts increases the compliance of the joint and reduces the sensitivity of the prestress to the variations of pressure in the pipe and thermal dilatations of the flanges. Sometimes, both ways, the reduction of $s_{x,k}$ and a_k is combined.

If several input variables vary, one must decide, which of them should be reduced. As the standard deviation equals the square root of the scatter, it is obvious that the reduction of scatter of a quantity, contributing to the total scatter by only 5% to 10%, will have negligible effect. Also, the costs of the pertinent improving operation must be considered, as they usually increase with tightening the tolerances.

After having obtained the corrected standard deviation s_{xi} , the lower (L) and upper (U) allowable limit for the input quantity x_j can be determined as

$$x_{i,L,U} = x_{i,0} \pm k_{L,U} s_{xi} \tag{25}$$

$x_{j,0}$ is the nominal (design) value and k is a constant (e.g. 5% quantile of standard normal distribution). k_L corresponds to the lower limit, whereas k_U corresponds to the upper limit.

The above optimization can be performed even if the scatter s_y^2 from the preliminary design is smaller than the allowable value. The optimization assigns such tolerances that the total costs are minimal. Sometimes, the tolerances may even be made wider, with lower costs.

Often, the scatter of some input quantities cannot be changed continuously. In such cases, the response must be evaluated for each possible value of every discontinuous quantity.

The determination of suitable tolerances will be illustrated on the following example, adapted from [9].

Example 1

A cantilever microbeam from Figure 2, with length $L = 10$ mm, width $w = 1.0$ mm and thickness $t = 50$ μ m, made of a material with elastic modulus $E = 200$ GPa, has compliance $C = 4L^3/(Ewt^3) = 0.16$ mm/mN. Each input quantity has coefficient of variation $v_i = v_w = v_t = v_E = v = 0.01 = 1\%$. The variation coefficient of the compliance, Equation (17), is $v_C = (9v_L^2 + v_E^2 + v_w^2 + 9v_t^2)^{1/2} =$

0.0447, and the standard deviation $s_c = Cv_c = 0.00716$ mm/mN. Such variation of compliance is unacceptably high and must be reduced to $s_c' = 0.004$ mm/mN.

Solution. The corresponding reduced variation coefficient is $v_c' = s_c'/C = 0.004/0.16 = 0.025$. It is possible to reduce the scatter of L , w , and t ; the material (E) remains unchanged. The easiest way is to reduce the scatter of L . However, even if this scatter were zero, the variation coefficient of compliance would be $v_c' = 0.033$, which is much more than demanded. Therefore, the variance of all three quantities (L , w , and t) must be reduced by more accurate manufacturing. If the new variation coefficients of L , w , and t would have the same value, $v_L' = v_w' = v_t' = v'$, this value v' can be calculated from the modified Equation (18):

$$v_c' = [9v'^2 + v_E'^2 + v'^2 + 9v'^2]^{1/2}. \quad (26)$$

With the variation coefficient of elastic modulus unchanged, $v_E = 0.01$, the new coefficients of variation of L , w , and t must be reduced to $v' \leq 0.005256$, which is approximately $v' = 0.005$. The corresponding allowable standard deviations, obtained by multiplying the variation coefficient v' by the nominal values of L , w , and t , are $s_L' \leq 0.05$ mm, $s_w' \leq 0.005$ mm, and $s_t' \leq 0.25$ μ m. In the limit case, $v_c' = 0.024$ and $s_c' = 0.0038$ m/N. However, the tolerances of individual dimensions could be adjusted with respect to the manufacturing possibilities, the principal condition being $s_c' \leq 0.004$ m/N.

4. Uncertainties in ensuring safety and lifetime using proof testing

If the high reliability of a certain object must be ensured, a proof-test is often used: the component is exposed to some overload, specified so that only sufficiently strong components survive it; the weaker ones are destroyed. In the same way, sufficient lifetime can be ensured for components made of brittle materials suffering by static fatigue. The minimum time to failure of a component that has passed a proof-test is [10 - 12]

$$t_{\min} = \frac{2\sigma_{pt}^{N-2}}{(N-2)AY^2K_{IC}^{N-2}\sigma_0^N}, \quad (27)$$

where K_{IC} is the fracture toughness of the material, N and A are the parameters of subcritical crack growth, Y is the geometrical factor of typical crack, responsible for fatigue failure, σ_0 is the characteristic operational stress (assumed constant), and σ_{pt} is the proof-test stress. A rearrangement of Equation (24) gives the formula for the proof stress needed to guarantee the minimum lifetime:

$$\sigma_{pt} = \sigma_0^{N/(N-2)} K_{IC} \left[\frac{N-2}{2} AY^2 t_{\min} \right]^{1/(N-2)}. \quad (28)$$

However, K_{IC} , N , and A were determined by measurement and are known only approximately and Y was estimated. Therefore, it is recommended to perform sensitivity analysis and correct

the proof stress appropriately. The pertinent theory, based on probabilistic analysis, is explained in [11, 12] or in [10]. For easier application, strength-probability-time diagrams were developed [13 – 15], in which the necessary proof stress can be found for the demanded time to failure and confidence level.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty, University of Pardubice, Czech Republic

References

- [1] Dieter G E. *Engineering Design*. 2nd ed. New York: McGraw-Hill; 1991. 721 p.
- [2] Spendley W, Hext G R, Himsworth F R. *Technometrics*. 1962; 4: No. 4, 441 – 461.
- [3] Menčík J. Efficient tools for reliability-based optimisation. In: 2nd ASRANET International Colloquium, Barcelona, 5-7 July 2004. Universities of Glasgow and Strathclyde, 2004.
- [4] Taguchi G. *Introduction to Quality Engineering: Designing Quality into Products and Processes*. Tokyo: Asian Productivity Organization; 1986. 191 p.
- [5] Ross P J. *Taguchi Techniques for Quality Engineering*. New York: McGraw-Hill; 1996. 329 p.
- [6] Taguchi G, Chowdhury S, Wu Y. *Taguchi's Quality Engineering Handbook*. Hoboken, New Jersey: John Wiley & Sons; 2005. 1696 p.
- [7] Fowlkes W Y, Creveling C M. *Engineering methods for robust product design. (Using Taguchi methods in technology and product development.)* New York: Addison-Wesley; 1995. 432 p.
- [8] Novák D, Teplý B, Shiraishi N. Sensitivity Analysis of Structures: A Review. In: Int. Conf. CIVIL COMP'93, August 1993. Edinburgh, Scotland: 1993. p. 201-207
- [9] Menčík J. Reliability-based parameter optimisation and tolerancing in structural design. In: 8th Int. Conf. ICOSSAR 2001. Newport Beach, USA. Structural Safety and Reliability (R B Corotis, G I Schuëller, M Shinozuka, editors. Lisse: A.A. Balkema Publishers; p. 188

- [10] Menčík J. Strength and fracture of glass and ceramics. Amsterdam: Elsevier; 1992. 357 p.
- [11] Jacobs D F, Ritter J E jr. Uncertainty in minimum lifetime predictions. J. Am. Ceram. Soc. 1976; 59: No. 11/12, 481 – 487
- [12] Wiederhorn S M, Fuller E R jr, Mandel J, Evans A G. An error analysis of failure prediction techniques derived from fracture mechanics. J. Am. Ceram. Soc. 1976; 59: No. 9/10, 403 – 411
- [13] Davidge R W. Mechanical behaviour of ceramics. Cambridge: Cambridge University Press; 1979. 174 p.
- [14] Davidge R W, McLaren J R, Tappin G. Strength probability time (SPT) relationships in ceramics. J. Mater. Sci., 1973; 8: No. 12, 1699 – 1705.
- [15] Evans A G, Langdon T G. Structural ceramics. Prog. Mater. Sci. 1976; 21: No. 3 – 4, 171 – 441.

Reliability Testing and Verification

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62377>

Abstract

This chapter describes various methods for reduction of uncertainties in the determination of characteristic values of random quantities (quantiles of normal and Weibull distribution, tolerance limits, linearly correlated data, interference method, Monte Carlo method, bootstrap method).

Keywords: Random quantity, uncertainty, normal distribution, Weibull distribution, tolerance limits, correlation, interference method, Monte Carlo method, bootstrap method

Reliability tests are often indispensable. The material properties, needed in design, can only sometimes be found in data sheets. If they are not available, they must be obtained by testing, for example the strength of a new alloy or concrete or the fatigue resistance of a vehicle part. Also, the manufacturers of electrical components must provide the reliability data for catalogs (e.g. the failure rate and the data characterizing the influence of some factors, such as temperature or vibrations). It is also impossible to predict with 100% accuracy the properties of a new bridge, an engine or a complex system consisting of many parts, whose properties vary more or less around the nominal values. In all these cases, tests are often necessary to verify whether the object has the demanded properties or if it conforms to the standards. Also, the information on loads (e.g. wind velocities in an unknown area) must often be obtained by measurement.

The reliability tests can be divided into two groups: those for providing detailed information on properties of new materials or components, and those for the verification of the expected values. The former are more extensive, as they must provide the mean value and statistical parameters characterizing the random variability. The extent of verification tests is smaller.

In this chapter, the reliability tests of mass-produced components will be described first, followed by the tests of large or complex structures or components and the tests of strength and fatigue resistance.

1. Testing of mass-produced electrical and mechanical components

The most important reliability characteristics are the mean failure rate and the mean time to failure or between failures. The tests can be done so that several components are loaded in a usual way (e.g. by electric current), and the times to failure of individual pieces are measured. As the time to failure of some samples can be very long, the test is sometimes terminated after failure of several pieces, at time t_t . The total cumulated time of operation to failure is calculated generally as

$$t_{\text{tot}} = \sum t_{f,j} + mt_t, \quad j = 1, \dots, r, \quad (1)$$

where $t_{f,j}$ is the time to failure of j -th piece, r is the number of failed specimens, and m is the number of pieces that have survived the test, whose duration was t_t . The total number of all checked samples is $n = r + m$. If all pieces have failed during the test, $m = 0$ and the term mt_t falls out. (Also other test arrangements are possible, for example with replacing the failed pieces by good ones; see [1] or the corresponding IEC standards listed in Appendix 2. The mean time to failure is calculated as

$$MTTF = t_{\text{tot}} / r, \quad (2)$$

The individual times to failure vary, and this must also be characterized. If failures occur due to various reasons, an exponential distribution of times to failure is often assumed. A simple check of this is the standard deviation σ . For exponential distribution, the standard deviation has the same value as the mean μ (in an ideal case; in real tests it can somewhat differ). If the difference between μ and σ is larger, a statistical test should be made to check whether the exponential distribution is suitable. Common for this purpose are the goodness-of-fit tests (e.g. Kolmogorov-Smirnov or the χ^2 test); see [2 - 4]. If exponential distribution is not suitable, another distribution can be better, e.g. Weibull.

If an exponential distribution is acceptable, the estimate of the mean failure rate can be obtained easily as the reciprocal value of the mean time to failure,

$$\bar{\lambda} = 1 / MTTF. \quad (3)$$

The two-sided confidence interval for the true mean failure rate λ is [1]:

$$\lambda_L = \frac{\chi^2_{1-\alpha/2}(2r)}{2r} \bar{\lambda} \leq \lambda \leq \frac{\chi^2_{\alpha/2}(2r)}{2r} \bar{\lambda} = \lambda_U. \quad (4)$$

if the testing was terminated at the r th failure, and

$$\lambda_L = \bar{\lambda} \frac{\chi^2_{1-\alpha/2}(2r)}{2r}, \quad \lambda_U = \bar{\lambda} \frac{\chi^2_{\alpha/2}(2r+2)}{2r} \quad (5)$$

if the testing continued some time after the r th failure. In Equations (4) and (5), λ is the calculated mean value of λ , the subscripts L and U denote the lower and upper confidence limit, $\chi^2_{1-\alpha/2}(2r)$ is the $(1-\alpha/2)$ -critical value of the chi-square distribution for $2r$ degrees of freedom, $\chi^2_{\alpha/2}(2r)$ is the $\alpha/2$ -critical value for $2r$ degrees of freedom, and $\chi^2_{\alpha/2}(2r+2)$ is the $\alpha/2$ -critical value for $2r+2$ degrees of freedom. The probability that λ will lie within this confidence interval is $\gamma = 1 - \alpha$. Often, we are interested only in the maximum expectable failure rate; the pertinent formula for the upper limit of one-sided interval is

$$\lambda_U = \frac{\chi^2_{\alpha}(2r)}{2r}, \quad (6)$$

the probability that the actual failure rate will be higher is now α . As the mean time to failure failure is the reciprocal of the failure rate, the corresponding two-sided confidence interval for the mean time to failure is obtained as

$$t_L = \frac{2r}{\chi^2_{\alpha/2}(2r)} MTTF \leq \bar{t} \leq \frac{2r}{\chi^2_{1-\alpha/2}(2r)} MTTF = t_U, \quad (7)$$

if the test was terminated after the r th failure (and analogously for a longer test).

