Determination of the Sample Distribution Law by Analysis of Multiple Measurements

Eugene O. Podchasov and Arina D. Terenteva
Department of robotics and complex automation, BMSTU, Russia
Email: podchasov@bmstu.ru, terentevaad@bmstu.ru

Abstract—**Determination of distribution law of measured series of values is important task in automated regulation and active control. In many cases this is a challenge, especially when type of the law is uncommon and differ from Gaussian. The method for determination of distribution law based on criteria of process stabilizing is proposed. The recommendation for Poisson, Exponential and χ^2 laws are given. Determination of distribution law of series is based on calculation of 4 criteria, each one being based on Mean, standard deviation or its derivative. In each case the series of 1000 computational experiments being held and on base of its results the distribution law being determined. Each case being described by the number of measurements that is necessary for each criteria level stabilize below value of 0,1. It is shown that for each main distribution laws the order in which each criteria stabilize in sequential measurement differs. The order of such stabilization is offered as a way to determine the distribution laws for measurements in active control algorithms, while the measurement amount necessary for stabilization may be used for estimation of distribution laws parameters. The calculation experiment results for some of the most common laws is given and criteria for each law definition is formulated.**

Index Terms— **mechanical experiment, technical measurement, precision, control automation, consecutive analysis**

I. INTRODUCTION

Experiment is the essential of any scientific researches. In modern conditions there is a necessity of experiment's, which includes an enormous amount of observations. It may be caused by observation of processes with very low probability of happening or by increment of measurements number for accuracy increment in cases when precision of equipment is insufficient [1]-[3].

Multynumbered measurements are of a high importance for usage in systems of continuous control. Data, which is collected by such systems, may become basement for automated control and adjustment systems synthesis after correct mathematical calculations [4]-[5].

The most perspective way of adaptive control is adjustment with moving average, cause it consists information about several pre-measured values of parameter controlled.

It is especially important in mechanical treatment control. To control the linear values, it is necessary to use measuring equipment with an accuracy that exceeds the measured value by a discharge. This condition of the intermediate control and acceptance tests can lead to such difficulties as the lack of necessary equipment, excessive increase in the cost of the process and the finished product, technological difficulties of control associated with the configuration of parts. In such cases, multiple measurements are used to control high-precision linear quantities using the existing universal measuring equipment.

By means of mathematical modeling in papers [1]-[4] shown that in cases of estimation of general populations or selections that close enough to them optimal way of adjustment depends on distribution law of value being measured. However, it still unknown on practice [5]-[8]. Moreover, practice works almost always is about processing selections with limited number of measured values. That's why there is a necessity of development method to estimate properties of selections and choose the most similar general population [9].

To find the most effected way to estimate properties of selection, a mathematical model of pseudorandom values was generated using Matlab. Each pseudorandom measured value consists of three parts: the first part is a nominal value [11]-[13], the second part is systematical error [14]-[16], which can de described by linear or periodical law.

In all of the technical measurements it is adopted to think that values being measured have Gaussian distribution law [12]. However, in researches in field of mathematical statistic in the technical measurements R.Storm [17] shows that "in practice cumulative distribution functions are nearly always unknown". It means that classical methods used for estimation of measurement results is uneffective for many cases, that should be researched separately.

Sometimes the value being measured distributed by sophisticated and rare distribution law, which is hard to determine. The method being described allows to estimate the distribution law by mean of set of single values analizis.

Manuscript received August 13, 2020; revised January 11, 2021.

II. PROCESS STABILIZING CRITERIA

Each selection may be described with four dimentionless criteria, which show process stabilizing with each next measure. this criteria are:
- criteria of variation of average value

$$T_1(n) = \frac{x(i)_{av} - x(n+1)_{av}}{x(n)_{av}}$$

where n is the current number of measurements, x(i) is the current value of measurable;
- criteria of variation of average value increment

$$T_2(n) = T_1(n) + T_1(n+1);$$

- criteria of variation of standard deviation

$$T_3(n) = \frac{S(n) - S(n+1)}{S(n)},$$

where S(n) - standard deviation;
- criteria of variation of standard deviation increment

$$T_4(n) = \frac{D(n) - D(n+1)}{D(n)},$$

where D(n) - dispersion.

