XVI IMEKO World Congress

Measurement - Supports Science - Improves Technology - Protects Environment ... and Provides Employment - Now and in the Future

Vienna, AUSTRIA, 2000, September 25-28

METROLOGICAL ISSUES IN WEAK SCALE TYPE EVALUATION

L. Mari

Università C. Cattaneo LIUC, Castellanza (Va), Italy

Abstract: The paper discusses the metrological conditions required for evaluations performed in weak scales to be considered measurements. The problem of uncertainty expression is taken into account, and the need of models applicable to ordinal and nominal scale measurement is highlighted.

Keywords: Foundations of measurement, Measurement theory, Nominal scale

1   INTRODUCTION

It is now a more and more recognized fact that evaluations performed in algebraically weak scale types, in particular ordinal and nominal ones, can be aptly considered as peculiar forms of measurements. The scientific and technological usage of these scale types is not new, grounding disciplines such as pattern recognition and estimation of preferences and utility. Previous works have already started to analyze the commonalities between such fields and measurement, with the explicit goal of applying already obtained results within the metrological context (e.g. [1]). Nevertheless the very metrological component in this research still requires further investigation, to make the application of old concepts to the new field unambiguous in its potentialities and limitations.

The fundamental issue is to characterize the conditions required for an evaluation performed in a weak scale to be adequately considered a measurement, thus assuring the objectivity and intersubjectivity of its results. Specifically, suitable models to express the uncertainty of such results have to be found.

2   WHY WEAK SCALE TYPE MEASUREMENT IS A PROBLEM

Measurement is an operation aimed at faithfully expressing empirical information by means of symbols so that inferential processes among symbols can be performed in place of empirical handling among things. Critical is therefore the characterization of such a concept of faithfulness, from which the criteria of correct substitution of things with symbols are derived. Indeed, the fact of associating a symbol with a thing does not guarantee in itself that the result of the symbol processing can be meaningfully back propagated to things.

According to the standard concept of measurement, as embodied in the so-called representational point of view to measurement theory [2], the formal properties of any measurement can be abstractly synthesized in terms of the scale type in which the measurement itself is performed. As a consequence, different "levels of measurement" have been identified, each of them characterized by the empirical structure that is homomorphically mapped into the chosen symbol set.

This perspective, that potentially allowed broadening the long-established classical concept of measurement, has not been undisputed, the objections tracing back to two basic, intertwined sources:

• According to the Euclidean standpoint, measurement is the empirical counterpart of unit counting, so that only absolute (the unit is fixed in the measurand definition) and ratio (a unit is empirically defined, although in a conventional way) scale evaluations are deemed as actual measurements; interval scale type (a unit is not empirically defined, but values still maintain a metric) was subsequently added, to include quantities such as temperature (before the Kelvin scale) and potential energy in the group of "measurable physical quantities". On the other hand, weaker scale types, for which a metric is not defined, were "too different" from this paradigm to be accepted in the frame of measurement;

• The mathematical expressions by which the physical laws are formalized include operators (such as sums and products) that can be meaningfully applied only to values embedded in a scale involving a unit or at least a metric, a condition that weak scale types do not fulfill; evaluations performed in such scales appear therefore useless in the common function of measurement, a fact that led to consider them as distinct from proper measurement.

The discussion on "the quantity of algebraic structure" that an evaluation should embed to be called a measurement is purely nominalistic, the scale type theory [3] having shown the gradualness of the

transition from strong to weak scale types. Any threshold from "actual measurements" to "non-measurements" is thus drawn on the basis of a convention [4], and therefore is immaterial. On the other hand, a distinction surely remains at the application level. A "physical law" applicable to nominal scale values is at most able to assert some equivalence relations, formalizing the undistinguishability of the considered things with respect to the specific application (the inferences that can be drawn are of the kind: for any given three things a, b, and c, if it is empirically known that eq(a,b) and eq(a,c) then the conclusion can be reached on purely formal bases that eq(b,c)), and to compare equivalence classes with respect to their cardinality (in this case inheriting the "extensive" structure of the natural numbers, so that, e.g., if #A=#B then a biunivocal relation can be defined between the classes A and B).