The determination and importance of confidence limits will be illustrated on the following examples.

Example 1

Ten electrical components were tested to determine the failure rate. The tests were terminated after $t_T = 500$ h. During this time, six components failed ($r = 6$), in times: 65, 75, 90, 120, and 410 h. Four components survived the test. Estimate the mean time to failure and failure rate and construct two-sided confidence intervals (for confidence $\alpha = 90\%$).

Solution. The mean value and standard deviation of the times to failure of the six failed components were 168.66 and 136.33 h, respectively. It is thus possible to assume exponential distribution.

The cumulated duration of tests, calculated after [1], was:

$$t_{tot} = \sum_{i=1}^6 t_i + 4 \times t_t = 60 + 75 + 90 + 120 + 250 + 410 + 4 \times 500 = 3010 \text{ h.}$$

The mean time to failure is $t_{\text{mean}} = t_{\text{tot}}/r = 3010/6 = 501.67 \text{ h}$, and the mean failure rate is $\lambda_{\text{mean}} = \lambda = 1/t_{\text{mean}} = 1/501.67 = 1.993 \times 10^{-3} \text{ h}^{-1}$.

The lower and upper confidence limits for λ were calculated, with respect that the tests were terminated before the failure of all samples, according to Equation (5). With $r = 6$ and $\alpha = 10\%$, the critical values are $\chi^2_{0.95}(12) = 5.226$ and $\chi^2_{0.5}(14) = 23.685$. Inserting them, together with $\lambda_{\text{mean}} = 1.993 \times 10^{-3} \text{ h}^{-1}$ into (5) gives $\lambda_L = 8.68 \times 10^{-4} \text{ h}^{-1}$ and $\lambda_U = 3.93 \times 10^{-3} \text{ h}^{-1}$. The confidence limits for the mean time to failure are $t_L = 1/\lambda_U = 254.4 \text{ h}$ and $t_U = 1/\lambda_L = 1152.1 \text{ h}$. The mean time to failure thus can lie within the interval $t_{\text{mean}} \in (254 \text{ h}; 1152 \text{ h})$.

As we can see from this example, the confidence interval obtained from only six failures is very wide. If it should be narrower (to get more accurate estimate), it is necessary either to make a longer test so that more parts of the tested group fail or to increase the number of parts tested simultaneously or both.

Example 2

The above testing has continued until the time $t_t = 1000 \text{ h}$. During this time, two more pieces failed, at the times $t_7 = 520 \text{ h}$ and $t_8 = 760 \text{ h}$.

Solution. The same procedure as above has given the following results: $T = 4290 \text{ h}$ and $r = 8$, so that the mean time to failure is now $t_{\text{mean}} = T/r = 4290/8 = 536 \text{ h}$ and the mean failure rate $\lambda_{\text{mean}} = 1/536 = 1.865 \times 10^{-3} \text{ h}^{-1}$. Also, the confidence interval will respect that more pieces have failed. The critical values now are $\chi^2_{0.95}(16) = 7.962$ and $\chi^2_{0.5}(18) = 28.869$. With all these values, the lower and upper limits of failure rate are $\lambda_L = 9.28 \times 10^{-4} \text{ h}^{-1}$ and $\lambda_U = 3.4 \times 10^{-3} \text{ h}^{-1}$. The mean time to failure t_{mean} thus can be expected to lie within the interval (297 h; 1078 h).

The whole test lasted twice as long as the previous one, but the new confidence interval is only slightly narrower. If significantly more accurate estimates should be achieved, much longer tests or with substantially higher number of tested pieces must be done. Thus, when preparing the tests for the determination of failure rate, one should estimate in advance the duration of the test, the number of tested pieces, and the number of pieces that can fail — all this for the acceptable probability α that the actual maximum failure rate would be higher than that obtained from the test.

The rearrangement of the expression for the upper limit of confidence interval for λ gives the following relationship between the expected failure rate λ_0 , the number of tested samples n , test duration t_v , and the number of failed components r [1]:

$$nt_i = \chi^2_{\alpha}(2r) / (2\lambda_0) . \tag{8}$$

If the number of failed samples does not exceed r , the actual failure rate is not higher than λ_0 , the risk of wrong prediction being α .

As it follows from the product $n \times t_i$ in Equation (8), the number of tested parts n is equivalent to the test duration t_i . This means that the same information can be obtained by testing, for example, 10 specimens for 1000 h or 1000 specimens for 10 h. If the tested objects are expensive, one would prefer testing fewer specimens for longer time. However, at least several pieces should always be tested to reduce the risk that the only piece chosen at random for the test was especially good or especially bad.

The following table, based on Equation (8), shows the values of the product $n \times t_i$ for the various numbers of failed parts during the tests; the probability of a wrong result is $\alpha = 10\%$.

$\lambda_0(h^{-1})$	$n \times t$ for $r = 3$	$n \times t$ for $r = 5$
0.001	5322	7994
0.0001	53,223	79,936
0.00001	532,232	799,359
0.000001	5,322,320	7,993,590

Table 1. Extent of tests for various failure rate and the number of failed pieces.

For example, the reliability testing of components with assumed exponential distribution, failure rate $\lambda = 10^{-4} h^{-1}$ and the test terminated after the fifth failure, needs $n \times t_i = 79,936 \approx 80,000$ pieces \times hour. Thus, for example, 100 components should be tested 800 h or 800 components for 100 h. If the expected failure rate were $\lambda = 10^{-6} h^{-1}$, then $n \times t_i \approx 8,000,000$ pieces \times hour, so that 10,000 components must be tested for 800 h or 100 components for 80,000 h. One can see that testing for proving the reliability of very reliable components becomes very difficult or impracticable. Therefore, various **accelerated tests** are often used. One way, suitable for the items working periodically with pauses between the operations, such as switches or valves, eliminates the idle times: the switch is permanently switched on and off.

Another way to obtain the demanded reliability information sooner uses a higher intensity of load (e.g. higher mechanical load, higher electric stress or electric current, or more severe environment (e.g. higher temperature or vibrations). If this approach should be effective, one must know the mechanism of degradation and the relationship between the load intensity and the rate of degradation. For example, the rate of chemical processes, which are the cause of some failures, often depends on the temperature according to the Arrhenius equation:

$$r = C \exp\left(-\frac{\Delta E}{kT}\right), \tag{9}$$

C is a constant, ΔE is the activation energy, k is the Boltzmann constant, and T is the absolute temperature (K). If the times to failure have exponential distribution, the failure rates or times to failure are related with the absolute temperatures as follows [1]:

$$\frac{\lambda_1}{\lambda_2} = \frac{t_2}{t_1} = \left[\frac{\Delta E}{k} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) \right], \quad (10)$$

Equation (10) can be used for the determination of necessary temperature change from T_1 to T_2 if the test duration should be reduced from t_1 to t_2 .

Similarly, the number of cycles to fatigue failure of periodically loaded components can be reduced by increasing the characteristic stress or load amplitude P . The basic relationship, based on the Wöhler-like curve [Equation (1) in Chapter 6], is

$$t_1 / t_2 = C (P_1 / P_2)^n; \quad (11)$$

C and n are constants for a given material and environment. Similar relationships can be used for finding the increased load for shortened tests of components exposed to creep or static fatigue (stress enhanced corrosion), with rates depending on some power of the load.

Today, mass-produced electronic and electrical components are tested in special chambers and under special conditions enabling acceptably short duration of the tests. More about these tests, denoted **HALT** (for highly accelerated life testing) or **HASS** (for highly accelerated stress screening), can be found in the literature, for example [5].

Sorting tests

These tests aim at sorting out “weak” items that could fail shortly after being put into service. However, they must not cause excessive degradation of properties in “good” components (i.e. they should not shorten their life significantly). Sorting tests can be nondestructive or destructive. **Nondestructive tests** use visual observation, X-ray, ultrasound or magnetic inspection, and special electrical or other measurements. **Destructive tests** can be arranged in several ways, for example **proof tests** that use short-time overloading by mechanical or electrical stress exceeding the nominal value so that the weak parts are destroyed during the test. Other ways for revealing the weak parts are artificial aging under increased temperature, cyclic loading by varying temperatures (this causes additional thermal stresses that can reveal hidden defects or weak joints), the use of burn-in period with 75% to 100% of nominal load acting several tens of hours before putting into service, special kinds of mechanical loading, such as impacts, vibrations of certain amplitude, and frequency, overloading of rotating parts by centrifugal forces and others.

2. Acceptance sampling

This operation, common in series production, ensures that only those batches of items will be released to the customer or to the next operation, which are either perfect or contain only very

small proportion of out-of-tolerance parts. Before this control is introduced, a test plan must be prepared, which contains:

1. Kinds of monitored indicators,
2. Number of tested items,
3. Duration of the tests,
4. Criterion for the decision on the acceptance.

Generally, three approaches are used:

- a. **100% control.** Every component or item must pass the inspection, and those that do not fulfill certain parameters are discarded. This control is most expensive, but it should be the safest. Nevertheless, if the evaluation depends on human senses (visual check, for example) here, a small probability of erroneous decisions also exists. The inspector can overlook a defect or, vice versa, he can denote a good item as defective, especially if the number of tested items is very high, which can lead to his fatigue. An example is the check for internal flaws using X-rays, with the images interpreted via observation by naked eye. Also, 100% control cannot be done if every test ends with the destruction of the tested piece, even if it is good (e.g. the check of the airbag deployment system in cars).
- b. **Random inspection.** Only several pieces, chosen at random, are tested (e.g. 1% of the batch). The entire lot is accepted or rejected according to the result of the inspection. This kind of acceptance is much less demanding than 100% control, but it has been criticized that it is rather subjective and not sufficiently reliable. If, for example, a batch of 10,000 pieces contains 1% of defective piece and it was decided that 1% will be tested, then 100 pieces must be checked. One percent of 100 is one piece. However, it can happen that the checked sample will contain not exactly one defective piece, but two or three or even none. This uncertainty has led to the development of the following method based on the probability theory.
- c. **Statistical acceptance.** Several variants exist. The principle will be explained on the so-called **single sampling plan**. A sample of n pieces is taken at random from the lot and tested. The number z of out-of-tolerance pieces, found in the sample, is compared with the so-called **decisive number** c . If it is lower, $z < c$, the whole lot is accepted; if $z > c$, it is rejected. The values of c for various expected proportions p of unsuitable pieces and extent n of the tested sample can be found in the standards for statistical acceptance [6] or calculated, with a consideration of further important parameters, **AQL** (acceptable quality level) and **LQL** (limiting quality level, also called the lot tolerance percent defective, **LTPD**). LQL gives the maximum fraction of defectives, acceptable, on average, in the batches denoted as good. The principle of determination of the decisive number c is as follows. If the fraction of defectives in the population is p , the number z of the defectives that can appear in a random sample of size n has binomial or Poisson's distribution (for low probabilities p in the latter case). It is thus possible to calculate the cumulative probabilities for $z = 0, 1, 2, 3, \dots$. The decisive number c is such that only very low probability β exists that the lot, whose test has given $z \leq c$, will contain higher percentage of defective

than LQL. The probability β is called **customer's risk** and means the risk that an unsatisfactory lot will be accepted as good. On the other hand, also a **producer's risk** α exists, such that a good lot, with less defective pieces than AQL, will be rejected. Usually, 5% or 1% is chosen for both α and β .

The curve showing how the probability of accepting the lot decreases with increasing proportion of defectives in the sample is called the **operating characteristic curve (OCC)**. Figure 1 shows examples of OCCs for two different decisive numbers.

The rejected batch is either discarded or 100% checked. In the latter case, the good pieces are added to other good items. This makes the average quality of the batches composed in this way better, so that the quality demands in the tests may slightly be reduced.

