The Mathematical Theory of Evidence and Measurement Uncertainty

Simona Salicone

The concept of measurement uncertainty and the way it should be treated are considered by the "Guide to the Expression of Uncertainty in Measurement," known as GUM [1] that follows an intrinsically probabilistic approach. Presently, the GUM has been universally accepted and its recommendations followed to associate an uncertainty value to the measured value. For this reason, in this paper, we start from the GUM and its fundamental concepts and definitions to introduce an interesting and promising generalization of these same concepts in terms of Theory of Evidence.

The Uncertainty Concept

Let us recall a few important definitions given by the GUM [1].

- Measurement uncertainty is defined *as a parameter, associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand [1].*
- The aim of uncertainty evaluation is: *to provide an interval about the measurement result that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the quantity subject to measurement [1],*
- and *the ideal method for evaluating and expressing measurement uncertainty should be capable of readily providing such an interval, in particular, one with a coverage probability or level of confidence that corresponds in a realistic way to that required [1].*

It is known that when a physical quantity (measurand) is measured more times, even with the same measurement set-up, different measured values are generally obtained. All the obtained or obtainable measured values are referred to, in the GUM, as *distribution of values* or *dispersion of the values*, and uncertainty is expected to somehow *characterize* and quantify this dispersion. This means that it is not possible to know the value assumed by the measurand (the true value) exactly, and uncertainty reflects this lack of exact knowledge in terms of an interval *expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the quantity subject to measurement* [1].

In general, the distribution of the measured values may be due to contributions of different nature: the random contributions to uncertainty and the systematic ones. As reported by the "International Vocabulary of Metrology" (VIM) [2], a random contribution to uncertainty is a *component of measurement error that in replicate measurements varies in an unpredictable manner*, while a systematic one is a *component of measurement error that in replicate measurements remains constant or varies in a predictable manner.* Therefore, if only random contributions affect the measurand, their effect on the measurement result is that the measured values vary randomly around a mean value that is supposed to be the best estimate of the value of the measurand. On the other side, if only systematic contributions affect the measurand, their effect on the measured values is that they are all biased. Of course, if both random and systematic contributions are present, the measured values vary randomly around a biased value, which is generally unknown.

Generally, both kinds of contributions affect a measurement procedure. Hence, both kinds of contributions should be taken into account. However, the GUM states *that it is assumed that the result of a measurement has been corrected for all recognized significant systematic effects and that every effort has been made to identify such effects* [1]. In the Author's opinion, there are two essential reasons for this requirement.

The first reason is a practical one and is related to the available mathematical tools. When the fundamentals of the GUM were discussed in the 1980s, the probability theory was very likely the only known mathematical theory able to handle incomplete information. However, the probability theory has been developed to represent random phenomena. Therefore, in order to refer to the probability theory, only random contributions to uncertainty should be considered, while the systematic ones should disappear.

The second reason is a more conceptual one and is related to the universally recognized best practice in measurement. This best practice requires that, when a systematic effect is recognized and significant, a proper compensation is applied, since this is an immediate (at least from the theoretical point of view) way to minimize uncertainty. The key point here is: when shall an effect be considered significant? To answer to this question, it is necessary to introduce the concept of **target uncertainty**, defined as the *measurement uncertainty specified as an upper limit and decided on the basis of the intended use of measurement results* [2]. In [3], a very simple example is given to intuitively explain that the target uncertainty is set by the intended use of the measurement results. In primary metrology, target uncertainty is set as small as possible and, to reach this goal, it is necessary that all systematic effects are recognized, identified, and compensated for. On the other hand, outside primary metrology laboratories, target uncertainty is not, in general, the smallest obtainable value. Usually, in the industrial world, target uncertainty is a trade-off between the cost of improving measurement uncertainty and the waste production. Therefore, it cannot be stated *a priori* that the compensation of all systematic effects is a necessary operation.

Since, generally, the practice of identifying and compensating the systematic effects might be very expensive, from both a time and economical perspective, it is very important to understand whether this practice is necessary or not, that is, when a systematic effect is actually *significant*. It can be stated that a systematic effect is significant if its missed compensation causes the final measurement uncertainty to exceed the target uncertainty. This statement is intuitive and generally accepted. However, it is somehow tautological: how can I state if a systematic contribution is significant or not if I do not consider it in the evaluation of the final uncertainty and I do not compare this value with the target uncertainty?

