
A New Statistical Tool Focused on Metrological Tasks

Eugene Charnukha

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.74872>

Abstract

A set of mathematical tools based on the principle of probability of origin are presented and intended to directly account for all a priori and experimental information. The principle of determining the probability of data origin, relatively the model of the experiment for evaluating the result of this experiment, is proposed. The application of this principle and its properties are described using the example of the trivial model of the direct experiment. Estimates of the result of the experiment are compared for various algorithms, including normative ones, and for various types of experiments.

Keywords: stochastic models of metrology, uncertainty, probability metrics, range measure, calibration experiment, repeated, multiple, work measurements, a priori information

1. Introduction

The key point of the text is the principle of the probability of the origin of the data. We believe that it is useful before exposition of this principle and its consequences spell out some general speculations about the situation in metrology. Metrology as a technology needs a simple, well-established and understandable procedure for implementing its tasks. Metrology as a business tries to canonize and protect its methods from strangers. These peculiarities prevent the use of new mathematical tools. Metrology in a narrow sense begins with the creation of the standard, continues by the construction of a calibration hierarchy, and ends with the calibration of the working instrument of measurement. Metrology in a broad sense is a component of the experiment everywhere where its main tools are used, namely, traceability to the standard and an estimation of uncertainty.

Uncertainty is estimated using statistical tools. The peculiarity of statistical instruments applicable in metrology is the essential role of a priori information in their work. The best way to obtain a priori information is a specially performed calibration experiment. In an ideal metrological experiment, the values of all model parameters are known and controlled except for one single parameter whose value is estimated.

Statistics without a priori information cannot be used as the metrological tool. But the origin of the a priori information can be different. For example, certain object does not in any way depend on the will of the observer, and, consequently, a calibration experiment is impossible. But it is possible to collect a lot of different data about this object and similar ones. Data can only be used to classify them and to monitor the evolution of the object. On the other hand, if we reformulate the accumulated database as a priori information for identifying an object class from new data, then this is already a metrological formulation of the problem. The estimation of the absolute value characterizing the object is difficult because there is no direct traceability to the standard. But recognizing an object and estimating the magnitude of relative changes from a small amount of data can be formulated as a metrological task.

Usually, the data of the working experiment on the subject of observation are not numerous, but there is a priori information obtained in the calibration experiment. It is assumed that by the time of the working experiment this information is still relevant. Comparing the data and the model, we can estimate the observed state of the object.

An effective method—to compare the model used and the available data—is to estimate the probability that the data is generated by a source corresponding to the model. This probability is interpreted, in particular, as an estimate of the reliability of a particular value of the investigated quantity, described in the a priori model as an adjustable parameter. In other words, as an argument for the criterion to choose, one of the many variants of the measurement model provides a description of the object under study.

In this text, an analysis of the features of traditional statistical tools [1] and some new tools to replace them is proposed. The dignity of new tools (in particular the rank measure) is significantly a better universality, but its disadvantage is a large computing expenses.

The rank measure was first proposed and intuitively grounded in [2]. In paper [3], it was formally justified. Some aspects of its application were discussed in Ref. [4]. Paper [5] describes the main tools and their applications for the method of converting the densities (MCD). In paper [6] the application of a rank measure to the type of experiment rarely used by metrology but widespread in technical disciplines is discussed. This is a simple interpretation of dynamic experiment. Its main features are as follows: enough data is collected, and a minimum number of observable factors are required to evaluate the values of many parameters of the model.

2. Models

Habitual models of the measurement experiment are constructed from the principal f and stochastic η components, formally $M(f, \eta)$. The principal component is a mathematical description

of the physical principle of the experiment. Usually, this is an arithmetic formula, but sometimes the algorithm decides equation or even numerical simulation.

The stochastic component is a description of the random (or considered to be) influence on the result of the experiment. Often, this description consists of a system of equivalent noise sources with some specified characteristics.

The components of the model are formalized as headings of procedures whose variables are divided into two parts—the variable values of which must be determined quite accurately by the time of the working experiment, and the variable X whose values are estimated from the data D is obtained as a result of the experiment $(M, D) \rightarrow X$, where both the experimental data and the evaluated variables can be either simple or complex data structures.

The main purpose of the model is to formulate a prediction. For metrological tasks, we set the value of the controlled parameters of the model, and from it we obtain a data structure modeling experimental data. Two modeling methods that can be compared with the definitions of probability have been distributed. The Monte Carlo method (MCM) is comparable to a countable probability, and the method of converting the densities (MCD) is comparable with the axiomatic probability.

In metrological statistics the most widespread one is the simple additive noise model (additive random error model) $d_i = x + \eta_i$, where d_i is the observed process; x is the value under measurement (measurand) (is constant throughout the experiment); and η_i is a random impurity, at each time of measurement i having a different value. It is the simplest model of a direct measurement experiment. It is also called the trivial model.

It is important that it is a priori known about a random component. It is usually assumed that only the form of distribution of probability of the source of chance is known. It is necessary to estimate the value of the constant component (as a shift parameter of a known distribution) over a small number of data affected by a random error with zero shift (for simplicity of interpretation) but with a scattering magnitude of unknown magnitude. It is also assumed that the time between measurements is so large that the data sampling elements are statistically independent.

3. Normative identification of the trivial model

3.1. Sectorial formula

A trivial model with an unknown scattering parameter in accordance with mathematical statistics and normative documents of metrology is identified according to the formula (we call it the sectorial formula) $\bar{x} = s(D) \pm kS(D)$, where \bar{x} is the estimate of the value of the measured quantity in the form of a confidence interval, D is experimental data, $s(D)$ is the statistics used to estimate the value of the shift parameter of the distribution given by the model, $S(D)$ is the statistics applied to estimate the scattering parameter and k is the coverage coefficient, which in general depends on the distribution law (both model and real) of the source of randomness, the number of repetitions of the experiment, both statistics, confidence and correction factors.

The property of the formula is illustrated in **Figure 1**. In this figure, by MCM the cloud of possible results of a multiple experiment is calculated and is delineated by means of a formula. The formula is linear, therefore divides the cloud of estimates into two regions by oblique boundaries.

The change in the coefficient of coverage will lead to a shift in the boundaries of the blue and red sectors, and a corresponding change in the confidence probability is due to a change in the ratio of the shares of estimates within and outside the confidence interval.

The advantage of the formula is that whatever the dispersion of the source of chance, you will still get your 95% of correct estimates. This is illustrated by the superposition of clouds with different dispersions.

