TEN THESES FOR A NEW GUM

Michael Grabe

Germany, 38104 Braunschweig, Am Hasselteich 5, michael.grabe@gmx.net

Abstract: Since the official appearance of the *Guide to the Expression of Uncertainty in Measurement*, GUM for short, in 1993, discussions concerning its appreciation never became silent. Nowadays, to argue against the GUM appears as heretical as naïve since the dice seem to have been cast however demotivating its implications may be. The following contribution elaborates a poster presentation of the same title pointing to the deficiencies of the GUM as seen by the author. Each thesis is briefly motivated and provided with a proposal for rectification. Though the GUM seems to be insistently moored, the author hopes to trigger a discussion aiming at the foundation of the GUM.

1. Introduction

The discussion, elaborating a poster presentation [5], is based on the uncertainty $u_{\bar{x}}$ assigned to an arithmetic mean

$$\bar{x} = \frac{1}{n}\sum_{l=1}^{n} x_l \tag{1}$$

out of n repeated measurements $x_1, x_2, \ldots, x_n$. The uncertainty as laid down by the GUM is given by

$$u_{\bar{x}} = k_P \sqrt{\frac{s_x^2}{n} + \frac{f_s^2}{3}}\,. \tag{2}$$

The extension factor k_P is recommended to be set equal to 2. While

$$s_x^2 = \frac{1}{n-1}\sum_{l=1}^{n}(x_l - \bar{x})^2 \tag{3}$$

denotes the empirical variance, $\pm f_s$ marks the endpoints of the interval

$$-f_s \leqslant f \leqslant f_s \tag{4}$$

confining the unknown systematic error f. Via postulate the GUM randomizes unknown systematic errors. May it suffice to allocate a rectangular density to f, the density providing the time-constant perturbation with a theoretical variance

$$\sigma_f^2 = \frac{f_s^2}{3}\,. \tag{5}$$

In the following we shall refer to the formulas (1) - (5).

2. Ten Theses for a New GUM

As much has been said about the GUM, the author summarizes his concerns in terms of theses.

Thesis 1: Localization of True Values

The uncertainty should lay down the smallest possible interval which localizes the true value of the measurand (quasi-) safely [1], [4].

Discussion: The GUM raises no claim to localize the true values of measurands. Metrology, however, should do just this, as the true values of measurands constitute the formulas of physics. Certainly, only the localization of true values safeguards traceability, metrology's basic concept.

Uncertainty quotations being padded up too largely might obscure meaningful physical effects. At the same time, quotations being too tight might predict effects which do not exist. Metrology should rely on uncertainties being neither "optimistic" nor "pessimistic" but objective.

Thesis 2: Separation of Random and Systematic Errors

Error propagation should map the physical operation of the measuring device. This proceeding would separate the flow of random and systematic errors, [2].

Discussion: Unknown systematic errors get fixed prior to measurements and, if everything works well, remain constant in time, at least during the taking up of repeated measurements. Though the actual value of some unknown systematic error is not known, but, whichever value this may be, subsequent repeated measurements are burdened by the same time-constant perturbation. As metrology has to be traceable, there is hardly leeway for randomization procedures.

Hence, the formalism of error calculus should map the physical time-constancy of systematic errors and treat systematic perturbations consistently as constants. This would separate the flow of random and systematic errors, thus allowing different, specific error treatments according to pertaining intrinsic properties.

Thesis 3: Randomizations versus Biases

Unknown systematic errors should not be "randomized", rather their interpretation as being time-constant perturbations should lead to the introduction of biases charging weighted and unweighted means.