The practical consequence is hence that the effect of systematic contributions on the final uncertainty must be always evaluated. Is probability the best mathematical theory to represent and propagate non-random effects also? In my opinion, a different, more general mathematical theory should be followed. Luckily, such a mathematical theory exists and was formulated by Shafer in 1976: it is the Theory of Evidence [4].

An Easy-to-Follow Approach to the Theory of Evidence

The Theory of Evidence was defined by Shafer with the aim to generalize and relax the rules to which the probability functions have to obey. In order to understand Shafer's reasons, we start from the definition of the probability functions. Let us first define the universal set X and its power set $P(X)$: the universal set X is a set which contains all possible objects of the considered framework, including itself; while the power set of a set X, written $P(X)$ or 2^X, is the set of all subsets of X, including the empty set $\varnothing$ and X itself. Probability functions are functions defined in $P(X)$, taking values between 0 and 1:

$$Pro : P(X) \rightarrow [0,1] \tag{1}$$

and obeying the three following rules:

$$Pro(\varnothing) = 0 \tag{2}$$

$$Pro(X) = 1 \tag{3}$$

$$\begin{aligned} Pro(A_1 \cup A_2 \cup \ldots \cup A_n) &= \sum_{i=1}^{n} Pro(A_i) \\ &- \sum_{i<j} Pro(A_i \cap A_j) + \ldots \\ &+ (-1)^{r+1} \sum_{i<j<\ldots<r} Pro(A_i \cap A_j \cap \ldots \cap A_r) + \ldots \\ &+ (-1)^{n+1} Pro(A_i \cap A_j \cap \ldots \cap A_n), \quad \text{for any } A_i \in P(X). \end{aligned} \tag{4}$$

Rules (2) and (3) require that the probability associated to the empty set is zero, while the probability associated to the universal set is one. Rule (4) is called the **additivity rule** for probabilities and expresses a strict relationship between the sets of $P(X)$. This general theorem is often simplified by considering only two sets. When only two sets A and B are considered, the additivity rule becomes:

$$Pro(A \cup B) = Pro(A) + Pro(B) - Pro(A \cap B), \quad \text{for any } A, B \in P(X). \tag{5}$$

This last formula (5) relates the probability of the union of two sets to those of the single sets and their intersection. It simplifies further on when the two sets do not overlap, so that $A \cap B = \varnothing$ and $Pro(A \cap B) = 0$:

$$Pro(A \cup B) = Pro(A) + Pro(B), \quad \text{for any } A, B \in P(X) \mid A \cap B = \varnothing \tag{6}$$

thus yielding, when the two sets are complementary, that is $B = \overline{A}$, $\overline{A}$ being the negation of A, to:

$$Pro(A) + Pro(\overline{A}) = 1, \quad \text{for any } A \in P(X). \tag{7}$$

According to (7), once the probability of an event A is given, the probability of the negation of this event is automatically set. In other words, the probability that an event occurs added to the probability that the same event does not occur is the certain event, whose associated probability is one.

Shafer states that the probability of a given event is not always known or available, but that a person can nevertheless always assign a degree of belief about the considered event. Of course, if the probability of an event is known, the degree of belief given by that person will be exactly that number; but if the probability is not known, the person can anyway provide his/her own degree of belief [3]–[6]. Moreover, while the probability (if known) is a precise number, the degree of belief is strictly related to the person that makes his/her judgment. In other words, different individuals, according to their own knowledge, can make different judgments and provide different degrees of belief about the same considered event:

> *Whenever I write of the "degree of support" that given evidence provides for a proposition or of the "degree of belief" that an individual accords to the proposition, I picture in my mind an act of judgment. I do not pretend that there exists an objective relation between given evidence and a given proposition that determines a precise numerical degree of support. Nor do I pretend that an actual human being's state of mind with respect to a proposition can ever be described by a precise real number called his degree of belief. Rather, I merely suppose that an individual can make a judgment … he can announce a number that represents the degree to which he judges that evidence to support a given proposition and, hence, the degree of belief he wishes to accord the proposition* [4].