The disadvantage is the strong dependence of the error probability on the standard deviation. If by will of chance the data is close, then the probability of error is large, greater than the confidence probability. If the data is very scattered, then the confidence interval is too wide, with that the actual probability of making a mistake is negligible. The confidence interval is located at the level value of statistics from the border blue/red to the border red/blue. But in the statistical limit, the confidence probability will be met. Intuitively, it is believed that, namely, the extreme values of the cloud of estimates are discarded, but in reality, it is not so. The paradox is that the probability of error is more there when the data seem better and vice versa.

The illustration is given for normal distribution and normative statistics. For other distributions and for other statistics, the scattering clouds of the results are different, sometimes quite bizarre. Coefficient of coverage should also have its own value different from Student; however, it is

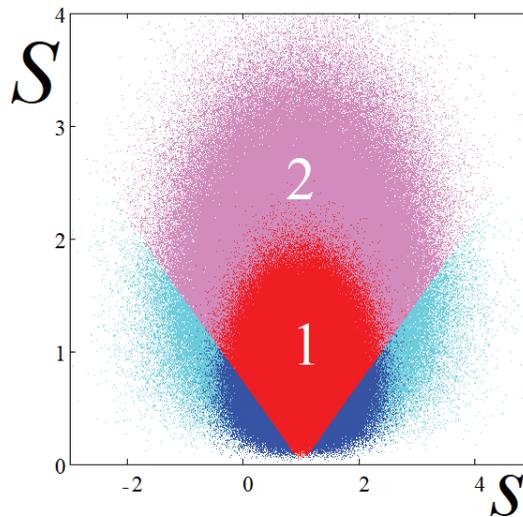


Figure 1. Clouds of scattering of results of estimates. The number of tests is 10^6 , the multiplicity of the experiment is 5, the source of chance has the normal distribution with $\mu = 1$ and $\sigma = 1$ and 2 (the notations in the figure by different transparency) and the color markings for the confidence probability of 0.95 are blue (erroneous estimates) and red (correct $\mu \in r$). Statistics are used (the arithmetic mean and the standard deviation) and the coverage coefficient is the Student's coefficient, now depending only on the number of repetitions and the confidence level.

quite simple to calculate. Here are just several simple illustrations. Let us replace the normal distribution to a very important uniform distribution. First, we apply to it normative statistics [Figure 2 (left and central)] and then more suitable statistics of extrema $s = \frac{1}{2}(\max(D) + \min(D))$ and $S = \frac{1}{2}(\max(D) - \min(D))$ [Figure 2 (right)].

Without going into numerical details, we give a few qualitative remarks on the illustrations given. Although the scale of both the distributions and clouds of assessments is comparable, coverage coefficients are distinctly different. It can be judged from the tilt of the colored borders.

Clouds differ not only in form but also in size. The most compact cloud gives set of a normal distribution with of normative statisticians [Figure 2 (left)] because this combination is optimal. The combination of a uniform distribution and normative statistics (central) is not optimal; hence, the cloud is scattered more. This loss of efficiency is not catastrophic, so this combination is used in practice. Normative statistics provide acceptable estimates for many finite distributions and many distributions with light tails, but there are such distributions where the efficiency is too small, for example, distributions with heavy tails. The combination of uniform distribution and statistics of extrema (right figure), although not optimally but somewhat more efficient than in the previous example. But in practice this combination is not used because the sectoral formula of the cloud cross section leads to an unacceptably overestimation of the confidence interval value. The reason is that the maximum cloud density is at the vertex, when, as in the previous examples, the maximum density is closer to the centres of the clouds. An effective algorithm for estimating the distribution of the scattering parameter could help, but because of the variability of the distribution form, mathematical statistic could not offer such an algorithm.

De facto, the distribution form and both statistics are used as a single set. The situation can be interpreted in two ways. On the one hand, having the form of distribution, we can choose or synthesize statistics more or less effectively. On the other hand, selecting statistics from a certain set of tools, we actually choose a class of distribution forms for which the statistics are still effective. However, neither the value of efficiency nor the form of distribution can be precisely determined.

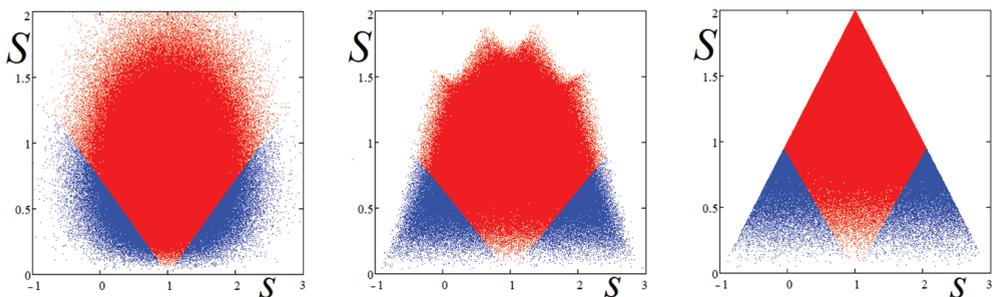


Figure 2. Clouds of scattering of results of estimates. For normal $N(x, \mu = 1, \sigma = 1)$ (left) and uniform $U(x, \min=-1, \max=3)$ (central and right) distributions. The number of tests is 10^6 , the multiplicity of the experiment is 5 and the color markings for the confidence probability of 0.95.

3.2. Corrections coefficients

The normative tool has yet a problem that we call a mysterious amendment to deviation. Deviation is recommended to be used not in a pure form, but with a correction coefficient (the so-called standard deviation). It is explained that this amendment allegedly eliminated deviation from the dispersion of the normal random source. But very few noticed that this is not quite true.

Firstly, the distribution of deviation is asymmetric, its form changes, and is especially strongly at small amounts of repeated experiments. And only to an infinite number of experiments it approximates to normality and, accordingly, to symmetry.

Secondly, because of the nonsymmetric form of deviation distribution, it is not entirely clear in which its characteristic should be adjusted. It is customary to correct the mode, but with the same success, it is possible to correct a centre of gravity or some kind of composite criterion composed of the moments of this distribution.

Thirdly, even for the mode, the recommended corrections only partially eliminate the problem. The reason lies in the desire to describe the correction factor by a simple formula. While its magnitude is simply calculated, the result does not fit into any of the proposed theoretical constructions (Figure 3). The reason is the complex and contradictory changes in the form and position of the cloud of estimates as the number of repeated experiments is changing.