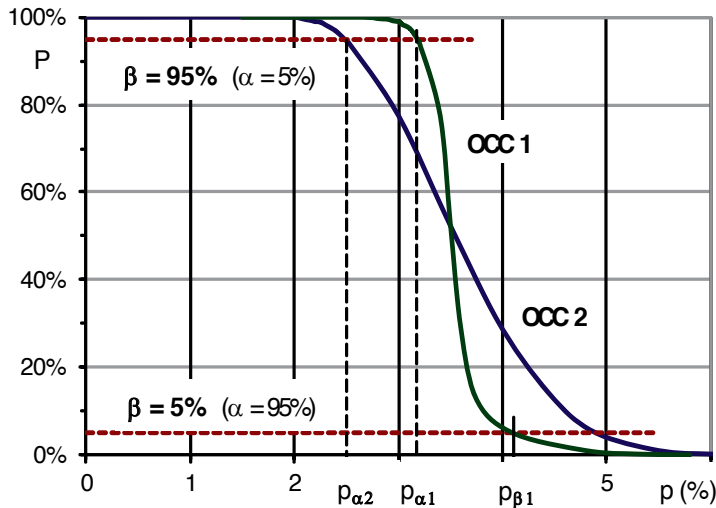


Figure 1. Operating characteristic curve (OCC). P - probability of acceptance; p - percentage of defectives in the population; α - producer's risk; β - customer's risk. Subscripts 1 and 2 denote curves OCC1 and OCC2.

Also other schemes exist. For example, a **double sampling scheme** uses two decisive numbers, c_1 and c_2 . If the number z of defectives in the first sample is smaller than c_1 , the lot is accepted, and if it is higher than c_2 , it is rejected. If $c_1 < z \leq c_2$, another sample is taken and the total number of defective in both samples is checked, etc. Further modifications, such as multiple sampling or sequential sampling, exist as well. For more, see [6].

However, doubts are sometimes cast on the cost-effectivity of statistical control. On the one hand, this control costs money. On the other hand, losses can arise due to possible defective pieces hidden in the batches checked as good. Deming [7] has pointed out that if the cost for inspection of one piece is k_1 and the average cost of a failure caused by not inspecting is k_2 and the average fraction of defective is p , then, if $pk_2 < k_1$, the lowest total costs (control costs plus costs caused by failures) will be achieved without any testing. If $pk_2 > k_1$, full (100%) inspection

should be used, especially for higher ratios pk_2/k_1 . However, the situation is often not so simple; the fraction p of defectives can vary, 100% testing can be impracticable for too high investment costs or if all tests end with destruction, etc.

The statistical acceptance was very popular in the second half of the 20th century but not so much today. There are two reasons: the demands on quality and reliability are much higher today than 50 years ago and the allowable probabilities are often of the order $1:10^6$, much lower than the degree of confidence common in statistical sampling. Moreover, the controlling devices are much more powerful today. The incorporation of automated test equipment (ATE) into production line enables 100% control.

3. Testing of large structures and complex components

These tests will be illustrated on two cases: bridges and large components exposed to fatigue, such as parts of heavy vehicles (e.g. locomotives).

The assumed service life of road and railway bridges is many tens of years and sometimes more. During this time, the structure deteriorates and its safety decreases. Also, the loading pattern can change in a long time (new kinds of vehicles and changed traffic demands). For these reasons, bridges must sometimes be repaired or reconstructed. In such case, thorough inspections are done at suitable time, including **load tests** in important cases. In these tests, the bridge is usually loaded by a group of trucks loaded by sand or concrete blocks as much as possible so that the load-carrying capacity of the bridge is attained. During the tests, deformations and stresses at selected points are measured and compared with the values obtained by computer analysis of the structure – to see if the actual response (e.g. deflection of some parts of the bridge) corresponds to the assumed response. In some cases, dynamic properties are also studied (i.e. the response to periodic or dynamic loading). If the actual condition is worse than allowed, measures must be taken for improvement.

Large parts of mechanical structures, such as vehicles or aircrafts (sometimes these objects as a whole), are mechanically loaded with the purpose to find whether the actual response (deformations and stresses at selected points) corresponds to the values assumed in design. Also, dynamic response is investigated. Exceptionally, the object is loaded until the destruction. In the past, the measurements were often the only reliable source of information of the stresses and behavior. Today, the methods of stress analysis are much better and much information can be obtained by computer simulation as early as in the design stage. Therefore, today the tests serve rather for confirmation whether the demanded parameters have been achieved.

The test loads are often imposed by electrohydraulic cylinders attached to the tested object. Often, special test stands are used, consisting of a massive frame with hydraulic cylinders, clamping equipment, and a controlling unit. The work of the stand is controlled by a computer. This enables one to program the demanded loading sequences. Sometimes, the load program is based on a record made during a test vehicle driving on real roads or on a test track containing

typical examples of road surfaces. The test vehicle is equipped with sensors (usually strain gauges fixed at certain points of the car body) and the measured data are recorded. These data must be transformed to the data for the control of the load cylinders of the testing stand. The reason is that these cylinders are often attached to the tested structure at different points than were those used in the test vehicle driven on the track. Also, the data recorded with one test vehicle are sometimes used for the testing of other types of vehicles. The test stand can repeat the recorded load sequence again and again, so that also fatigue resistance can be tested in this way.

4. Tests of strength and fatigue resistance

These tests are often arranged according to various standards. In this paragraph, we thus limit our attention to some probabilistic aspects of these tests.

Strength tests. The individual values vary, so that the number of tests should be adjusted to the purpose of the measurement and to the scatter of individual values. If only approximate information on the average strength is needed, three tests may be sufficient; the standard deviation can serve for the estimation of confidence interval of the mean strength. However, especially for brittle materials with high scatter of individual values, the knowledge of the "minimum" strength is often demanded. This is determined as a low-probability quantile. For this purpose, more tests must be done, often several tens or more. From these tests, the parameters of strength distribution are determined. Often, Weibull distribution, but also log-normal distribution, is assumed. The determination of parameters and quantiles of Weibull distribution was described in Chapters 11 and 18. The parameters of log-normal distribution are found in several steps. In the first step, logarithms are taken from the measured values, then the average and standard deviation are calculated from the transformed data, and finally they are transformed back to the original system of units. The question of which distribution is better can be solved by means of statistical tests of goodness of fit [2 – 4].

Generally, many values are necessary to obtain reliable values of low-probability quantiles of strength. (Remember that 1% quantile corresponds to the minimum of 100 values.)

Fatigue tests. The main purpose of fatigue tests is the determination of fatigue limit (if it exists) and finding the relationship between the characteristic stress (S) and the time or number of cycles to failure (N_f). As for the fatigue limit, everything from the above paragraph on strength tests remains valid. The S N relationship is obtained by making the tests under various characteristic stress amplitudes and fitting the data by a suitable function, for example [8, 9]:

$$N_f = A S^{-w}, \quad (12)$$

or a similar expression. Now, two possibilities exist depending on the number of tests that were done or could be done with respect to the available money and time. If only several tests have been performed, all measured $N_f(S)$ values are fitted by the regression function (12). The

consequences of the scatter of individual values are depicted in Figure 1 in Chapter 18. The regression function, obtained by the least-squares method, gives such N_f values that probability 50% exists that the true number of cycles to failure under a chosen stress will be 50% lower (!) than the number obtained from the regression function. The "safe" $N_{f,\alpha}$ values, for which acceptably low probability α would exist that the component or construction can fail earlier, may be found as boundary values of the pertinent confidence band for all $S-N$ data; see Chapter 18.

If more values (e.g. tens) are available for each stress level, a more accurate procedure can be used. The data for individual stress levels are rank-ordered in ascending order. Each value corresponds to some quantile of time to failure for a given stress level. For example, the shortest time of 10 values obtained for the same stress corresponds approximately to 10% quantile of the time to failure. Now, only the $N_{f,\alpha}(S)$ values, corresponding to the same quantile α , are fitted by regression function (12). The "safety" of the prediction of the time to failure with this function equals $1 - \alpha$. It is also possible to fit all measured data by function of type (12) with additional parameters characterizing the probability that the actual number of cycles to failure will be lower than that calculated via modified Equation (12).

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

References

- [1] Bednařík J et al. Reliability techniques in electronic practice (In Czech: Technika spolehlivosti v elektronické praxi). Praha: SNTL; 1990. 336 p.
- [2] Freund J E, Perles B E. Modern elementary statistics. 12th ed. New Jersey: Prentice-Hall; 2006. 576 p.
- [3] Suhir E. Applied Probability for Engineers and Scientists. New York: McGraw-Hill; 1997. 593 p.
- [4] Montgomery D C, Runger G C. Applied Statistics and Probability for Engineers. 4th ed. New York: John Wiley; 2006. 784 p.
- [5] Levin M A, Kalal T T. Improving Product Reliability. Strategies and Implementation. Chichester, England: John Wiley & Sons; 2003. 313 p.

- [6] Schilling E G, Neubauer D V. Acceptance Sampling in Quality Control. Boca Raton: CRC Press (Chapman & Hall); 2009. 700 p.
- [7] Deming W E. Out of the Crisis. Reprint Edition. Cambridge MA: The MIT Press; 2000. 485 p.
- [8] Fuchs H O, Stephens R I. Metal Fatigue in Engineering. New York: Wiley and Sons; 1980. 336 p.
- [9] Stephens R I, Fatemi A, Stephens R R, Fuchs H O. Metal fatigue in engineering. 2nd ed. New York: John Wiley & Sons; 2001. 473 p.

Multicriterial Condition Evaluation and Fuzzy Methods

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62380>

Abstract

This chapter describes various methods for reduction of uncertainties in the determination of characteristic values of random quantities (quantiles of normal and Weibull distribution, tolerance limits, linearly correlated data, interference method, Monte Carlo method, bootstrap method).

Keywords: Random quantity, uncertainty, normal distribution, Weibull distribution, tolerance limits, correlation, interference method, Monte Carlo method, bootstrap method

Large and long-life engineering structures, such as bridges, dams, buildings, or cooling towers, deteriorate gradually due to corrosion, mechanical damage, fatigue, and other processes. As a consequence, their safety slowly decreases, and after some time, they must be repaired. Every decision of a repair must be based on a good knowledge of the actual technical condition. This is gained from regular inspections. However, it is impossible to characterize the overall condition of a complex object by means of only one simply measurable quantity. It is influenced by many factors, and most of them can be characterized only verbally (e.g. slightly corroded reinforcement or many short cracks in the wall). Probabilistic and exact methods cannot be applied everywhere, often because of the lack of data and the vagueness of the characteristic criteria and way of their evaluation, and also because of the lack of appropriate models relating the extent of the defects to the load-carrying capacity or lifetime of the object. The overall safety and reliability characteristic is obtained by a suitable processing of many partial ratings, each for every criterion. Such evaluation is usually based on a judgement by a person with long practice – an expert. The results of this approach, using his experience and intuition, are usually reasonably good. Nevertheless, methods that are more objective and less sensitive on the person of the inspector are needed. In this chapter, two procedures will be described: a simple method that assigns weights to the individual criteria and uses a simple rule for their aggregation, and an advanced method for more complex cases, which uses fuzzy logic tools.

Simple multicriterial condition assessment

The method can be explained on the evaluation of bridges, but a similar approach may be used for other objects, too. The decision on a repair is based on the results of inspections. Various systems exist for the classification of bridge condition, usually based on a scale with several degrees. For example, in Czech Republic a three-degree scale is used for railway bridges (1 = good, 2 = satisfactory, 3 = unsatisfactory), whereas a seven-degree scale is used for road bridges (1 = faultless, 2 = very good, 3 = good, 4 = satisfactory, 5 = bad, 6 = very bad, 7 = emergency, danger of collapse). In Poland, a continuous scale between 0 and 5 is used.

The condition evaluation is based on several criteria. For small concrete railway bridges, they pertain to: 1 = condition of the beams, 2 = condition of water insulation, 3 = condition of the cornice, and 4 = response to train passage. The bridge inspection protocol with verbal description of the situation serves for assigning weights to the individual criteria. This is facilitated by the catalog relating the weights to various degree of damage. The sum of the weights for individual criteria forms the resultant characteristic of the overall condition (R) and serves for the decision whether or not the bridge should be repaired.

The procedure will be explained here on an example of a simple bridge [1].

Example 1

A concrete bridge was inspected to evaluate its overall condition. The results of the inspection, written into a protocol, were as follows:

1. Concrete beams. The concrete plastering between the beams at the bridge bottom is with cracks, the steel beams have begun corroding.
2. Water insulation. The insulation is damaged, the slab and supports leak water through, lime leaches are formed.
3. Cornice. Hair cracks exist along the height on the left side, the lower edge above the support no. 2 has fallen away, the reinforcement is bare.
4. Response during train passage. Quiet.

These results were then compared with the catalog of weights (W) for various conditions. An extract from the list is shown below. The first subscript denotes the criterion and the second subscript denotes the classification. The weights range from 0 to 1; higher values correspond to better condition.