The choice of effective criteria leads to a reduction in the number of measurements. However, the stabilization of the process for each of the criteria for random samples is not uniform [6]. This necessitates the study of these equations, both independently and jointly, in order to analyze the influence of the characteristics of random processes on their stabilization for each of the criteria.

III. PROCESS MODELING AND RESEARCH METHODOLOGY.

To assess the effectiveness of the method

A thorough analysis based on these criteria is carried out mathematical modeling of a sample of 50 pseudorandom numbers in the range from 0 to 10, taken according to the uniform distribution law, after which the changes in the studied criteria T1, T2, T3, T4 are calculated in a sequential analysis.

The results of studies of various sequences of random numbers generated by the uniform law are presented in Table I, from which it is clear that the fluctuations in the average number of measurements required to stabilize the process are small. As soon as the process ceases to go beyond the boundaries of the interval specified by the user, it can be considered stable, and the number of measurements n (Ti) following the jump in the process that has gone beyond the limits of the interval is sufficient. If all four criteria T1, T2, T3, T4 are taken into account simultaneously, the result will be even more accurate.

TABLE I. AVERAGE VALUES OF THE NUMBER OF MEASUREMENTS n REQUIRED FOR STABILIZATION PROCESS, WITH DIFFERENT SEQUENCES OF RANDOM NUMBERS

№ of selection	n(T1)	n(T2)	n(T3)	n(T4)
1	8,341	12,822	9,662	10,899
2	8,955	12,850	9,554	10,893
3	8,355	13,028	9,502	10,895
4	8,876	12,951	9,559	10,896
5	8,312	12,873	9,660	10,890

An analysis of the results of a mathematical experiment revealed the convergence of solutions for similar source data and showed that the required minimum number of measurements to stabilize the process, in which there is an exclusively random error, is determined by the criterion T2 of the increment of average fluctuations. Similar experiments are given for various values of amplitude, the dimension of which coincides with the dimensions of the measured quantity and the random component of the error (Table II).

TABLE II. THE AVERAGE VALUE OF THE NUMBER OF MEASUREMENTS n, NECESSARY FOR STABILIZATION, AT VARIOUS AMPLITUDES OF A RANDOM VALUE

Range	n(T1)	n(T2)	n(T3)	n(T4)
0…0,1	8,341	12,822	9,662	10,899
0…1	8,340	12,900	9,670	10,879
0…2	8,370	12,827	9,597	10,896
0…25	8,341	12,987	9,477	10,887
0…100	8,350	12,885	9,779	10,915

When comparing the data from the Table I and II, it was revealed that the average values of the number of measurements remained almost unchanged, on the basis of which the following conclusion can be drawn: if there is only a random error in the process, the amplitude does not affect the stability of the measurement process. Based on this, the assumption was made about the constancy of the required number of measurements for the sample with random error. For further calculations, are given in table. 1 values are taken as reference, which indicate the presence of only a random error in the process.

When processing experimental studies, it is important to identify and eliminate systematic error. The reasons for its appearance during multiple measurements are varied, for example, vibration or non-return of the sensor to zero. To eliminate them, it is necessary to study the patterns of mutual influence of random and systematic errors and evaluate the measurement results.

For a sample containing both random and systematic components of the error, distributed according to a linear law, we can assume that the required minimum number of measurements determines only criterion T3. For research advanced assumptions about the influence of criterion T3 on the number of changes n, additional mathematical experiments were carried out that revealed the simultaneous influence of the amplitude of the random component of the error A and the coefficient k of the linear component of a systematic error in the absence of a systematic component distributed according to the periodic law (Fig. 1).
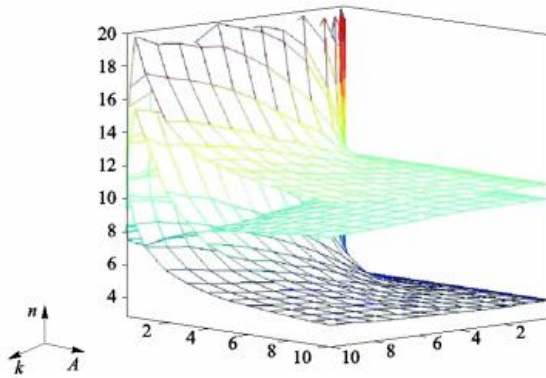
Figure 1. Dependence of the influence of the amplitude of the random component of the error and the coefficient k of the linear component of the systematic error