Discussion: The randomization of unknown systematic errors evokes a complicated mathematical formalism based on convolution procedures which, unfortunately, lack a physical basis: The experimenter cannot carry out the simplest of all convolutions of, say, a normal and a rectangular density, simply because he does not know the theoretical standard deviation of the normal density. However, there are more concerns: To illustrate this, let us cast (2) into

$$u_{\bar{x}} = \sqrt{(k_P \frac{s_x}{\sqrt{n}})^2 + (k_P \frac{f_s}{\sqrt{3}})^2}$$

which discloses the observation that the uncertainty as specified by the GUM is due to a quadratic combination of two random components, namely

$$k_P \frac{s_x}{\sqrt{n}} \quad \text{and} \quad k_P \frac{f_s}{\sqrt{3}}.$$

Obviuosly, the expanded $k_P = 2$ uncertainty shifts the unknown systematic error f by a factor of $2/\sqrt{3}$ outside the domain $-f_s \; ... \; +f_s$ as assessed by the experimenter thus overriding his estimation. In order to prevent f from crossing its boundaries, the k_P- factor should be limited to $\sqrt{3}$. But this would be equivalent to putting $f = f_s$ rendering the procedure to randomize unknown systematic errors superfluous right from the outset, as $f = f_s$ is equivalent to a worst case estimation. However, at the same time, $k_P = \sqrt{3}$ would be too small for metrological purposes.

On the other hand, the introduction of biases would offer a natural way to guide the GUM out of the deadlock. Also, biases would not render error calculus more intricate. Rather, the formalism would get simplified and more transparent as compared with a series of convolutions being, incidentally, impracticable. In particular, the introduction of biases would provide traceability.

Thesis 4: Equal Numbers of Repeated Measurements

Error propagation should be based on variables endued with equal numbers of repeated measurements as this would allow to bring empirical covariances to bear — be the measurements dependent or not.

Discussion: Let us consider two normally distributed variables. Then, there is a three-dimensional density of the empirical variances and the empirical covariance. In case we wanted to exploit this density, we should not dismiss the empirical covariance, be its expectation zero or not. Dismissing the empirical covariance would mutilate the distribution density and would impose so-called effective degrees of freedom on experimenters , a concept which is difficult to handle and breaks down, given the variables are dependent. Empirical covariances clearly imply that we should have the same number of repeated measurements in each of the two series of input data as with unequal numbers there in no empirical covariance.

Thesis 5: Empirical Covariances

To consider empirical covariances allows the influence of random errors to be assessed by confidence intervals according to Student. This would apply to any kind of error propagation including least squares.

Discussion: The benefit of equal numbers of repeated measurements may be seen from the empirical variance which is commonly assigned to the series expansion of the function concatenating the variables of error propagation. The said statistic includes, of course, the empirical variances of the two series of input data, but it should also include their empirical covariance, be there a dependence or not. Only if this happens to apply are we in a position to say that this very statistic, properly written, is χ^2 - distributed. This, in the end, will lead us to confidence intervals according to Student and appears a considerable progress as compared with the application of effective degrees of freedom.

We should not flinch from considering the empirical covariance given the two series of input data are independent. Though it is true that a particular empirical covariance will influence the χ^2 - variable and will also influence the length of the confidence intervall to be designed. These observations, however, will not bother us as the actual length of confidence intervals are alway sample-dependent.

Thesis 6: Worst Case Estimations

The influence of systematic errors should be assessed through worst case estimations.

Discussion: GUM‘s 2σ - uncertainties cause the unknown systematic errors to cross the interval limits as layed down on the part of experimenters. Unfortunately, even this surcharge is lost by GUM‘s geometrical error combination and by GUM‘s inability to assign confidence levels to overall uncertainties.

If, by contrast, the flow of random and systematic errors is separated as proposed in thesis 2, the uncertainty component which expresses the influence of systematic errors may be submitted to a worst case estimation. This would not be an overestimated but a reasonable account of systematic errors.

Thesis 7: Overall Uncertainties

The uncertainy components of random and systematic origin should be added arithmetically.
Discussion: The dilemma of the GUM is that random and systematic errors are combined geometrically. While GUM's expanded 2σ uncertainties overestimate the influence of systematic errors, they underestimate the influence of random errors since for them, the k_P - factor should be larger than 2. The situation is worse, however, as k_P should not exceed $\sqrt{3}$. These deficiencies cannot be corrected as there is no way to assign confidence levels to GUM's geometrically devised overall uncertainties. Consequently, the experimenter cannot know whether or not a given uncertainty localizes the true value.

By contrast, combining the uncertainty components due to random and systematic errors linearly unleashes the intrinsic properties of these unequal kinds of measurement errors. Due to the linear combination, the differing impacts of random and systematic errors are no longer blurred, as they would be were they combined geometrically. Hence, experimenters may attribute something like "quasi-safeness" to linear combinations of random and systematic uncertainty components [1] - [3].

