Examples of Robust Estimation with Small Number of Measurements

Evgeniy T. Volodarsky[1] and Zygmunt L. Warsza[2]

[1] State Technological University - Kiev Politekhnika, Ukraine
vet-1@ukr.net
[2] Industrial Research Institute of Automation and Measurement (PIAP) Warszawa Pl
zlw@op.pl

Abstract. Two robust methods of assessing the value and the uncertainty of the measurand from the samples of small number of experimental data are presented. Those methods should be used when some measurements results contain outliers, i.e. when the values of certain measurement significantly differ from the others. They allow to set a credible statistical parameters of the measurements with the use of all experimental data. The following considerations are illustrated by the numerical example of the interlaboratory measurement data key comparison. Compared are the results obtained by a classical method with rejection of outliers with two robust methods: a rescaled median absolute deviation MAD_S and an iterative two-criteria method.

Keywords: robust statistics, outliers, uncertainty of measurements, interlaboratory comparisons.

1 Introduction

In many experimental studies in various fields, including the technical and scientific research, interlaboratory comparison and laboratory proficiency testing the measurement samples can contain few number of elements only. This occurs because of the high costs of measurements, the use of destructive methods, the poor availability of objects for testing, or the inability of multiple tests due to long or limited time of their execution. For small samples the measurement result and its uncertainty u_A evaluated by the GUM recommendations [1], significantly depends on the outliers. Therefore the obtained values sometimes may be even unreliable or unrealistic. Removing one observation only from a small sample significantly reduces the credibility of the evaluation results. For example for a very small sample of 4 elements the relative standard deviation of uncertainty $s(u_A)/u_A$ is as high as 42%, and for n=3 it will increase even up to 52% (GUM [1], Table E.1 in Appendix E.1). The removal of only one observation from a such small sample increases the relative standard deviation of uncertainty approximately on 24%. So the general tendency for small samples with outliers is to use the robust statistical methods, which applying all data obtained experimentally, including outliers. These methods are developed and to be used from the late 70's of

R. Szewczyk et al. (eds.), *Progress in Automation, Robotics and Measuring Techniques*,
Advances in Intelligent Systems and Computing 352, DOI: 10.1007/978-3-319-15835-8_31

the twentieth century. They are highly resistant to the influence of outliers. Such data considered before in conventional methods as to be "bad" can be successfully used now. Literature on these methods is quite rich. An overview of the basic items are in the bibliography of [5] - [8]. Robust methods provide less than conventional methods the impact of too high errors caused by different usually unrecognized sources. The term *robust* means resistance immunity to irregularities and inhomogeneities of the sample data.

In the robust statistics the outlier data are not removed, but are used different ways to modify their values, or their participation in procedures to estimate the statistical parameters of the sample. A number of robust statistical methods, (among others) are programmed in MatLab. Two of them are recommended in ISO 13528-2 [3] for proficiency testing by interlaboratory comparisons. These methods should be added in the new upgraded GUM. Various data processing tasks appear constantly in the new applications of robust statistical methods including such one as calibration of multi-parameter measurements in chemometrics. One of the areas where robust methods could also be usefully applied is estimation of accuracy of results obtained by the some measurement method in inter-laboratory comparison experiments [6] - [8].

2 Method of Rescaled Median Deviation

In the simplest robust method for a sample of n elements used is the Median Absolute Deviation

$$\mathrm{MAD}_n = \mathrm{med}\{|x_i - M_n|\} \tag{1}$$

where: x_i - i-th element of the sample, $M_n = \mathrm{med}\{x_i\}$ - the median.

This simple robust procedure is as follows:

- for all n data x_i ordered by values determined is the median *med* and considered to be the estimate of the measurement result value,
- the deviations of the sample data sets from this median the median absolute deviation MAD is calculated,
- standard uncertainty $s(x)$ of the measurand is considered the rescaled median deviation $\mathrm{MAD_S}$

$$s(x) \equiv \mathrm{MAD_s} = \kappa(n)\ \mathrm{MAD} \tag{2}$$

For a normal distribution the value of κ_∞=1,483 is the asymptotic limit of the ratio of $s(x)$/MAD when $n \to \infty$, i.e. for the general population. Use of κ_∞ for samples with a finite number n of measurements gives too low the assessment of uncertainty, as $s(x_n) > s(x_\infty)$. Then for the more accurate estimation coefficient $\kappa(n)$ as dependent on the number of elements n in the data sample has to be applied. Randa of NIST published values of the coefficient $\kappa(n)$ in internet [4].

