A Fiducial Approach To Comparing The Electric Power Objects Of The Same Type

E.M. Farhadzadeh*, A.Z. Muradaliyev, Y.Z. Farzaliyev, T.K. Rafiyeva, S.A. Abdullayeva

Azerbaijan Scientific-Research and Design-Prospecting Power Engineering Institute, Baku, Republic of Azerbaijan

***Abstract* — An increase in service life of equipment and plants (objects) in electric power systems makes it more appropriate to relate the organization of a system of maintenance service and restoration of wear and tear to their technical condition. This, in turn generates the need to quantitatively estimate the indices of their individual reliability. There can be no data on failures and defects of concrete objects, therefore, in practice we often calculate generalized reliability indices. An intuitive understanding of the varied significance of varieties of attributes is reflected by classifying statistical data for some varieties of attributes. For example, they can be classified according to voltage class, design, service life, etc. At the same time, the question on the appropriateness of the statistical data classification is not considered. Initial assumptions of known methods and criteria of checking if it is expedient to classify the statistical data on failures of the electric power system objects in most cases are unacceptable, since they are not relevant to this data set. We have developed a new method and an algorithm to assess the appropriateness of the statistical data classification. Their novelty lies in the application of a fiducial approach to estimation of critical values of a sample from a set of multivariate statistical data.**

***Index Terms* — reliability indices, varieties of attributes, classification, expediency, risk of the erroneous decision.**

I. Introduction

The need to improve the methods for a quantitative estimation of reliability indices of electric power system equipment and devices (objects) has become increasingly more pressing over time [1]. This is largely associated with a recommended transition from a system with regulated term and scope of planned maintenance to wear and tear restoration depending on technical condition of an object [2]. Thus, it is obvious, that this concerns reliability indices of a certain object, in other words, an individual reliability of an object with set varieties of attributes (VA). The varieties of attributes are established on the basis of nameplate data, operation conditions, statistical data on operation, results of tests and repair.

Since the data on failures and defects of concrete objects can simply be absent, in practice we calculate generalized reliability indices that are used for approximate calculations rather than the individual ones. The intuitive perception of the interrelation between reliability indices and varieties of attributes, however, leads to the understanding that it is sensible to classify the statistical data by the varieties of attributes. The practice of classifying data by one of the set of varieties of attributes is widespread. For example, classification according to a voltage class, or rated power or design or other attributes. Occasionally, data are classified according to two, and sometimes three varieties of attributes. In this case, the question whether it is appropriate to classify the statistical data is not considered, i.e. the random nature of reliability indices estimates is not taken into account. It is worth reminding that these data are called multivariate, i.e. depending on the set of the variety of attributes.

We will be surprised at such parameters as average temperature of patients in hospital (for continuous random variables) and average academic performance of students at university or at school (for discrete random variables). However, we calculate average duration of idle time of objects in emergency repair, or an average number of disconnected short circuits a year, or availability factor of a set of objects, and use these parameters in further calculations. We unequivocally characterize such parameters, as average temperature of patients from a surgery department in

* Corresponding author.
E-mail: elmeht@rambler.ru

http://dx.doi.org/10.25729/esr.2018.04.0008
Available online February 28, 2019.

hospital or average academic performance of sophomore students as strange. Nevertheless, we are absolutely confident when we calculate the reliability indices, for example, for the objects of various voltage classes and analyze this dependence. In the end, we do not change our mind about strangeness of the estimates of average academic performance of sophomore students of the Energy Department of university or average temperature of patients of a female surgery department in hospital. At the same time, without any doubts, we use the estimates of reliability indices for the objects of a set voltage class and design in calculations. Certainly, someone can object and say that we use the varieties of attributes of different significance. Yes, possibly, but it confirms even more the necessity of a quantitative estimation of the significance of the varieties of attributes.

Application of known methods for checking the appropriateness of classification of the statistical data characterizing reliability of electric power system objects, in most cases is unacceptable, since initial preconditions of these methods are not relevant to them. These preconditions include, first of all, a great number of realizations of samples and the normal law of their distribution [3]. The difficulties also arise when solving practical problems related to the objects comparison and ranking. Fiducial approach in many respects helps overcome these difficulties.