1. **Concrete beams** (7-degree classification). $W_{1,7} = 0.9 - 1.0$: Concrete plastering is without cracks, the steel reinforcement is fully covered, the edges of beams are without rust, protective painting is in order. $W_{1,6} = 0.8$: Concrete plastering at the bottom contains hair cracks, the steel reinforcement is bare at lengths less than 0.05 m (i.e. not substantial), the edges of beams are with slight rust, the protective painting starts flaking off. $W_{1,1} = 0.0 - 0.1$: The plastering has significantly fallen away, the edges of steel beams are very rusty, with the thickness reduced by 2 to 3 mm, cracks are present in the concrete, and the concrete crumbles up to the depth 60 mm.

2. **Water insulation** (5 degrees). $W_{2,5} = 1.0$: Good, water seepage not found. $W_{2,1} = 0.0 - 0.3$: Heavily damaged, intensive seeping over the total area, lime stalactites have been formed.
3. **Cornice** (5 degrees). $W_{3,5} = 1.0$: Concrete plastering is without cracks, the steel reinforcement is fully covered, the edges of beams are without rust, protective painting is in order. $W_{3,1} = 0.0 - 0.3$: The plastering has significantly blown and fallen away, the edges of steel beams are heavily rusty and with the thickness reduced by 2 to 3 mm, cracks appear in the concrete, the concrete crumbles up to the depth 60 mm.
4. **Response during train passage** (3 degrees). $W_{4,3} = 0.8 - 1.0$: Quiet. $W_{4,2} = 0.5 - 0.7$: Neither quiet nor unsteady. $W_{4,1} = 0.0 - 0.4$: Unsteady.

Based on this list, the individual criteria of the investigated bridge were assigned the following weights: $W_1 = 0.65$, $W_2 = 0.45$, $W_3 = 0.85$, and $W_4 = 0.90$. The resultant characteristic R of the overall condition is obtained as the sum of the individual weights:

$$R = W_1 + W_2 + W_3 + W_4 = 0.65 + 0.45 + 0.85 + 0.90 = 2.85.$$

This result can be interpreted using the following 3-degree scale based on experience:

$R = 3.0 - 4.0$, Degree 1. Condition: Good. Load-carrying construction needs only common maintenance.

$R = 1.8 - 3.0$, Degree 2. Condition: Satisfactory. Load-carrying construction needs repair (more extensive than common maintenance), but the defects do not endanger the safety.

$R = 0.0 - 1.8$, Degree 3. Condition: Unsatisfactory. Load-carrying construction needs total reconstruction or exchange of the load-carrying construction or substantial repair.

The above value $R = 2.85$ can thus be interpreted as "satisfactory condition, but a repair is necessary". The bridge manager will decide about the repair (also with respect to the money available, condition of other bridges in the network, etc.; see Chapter 17).

Note: The method can be improved by assigning weights to the groups of individual criteria to better respect the influence of each criterion on the safety of the whole structure.

2. Fuzzy logic approach to condition assessment

As we have seen, the above characterization of bridge condition was based not on exact values but on rather vague terms. There are many situations like this. In daily life, we often describe the situation as "slightly increased temperature" or "the girders are very rusty". Even a driver controls his car in terms "fast-slow" and "near-far". The necessity of working with such "fuzzy" quantities has led to the development of methods based on fuzzy sets [2, 3]. These methods enable work with linguistic and numerical quantities and allow their combination as well as the use of mathematical and logical operators (IF, AND, OR, THEN,...). The procedures for application of fuzzy methods are principally similar as the above multicriterial condition assessment. However, instead of one single ("sharp") value for each criterion, they use so-

called membership functions and offer more flexibility and better characterization of the situation (Figure 1A). The application of fuzzy logic in decision processes consists of three steps: fuzzification, fuzzy inference, and defuzzification.

STEP 1. Fuzzification

The real values of input variables are transformed into fuzzy values of linguistic variables. This is done by assigning a suitable attribute to each basic variable. An example of such variable is "deflection of a beam" and an example of an attribute is "small". Often, three to seven attributes are used (e.g. positive big, positive medium, positive small, zero, negative small, negative medium, and negative big). The fuzzy approach uses membership functions that express the degree of correspondence of individual quantities to their definitions. For example, usual operating temperature of a machine is from 40°C to 70°C. The temperature 75°C can be considered as increased, but still also as operating. Its appropriateness to operating conditions, however, is not so high as if it were within the above interval. The fuzzy approach enables dealing with just such cases. Examples of various membership functions are shown in Figure 1A; $\mu(x) = 1$ means full correspondence of x with its definition.

STEP 2. Fuzzy inference

In the second step, mathematical and logic operations are performed with the fuzzified input variables. For example, "If 'A' is small and 'B' is high, then 'C' is small". The output is also fuzzy or in a form of a linguistic variable. A suitable processing of membership functions for several input variables gives the membership function of the result. For example, if a load "about 5 kN" acts on a structure and also a load "about 10 kN", then the total load is "about 15 kN". Figure 1B shows this simple case for triangular membership functions.

STEP 3. Defuzzification

In this step, the fuzzy result is transformed into a sharp value of the output variable, characterizing the overall condition, e.g. "the damage degree is 4.3". Various methods exist for this purpose: position of the centroid of the resultant membership function, the first of maxima, etc. If the technical condition of a structure is evaluated, the resultant statement can be "the condition is good (satisfactory, bad,...)". This serves for the decision about the further operation or repair of the object.

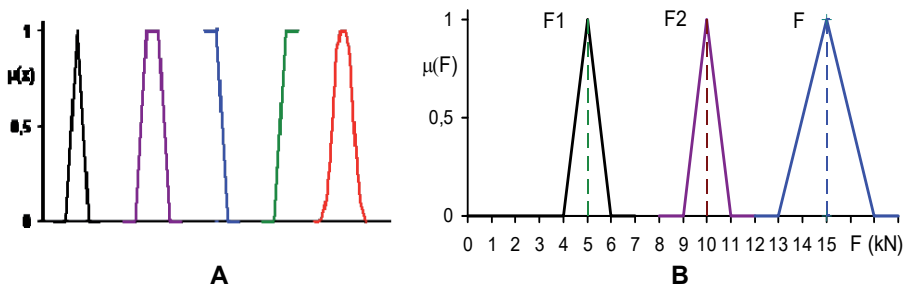


Figure 1. (A) Examples of membership functions and (B) example of composition of two fuzzy quantities ($F = F1 + F2$).

The condition assessment using fuzzy logic needs computer support. Special programs may be created, but commercial software can also be used. For example, Matlab offers universal Fuzzy Logic Toolbox [4]. It enables the definition of various membership functions (e.g. for the intensity of damage and extent of damage or other quantities relevant to the particular problem). The user can also choose the rules for the inference process from a database. The solution is controlled by the editor of fuzzy inference system (Fig. 2), and the results can be presented in graphic form.

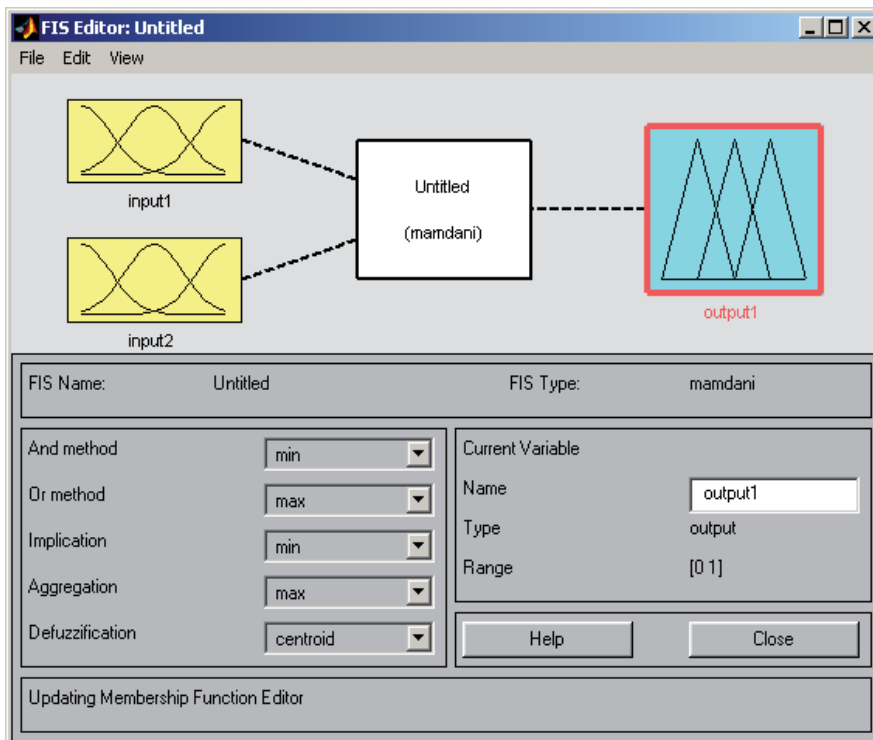


Figure 2. Editor of fuzzy inference system in Matlab®.

The main parts of a fuzzy-logic tool are: an editor, databases of membership functions and rules for work with them, and a viewer on the resultant membership function. Before computer-aided condition evaluation with fuzzy methods may be applied on an engineering object, the following steps must be done:

1. Definition of quantities, which have influence on the condition of the object.
2. Definition of membership functions for individual attributes of each quantity.
3. Definition of rules for the construction of the resultant membership function.
4. Definition of rules for defuzzification (= for assigning a sharp value to the characteristic quantity), allowing unambiguous decision about operation or repair.

There are many publications on fuzzy methods. Their use for reliability assessment is explained in [2, 3, 5 - 7]. Rudolf [8] developed an application of computer aided fuzzy inference for the evaluation of bridges; see also [1].

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

References

- [1] Menčík J, Křupka J, Rudolf P. Evaluation of technical condition, safety and reliability of existing engineering structures. Int. Conf. ESREL, Rhodos. 5.-9. 9. 2010. ESRA. Book of Abstracts, p. 76, full text (8 pages) on CD-ROM. Taylor & Francis, ISBN 978-0-415-60427-7.
- [2] Möller B, Beer M. Fuzzy Randomness. Berlin: Springer; 2004. 336 p.
- [3] Möller B, Reuter U. Uncertainty Forecasting in Engineering. Berlin: Springer; 2007. 202 p.
- [4] Fuzzy Logic Toolbox User's Guide [online]. 2008. The MathWorks, Inc. Available from: http://www.mathworks.com/access/helpdesk/help/pdf_doc/fuzzy/fuzzy.pdf. [Accessed: 2015-12-10]
- [5] Biondini F, Bontempi F, Malerba P G. Fuzzy reliability analysis of concrete structures. Computers and Structures. 2004; 82; 1033–1052.
- [6] Kala Z. Fuzzy Probabilistic Analysis of Steel Structure Focused on Reliability Design Concept of Eurocodes. Nonlinear Analysis: Modelling and Control. 2007; 12; 371–382.
- [7] Kala Z. Fuzzy-random analysis of steel structures. Engineering Mechanics, 2007; 14; No. 1, pp. 1 – 12.
- [8] Rudolf P. Reliability and the Service Life of Existing Bridges and Their Assessment Using Modern Computational Tools. [PhD thesis]. Pardubice: University of Pardubice, Jan Perner Transport Faculty; 2009.

Bayesian Methods

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62378>

Abstract

Probabilistic Bayesian methods enable combination of information from various sources. The Bayes theorem is explained and its use is illustrated on several examples of practical importance, such as revealing the cause of an accident or reliability increasing of non-destructive testing. Also its use for continuous quantities and for increasing the reliability of the parameters of normal or Weibull distribution is shown.

Keywords: Statistics, probability, Bayes, Bayes theorem, reliability, non-destructive testing, normal distribution, Weibull distribution, combination of information

The term “Bayesian methods” denotes probabilistic methods that enable the combination of information on some event or quantity with previous information from measurement or experience. The use of additional information can increase the reliability of our information or reduce the extent of measurements needed for making conclusions on certain event. Examples of application are the determination of the most probable cause of a failure, increasing the reliability of diagnostic methods or increasing the accuracy of the determination of distribution parameters of random quantities.

Bayesian methods are based on the so-called Bayes theorem [1 – 6]. It was originally formulated for discrete quantities, but extended later for continuous quantities as well. These methods have also been included into standards. In this chapter, their principle will be explained, and the use is shown on several practical examples.