Hence, he defines the belief functions *Bel* as:

$$Bel : P(X) \rightarrow [0,1] \tag{8}$$

such that:

$$Bel(\varnothing) = 0 \tag{9}$$

$$\mathrm{Bel}(\mathrm{X}) = 1 \tag{10}$$

$$\begin{aligned} &Bel(A_1 \cup A_2 \cup \ldots \cup A_n) \geq \\ &\sum_{i=1}^{n} Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + \ldots + (-1)^{r+1} \sum_{i<j<\ldots<r} Bel(A_i \cap A_j \cap \ldots \cap A_r) \\ &+ \ldots + (-1)^{n+1} Bel(A_i \cap A_j \cap \ldots \cap A_n), \text{ for any } A_i \in P(X) \ . \end{aligned} \tag{11}$$

While (9) and (10) are equal to (2) and (3), rule (11) is less strict than the additivity rule (4) and is called **superadditivity rule**. When only two sets *A* and *B* are considered, (11) becomes:

$$Bel(A \cup B) \geq Bel(A) + Bel(B) - Bel(A \cap B), \quad \text{for any } A, B \in P(X) \ . \tag{12}$$

This last formula relates the belief of the union of two sets to those of the single sets and their intersection. It simplifies further on when the two sets do not overlap:

$$Bel(A \cup B) \geq Bel(A) + Bel(B), \quad \text{for any } A, B \in P(X) \mid A \cap B = \varnothing \tag{13}$$

thus yielding, when the two sets are complementary, to:

$$Bel(A) + Bel(\overline{A}) \leq 1, \quad \text{for any } A \in P(X) \ . \tag{14}$$

In Shafer's original approach, the belief function about an event A ($Bel(A)$) is an act of personal judgment [4] and represents the degree of credibility that the event A occurs assigned by a person. The degree of belief defined by Shafer is hence subjective. This subjectivity is not in contrast with the GUM [1]. Indeed, the GUM defines two different methods to evaluate measurement uncertainty, named type-A and type-B evaluation of uncertainties. Type-A and type-B evaluations are defined, respectively, as: "*method of evaluation of uncertainty by the statistical analysis of series of observations*" and "*method of evaluation of uncertainty by means other than the statistical analysis of series of observations*" [1, 7]. Hence, also the GUM introduces, with the type-B method, subjectivity and assigns to this method the same importance as to the one based on the statistical analysis [1]. Moreover, as far as the type-B method is concerned, GUM article 4.3.2 states: "*The proper use of the pool of available information for a type-B evaluation of standard uncertainty calls for insight based on experience and general knowledge, and is a skill that can be learned with practice. . .*" [1]. It is interesting to observe how much similar are Shafer's words and the GUM ones. The close correspondence between the "*given evidence*" and "*the pool of available information,*" as well as that between the "*degree of belief*" and the "*insight based on experience and general knowledge*" is quite evident [6] and suggests that Shafer's belief functions are perfectly compliant with the requirements of the GUM and can be usefully employed to represent measurement uncertainty.

Even if Shafer starting point is completely subjective, and his original definition of belief functions is based on an "*act of personal judgment*", it will be shown that belief functions can be also used in an objective manner, according to the available knowledge. Thus, as well as the type-A and type-B methods of the GUM, objectivity and subjectivity can be fully represented by belief functions.

Coming back to Shafer's original approach [4], [5] similarly to when a person can provide a degree of belief, or credibility, about an event A, he or she can also assign a degree of doubt about the event, that is, his or her own degree of credibility that the event A does not occur: $Bel(\overline{A})$. Since both $Bel(A)$ and $Bel(\overline{A})$ are acts of personal judgment, the superadditivity rule does not strictly require to assign 1 to their sum, although the union of A and $\overline{A}$ is the certain event X [3]–[5]. This means that, in order to give a complete description about an event A, both the belief function $Bel(A)$ and the doubt function $Bel(\overline{A})$ should be given.

The doubt function is very seldom used, and, in general, a complete description of an event A is given by the belief function $Bel(A)$ and a function $Pl(A)$, related to the doubt function [4]–[6] by:

$$Pl(A) = 1 - Bel(\overline{A}) \tag{15}$$

The above relationship states that function Pl associated to the event A expresses the extent to which one fails to doubt A, in other words, the extent to which one finds A plausible. Thus, Pl

is called plausibility function and $Pl(A)$ represents the plausibility of A [4]–[6]. Plausibility functions are defined as:

$$Pl : P(X) \rightarrow [0,1] \quad (16)$$

and obey:

$$Pl(\varnothing) = 0 \quad (17)$$

$$Pl(X) = 1 \quad (18)$$

$$Pl(A_i \cap A_j \cap \ldots \cap A_n) \leq \sum_{i=1}^{n} Pl(A_i) - \sum_{i<j} Pl(A_i \cup A_j) + \ldots + (-1)^{r+1} \sum_{i<j<\ldots<r} Pl(A_i \cup A_j \cup \ldots \cup A_r) + \ldots + (-1)^{n+1} Pl(A_1 \cup A_2 \cup \ldots \cup A_n), \text{ for any } A_i \in P(X) \ . \quad (19)$$