The idea of the correction is that, a priori knowing its magnitude, we correct the estimate made by the statistics that measures the scattering parameter so that in the statistical limit the estimate coincides with the value of the dispersion. The question arises: what for? The quality of the estimate of the measured quantity is determined by the sectoral formula, and the coefficient

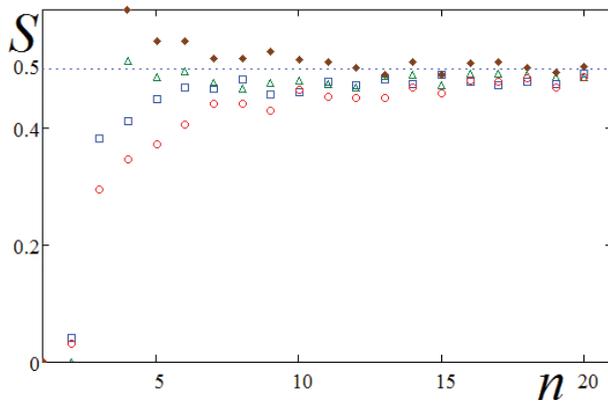


Figure 3. Estimates of the scattering parameter and the effect of corrections as a function of the number of repetitions of the experiment. The source of randomness is the normal distribution with $\mu = 2$ and $\sigma = 0.5$. The statistics for estimating the scattering parameter is the deviation. MCM is used for obtaining data by two 10^7 tests. Each point is the result of an independent experiment. Legend on the figure field: ○ is estimate without correction, □ with correction factor $\sqrt{\frac{n}{n-1}}$ (standard deviation), △ with correction $\sqrt{\frac{n-2}{n-3}}$ and ◆ with correction $\sqrt{\frac{n-1}{n-3}}$.

of coverage of which is calculated even more easily than the correction. A reasonable way is to abandon the amendments and the coefficients of coverage numerically computed, but this will no longer be the coefficients of the Student.

The sectoral formula is useful, but the rank measure copes with similar tasks of metrology better.

4. The principle of measuring of probabilities of origin

The principle says that the important instrument of metrological research should allow to estimate the probability of obtaining a certain sample of data from the selected model.

According to the principle—using the model and experimental data—the joint probability distribution for all values of each of the estimated variables is calculated. Each point of this distribution is interpreted as the probability that the data is obtained in accordance with the model and, moreover, with specific values of its parameters. Evaluation of the result of the experiment is given as $\hat{X} := \{(x, p)\}$ (the value of each of the estimated variables, the corresponding probability of this value). Of course, differences in the parameters of the model lead to different probabilities for a particular value of the evaluated value; the same can be said if the model is the same and the experimental data are different.

The task of constructing the estimation algorithm is solved in the general form of both MCM and CDM. The results are comparable, although the algorithms are different. To solve this, we need a consistency of the numerical model and also a metric for the data structures that model the results of the experiment.

Formally, this sequence of operations must be performed: $\bar{x} \xrightarrow{dis} \{x\} \rightarrow \{M(x)\} \rightarrow \{Pr|x\} \rightarrow \{\mu(Pr, D)|x\} \rightarrow u(x)$. The range of possible values of the estimated parameter \bar{x} must be broken one way or another into a set of possible values $\{x\}$. Using the model for each possible value, a prediction of the possible data values $\{Pr\}$ (it also is a set) should be obtained. Each prediction is compared with the experimental data by means of the metric μ . The results of the comparison are collected in the uncertainty function $u(x)$. And, only after this based on the uncertainty function, simplified formal estimates are performed.

The numerical consistency of the model is understood as the ability of the model (if all the adjustable variables are given) in a numerical experiment to generate model data indistinguishable (quite similar) from the data obtained in the experiment.

The metric should evaluate the magnitude of the difference between the same type of data in both experimental and simulation origin. The metric is constructed based on the modeling method and also on features of the application where it is used.

When using MCM, the ‘natural’ metric consists of counting the (approximate) matches of the data set to be checked and the extensive database generated for the given parameter values. In order to estimate the probability to the value of the parameter being evaluated, the model is launched many times (at example N), at this value of the parameter x , and the fraction of coincidences

with the experimental data is counted in this series of numerical experiments, formally $\{M(x) \rightarrow Pr = D\}_N \xrightarrow{\text{count}} C$; sequence metric is $\mu \stackrel{\text{def}}{=} \frac{C}{N} |x$. Repeating a series of experiments for other values of the same parameter, we obtain the required estimate, which looks like a density of probability distribution. If it is required to evaluate number of the factors more than have been considered, then they are evaluated by the same algorithm by performing similar operations.

When using MCD, the estimation algorithm solves the deconvolution problem in the general formulation $M(u(x)) \rightarrow Pr = D \rightarrow u(x)$. The model parameters are specified as densities for both the stochastic component and the parameters to be evaluated (as objective function $u(x)$). The prediction of the model will be obtained as a certain n -dimensional density describing the possible values of the data. It is required to choose both the dimensions and the form of the density of the evaluated parameters so that the metric points out the maximum similarity of the experimental data and the prediction of the model. The natural metric in this approach is the magnitude of the overlap of the prediction density and the actual experimental data, namely, $\mu \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} Pr(x, D(x)) dx$ in general and in a case of point data as $\mu \stackrel{\text{def}}{=} Pr(D)$.

Obviously, the solution in general form, without taking into account the structure of the model and data, is very labour-consuming by both methods. But for simple models and data, the situation is so simplified that it leads to simple algorithms.

5. Rank measure

The concept of a rank measure was proposed years ago and analyzed from both the intuitive and the formal points of view. Here, we propose an approach which can be regarded as justification as rationale in constructive style.

Statement. For a trivial metrological model, if the source of randomness is described only by its distribution, and the data elements are statistically independent, the implementation of the ‘principle of measuring the probability of origin’ leads to a simple ‘rank measure’.

Proof. From the assumption of data independence, the value of the metric is independent of the permutation of the data elements in the data sample used to identify the trivial model.

In fact, suppose that for two data samples of the same length, all elements are the same. Should the metric distinguish them? It is obvious enough that it is not necessary to distinguish and there is no possibility to do this.

Now, in each sample, one element by element of a different but identical value and in the same position is replaced. As before, the samples are indistinguishable.

Now, in one of the data samples, we change the positions of any two elements. If the data elements are equal, then the samples are indistinguishable. If the data elements are different, then the samples can be distinguished, but should this be done?

