Method for overcoming the heteroscedasticity of statistical values of indicators when assessing the quality of IETMs with elements of artificial intelligence

Yan A. Ivakin [1,2,3], Maria S. Smirnova [1] and Elena A. Frolova [1]

[1] *Saint-Petersburg State University of Aerospace Instrumentation, Bolshaya Morskaia str. 67, A, St. Petersburg, 190000, Russian Federation*
[2] *Saint-Petersburg Federal Research Center of the Russian Academy of Sciences, 14th line V.O., 39, St. Petersburg, 199178, Russian Federation*
[3] *Concern OCEANPRIBOR JSC, Chkalovsky prospect, 46, St. Petersburg, 198226, Russian Federation*

Abstract

In the last decade, mobile interactive electronic technical manuals - IETM - have become a modern means of competence support for personnel operating aircraft. The system of technical regulation distinguishes several classes of IETM according to the degree of their functional equipment. The highest classes of functional development of IETM presuppose their deep integration into on-board automation systems, the possibility of direct interface interaction with electronic diagnostic modules for accompanying products and wide integration of artificial intelligence tools. In turn, the inclusion of elements of artificial intelligence leads to a change in statistical approaches and principles for assessing the quality of the IETM themselves. This is due to the fact of continuous change in the consumer properties of IETM in the process of their use, the observed heteroscedasticity of the recorded values of indicators when assessing their quality. This article is devoted to the description of the methodology that allows to overcome the described specifics of the procedures for assessing the quality of IETM with elements of artificial intelligence.

Keywords
Quality assessment, interactive electronic technical manuals, deep neural networks

1. Introduction

In accordance with state standards [1-4], there are several classes of interactive electronic technical manuals (IETM), each of which is characterized by a certain level of development of functionality and software adaptability during implementation. The highest classes of functional development of IETM presuppose their deep integration into on-board automation systems, the possibility of direct interface interaction with electronic diagnostic modules for accompanying products and wide integration of artificial intelligence tools. Modern technologies of artificial intelligence are practically fully implemented on the basis of the so-called. deep neural network software (hardware and software) solutions. Such software, intelligent solutions are characterized by their constant properties of information plasticity, adaptability and self-adjustment. In turn, the inclusion of elements of artificial intelligence leads to a change in statistical approaches and principles for assessing the quality of the IETM themselves. This is due to the fact of continuous change in the consumer properties of IETM in the process of their use, the observed heteroscedasticity of the recorded values of indicators when assessing their quality. The heteroscedasticity of statistical values of indicators when assessing the quality of IETM with elements of artificial intelligence based on deep neural networks leads to

inefficiency of the traditional apparatus of statistical analysis, traditionally used in research of technical systems. Variable (changing) values of the main parameters of the estimated values of a random nature, such as variance, standard deviation, etc., require the use of statistical approaches that are more characteristic of socio-political and information-psychological modeling.

As a result of a complex of studies, a study of a number of scientific results described in [5-14], the authors propose a method for obtaining statistical values of qualimetric indicators with a probabilistic measure applicable to assess the quality of IETM with elements of artificial intelligence based on deep neural networks.

