What can we learn from the GUM of 1993/5?

Robin Willink
robin.willink@gmail.com

AMCTM 2014, St. Petersburg, 10-12 September

Reference:
Measurement Uncertainty and Probability
R. Willink, Cambridge, 2013
Chapters 1, 4, 9, 14
Sections 2.3, 3.1, 15.5

Origin

In 1980, the BIPM 'convened a meeting for the purpose of arriving at a uniform and generally acceptable procedure for the specification of uncertainty'.

- Working Group produced a report, with Recommendation INC-1 (1980)
- CIPM approved/reaffirmed Recommendation INC-1 (1980) in 1981/86
- Acknowledgement in GUM
  - GUM [0.5] The approach upon which this guidance document is based is that outlined in Recommendation INC-1 (1980) ... Because Recommendation INC-1 (1980) is the foundation upon which this document rests, ...

Three controversies to encounter

1. Is there a true value?

   If so, what is "measurement uncertainty"?

   And if not, what is "measurement uncertainty"?

2. **Can variances be associated with constant errors?**

   If so, is it always appropriate to do so?

   Is this permissible in the field of testing?

3. Can the measurand be treated as a random variable? Can it be assigned a probability distribution?

   If so, where will it take us?

1. Measurement: estimation or declaration?

There seem to be different kinds of measurement:

1. estimating the unknowable ‘true’ value
2. stating or declaring ‘the value’ of the measurand
3. representing something not definable in other terms (e.g.abilities)

- ‘Uncertainty of measurement’ will mean different things in each case.
- A document should acknowledge this and restrict its scope accordingly.
- The GUM implies the general existence of ‘true values’ but then seems to ignore the idea.

Do we acknowledge other concepts of ‘measurement’?
How does this idea affect our documents?

VIM [2.11, note 3] When ... , the measurand may be considered to have an “essentially unique” true quantity value. This is the approach taken by the ... [GUM] ...

GUM [3.1.2] In general, the **result of a measurement** is only an approximation or **estimate** of the value of the measurand ...

But GUM[2.3.5] **expanded uncertainty**

quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand.

Estimation of a value implies ‘error’. So uncertainty represents the likely magnitude of the measurement error.

Why not involve the concepts of ‘accuracy’ and ‘error’ when our documents are based on the concept of ‘true values’?

The meaning of ‘uncertainty’

M. Mandelkern, *Statistical Science*, **17** (2002) 149–159, and respondents

Mandelkern (a physicist): physicists would like the interval to ‘convey an estimate of the experimental uncertainty’
Wasserman: what is actually meant by experimental uncertainty?

- Is it a parent standard deviation? If so it is unknown but estimated.
- Is it the standard deviation of a probability distribution describing your belief about the true value? If so it is known and calculated.

- Unless a guide takes a coherent position on what ‘uncertainty’ is, it cannot associate a practical meaning with statements of uncertainty.
- A methodology must be able to be validated using realistic simulations.

2. Report of the Working Group, 1980

Report of the BIPM Working Group on the Statement of Uncertainties (1st meeting - 21 to 23 October 1980) to the Comite International des Poids et Mesures

This then leads us ... to distinguishing between the following two major categories of uncertainty components:

- Group A: the evaluation can be based on (objective) statistical methods,

- Group B: the estimation must rely on “other methods”; this inevitably implies some element of subjective appreciation.

The traditional distinction between “random” and “systematic” uncertainties (or “errors”, as they were often called previously) is purposely avoided here, ...

Some utility lost?

- Random/Systematic - errors with different practical effects
- Type A/Type B - errors with different means of assessment

Are both these dichotomies relevant?

- Repeated measurement requires clear separation of systematic errors from random errors.
- In legal contexts a testing method must itself be testable: its performance relates to frequency of success.
- In general, testing seems to require a one-sided statement of uncertainty.

Should our documents give more attention to the purposes of measurements?

American Society of Mechanical Engineers

The Purpose for Measurements and Understanding Their Uncertainties
(The Basic, ISO and ASME Uncertainty Models Compared),
Ronald H. Dieck, Ron Dieck Associates, Inc. rondieck@aol.com
(60th International Instrumentation Symposium, London, June 2014)

ASME Model recognizes at the outset that the physical universe contains random and systematic effects and errors. It seeks to use those intuitive classifications for the error source and uncertainty groupings.

5. The differences between the ASME and ISO models may be reduced to only two significant items each: the pedigree and effect of error sources.

7. The ASME and ISO Models may be brought into complete harmony by adopting the subscripting approach to the elemental uncertainties.