If the data is independent, then any position of each element is equally probable. Thus, the probability of origin is unchanged. The metric must be such that a simple permutation of data elements within one of the samples does not change the value of the metric. Consequently, neither the number nor the step of internal permutations on the value of the metric is affected.

This creates an equivalence class for data samples formally different as records of the data acquisition process, but within the class, those samples are indistinguishable by the metric. Data sample after simple sorting in ascending order (rank statistics) is a natural representative of each of these classes and can be used instead.

Each of the data sample elements $d_k | k = 1 \dots n$ in its ordered sample has its own order density of distribution $p_{v_k}(x) = \frac{n!}{(k-1)!(n-k)!} P(x)^{k-1} (1-P(x))^{n-k} p(x)$ different from the distributions in each of elements in other positions, where $p(x)$ and $P(x)$ are the probability distribution density and the cumulative distribution function of a model random source, respectively.

The probability of the origin of the value of each data element d_k is calculated from the corresponding order distribution density as $p_{v_k}(d_k)$. The probability associated with the entire sample of data $\{d_k\}_n$ is naturally calculated as the multiplication of the origin probabilities of each of the elements of data $m(\{d_k\}_n) = \prod_{k=1}^n p_{v_k}(d_k)$ because the event of obtaining a sample of data is considered single. We call this result the 'rank measure'. *End of proof.*

An important feature of the algorithm for identifying a trivial statistical model with the assumptions made is that there is no need to explicitly define the metric. You can immediately go to the estimation of the demanded probability of origin by comparing the prediction of the model in the form of the densities of the distributions of each of the data elements and the sorted experimental data. The formula of a rank measure can be dissected to three factors:

$$m(\{d\}_n, \mu, \sigma) = \left(\prod_{k=1}^n \frac{n!}{(k-1)!(n-k)!} \right) \times \left(\prod_{k=1}^n \left(P(d_k, \mu, \sigma)^{k-1} (1 - P(d_k, \mu, \sigma))^{n-k} \right) \right) \left(\prod_{k=1}^n p(d_k, \mu, \sigma) \right)$$

Their interpretation is obvious: the latter is the formula of the likelihood method, the second is the correction to the likelihood method and the first is the normalizing factor. For this reason, the rank measure can be considered as a corrected likelihood method.

The rank measure is the simplest solution of the identification problem for the simplest model that can be obtained within the framework of calculating the probability of origin. The reason is in the availability of an analytical formula for calculating the model's prediction. For more complex models, there is no such formula. At least we need to compute the prediction of the model numerically. Studies were conducted and it was revealed that for two important particular models' explicit formulation of a metric is not required too. It is multifactor expansion of the trivial model and model where the parameters of the dynamic deterministic function are identified against the background of noise.

6. Using rank measure in metrology

In this section we give examples of the application of a rank measure in some basic types of experiments. Let us compare the results obtained by algorithms using a rank measure and the results of normative algorithms. In this section, several varieties of direct measurement experiment and one generalization are considered.

6.1. Calibration experiment

Calibration experiment is main type of experiments in metrology. There is no means of measurement which one way or another would not undergo calibration. The purpose of the calibration experiment is to compare the measuring instrument with the standard, collect the data and describe a correction function that will be used as a priori information in the working measurement experiment.

In the calibration experiment, the values of the standard and the readings of the measuring instrument are juxtaposed. In this case, the measuring means is used to estimate the value of the standard used. The results are collected and form a data structure, for example, as in **Figure 4** (left).

The correction function is constructed as a regression at the calibration data. The obvious representation is the density stretched over the whole measurement range and accumulating all the calibration information [**Figure 4** (right)]. The more calibration data and the more carefully the regression, the more reliable the results. The replacement of the abscissa axis from the value of the reference value to the unknown means that the probability of the value of the standard corresponding to the experimental data is estimated.

The quantity and quality of the information collected in the calibration experiment and the information stored in the correction function largely determine the capabilities of the working measurement experiment. Although modern regulatory documents allow the use of a correction function in this form, for example, IEEE 1451, historically, the systematic error is eliminated separately, and the uncertainty of the measurement tool is described as an interval approximation of the density function in the form of a two-term formula or its simplifications.

6.2. Single experiment

The correction function is used in a working experiment to fully evaluate the result of the experiment. If the data comes in the form of a point estimate (number), then the corrected measurement result is calculated as cross section of correction function, which is interpreted

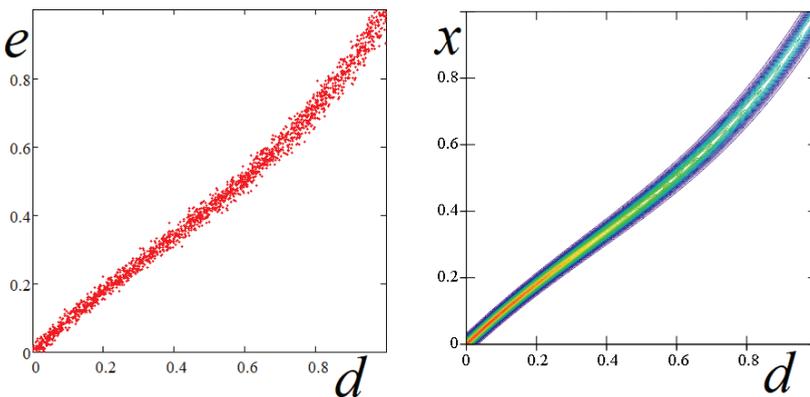


Figure 4. The structure of the calibration data in graphical form (left) and the correction function (right) as the regression of these data $p(x) = \hat{f}(d)$.

as the distribution density of the possible values of the measured value. That is, the systematic error is eliminated, and an estimate of the uncertainty of the values of the measurand is given (Figure 5).

On the other hand, the data may already contain a description of the uncertainty, for example, in the form of a probability density $g(d)$. In this case, the joint probability density distribution of the data and the estimate is calculated $p(d, x) = g(d) \cdot (\hat{f}(d) = x)$. Note that this is a joint distribution and does not refer to independent distributions because of the large correlation (Figure 6). The projection of this joint density will lead to a final evaluation of the measurement result $p(x) = \int_{-\infty}^{\infty} (g(d) \cdot \hat{f}(d)) dd$. Thus, a complete and natural synthesis of the available a priori and a posteriori information about this measurement experiment was made without any assumptions and approximate calculations.