The Author suggested in a previous paper [5] that the transition from strong to weak scale types can be interpreted as a progressive reduction of the quantity of semantic information conveyed by measurement results. This explains the emphasis towards strong scale types (in this view [6] has hypothesized an evolutionary path from nominal to absolute scale type), but at the same time confirms that nominal evaluations could be actual measurements: their metrological characterization must be found not in formal assumptions but in their empirical features. Therefore: can (and under which conditions) classifications and orderings be considered measurements?

3   THE CONDITIONS FOR WEAK SCALE TYPE MEASUREMENT

Peculiar to measurement is the claim of being an operation both objective (i.e. able to produce information in principle related to the measurement object, and not the measuring subject) and intersubjective (i.e. able to produce information that in principle any subject interprets in the same way) [7]. Moreover, a wide agreement is emerging on the fact that measurement results should include the indication of not only the measurand value, but also the uncertainty with which such a value has been evaluated [8], [9]. The issue is therefore to make the conditions explicit that assure that these characteristics are fulfilled also in the case of nominal and ordinal evaluations.

Both objectivity and intersubjectivity are guaranteed by the adoption of an empirical measuring system suitably referred to a measurand standard through a traceability chain [7]. An instrument with these characteristics and producing nominal scale values (in the following for the sake of generality this type will be mainly considered) acts as a comparator operating in reference to a standard made of a set of "prototypes", each of them being the representative of one of the equivalence classes that the instrument itself is able to discriminate. A label is univocally associated with each prototype, so that a measurement result indicates the label corresponding to the prototype empirically determined as undistinguishable to the measured thing. Since a nominal scale does not include any empirical operation but only an equivalence relation, the standard cannot be obtained by means of the usual "self-generative" procedures (such that, given the unit u and a combination operation op, the standard is given by iterating op on u: 2u=op(u,u), 3u=op(op(u,u),u), …) and therefore has to be defined in extensive way. It follows that the scale is inherently finite.

Moreover, the lack of a metric implies that some of the parameters by which the measuring systems are usually characterized, such as sensitivity and resolution, cannot be employed, and must be replaced by the weaker concept of "discrimination power". Indeed, each measurement procedure induces a partition on the set of the measured things, and the collection of such partitions is structured, as usual, as a partially ordered set (a "poset"). The refinement between partitions corresponds to an increase of the discrimination power of the corresponding procedures: a procedure is enhanced by enabling it to discriminate things that were previously dealt with as undistinguishable.

Nevertheless, the metrologically fundamental procedure of calibration is still required, to assure that an unbroken traceability chain links the standard and the measuring system.

The non-ideality of such a chain empirically emerges in two complementary ways: reduction of discrimination and uncertainty.

The first case refers to the situations in which distinct prototypes of the assumed standard cannot be discriminated by the measuring system, so that the traceability chain can be formalized by a non-injective function from the standard set into the symbol set of the measuring system. This case can be easily dealt with from both the empirical and the formal viewpoints in terms of the quantity of syntactic information conveyed by the evaluation [5]. If, for the sake of simplicity, each of the n elements x_i of the standard set is assumed to have the same a priori probability (so that the Hartley definition can be applied: $QInf(x_i) = \log_2(n)$ bit), the value $\log_2(n)$ bit itself is the maximum information quantity that a measurement result conveys. A poor measuring system, with a reduced discrimination power, will give a correspondingly reduced information quantity.

Different, and more challenging, is the case of uncertainty (whose indication should be always given, as recommended by the ISO "Guide to the expression of uncertainty in measurement" (GUM) [8], to document the quality that is attributed to the evaluation), here manifesting a doubt on the

correctness in the classification of the evaluated thing. Various techniques to express such an uncertainty can be applied: all share the fact that the rule f mapping the set A of the evaluated things into the set B of the symbols by which the elements of the standard set are labeled cannot be formalized as a function A→B. A more general strategy must be adopted: for example the range of f could be chosen as the set of subsets of B, or the set of fuzzy subsets of B, or the set of the probability distributions on B, or the set BxR of couples (measurand value, estimated standard deviation of it) as the GUM basically suggests.