3. It is further recommended that future Standards advance the concept of allowing subscripting elemental uncertainty estimates with classifications of the other Model. For example, the ISO Model elemental uncertainties would all include subscripts of “s” if the elemental source is systematic and “r” if random, and applied with reference to a single chosen hierarchal level. The ASME Model would include subscripts of “A” or “B” depending on the pedigree of the elemental uncertainty estimate. In fact, the ASME Model already allows such subscripting. The combined classification is also used in the governing NIST document (Ref. 3) and in an instructional book by Dieck (Ref. 5).

Ref. 3 is NIST Technical Note 1297, Taylor and Kuyatt, 1994
Ref. 5 is *Measurement Uncertainty, Methods and Applications*, 4th ed., Dieck, ISA, 2006
(Ron Dieck is the longstanding chairman of the committee ASME PTC19.1, Measurement Uncertainty.)

3. Report of 1980

The new approach ... recommends ... estimation of ... approximations to the variances ... needed in the ... law of “error propagation”.

The essential quantities appearing in this law are the variances (and covariances) of the variables (measurements) involved.

In these approaches it is necessary to make (at least implicitly) some assumption about the underlying population.

Recommendation INC-1

2. The components in category A are characterized by the estimated variances, s_i^2 ... and the number of degrees of freedom, ν_i. ...

3. The components in category B should be characterized by quanitities u_j^2, which may be considered as approximations to the corresponding variances, the existence of which is assumed. ...

Mixing the unmixable

- Type A analysis - the random variables alluded to are the potential results of measurements; the true value is the mean of the population

  GUM [4.2.1]... the best available estimate of the ... expected value μ_q of a quantity q that varies randomly is the **arithmetic mean** ... of the n observations

- Type B analysis - in some places, the text suggests that the true value of a quantity can be considered a random variable. But there is no mention of Bayesian statistics or fiducial probability.

  GUM [4.3.5] the probability that [(the value of) the input quantity] X_i lies within this interval [i.e. the interval a_- to a_+] is 0,5 or 50 percent

- Combination - the W-S formula is correct in a frequentist analysis. (*Metrologia* **40**(2003) 9-17 gives analogous Bayesian/fiducial formulae.)

Gleser, 'Assessing Uncertainty in Measurement',
Statistical Science **13** (1998) 277-290 (With my emphasis.)

If the BIPM/ISO approach is followed, there will be international standardization of presentations of uncertainty, at least by national laboratories; such standardization is certainly highly desirable. **The methods advocated for expressing uncertainty in the *Guide*, however, seem to form a new paradigm for statistical inference that is neither completely frequentist nor completely Bayesian, and consequently lack a firm theoretical basis.** Thus, there is concern among statisticians that the methods advocated by the *Guide* could prove to be misleading or inaccurate. Widespread acceptance of the paradigm advocated in the *Guide* within the physical science community also has the potential to increase confusion about statistical concepts and thus impede communication between statisticians and nonstatisticians.

Should our documents be usable in a course on applied statistics? Do they meet the standards of the world's best experimentalists?

An influence of the GUM?

Data Reduction and Error Analysis for the Physical Sciences 3rd ed.
Bevington and Robinson (2003) (With my emphasis.)

... we should expect that approximately 68% of our determinations of $\bar{x}$ should fall within the range $\mu - s_\mu < \bar{x} < \mu + s_\mu$. ... **we make a slight logical leap** to state that there is approximately 0.68 probability that the true value of the mean μ lies in the range $\bar{x} - s_\mu < \mu < \bar{x} + s_\mu$ or that the specified range is the 68% confidence interval.

A means of resolution and a return to INC-1

- Frequentist Type A analysis is well accepted. ('Objective Bayesian' Type A analysis offers no advantage in statistical accuracy or efficiency.)
- **The essence of Type B analysis is expert knowledge and belief. These concepts should be retained but applied to the population of errors alluded to in the report of 1980, not to the true values of influence quantities.**
- We can recover a fully frequentist description if we make the error not the 'input quantity' the subject of our belief.

