Uncertainty in measurement: some thoughts about its expressing and processing

Luca Mari
Università C.Cattaneo - LIUC
C.so Matteotti 22, 21053
Castellanza (Va), Italy
email: lmari@liuc.it

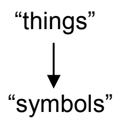
Leon Reznik
School of Communications and Informatics
Victoria University, PO Box 14428,
Melbourne City MC, VIC 8001,
Australia
email: Leon.Reznik@vu.edu.au

Summary: The paper discusses the concept of non-exactness of measurement results, and analyzes it by clearly distinguishing between (i) the way the results are expressed to make their uncertainty explicit; (ii) the way the chosen expression is interpreted as a suitable combination of non-specificity and uncertainty; (iii) the way the interpreted results are formally dealt with. In this perspective the merits and flaws of the ISO Guide to the expression of uncertainty in measurement are highlighted.

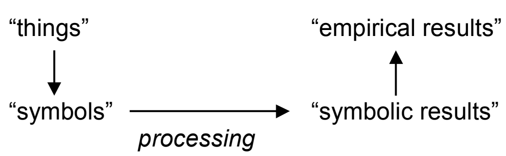
Keywords: *Foundations of Measurement, Measurement Uncertainty*

1. Why non-exactness is an issue in measurement

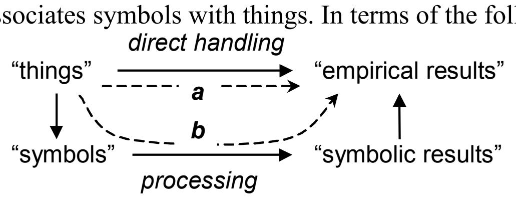
Measurement is a means of setting up a bridge between the empirical world (to which the measured thing belongs) and the linguistic/symbolic world (to which the measurement result belongs):



The pragmatic aim of measurement is to enable symbolic processing of data drawn from the empirical world, so that any result obtained in data processing can be re-interpreted in terms of the measured things:



Crucial for the validity of this re-interpretation is therefore the *faithfulness* of the operation that associates symbols with things. In terms of the following diagram:



the issue is whether the procedures *a* and *b* would lead to the same result.

In the case of measurement such a notion of faithfulness is peculiar, since the result of any measurement is defined in reference to a given scale, playing the role of both conceptual and operational context (i.e.: model) in which the result is interpreted [Mari, 1999]. To be faithful, measurement is then required not only to keep a stable association among things and symbols, but also to preserve any structure (e.g., ordering) empirically observed among things.

The fact is that these two worlds, the one of empirical things and the one of symbols, are inherently different. According to Bridgman, «there are certain human activities which apparently have perfect sharpness. The realm of mathematics and of logic is such a realm, par excellence. Here we have yes-no sharpness. But this yes-no sharpness is found only in the realm of things we say, as distinguished from the realm of things we do. Nothing that happens in the laboratory corresponds to the statement that a given point is either on a given line or it is not» [Bridgman, 1959].

As Pattee pointed out [Pattee, 1989], the basic measurement problem is that semantic grounding of symbols by measurement is a controlled action by an observer that could not functionally be described by any laws. More accurately, if a measuring device, which is certainly a physical system obeying laws, is actually described by these laws by combining the device with the original system being measured, then the initial conditions are no longer separated, and additional new measuring devices are required to establish the initial conditions for this combined system. Therefore, syntactic description of measurement by laws destroys the semantic function of the measurement.

L.Zadeh [ ] classifies all the information commonly available into three groups:

- factual information which is numerical and measurement-based;
- pseudo-measurement based and pseudo-numerical (e.g. "checkout time is 11.00") information;
- perception based information which is mainly linguistic (e.g. "Robert is honest").

Those three groups are commonly assumed to differ in the degree of uncertainty, which the corresponding information has. However, does it mean that those groups have a fundamental, ontological difference as one traditionally assumes?

The classification follows the traditional division between quality and quantity, hard and soft sciences. This division could be illustrated with two quotations: “Qualitative is nothing but poor quantitative” (Ernest Rutherford) and “A social scientist is a person who counts telephone poles” (Robert Hutchins).

For Rutherford everything we call a quality or a percept is expressible in terms of numerical magnitudes, without loss or distortion. Therefore, for him every quality can be quantified and hence measured and computed. For Rutherford science does not begin until quantification is made, until crude and inexact talk about quality is replaced by precise, exact, and completely equivalent talk about numbers. Here he follows another famous quotation by Dmitry Mendeleev: “Science begins where measurements are started”.