Thesis 8: The Method of Least Squares

Unknown systematic errors result in a breakdown of the Gauss-Markoff theorem depriving experimenters of weigh-factors. Whenever weightings are indispensable, weight-factors may be chosen by trial and error only, given the localizations of the true values of the least squares estimators are known to be maintained.

Discussion: In least squares, the weight-factors are to be drawn from the Gauss-Markoff theorem. To recall, weighting procedures imply two things:

- they shift the adjusted parameters and
- they reduce the uncertainties of the shifted parameters.

On the other hand, the Gauss-Markoff theorem is known to break down, given the input data are biased, i.e. if systematic errors are effective as constants. Hence, experimenters have no longer weight-factors at hand. In particular they cannot know whether the two proceedings, namely, the individual shifts of the adjusted parameters and the individual reductions of pertaining uncertainties have abrogated the localizations of true values - given localizations have existed at all prior to the setting of weights.

Unfortunately, the GUM cannot safeguard the localization of true values, whether or not the linear system has been weighted. This observation may be proved by computer simulations: The more the unknown systematic errors tend to exhaust their intervals, the more often the localizations of the true values may break down.

The theses suggested, render the localizations of true values quasi-safe, i.e. localizations do not depend on the particular choice of weight-factors. Consequently, it is admissible to control choices by considering directly the uncertainties of the least squares estimators, whereat trial and error will assist to reduce uncertainties without raising the menace of delocalizations, [1] .

Thesis 9: Uncertainty Spaces

The uncertainty spaces of a set or subset of least squares estimators should be composed of

- confidence ellipsoids, due to Hotelling, expressing random errors and
- security polytopes, as proposed by the author, expressing the influence of unknown systematic errors.

Discussion: Consider a least squares adjustment concerning *r* unknowns aiming at *r* pertaining true values. Then, let there be an application which asks the experimenter where to find the *r*-tuple of true values with respect to the *r*-tuple of least squares estimators.

To express the influence of random errors, a confidence ellipsoid based on Hotellings‘ distribution density should be used. Such an ellipsoid is based on the empirical variances and empirical covariances of the least squares estimators and "breathes" in a statistically well defined way – quite in contrast to the confidence ellipsoid often referred to in classical error calculus being built up by theoretical, i.e. unknown theoretical variances and covariances.

To express the influence of systematic errors, it is suggested to introduce, what the author has called "security polytopes". Such solids, hitherto unknown in error calculus, result from singular linear mappings of the equations which convey propagated systematic errors. The linear combination of a confidence ellipsoid and a security polytope produces a point symmetric overall uncertainty space which may be expected to localize the *r*-tuple of true values with respect to the *r*-tuple of least squares estimators quasi-safe [1].

Thesis 10: Analysis of Variance

Unknown systematic errors burdening empirical data result in a breakdown of the analysis of variance.
Discussion: Given the error model considers and treats unknown systematic errors as time-constant quantities, the analysis of variance clearly breaks down. But then, at least with respect to the purposes of metrology, this for decades higly esteemed tool to investigate empirical data would cease to exist.

To test the mutual compatibility of a set of means, uncertainties might be compared directly.

3. Conclusion

In view of the author, the deficiencies addressed do not appear remediable and call for a new GUM .

Literature

[1] Grabe, M.: *Measurement Uncertainties in Science and Technology*, Springer-Verlag, Berlin-Heidelberg, 2005

[2] -, *Principles of "Metrological Statistics"*, Metrologia 23 (1986/87) 213-219

[3] -, *Estimation of Measurement Uncertainties – an Alternative to the ISO-Guide*, Metrologia 38 (2001) 97-106

[4] -, *The Alternative Error Model and its Impact on Traceabilty and Key Comparison*, Joint BIPM-NPL Workshop on the Evaluation of Interlaboratory Comparison Data, NPL Teddington September 19, 2002

[5] -, *Ten Theses for a New GUM*, Posterpaper, PTB-BIPM Workshop on the Impact of Information Technology in Metrology, Berlin 4-8- June 2007