According to the data in Fig. 1 shows that the change in the random component

The corresponding error with a constant value of the systematic error has an insignificant effect on all four criteria T1, T2, T3, T4 in a sequential analysis (Fig. 2–5).
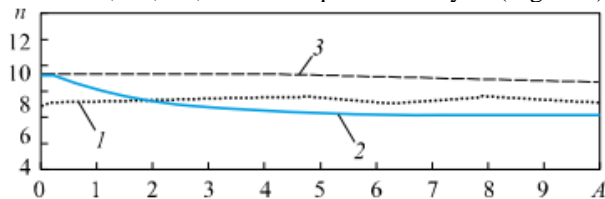

Figure 2. The dependence of the change in the amplitude of a random component of the error at a constant value systematic error for the T1 criterion: 1 - k = 0.001; 2 - k = 0.5; 3 - k = 10
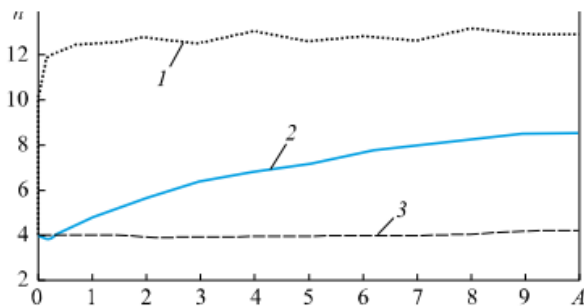
Figure 3. The dependence of the change in the amplitude of a random component of the error at a constant value systematic error for criterion T2: 1 - k = 0.001; 2 - k = 0.5; 3 - k = 10
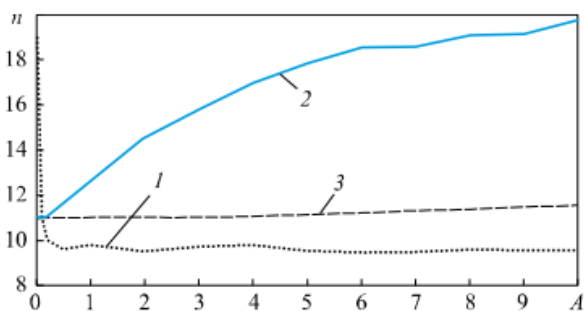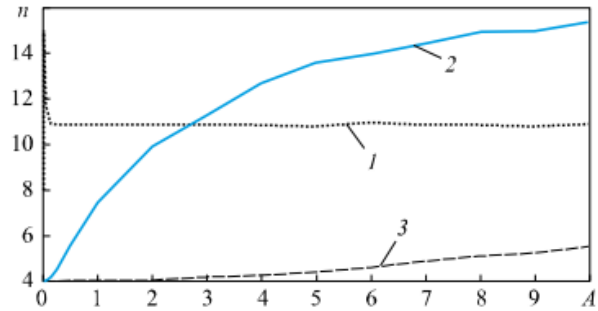

Figure 4. The dependence of the amplitude change of random component of the error at a constant value systematic error for criterion T3: 1 - k = 0.001; 2 - k = 0.5; 3 - k = 10

Figure 5. The dependence of the change in the amplitude of a random component of the error at a constant value systematic error for the T4 criterion:1 - k = 0.001; 2 - k = 0.5; 3 - k = 10

On the other hand, the distribution law of random part of bias may not have a uniform distribution law. It has the most importance when value being measured in not a linear size, but a number of failures, which has a great significance for statistical modelling.