6.3. Multiple experiment

Measuring the same physical quantity repeatedly, in principle, we get the opportunity to deal with errors and thereby improve the accuracy of the evaluation of the result. The problem of normative statistical tools is that it was far from always possible to use data efficiently, and sometimes efficiency was reduced to zero. From this point of view, since the rank measure uses the form of a specific distribution, it will always be optimal in efficiency with respect to this distribution.

6.3.1. The scattering parameter is unknown and is estimated from experimental data

The greatest effect of using the rank measure as statistics for estimating the distribution parameters is observed in a multiple experiment with unknown scattering. According to the principle of probability of origin, the probability of obtaining experimental data from a

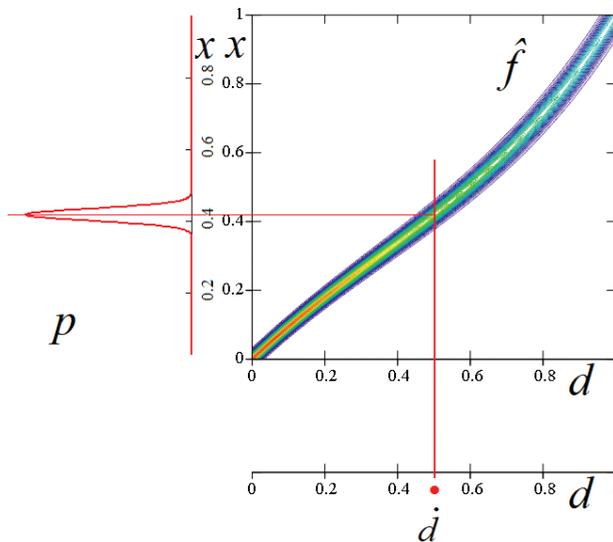


Figure 5. The transition from point experimental data \hat{d} to a full evaluation of the experimental result, taking into account a priori information about the property of the measuring instrument $p(x) = \hat{f}(\hat{d})$.

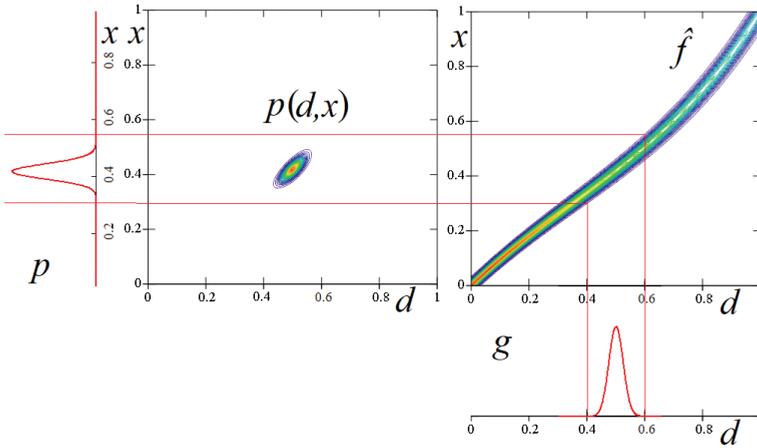


Figure 6. The transition from experimental data with uncertainty $g(d)$ to a full evaluation of the experimental result $p(x) = \int_{-\infty}^{\infty} p(d, x) dd$. In the middle of the figure, an intermediate result is presented $p(d, x) = g(d) \cdot \hat{f}(d)$.

random process model with a known form of the distribution density is estimated, but the parameters of the shift μ and scattering σ must be estimated from the experimental data. Note that the form of the distribution can be arbitrary, but it shall be a priori known, for example, obtained from a calibration experiment. We seek a joint distribution of the values of parameters that are estimated $u(\mu, \sigma) = m(\{d\}_n, \mu, \sigma)$, and the distribution density of error source is also described in terms of the values of these parameters $p(x, \mu, \sigma)$.

For example, we estimate the shift parameter from the data for normal and uniform distributions $\{d\}_n = \{-0.125, -0.044, 0.183, 0.349, 0.404\}$. It is convenient to designate the desired joint distribution as $u(\mu, \sigma) | p$ with an explicit indication of the distribution form used in the model and interpret it as a function of the uncertainty of estimates with respect to the distribution used (**Figure 7**).

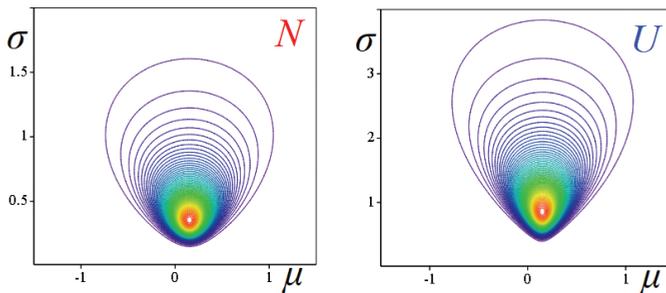


Figure 7. Uncertainty functions $u(\mu, \sigma) | N$ for the normal distribution (only the form is used) and $u(\mu, \sigma) | U$ for the uniform distribution. For a correct comparison, distribution densities are scaled to the law 2σ .

The uncertainty functions for different distributions differ in varying degrees by form but mainly by the scattering estimate. The distributions used in the example are both symmetric for this reason, and the difference in the estimation of the shift parameter is small.

Now, it became possible to move from a joint estimation of parameters to only an estimate of the shift parameter (usually interpreted as an estimate of the measured quantity). At this stage, it is possible to take into account a priori information about the scattering parameter. This information can be different. One of the polar cases is its complete absence; the scattering can be any $u(\mu) = \int_{-\infty}^{\infty} u(\mu, \sigma) d\sigma$ (Figure 8).

If, for joint uncertainty function, the influence of the form of the model distribution is obvious, then the integral estimates of only the shift parameter differ insignificantly. Small differences can be interpreted as evidence of the prevalent thesis ‘if there is a small number of data the form of the distribution is unimportant’. More precisely, when identifying only the shift parameter for a small number of data, the form of the distribution has no important significance and does not introduce significant errors in addition for a wide class of distributions. However, it is possible to construct counterexamples that show that this is not always so, for example, using distributions having a significant displacement.

The form of the uncertainty function of the result for a number of reasons has heavier tails than the original distribution. Briefly, there are two main reasons. There is still a high probability of obtaining compact data from the distribution with a large value of the scattering parameter, which heavies the tails of the uncertainty function. On the contrary, the probability of compact distributions is concentrated in a small space, which leads to a high probability density near the vertex of the uncertainty function and sharpens it.

