10th IMEKO TC7 International Symposium
June 30–July 2, 2004, Saint-Petersburg, Russia

MINIMIZATION OF UNCERTAINTIES IN MEASUREMENTS WHITH REPEATED OBSERVATIONS

Igor Zakharov, *Natalya Shtefan*

Metrology and Measuring Techniques,
National University of Radioelectronics, Kharkov, Ukraine

Abstract – Algorithms minimizing A-type uncertainties when processing results of measurements with repeated observations distributed according to the laws unlike the normal one are given. A comparative estimation of the developed and traditional algorithms is cited.

Keywords: uncertainty in measurements, abnormal law of distribution, Monte-Carlo method.

1. INTRODUCTION

Development of modern technology and scientific research sets before designers of measuring equipment a contradictory task of ensuring both high operating speed and low measurement errors. Solution of these problems requires creating algorithms for measurement information processing which make it possible to reduce the number of observations needed to get the prescribed accuracy and reliability. In this case it should be taken into account that the traditional processing algorithms were obtained assuming the normal law of distribution of the observation results; in practice it often doesn't correspond to reality.

The aim of the present paper is to obtain efficient estimates of mathematical expectation and dispersion when processing observation results distributed according to arbitrary laws.

2. SUBJECT&METODS

Algorithm for processing the results of measurements with repeated observations consists in the following:

1. Removing blunders from the observation results.
2. Obtaining efficient estimate of mathematical expectation of observation results taken as a measurement result.
3. Finding efficient estimate of the standard A-type uncertainty.
4. Determining efficient estimate of the extended uncertainty.

The above-mentioned efficient estimates were determined using the Monte-Carlo method.

3. RESULTS

3.1. Efficient estimate of the expectation

The result of measurements with repeating observations is usually determined as their average [1]

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \ . \qquad (1)$$

However such estimation is effective[1] only for the normal law of distribution [3].

The center of range

$$\breve{x} = \frac{x_{\min} + x_{\max}}{2} \qquad (2)$$

is effective estimation for rectangular distribution.

Here $x_{\min}$ is a minimum term of sample, $x_{\max}$ - a maximum term of sample.

The estimation of a median

$$\hat{x} = \begin{cases} x_{\frac{n+1}{2}}, & n - \text{odd}; \\ \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & n - \text{even}. \end{cases} \qquad (3)$$

is effective estimation for double exponential distribution.

The estimate

$$x_{eff} = k_1\bar{x} + k_2\breve{x} + k_3\hat{x}\,, \qquad (4)$$

is offered as the efficient estimate of the expectation, where k_1, k_2 and k_3 - are coefficients depending on the kurtosis E of distribution of observation results [2].

The Monte-Carlo method was applied for determination of coefficients k_1, k_2 and k_3. It was used in the following way:

[1] The effective estimate is the one having the least dispersion in comparison with the others.

1. With the means of the generator of random numbers for the given law of distribution m samples each of the given size n of values of the measured parameter x were received. In so doing not less than 20 laws of distribution were taken, the sample size n varied from 3 up to 21, and the number of samples m of each volume was not less than 1,000.

2. For every j-th sample the parameters $\bar{x}_j$, $\hat{x}_j$ and $\breve{x}_j$ were calculated according to the expressions (1-3).

3. By changing values of factors k_1, k_2 and k_3 according to expression (4) samples of effective estimation of x_{eff} each of volume m were obtained.

4. For each sample the value of dispersion of estimation of x_{eff} was evaluated. The combination of factors corresponding to a minimum of the dispersion of x_{eff} was registered as optimum for the given sample size n and the law of distribution.

The research has shown, that optimum combinations of factors k_1, k_2 and k_3 do not practically depend upon the volume of samples n. The dependence of values of factors on a kurtosis of distribution is shown in fig. 1.

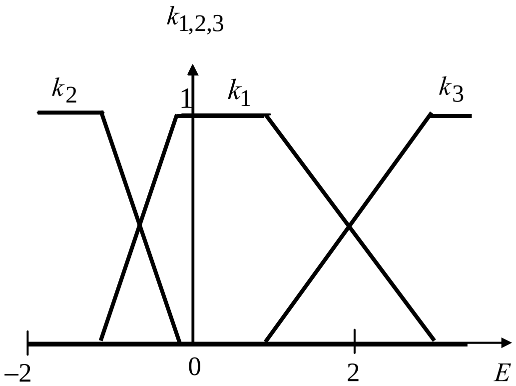

Fig. 1. The relationships between values of factors k_1, k_2 and k_3 and a kurtosis of distribution

These relationships were approximated by the expressions shown in table 1 with an error not exceeding 2 %.

For definition of comparative effectiveness[2] $Eff[x_{eff}/\bar{x}]$ of the proposed estimation and the average the ratio $\frac{D[x_{eff}]}{D[\bar{x}]}$ was calculated depending on the number of observations n for distributions with various kurtosis (Fig. 2).