4   BEYOND THE GUM?

The general fact remains that the GUM recommendations cannot be applied for nominal or ordinal scales, since neither mean values nor standard deviations are meaningful in this case (for the concept of meaningfulness as formalized by the representational point of view, see [10]). In fact, the whole frame of the GUM depends on the (implicit) assumption that the reference scale be at least of interval type. Due to this fact, the GUM itself cannot be applied for weaker scales. The need arises therefore to complement the GUM with a wider set of recommendations, also applicable to ordinal and nominal scale measurement.
From the analysis of the hypotheses on which the GUM is based some teaching can be achieved for this goal:

A. the GUM imposes that the measurand value and its uncertainty must be expressed and dealt with in distinct ways;

B. by taking into account both "type A" and "type B" uncertainties and recommending to express the uncertainty as a standard deviation in both cases, the GUM leaves substantially undefined the conceptual and operational bases of type B uncertainties;

C. as already noted, by recommending to express the uncertainty as a standard deviation, the GUM restricts its scope to at least interval scale measurement.

The hypothesis A is reasonable whenever the measurement results are to be treated by "classical tools", i.e. differential equations. Indeed, physical laws leave uncertainties apart, and handle measurand values only. But if the available formal tools change, and this is the case of weak scale types, why not to formalize the measurand value and its uncertainty as a single entity?
The hypothesis B surely creates a powerful, because unified, framework to propagate uncertainty, but forces the adoption of probability even in non-statistical situations. When subjective information has to be formalized, probability is just one of the possible "uncertainty measures", and specifically the only one characterized by the additivity axiom. Therefore the GUM implies that any uncertainty evaluation must be additive, whereas this assumption could be relaxed (why should subjective information always be additive?) by adopting, e.g., plausibility and credibility (i.e. "lower" and "upper probability") pairs [11].
The hypothesis C grounds the uncertainty expression on a formally solid basis, but leaves completely unanswered the issue on how to express and process the uncertainty in the case of nominal or ordinal scale measurement.
These cases share a common point: only "strong information" situations are dealt with, in which the classical body of mathematics can be suitably applied.
It is a matter of fact that also "weak information" situations are experimentally encountered. Therefore, an interesting research topic is to find some general technique to complement the GUM recommendations in such cases.

REFERENCES

[1] H. Watanabe, Theory of classification of objects by predicates, *Measurement* **18** (1) (1996) 59-69.
[2] D. Krantz, R. Luce, P. Suppes, A. Tversky, *Foundations of Measurement*, Academic Press, New York, Vol.1 (1971), Vol. 2 (1989), Vol. 3 (1990).
[3] S. Stevens, On the theory of scales of measurement, *Science* **103** (2684) (1946) 677-680.
[4] S. Stevens, Measurement, psychophysics, and utility, in: West Churchman, Ratoosh (eds.), *Measurement: definitions and theories*, J. Wiley, New York, 1959.
[5] L. Mari, Notes towards a qualitative analysis of information in measurement results, *Measurement* **25** (3) (1999) 183-192.
[6] R. Carnap, *Philosophical foundations of physics*, Basic books, New York, 1966.
[7] L. Mari, Beyond the representational viewpoint: a new formalization of measurement, under publication in the journal *Measurement*.

[8] ISO et al., *Guide to the expression of uncertainty in measurement*, International Organization of Standardization, Geneva, 1993.

[9] L. Mari, G. Zingales, Uncertainty in measurement science, *Proceedings of the International Workshop Imeko TC7*, Kyoto, 1999, p.177-189.

[10] F. Roberts, *Measurement theory*, Addison-Wesley, Reading (Ma), 1979.

[11] G. Shafer, *A mathematical theory of evidence*, Princeton Univ. Press, Princeton, 1976.

AUTHOR: Dr. Luca MARI, Università C. Cattaneo LIUC, C.so Matteotti 22, 21053 Castellanza (Va), Italy, Phone: ++39 0331 572228, Fax: ++39 0331 572320, Email: lmari@liuc.it