Rule (19) can be simplified by considering only two sets A and B:

$$Pl(A \cap B) \leq Pl(A) + Pl(B) - Pl(A \cup B), \text{ for any } A, B \in P(X) \ . \quad (20)$$

This last formula relates the plausibility of the intersection of two sets to those of the single sets and their union. It simplifies further on when the two sets do not overlap:

$$Pl(A \cup B) \leq Pl(A) + Pl(B), \text{for any } A, B \in P(X) \mid A \cap B = \varnothing \ . \quad (21)$$

Thus, when the two sets are complementary, yielding:

$$Pl(A) + Pl(\overline{A}) \geq 1, \quad \text{for any } A \in P(X) \ . \quad (22)$$

The given definitions of the belief and plausibility functions show how they generalize probability and are, therefore, the fundamental brick on which we can state that the Theory of Evidence is a mathematical generalization of probability.

In case a person knows the probability of an event, he will assign the same numerical number to his or her degree of belief: $Bel(A) = Pro(A)$. In such a situation, because of (7) and (15), it also follows that $Pl(A) = Pro(A)$. This means that, if the probability of an event is known, the belief, the plausibility, and the probability of this event are the same: $Bel(A) = Pro(A) = Pl(A)$. In general, however, the relationship among the belief, the plausibility, and the probability of an event [4], [8] is:

$$Bel(A) \leq Pro(A) \leq Pl(A) \quad (23)$$

and this is the reason why belief and plausibility functions are also called, respectively, **lower** and **upper probability functions**.

This is an important conclusion, because there are several situations, as the one considered later in this paper, where it is known that an event belongs to a given set, but nothing else is known about the probability of this event. In such a case, from a strict probabilistic point of view, all possible probability functions in a given family should be considered [9], and this implies a very complicated formulation of the problem. On the other hand, (23) allows a great simplification: whenever there is not enough evidence to select a single probability function to represent an event, but a family of probability functions over the given universal set X should instead be considered, it is possible to consider only the lower and upper probability functions, that is, the belief and plausibility functions [9]. Since these functions can be assigned starting from a given family of probability functions, this means that they can be defined in an objective way, even if there is very little or no knowledge about the event, thus overcoming the initial subjective approach by Shafer.

An Example

Let us see with a simple practical example how the mathematical concepts briefly recalled in the previous section can be usefully applied.

The considered metrological example consists in the evaluation of the uncertainty associated with a dc voltage value measured with a multimeter. A dc voltage of 24 V is generated by a Fluke 5500 A calibrator and measured by eight different Leader 856 digital multimeters, providing the eight measured values: V_{m1} = 24.008 V; V_{m2} = 23.998 V; V_{m3} = 23.999 V; V_{m4} = 24.004 V; V_{m5} = 23.996 V; V_{m6} = 23.998 V; V_{m7} = 24.007 V; and V_{m8} = 24.005 V.

The calibrator's accuracy is much higher than that of the multimeters', so that the uncertainty associated to the generated voltage signal is negligible with respect to the uncertainty associated to the measured values. Therefore, the 24 V signal can be considered as the reference true value, with no uncertainty. On the other hand, from the manufacturer specifications of the multimeters, it is possible to retrieve an uncertainty interval to associate to the measured values. In the 30 V range, the instrument features a resolution of 1 mV and a measurement accuracy of ±±2 mV. Hence, according to these specifications, the eight multimeters provide the following measurement results: V_1 = 24.008 V ± 14 mV; V_2 = 23.998 V ± 14 mV; V_3 = 23.999 V ± 14 mV; V_4 = 24.004 V ± 14 mV; V_5 = 23.996 V ± 14 mV; V_6 = 23.998 V ± 14 mV; V_7 = 24.007 V ± 14 mV; and V_8 = 24.005 V ± 14 mV. The accuracy intervals depend on the measured values. In this example, however, since the eight multimeters provide eight readings very close to each other, the widths of the intervals turn out to be all equal.