[2] The comparative efficiency of an estimate is calculated as a relation of its dispersion to the dispersion of another estimate.

TABLE I. Dependence of factors of expression (4) upon a kurtosis

E	k_1	k_2	k_3
$\le -1{,}15$	0	1	0
$(-1{,}15; -0{,}2]$	$1{,}05E + 1{,}22$	$-0{,}05E - 0{,}22$	0
$(-0{,}2;\ 0{,}9]$	1	0	0
$(0{,}9;\ 3)$	$-0{,}47E + 1{,}41$	0	$0{,}47E - 0{,}41$
≥ 3	0	0	1

From Fig. 2 it is seen, that at the number of observations equal to 21 application of x_{eff} allows to reduce the dispersion of the estimation of a measurement result in comparison with the average up to 4.4 times for distribution with a kurtosis –1.2 and up to 1.7 times for distribution with a kurtosis 3.

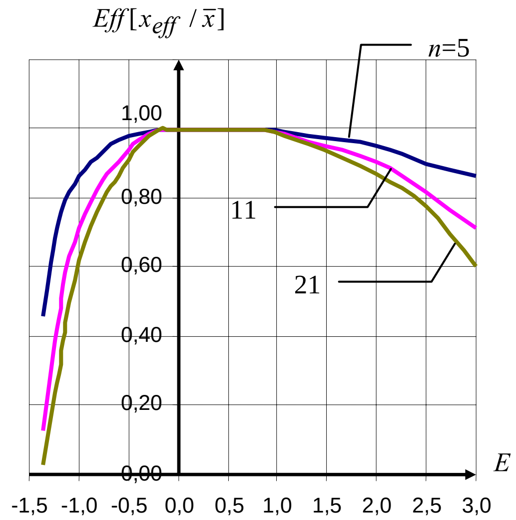

Fig. 2. Relationship of relative effectiveness of the proposed estimation to a kurtosis of distribution

3.2. Effective evaluation of the standard deviation

The estimation of standard deviation of an average is usually obtained from formula [1]

$$S[\bar{x}] = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2}\ . \qquad (5)$$

However such estimation is effective only for the normal law of distribution.

The estimation

$$S[x_{eff}] = \sqrt{Eff[x_{eff}/\bar{x}] \cdot \hat{D}[\bar{x}]} = \sqrt{\frac{Eff[x_{eff}/\bar{x}]}{n}} S \quad (6)$$

is proposed as an effective evaluation of the standard deviation of x_{eff}.

Here S is an experimental standard deviation of a single observation result, as which the following are investigated:

- the experimental standard deviation S_B, calculated by the Bessel formula

$$S_B = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - x_{eff}\right)^2}\,; \quad (7)$$

- the experimental standard deviation calculated through the range R (for limited distributions):

$$S_R = \frac{R}{2k_{p=1}} = \frac{x_{\max} - x_{\min}}{2k_{p=1}}\,, \quad (8)$$

where $k_{p=1}$ is a coverage factor for the level of confidence $p = 1$;

- the experimental standard deviation calculated through the average absolute deviation by the formula

$$S_A = \frac{1}{nd}\sum_{i=1}^{n}\left|x_i - x_{eff}\right| = \frac{S_a}{d}\,, \quad (9)$$

where $\hat{S}_a$ is the estimation of an average absolute deviation; d is a parameter characterizing the form of distribution.

All estimations of standard deviation are biased, therefore their use for small samples demands introduction of correction factor.

The study of comparative effectiveness of estimations S together with correction factors by the Monte-Carlo method has made it possible to receive the following expression for $S[x_{eff}]$:

$$S[x_{eff}] = \begin{cases} K_S S_R, & -1{,}5 \le E \le -0{,}55; \\ K_S S_B, & -0{,}55 < E \le 0{,}5; \\ K_S S_A, & 0{,}5 < E \le 3. \end{cases} \quad (10)$$

Here factor K_S, which takes into account the law of distribution, the number of results of measurements and the bias of experimental standard deviation, was approximated by the expression

$$K_S(E,n) = \begin{cases} 1{,}83n^{-0{,}66}E + 4{,}72n^{-0{,}78}, & E \le -0{,}55; \\ 1{,}23n^{-0{,}58}, & E > -0{,}55, \end{cases} \quad (11)$$

with an error not exceeding 10 %.

3.3. Extended uncertainty

The extended uncertainty of average values is usually calculated by the formula:

$$U = t_p(\nu)\,, \quad (12)$$

where $t_p(\nu)$ is Student's factor for degrees of freedom $\nu = n - 1$ and level of confidence p.

However such estimation is authentic only for the normal law of distribution.