3 The Robust Iterative Method

More reliable statistical parameters than by above method can be obtained by an iterative robust methods. In the method of robust statistics considered here the outlier data is downloading to positions closer to the center of the distribution. This operation is called winsoryzation after the name of American mathematician Winsor. Samples with the outlier data should not be simulated by a model of single normal distribution and the least squares method (LSM) is not useful, as shares of single data in it increases with the square of its distance from the center of concentration. More resistant to large deviations is the criterion of minimum modules (LMM) given by Laplace. So, in robust methods many ways of both criteria "symbiosis" are used. It is assumed that only the central part of the PDF (Probability Density Function) of sample data distribution, i.e. for small deviations from the estimate of measurand value, does not differ from the normal distribution. Only for them the least-squares criterion LSM can be used. Beyond the limits of this range the criterion of minimum module LMM is used to reduce the impact of outliers. After Tukay and Huber [5] works it is possible to apply for data processing the iterative robust method under acronym IRLS (iteratively reweighted least squares). In this method the following functional is used for the sensitivity

$$\sum_{i=1}^{n} \rho(x_i - \mu) \tag{3}$$

where: $\rho(x_i - \mu)$ - function depended on the selected parameter c.

For observations of the deviation values of $|\varepsilon| < c\sigma$ (where σ is the standard deviation, c - factor) a square function is used and for larger deviations the modules $|\varepsilon| = |x_i - \mu|$ are minimized. So the function $\rho(\varepsilon)$ is "more mild" for data outliers with values $|\varepsilon| > c\sigma$ from the center of the sample distribution. Constant c determines the degree of "robustness". The value of the constant c depends on the percentage of "contamination" of the sample distribution. For 1% c=2, and for the 5% c=1.4. Commonly c=1.5 is used. Experimental data are modified in accordance with the selected criterion as follows

$$x_i{}^* = \begin{cases} x_i & \text{for} \quad |x_i - \hat{\mu}| \le c\sigma \\ \hat{\mu} - c\sigma & \text{for} \quad x_i < \hat{\mu} - c\sigma \\ \hat{\mu} + c\sigma & \text{for} \quad x_i > \hat{\mu} + c\sigma \end{cases} \tag{4}$$

where $\hat{\mu} = med\{x_i\}$ from data x_i ranked in ascending order.

"Treatment" of the data by (4) is one of the ways of winsoryzation. As resistant to outliers the estimate of the sample data grouping center $\hat{\mu}$ the median $med\{x_i\}$ shall be preliminary adopted. Huber [5] finds that the best assessment of the distribution center is the midrange between the lower first (p=1/4) and the higher third (p=3/4) of the sample quartiles (inter-quartile midrange) - Fig 1.

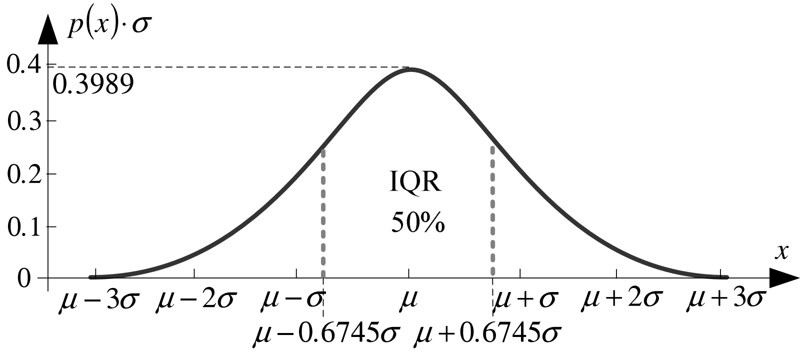

Fig. 1. Definition of inter-quartile mid-range: dotted lines – ordinates of first and third quartile $a = \mu$ - 0.6745, $b = \mu$ + 0.6745

The iterative procedure starts after arranging the elements of the sample according to their values $x_1, x_2, \ldots x_n$. Then the center of grouping data is

$$x^* = med\{x_i\},\ i = 1, \ldots, n \tag{5}$$

In this case the standard deviation is

$$s^* = 1{,}483 \cdot MAD_n \tag{6}$$

Then for c=1,5 with $\varphi = 1{,}5 s^*$ can be determined boundaries of the range $x^* \pm \varphi$ to which are compared the original data x_i. Data protruding beyond this range are pulled on this boundaries and whole procedure is repeated. In any step (j) of an iterative procedure, after the modified value from the step (j-1) according to the conditions (4), is in turn fined a new mean value and new standard deviation of the sample

$$\bar{x}_{(j)} = \sum_{i=1}^{n} x^*_{i(j)} / n; \tag{7}$$

$$s_{(j)} = 1{,}134 \sqrt{\sum_{i=1}^{n} (x^*_{i(j)} - \bar{x}_{(j)})^2 / (n-1)} \tag{8}$$

Factor 1.134 is used when c = 1.5.

The resulting value s_j is used to calculate a new distance φ_j=1.5s to boundaries of inter-quartile interval and again data coming off as outliers are pulled on them, and the procedure as above is continued. Convergence of the algorithm is determined by comparing the calculated values $\bar{x}^*_j$ and $\bar{x}^*_{j-1}$ of the current and the previous iteration step. The procedure is repeated until changes of $\bar{x}_j$ and s_j between successive steps will be minimal. The procedure is stopped after j=m steps, where the difference of standard deviations $s_{(m)} - s_{(m-1)}$ for two successive steps is acceptably small.

Robust iterative double-criteria method IRLS has no defects of the median method. It allows in the calculation of the standard deviation of the sample also to include the outliers, i.e., data of the maximum absolute deviation, bringing them to the borders of the inter-quartile diapason of normal probability distribution of the data. An example of use this procedure in the inter-comparison measurements are presented in section 4. Numerical example of the data homogeneity conditions test and determination of limits of extreme deviations for small samples are given in [8].

