Michał URBAŃSKI[1], Kinga WÓJCICKA[1], Paweł WÓJCICKI[2]

Warsaw University of Technology, Faculty of Physics (1), Faculty of Mathematics and Information Sciences(2)

doi:10.15199/48.2022.12.37

Probability to possibility transformation as a way to standardize description of systematic and randoms errors

Abstract. *From the point of view of the probabilistic model, measurement is described as a selection from sample set (set of elementary events). In the framework of fuzzy variable approach the possibility measure describes the degree of adjustment of the standards to the object being measured. On the other side fuzzy approach is better than probabilistic to describe the measurement process. To comply with the GUM recommendations, a transformation of probability to possibility is proposed, in which expanded uncertainty is maintained. In this way, both types of uncertainties are determined within one fuzzy model.*

Streszczenie. *Z punktu widzenia modelu probabilistycznego pomiar opisujemy jako losowanie ze zbioru zdarzeń elementarnych. Pomiar w języku zbiorów rozmytych można opisać jako proces dopasowania do mierzonego obiektu odpowiedniego wzorca. Miarą jakości dopasowania jest stopień zgodności tego dopasowania. W celu stworzenia modelu zgodnego z rekomendacją Przewodnika Wyrażania Niepewności Pomiaru proponujemy transformację rozkładu prawdopodobieństwa na funkcję przynależności zbioru rozmytego zachowującą niepewność typu A. W ten sposób oba rodzaje niepewności opisuje się jednym modelem z miarą rozmytą.* ***(Transformacja prawdopodobieństwo–możliwość jako sposób na ujednolicenie opisu błędów systematycznych i losowych)***

Keywords: fuzzy sets, fuzzy variable, uncertainty, theory of measurements, probability to possibility transformation
Słowa kluczowe: zbiory rozmyte, zmienna rozmyta, niepewność, teoria pomiarów, transformacja prawdopodobieństwo - zbiór rozmyty

Introduction

In accordance with the recommendations of the „Guide to the Expression of Uncertainty in Measurement" (in short GUM [4]) estimation methods of uncertainty can be divided into two groups depending on the source of information: A – from statistical analysis of repeated measurements, i.e. using a posterior distribution, B – from analysis of other data basis on a-priori distribution.

To mathematically describe these two methods, some analysis of assumptions of a probabilistic model will be done. We assume here that type B should be treated as an expert type decision. For the mathematical formalization of expert decision-making [15] fuzzy logic was introduced.

Measurement one can define as assigning values, most often numbers to the objects, what is described as a mapping:

$$\Phi : \Omega \to \mathbb{R}, \tag{1}$$

where Ω is a set off all measurable (real life) objects, Φ is a mapping which represents the measurement scale, $\mathbb{R}$ stands as usually as the set of all real numbers that represent the measurement outcome.

From the measurement theory point of view the mapping Φ should be a homomorphism [9]; this mapping represents the measured variable (physical quantity). The operation to assign numbers to the objects is realized by comparing the measured object ω with the reference objects (standards). If a set of reference objects is γ_i, and nominal value of measured quantity of a reference object is $\Phi(\gamma_i)$ then comparison without measurement errors leads to the claim that $\Phi(\omega) = \Phi(\gamma_n)$ for some γ_n. Then we say that the outcome of a measured quantity Φ for a reference object ω is $x = \Phi(a_n)$, where $x = \Phi(\omega)$. This means that in the case of an ideal measurement the equality holds just for one reference object, in other cases inequality holds (null method).

In order to include a measurement error one needs to do a series of comparisons for a reference set γ_i, $i = 1, \ldots, N$ (N - number of comparisons), determining individually for each single comparison the degree of adjustment to the equality $\Phi(\omega) = \Phi(\gamma_i)$. The general comparison process can be described algebraically without determining the degree of adjustment (see [14]). In our considerations we assume that a resulting outcome of a series of comparisons is a special measure of the possibility μ for the equality $\Phi(\omega) = \Phi(\gamma_i)$, and denote it by

$$\mu\left(\Phi(\omega) = \Phi(\gamma_i)\right). \tag{2}$$

In the probabilistic model we assume that a measure is additive and can be estimated by frequency of events in the random sample. In the fuzzy model this measure is maxitive and is determined by other methods known as „expert decision making" [15]. These two methods correspond to categories A and B according to the Guide. In the A method, the μ measure is estimated with an empirical estimator of the probability distribution, and in the B method, it is determined by the expert decision. The combination of both into a uniform model is possible in at least two ways:

1. one can describe the B method in terms of classical probability, so to assume that the expert decision of choosing a probability distribution for B can be described by additive measure,
2. one can describe the A method in terms of fuzzy sets, so to assume that B component of measurement uncertainty can be described by means of fuzzy sets.