2. The essence of the problem of heteroskedasticity of statistical values of indicators in the estimation of IETM

In the course of research [11-14], as well as during statistical processing of the results of studies of the applicability of neural network technologies in IETM, specific features of the application of the statistical apparatus of experimental research in the subject area of the corresponding deep neural networks (DNN) were revealed. In particular, it was found that the use of a classical for statistics set of samples of single research trials (i.e., training, test and control samples) for training DNN and experimentation leads to the fact that the properties of the neural network itself in IETM constantly changing. From the point of view of the statistical foundations of experimental studies to assess the technical and functional characteristics of intelligent IETM, this means that with the statistical accumulation of the results of single tests, the conditions of the classical model of a statistical experiment are not fully satisfied, i.e. are violated. These violations relate to the premise of uncorrelated disturbances and the absence of constancy in the value of the variance of disturbances of the observed random variables. Thus, neural networks in IETM as an object of a statistical experiment, due to their constant customizability and logical plasticity, violate the fundamental condition for the application of the standard mathematical apparatus of a computational experiment: the conditions of homoscedasticity are the conditions for the constancy of the variances of the random component. Failure to meet this condition is called heteroscedasticity (i.e., variance of variance of deviations). The essence of the heteroscedasticity of the statistically accumulated results of unified tests of the DNN as part of the IETM is clearly understood from the comparison of the pictograms in Figure 1:
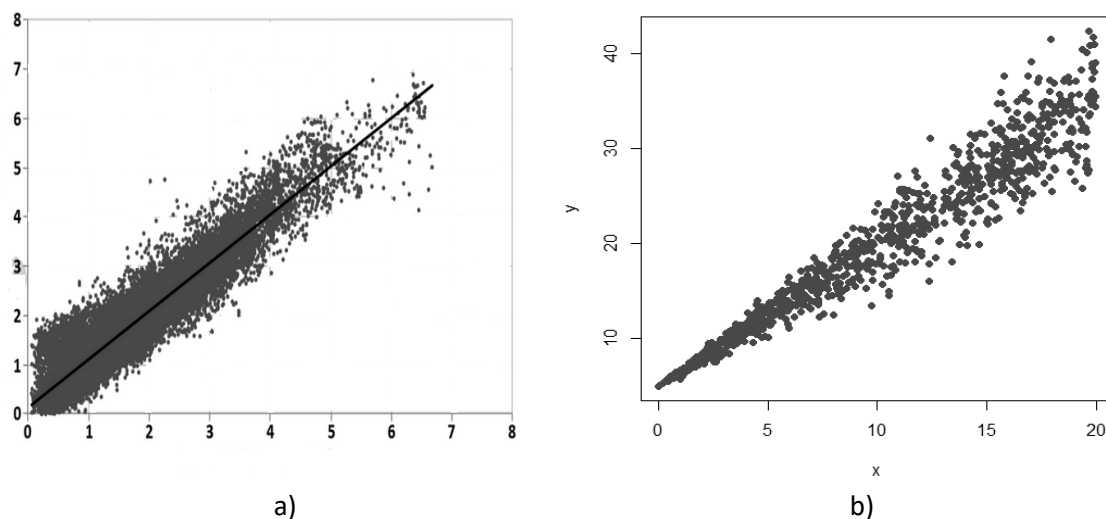

| a) | b) |

Figure 1: The essence of the heteroscedasticity of the statistics of single tests when experimenting on the DNN as part of the IETM (a) homoscedasticity of single trials; b) heteroscedasticity of single tests)

Traditionally, the problem of heteroscedasticity is mainly characteristic of samples in statistical studies related to social observations, where the objects of observation are: a person, social groups, society, etc. That is, entities that also change already in the process of experiment and / or preliminary testing. In the study of technical objects, as a rule, the samples are of a homoscedastic nature. It is due

to the constant customizability and logical-semantic plasticity of the DNN as part of the IETM that there is reason to believe that the probabilistic distributions of perturbations of the values observed in the experiment will be different for different observations.

In the case of heteroscedasticity, the estimates of the studied values are still unbiased, but the use of the mathematical apparatus for assessing the level of confidence in the results obtained has some peculiarities:

1. Estimates will not be efficient (that is, they will not have the least variance compared to other estimates for a given parameter).
2. The variances of estimates will be biased. Bias is due to the fact that the variance estimate S^2_{ost} used to calculate the variance of the estimates is no longer unbiased.
3. As a consequence of the above, all conclusions drawn from the relevant statistics, as well as interval estimates, will be unreliable. Consequently, statistical inferences from standard quality checks on assessments can be erroneous and lead to inaccurate conclusions. It is likely that the standard errors will be underestimated, and therefore the associated risks will be overestimated.

This can lead to the recognition of statistically significant values, which in fact are not.

To date, the problem of heteroscedasticity of statistically accumulated results of unified tests of DNN as part of intelligent IETM has been identified and substantiated. One of the options for its resolution is the below presented methodology, substantiated and tested in the course of relevant studies to substantiate the entire set of methods for the practical application of DNN and other information technologies of artificial intelligence as part of the IETM of the fifth class.

3. Structure of the methodology for obtaining statistical values of indicators when evaluating the quality of DNN in the composition of IETM

Evaluation of the DNN as part of the IETM is carried out by comparing the value of the mathematical expectation of the proportion of correctly recognized objects (targets) during the execution of the DNN of the current functional task, according to the statistically significant number of control samples for the accepted control dataset obtained on the experimental IETM sample with a priori specified criterion value. At the same time, an ensemble of DNN implementations is produced, which allows one to estimate the mathematical expectation of the proportion of correctly recognized objects (targets) when the DNN is executed for the corresponding functional task with a given level of statistical stability (significance), and then correlate it with the specified criterion value.

A brief description of the methodology for obtaining statistical values of indicators when assessing the quality of the gas pumping station as part of the IETM:

The DNN is trained on an experimental IETM sample on one dataset for which a statistically significant number of control samples is determined. In this case, the volume of all control samples by the number of elementary object recognition should be the same, which should ensure the unbiased nature of the final statistical estimates.