Fiducial probabilities and intervals. Fiducial distributions were proposed by R.A. Fisher in 1935. They determine "to what extent, we can trust (fiducial means based on or having trust) any set value of an unknown index (parameter) of this distribution", and in essence, this is a distribution of possible realizations of distribution parameters of a random sample from a population [4]. According to Fisher:

1. We should trust only the decisions based on empirical data, to be more precise, distribution of an observable sample [5];
2. An acceptable way of constructing fiducial intervals is calculation of probability distribution of possible variable values [5];
3. Confidence and fiducial intervals are identical but only when a single parameter is estimated. If complex parameter is estimated by two or a greater number of parameters, the results can differ [6].

The difficulties in analytical representation of distributions of parameters calculated for small samples from finite population of multivariate data (FPMD) [7] and for complex indices are well known. However, even in 1942, Kolmogorov A.N. noted that at a small size of sample (ns) the best interval estimates are provided by fiducial probabilities [8].

New unlimited opportunities for calculation of statistical functions of fiducial distributions were brought about by the advent of computer equipment and development of computer technologies.

Prior to the algorithm for calculation of statistical function of fiducial distributions, let us specify the specific features of a relationship between confidence and fiducial intervals [9]. Confidence intervals of parameters are determined analytically by the known formulas for some initial preconditions. They characterize a set of possible realizations of a concrete parameter. For example, if to take a random sample of random variables with uniform distribution in an interval [0,1], we can easily enough identify the confidence borders of the arithmetic mean $M^*(\xi)$ with a set significance value. Between lower $\underline{M(\xi)}$ and upper $\overline{M(\xi)}$ boundary values of a confidence interval there is a set of possible realizations . This set does not necessarily include the true value of $M(\xi) = 0.5$. However, if we repeat these calculations a set of times (N), then $N(1-\alpha)$ confidence intervals will contain the value of $M(\xi) = 0.5$. This is a known engineering interpretation of the confidence interval.

In real operation of electric power system objects, there is certainly no opportunity to repeat the «tests». In addition, it is impossible to disagree with Fisher that empirical data characterize these objects. All uniform objects form a FPMD, for example, related to the duration of emergency idle time. A sample from this population characterizes the significance of an attribute according to which the classification is done. The set of possible realizations is determined similarly to an estimation of a set of realizations of a confidence interval (see an example with an estimation of arithmetic mean $M^*(\xi)$), the only difference being that for the confidence interval this set is determined analytically, while for fiducial interval - by simulation modelling. The simulation modelling should be carried out so that the statistical function of fiducial distribution of a calculated parameter completely coincides with distribution of the parameter within the confidence interval.

Thus, the boundary values of confidence and fiducial intervals of parameters coincide when the law of parameter distribution inside of a confidence interval is known, in other words, the law of fiducial distribution. This requirement provides objectivity of simulation modelling algorithm, an opportunity to control operability and estimate the accuracy when the initial preconditions are not met [10]. For example, the boundary values of Pearson correlation coefficient are calculated at a large size of samples and normal law of their distribution. The result of calculation of boundary values of Pearson correlation coefficient based on fiducial approach should completely coincide with its tabulated values. In these conditions, at super small samples [ns=(3÷10)], when boundary values of a confidence interval are erroneous, the fiducial approach provides objective calculation of accuracy.

The condition of correspondence between an analyzed parameter and a group of single parameters [11] is optional, since confidence intervals can be calculated for some complex parameters as well. Vivid examples of such parameters are the correlation factor, linear regression coefficient, etc.

Figure 1 demonstrates a simplified block diagram of modelling a statistical function of fiducial distribution (s.f.f.d.) for an arithmetic mean of random variables with uniform distribution in an interval [0,1] [12].

II. Modeling of possible realizations of objects reliability indices

We will model the reliability index realizations on an example of an average duration of forced idle time for eight circuit breakers of 300 MW oil-and-gas fired units [$M^*(\tau_c)$]. The realizations τ_c are given for illustration in Table 1.All these data are called finite population of multivariate data (FPMD). Their number n_Σ=43 and arithmetic mean $M^*(\tau_{c,\Sigma}) = \sum_{i=1}^{c}\sum_{j=1}^{n_i}\tau_{c,i,j}\Big/43 = 72{,}3\,h$ Estimates of this parameter for each power unit [$M^*(\tau_{c,i})$] are presented in and allows two groups to be identified. The first group