Now, we can write an interval estimate of the measurement result as a quantile of the uncertainty function. For the confidence probability of 0.95 by the normal distribution model, result estimation with uncertainty is 0.153 ± 0.869 and by the uniform distribution model is 0.149 ± 0.94 . Uncertainty function has less scattering than the original distribution (at example for normal distribution ± 1.96 and for uniform ± 2.0), which is actually the goal of increasing the multiplicity of the experiment. The recording of the result by the form is the same as the

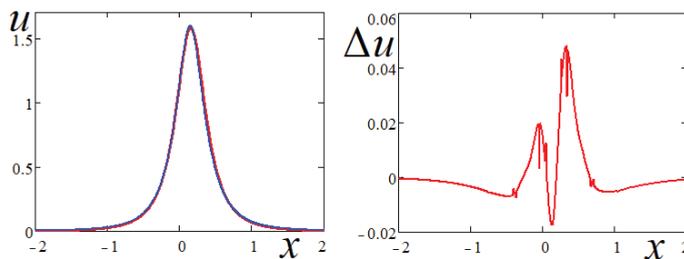


Figure 8. The uncertainty functions of estimating the shift parameter (left) and their difference (right). The notations on the left figure are a red line for the normal distribution and blue for the uniform, respectively. For a correct comparison, the uncertainty functions are normalized, which is interpreted as an assumption of the validity of both models simultaneously.

normative one, but in fact it has a more rigorous meaning. Tails of joint distributions (as well as clouds of estimates) are cut vertically, but not by the sector as in the normative case.

6.3.2. The scattering parameter is known fully or partially

There are many cases when the scattering parameter is known a priori with greater or lesser accuracy. The direct way to take into account information about the value of the scattering parameter is to solve the estimation problem for an unknown parameter and only then to use a priori information $u(\mu, \sigma) \xrightarrow{\sigma^2} u(\mu)$. The algorithm for solving the problem formally depends on the form of the representation of this information but at the heart of all algorithms lies an integral that somehow projects the joint uncertainty function to the shift parameter uncertainty.

The most often known is the range of possible values of the scattering parameter σ . The solution reduces to a simple integral $u(\mu) = \int_{\sigma} u(\mu, \sigma) d\sigma$. Two polar variants are evident. It is complete ignorance $\bar{\sigma} = \pm\infty$ and exact knowledge $\sigma = \sigma \pm 0$ which are solved analogically. The case of a known density distribution $g(\sigma)$ of the scattering parameter $u(\mu) = \int_{\pm\infty} g(\sigma)u(\mu, \sigma) d\sigma$ is slightly more complicated (**Figure 9**).

6.3.3. Repetitive experiment

Under favorable conditions, instead of the joint uncertainty function of the parameters, one can use the fact that the correction function itself is a distribution. Consequently, one complete correction function can be replaced by a set of ordinal correction functions with the same external characteristics. This is done either experimentally in a calibration experiment or analytically from the formulas of the densities of ordinal distributions for each value of the measured quantity in the entire measurement range. We obtain a family of correction functions passing along and partially overlapping $\{f_{i/n}(d)\}_n$. Each of these functions is used to correct its data element from an ordered data sample [**Figure 10** (left)]. The uncertainty function with

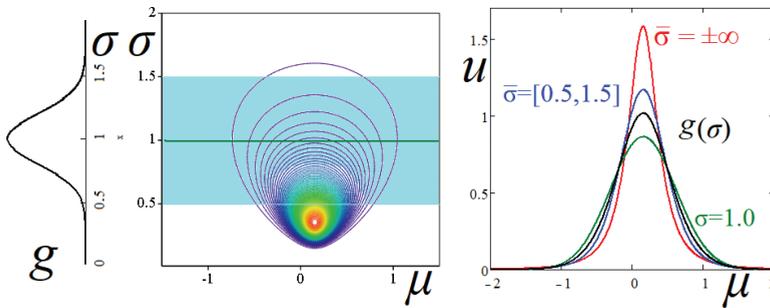


Figure 9. Illustration of the use of a priori information on the scattering parameter in order to convert joint uncertainty to uncertainty function of the shift parameter. Legend on the right picture field: The green line is the exact knowledge of the scattering parameter $\sigma = 1.0$; black line corresponds to the normal distribution $g(\sigma) = N(\sigma, 1, 0.2)$ (graph on the left); blue line corresponds to the interval number $\bar{\sigma} = [0.5, 1.5]$; and red line for complete ignorance $\bar{\sigma} = \pm\infty$. Legend on the Centre picture field: The joint uncertainty distribution with respect to both parameters is the same as in **figure 7** on the left; the green line is the exact knowledge of the scattering parameter; and blue wideband is the image of the scattering parameter interval.

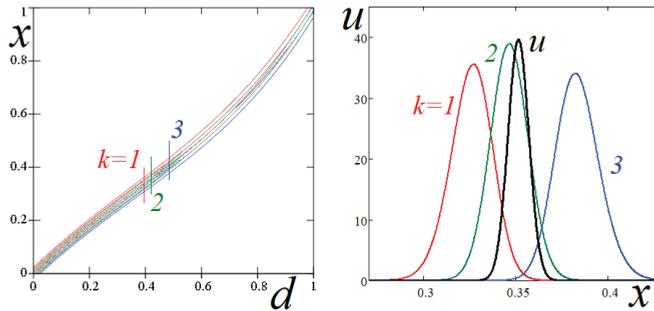


Figure 10. Illustration of the use of the set of ordinal correction functions. On the left is a set of correction functions for a triple experiment, and on the right is an example of estimating the value of the measured parameter for data {0.4, 0.41, 0.44} each of the three ordinal estimations (color lines) and resultant estimation (black line).

respect to the shift parameter is calculated directly as the product of these functions $u(x) = \prod_{k=1}^n f_{i_k}(d_k)$. The formula is interpreted as uncertainty of a repetitive measurement of the same physical quantity by an imperfect means of measurement but with a known scattering parameter of its stochastic model. It is assumed that new sources of randomness, not accounted for by the calibration experiment, are not added. This is what distinguishes the repeated experiment from a multiple experiment.

This tool is more refined because it can take into account the change in the form of the distribution of the correction function for different elements from the data set. But it is more vulnerable because it does not provide for any additional sources of randomness that cannot be taken into account in the calibration experiment.