Fig. 1 shows that the 24 V reference value (pink vertical line) is included in each of the 8 accuracy intervals (blue lines around the blue dots, which represent the eight readings), thus confirming that the multimeters are working properly. The same figure also shows that the eight readings distribute around the 24 V reference value. The mean of the eight readings is V_{mean} = 24.002 V. V_{mean} can be considered as the best estimate of the measurand [1]. It is represented by the orange dashed vertical line in Fig.1.

Repeated readings have been also performed for each multimeter. This showed that the variability of each instrument is always within ±1 digit; that means, in the considered range,

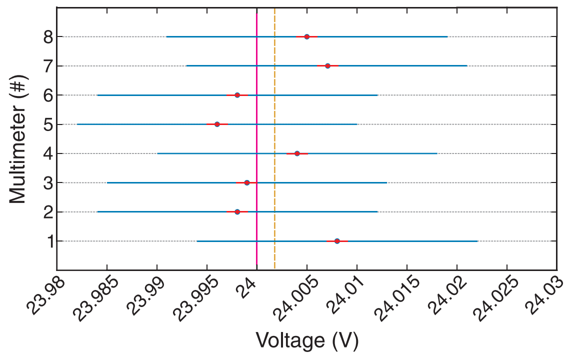

Fig. 1. Experimental results obtained with eight multimeters 4 ½ LEADER 856. The pink vertical line represents the 24 V voltage reference value. The multimeters are numbered from 1 to 8 (vertical axes). For each multimeter, the dot represents the single measured value, the blue interval represents the accuracy interval around the measured value obtained from the data sheet, while the red interval represents the random variation of the last digit on the instrument's display. The orange vertical dashed line represents the mean value of the eight different readings provided by the eight multimeters.

a variability of ±1 mV around the measured value. This variability leads to intervals of variation (red lines around the blue dots in Fig. 1) which are smaller than the accuracy intervals (blue lines around the blue dots in Fig. 1).

The obtained results can be interpreted as follows. Since the different readings of the same instrument fall in a very small interval within the accuracy one, it can be stated that the ±1 digit variability is due to random phenomena. These random phenomena act in a similar way on the different multimeters, since the same ±1 digit variability is observed (same width of the red intervals in Fig. 1). Moreover, since the blue dots are positioned in different positions with respect to the pink vertical line, it can be stated that the different multimeters introduce different systematic effects.

What can be concluded from the observation of these measurement results? Let us consider a situation in which all the eight measurement values provided by the eight different multimeters are available. In this case, it is possible to consider the mean of the eight measured values, infer a probability distribution, and evaluate a standard uncertainty value. Since, in this case, a statistical analysis is performed, there is no real need to use a different mathematical theory than the probability theory.

However, it is quite an impractical situation to employ more than one measurement instrument to measure the same physical quantity. The more general situation is that only one multimeter is employed to measure a voltage value, because only one multimeter is available and, in general, the cost of repeating the measurement procedure with more multimeters, provided that they are available, is too high. In this more general case, taking different measured values does not help to improve the estimate of the measurand. In fact, the different measured values are affected by the same systematic effect, which is preponderant with respect to the random one. Furthermore, the measured values differ from each other very slightly and it is not guaranteed that their average provides an estimate of the measurand better than the single measured value (Fig. 1). In such a situation, no statistical analysis is useful and the entire accuracy interval provided by the manufacturer must be considered. But is it possible to associate a probability distribution over this interval? As a matter of fact, only the manufacturer could have the elements to suggest a particular probability distribution, having tested the entire class of the considered multimeters, but, generally, he does not provide any information about it. Hence, which probability distribution should be chosen? Normal? Uniform? U? …? The choice of whichever probability distribution is completely arbitrary and no supported by any evidence. In fact, the only available evidence is that the value of the measurand is in one point of the given accuracy interval and nothing else is known. It follows that, in this situation, it is not possible to assign correctly a probability distribution over the given interval.

As already stated in the previous section, the correct mathematical approach would require one to consider the whole family of possible probability distributions over the given interval [9], but it can be readily perceived that this approach is quite cumbersome to follow and time consuming. However, the Theory of Evidence offers a simpler solution that considers only the lower and upper probabilities in the given family, that is, the belief and plausibility functions [9].