1. Bayes theorem

Let us assume that an event (B) can occur if another event (A) has occurred. The event A , however, could occur by several ways (A_1, A_2, \dots, A_n), which are mutually exclusive. The probability of simultaneous occurrence of both events A_j and B is calculated as

$$P(BA_j) = P(A_j) \times P(B|A_j), \quad (1)$$

where $P(A_j)$ is the probability of event A_j , and $P(B|A_j)$ is (conditional) probability that event B can occur provided that event A_j has happened. The total probability of event B is

$$P(B) = \sum P(BA_j); \quad (2)$$

the summation is done for all possible cases $j = 1, 2, \dots, n$. Bayes theorem looks at the issue from the opposite side: "If event B has happened, what is the probability that it was as a consequence of (or after) event A_j ?" With the use of Equations (1) and (2) and the fact that $P(BA_j) = P(A_jB)$, this probability can be expressed as [2 – 6]:

$$P(A_j|B) = P(A_j) \times P(B|A_j) / P(B), \quad (3)$$

where the total probability $P(B)$ in the denominator is calculated from individual probabilities via Equations (2) and (1). Equation (3) is the simplest form of Bayes theorem. Its use will be shown on three examples. The first example, adapted from [4], does not solve a reliability problem, but is very instructive.

Example 1. Identification of origin of a sample from several possible sources.

The materials for road building are delivered from two plants with daily capacities 300 t (plant 1) and 700 t (plant 2). The long-term monitoring of quality shows that plant 1 has 2% of all batches faulty and plant 2 has 4% faulty batches. If now a sample is chosen at random at the building site, and if this sample is faulty, which plant is the batch from?

From the total amount of $300 + 700 = 1000$ t/day, plants 1 and 2 produce 30% and 70%, respectively. Let us denote event A_1 : the sample is from plant 1; event A_2 : the sample is from plant 2. The corresponding probabilities are $P(A_1) = 0.3$; $P(A_2) = 0.7$. Event B : the sample is defective. The probability of defective sample from plant 1 is $P(B|A_1) = 0.02$, and from plant 2, it is $P(B|A_2) = 0.04$. The total fraction of faulty production is: $P(B) = 0.3 \times 0.02 + 0.7 \times 0.04 = 0.034 = 3.4\%$. The defective material from plant 1 represents $0.02 \times 0.3 = 0.006$ from the total production of both plants. This is $0.006/0.034 = 0.176 = 17.6\%$ from the total faulty production. Similarly, plant 2 produces 82.4% of the scrap. These numbers also say that if the randomly chosen sample was faulty, a probability of 17.6% exists that it is from plant 1 and 82.4% that it is from plant 2.

Using Bayes rule (3), one can express the probability that the defective specimen is from plant 1 as $P(A_1|B) = P(A_1) \times P(B|A_1)/P(B)$. The values $0.3 \times 0.02/0.034$ yield the same result 0.176 as

above. Similarly, the probability that the faulty sample is from plant 2, $P(A_2|B) = 0.7 \times 0.04 / 0.034 = 0.824$ ($= 1 - 0.176$).

If the quality is not considered, the probability that a randomly chosen sample comes from plant 1 equals 30% (i.e. the fraction of production from plant 1). If, however, additional information "the sample was defective" was used together with the information on quality in both plants, this probability has dropped to 17.6%. The same information has increased the probability of the sample being from plant 2 from 70% to 82.4%. Although the probability that a sample is from plant 2 was higher even without the Bayes rule (70%), the strengthening of this hypothesis is obvious.

Further strengthening of the hypothesis by using more tests

The hypothesis "the material is from plant 1 (or 2)" can be strengthened (or mitigated) by checking more specimens. If n specimens are taken from one batch, and if all appear to be defective (= event B'), then the expression $P(B|A_i)$ in Bayes rule (3) must be replaced by the expression $P(B'|A_i) = P(B|A_i)^n$. For example, if three specimens were taken from a batch from the above example, and if all were faulty, then $P(B'|A_1) = 0.02^3$, $P(B'|A_2) = 0.04^3$, $P(B') = 0.3 \times 0.02^3 + 0.7 \times 0.04^3 = 0.0000472$, and $P(A_1|B') = 0.3 \times 0.02^3 / 0.0000472 = 0.05$. Similarly, $P(A_2|B') = 0.95$. In such case, it is nearly sure that the batch was from plant 2.

Example 2. Revealing the most probable cause of an accident.

This example is adapted from [2]. An explosion occurred during a repair of a tank for liquid natural gas. The accident could have happened due to (1) static electricity, (2) fault in the electric equipment, (3) work with open flame during the repair, or (4) intentional act (sabotage). Engineers for risk analysis estimated that the accident could happen with a probability of 25% due to static electricity, 20% due to a fault in the electric equipment, 40% due to work with open flame, and 75% due to a sabotage. The discussion with them also gave the following subjective assessment of probability of individual causes: 0.30, 0.40, 0.15, and 0.15. What is the most probable cause of the explosion in view of all this information?

Solution. Event A : presence of conditions for explosion: $P(A_1) = 0.30$; $P(A_2) = 0.40$; $P(A_3) = 0.15$; $P(A_4) = 0.15$ (note: $\sum P(A_i) = 1.00$). Event B : explosion. The probabilities of explosion under particular conditions are $P(B|A_1) = 0.25$; $P(B|A_2) = 0.20$; $P(B|A_3) = 0.40$; $P(B|A_4) = 0.75$. Total probability of the accident: $P(B) = 0.30 \times 0.25 + 0.40 \times 0.20 + 0.15 \times 0.40 + 0.15 \times 0.75 = 0.3275$. Probability that the explosion has happened due to: (1) static electricity: $P(A_1|B) = 0.30 \times 0.25 / 0.3275 = 0.229 = 22.9\%$, (2) electric appliance: $P(A_2|B) = 24.4\%$, (3) open flame: $P(A_3|B) = 18.3\%$, and (4) sabotage: $P(A_4|B) = 34.3\%$. [Compare these updated probabilities with the original estimates $P(A_i)$.]

Example 3. Increasing the reliability of nondestructive testing.

Welded components are tested for the occurrence of defects (cracks). The device used for nondestructive testing is not perfect. It classifies defect correctly (as defect) only with probability 98%, whereas, in 2% of all cases, it does not recognize the crack and classifies the

component as good. On the contrary, the device marks 96% of good parts as good, but 4% classifies as with a crack. According to long-term inspection records, 3% of all tested components contain cracks. The questions are: If the tested part was classified as "wrong" (i.e. with a defect), what is the probability that it is actually (a) wrong or (b) good? And what about if the component was classified as "good"?

Solution. Event A_1 : Component contains a defect, A_2 : component is good. $P(A_1) = 0.03$; $P(A_2) = 0.97$. Event B : component is classified as wrong. $P(B|A_1) = 0.98$; $P(B|A_2) = 0.04$. The fraction of tested components marked as wrong: $P(B) = 0.03 \times 0.98 + 0.97 \times 0.04 = 0.0682$.

Case 1a. Probability that the component marked as wrong is actually wrong, is $P(A_1|B) = P(A_1) \times P(B|A_1)/P(B) = 0.03 \times 0.98/0.0682 = 0.431 = 43.1\%$. Case 1b. Probability that the component marked as wrong, is actually good, is $P(A_2|B) = 0.97 \times 0.04/0.0682 = 0.569 = 56.9\%$. (Remark: Due to the high proportion of good parts (98%), the proportion of good but rejected parts is high.)

Event B' : Component is classified as good. $P(B'|A_1) = 0.02$; $P(B'|A_2) = 0.96$. The total fraction of components, denoted as good, is $P(B') = 0.03 \times 0.02 + 0.97 \times 0.96 = 0.9318$. Case 2a. Probability that the component marked as good is actually wrong, is $P(A_1|B') = 0.03 \times 0.02/0.9318 = 0.00064 = 0.06\%$. Case 2b. Probability that the component marked as good is actually good is $P(A_2|B') = 0.99936 = 99.94\%$.

Recommendation: All rejected components could be tested once more to reduce the number of discarded good components.

A similar approach can be used in medicine (e.g. in cancer screening).

2. Bayes rule for continuous quantities

If the probability of event B depends on the value of a continuous quantity A , described by the probability density $f(A)$, it is possible to calculate the total probability of this event as

$$P(B) = \int [P(B|A)f(A)] dA; \quad (4)$$

the integration is performed over the whole domain A . [The integration has replaced the summation in Equation (2).] An example is the nondestructive detection of cracks in welds: $P(B)$ is the probability of crack detection, $f(A)$ is the probability distribution of cracks of size A , and $P(B|A)$ is the probability of detection of a crack of size A , the so-called "probability of detection" curve (shortly POD curve) of the device.

Now, a question can be asked: If event B (result of the test) has occurred, what is the actual distribution of random variable A ? Bayes rule (3) can be modified also for this case; the formula for updated distribution of quantity A is [1, 3]:

$$f(A|B) = f(A) \times P(B|A) / P(B), \tag{5}$$

where $P(B)$ is given by Equation (4). For example, the updated distribution of crack lengths can be used for the estimation of time to fatigue failure by the Monte Carlo method [1, 2].

2. Other applications

Bayesian methods can also be used for the improvement of parameter estimate of various probability distributions. Three examples follow.

Parameters of normal distribution

The mean value μ and standard deviation σ of a population with normal distribution are usually unknown, so that they are replaced by their estimates m and s from a sample of size n . The estimate of the mean value can be refined via confidence interval:

$$m - t_{\alpha,\nu} \sqrt{(s/n)} \leq \mu \leq m + t_{\alpha,\nu} \sqrt{(s/n)}, \tag{6}$$

where $t_{\alpha,\nu}$ is α -critical value of Student's t -distribution for $\nu = n - 1$ degrees of freedom.

The estimate can be made more accurate if additional information is available (e.g. estimates of m_0 and s_0 from previous measurements or records). If the number n_0 of these values is known, and if the assumption can be made that all samples (new and old) belong to the same population, the following procedure may be used. The updated average is calculated as the weighted average of both sample averages:

$$m_u = (nm + n_0m_0) / n_u; n_u = n + n_0, \tag{7}$$

where n_u is the updated number of values. The updated standard deviation is

$$s_u = \sqrt{\frac{(n-1)s^2 + (n_0-1)s_0^2 + nm^2 + n_0m_0^2 - n_um_u^2}{n_u - 1}}. \tag{8}$$

Then, the updated confidence interval for μ can be calculated with m , s , and n in (9) replaced by the updated values m_u , s_u , and n_u . If n_0 is unknown, the literature on Bayesian methods recommends an approximate formula [1, 5, 6]:

$$n_0 = s^2 / s_0^2, \tag{9}$$

based on the idea that m_0 and s_0 carry information corresponding to a fictitious sample of certain size n_0 . The smaller the scatter s_0^2 compared to s^2 , the more important are the original results and the larger is the size of the fictitious sample. (An important condition for this estimate is that the "a priori" values of m_0 and s_0 were obtained from large samples.)

Quantiles of normal distribution

The ISO 12491 standard "Statistical methods for quality control of building materials and components" recommends the following formula for the Bayesian estimate of p -quantile of normal variable x :

$$x_{p,B} = m_u + t_p s_u \sqrt{1 + 1/n_u}, \quad (10)$$

where $t_p = t_p(\alpha, p, \nu_u)$ is p -quantile of Student's t -distribution for $\nu_u = n_u - 1$ degrees of freedom. If no additional information is available, the standard recommends the original values m , s , and n .

Weibull distribution

Some quantities, such as the strength or time to failure due to fatigue, can often be approximated by Weibull distribution:

$$F(x) = 1 - \exp\left\{-\left[(x-x_0)/a\right]^b\right\}. \quad (11)$$

The parameters a , b , and x_0 are determined from tests. (The threshold value x_0 is often assumed equal zero.) Sometimes, the number of tests is too low for obtaining reliable values of all parameters. Fortunately, the investigated component or structure is often not a quite new solution but rather an improvement of the current conception. In such case, one can expect that the failure mechanism will be similar as in the previous construction. As the parameter b is closely related to the character of failures, one can assume that the value b will be approximately the same as for the previous components and use it as a known constant. Under this assumption, the finding of the remaining parameters a and x_0 from small amount values is more reliable. This approach is called "Weibayes" [7]. The assumed value b is more reliable if it was determined from many tests. It is thus suitable to keep records from all tests – for possible use in the future!

