IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

A comparison of the expected and statistical probability distribution of system failures

To cite this article: A S Dulesov *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **537** 062027

IOP Conf. Series: Materials Science and Engineering **537** (2019) 062027 doi:10.1088/1757-899X/537/6/062027

A comparison of the expected and statistical probability distribution of system failures

A S Dulesov[1], D J Karandeev[1], R I Bazhenov[2], T G Krasnova[1], N V Dulesova[3]

[1] Katanov Khakass State University, 92, Lenina ave., Abakan, 655017, Russia
[2] Sholom-Aleichem Priamursky State University, 70A Shirokaya street, Birobidzhan, 679015, Russia
[3] Khakas Technical Institute of Siberian Federal University, 15, Komarova ave., Abakan, 655017, Russia

E-mail: den_dr_house_1991@mail.ru

Abstract. The possibility of applying the information theory in the problem of comparing the expected and statistical probability distribution of failures of a technical system are considered. The paper presents a brief analysis of the processes of additive and multiplicative growth of the system indicators, among which the probability of failure-free operation and failure rate were considered. These indicators were considered in order to analyze the reliability of the system. The increase in reliability of the indicators is associated with the fixing of the failure rate of the system elements and the construction of probability distributions. In order to compare the two distributions, a method for measuring uncertainty is proposed, which includes Shannon's measure of uncertainty, cross-entropy and Kullback-Leibler divergence. Together, they make it possible to determine the connection between the two different probability distributions of failures, to calculate the distance between the distributions, to identify the degree of difference between the real and desired state of the system during operation. An example of calculation confirming the importance of the participation of the offered method for measuring uncertainty in the problem of comparison of the expected and statistical probability distribution of system failures is given.

1. Introduction

When considering compliance with the high level of reliability of complex technical systems, the analysis of indicators and statistics obtained during testing or operation is not excluded from consideration. The statistical data obtained in the process of testing and operation of the system elements make it possible to track the dynamics and statics of changes in the quantitative characteristics of reliability [1]. In most cases, when testing or operating elements or equipment, the time of operation before until a complete failure occurs is considered to be an important characteristic. At the same time, engineering practice takes into account the characteristics that are in demand for determining and subsequent application of probabilistic indicators [2, 3].

Since we have to deal with failures, the nature of which is random, reliability calculations are usually based on probabilistic estimates of the state of the system [3, 4]. At the same time, it is possible to use the information theory because of a probabilistic measure p_i [5, 6] In all cases, p_i – the relative number of discrete states of the system, that is, failures related to each of the *i*-th elements. According to information theory [5, 7], signals (in fact are failures) can be considered as discrete states. They are

IOP Conf. Series: Materials Science and Engineering **537** (2019) 062027 doi:10.1088/1757-899X/537/6/062027

distributed according to the frequency of their occurrence during testing and operation of equipment. Such definition of probability spaces allows to use the model of K. Shannon [8] for determining the entropy of the system state. The meaning and values of the entropy are entirely determined by the choice of the state space and its characteristics.

Application of the Shannon's formula allows to determine entropy for both the natural and statistical probability distribution of system failures [9]. The natural distribution is essentially "true", (that is, desired) related to the known laws of the distribution of a random variable and for which the equipment availability index is determined. As for the statistical distribution, it can be built by monitoring the state of the equipment for a certain period of time. Moreover, it is no longer possible to guarantee the fact of complete coincidence of the obtained statistical distribution with the natural distribution of the probability of failures due to the impact on the system of hard-to-predict factors of probabilistic nature. In this case, it is important to have information on the divergence between the distributions in order, during operation of the equipment, to assess the similarity of the quantitative characteristics of reliability with the expected ones.

2. Nature of the failure distribution

Next, one explain the simplified nature of the failure distribution.