In this paper, we believe that the expert decision should be made using the fuzzy measurements [15], since humans intuitively select better choice and this, in some sense, is described by the maximum measure.

It is possible to describe the method A in the language of fuzzy sets by using the transformation probability to possibility which converts the confidence interval to α-cut, as described in [12, 13]. It is shown there how to define a transformation probability to possibility which is preserving the expanded uncertainty (so that the extended uncertainty of the type A determined by the probabilistic methods is the same as the one determined using the fuzzy methods).

We propose to replace the probabilistic model consistently with the fuzzy one [12]. A fuzzy variable is defined analogously to a random variable but with two differences [8]:

1. a fuzzy variable is defined on a set of patterns and not on a set of elementary events.
2. a measure of possibility defined on the family of subsets of the standard set is maxitive and not additive as a probabilistic measure.

This approach allows a uniform description of random and systematic errors. Systematic errors are described with a uniform distribution of possibilities, i.e. a distribution that mathematically represents the properties of interval arithmetic. The properties of the propagation of random errors are described by the appropriate selection of the t-norm. One of the most important effects of this model is that when estimating methods A and B, the extended uncertainty of type A is consistent with the probabilistic model but the compound uncertainty is the arithmetic sum of components A and B, provided that the fuzzy variable describing component B has a uniform capacity distribution (see [14]).

Measurement

From the mathematical point of view, measurement is an assignment of objects with numbers, i.e. it can be described in a form as (1).

Such a measurement model describes an ideal measurement with measurement errors equal to zero. It must be assumed that a linear order and an additive operation for composing objects can be defined in the event set Ω. The representation of Φ should be a representation of the properties of a set of objects in real numbers (there are many monographs about representative measurement theory such as [6]). The consideration of measurement errors is a generalization of this model. To describe systematic errors, an interval representation is used, i. e. the mapping Φ in the formula (1) assigns a series of intervals to the objects:

$$\Phi : \Omega \to \mathbb{I} \tag{3}$$

where $\mathbb{I}$ is a set of intervals.

We assume that if the measurement can be described by the interval representation, then only systematic errors occur in the measurements.

The probabilistic model is based on the assumption that the measurement mapping (1) is a random variable. This seems to be natural, although there are other solutions in the literature (for example see [11, 2]). Next, we will consider the consequences of this assumption by examining the basic axioms of probability theory.

Similarly, the fuzzy model assumes that the equation (1) is considered as the definition of fuzzy variable.

Probabilistic model of measurement

The description of the measurement in the language of random variables requires basic concepts of probability theory and the answer to the question with which concepts of measurement theory these probabilistic concepts correspond.

Let's use the standard notation:

1. a sample space Ω which is the set of all possible outcomes,
2. an additive measure $P : \mathcal{B} \to [0, 1]$, where $\mathcal{B} \subset 2^{\Omega}$ ($\mathcal{B}$ is the so called σ-algebra of Ω subsets),
3. a random variable is $X : \Omega \to \mathbb{R}$, a random variable X represents the measurement and is a generalisation of equation (1).

Ω is a set of possible events or a set of objects. We should assume that such a set exists and properly describes a measurement experiment that involves assigning numbers to objects. To describe the process of assigning considered objects to standards, we assume that the equation (1) is defined by a random variable which occurs with a certain probability.

Correlation (dependence) of fuzzy variables is described by a copula function $C : [0, 1] \times [0, 1] \to [0, 1]$ such that

$$F(x, y) = C(F_X(x), F_Y(y)) \tag{4}$$

where F_X and F_Y are cumulative distribution function of a random variable X and Y respectively, $F(x, y)$ is a joint cumulative distribution function of random variables X and Y.

It follows from the Sklar's theorem that for any joint cumulative distribution function of two variables X and Y exists function $C : [0, 1] \times [0, 1] \to [0, 1]$ such that holds (4) [10]. The copula function contains all information on the dependence between random variables X and Y.

In the papers and books on measurement sciences the sample space or the set of elementary events, is usually not defined and only random variable distributions are considered. The set of reference object is identified with real numbers. The sample space is identified with the set of possible measurement results that is with the set of real numbers. One has to assume that $X^{-1}(S)$ is well defined for any set of values $S \subset \mathbb{R}$, so there are real objects that are measurable.