$$F_S^*(\tau_f) = \begin{cases} 0 & \text{if } \tau_f \le \tau_{f,1} \\ \dfrac{s-1}{n_i+1} + \dfrac{(\tau_c - \tau_{c,s})}{(n_i+1)\cdot(\tau_{c,(s+1)} - \tau_{c,s})} & \text{if } \tau_{f,1} < \tau_f < \tau_{f,(n_i+1)} \\ 1 & \text{if } \tau_f \ge \tau_{f,(n_i+2)} \end{cases} \quad (1)$$

where $s = 1,(n_i+1)$

will include the estimates [$M^*(\tau_{c,i})$] that exceed [$M^*(\tau_{c,\Sigma})$], and the second group will include estimates for which [$M^*(\tau_{c,i})$]<[$M^*(\tau\ \xi_{c,\Sigma})$] Our task is to determine estimates [$M^*(\tau_{c,i})$] that randomly differ from [$M^*(\tau_{c,\Sigma})$]

Let us consider three assumptions (hypotheses) H:

1. Estimate [$M^*(\tau_{c,i})$] randomly differs from [$M^*(\tau_{c,\Sigma})$] This will be denoted as $H \to H_1$;
2. Estimate [$M^*(\tau_{c,i})$] is non-randomly larger than [$M^*(\tau_{c,\Sigma})$] This will be denoted as $H \to H_2$;
3. Estimate [$M^*(\tau_{c,i})$] is non-randomly lower than [$M^*(\tau_{c,\Sigma})$] This will be denoted as $H \to H_3$.

where conformity is denoted by →

To make a decision with the minimal risk of erroneous decision, it is necessary to be able to calculate critical values of these parameters [$M^*(\tau_{c,\Sigma})$] and [$M^*(\tau_{c,i})$] with i=1.8.

Possible realizations of duration of the forced idle time of circuit breakers $\tau_{f,i}$ are modelled by statistical distribution functions $F^*(\tau_{c,\Sigma})$ and $F^*(\tau_{c,i})$ with i=1,n8. It is worth noting, that one of the basic reasons why the confidence and fiducial intervals are different is the discrepancy between the modelled set of possible realizations of calculated parameters and the real set.

The traditional approach to modeling possible realizations τ_c by $F^*(\tau_{c,i})$ for super small sizes of samples is unacceptable. In [13], we propose a new method. The statistical distribution function $F_s^*(\tau_{c,i})$ is represented by the equation (1):

Realization is calculated by the formula:

$$\tau_c = \tau_{c,s} + (\tau_{s+1} - \tau_{c,s}) \cdot [\xi(n_i + 1) - (s - 1)] \quad (2)$$

Realization τ_c of average duration of forced idle time $M^*(\tau_c)$ is calculated by n_i realizations τ_f.

III. Formation of statistical function of fiducial distribution

According to possible hypotheses we will distinguish three statistical functions of fiducial distributions:

$F^*[M^*(\tau_c / H_1)]$, $F^*[M^*(\tau_c / H_2)]$ and $F^*[M^*(\tau_c / H_3)]$.

For $F^*[M^*(\tau_c/H_1)]$ the sample should be representative.

For $F^*[M^*(\tau_c / H_2)]$ and $F^*[M^*(\tau_c / H_3)]$ the samples are modeled similarly to $F^*[M^*(\tau_c / H_1)]$ with the essential difference being that modeling of realizations τ_c is performed not by statistical distribution function $F^*(\tau_{c,\Sigma})$, but by $F^*(\tau_{c,i})$, where i=1,n8.

Table 1. Data on duration of forced idle time of circuit breakers, h.

i	Conventional numbers of circuit breakers of power units							
	1	2	3	4	5	6	7	8
1			78.59					
2			3.36				66.29	
3	64.42		3.48				47.02	
4	15.31	46.12	42.05	61.36		49.15	93.13	
5	53.5	46.27	45.15	236.3	63.5	91.17	54.03	36.05
6	94.55	298.58	62.36	123.59	38.07	99.51	79.21	6.23
7	69.37	134.12	18.15	358.15		39.11	57.2	15.35
8	5.48	35.51	29.42			133.24	66.1	
9	185.0		7.43				1.3	
10			25.5					
$\sum\tau_{c,i}$, h.	487.1	560.5	320.0	780.0	102.0	412.0	464.0	57.6
$\left[M^*(\tau_{c,i})\right]$	69.6	112.1	32.0	195.0	51.0	82.4	58.0	19.2