The situation where the scattering parameter is known sufficiently accurately is not so rare, although it is hidden inside the measuring instrument. At best, the user can adjust the ‘accumulation time’. If the accumulation of information is made in digital form, then this is a direct analogue to the number of repeated measurements, but in the analogue form, the accumulation is not fundamentally different from the effect of repeated measurements.

6.3.4. The uncertainty of the experimental data is known

The abstraction of point data is very useful from a practical point of view. Its application seriously simplifies both calculations and their interpretation, and the results are of quite satisfactory quality. In most cases, it should be used. However, in the strict approach, each data element must be assigned to its own individual uncertainty. For many applications, including the case of multiple measurement experiments, an adequate form of describing the uncertainty of the experimental data is the probability density of the obtained value $D := \{g_k(\delta)\}_n$ interpreted as the reliability of the point fragment of this estimate. It is because the basis is the single experiments in which the initial data are obtained.

Normative documents including GUM solve this problem taking into account uncertainties apart, for example, preliminarily dividing the uncertainties into type A and type B and then

combining them in a specific way. The method is simple but strictly adequate only for normal distribution and simple models. For distributions similar to normal distribution, the deterioration in the result still is quite acceptable.

To strictly take into account the uncertainty of the measuring instrument, it is sufficient to slightly upgrade the rank measure to.

$$m(D) = \int_{\mathbb{R}^n} (m(\text{sort}(\{\delta\}_n)) \cdot \prod_{i=1}^n g_i(\delta_i)) d\delta^n.$$

The formula is interpreted as an n-fold integral of a rank measure from deviation to point data with their joint probability. The complexity of applying the formula is the multiplicity of the integral and the need to constantly check the order of the data if the density of the data distribution overlaps. When the distribution density of data is reduced to the delta function, the upgraded measure reduces to the original measure. The delta function is the model of point data. From this point of view, uncertainty function for point data is the most likely, but for data deviations it is a less likely alternative.

In a more general case, all sources of uncertainty are taken into account in a natural way when calculating the model's predictions and when a comparison of the prediction and an adequate data model is made.

Let us explain this with an example (**Figure 11**). The data is the same as for **Figure 10**. We will supplement the data with uncertainty ± 0.05 . The uncertainty is the same for all data elements, but it can also have an individual value. The law of distribution of uncertainty will be assumed to be uniform. The model of the measurement experiment being studied differs from the trivial model only in the presence of two sources of randomness. One source has a normal distribution law, for example, the error of manufacturing samples from the same material whose property is being investigated. Another source has a uniform distribution of, for example, uncertainty of a digit measuring instrument.

The work of the algorithm can be interpreted as the creation of a film. Each frame is an estimate of the parameters from a given set of point data $\{\delta_k\}_n$. Each frame is similar to the one in **Figure 11** (left). The difference between frames is a consequence of the differences in the data. Data is selected from specified distributions either randomly (MCM) or according to a

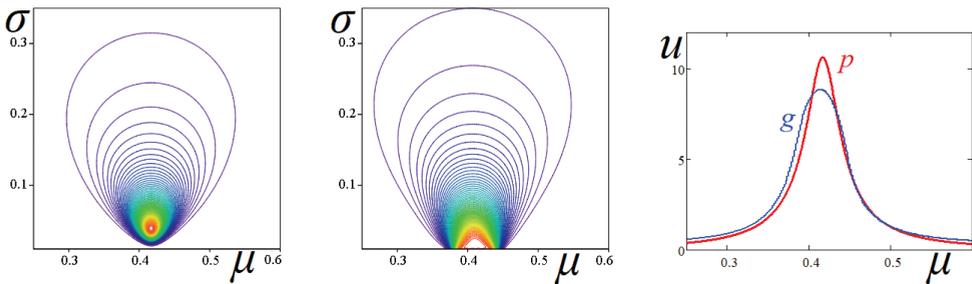


Figure 11. Illustration of identification of a trivial model with two different sources of randomness. The left figure is obtained for point data, and the central figure is obtained for data with uncertainty. The right figure shows the uncertainty functions of the estimates for the two preceding figures: p , without data uncertainties, and g , with uncertainties.

regular grid (MCD) (in example used 12 grid knots for each data distribution). Each frame corresponds a probability that is calculated by formula $\prod_{k=1}^n g_k(\delta_k)$. When all data distributions are uniform and equal, then probabilities of frames are equal. At the end of the algorithm, all frames (in example $12^3 = 1728$ frames) are summed according to their probability. The result is shown in **Figure 11** in the centre.

The uncertainty is large compared to the distance between data; hence, the probability of accidental coincidence of data is large, which leads to a touch of the uncertainty function of the estimates to the abscissa axis [**Figure 11** (centre)]. The uncertainty of the data, as it was, ‘smears out’ the uncertainty function of the estimate. Uncertainty is greater in all respects but especially strongly affects the top of the uncertainty function of the estimate of the measured parameter and often changes the form of the evaluation function.

This allows us to build a logical chain from the interpretation of data by interpreting possible estimates to the final estimate of the uncertainty of the measurand. For example, $D \xrightarrow{\hat{f}(d)} \{d(x)\} \xrightarrow{\{p(x)\}} \{u(x)\} \rightarrow \bar{x}$, where D is the initial experimental data given in point form, $\hat{f}(d)$ is the correction function obtained from a calibration experiment, $\{d(x)\}$ is the data in the form of densities that have adjusted by the calibration experiment, $\{p(x)\}$ is the set of admissible types of distributions in measurement model, $\{u(x)\}$ is the set of densities of estimates measurand and \bar{x} is the final evaluation, for example, obtained for the worst case.

6.3.5. Multifactor multiple experiment

The purpose of the multifactorial experiment is to estimate the value of several quantities in the form of a joint uncertainty function by factors. The number of factors considered varies easily, so in the examples we confine ourselves to two. And so, $u(\xi, v)$ is estimated by the data structure $\Xi\Upsilon$ for each of the factors. The result of the evaluation and the complexity of the algorithm are essentially determined by the relationships within the data structure. The simplest solution is obtained when the data for different factors are not related to each other. For example, a data structure is simply a list of independent data differing only belonging to its factor $\Xi\Upsilon := \Xi, \Upsilon$. The solution consists of a multiplication of uncertainty functions for each of the factors calculated independently $u(\xi, v) = u(\xi)u(v)$. The number of data for each factor can be different.