Hutchins tacitly accepts Rutherford’s equation of science with quantitative but for him this makes the phrase “social scientist” a contradiction in terms. For Hutchins, the features or qualities of a social structure which are of interest or importance are precisely those which are unquantifiable, and conversely anything that can be counted is trivial or irrelevant.

One of the not-so-commonly discussed consequence of such assumptions is that some classical distinctions, such as “linguistic” vs. “numerical”, and “qualitative” vs. “quantitative”, become rather ... fuzzy, in the proper sense that no clear-cut threshold can be drawn to define them, but by mere convention. Explicitly: even admitting that nominal scale measurement is inherently non-numerical [Mari 2000], ... what are “numbers”? On this subject some interesting lessons would come from the theory of algebraic structures (mainly focused on the concepts of relational systems - a generalization of universal algebras - and morphisms among them). In terms of scale types, the status of the so-called interval type could be considered, measuring quantities such as temperature (before Kelvin scale) and potential energy. According to a strong tradition, any kind of (integer, rational, real, complex) numbers derive from the natural sequence (for example, Kronecker said that natural numbers were given to manhood by God, while all other numbers have been created by us). Therefore the emphasis is on the existence of a unit, from which the whole sequence is inductively generated (as in the axiomatic approach by Peano). But the formal entities embedded in an interval scale do not empirically obey to any unit identification (consider the case of temperature: surely the unit degree is not an empirically primitive entity in the measurand definition). On the other hand, who would dare not to consider temperature degrees in Celsius scale as “numbers”? As a consequence, it seems that the very concept of number was implicitly extended, to embody “entities with a total order and a compatible metric on it”. Is this the “final” extension? Or even “entities with a total order”, i.e. in an ordinal scale, are “numbers”?

R. Rosen gives another example [Rosen, 1987]. It is a fact of experience that 2 sticks + 3 sticks = 5 sticks. On its face, this is a proposition about sticks. But it is not the same kind of proposition as “sticks burn” or “sticks float”. It differs from them because it is something else besides sticks, and that “something else” according to Rosen is the mathematics. The mathematical world is embodied in percepts, but exists independent of them. “Truth” in the mathematical world is likewise manifested in, but independent of, any material embodiment, and is thus outside of conventional perceptual categories like space and time.

With a purely pragmatic position, one might see the distinction between “numbers” and “non-numbers” as purely conventional, and definitely useless. Furthermore, “numbers”, whatever they are, are surely particular linguistic entities (characterized by specific algebraic properties), so that in any case the distinction should be “numbers” vs. “non-numbers”, and not “numbers” vs. “linguistic entities”.

The non-exactness (in the following the difference between non-exactness and uncertainty will be maintained and discussed) of measurement results accounts for such a distinction, although «by forcing the physical experience into the straight jacket of mathematics, with its yes-no sharpness, one is discarding an essential aspect of all physical experience and to that extent renouncing the possibility of exactly reproducing that experience. In this sense, the commitment of physics to the use of mathematics itself constitutes, paradoxically, a renunciation of the possibility of rigor» [Bridgman, 1959].

2. The expression of non-exact measurement results

Taking into account the linguistic side of the problem, the first decision to be made is related to the form a measurement result should be given to make its non-exactness explicit. According to the ISO *Guide to the expression of uncertainty in measurement* (GUM) ([ISO, 1993]; a useful synthesis of the Guide can be found in [Taylor, Kuyatt, 1997]), any measurement result must account for both the measurand value and its estimated uncertainty, and is therefore expressed as a couple:

measurement result = $<x,s(x)>$

where the first term, representing the measurand value, could be obtained as the average of the population of the instrument readings, and the second term, representing the uncertainty value, as its estimated standard deviation.

This position has the fundamental consequence that the concepts of measurement result and measurand value cannot be dealt with as equivalent, as instead, often implicitly, assumed in traditional approaches: measurement results express information not only on the measurand value, but also on its uncertainty. In other terms, “exact” measurement becomes the exception, not the rule.

If innovative in this *conceptual* assumption, the GUM maintains a strong link with the tradition in the *formal* side of its recommendations: indeed different, and somehow more general, representations could be chosen, for example subsets, or fuzzy subsets, or probability distributions. All of these are more widely applicable than the representation suggested by the GUM, that can be employed only for algebraically strong scales (although the generality is understood, because of the applicability of the Chebyshev's inequality, of expressing measurement results as couples $<x,s(x)>$, being x and $s(x)$ the average and standard deviation of the (unknown) assumed probability distribution).