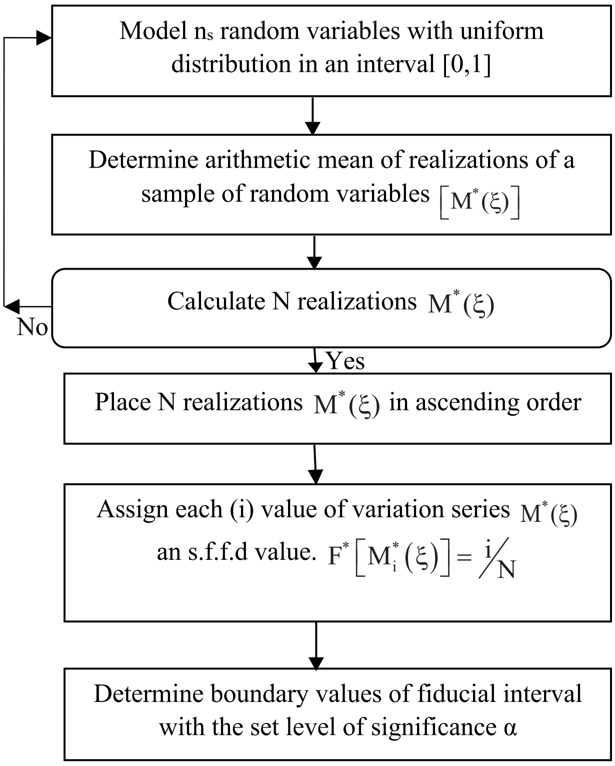

Fig.1. A simplified block diagram of an algorithm for modelling a s.f.f.d. $F^*\left[M_i^*(\xi)\right]$

It is obvious, that number of possible realizations $M^*(\tau_c)$ equal to N should satisfy the requirement of stability of estimates of quantiles of fiducial distributions for set values of Type I and Type II errors.

We will consider the quantiles of these distributions to be steady if the divergence of realizations of critical values does not exceed 1% with an increase in the number of realizations N. Let us note one more feature of the algorithm for decision-making on significance of varieties of attributes:

if $[M^*(\tau_{c,\Sigma})] < [M^*(\tau_{c,i})]$, we compare the distributions $R^*[M^*(\tau_c / H_1)] = [1 - F^*[M^*(\tau_c / H_1)]]$ and $F^*[M^*(\tau_c / H_2)]$;

if $[M^*(\tau_{c,\Sigma})] > [M^*(\tau_{c,i})]$ we compare $R^*[M^*(\tau_c / H_3)] = [1 - F^*[M^*(\tau_c / H_3)]]$ and $F^*[M^*(\tau_c / H_1)]$

Accordingly, we consider:

critical values $\overline{M^*(\tau_{c,(1-\alpha)})}$ and $\underline{M^*(\tau_{c,\beta_k})}$ at $[M^*(\tau_{c,\Sigma})] < [M^*(\tau_{c,i})]$;

critical values $\overline{M^*(\tau_{c,(1-\beta)})}$ and $\underline{M^*(\tau_{c,\alpha_k})}$ at $[M^*(\tau_{c,\Sigma})] > [M^*(\tau_{c,i})]$.