Also, the criterial value of the required execution parameter of the DNN for the adopted IETM functional task, which corresponds to the current dataset, is set a priori. Such a criterion value is a fixed, minimum value of the proportion of correctly recognized objects (targets) in the implementation of the DNN within the framework of the current (modeled) functional task.

An assessment of a statistically significant number of control samples is made to ensure the required low levels of risk and the required confidence level for the estimates of the quality of the DNN as part of the IETM. In this case, each control sample is considered as a single trial of a statistical experiment, and their total number is considered as the total number N of single trials (observations) within the framework of such an experiment. It should be borne in mind that each element of the set of the control sample in this case is considered as a one-time implementation in calculating the mathematical expectation of the proportion of correctly recognized objects and therefore the parameters of statistical stability are determined precisely by the prepared number of control samples.

It also analyzes the acceptability of the current risks of making final decisions in this computational experiment (a single experimental study) according to the indicators:

- α- risk of incorrect acceptance of the observed value, test result;
- β- risk of incorrect deviation when it is necessary to accept the observed value.

Obviously, due to the objectively high labor intensity of preparing the necessary and sufficient number of control samples with a previously justified volume of data sets of 20% of the volume of the current dataset, it is a priori rational to accept:

$$\alpha = \beta = 0.2. \tag{1}$$

Based on the provided number N, equal to the number of control samples prepared according to the current dataset, and the a priori accepted value of risks from (1), the data of the level of confidence in the obtained statistical values is calculated. The output data of such calculations will be the value of the confidence probability P_a, which is provided by the current number of control samples.

If the values of P_a, α, β corresponding to the current value of the prepared control samples N do not satisfy the external conditions of experimentation, then it is necessary to carry out the specified calculation according to the required (necessary and sufficient for external requirements for the experiment) the above parameters and select the numerical value of the number of single tests N that will provide them.

After these manipulations, it is necessary to make sure that the power of all control samples of single tests from N is the same.

A sequential multiple implementations of the trained DNN is performed using N control samples on an experimental IETM sample. At the same time, within the framework of the implementation of each control sample using an expert "teacher", the correctness of recognizing the object of the trained DNN is assessed, and with the help of software and hardware tools, the fact of success / failure of implementation is recorded for each single element of the set of the control sample. This makes it possible to estimate the mathematical expectation of the proportion of correctly recognized objects (targets) as the ratio of the number of successful tests to the total number of tests in the current control sample. Then the statistical stability of the value of the specified mathematical expectation is determined on the total volume of N control samples.

Each ensemble of implementation on the experimental IETM sample, obtained according to clause 3 of this methodology, is accumulated for each control sample and averaged to obtain the mathematical expectation (ME) of the proportion of correctly recognized objects within each of the samples, as an average (weighted by the probabilities of possible values) values of a random variable. The results of the implementation are recorded in a table, the form of which is shown in Table 1.

Table 1
Format of the table for registering the values of the mathematical expectations of the proportion of correctly recognized objects during the experiment

Control sample index	The number of single trials in the implementation of the sample	Meaning of expected value
1	Const.	$ME_1[X]$
2	Const.	$ME_2[X]$
3	Const.	$ME_3[X]$
4	Const.	$ME_4[X]$
...	Const.	...
$N-2$	Const.	$ME_{N-2}[X]$
$N-1$	Const.	$ME_{N-1}[X]$
N	Const.	$ME_N[X]$

The values of the mathematical expectation obtained for each of the control samples are averaged over the number N, due to which the sample average value of the ME of the proportion of correctly recognized objects in the entire ensemble of experiment realizations is obtained $\overline{ME}[X]_N$.

On the same ensemble of realizations, the sample standard deviation σ is calculated for $\overline{ME}[X]_N$, as a measure of the spread of values of a random variable relative to its mathematical expectation, according to the relation [8]:

$$\sigma = \sqrt{D} = \sqrt{\left(\frac{\sum_{i=1}^{N}(ME_i[X] - \overline{ME}[X]_N)_i}{N}\right)^2}. \tag{2}$$

A final histogram is formed, which displays:

- on the abscissa axis - identifiers of alternative options for obtaining values (an ensemble of implementations on an experimental IETM sample and the criterion value of the accuracy parameter of the DNN);
- along the ordinate axis - the sample mean value of the ME of the proportion of correctly recognized objects in the entire ensemble of experiment implementations $\overline{ME}[X]_N$ and the corresponding criterion value.

The final histogram obtained in 6 is analyzed for statistical stability and correctness.