A set of measured objects can be written as a set:

$$A = X^{-1}(S) \tag{5}$$

where S is a set of all possible measurement outcomes. So that A is a set of real events that correspond to these outcomes (subset of sample space). Essentially one needs to assume that there exists a set of ideal objects $X^{-1}(x)$, for every $x \in \mathbb{R}$, and a measurement error is given randomly. The probability distribution μ of the random variable X is a measure that is induced to a subset of real numbers, as shown on the figure (1). In fact:

$$\mu(S) = P(X^{-1}(S)) \tag{6}$$

An example of such a measure is the cumulative distribution function F_X:

$$F_X(x) = P(X^{-1}((-\infty, x]) \tag{7}$$

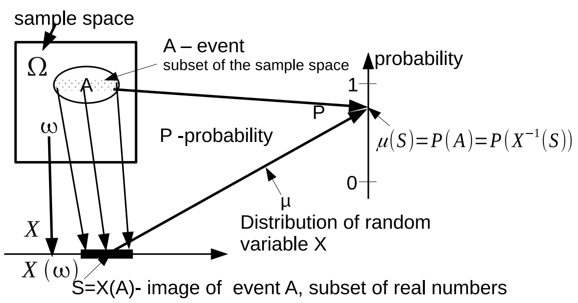

Fig. 1. Scheme of probability distribution, arrows represent mappings

From the probabilistic point of view all experiments consist in selection from a sample space, therefore the measurement consists in selection from a sample space (set of elementary events) some events and assigning them numerical value. The dispersion of the measured data is modelled with the probability distribution but the question of how the set of elementary events exists needs to be answered. According to the definition of random variables, elementary events are given by

$$\omega(x) = X^{-1}(x) \tag{8}$$

where $x \in \mathbb{R}$ represents a value of measured quantity, $\omega(x)$ is an element of the sample space of all ideal objects.

The probabilistic method for estimating uncertainty requires the following assumptions:

1. measurement errors which are observed both as dispersion of the data and as constant error during the experiment, are represented by random variables,

2. the error components represented by random variables are additive,
3. the rules for transmitting uncertainties (complex uncertainty) are derived from the rules for determining the probability distribution of a random variable representing the total error,
4. uncertainty is described by a measure of dispersion of measurement results: standard deviation or confidence interval of the estimator of the measured quantity.

Such assumptions mean that errors considered to be systematic in the old error theory are described in the GUM as random errors that are represented by random variables (for problems of systematic errors see [1]). This is an assumption that gives rise to doubt. The classification into methods A and B according to the Guide does not correspond to the traditional classification into systematic and random errors, especially since this classification depends on the definition of the sample space at each stage of the construction of a probabilistic model.

For example, let study the measurement with a multimeter that was bought in a factory. At the time of its purchase, the sample space includes all multimeters, then the calibration error is a random error because we pull the instrument from the set that was made in these factory. When a measurement is carried out with a single measuring instrument, the sample space is limited to the states of one multimeter and the calibration component remains unchanged for subsequent measurements and should be considered a systematic error.

The rules for the propagation of uncertainty are derived from the equations that determine the distribution of the sum of random variables, i. e. they are derived from the equation that determines the total distribution (4). The general measurement model can be described by the following equation:

$$X = x_0 + \Delta X, \tag{9}$$

where X is the random variable used to describe the measurement, x_0 - true value, ΔX - random variable representing the measurement errors.

Usually errors are assumed to be additive:

$$\Delta X = \sum_{i=1}^{K} \Delta X_i \tag{10}$$

where: ΔX_i-error components.

Extended uncertainty is equal to the radius of a confidence interval of an estimator of probability distribution parameter of the random variable X. For example it is equal to the radius of a confidence interval of average value. Distribution of X depends on the distribution of error components and correlation between these components. This correlation one can describe using a copula function [5]. Cumulative distribution function of ΔX given by (10) has the form :

(11)

$$F_{\Delta X}(z) = \int\limits_{\sum_{i=1}^{K} \Delta x_i = z} F(\Delta x_1, \ldots, \Delta x_K) dx_1 \ldots dx_K$$

where joint cumulative distribution $F(\Delta x_1, \ldots, \Delta x_K)$ is given by (4) for a given copula function.

Fuzzy variable

Fuzzy variable is defined in a similar manner as a random variable and based on the following definitions [8]:

1. pattern set Γ, a set of standard objects, a collection of reference objects with duplicated objects to measure a certain physical quantity,
2. possibility measure $\Pi : 2^{\Gamma} \to [0, 1]$, possibility measure is maxitive measure,
3. fuzzy variable is a mapping $\xi : \Gamma \to \mathbb{R}$, we treat the fuzzy variable as a generalization of measurement mapping (1).