IV. SELECTION OF MODELING PARAMETERS

Previously held researches [8], [9] shown the efficiency of such a method for determination of systematicsl parts of biases. In this work mathematical modeling of large amount of selections is considered, so the assumption of compliance with the parameters of the selections and general population was considered as adequate.

Most processes requires not much than 50 measurement at the same section, that's way it is enough to limit the number of measurement at this level n = 50. For averaging the obtained results of each experiment, repeated measurements are carried out with the number of repetitions 1000.

For statistical evaluation of the results, the standard deviation is selected js.

In the subsequent figures, the color indicates the criteria: T1 – red; T2 – green; T3 – blue; T4 – magenta.

The Poisson model usually describes a pattern of rare events: under certain assumptions about the nature of the process of occurrence of random events, the number of events that occur over a fixed period of time or in a fixed region of space is often subject to a Poisson distribution.

Examples are the number of radioactive decay particles recorded by the counter for some time t, the number of calls received at the telephone exchange during time t or the number of defects in a piece of cloth or tape of a fixed length.

For the Poisson distribution law, the scale parameter λ = 1, 3, 6 is changing. The results are presented in Table I.

TABLE III. AVERAGE VALUES OF THE NUMBER OF MEASUREMENTS N NECESSARY FOR STABILIZATION OF THE PROCESS, WITH DIFFERENT PARAMETERS OF THE POISSON DISTRIBUTION

№	$\lambda = 1$	$\lambda = 3$	$\lambda = 6$
$n(T_1)$	18.5900	8.9130	6.2390
$n(T_2)$	25.7390	13.0690	9.2260
$n(T_3)$	21.0130	18.8610	17.7430
$n(T_4)$	18.9860	17.1230	16.1980

For clarity and ease of interpretation of the results obtained in Fig. 1, the values of the criteria T_1, T_2, T_3, T_4 by colors are shown. The figure also has control limits for admission to criteria T_1, T_2, T_3, $T_4 = 0.1$.
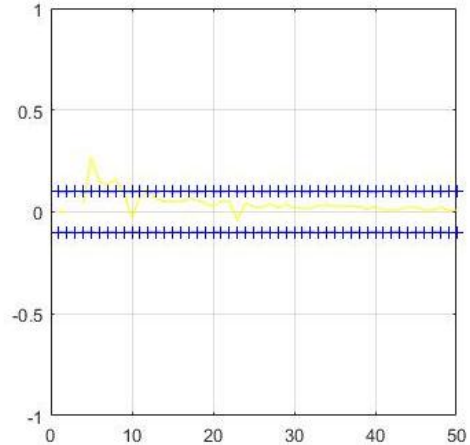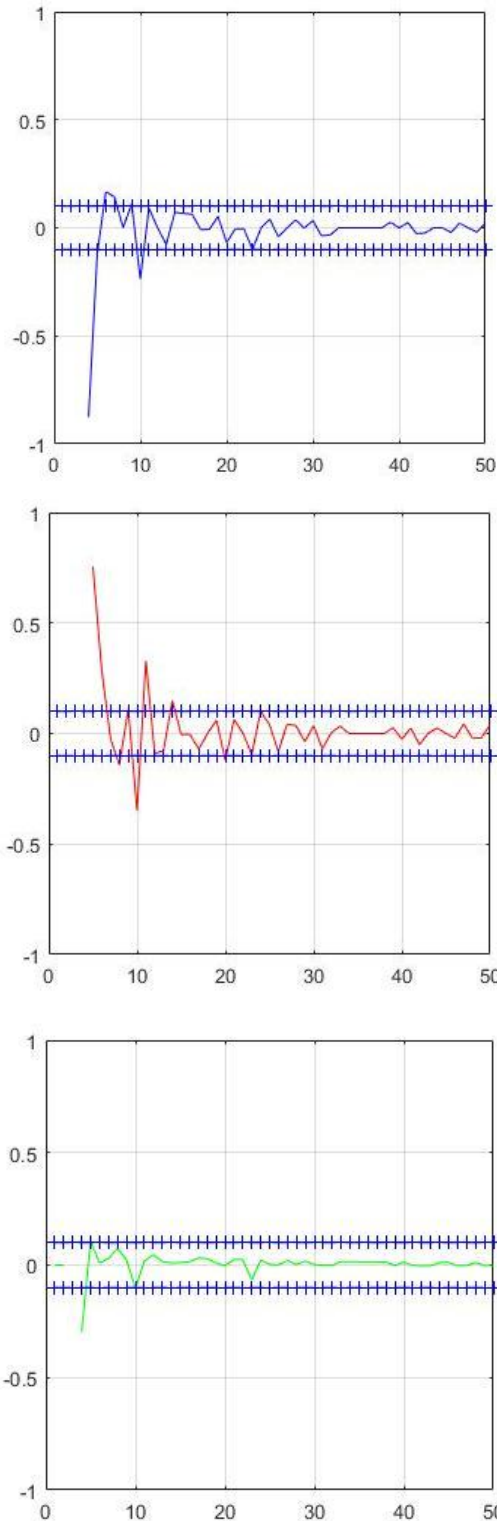