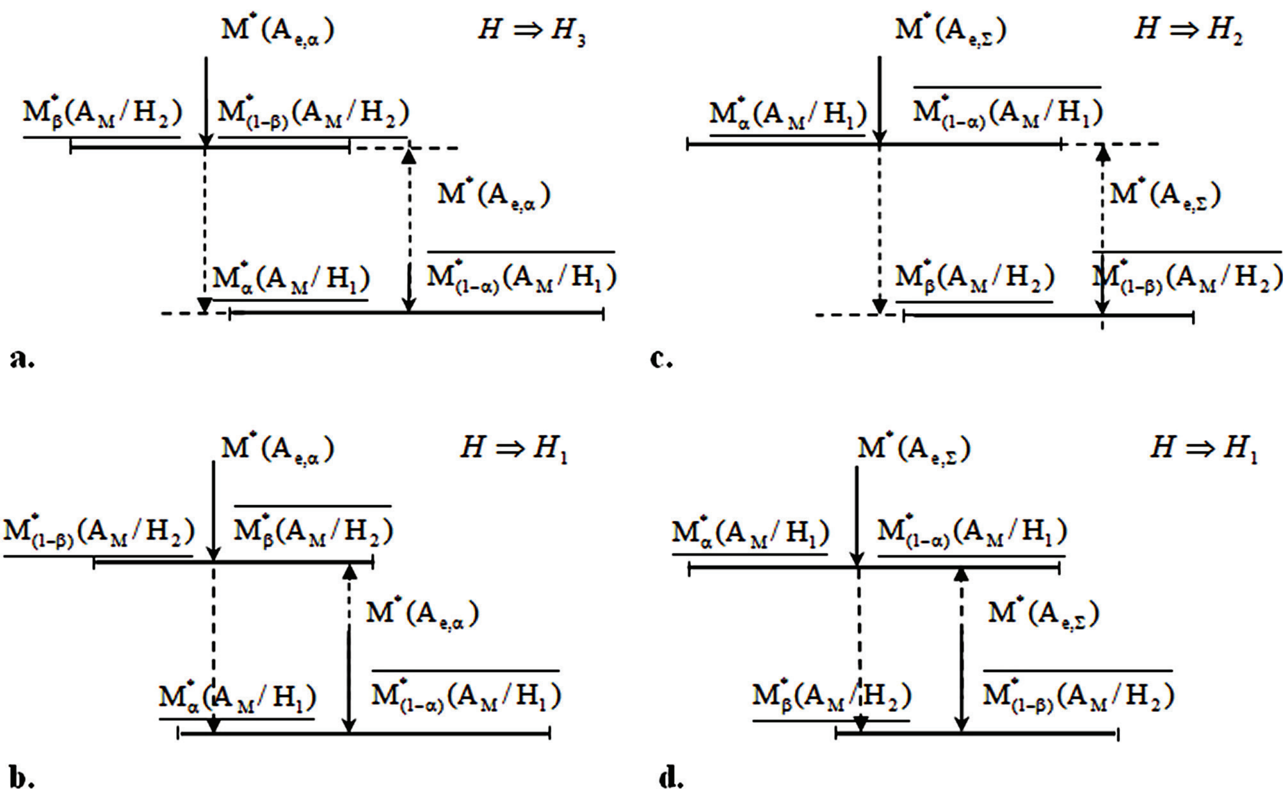

Fig. 2. Graphic illustration of relationships between experimental estimates of reliability indices (A_e) ***and their critical values:*** (a and b) → $M^*(A_{e,i}) > M^*(A_{e,\Sigma})$; (c and d) $M^*(A_{e,i}) < M^*(A_{e,\Sigma})$

IV. A criterion of decision-making on significance of attribute varieties

The appropriateness of the statistical data classification with respect to a set variety of attributes is assessed to specify the reliability indices of objects. Classification is considered to be appropriate if an estimate of reliability indices calculated after classification of data in a sample non-randomly differs from a reliability index calculated based on an initial data set. A condition, that determines the appropriateness is called criterion. The following criterion is proposed:

$$\left.\begin{array}{l|l} \text{1. If } M^*(A_{e,i}) < M^*(A_{e,\Sigma}) & \text{2. If } M^*(A_{e,i}) > M^*(A_{e,\Sigma}) \\ \text{and } M^*(A_{e,i}) < \underline{M^*_{\alpha}(A_M/H_1)}, & \text{and } M^*(A_{e,i}) > \overline{M^*_{(1-\alpha)}(A_M/H_1)} \\ \text{then } H \Rightarrow H_3 \rightarrow \text{exit} & \text{then } H \Rightarrow H_2 \rightarrow \text{exit} \\ \text{if } M^*(A_{e,i}) > \underline{M^*_{\alpha}(A_M/H_1)} & \text{if } M^*(A_{e,i}) < \overline{M^*_{(1-\alpha)}(A_M/H_1)} \\ \text{then } H \Rightarrow H_1 \rightarrow \text{exit} & \text{then } H \Rightarrow H_1 \rightarrow \text{exit} \end{array}\right\} \quad (3)$$

Here $M^*(A_{e,\Sigma})$ and $M^*(A_{e,i})$ are estimates of any reliability index, A_e, calculated accordingly by set (Σ) of experimental (e) data and sample (v) for the i-th object; $\underline{M^*_{\beta}(A_M/H_2)}$ and $\overline{M^*_{(1-\alpha)}(A_M/H_1)}$ are the lower and upper boundary values of fiducial interval, respectively, calculated by modelled (м) estimates of reliability index on the basis of representative samples of random variables; $\underline{M^*_{\beta}(A_M/H_2)}$ and $\overline{M^*_{(1-\beta)}(A_M/H_2)}$ are respectively the lower and upper boundary values of fiducial interval calculated by modelled estimates of reliability index on the basis of statistical distribution function of experimental sample of random variables.