Possibility measure Π fulfills the following conditions: $\Pi(\emptyset) = 0, \Pi(\Gamma) = 0$ and $\Pi(D \cup B) = \max(\Pi(D), \Pi(B))$ for any $D, B \in \Gamma$. The interaction between fuzzy variables is described by a t-norm T:

$$\Pi(A \cap B) = T(\Pi(A), \Pi(B)) \tag{12}$$

T-norm represents the logical connective „AND" in fuzzy logic. A T-norm has the same algebraic properties as a copula function C.

A fuzzy set is described by means of its membership function and we identify a fuzzy set with the membership function. In the framework of fuzzy variable approach the membership function is defined as the distribution function of a fuzzy variable ξ:

$$\bar{A}_{\xi}(x) = \Pi(\xi^{-1}(x)) \tag{13}$$

From the definition the distribution of a fuzzy variable ξ is equal to the one induced by a function ξ and possibility measure Π.

For every fuzzy set we define section of a fuzzy set. The section of a fuzzy set $\bar{A}$ at the level α is called α-cut for short and is defined as:

$$\bar{A}^{\alpha} = \left\langle x \in \mathbb{R} : \bar{A}(x) \geq \alpha \right\rangle \tag{14}$$

In the measurement theory based on fuzzy set model there is no concept of a sample space, because the essential part of the measurement is to have a series of comparisons with the set of standards. The result of the measurement are numbers but they are obtained by comparing with set of measurement standards (this set contains the basic standards and its duplications). As already mentioned, each comparison is assigned a degree of possibility, which is a maxitive measure, since the degree of possibility is a heuristic method and not an estimation method based on the frequency of events.

The measurement consists in assigning an appropriate measurement standards to the studied objects by comparing them correspondingly to the measured quantity. In the fuzzy measurement model, each assignment is made with a certain degree of compliance of the object to be tested with the standard.

Since the reference objects are numbered with nominal values the membership function is defined on the set of real numbers.

Model of uncertainty and uncertainty propagation

In this paper we study only expanded uncertainty due to the fact that in the framework of fuzzy model there is no conception of standard deviation. In the framework of probabilistic theory the expanded uncertainty is defined as the radius of a confidence interval (for some confidence level) of an estimator of mezurand which most often is the estimator of the expected value, there is the average value. The confidence level is the probability that the true value of mezurand is in the confidence interval. In the fuzzy theory of measurement the expanded uncertainty is defined as the radius of

α-cut defined in (14) [7, 13]. The relation between probability and possibility levels requires additional work and will be published later. In this section we want to emphasize that the principles of uncertainty propagation are consequence of the arithmetic of a random variable and a variable fuzzy respectively.

Probability to possibility distribution

A probability space is a mathematical triple $(\Omega, \mathcal{B}, P)$ where Ω is a sample space, $\mathcal{B}$ is a σ-algebra of subsets of Ω and P is a probability measure.

A possibility space is mathematical 2-tuple (Γ, Π), where Γ is a set of standards and Π is a possibility measure. Note that sigma algebra is not needed here since a possibility measure is maxitive.

In order to construct the transformation of a probabilistic structure into a fuzzy one structure, one needs to note that both can be considered as algebraic structures $(\mathcal{F}, \oplus)$, where $\mathcal{F}$ is set of random variables or fuzzy variables and $\oplus$ is addition of random variables or fuzzy variables. The addition operation depends on properties of a random measure or fuzzy measure respectively for random and fuzzy variables. There is a variety of such transformations in literature (see e.g. [3]) but in this paper we assume that this transformation preserves the addition and radius of confidence interval of the median estimator as well [13, 12, 14].

The confidence interval at the probability level can be found by means of the equation:

$$P(I_p) = p \tag{15}$$

Expanded uncertainty at confidence level p is equal to the radius (rad) of a confidence interval. Denote $U_p = \text{rad}(I_p)$. Definition of expanded uncertainty in the Guide is ambiguous: it is assumed that the confidence interval can be applied only to the method A. This is an inconsistency that we want to avoid. Instead we proposed a uniform description in the structure of fuzzy sentences. We assume that a confidence interval corresponds to α-cut of fuzzy set representing the measurement result. In some papers it is assumed that one can relate α-cut to the confidence level given by $p = 1 - \alpha$. This may be generalized for other cases as well. If we assume that the relation between confidence level and fuzzy level have the form $p = 1 - \alpha$ we well obtained the fuzzy set $\bar{A}_X$ which represents the random variable X. The fuzzy set $\bar{A}_X$ can compute from the cumulative probabilistic distribution function F_X according to the equation:

$$\bar{A}_X(x) = \begin{cases} 2F_X(x) & for \quad x \leq M(X) \\ 2 - 2F_X(x) & for \quad x > M(X) \end{cases} \tag{16}$$

where $M(X)$ is median of random variable X and F_X is cumulative probabilistic distribution of random variable X.