Figure 6.   Values of criteria T1 (blue), T2 (red), T3 (green), T4 (yellow), for the Poisson law

The exponential distribution models the time between two successive occurrences of the event, and the parameter λ describes the average number of occurrences of the event per unit time. Usually using this law they describe: the duration of customer service, the life of the equipment to failure, the time interval between breakdowns, etc.

For the exponential distribution law, the scale parameter λ = 0.1, 1, 3 is changing. The results are presented in Table II.

TABLE IV.   AVERAGE VALUES OF THE NUMBER OF MEASUREMENTS n NECESSARY FOR STABILIZATION OF THE PROCESS, WITH DIFFERENT PARAMETERS OF THE EXPONENTIAL DISTRIBUTION

№	λ = 0,1	λ = 1	λ = 3
n(T₁)	21.6140	21.5100	22.3950
n(T₂)	26.9210	27.2820	27.0100
n(T₃)	27.3330	27.2670	27.0870
n(T₄)	26.0730	25.9510	25.8300

For clarity and ease of interpretation of the results obtained in Fig. 2, the values of the criteria T_1, T_2, T_3, T_4 by colors are shown. The figure also has control limits for admission to criteria T_1, T_2, T_3, $T_4 = 0.1$.
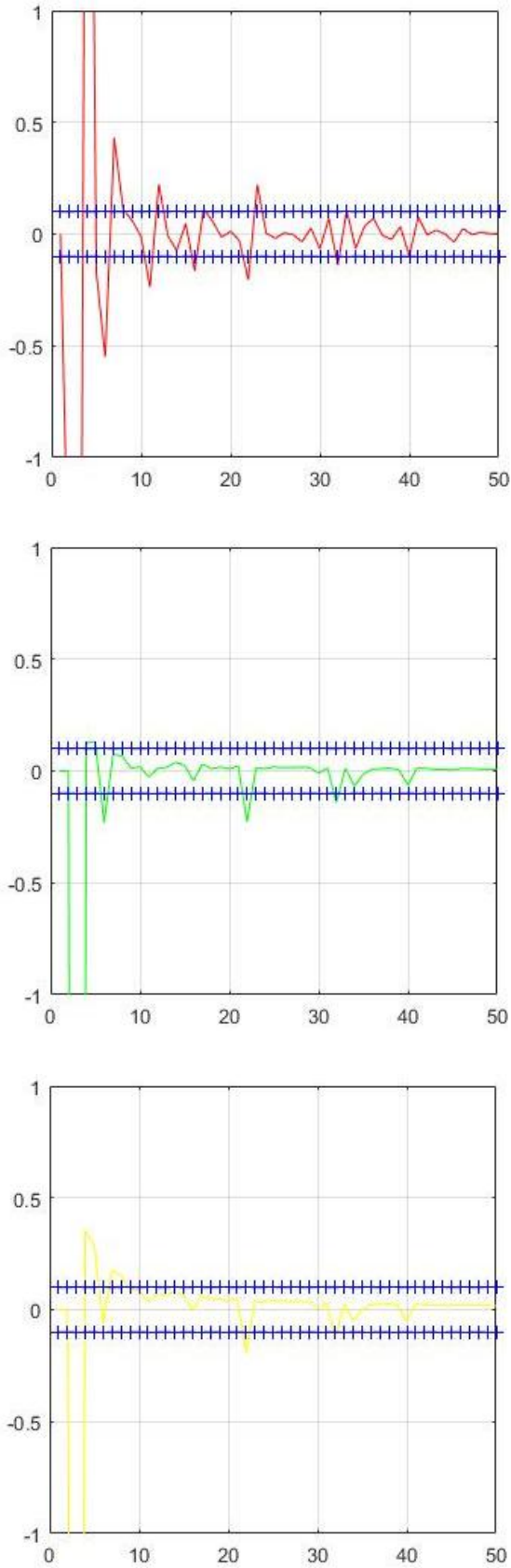
TABLE V.   AVERAGE VALUES OF THE NUMBER OF MEASUREMENTS N NECESSARY FOR STABILIZATION OF THE PROCESS, WITH DIFFERENT PARAMETERS OF THE χ^2 DISTRIBUTION