4. Software for Bayesian methods

The problems from the above first three examples can be solved easily using Excel and standard Bayesian notation. Some simple programs can be found in the literature, for example [2, 3]. At ETH Zürich, a PC program Combinfo was created, which enables the combination of data from various sources [8], including vague information, such as probability estimates by experts or

by judgment. The program allows assigning various weights to individual information. Bayesian methods are also incorporated into software packages for reliability analysis, such as www.reliability.com, www.weibull.com, www.reliasoft.com, or www.itemsoft.com.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

References

- [1] Ang A H S, Tang W H. Probability Concepts in Engineering Planning and Design. Vol. 1, „Basic Principles“. New York: John Wiley; 1975. 574 p.
- [2] Martz H F, Waller R A. Bayesian Reliability Analysis. New York: John Wiley; 1982. 745 p.
- [3] Press S J. Bayesian statistics: Principles, models and applications. New York: John Wiley & Sons; 1989. 256 p.
- [4] Harr M E. Reliability-based design in civil engineering. Mineola, NY: Dover Publications; 1996. 291 p.
- [5] Holický M. Principles of verification of reliability and life endurance of structures (in Czech: Zásady ověřování spolehlivosti a životnosti staveb). Praha: ČVUT; 1998. ISBN 80-01-01880-6. 101 p.
- [6] Holický M, Marková J. Principles of reliability theory and risk evaluation (in Czech: Základy teorie spolehlivosti a hodnocení rizik). Praha: ČVUT; 2005. ISBN 80-01-03129-2. 115 p.
- [7] Abernethy R B. The New Weibull Handbook. 4th ed. North Palm Beach, Florida: Robert B Abernethy; 2001. ISBN 0-9653062-1-6.
- [8] Scheiwiller A. Informationsverknüpfung im Bauwesen. IBK Bericht Nr. 241, ETH Zürich, Basel: Birkhäuser Verlag; 1999. 126 p.

Epilogue – Ways to Higher Reliability

Epilogue — Ways to Higher Reliability

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62379>

Abstract

This last chapter summarizes the means and recommendations for reliability increasing, which are suitable for design and operation.

Keywords: Reliability, materials, design, control, standards, maintenance, computers, measurements, diagnostics, standards, failure analysis

Many current products are much more reliable than in the past, and such trend should continue. This book summarized the methods that contribute to higher reliability. The first part (Chapters 1 to 10) explained the basic terms and methods, whereas the second part (Chapters 11 to 22) explained more advanced tools for reliability evaluation and optimization. This epilogue summarizes the means that have enabled growth of reliability and gives a brief list of recommendations for reliability increasing.

1. Experience from failures and accidents

All big accidents of aircrafts, ships, big structures, or chemical plants have been thoroughly investigated. Intensive attention has also been devoted to the frequent failures. The analysis of causes and time course of failures contributed to the improvement of the pertinent objects or constructions, to the improvement of manufacturing and building processes, and to the creation of various standards and procedures for increasing reliability and safety.

2. Reliability theory and statistical methods

The basic concepts and quantities for reliability characterization and measurement have been defined. Gradually, the main kinds of failures and their causes were identified and the

characteristic course of failure rate during the life of various objects (bathtub curve) was explained. Statistical analysis enables a better understanding and prediction of failures. The theory of probability has led to the measures for increasing reliability (e.g. the use of redundancy or the optimization of reliability of complex systems by allocating various reliabilities to the individual components). It also enables one to formulate reasonably reliable conclusions from limited information (e.g. minimum strength of a material or statistical acceptance). However, efficient methods for reliability increasing have also been developed, which do not work with probability, such as failure modes and effects analysis (FMEA).

3. Better materials, better components, and better technologies

Due to systematic research, many new materials have been developed since World War II, with outstanding properties (e.g. plastics, such as Teflon, carbon fibers, and synthetic diamond). Also, various methods of surface treatment, which increase the resistance to corrosion, high temperatures or wear (e.g. hard TiN layers on machine tools), or strength (e.g. glass strengthening by ion exchange in the surface layer). A great variety of components are available on the market. The manufacturers of various machines, appliances, and other products can buy components tailored for particular purposes and thus bring their products to perfection. Also, high-precision tools and technologies exist, which enable a better achievement of the demanded parameters of components or products.

4. Better knowledge in mechanics, materials science, and other branches

During the last 50 years, various branches of engineering sciences have made significant progress (e.g. strength of materials, fatigue, fracture mechanics, dynamics, heat transfer, and flow of liquids and gases as well as control). In design and dimensioning, one can use better models for the response of structures and appliances to operation loads. Today, a much better knowledge of materials and the causes and time course of their deterioration and failure also exists.

5. Possibility to analyze, simulate, and test the objects via computer models

The improvement thanks to computers is enormous. Computers can quickly process a huge amount of information. In the past, stresses and deformations could be analyzed only in components of simple shapes and loads, and the results were often only approximate. Very important at those times thus was the testing of physical models and actual constructions, which is cumbersome and expensive. Today, computers allow the analysis and solution of very complex problems. For example, the programs for finite-element analysis can relatively and accurately determine the stresses in complicated bodies and reveal their critical parts. As

early as in the design stage (which is crucial for reliability), it is possible to reveal the behavior under many load variants and conditions, including extreme ones. Unsuitable solutions can thus be excluded in advance. This facilitates the finding of the optimum shape or configuration, especially if computer programs for optimization are used. All this reduces the necessary extent of tests of prototypes (which are, nevertheless, still important). Computers also allow one to store and process information from the operation, which can be used for the optimization of maintenance and gradual improvements of the object.

6. Obtaining reliable design data by measurements and tests

The properties of materials or standard components can be obtained from material data sheets or manufacturers' catalogues. If they are missing, or in very demanding cases, they are gained via special tests. The important parts or prototypes are tested during the development. Overload tests done before putting the component or object into service can reveal weak pieces. Examples are load tests of bridges, overpressure tests for pressure vessels, tests of rotating parts under significantly higher velocities, and high voltage tests of electrical components and appliances. Special kinds are proof tests, in which all "weak" parts with unacceptably large defects are destroyed by controlled overloading.

7. Better techniques for the measurement of various quantities and for control; the use of diagnostics and design of intelligent devices

There has been a great improvement in measuring technologies, sensors, and devices for the analysis and processing of various quantities and signals (e.g. vibration diagnostics). All these are significantly enhanced by computers. Today, it is possible to measure gradual changes and deterioration of a component or machine and the changes of the operating conditions. In this way, the appliance can be switched off and repaired before a serious failure happens. Many production processes are 100% monitored. For example, in the production line for glass bottles (with the rate one bottle per second), all important parameters are measured at every bottle and also their changes with time. This, together with the identification of individual moulds, makes possible an early intervention targeted only at the problematic mould. The evolution proceeds towards smart devices with self-control. Two examples of intelligent elements from everyday life can be named: indication of unfastened seatbelts in a car and automatic dynamic balancing of the content in a home washing machine before spinning.

8. Codes and mandatory procedures to ensure reliability and safety

Experience, gathered continuously for a long time, has been incorporated into standards and regulations. These include a variety of proven procedures and practices that guarantee a

universally acceptable level of reliability; see, for example, the codes for the design of steel structures or standards for production and acceptance control. Codes also represent certain etalons in disputes arising due to malfunction or accident.

9. Organizational measures, consistent control of processes, and operation

Even the best technical ideas, solutions, and regulations are useless if their application is not ensured. In complex processes, this must be achieved organizationally. Where necessary, checks must be done consistently at the input, during the process, and at the output. It is reasonable to seek ways for the elimination of human errors (e.g. by replacing physical or mental work by machines and computers). If this is not possible, emphasis must be put on personal responsibility. For example, in manual welding or inspection of welds, every qualified worker has his personal stamp to confirm that it was he who has done the operation.

10. Better approaches to maintenance

From the originally used maintenance after failure, the development went over preventive maintenance, done in fixed intervals, to on-condition maintenance, which strongly uses technical diagnostics and decides with respect to the actual condition. The newest trend is reliability centered maintenance, which reduces the pertinent costs by the elimination of all unnecessary maintenance works as revealed by a thorough analysis.

11. Competition and legal responsibility for defects and failures

In economic systems where supply exceeds demand and a possibility of choice exists, emphasis is put on reliability. Free market and competition make permanent pressure on manufacturers to improve their products. If several firms can make certain products at a similar price, the firm, whose products are more reliable, will win, as the losses due to the failures of its products will be lower. The pressure toward increasing reliability is also supported by legislation, with legal responsibility for any defect, failure, or damage caused by them.

12. Recommendations for reliable design and operation

1. Use experience from failures and accidents of similar objects in the past.
2. Obtain all demands on service life and reliability (dependability).
3. Obtain the relevant information on the conditions of operation and other demands.

4. Obtain all relevant information on the loads acting during operation and in extraordinary or extreme situations (e.g. during building and due to climatic events).
5. Select proper materials and components with guaranteed quality and reliability; use well-proven manufacturing technologies.
6. Use reliable data on materials and components. Obtain them by testing if they are missing.
7. Use reliability methods, including probability and statistics.
8. Use standards and well-proven procedures for reliability and safety ensuring.
9. Propose the optimum arrangement of the object.
10. Use redundancy for reliability increasing. Optimize reliability allocation to individual parts and components.
11. Use the failure modes and effects analysis (FMEA) and the fault tree analysis.
12. Use the robust design approach. Make analysis of the sensitivity to variations of input quantities. Use design of experiments (DOE). Set appropriate tolerances.
13. Use proper dimensioning for the assumed loads and demanded load response.
14. Use up-to-date methods from mechanics and other relevant branches.
15. Use computer-optimized design. Use simulation and study the response in common and extreme situations. Do it as early as in the design stage.
16. Diminish the influence of human factor. Use "intelligent" appliances.
17. Control the purchased parts and operations in manufacture and building.
18. Monitor all important parameters in operation to predict the best time for renewal.
19. Use diagnostics.
20. Prescribe and use suitable system of maintenance and repairs.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

Sources of Information on Reliability

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62365>

This appendix gives the names of publishing houses, journals, web servers, and professional bodies related to reliability. The web addresses correspond to December 2015.

1. Books

Useful books on reliability, probability, risk, maintenance, and related subjects are listed in the References in this book. Many others can be found in various libraries. Some books are freely accessible via <http://www.knovel.com>; it is sufficient to write the book name or a keyword into a “search window” there. The primary sources of books on reliability are various publishing houses. Among the best known, the following can be named:

Elsevier Science(<http://www.elsevier.com>)

Springer(<http://www.springer.com>)

John Wiley & Sons(<http://wiley.com>)

Taylor and Francis(<http://www.tandfonline.com>)

2. Journals

The web pages presented below offer information on the reliability-related journals, contents of individual volumes and abstracts, as well as other useful information. For the unsubscribed readers, often it is possible to buy the pertinent article and sometimes even to get free access to the full text via web.

Reliability Engineering & System Safety(www.journals.elsevier.com/reliability-engineering-and-system-safety)

A scientific journal devoted to the development and application of methods for the enhancement of the safety and reliability of complex technological systems. It is published by Elsevier in association with the European Safety and Reliability Association and the Safety Engineering and Risk Analysis Division.

Reliability Edge(<http://www.reliabilitynews.com>)

A journal published four times a year by ReliaSoft Corporation brings articles related to the reliability engineering theories and principles along with useful information on ReliaSoft's upcoming training seminars and product updates.

Reliability HotWire(<http://weibull.com/hotwire/index.htm>)

An Internet journal bringing news from reliability and solutions of various practical problems (see also the title "Weibull" below).

Quality and Reliability Engineering International(<http://www3.interscience.wiley.com/cgi-bin/jhome/3680>) ([http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1099-1638](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1099-1638))

A scientific journal devoted to the problems of quality and reliability. Published by John Wiley & Sons, Ltd.

IEEE Transactions on Reliability(<http://www.ieee.org>, <http://ieeexplore.ieee.org>)

A scientific journal devoted to the principles and practices of reliability, maintainability, and product liability pertaining to electrical and electronic equipment.

IEEE Transactions on Device and Material Reliability(<http://ieeexplore.org/xpl/RecentIssue.jsp?punumber=7928>)

A scientific journal devoted to the reliability of electronic elements.

Microelectronics Reliability(<http://www.journals.elsevier.com/microelectronics-reliability>)

A scientific journal devoted to the reliability of electronic components. Published by Elsevier.

Software Testing, Verification, and Reliability(<http://www3.interscience.wiley.com/cgi-bin/jhome/13635>)

A scientific journal devoted to the problems of testing and reliability of software. Published by John Wiley & Sons, Ltd.

Maintenance Technology(<http://mt-online.com>)

An electronic journal devoted to maintenance.