1. Let over time to the existing failures of the element will be added to one event/failure (stationary process). In this case, the failure statistics will be considered for each of the same type of elements separately. Over time, during operation or testing, the difference in the number of failures between them will remain little discernible, indicating approximately the same levels of reliability of the elements. Therefore, the absolute difference between the number of failures will remain approximately constant, and the relative difference will tend to zero. If we consider the number of failures of the element as a share in the total number of failures of all considered elements, this share will tend to $1/n$, where n is the number of similar elements. For the entire set of k events, the probability of occurrence of one event $p=1/k$. It becomes obvious: we have a simple process of quantitative growth, since on average one more is added to the already existing events. The simplest process of growth of events testifies to independence of events of each of elements as the assumption is accepted: occurrence of events doesn't depend on the prevailing external and internal circumstances. Thus, we have additive growth associated with the addition operation. This growth indicates an increase in the number of events in the system, for which the difference in the number of failures between the same type of independent elements over time approaches zero, and their shares between the considered elements are equalized. Consequently, elements of the same type (due to the independence of the occurrence of events) will have approximately the same levels of their own reliability, which simplifies the calculation of reliability.

2. Let's consider the nature of events when not one but several failures are added to the already existing failures, the number of which is determined each time by multiplying the existing failures by a certain number. This nature of the event flow involves the influence of external and internal factors on the state of the element. In this case, the number of failures of each of the elements will grow unevenly. Then at each considered time interval the relative difference between the events will remain unchanged, and the absolute difference will increase. In this case, the connection between the elements are not taken into account, and events are considered as independent. For this case, we propose the following example. Let the occurrence of failure for any element the more likely it is, the higher its share in the total number of failures and each time their number increases. So the probability of the occurrence of an event j for the element i having n_j failures will be equal to: $p_j = n_j/(n_1 + n_2 + n_3 +\ldots+ n_k)$. It can be seen from the expression: with increasing n_j probability increases, that is, the share of p_j in the total mass of probabilities (equal to 1) will increase. However, despite occasional fluctuations, the overall shape of the distribution of failures for all elements remains the same - objects of the set proportionally increased. In this case, there is a multiplicative growth associated with the multiplication operation. For multiplicative growth, there is no absolute difference between element failures. On the other hand, the relative difference between element failures remains (the share of failures of each element) since growth is inherent in all elements of the system.

Let's summarize the above**.** Considering additive growth, addition indicates the existence of the principle of growth and as a consequence of the independence of events between the elements of the system. On the contrary, the multiplication operation relating to multiplicative growth confirms the existence of events due to the occurrence of external and internal factors. The analysis of multiplicative growth in the system can be useful in identifying weak connection in its structure. However, one should not forget that the above about additive and multiplicative growth refers to ideal cases.

3. Measure of information uncertainty in estimating the expected and statistical distribution of failure rates

Multiplicative growth can be seen on the example of the frequency of occurrence of events. We will consider the growth through a quantitative characteristic: $\lambda_i(t)$ is the failure rate of the element *i* in the time interval *t*. In the selected interval, the intensity is considered as a mathematical expectation of a random variable. Considering the intensity as a statistical series, this frequency characteristic of the system reliability allows to determine the probability of occurrence of events for each of the elements:

$$p_i(\lambda) = \frac{\lambda_i}{\Lambda}, \tag{1}$$

where Λ – the sum of frequencies of all system elements.

This value is not difficult to determine during the operation of the system, since the distribution of failures is statistical. As for the expected or desired failure distribution, it can be derived from the data obtained during the failure tests for new equipment, the preliminary monitoring of failures at the initial stage of operation and expert assessments. The desired distribution can be considered a postulated a priori distribution $p_i(\lambda)$, and the statistical distribution $q_i(\lambda)$ (obtained from operating experience) can be considered verifiable. Comparison of these two distributions between themselves is possible through a measure of information uncertainty [10, 11].

Entropy [12, 13] is taken as a measure of uncertainty, which in the case of considering the multiplicative growth is calculated by the Shannon's formula [14, 15]:

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \text{ bit, by } \sum_{i=1}^{n} p_i = 1, \tag{2}$$

where n – the number of elements in the system.

Formula (2) is valid for determining the entropy of various kinds of distributions. The Shannon's entropy will increase as the density of the distribution decreases p_i ($p_i \to 1$, $H \to 0$) [16, 17]. However, Shannon's entropy doesn't take into account the situation when in the considered time interval there are no equipment failures in the statistics, that is, $\lambda_i = 0$, during the operation of the system. Then $p_i = 0$, and entropy $H \to \infty$. It is possible that the events that disrupt the operation of the equipment, didn't lead him to a state of complete failure. In this case, λ_i can be determined by averaging: $\lambda_i = (\lambda_{i-\Delta}+\lambda_{i+\Delta})/2$, where $\lambda_{i-\Delta}$ and $\lambda_{i+\Delta}$ are, respectively, the failure rates to the left and right of the interval *t*.