№	V = 1	V = 3	V = 6
n(T₁)	30.7140	16.4700	9.9650
n(T₂)	35.4150	21.7790	14.0650
n(T₃)	29.3640	25.9490	22.6570
n(T₄)	28.1130	24.2320	21.3300

For clarity and ease of interpretation of the results obtained in Fig. 3, the values of the criteria T_1, T_2, T_3, T_4 by colors are shown. The figure also has control limits for admission to criteria T_1, T_2, T_3, $T_4 = 0.1$.
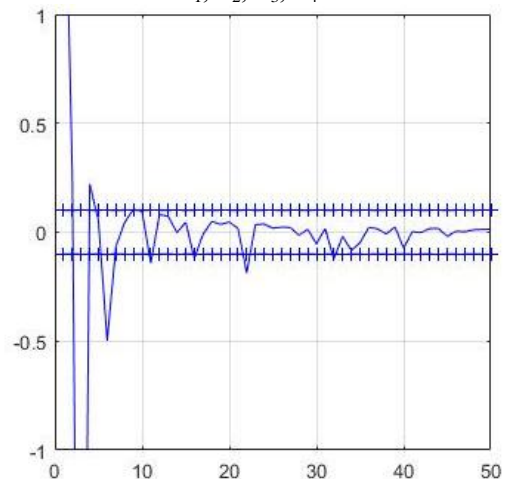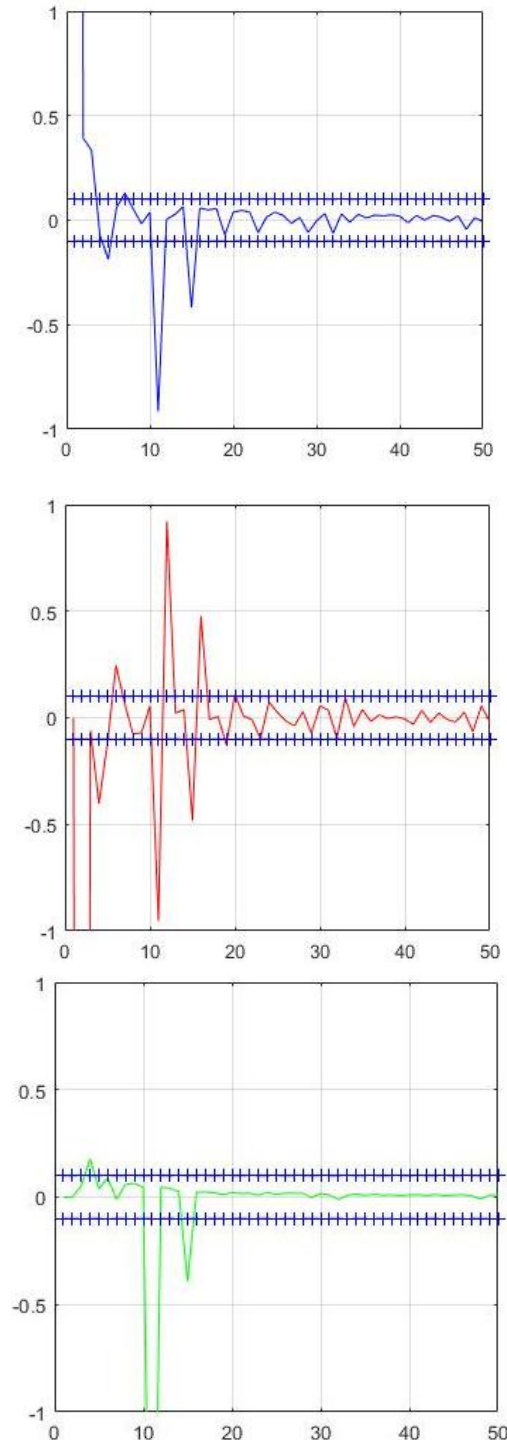