Maintenance Resources(<http://www.maintenanceresources.com/productsshowcase/index.htm>)

A journal for professionals on reliability and maintenance.

Warranty Week(<http://www.warrantyweek.com>)

The newsletter for warranty management professionals.

3. Internet

The Internet is a very important source. This paragraph presents the links to several servers devoted to reliability and related topics. However, the named sources correspond to the date when this book was published, and changes cannot be excluded (as usual with Internet).

Reliability Web(<http://www.reliabilityweb.com>)

This web contains lot of useful information, including links to the specialized papers on reliability (freely accessible), calendar of events, info on books to buy, or discussion forum.

ReliaSoft Corporation(<http://www.reliasoft.com>)

This corporation offers various software and courses oriented to the automotive industry but not only to this. It also operates the portal <http://www.reliawiki.com>, with useful resources freely downloadable, including comprehensive handbooks, such as Life Data Analysis Reference, Accelerated Life Testing Reference, System Analysis Reference – Reliability, Availability & Optimization, Experiment Design & Analysis Reference, and Reliability Growth Analysis Reference. **Reliawiki** also provides the access to the magazine Reliability HotWire.

Barringer & Associates(<http://www.barringer1.com>)

Consultancy firm offering various courses from reliability area and many interesting texts (including some standards of Military Handbook) and software products, some of them for free. It also offers books for sale.

System Reliability Center(<http://src.alionscience.com>)

The Center provides expert services, information support and education for people engaged in reliability and also a forum for the exchange of information from reliability, information on literature, software tools, standards, publications, and older issues of the journal published by SRC. It operates a Toolbox with answers to many practical problems.

Weibull(<http://www.weibull.com>)

The reliability engineering web site devoted to theory, data analysis, and modelling. It includes sections on life testing, system reliability and maintainability, reliability growth analysis, FMEA and FMECA, reliability-centered maintenance (RCM), and design of experiments (DOE). It also contains the info on books and free access to the magazine *Reliability HotWire* for reliability professionals.

Quanterion Solutions, Inc.(<https://quanterion.com>)

A firm offering (among other products) courses, consulting services, special books on reliability, and downloadable informative texts and data related to various reliability problems.

Maintenance Resources(<http://www.maintenanceresources.com>)

Web pages related to maintenance and maintainability.

Maintenance World(<http://www.maintenanceworld.com>)

A source for reliability and maintenance management and professionals for the exchange of experience. It informs on conferences and courses and contains freely accessible articles dealing with solution of many practical problems related to operation and maintenance.

Plant Maintenance Resource Center(<http://www.plant-maintenance.com>)

A portal for industrial maintenance, informing about issues on maintenance and reliability, including books. It contains many articles devoted to practical problems. Some links, unfortunately, are no more active, as this service was understood as a competition to Google.

Google, Wikipedia, and Wikimedia(<http://www.google.com>, <http://www.wikipedia.com>, <http://www.wikimedia.com>)

Web portals enabling search for information on any topic, including reliability, probability, mathematics, maintenance, and many others.

Information on failure rates

Very important for the assessment of reliability and availability of various appliances and systems are the data on reliability of mass-produced components. Some values of failure rate can be found, for example, in the databases **Electronic Parts Reliability Data** and **Nonelectronic Parts Reliability Data**. The data contained in these databases represent a compilation of field experience in military, commercial, and industrial applications. Both databases were created by Reliability Information Analysis Center (RIAC), originally for the U.S. Department of Defense. Some of the data from the database Electronic Parts Reliability Data (EPRD) are available at <http://theriac.org/productsandservices/products/downloads/content/EPRD%20Sample.pdf>, and some data from the database Nonelectronic Parts Reliability Data (NPRD) are available at <http://theriac.org/productsandservices/products/downloads/content/NPRD-2011SamplePages.pdf>.

More comprehensive versions of these databases can be purchased, for example via Quantarion Solutions, Inc., or other providers. Some data on the reliability of electric components can be found in the Military Handbook Reliability Prediction of Electronic Equipment (MIL HDBK 217), 1991 issue; see <http://www.sre.org/pubs/Mil-Hdbk-217F.pdf>.

Note: Some of the data available in the above databases are not “up-to-date” but still contain useful information.

4. Professional bodies

European Safety and Reliability Association (ESRA)(<http://www.esrahomepage.org>)

A nonprofit professional organization aimed at the advancement of safety and reliability technology in all areas. It informs on various activities and conferences, publishes ESRA Newsletters, and organizes the annual European Safety and Reliability Conference ESREL.

European Safety, Reliability & Data Association (ESReDA)(<http://www.esreda.org>)

This Association provides a forum for the exchange of information, data, and current research in Safety and Reliability.

International Association for Probabilistic Safety Assessment and Management(<http://www.iapsam.org>)

The main purpose of IAPSAM is to sponsor and oversee the organization of the International Conferences on Probabilistic Safety Assessment and Management (PSAM).

IEEE Reliability Society(<http://www.ieee.org/portal/site/relsoc>)

A division of the important professional organization IEEE publishes Transactions on Reliability and other journals; organizes international symposia on reliability, availability, quality, and system safety; and informs on various events, standards, and literature.

American Society for Quality, Reliability Division(<http://asq.org/reliability>)

A section for reliability within the American Society for Quality. It informs on various activities, conferences, literature, journals, newsletters, and courses.

Safety Engineering and Risk Analysis Division (SERAD)(https://community.asme.org/safety_engineering_risk_analysis_division/w/wiki/3574.about.aspx)

A division of the American Society for Mechanical Engineers (ASME), which stimulates an interest in risk analysis and safety information applied to mechanical engineering.

Society of Automotive Engineers, Reliability, Maintainability, Supportability, and Logistics Division (G-11)(<http://www.sae.org/standardsdev/aerospace/g11.htm>)

A division of SAE providing an industry/government forum to review RMS technology and its interfaces with logistics support, engineering design and development, maintainability, reliability, and diagnostics, especially for automotive and aerospace industries.

Safety and Reliability Society(<http://www.sars.org.uk>)

International professional organisation. Among other activities, it publishes a quarterly journal *Safety and Reliability*, each volume devoted to certain topic.

Society for Maintenance and Reliability Professionals(<http://www.smrp.org>)

A U.S./international nonprofit professional society aiming at the advancement of reliability and physical asset management industry. It is valuable for practitioners looking to expand their knowledge and skills in maintenance and reliability and build business connections with other professionals.

Society of Reliability Engineers(<http://www.sre.org>)

A professional society. The web site contains info on activities, articles, and references.

In addition to those mentioned above, national organizations for reliability and quality exist in various countries and can be found via the Internet. An example follows.

Czech Society for Quality, Reliability Section(<http://www.csq.cz>)

CSQ is an association bringing together individuals and organizations engaged in quality management, reliability, risk and security, environmental management, automotive industry, technical standardization, and others. The special sections organize courses and seminars and publish the corresponding materials (mostly in Czech language).

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

Standards Related to Reliability

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62366>

1. IEC – International Electrotechnical Commission (<http://iec.ch>)

Due to big importance, reliability issues are covered by many standards, which provide well-tested procedures and design values. The standards are the result of cooperation of many experts from various countries and contribute worldwide to the high level of quality, reliability, and safety of products and services and to the reduction of failures and accidents. Standards also facilitate legal disputes related to the compensation of various damages. In some cases, the use of standards is compulsory, but sometimes it is a matter of agreement between the supplier and the customer. However, competition among many manufacturers leads to high emphasis on quality and reliability so that the use of pertinent standards gradually becomes the norm. Generally, standards offer safe rules and values, though somewhat conservative, so that they are also updated from time to time, with respect to new information or methods.

There are many associations for standardization around the world. The names of many of them as well as the numbers and titles of numerous individual standards can be found, for example, at the website of IHS Global (<http://global.ihs.com>) and can be purchased here. They can also be found (and bought) via various standardization bodies given below. In this chapter, several institutions will be listed, whose standards for reliability are used internationally. Then, some of these standards will be named as examples.

The technical commission TC 56 “Dependability” of this organization for standardization prepares the international standards related to reliability. An overview of all valid IEC standards can be found via the above web site.

2. ISO – International Commission for Standardisation (<http://www.iso.org>)

This commission prepares and approves the standards for quality and many other subjects.

Reliability and quality standards can also be found under BS (British Standards), DIN (Deutsches Institut für Normung), ASME (American Society of Mechanical Engineers), NBS (National Bureau of Standards, USA), JSA (Japanese Standards Association), KSA (Korean Standards Association), GOST (Gosudarstvennyye Standarty – Russian state standards), and various others. Standards for civil engineering constructions can be found, for example, under EN (Euronorms, i.e. European standards or Eurocodes; see <http://www.eurocodes-online.com>).

Standards for military applications, but not only for them, can be found via <http://www.dstan.mod.uk> (United Kingdom defence standards), <http://nso.nato.int/nso> (NATO Standardization Office), and <http://www.defense.gov> (DoD or U.S. Department of Defense) or via websites such as <http://quicksearch.dla.mil>, <https://assist.dla.mil>, or <http://everyspec.com>. The access to some of them is open only for authorized people.

A selection of some international standards related to reliability follows.

IEC/ISO 31010. Risk management. Risk assessment techniques.

IEC 60050-191. International electrotechnical vocabulary. Part 191: Dependability and quality of service.

IEC 60050-192. International electrotechnical vocabulary. Part 192: Dependability.

IEC 60300-1. Dependability management. Part 1: Guidance for management and application.

IEC 60300-3-1. Dependability management. Part 3-1: Application guide: Analysis techniques for dependability – Guide on methodology.

IEC 60300-3-2. Dependability management. Part 3-2: Application guide: Collection of dependability data from the field.

IEC 60300-3-3. Dependability management. Part 3-3: Application guide: Life cycle costing.

IEC 60300-3-4. Dependability management. Part 3-4: Application guide: Guide to the specification of dependability requirements.

IEC 60300-3-5. Dependability management. Part 3-5: Application guide: Reliability test conditions and statistical test principles.

IEC 60300-3-6. Dependability management. Part 3: Application guide Section 6: Software aspects of dependability.

IEC 60300-3-10. Dependability management. Part 3-10: Application guide: Maintainability.

IEC 60300-3-11. Dependability management. Part 3-11: Application guide: Reliability centred maintenance.

IEC 60300-3-12. Dependability management. Part 3-12: Application guide: Integrated logistic support.

IEC 60300-3-14. Dependability management. Part 3-14: Application guide: Maintenance and maintenance support.

IEC 60300-3-15. Dependability management. Part 3-15: Application guide: Engineering of system dependability.

IEC 60300-3-16. Dependability management. Part 3-16: Application guide: Guidelines for specification of maintenance support services.

IEC 60319. Presentation and specification of reliability data for electronic components.

IEC 60605-2. Equipment reliability testing. Part 2: Design test cycles.

IEC 60605-4. Equipment reliability testing. Part 4: Statistical procedures for exponential distribution: Point estimates, confidence interval, prediction intervals, and tolerance intervals.

IEC 60605-6. Equipment reliability testing. Part 6: Tests for the validity and estimation of the constant failure rate and constant failure intensity.

IEC 60706-2. Maintainability of equipment. Part 2: Maintainability requirements and studies during the design and development phase.

IEC 60706-3. Maintainability of equipment. Part 3: Verification and collection, analysis, and presentation of data.

IEC 60706-5. Maintainability of equipment. Part 5: Testability and diagnostic testing.

IEC 60812. Analysis techniques for system reliability. Procedure for failure mode and effects analysis (FMEA).

IEC 61014. Programmes for reliability growth.

IEC 61025. Fault tree analysis (FTA).

IEC 61070. Compliance test procedures for steady-state availability.

IEC 61078. Analysis techniques for dependability. Reliability block diagram and Boolean methods.

IEC 61123. Reliability testing. Compliance test plans for success ratio.

IEC 61124. Reliability testing. Compliance tests for constant failure rate and constant failure intensity.

IEC 61160. Design review.

IEC 61163-1. Reliability stress screening. Part 1: Repairable assemblies manufactured in lots.

IEC 61163-2. Reliability stress screening. Part 2: Electronic components.

IEC 61164. Reliability growth — Statistical test and estimation methods.

IEC 61165. Application of Markov techniques.

IEC 61649. Weibull analysis.

IEC 61650. Reliability data analysis techniques: Procedures for comparison of two constant failure rates and two constant failure (event) intensities.

IEC 61703. Mathematical expressions for reliability, availability, maintainability, and maintenance support terms.

IEC 61709. Electric components: Reliability. Reference conditions for failure rates and stress models for conversion.