In order to determine the extended uncertainty x_{eff} for any law of distribution the expression

$$U = kS[x_{eff}] \quad (13)$$

has been proposed, where k is a coverage factor, having the distribution of value

$$t = \frac{x_{eff} - M(x)}{S[x_{eff}]}\,. \quad (14)$$

This distribution for the initial laws of distribution with various values of the kurtosis, and also the numerical values of coverage factors k at small number of observations n were obtained by a Monte-Carlo method, as follows. After deriving model of distribution of $t_p(n)$ the empirical integral distribution function $F_p(t)$ is created. For the given levels of confidence p the values of factors k_- and k_+ are determined as follows

$$k_- = F^{-1}\left(\frac{1-P_\partial}{2}\right), \quad k_+ = F^{-1}\left(\frac{1+P_\partial}{2}\right), \quad (15)$$

after which the coverage factor value is determined by the formula

$$k = \frac{k_+ - k_-}{2}\,. \quad (16)$$

The calculated numerical values of k are shown in Fig. 3 for some considered distributions and various values of confidence level p.

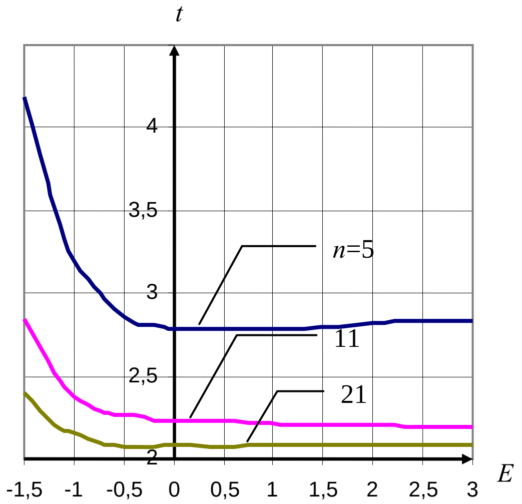

Fig. 3. Coverage factor

The elimination of blunders is carried out at the stage of preliminary processing of the results of measurements. This procedure can be carried out by application of either truncated estimations or special statistical criteria. A disadvantage of truncation is the increase of dispersion and bias of obtained estimations.

3.4. Elimination of blunders

For elimination of blunders the median criterion has been developed

$$\frac{x_{\max(\min)} - \hat{x}}{s_a} \geq \alpha , \qquad (17)$$

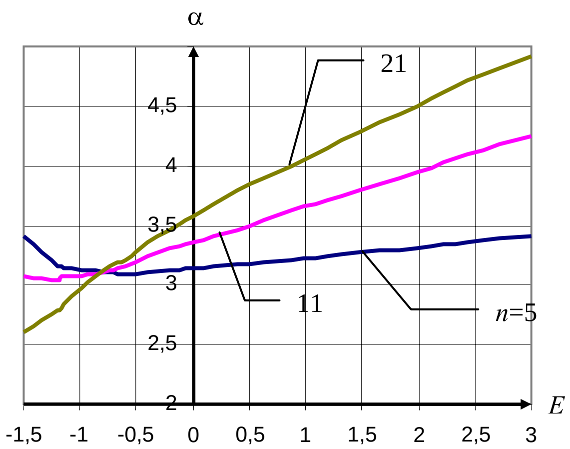

Fig. 4. Dependence of parameter α on the form of distribution, the number of observations and the level of significance

where $x_{\max(\min)}$ is a maximum or minimum result of observations, which can contain a blunder; s_a is an estimation of an average absolute deviation of results of observations, which like the median is least sensitive to blunders; α is a parameter depending on the form of distribution, the number of observations and the level of significance, for which the corresponding statistical data (Fig. 4) have been calculated.

4. CONCLUSIONS

Investigation of x_{eff} has shown that its comparative efficiency changes depending on E from 1 to 0.03 when compared to $\bar{x}$, from 1 to 0.07 when compared to $\breve{x}$, and from 1 to 0.005 when compared to $\hat{x}$.

It is shown that with small samples the median criterion is more efficient than the truncation algorithms used in robust methods.

It is indicated that the offered estimation of the standard uncertainty is more efficient than the estimation of the standard deviation obtained from the expression for the sample dispersion, and the obtained values of the coverage factor differ significantly from Student's factor with a small number of observations.

REFERENCES

[1] Guide to the Expression of Uncertainly in Measurement/ First edition – ISO/ Switzerland. 1993 - 101 p.

[2] Zakharov I.P., Shtefan N.V. Determining efficient estimates of the distribution center in statistical processing of observation results. In Radio electronics and informatics (Kharkov), 2002, No. 3 (20) (In Russian)

[3] Zakharov I.P., Kukush V.D. Theory of uncertainty in measurements. Manual: Kharkov, Consum, 2002 - 256 p. (In Russian).

Authors: prof. Igor Zakharov, assistant Natalya Shtefan, department of Metrology and Measuring Techniques, Kharkov's National University of Radioelectronics, Lenin's av., 14, 61166, Kharkov, Ukraine, phone: +380577021331, fax :+380577021013, e-mail. zip@bi.com.ua