The considered example, where the only available information about the uncertainty to associate to a measured value is the accuracy interval provided in the manufacturer specifications, is very often met in measurement procedures. For this reason, the provided example, though very simple, is very general. In this case, it is only known that the value of the measurand lies in one point of the given interval, but nothing else is known. To be more specific, it is not known any probability distribution associated to the dispersion of the possible measured values. This situation corresponds exactly to the definition of **total ignorance**, or **complete ignorance**, given by Shafer: "*the situation where we have no evidence about X at all*" [4]; X being the universal set, that is, in our example and in all measurement applications, the accuracy interval provided by the manufacturer. In this situation, no evidence, no knowledge, no information is available about X, so that it is not possible to assign to it any probability function [4], [9]. As a matter of fact, in Shafer's opinion, probability functions are not able to handle correctly the case of total ignorance and this is the reason why he had to generalize the additivity rule and define the belief functions [4], [5]. On the other hand, total ignorance can be fully mathematically represented within the theory of evidence. The belief and plausibility functions associated to total ignorance over an interval (or, in general, a universal set) X are [3]–[5]:

$$\begin{cases} Bel(X) = 1 \\ Bel(A) = 0 \quad \forall A \neq X \mid A \in P(X) \end{cases} \tag{24}$$

called the *vacuous belief function*, and

$$Pl(A) = 1 \ \forall A \in P(X) \ . \tag{25}$$

In the next section, new functions will be defined so that a complete mathematical description of total ignorance can be given.

Probability and Possibility Theories as Particular Cases of the Theory of Evidence

In addition to belief and plausibility functions, a third function is defined within the theory of evidence, called *basic probability assignment function* and defined in [4], [5], and [10] as:

$$m : P(X) \to [0,1] \tag{26}$$

$$m(\varnothing) = 0 \tag{27}$$

$$\sum_{A \in P(X)} m(A) = 1 \tag{28}$$

where $m(A)$ represents the degree of belief that an element x belongs to set A: no belief is assigned to the empty set and the value assigned to the total belief is one. It is possible to define the belief and the plausibility functions from the basic probability assignment function [4]–[6], [10] as:

$$Bel(A) = \sum_{B|B \subseteq A} m(B) \tag{29}$$

$$Pl(A) = \sum_{B|A \cap B \neq \varnothing} m(B) \ . \tag{30}$$

The basic probability assignment function allows to define the focal elements [4]–[6], [10] that is, those sets $A \in P(X)$ for which $m(A) > 0$. It can be also said that the focal elements are those sets where the degree of belief focuses. The particular form of the focal elements is very important, since it determines some characteristics of the belief and plausibility functions. Of great importance are the cases in which the focal elements are singletons and the focal elements are all nested.

In the first case, the basic probability assignment function is greater than zero for any point of the universal set and it is possible to write:

$$m(\{x\}) > 0 \ . \tag{31}$$

It is possible to prove that, under this assumption [4], [5]:

$$Bel(A) + Bel(\bar{A}) = 1 \tag{32}$$

Equation (32) is similar to (7) and shows that, when the focal elements are singletons, the belief functions obey the rule of additivity. In this case, belief functions are also called Bayesian belief functions [4]. Since the rule of additivity is valid for the probability functions, the obtained result suggests that, when the focal elements are singletons, the belief functions are also probability functions. In other words, (32) suggests that, when the focal elements are singletons, we are dealing with the probability theory and $Bel(A) = Pro(A) = Pl(A)$ holds.

It is well known that the probability functions *Pro* are fully represented by probability distribution functions p. It can be proved that, if (31) holds, the basic probability assignment function is also a probability distribution function [4], [5], that is:

$$p(x) = m(\{x\}) \tag{33}$$

This result is very important, since it demonstrates that the approach to measurement uncertainty in terms of the theory of evidence is a generalization of the approach in terms of the probability theory, suggested by the GUM [1].

A second particular case of the theory of evidence is the possibility theory, obtained every time the focal elements are nested sets. A simple way to define nested focal elements is the following, as shown in Fig. 2: $X = \{x_1, x_2, \ldots x_n\}$ is the universal set, and $A_1 = \{x_1\}, \ldots A_i = \{x_1, \ldots x_i\}, \ldots A_n = X$ are the focal elements. Therefore:

$$A_1 \subset A_2 \subset \ldots \subset A_n \ . \tag{34}$$

When (34) holds, belief and plausibility functions are called, respectively, necessity and possibility functions *Nec* and *Pos* and (29) and (30) simplify as follows [4]–[6], [10]:

$$Nec(A_j) = \sum_{k=1}^{j} m(A_k) \tag{35}$$

$$Pos(A_j) = Pl(A_j) = \sum_{A_k | A_k \cap A_j \neq \varnothing} m(A_k) = \sum_{k=1}^{n} m(A_k) = 1 \ . \tag{36}$$