IEC 61710. Power law model: Goodness-of-fit tests and estimation methods.

IEC 61882. Hazard and operability studies (HAZOP studies): Application guide.

IEC 62198. Managing risk in projects: Application guidelines.

ISO 61713. Software dependability through the software life-cycle processes: Application guide.

ISO 9000. Quality management.

This is a group of standards that addresses various aspects of quality management. It provides guidance and tools for companies and organizations that want to ensure that their products and services consistently meet the customer's requirements and that quality is permanently improved.

The updated version ISO 9001:2008 – Quality Management Systems has several parts: Principles, Vocabulary, Requirements, etc. This standard sets out the criteria for a quality management system and is the only standard in this family that can be certified to. It can be used by any organization regardless of its field of activity. ISO 9001:2008 is implemented by more than 1 million companies and organizations in more than 170 countries.

There are also branch-related groups of standards based on internationally recognized standard ISO 9001. For example, standards AS/EN 9100, AS/EN 9110, and AS/EN 9120 are devoted to quality management systems with specific requirements on aviation, space, and defense industries. They are published by the International Aerospace Quality Group (IAQG). Especially, the AS/EN 9120 focuses on product safety and reliability and addresses critical product performance, conformity to specifications, and airworthiness.

ISO 2394. General principles on reliability for structures.

ISO 12491. Statistical methods for quality control of building materials and components.

ISO 13822. Bases for design of structures: Assessment of existing structures.

ISO/IEC Guide 51. Safety aspects: Guidelines for their inclusion in standards.

ISO-IEC Guide 73. Risk management: Vocabulary.

CAN/CSA – Q634-91. Risk Analysis Requirements and Guidelines.

EN 1990. Eurocode: Basis of structural design.

EN 50126. Railway applications. The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS). This is a group of the standards of the CENELEC (European Committee for Electrotechnical Standardization) for the reliability and safety in rail

industry in Europe and other countries. It has several parts: EN 50126-1: Part 1: Generic RAMS process, EN 50126-2: Part 2: Systems approach to safety, EN 50126-3: Guide to application of EN 50126-1 for rolling stock RAM, EN 50126-4: Functional safety: Electrical/electronic/programmable electronic systems, and EN 50126-5: Functional safety: Software.

The standards for reliability and quality in the railway industry can also be found under the International Railway Industry Standard (IRIS). Those for road vehicles industry can be found, for example, under SAE (Society of Automotive Engineers) or VDI (Verein Deutscher Ingenieure).

As said above, the standards are updated from time to time, and some can even lose its validity, be withdrawn, or replaced by others. Thus, when looking for a certain standard, one must make sure whether it is valid or if it has undergone some changes.

3. MIL-HDBK and MIL-STD

Various volumes of Military Handbook (MIL-HDBK) or Military Standards (MIL-STD), issued by the U.S. Department of Defense (DoD), deal with reliability (among other things) and are worldwide recognized and become international standards of their own. Some titles follow. However, the reader must always check whether the pertinent issue is valid. Nevertheless, even “cancelled” issues can contain useful information.

MIL-HDBK-189. Military Handbook: Reliability growth management.

MIL-HDBK-217. Military Handbook: Reliability prediction of electronic equipment. It contains failure rate models for numerous components, usually more conservative than in other standards.

MIL-HDBK-338B. Military Handbook: Electronic reliability design handbook.

MIL-HDBK-344. Military Handbook: Environmental stress screening of electronic equipment.

MIL-HDBK-472. Military Standardization Handbook: Maintainability prediction.

MIL-STD-105E. Military Standard: Sampling procedures and tables for inspection by attributes. This standard was officially cancelled in 1995. A similar topic is dealt with in MIL-STD-1916: DoD preferred methods for acceptance of product.

MIL-HDBK-781D. Military Handbook: Reliability test methods, plans, and environments for engineering development, qualification, and production.

MIL-STD-785B. Military Standard: Reliability program for systems and equipment development and production.

MIL-STD-1629A. Military Standard: Procedures for performing a failure mode, effects, and criticality analysis.

MIL-STD-2074. Military Standard: Failure classification for reliability testing.

MIL-STD-2155. Military Standard: Failure reporting, analysis, and corrective action system (FRACAS).]

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

Software for Reliability Analysis

Jaroslav Menčík

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62367>

This short overview gives only some of the commercial programs for reliability assessment. They range from simple ones at moderate prices and suitable for a limited range of problems to program systems for the analysis of complex problems. For getting practice and for the solution of simple problems, the reader can create own programs based on universal software such as Excel, although their possibilities are limited.

VaP – Variables Processor (<http://www.petschacher.at>)

A simple Monte Carlo simulation program VaP enables the probabilistic analysis of a user-defined function $G(x)$ depending on one or more random input variables. Several types of probability distributions can be used. VaP calculates the expected value, standard deviation, skewness, and curtosis of G . It shows the histogram of the function G and calculates the probability that G gets less than zero. (The program was originally developed for the students of civil engineering at ETH, with G denoting reliability margin, cf. Chapter 14, and can also use the First Order Reliability Method FORM.) The main results are saved and printed. Favorable price-performance ratio.

Anthill(<http://www.sbra-anthill.com>)

Anthill is a computer program for the calculation of reliability and other statistical analyses based on the Monte Carlo method. The user-defined model for analysis can use various mathematical and logical functions and predefined histograms. After the trials with random values of input quantities have been performed, a statistical analysis of sampled values is performed, and the results are displayed. The resulting histograms and statistical parameters can be saved for further postprocessing. Favorable price-performance ratio.

Feasible Reliability Engineering Tool (FReET) (<http://www.freet.cz>)

FReET is a multipurpose probabilistic software for statistical, sensitivity, and reliability analysis of engineering problems, developed at the Brno University of Technology, Institute

of Structural Mechanics. It allows the simulation of random uncertainties in various problems, especially in civil and mechanical engineering (material properties, loading and geometrical imperfections). It uses the following probabilistic techniques: crude Monte Carlo simulation, Latin hypercube sampling, simulated annealing, first-order reliability method (FORM), and others. Favorable price-performance ratio.

Strurel (<http://www.strurel.com>)

Strurel is a set of programs for the reliability analysis of constructions (especially in civil engineering). It consists of three programs: Statrel (reliability-oriented statistical analysis), Comrel (time-invariant and time-variant analysis), and Sysrel (a program for system reliability analysis). It can work with analytical functions and performs reliability analysis using various methods, such as Monte Carlo or FORM and SORM (first-order or second-order reliability methods), used for the solution of the problems of load-resistance type.

In addition to the above software, oriented mostly on the determination of failure probability of one component or construction, also software systems exist, which use various tools and enable comprehensive reliability analysis of very complex objects, such as aircrafts or weapon systems. Here, four brands will be mentioned.

Item Software (<http://www.itemsoft.com>)

This software firm (USA) offers various products, such as ITEM ToolKit. This is a suite of several analytical and reliability prediction modules in one integrated environment, such as

MIL-HDBK-217 module. A reliability prediction program based on the internationally recognized method of calculating electronic equipment reliability defined in military handbook MIL-HDBK-217 (published by the U.S. Department of Defense).

IEC 62380 Electronic Reliability Prediction module. It supports reliability prediction methods based on the latest European reliability prediction standard IEC.

NSWC Mechanical Reliability Prediction module. It uses a series of models for various types of mechanical devices including actuators, springs, bearings, seals, electric motors, pumps, compressors, brakes, and clutches to predict failure rates based on temperature, stresses, flow rates, and various other parameters. The module is based on the Naval Surface Warfare Center Handbook of Reliability Prediction Procedures for Mechanical Equipment.

China 299B Electronic Reliability Prediction module: A reliability prediction program based on the internationally recognized method of calculating electronic equipment reliability provided in the Chinese Military/Commercial Standard GJB/z 299B.

Telcordia Electronic Reliability Prediction module: Based on the Telcordia (Bellcore) TR-332 and SR-332 standards, calculates the reliability (steady-state failure rate) for various categories of electronic, electrical, and electromechanical components for various quality levels, environmental conditions, electrical stress conditions, and other parameters.

In addition to these modules, the ITEM ToolKit contains several other modules, for example for failure modes, effects, and criticality analysis (FMECA); for fault tree analysis (FTA); for

construction of reliability block diagrams (RBD); for Markov analysis; and for maintenance and others.

ReliaSoft (<http://www.reliasoft.com>)

This U.S. software firm offers a group of programs in one integrated environment, such as Weibull $\pm\pm$ [®] for reliability analysis; ALTA[®] for Accelerated Life Testing Data Analysis; DOE \pm [±] for design of experiments; BlockSim[®] for the creation of reliability block diagrams based on fault tree analysis; RENO[®] simulation software for risk and decision analysis; Xfmea[®] – software for facilitating data management and reporting for all types of FMEA and FMECA; RCM $\pm\pm$ [®] – software for the support of reliability-centered maintenance; Lambda Predict[®] – for reliability assessment based on standards; RBI[®] – risk-based inspection analysis for oil, gas, chemical, and power plants in adherence to the guidelines presented in the American Petroleum Institute’s publications API RP 580 and RP 581, as well as the American Society of Mechanical Engineers ASME; XFRACAS[®] – software system for web-based incident/failure/data reporting, analysis, and corrective action; and RGA[®] – software for the analysis and support of reliability growth.

Isograph (<http://www.isograph.com>)

Isograph offers various software for reliability analysis, such as:

Availability Workbench: A system for availability simulation and reliability-centered maintenance RCM. It is used to optimize maintenance and spare parts, predict system availability and throughput, and estimate life-cycle costs. It includes Weibull analysis and life cycle costing modules as well as modeling methods such as FMECA, reliability block diagram analysis, and fault tree analysis.

Reliability Workbench: An integrated visual environment in which failure rate and maintainability prediction, FMECA, reliability allocation, reliability block diagram, fault tree, event tree, and Markov analysis are combined. Failure rate predictions are calculated from the Telcordia, MIL-HDBK-217, 217 Plus, and IEC TR 62380 standards for electronic equipment and the NSWC-98/LE1 Handbook for mechanical parts. FMECA, reliability block diagram, and fault tree analysis are performed to well-known standards such as MIL-STD 1629 and IEC 61508.

Hazop \pm : software for hazard and operability studies, with visual environment using the forms for entering Hazop information. Extensive reporting facilities are available.

More information, including other products, is available at the web site.

PTC Windchill, formerly **Relx Software** (<http://www.ptc.com/product/windchill/quality>)

PTC Windchill Quality Solutions combines quality, reliability, and risk management into an integrated toolset with the following products:

PTC Windchill CAPA: corrects and prevents actions and demonstrates compliance.

PTC Windchill Nonconformance: manages, corrects, and tracks internal quality issues.

PTC Windchill FRACAS: identifies and prioritizes failure-related trends.

PTC Windchill FMEA: identifies and mitigates potential failures.

PTC Windchill MSG-3: manages aircraft reliability according to industry standards.

PTC Windchill FTA: utilizes fault tree analysis to investigate safety and reliability issues.

PTC Windchill Prediction: predicts failure rate of components and system reliability.

PTC Windchill RBD: Reliability Block Diagrams manage quality in complex systems.

PTC Windchill Maintainability: predicts maintenance and repairs.

PTC Windchill LCC: Life Cycle Cost software analyzes the lifetime cost of a product.

PTC Windchill Weibull: Life data analysis or Weibull estimates on the life data of a product.

PTC Windchill ALT : Accelerated Life Testing software predicts product reliability.

PTC Windchill Markov: Visual analysis software that models complex systems.

PTC Windchill Customer Experience Management: reports and manages quality-related field issues.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic

Authored by Jaroslav Mencik

Our life is strongly influenced by the reliability of the things we use, as well as of processes and services. Failures cause losses in the industry and society. Methods for reliability assessment and optimization are thus very important. This book explains the fundamental concepts and tools. It is divided into two parts. Chapters 1 to 10 explain the basic terms and methods for the determination of reliability characteristics, which create the base for any reliability evaluation. In the second part (Chapters 11 to 23) advanced methods are explained, such as Failure Modes and Effects Analysis and Fault Tree Analysis, Load-Resistance interference method, the Monte Carlo simulation technique, cost-based reliability optimization, reliability testing, and methods based on Bayesian approach or fuzzy logic for processing of vague information. The book is written in a readable way and practical examples help to understand the topics. It is complemented with references and a list of standards, software and sources of information on reliability.

Photo by agsandrew / AdobeStock

IntechOpen

ISBN 978-953-51-6653-5



9 789535 166535