Figure 2 presents a graphical illustration of the criterion for comparison of estimates of average duration of the forced idle time of the power unit circuit breakers. The estimates were obtained by classifying the statistical data according to the dispatcher numbers of circuit breakers. This method allows us to transition to the estimates of individual reliability indices and reliability indices of clusters of objects. Their basic difference is that the individual reliability indices are calculated by classifying the data according to the significant varieties of attributes out of those set, whereas the reliability indices of clusters are calculated by classifying the statistical data according to the significant varieties of attributes from a specified set of attributes and their varieties.

Table 2 presents the results of an analysis of the appropriateness of the classification of statistical data (Table 1) on the duration of the forced idle time of the 300 MW power unit circuit breakers.

If $H \Longrightarrow H_1$, classification is considered to be inappropriate, while at $H \Longrightarrow H_2$ or $H \Longrightarrow H_3$, it is appropriate. Thus, irrespective of the relationship between the experimental estimates of average idle time of power unit circuit breakers according to the data population and samples, classification of data is inappropriate for four of eight circuit breakers (1, 5, 6 and 7). An analysis of this vivid example confirms the appropriateness of monitoring the significance of a divergence of reliability index estimates before and after classification of statistical data.

V. Conclusion

1. The accuracy and reliability of calculation of the reliability indices of electric power system equipment and devices can be increased by:
   - employing computer technologies based on simulation modeling of fiducial distribution in the statistical analysis of experimental data;
   - determining (based on these distributions) the critical values of reliability indices;
   - applying the recommended criterion of estimation of the data classification appropriateness;
2. Fiducial distributions of reliability indices are modeled for analyzed assumptions: classification, accordingly, is appropriate and inappropriate. This provides the minimal risk of erroneous decision;

Table 2. An analysis of appropriateness of the classification of statistical data on forced idle time of 300 MW power unit circuit breakers according to their conventional number

№	Parameters	Conventional number of circuit breakers							
		1	2	3	4	5	6	7	8
1	n_i	7	5	10	4	2	5	8	3
2	$M^*_{e,i}(\tau_{em})$, h.	69.7	112.1	32	195	51	82.4	58	19.2
3	$\overline{M^*_{M,(1-\alpha)}(\tau_{em})}$, h.	-	108.9	-	113.3	-	108.9	-	-
4	$\underline{M^*_{M,\alpha}(\tau_{em})}$, h.	41.5	-	46.4	-	15.2	-	43.4	24.6
5	$\overline{M^*_{M,(1-\beta)}(\tau_{em})}$, h.	93.4	-	-	-	91.3	-	81.2	-
6	$\underline{M^*_{M,\beta}(\tau_{em})}$, h.	-	-	-	-	-	40.9	-	-
7	H	H_1	H_2	H_2	H_2	H_1	H_1	H_1	H_2

3. Classification of statistical data based on the set varieties of attributes is done until the obtained reliability index estimate randomly differs from an estimate calculated in the preceded stage of the classification;
4. The recommended method makes it possible both to control the statistical data classification when complex reliability indices are calculated, which cannot be done by any of the existing methods, and to operate small samples of multivariate data;
5. For small samples, with the number of realizations varying from 2 to 10, one of the pitfalls is a discrete nature of fiducial distribution. In this case, we propose switching from comparison of the fiducial distribution quantiles, to comparison of experimental values of Type I and Type II errors with their critical values.

References

[1] N.I. Voropai, "A concept of SMART-GRID and reliability of electric power systems. Methodological problems in reliability study of large energy systems." Issue 62, Ivanovo. Press Hundred, 2011, p.321-325.

[2] "Rules of technical operation of power plants and networks of the Russian Federation." Ministry for the Electric Power Industry of the Russian Federation. 2003, 160 p.