When comparing these distributions, it is impossible to say unequivocally that in the future the probability of occurrence of q_i events will coincide with the probability p_i. The statistical distribution $q_i(\lambda)$ obtained from (1) is verifiable and serves as an approximation of the distribution $p_i(\lambda)$.Turning to entropy as a measure of information uncertainty, it is possible to calculate an approximation that reflects the amount of loss (unaccounted for amount) of information when moving from the expected distribution $p_i(\lambda)$ to statistical $q_i(\lambda)$, due to the Nature and capabilities of the system control. The measure of information considered in this case is called cross-entropy [6].

For two distributions of $p_i(\lambda)$ and $q_i(\lambda)$ and discrete values of *p* and *q*, the cross-entropy is determined as follows:

IOP Conf. Series: Materials Science and Engineering **537** (2019) 062027 doi:10.1088/1757-899X/537/6/062027

$$H_p(q) = -\sum_{i=1}^{n} p_i(\lambda)\log_2 q_i(\lambda), \text{ bit.} \quad (3)$$

If we consider the additive growth process [18] as the occurrence of a single event from the whole set of k events in the system, the probability of occurrence of the single event p=1/k. Since all probabilities are equal, it is possible to speak about equality of all values λ of the expected distribution $p_i(\lambda)$. Distributions of this kind are typical for a steady flow of failures. Then cross-entropy according to (3) is determined as:

$$H(q) = -\frac{1}{k}\sum_{i=1}^{n} \log_2 q_i(\lambda), \text{ bit.} \quad (4)$$

There is a difference between entropy (2) and cross-entropy (3), which is called Kullback-Leibler divergence [19] – the divergence of the distribution q with respect to p:

$$D_p(q) = H_p(q) - H(p). \quad (5)$$

When substituting expressions (3) and (2) into (5) by performing simple mathematical transformations, the divergence or Kullback-Leibler divergence has the form:

$$D_p(q) = \sum_{i=1}^{n} p_i(\lambda)\log_2 \frac{p_i(\lambda)}{q_i(\lambda)}, \text{ bit.} \quad (6)$$

Kullback-Leibler divergence is the distance between the two distributions. It shows how different the distributions of random variables are [20]. It should be borne in mind that the functional (6) is not a metric in the space of distributions, since it does not satisfy the axiom of symmetry: $D_p(q) \neq D_q(p)$.

4. Example of determining the amount of information

Let there be a technical system consisting of 5 elements, n=5. Based on the incompatibility of events in the distributions $p_i(\lambda)$ and $q_i(\lambda)$, the elements failures will be considered as random discrete values p and q. It is necessary to determine the cross-entropy of the system and the Kullback-Leibler divergence.

There is a number of data (obtained during testing and operation of the equipment), including the failure rate λ_i distributed over the selected time interval t (year) (table. 1).

Table 1. Data on failures during testing and failure rate during operation.

№ element	1	2	3	4	5
Failure rate during testing, λ_i	0.1	0.3	0.35	0.25	0.15
Failure rate during operation, λ_i	0.15	0.31	0.33	0.26	0.18

Table 2. Probabilities (according to expression (1)).

№ element	1	2	3	4	5
Probability during testing, p_i	0.087	0.261	0.304	0.217	0.130
Probability during operation, q_i	0.122	0.252	0.268	0.211	0.146

Table 3. Cross-entropy of system elements (according to (3)) and Kullback-Leibler divergence (according to (6)).

№ element	1	2	3	4	5
Cross-entropy, bit	0.264	0.519	0.578	0.487	0.362
Kullback-Leibler divergence, bit	- 0.042	0.013	0.055	0.009	- 0.022

Cross-entropy of the system – $H_p(q)$ = 2.209 bit.

Kullback-Leibler divergence – $D_p(q)$ = 0.013 bit.