Similarly to the probability functions *Pro* that may be fully represented by a probability distribution function p, the possibility functions *Pos* may be fully represented by a distribution function called possibility distribution function r and defined [4], [5] as:

$$r : X \to [0,1] \tag{37}$$

such that

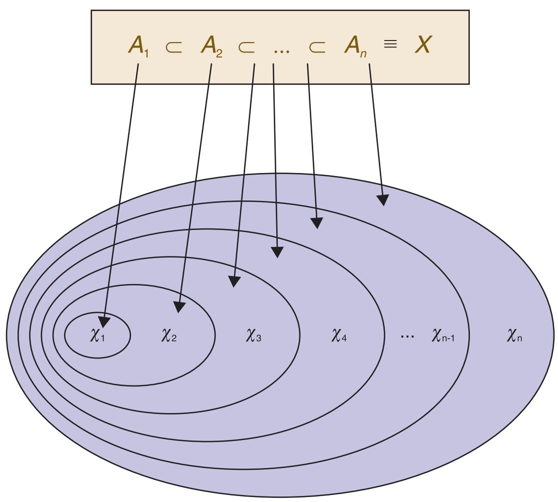

Fig. 2. Universal set and nested focal elements.

$$\sup_{x \in X} r(x) = 1 \quad . \tag{38}$$

The relationship between *Pos* and *r* is [4], [5]:

$$r(x) = Pos(\{x\}) \tag{39}$$

and

$$Pos(A) = \sup_{x \in A} r(x) \quad . \tag{40}$$

It can be proved that, if (34) holds, the relationship between the basic probability assignment function *m* and the possibility distribution function *r* [4], [5] is:

$$r_i = \sum_{k=i}^{n} m_k \tag{41}$$

which leads to $r_1 = 1$, and

$$m_i = r_i - r_{i+1} \tag{42}$$

where $r_{n+1} = 0$ by definition, where $r_i = r(x_i)$ and $m_i = m(A_i)$.

More about Shafer's Total Ignorance

Let us now see how the additional mathematical derivations given in the previous section can be usefully exploited to represent Shafer's total ignorance. Equations (24) and (29) provide the basic probability assignment function associated to total ignorance as:

$$\begin{cases} m(X) = 1 \\ m(A) = 0 \quad \forall A \neq X \end{cases} \quad . \tag{43}$$

This means that, when total ignorance is represented, there is only one focal element: the universal set *X* itself.

Is it possible to represent total ignorance also with the probability and possibility distributions? First of all, let us recall that Shafer defines total ignorance within the theory of evidence as "*the situation where we have no evidence about X at all*" [4]. If a probability distribution is considered, we are not in the situation where we have no evidence about *X* at all, because we are adding the constraint that the focal elements are singletons, that is, that the basic probability assignment function is greater than zero over singletons. This assumption conflicts with (43), which does not allow to assign a basic probability *m* greater than zero to any subset of *X*. Therefore, if a probability distribution is assigned over *X*, this situation cannot be considered as total ignorance.

Similarly, if a possibility distribution is considered, at a first extent, we may think that we are not in the situation where we have no evidence about *X* at all, because we are adding the constraint that the focal elements are all nested. However, to satisfy (34), it is not necessary that all sets $A_i = \{x_1,...x_i\}$ are focal elements: there could be only some of them and, in the extreme case, only one of them; that is *X*. Therefore, this different assumption does not conflict with (43) and, hence, it is possible to assign a possibility distribution over *X* representing total ignorance.

Hence, we arrive to Shafer's firm conviction that probability functions are not able to handle correctly the case of total ignorance, while possibility functions are. The common use of a uniform probability distribution to represent total ignorance is not representative of the real Shafer's total ignorance, "*the situation where we have no evidence about X at all*", since it represents the situation where the evidence shows that every element of *X* has the same probability. On the other hand, total ignorance can be represented by a possibility distribution. Equations (41) and (43) yield [4], [5]:

$$r_i = r(x_i) = 1 \quad \forall i \quad . \tag{44}$$

Equation (44) shows that total ignorance can be represented by the uniform, or rectangular, possibility distribution. This is the greatest possibility distribution: all elements of the universal set are equally possible, with no preference of one element with respect to the others.