Figure 7.   Values of criteria T1 (blue), T2 (red), T3 (green), T4 (yellow), for the Exponential law

For the χ^2 distribution law, the scale parameter V = 0.1, 1, 3 is changing. The results are presented in Table III.
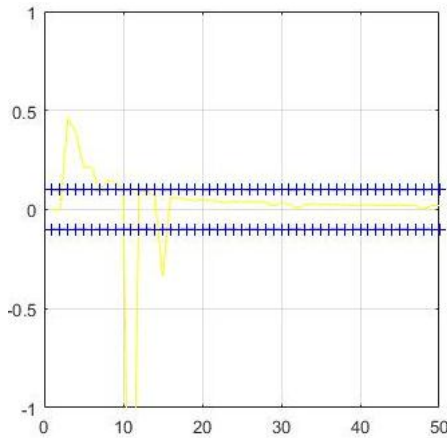
Figure 8. Values of criteria T1 (blue), T2 (red), T3 (green), T4 (yellow), for the χ^2 law

V. CONCLUSION

The results of mathematical modeling of repeated measurements of a quantity show:

1 The distribution of random bias according to the law of Poisson or χ^2 distribution with high scale parameter values can be determined by the order of criteria T3>T4>T2>T1.

2 The distribution of random bias according to the law of χ^2 distribution with high scale parameter value can be determined with T4 being more than 20 in all cases.

3 Distributions of random bias according to the exponential law is characterized by almost constant values of all four criterias while T2 and T3 are equal and the maximum of all.

4 Apart from Gaussian or rectangular distribution laws being described are unstable for Multynumbered measures with low number of measurements.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

E.O. Podchasov has developed a mathematical model and calculation program for method study. He held a calculations for this paper.

A.D. Terenteva has analyzed data, made a conclusions and wrote the article.

Both authors accepted the paper.

REFERENCES

[1] Y. B. Kolyada, Y. S. Korolev, N. T. Krushnyak, V. G. Razgulin, and V. N. Yanushkin, "Optimization of photovaltic measurement system," *Measurement Techniques*, vol. 54, no. 12, pp. 1353-1356, 2012.

[2] V. N. Yanushkin, Y. B. Kolyada, and N. T. Krushnyak, "High-precision digital photoelectric deflection measurement heads," *Measurement Techniques*, vol. 58, no. 12, pp. 1312-1316, 2016.

[3] V. N. Yanushkin, Y. B. Kolyada, and N. T. Krushnyak, "Digital photoelectric raster comparator," *Measurement Techniques*, vol. 59, no. 9, pp. 929-932, 2016.

[4] S. Ross, "Average run lengths for moving average control charts," *Probably Engineering and Information Science*, vol. 13, no. 2, pp. 209-220, 1999.

[5] S. H. Sheu and T. C. Lin, "The generally weighted moving average control chart for detecting small shifts in the process mean," *Quality Engineering*, vol. 16, pp. 209-231, 2004.