3. Evaluation methods and related meanings

The choice to express measurement results as $<x,s(x)>$ still leaves open the decision about the evaluation methods that can be adopted to obtain such results and the meaning to be attributed to both the terms of the couple. Once more, the GUM standpoint at this regard is a mix of innovation and tradition. With respect to the evaluation methods, the GUM embodies a recommendation issued by the CIPM in 1981 [CIPM, 1981] and admits both statistical (so-called "type A") and non-statistical ("type B") methods. The condition to make this pluralism operatively acceptable is that the suggested techniques to formally deal with the results are independent of the "type" of the evaluation method and therefore the same in both cases. From the conceptual point of view this position is an important step against the radical objectivism of some classical interpretations of measurement: some subjective information, in the form of "degrees of belief" (to quote the GUM) is present and required even in the case of an "objective" operation as measurement [Mari, Zingales, 1999].

On the other hand, not so pluralistic is the position of the GUM in reference to the meaning of the term $s(x)$. Its interpretation is meant to be statistical, as a standard deviation of x, thought of as a random variable in itself and therefore randomly extracted from a, typically unknown, probability distribution. The main application suggested by the GUM for this so-called "standard uncertainty" is to compute the "law of uncertainty propagation" (what is classically called "error propagation") through functional relations, thus essentially a problem of derived measurement. Only for specific applications a further interpretation is recognized as useful, in which $s(x)$ (and more precisely $ks(x)$, being k a proportionality factor usually in the range [1,3]), in this case called "expanded uncertainty", is considered to express the half-width of the interval of which x is the center point. This set-theoretic interpretation is however deemed as explicitly dependent on the statistical one and formally derived from it.

4. Two categories of applications

The position of the GUM is clearly conservative in this regard. A more general standpoint recognizes that the same result $<x,s(x)>$ could admit distinct interpretations in distinct applications. Given a set $\{<x_i,s(x_i)>\}$ of such measurement results, two basic areas of applications can be identified:

- "derived measurement": a quantity Y is known as analytically dependent of the quantities $X_1,\ldots,X_n$ through a function f, $Y=f(X_1, \ldots,X_n)$, and each $<x_i,s(x_i)>$ is the measurement result of a quantity X_i; the function f must be then somehow applied to the terms $<x_i,s(x_i)>$ to compute a measurement result $<y,s(y)>$ for Y;
- "measurement results comparison": all the $<x_i,s(x_i)>$ are repeated measurement results of the same quantity X, and must be compared to each other via a relation r, of which they are arguments, to establish whether such a relation holds among them or not.

Uncertainty propagation is clearly related to the first area of application, a case in which the statistical interpretation is plausibly the preferred one. The GUM suggests to compute y and $s(y)$ with distinct procedures, only in dependence on the terms x_i and $s(x_i)$ respectively:

$$y = f(x_1,\ldots x_n) \qquad\qquad s(y) = [c_i^2\, s^2(x_i)]^{1/2}$$

(in the case the quantities X_i are statistically correlated further terms should be introduced) having the factors c_i the role of "sensitivity coefficients", $c_i = \partial\tilde{f}/\partial x_i$, to weigh the corresponding standard uncertainties.

The most important example in the second area of application is the computation of what could be called "generalized equality", aimed at establishing whether two or more measurement results expressed as $<x_i,s(x_i)>$ can be considered undistinguishable with each other. The formal identity $<x_i,s(x_i)>=<x_j,s(x_j)>$, i.e. $x_i=x_j$ and $s(x_i)=s(x_j)$, is clearly a "too exact" criterion in this case, and different, more general, principles have been proposed, typically based on the set-theoretical interpretation of the terms $<x_i,s(x_i)>$. For example, if the couple $<x_i,s(x_i)>$ is interpreted as the interval $[\tilde{x}_i s(x_i), x_i s(x_i)]$ then two results could be judged "compatible" with each other in the case their intersection is non-null, as the Italian standard [UNI, 1984] suggests.

It is worth to note that the GUM does not even mention this second category of applications. For important problems such as the definition of the procedures to compare national standards and express the results of the comparison an agreed position is still an open issue.

5. (Non-)exactness as (non-)specificity and (un)certainty

The innovative conceptions of the GUM are surely of the greatest importance for both scientific advances and practical applications, and hopefully will finally remove some still widespread metaphysical assumptions such as the ones related to the concept of "true value". On the other hand, the GUM appears to be still very traditional in some of its groundings, and in particular in its assumption that the measurand value and its estimated uncertainty be definitely distinct values, both expressed as scalar and interpreted within the probabilistic framework.