As stated in the previous sections, Shafer's concept of total ignorance perfectly represents a situation very often met in the measurement procedures, that is, the situation when the only available information is that the measurand lies in a given interval. If the example of the multimeters is considered again, equation (44) allows to associate a possibility distribution to each measured value V_{m1}, V_{m2} ...V_{m8}. As an example, Fig. 3 shows the measurement result provided by multimeter 1 in terms of a possibility distribution. The support of the possibility distribution (the universal set *X*) is the accuracy interval provided by the manufacturer (V_1); the mean value of this interval is the measured value (V_{m1}); the rectangular shape of the possibility distribution shows that all values within the interval are equally possible, since no further information is available, so that no value is preferable with respect to the others. In the same figure, the measured values by multimeters 2 to 8 are

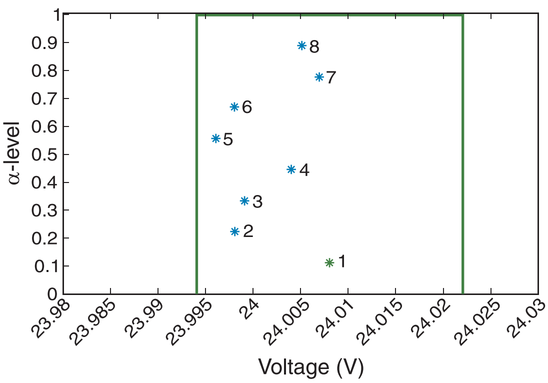

Fig. 3. Measurement result provided by Multimeter 1: the green star represents the single measured value while the green line represents the associated possibility distribution function. The support of this possibility distribution function is interval (24.008±0.014) V, according to the manufacturer specifications. The shape of this possibility distribution function is rectangular, showing that any value within interval (24.008±0.014) V is equally possible. In other words, despite the single measured value, the voltage value could be any of the possible values in interval (24.008±0.014) V, with equal possibility. The blue stars represent the measured values of Multimeters 2 to 8.

also reported, showing that all of them are included within the given possibility distribution and are equally possible.

Conclusion

This paper has provided a simple introduction to the theory of evidence and has shown how practical measurement applications agree with Shafer's concept of total ignorance. In the next paper, the meaning of the random-fuzzy variables will be explained, the way they can be built, according to the available metrological information, and the way they can be composed. The aim of this second paper is hence to show how measurement uncertainty can be represented and propagated through the measurement procedure, taking into account both the nature of the uncertainty contributions and the possible correlation between the different contributions. Finally, in the third paper, a method to compare the random-fuzzy variables with each other will be shown. Since the final aim of any measurement procedure is to take a decision, the aim of this last paper is to show how this last step can be performed, by taking into account the measurement uncertainty.

References

[1] JCGM 100:2008, "Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement," Joint Committee for Guides in Metrology, 2008.

[2] JCGM 200:2012, "International Vocabulary of Metrology - Basic and General Concepts and Associated Terms," Joint Committee for Guides in Metrology, 2012.

[3] S. Salicone, "The theory of evidence: A new promising approach to the evaluation and expression of measurement uncertainty," *IEEE Instrum. Meas. Mag.*, vol. 16, no. 1, pp. 18-23, Feb. 2013.

[4] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[5] S. Salicone, "*Measurement Uncertainty: An Approach via the Mathematical Theory of Evidence*", *Springer series in reliability engineering*. New York: Springer-Verlag, 2006.

[6] A. Ferrero, and S. Salicone, "Uncertainty: only one mathematical approach to its evaluation and expression?" *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2167–2178, Aug. 2012.

[7] A. Ferrero and S. Salicone, "Measurement uncertainty," *IEEE Instrum. Meas. M*ag., vol. 9, no. 3, pp. 44-51, Jun. 2006.

[8] D. Dubois, L. Foulloy, G. Mauris, and H. Prade, "Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities", *Reliable Computing. Kluwer Academic Publishers*, vol. 10, pp. 273–297, 2004.

[9] G. Coletti, R. Scozzafava, and B. Vantaggi, "Inferential processes leading to possibility and necessity" *Information Sciences*, vol. 245, pp. 132-145, 2013.

[10] A. Ferrero and S. Salicone, "Fully comprehensive mathematical approach to the expression of uncertainty in measurement," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 3, pp. 706-712, Jun. 2006.