Another solution is obtained if the experimental data are obtained synchronously $\Xi\Upsilon := \{\xi, v\}_n$ | $\xi \in \Xi, v \in \Upsilon$. If the statistical relationships between the factors do not manifest themselves $r(\xi, v) = p(\xi) \cdot g(v)$, then it is possible to express the densities of order distribution $r_{i,j}(\xi, v) = p_{i,j}(\xi) \cdot g_{i,j}(v)$.

For example, if the multiplicity of experiment is 3, the number of factors is 2, $p(\xi)$ is a normal distribution and $g(v)$ is the uniform distribution, then the figure of the set of ordinal distributions will look like in **Figure 12**.

The rank measure is constructed as follows. The data structure (in the example this is three data pairs) is ordered by one of the factors, for example, by ξ . This predefines the selection of the columns of the set of ordinal distributions. The rows are selected in accordance with the order of the data for the second factor. As a result, for each experiment of the nine distributions, three will be chosen. Using them as a function of the data values, we get three probabilities for each

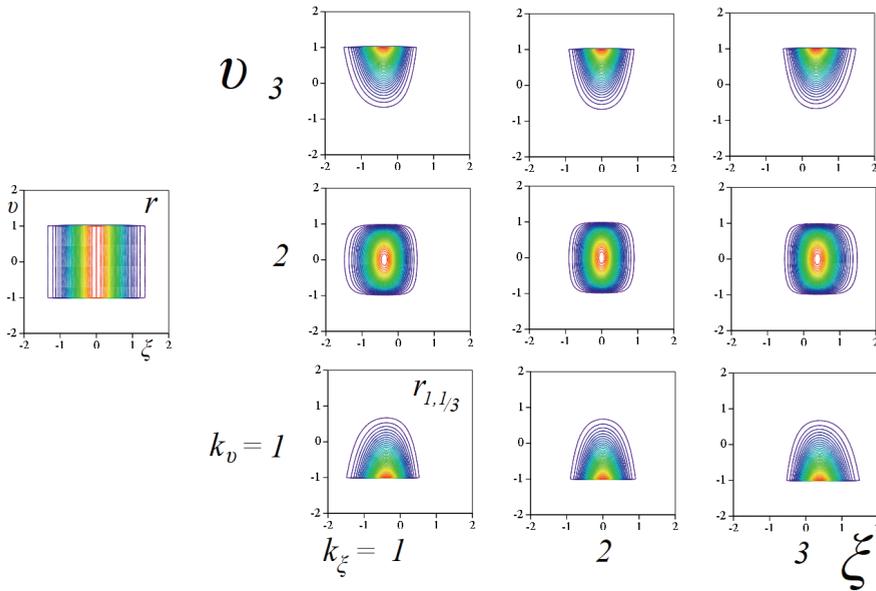


Figure 12. Initial joint distribution and set of ordinal joint distributions.

rank. Multiplying them we get the value of a rank measure. You can take advantage of this in a working experiment when the data about the same physical quantity comes in completely different ways.

For example, let's use the model whose distribution is shown in **Figure 12**. The received data is $\{(-0.5,0.5), (0.1, -0.4), (0.5,0.7)\}$. Their order is $\{(1, 2), (2, 1), (3, 3)\}$. The values of probabilities from the densities of ordinal distributions are $\{(0.578), (0.841), (1.114)\}$. Hence the value of a rank measure is 0.542. Next, it is possible to identify the shear and scattering parameters of the model in the usual way.

In the event that the statistical links between the factors are significant, the task is solved only numerically. For MCM, this is a direct numerical experiment. MCD is a search for direct and inverse transformations of such that make the distribution of the model independent by factors.

6.4. Indirect experiment

In order to pass from the model of direct measurement to the model of the indirect measurement experiment, it is necessary to replace the measurand of trivial model by a more complex measurement principle model $x = f(\xi, v)$, where ξ and v are the quantities measured from the direct measurement experiment. In a simple formulation, the problem of an indirect experiment consists in calculating the uncertainty function of the new measurand $u(x)$, starting from the uncertainties obtained from the experimental data $u(\xi)$ and $u(v)$. As a rule, the problem is easily solved by both MCM and MCD.

Although in the natural sciences and in technology one can find very complex principal models of the experiment, metrology strives to avoid indirect experiments. This is achieved through the creation of new standards and the construction of suitable calibration schemes (calibration hierarchy). Even if the measurement tool uses inside the complex indirect model but being calibrated in the target units, then it realizes direct experiment. All that metrology can afford is the use of an indirect experiment as a temporary means in cases where a direct reference to the standard is not yet possible. Of course, one can complicate the formulation of the problem of indirect experiment in different ways, for example, in the analogy of Section 6.3.5, complicating the data structure, but it is unlikely that metrologists will be interested in this.

7. Conclusion

The tools that metrology now uses have been created by statisticians at the beginning of the last century. By the middle of the century, metrology had mastered them. Over the years, the goals and circumstances of their creation and some of the properties have been forgotten. This creates some misunderstandings when interpreting the results of their application. Attempting to implement the GUM has been useful by simplifying and standardizing their application, but the tools themselves remained the same.

As a result of the application of new tools, a direct and obvious chain of information gathering and use is built up in the performance of metrology tasks from calibration to the final result.

Author details

Eugene Charnukha

Address all correspondence to: charnukha@tut.by

A. V. Luikov Heat and Mass Transfer Institute, National Academy of Sciences of Belarus, Minsk, Belarus

References

- [1] Uncertainty in Measurement (GUM). <https://www.bipm.org/en/publications/guides/>
- [2] Chernukho EV. Estimation of arbitrary-distribution parameters from the data of a repetitive experiment. *Journal of Engineering Physics and Thermophysics*. 2010;**83**(2):431-437
- [3] Chernukho EV. Substantiation of the rank measure as an efficient statistic for estimating distribution parameters of arbitrary form. *Journal of Engineering Physics and Thermophysics*. 2012;**85**(1):239-248

- [4] Chernukho EV. Uncertainty principle in metrology and statistics: Algorithms and applications. *Journal of Engineering Physics and Thermophysics*. 2014;**87**(2):498-506
- [5] Chernukho EV. Analysis of uncertainties with the example of direct and inverse heat metering problems. *Journal of Engineering Physics and Thermophysics*. 2016;**89**(2): 518-527. DOI: 10.1007/s10891-016-1405-9
- [6] Chernukho EV. Regression algorithm using the rank measure. *Journal of Engineering Physics and Thermophysics*. 2015;**88**(4):1034-1043. DOI: 10.1007/s10891-015-1282-7