Fuzzy sets derived from the transformation probability to possibility describe the components of uncertainty of type A, whereas the form (16) results from being fuzzy sets of type $L - R$ (for monomodal distributions).

The distribution of fuzzy variables representing type B errors is assessed by an expert decision, i. e. a fuzzy set representing type B errors is assessed on the basis of past experience and on general knowledge as well. For example, if we consider a type B error to be systematic, we can describe it with a fuzzy factor within an interval. The reason for this is that systematic errors are modeled in intervals, while the uniform distribution represents an interval itself. A fuzzy variable that represents the sum of errors of type A and B has a distribution that represents the sum of fuzzy sets:

$$\bar{A}_Z(z) = A_X \oplus_T \bar{A}_Y \tag{17}$$

This is a fuzzy generalization of (11) and (4). If a type B error has an even distribution, we get an arithmetic sum of the uncertainty for the combined uncertainty.

Conclusion

The measurement consists in comparing the properties of the studied objects with a set of standards (including the copies of standards). And that is how every analog-to-digital converter works. There is no reason to describe such a process as a random process in terms of the classical probability theory. Therefore, we believe the fuzzy model is better fit to the problem of measurement. In order to preserve the method for determining type A uncertainty as a measurand estimator, we propose the transformation of probability distribution into possibility distribution. The proposed transformation that transforms confidence intervals into appropriate $\alpha-$cuts of fuzzy sets which represent the scattering of the measured data.

Authors: *dr hab. inż. Michał Urbański, email: michal.urbanski@pw.edu.pl, mgr inż. Kinga Wójcicka, email: kinga.wojcicka.dokt@pw.edu.pl, Faculty of Physics, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warszawa, Poland, dr Paweł Wójcicki, email: Pawel.Wojcicki@pw.edu.pl, Faculty of Mathematics and Informational Sciences, Warsaw University of Technology.*

REFERENCES

[1] Attivissimo F., Cataldo A., Fabbiano L. , Giaquinto N., Systematic errors and measurement uncertainty: An experimental approach, Metrologia 44,pp. 1781–1789, 2011.

[2] Cox M.G., Rossi G.B., Harris P.M. and Forbes A., A probabilistic approach to the analysis of measurement processes, Metrologia 45, pp. 493–502, 2008.

[3] Dubois D., Foulloy L., Mauris G., Prade H., Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities, Reliable Computing, 10 273–297, 2004.

[4] Guide to the Expression of Uncertainty in Measurement [web page] `https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf`, [Accessed on 31 Jun. 2022.]

[5] Piotr Jaworski, Fabrizio Durante, Wolfgang Karl Härdle, Tomasz Rychlik (Editors): Copula Theory and Its Applications Lecture Notes in Statistics, Springer, 2010.

[6] Krantz D.H., Luce R.D, Suppes P., Tversky A., Foundations of Measurement, vol. 1. Academic Press, New York, 1971.

[7] Mauris G., Lasserre V., Foulloy L., A fuzzy approach for the expression of uncertainty in measurement, Measurement 29 pp. 165–177, 2001.

[8] Nahmias S. Fuzzy Variables, Fuzzy Sets and Systems, 1 pp. 97–110, 1978.

[9] Narens L., Abstract Mesurement Theory, MIT Press, Cambridge, 1985.

[10] Nelsen Roger B., An Introduction to Copulas, Springer, 1999

[11] Rossi G.B., Crenna F., A probabilistic approach to measurement-based decisions, Measurement 39 pp. 101–119, 2006.

[12] Urbański M. K., Modelowanie pomiarów w algebraicznych strukturach rozmytych, Prace Naukowe Politechniki Warszawskiej, Fizyka z. 55, Warszawa, 2011.

[13] Urbański M.K., Bzowski A., Error analysis using fuzzy arithmetic based on t-norm, 12th IMEKO TC1 and TC7 Joint Symposium Man Science & Measurement, l'Universitde Savoie, September 3–5, Annecy, France, 2008.

[14] Urbański M. K.. Wąsowski J., Algebraic approach to extensive measurement based on direct comparison, Measurement 42, pp. 1341–1351, 2009.

[15] Zadeh, L.A.: The Role of Fuzzy Logic in the Management of Uncertainty in Expert Systems, Fuzzy Sets and Systems 11, pp. 199–227, 1983.