On the contrary of the GUM viewpoints, we suggest that the non-exactness of a measurement result should be formalized as a combination of *two* distinct conceptual components, which can be called "uncertainty" and "non-specificity". An example is helpful to introduce the meaning of such concepts and their relations. Let us consider the following two statements:
A = "this is a 120-page book"
B = "this is a book"
aimed at expressing the knowledge of an observer on a given thing under examination. On the basis of the form of A and B, two conclusions can be immediately drawn:

- A entails B: if A is true then also B must be true (in set-theoretical terms, A is a subset of B); therefore A is more specific than B;
- regardless of the particular uncertainty assignment chosen, A is at most as certain as B, and plausibly more uncertain than it.

Hence the same formal expression, $<x,s(x)>$, admits two distinct, and actually opposite, meanings:

- the measurand value, the singleton x, is maximally specific but uncertain, with uncertainty $s(x)$;
- the measurand value, the interval $[\tilde{x}s(x), \underset{\cdot}{x}s(x)]$, is not completely specific but considered certain.

The same empirical information, as obtained by the measurement, can be represented by suitably balancing the specificity and the certainty of the result (for example, if the statement A cannot considered certain on the basis of the available information, then:
C = "this is a more-than-100-page book"
could be adopted, less specific but more certain than A. In the case the object is factually a 120-page book, then A would not be "truer" than C, but more specific, and therefore more informative, than it).

Here one might see the relevance of Zadeh's classification referred to in section 1, which is considering both (non-)specificity and (un)certainty.

6. Conclusion

The traditional standpoint asserts that each quantity measured on a given thing has a "true" value, expressed as a singleton, i.e. a maximally specific value. A less-than-complete knowledge on that quantity would generate an uncertainty on such a true value, which, as a consequence, remains hidden to the observer. A clear distinction between the empirical world, to which the measured thing belongs, and the linguistic/symbolic world, to which the measurement result belongs, shows that this standpoint cannot be maintained against any operational analysis of the epistemic role of measurement and the way it is performed. In this view the ISO Guide to the expression of uncertainty in measurement is a mix of innovation and conservation. It has removed the metaphysical assumption of the existence of a "true" value, but it has substantially preserved the requirement that the measurement value be formalized as a singleton, coupled with an estimation of its uncertainty.

A more general position is based on a clear distinction between the evaluation of the measurement result and its expression. The empirical component of the measurement (the one usually performed by a sensor) is aimed at gathering information from the field on the measurand. Once this operation has been completed, the observer is able to evaluate such information, whose quantity depends on the quality of the measurement system and procedure adopted. The same information can be then formally expressed and interpreted in different ways, in view of specific applications. In particular, the *same* measurement result could be formalized as an uncertain singleton or an interval, or (if we allow to loose the compatibility with the GUM) a more general entity, such as a probability distribution or a fuzzy set. Because of this distinction, we can conclude that measurement results are neither specific nor non-specific, and neither certain nor uncertain. They are neither singletons-with-uncertainty-degrees nor intervals.

Entities such as singletons, intervals, probability distributions, fuzzy sets, ... belong to the linguistic/symbolic world, not to the empirical one: that is why the same empirical state (e.g. a measurand evaluated on a given thing) can be formally expressed in so different ways. In such a formalization the issue is not truth, but adequacy to given goals.

7. References

Bridgman P.W.(1959) How much rigor is possible in physics? In: Henkin et al. (eds.), The Axiomatic method, North-Holland.
CIPM, BIPM (1981) Proc. Com. Int. Poids et Mesures, 49, 8-9, 26 (in French) (also reported in the Introduction of the GUM).
ISO, et al., (1993) Guide to the expression of uncertainty in measurement (GUM). International Organization of Standardization, Geneva.
Mari L. (1999). Notes towards a qualitative analysis of information in measurement results, Measurement, 25, 3, 183-192.
Mari L.(2000) Metrological issues in weak scale type evaluation, Proc. IMEKO XVI.

Mari, L., Zingales, G.(2000) Uncertainty in measurement science. In: Measurement Science – A discussion, K.Karija, L.Finkelstein (eds.), Ohmsha, IOS Press, p. 135-147
Pattee H.H. (1989) The measurement problem in artificial world models, BioSystems, 23: p.281-290
Rosen R. (1987) On the scope of syntactics in mathematics and science: The machine metaphor. In: Casti J. and A.Karlqvist, Eds. Real brains, artificial minds, North Holland, Amsterdam, p. 1-23
Taylor N., Kuyatt C. (1997). Guidelines for evaluating and expressing the uncertainty of NIST measurement results. NIST Technical Note 1297 (also available in the web: physics.nist.gov/Pubs/guidelines/contents.html).
UNI, 1984. Misure e misurazioni. Termini e definizioni fondamentali. Italian Standard UNI 4546 (in Italian).
Zadeh ***
Add here and pages for Mari's paper