[3] M. Kendal, Stuart P.A. "Statistical inference and relations." Publishing House "Nauka", 1973, 900 p.

[4] A.I. Orlov, "Applied statistics." M.: Publishing House "Examen", 2004, 656 p.

[5] R.A. Ficher, "The concepts of inverse probability and fiducial probability referring to unknown parameters" // "Proc.Roy.Soc. (London)" A139, p. 343-348. 1939

[6] R.A. Ficher, "The fiducial argument in statistical inference" // "Annals of Eugenics", 6, p. 391-398, 1935

[7] D.A. Frazer, "The fiducial method and invariance" // Biometrika, 1961,v.48, p. 261-280

[8] A.N. Kolmogorov, "Determination of the center of dispersion and a degree of accuracy for a limited number of observations." Izv. of the AS of the USSR. Mathematics. 1942, 6, p. 3-32

[9] E.M. Farhadzadeh, "About an arrangement of boundary values of confidence and fiducial intervals of reliability indices of systems." Izv. of the AS of the USSR, Technical cybernetics No.5, 1979, p.196-199.

[10] E.M. Farhadzadeh, "Estimation of the influence of differentiation of the initial information on system reliability," *Methodological problems in reliability study of large energy systems*. Issue 7. Irkutsk, 1976, p.24-34.

[11] E.M. Farhadzadeh, "About an estimation of statistical distribution function of duration of serviceable work of power system components," *Journal "Elektrichestvo"*, 1976, No. 6, p.78-80.

[12] E.M. Farhadzadeh, "Statistical modeling of duration of equipment failure-free operation on electronic digital computer," Izv. of AS of the USSR. "Energy and Transport", 1977, No.6, p.174-178.

[13] E.M. Farhadzadeh, Y.Z. Farzaliyev, A.Z. Muradaliyev, "Comparison of methods for modelling of continuous random variables by empirical distributions," Baku, "Energy problems", No.1, 2013, p.25-31.

Elmar M. Farhadzadeh (Professor, Dr. Sc.) was born in 1939 in the city of Baku of the Republic of Azerbaijan. He graduated from the Energy Department of the Azerbaijan Oil and Chemistry Institute in 1961. In 1982, he defended the doctoral thesis "Accuracy and reliability of characteristics of reliability of power plants" in Novosibirsk Electro-Technical Institute. Currently he is the head of the Laboratory for Reliability of the Equipment of Electric Power Systems at ASRDPPEI in Baku. His research interests are: reliability and efficiency of electric power systems.

Aydin Z. Muradaliyev (PhD) was born in 1960 in the city of Baku of the Republic of Azerbaijan. He graduated from Energy Department of the Azerbaijan Oil and Chemistry Institute in 1982 year. In 2013, he defended the doctoral thesis "Development of methods and algorithms for calculation of reliability indices of the electric power system equipment and devices". He is the head of the Department for Reliability of Equipment of Electric Power Systems at ASRDPPEI in Baku. His research interests include: Quantitative estimation of reliability of equipment and devices of electric power systems.

Yusif Z. Farzaliyev (PhD) was born in 1958 in the city of Fizuli of the Republic of Azerbaijan. He graduated from the Azerbaijan State University in 1985, as an applied mathematician. In 2008, he defended the master's thesis “An increase in the accuracy and reliability of calculation of reliability indices of power units”. He is the main expert of the Department for Reliability of Power Equipment at ASRDPPEI in Baku.

Tamara K. Rafiyeva was born in 1950 in the city of Baku of the Republic of Azerbaijan. In 1973 she graduated from Energy Department of the Azerbaijan Oil and Chemistry Institute. In 2007 she defended the master's thesis “Simulation modeling of individual reliability of power units”. She is a senior scientific researcher of the Department for Reliability of Power Equipment at ASRDPPEI in Baku.

Her research interests include: Quantitative estimation of a technical condition of the equipment and devices of electric power systems

Samira A. Abdullayeva was born in 1968 year in the city of Baku of the Republic of Azerbaijan. She graduated from the Energy Department of the Azerbaijan Oil and Chemistry Institute in 1990. She is a leading engineer of the Department for Reliability of Equipment of Electric Power Systems at ASRDPPEI in Baku. She is a post-graduate student.