4 Numerical Example

In this example, the mean value of measurement results of nine laboratories and its estimated uncertainty are calculated by two classic and two above robust methods. Results will be compared. Measurement data is taken from [2]. Nine laboratories conducted a joint experiment involving comparative measurements by a tested method to assess its accuracy. It was assumed initially that the credibility of all laboratory measurements are the same. From measurements made by tested method in n = 9 laboratories received are mean values $x_{1 \dots} x_9$ ordered below

$\underline{17.570}$ 19.500 20.100 20.155 20.300 20.705 20.940 21.185 $\underline{24.140}$.

Two underlined results $\underline{x_1}$ and $\underline{x_9}$ are the significant outliers. Results obtained by various methods are shown in Table 1

Table 1. Comparison of the results obtained by four methods

Method	For all data	Rejected x_1, x_9 by Grabbs crit.	Robust MAD$_s$	Robust iterative
Result value	$\bar{x}_0 = 20.511$	m =20.4	med=20.3	$\bar{x}_5^* = 20.412$
Std. uncertainty	$s_0 = 1.727$	s = 0.501	$s(x_9)$=1.045	$s^* = 1.039$

For the all 9 initial data $x_{i=}x_{i(0)}$ the mean value $\bar{x}_0 = 20{,}511$ and the sample standard deviation $s_0 = 1{,}727$. In the traditional model (cross-contamination) it is assumed that only valid observations are derived from a normal distribution. A consequence of that is to use the proper test, e.g. Grubbs test to find the outliers

$$G_{n\max} = (x_n - \bar{x}) / s \tag{9}$$

After rejection outliers $x_{1(0)} = 17.570$ i $x_{9(0)} = 24.140$, for the remaining data is obtained the average value $\bar{x}$ = 20,41 and much lower than previous the standard deviation s = 0.50 as results common to the whole experiment. Both are calculated from measurements in 7 laboratories only. These assessments are of the lower statistical reliability.

In the classical approach the average values $\bar{x}_0$, m calculated by both methods differ relatively little. Standard uncertainty s_0 of the data of all nine laboratories is very high. After elimination of two outliers by the Grabbs criterion, the uncertainty calculated for seven laboratories is almost 3.5 times lower. However, measurements are unreasonably idealized here. The reliability of the averaged data for 7 labs is decreasing as the formula $s(u_A)/u_A = 1/\sqrt{2(n-1)}$ (Table E.1 GUM [1]) showed that the relative standard deviation of the measurement uncertainty will increase from 25% to 29%.

For both robust methods values of the data grouping center are nearly similar.

Their uncertainties differ each other only by 9% and are between these two of the classical method. For the iterative method achieved is $s^*= 1.039 > s$. The mean value and standard deviation determined by this method is based on the data of all laboratories and seems to be as closer to the data which would be for a larger number of independent measurements treated as general population.

5 Summary

The rescaled median deviation method given in section 2, is very simple but it does not give correct results when the outlier is far from the rest of the data.

Iterative method of section 3 is more complicated, but easier to automate the algorithm. With the introduction of the threshold $\pm c\sigma$ decreasing sensitivity to data outliers, oriented is mainly to determine robust assessment of uncertainty.

Carried out in section 4 results of calculation showed the usefulness of the application of two criteria iterative robust method resistant to determine the statistical parameters of samples with a small number of data when they are taken from the general population of the assumed normal distribution, but include the results significantly different from the others. It allows you to more objectively assess the value of the result and the accuracy of the test methods.

The analysis shows that for the evaluation of results presented in controlled laboratories, should take into account the number of samples n obtained for the investigated objects. When a sample is of a small number of items, to evaluate the performance of results you can use the robust method of an iterative process of data with winsoryzation of outliers. In this case received is a much smaller variance and greater credibility than by the standard methods.

References

1. Guide to the Expression of Uncertainty in Measurement GUM. BIPM JCGM 100 (2008)
2. ISO 5725-2:1994 -Accuracy (trueness and precision) of measurement methods and results
3. ISO 13528-2005 Statistical methods for use in proficiency testing by interlaboratory comparisons. Annex C 1
4. Randa, J.: Update to Proposal for KCRV & Degree of Equivalence for GTRF Key Comparisons. NIST, GT-RF / 2005-04 Internet (2005)
5. Huber, P.J., Ronchetti, E.M.: Robust Statistics, 2nd edn., p. 380. Wiley (2011)

6. Wilrich, P.T.: Robust estimates of the theoretical standard deviation to be used in inter-laboratory precision experiments. Accred. and Quality Assurance 5, 231–240 (2007)
7. Belli, M., Ellison, S.L.: Implementation of proficiency testing schemes for a limited number of participants. Accreditation and Quality Assurance 2, 391–398 (2007)
8. Volodarski, E., Warsza, Z.: Applications of the robust statistic estimation on the example of inter-laboratory measurements, pp. 260–267 (2013) (in Polish)