First of all, the ratio of the values $\alpha = \beta$ adopted according to 2 of this methodology and the value obtained according to (2), the sample standard deviation is analyzed. Ideally, the boundaries of the maximum scatter of the random variable specified by the risk values α, β and the obtained value of the sample standard deviation should coincide. In practice, this means that they should not differ significantly (ie, by more than 25-30%). Otherwise, it is necessary to increase the volume of statistical tests - N, to increase the level of risk of statistical estimation, revising the level of confidence in the results obtained, etc., in order to include the range of scatter actually obtained in the experiment according to (2) into the initially set interval for α, β according to clause 2 of this method.

Secondly, the fact of the presence of a difference between the criterial value of the estimated performance parameter of the DNN in the IETM and the sample mean value of the ME of the proportion of correctly recognized objects (targets) in the entire ensemble of experiment implementations $\overline{ME}[X]_N$ is revealed. The presence of a difference is recognized as statistically significant if the specified difference exceeds (goes beyond ...) the limits of the maximum scatter of the random variable $ME_i[X]$.

The presence of this positive difference constitutes the effect of the use of DNN as part of the IETM, estimated by the considered indicator with a probabilistic measure.

A specific version of the implementation of the DNN as part of the IETM is recognized as having passed the test and meeting the efficiency requirements for the estimated quality indicator of the DNN if the sample mean value of the ME of the proportion of correctly recognized objects in the entire ensemble of realizations on the experimental IETM sample is statistically significant (i.e., with the current parameters P_a, α, β) is greater than the analogous value of the accuracy of the DNN performance, determined a priori as a criterion for the current stage of research.

The resulting paired histogram obtained according to 1 - 7 is transmitted further for semantic interpretation with the set unbiased parameters of the statistical significance P_a, α, β.

4. Conclusion

The toolkit (methodology) proposed in this article for overcoming the problem of heteroscedasticity of statistical values of indicators for assessing the quality of IETM with elements of artificial intelligence is not the only and universal one. Today, the search for means for working with the results of statistical observations of the DNN as part of the IETM, characterized by heteroscedasticity, is the most urgent direction of modern qualimetry in the field of artificial intelligence software. In view of the above, further research on this topic should be aimed at studying the general causes, mechanisms of formation and patterns in the representation of the indicated heteroscedasticity; synthesis of mathematical models of its adequate presentation and accounting when conducting appropriate experiments and research.

5. References

[1] GOST R 50.1.030-2001, Information technology to support the life cycle of products. Interactive electronic technical manuals. Requirements for the logical structure of the database, Standartinform, Moscow, 2001.

[2]   GOST R 54088 – 2010, Integrated logistics support. Interactive electronic operational and maintenance documents, Basic provisions and general requirements, Standartinform, Moscow, 2012.

[3]   GOST R 53393 – 2017, Integrated logistics support. Basic Provisions, Standartinform, Moscow, 2017.

[4]   GOST R 53394 – 2017, Integrated logistics support. Basic terms and definitions, Standartinform, Moscow, 2017.

[5]   A.V. Shatohin, Information and support network - a new approach to the operation of equipment and technology, Nacionalnaya oborona 1(82) (2020) 62-67.

[6]   A. V. Shatohin, Ya. A. Ivakin, V. S. Neshtenko, Coordination of services of enterprises of marine instrumentation in the interests of the system of operation of hydroacoustic weapons of the Navy, Morskoj sbornik, 11 (2020) 12-54.

[7]   Ya. A. Ivakin, A. G. Varzhapetyan, E. G. Semenova, E. A. Frolova, Information and accompanying network of aircraft engineering products as an information basis for manufacturers' quality policy, Nauka i biznes: puti razvitiya 8(110) (2020) 102-117.

[8]   B. Ya. Sovetov, S. A. Yakovlev, System modeling, Izdatelstvo Yurajt, Moscow, 2019, p. 343.

[9]   R. M. Yusupov, V. P. Zabolotskij, Conceptual and scientific-methodological foundations of informatization, Nauka, SPb, 2009, p. 541.

[10] B. Ya. Sovetov, V. V. Cekhanovskij, Information Technology, Izdatelstvo Yurajt, Moscow, 2016, p. 263.

[11] S. Makkonnell, Perfect code. Master Class, Izdatelstvo «Russkaya redakciya», Moscow, 2010, p. 896.

[12] V. Kozlovskij, G. Yunak, S. Klejmenov, D. Blagoveshchenskij, Digitalization of production: a new format for statistical quality management tools, 2020. URL: https://ria-stk.ru/stq/adetail.php?ID=190419.

[13] N. Bykova, Russia needs a unified industrial digitalization policy, 2020. URL: https://ria-stk.ru/stq/adetail.php?ID=191184.

[14] L. V. Hlebenskih, M. A. Zubkova, T. Yu. Saukova, Industrial automation in the modern world. Molodoj uchenyj 16(150) (2017). URL: https://moluch.ru/archive/150/42390/.