[6] A. D. Limorenko and Y. A. Shachnev, "Possibility of increasing the precision machining of parts through the use of a control algorithm," *Science and Education of BMSTU*, vol. 11, pp. 1-26, 2012.

[7] O. Baryshnikova, E. Podchasov, and A. Terenteva, "Selection of the most effective moving average control," in *Proc. 2018 ICIEAM*, 2018.

[8] M. E. Limorenko, E. O. Podchasov, and A. D. Terenteva "Accuracy increasing of automatic control responsible products," *Engineering Journal: Science and Innovation*, vol. 66, pp. 1-11, 2017.

[9] M. Limorenko, E. Podchasov, and A. Terenteva, "Development of systematical biases exclusion methods for technical measurements," in *Proc. 2018 ICIEAM*, 2018.

[10] E. Podchasov and A. Terenteva, "Theoretical methods for precision increment of earthwork made by power shovel actuator," *Lecture Notes in Mechanical Engineering: Proceedings of the 4th International Conference on Industrial Engineering*, 2018.

[11] G. N. Zaitsev, U. Kachestvom, "Quality control. Technological methods of quality control of products," St. Petersburg: Piter, p. 266, 2014.

[12] A. Y. Vukolov, A. V. Antonov, S. A. Vorotnikov, G. V. Shashurin, and D. V. Saschenko, "Mathematical model of 3-P wheel-legged mobile robotic platform," *International Review of Mechanical Engineering*, vol. 11, pp. 1-9, 2017.

[13] M. S. Nevelson, "Automatic control of precision processing machine tools," *Leningrad: Mechanical Engineering*, 1982.

[14] A. Y. Vukolov, and O. V. Egorova, "New challenges of 3D mechanisms models in TMM training," *International Review of Mechanical Engineering*, vol. 10, pp. 303-311. 2016.

[15] A. I. Rahmatullin and V. S. Moiseev, "Mathematical models and methods of optimization of non-stationary care systems," Kazan: Scholl., 211 p. 2006.

[16] D. P. Kim, "Automatic Control Theory: a textbook manual for high schools. vol. 2: Multi-dimensional, nonlinear, optimal and adaptive systems," Moscow: Fizmathlit, 463 p., 2004.

[17] N. I. Sidnyaev, Theory of Experiment Planning and Experiment Results Analysis, Moscow: Urait, 399 p., 2011.

[18] A. D. Terenteva, "Increase the accuracy of automated control for technical measurements", *IJMERR,* vol. 9, no. 4, 2020.

[19] ISO 21747:2006 "Statistical methods - Process performance and capability statistics for measured quality characteristics".

[20] IEC 61649:1997 "Goodness-of-fit tests, confidence intervals and lower confidence limits for Weibull distributed data".

[21] ISO 2854:1976 "Statistical interpretation of data - Techniques of estimation and tests relating to means and variance"

Eugene O. Podchasov was born in Moscow, Russia, at 05.02.1994. He gained his bachelor degree in field of metal cutting machines and tools in Bauman Moscow State Technical University in 2015, Master degree in field of Automation of technological processes in 2018.

He worked at Bauman Moscow State technical university as laboratory assistant in 2013-2016. Since 2016 he works as lecturer.

His scientific interest lay in field of mechanisms and software design of insulin pumps.

Mr. Podchasov has been awarded with Moscow government prize for best technical project by young researcher in 2010, and by Russian Fund for Promoting Innovations in 2017.

Arina D. Terenteva was born in Moscow, Russian Federation, at 26 April 1990. Gained her university degree in field of metrology and interchangeability in Bauman Moscow State Technical University in Moscow, Russian Federation in 2013. She worked as a teacher in community physics and mathematics school in Bauman Moscow State Technical University in 2008-2010; as laboratory assistant in department of robotics and complex automation in Bauman Moscow State Technical University in 2010-2013; as lecturer since 2013. She published more than 20 papers in Russian and international journals, mainly about automation control. Mrs. Arina D. Terenteva won competition as the best tutor of Bauman Moscow State Technical University in 2016/2017 year.