AMT-2021-157: 2nd Review

Antonio Possolo

January 15th, 2022

1 Summary

The article under review, *Truth and Uncertainty. A critical discussion of the error concept versus the uncertainty concept*, by Thomas von Clarmann, Steven Compernolle, and Frank Hase, deserves to be published because it is provocative and questions what has become uncritically "accepted wisdom" among many who follow the guidance in the GUM and in its supplements.

I am of the opinion that reviewers ought not to attempt to force authors to express only views that the reviewers share. It should be even less so in the case of a contribution like this one, of von Clarmann's *et al.*, which is more an opinion piece than a technical article, welcome nonetheless.

Below I offer assorted comments that do not add up to a thorough and complete review, which the authors and the editor may like to consider, and, in dialog with one another, possibly reach a settlement about next steps. I believe that the review I submitted of the original submission was quite thorough and detailed, and I appreciate the attention that the authors have obviously given to it.

I will abstain from making detailed comments about how the article is written in English, or offering specific suggestions for how it can be improved on this count. But I will make the following general remarks: the facts and arguments the authors present can, by and large, generally be understood by anyone with a modicum of familiarity with the relevant subjects; however, the article is far from being written in good scholarly English, the punctuation being particularly defective, which occasionally actually hinders understanding.

2 Comments

2.1 L078 *The error has been conceived as an attribute of a measurement or an estimate, while the term 'uncertainty' has been used as an attribute of the true state*

> I would say that the uncertainty is an attribute of the relationship between the person making the measurement and the true value of the measurand, characterizing the incomplete knowledge that person is left with following the measurement.

2.2 L104 *The estimated error is understood as a measure of the width of a distribution around the measured value which tells the data user the probability density of a certain value to be measured if the value actually measured was the true value.*

> The simplest measurement model expresses the measured value as $x = \mu + \varepsilon$, where μ denotes the true value of the measurand and ε denotes measurement error. The error can be positive, negative, or null, and it is modeled as a (non-observable) random variable. The spread of the corresponding probability distribution, gauged using the standard deviation or any other similar metric, is indicative of the typical magnitude of the error, but it is not the error itself.

2.3 L244 *In GUM the error concept is discarded because the capability of conducting an error estimate allegedly depends on the knowledge of the true value.*

> The theory of statistical estimation [Lehmann and Casella, 1998] has never shied away from entertaining quantities whose values are unknown, be they parameters of interest, or non-observable errors. There are statistical procedures that provide estimates of individual errors, not just summary estimates of their typical magnitude: for example, cross-validation [Stone, 1978] of a (linear or nonlinear) regression model.

2.4 L419 *Some interpretations of GUM-2008 [...] seem to suggest that error estimation and uncertainty analysis are best distinguished in the sense that the former relies on frequentist statistics while the latter is founded on Bayesian statistics.*

The authors are quite right when they dismiss this myth, variously and repeatedly throughout the article.

2.5 L462 *The uncertainty concept relies on the possibility of evaluating uncertainties caused by measurement errors and "systematic effects" without knowledge of the true value. This is certainly granted for linear problems. Here the uncertainty estimates do not depend on the value of the measurand. This is because in the linear case Gaussian error propagation holds.*

I do not understand what the authors mean when they say "This is certainly granted for linear problems." And regarding the third sentence: there are nonlinear models where the typical size of the errors does not depend on the value of the measurand, and there are linear models where it does. Finally, what is "Gaussian error propagation?"

2.6 L504 *The situation is more difficult for biases. Biases between different measurement systems do not tell us what the bias of one measurement system with respect to the — unfortunately unknowable — truth is. Even if the number of measurement systems is quite large, it is not guaranteed that the mean bias of all of them is zero.*

Bias (differently from how the VIM [Joint Committee for Guides in Metrology, 2012] defines it) is the difference between the mathematical expectation of an estimator and the quantity being estimated.

The situation is not as desperate as the authors paint it: the statistical jacknife [Mosteller and Tukey, 1977] [Efron, 1982], the statistical bootstrap [Efron and Tibshirani, 1993], and cross-validation [Mosteller and Tukey, 1977] can provide useful estimates of bias without assuming that the target of estimation is known.

As the authors note around L513, by intercomparing independent measurement results for the same measurand, one can ascertain at least whether all important sources of uncertainty have been accounted for in the respective uncertainty budgets or not.

If there should be lurking uncertainty components that manifest themselves only when measurement results (measured values and their uncertainties) are compared — so-called *dark uncertainty* [Thompson and Ellison, 2011] —, then this may be attributed to yet undetected and still unexplained effects, which can be persistent — that is, biases —, or volatile.

The so-called *Hubble Tension* illustrates such situation, in relation with which Riess et al. [2019] makes the following remark: "pinpointing the cause of the tension requires further improvement in the local measurements, with continued focus on precision, accuracy, and experimental design to control systematics."

References

B. Efron. *The jackknife, the bootstrap, and other resampling plans*, volume 38 of *CBMS-NSF Regional Conferences in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1982. ISBN 0-89871-179-7.

B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, London, UK, 1993.

Joint Committee for Guides in Metrology. *International vocabulary of metrology — Basic and general concepts and associated terms (VIM)*. International Bureau of Weights and Measures (BIPM), Sèvres, France, 3rd edition, 2012. URL `https://jcgm.bipm.org/vim/en/`. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, JCGM 200:2012 (2017 version with minor corrections and informative annotations).

E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer-Verlag, New York, NY, 2nd edition, 1998. ISBN 978-1-4419-3130-6.

F. Mosteller and J. W. Tukey. *Data Analysis and Regression*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1977. ISBN 0-201-04854-X.

A. G. Riess, S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic. Large Magellanic Cloud Cepheid Standards provide a 1% foundation for the determination of the Hubble Constant and stronger evidence for physics beyond ΛCDM. *The Astrophysical Journal*, 876(1):85, May 2019. doi: 10.3847/1538-4357/ab1422.

M. Stone. Cross-validation:a review. *Series Statistics*, 9(1):127–139, 1978. doi: 10.1080/02331887808801414.

M. Thompson and S. L. R. Ellison. Dark uncertainty. *Accreditation and Quality Assurance*, 16:483–487, October 2011. doi: 10.1007/s00769-011-0803-0.