To summarize, we note the following. We have two probability distributions, one of them is considered true, which should be confirmed during the operation of the technical system. The second distribution is obtained as a result of the operation of the system. If the two distributions are exactly the same, $D_p(q)$ must be zero.

Considering the results of divergence, one can control confidence in the comparison of distributions. If the differences are large, it will not take long to determine the desired distribution, since the influence of factors on the change in the system state is clearly visible. However, if their differences are insignificant, as is the case in our example, when the system state slightly deviates from the expected or predicted, you will have to look through a lot of data in search of insignificant factors.

5. Conclusion

As a result, it can be stated that information theory is also manifested in the field of reliability analysis of technical systems. This theory offers a concrete, formal description of many things, such as comparing the probabilities of the desired and statistical distributions. Having methods for measuring uncertainty and two sets of distributions, we can understand the following:

- what is the connection between two different probability distributions;
- what is the distance between the probability distributions;
- how much the desired state of the system differs from the predicted one;
- how closely two variables are dependent.

The ideas of using models from information theory are understandable, because they have good properties and fundamental origin.

Acknowledgments

The research was carried out with the financial support Program of the Russian Foundation for Assistance to Small Innovative Enterprises in Science and Technology by the grant "UMNIK" №13138GU/2018, code № 0040353.

References

[1] O'Connor P D T 2002 *Practical Reliability Engineering* (John Wiley & Sons)

[2] Ebeling C E 1997 *An Introduction to Reliability and Maintainability Engineering* (Boston: McGraw-Hill Companies)

[3] Ross Sheldon M 2009 Introduction to probability and statistics for engineers and scientists *Associated Press* p 267

[4] Lienig J and Bruemmer H 2017 Reliability Analysis *Fundamentals of Electronic Systems Design* pp 45–73 doi:10.1007/978-3-319-55840-0_4

[5] Cover T M and Joy T A 2006 *Elements of Information Theory* (New Jersey: Wiley and Sons)

[6] MacKay D J C 2003 *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press)

[7] Kolmogorov A N 1965 Three approaches to the quantitative definition of information

International Journal of Computer Mathematics **1** 3-11
[8] Shannon C E 1948 Mathematical Theory of Communication *Bell System Tech. J.* **27** 379-423
[9] Lawless J F and Fredette M 2005 Frequentist predictions intervals and predictive distributions *Biometrika* **92(3)** 529–42
[10] Dulesov A S, Karandeev D J and Dulesova N V 2017 Optimal redundancy of radial distribution networks by criteria of reliability and information uncertainty *IEEE 3nd International Conference on Industrial Engineering, Applications and Manufacturing* (*ICIEAM*) pp 1-4 doi: 10.1109/ICIEAM.2017.8076467
[11] Dulesov A S, Karandeev D J and Dulesova N V 2017 *IOP Conf. Ser.: Earth Environ. Sci.* **87** 032008
[12] Dulesov A S, Karandeev D J and Dulesova N V 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **450** 072004
[13] Dulesov A S, Karandeev D J and Krasnova T G 2017 The evaluation of the correlation between entropy and negentropy in the structure of a technical system *MATEC Web Conf* **129** pp 1-4
[14] Shannon C E 1949 Communication Theory of Secrecy Systems *Bell System Tech. J.* **28** 656-715
[15] Hartley R V L 1928 Transmission of Information *Bell System Tech. J.* **7** 535–63
[16] Dulesov A S, Eremeeva O S, Karandeev D J and Dulesova N V 2018 Approaches to Information Measurement of the Structure State of Technical Systems *IEEE FarEastCon* pp 1-6
[17] Dulesov A S, Karandeev D J and Dulesova N V 2018 Improving the operation quality of technical systems using information theory models *MATEC Web Conf.* **224**
[18] Dulesov A S, Eremeeva O S, Karandeev D J and Krasnova T G 2018 Analytical notes on growth of economic indicators of the enterprise *Advances in Economics, Business and Management Research* **47** pp 327-32
[19] Kullback S and Leibler R A 1951 On information and sufficiency *The Annals of Mathematical Statistics* **22(1)** pp 79-86
[20] Kullback S 1987 Letter to the Editor: The Kullback–Leibler distance *The American Statistician*, **41 (4)** pp 340–1