The difference between frequentist and Bayesian statistics is not the idea of quantifying belief; it is the subject of that belief. Frequentist models merely describe belief about potential measurements. (See *Measurement Uncertainty and Probability*, sec. 3.1)

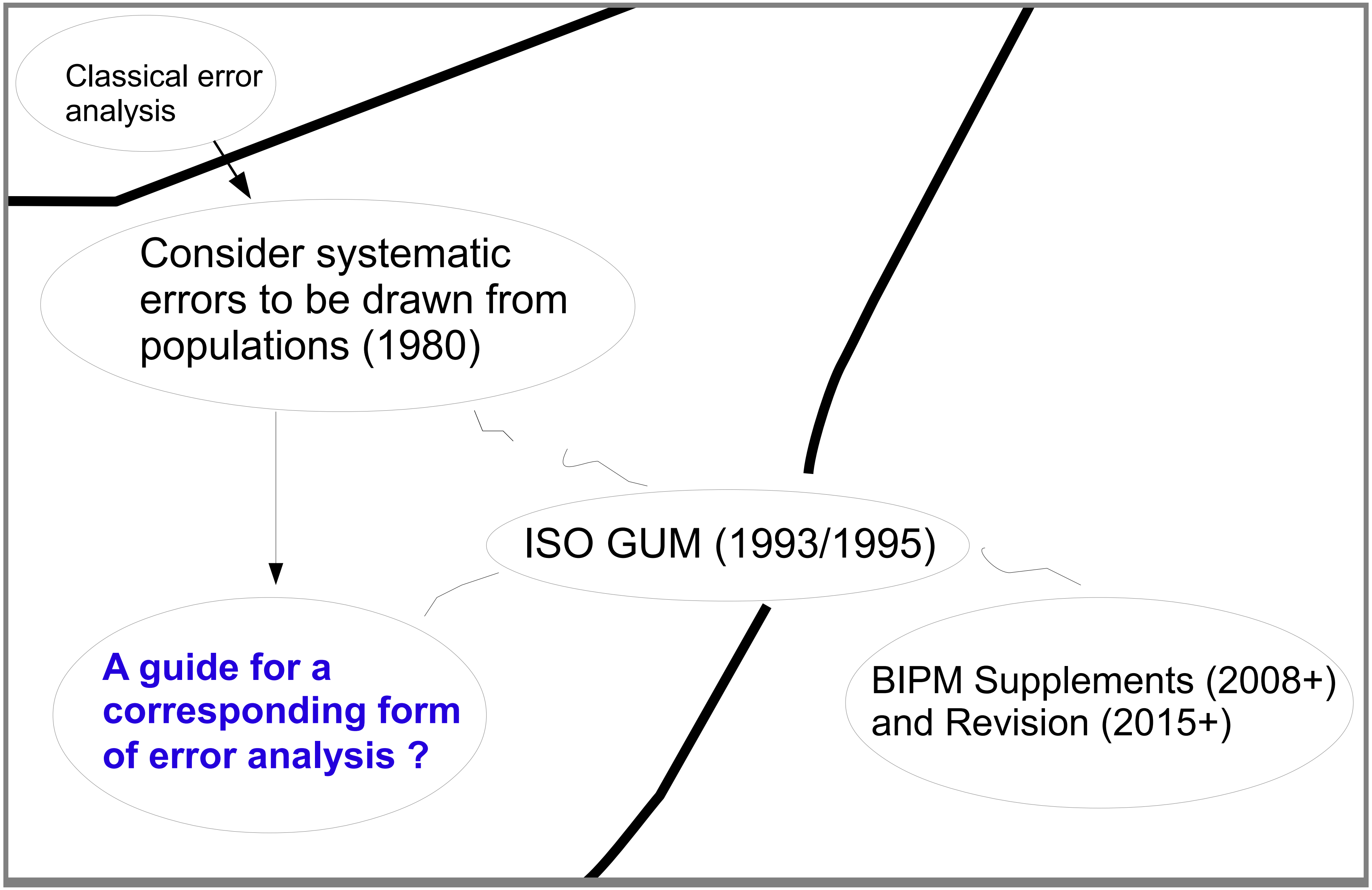
Issues of notation

- In the GUM X is an unknown true value and x is its known estimate
- If uncertainty is about X then perhaps we should write $u(X)$, not $u(x)$
- Examples of distinctions necessary for complete understanding:

$$
\begin{aligned}
G &= 6.673... \times 10^{-11}\,\mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{s}^{-2} \\
G &\approx 6.673 \times 10^{-11}\,\mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{s}^{-2} \\
\hat{G} &= 6.673 \times 10^{-11}\,\mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{s}^{-2}
\end{aligned}
$$

- We must avoid any statement like $G \approx E(G)$ in symbols or words.

Does our notation help clarify concepts?



The methods should work. But what do we mean by "work"?

How should a method perform and how can it be validated?

Paradigm shifts and residual confusion

- **bias:** The idea of seeing systematic errors as being drawn from distributions with mean zero implies that no measurement is biased!
- **correlated quantities:** True values cannot be described as correlated (or dependent). In a Bayesian analysis, we might instead say that the random variables describing our belief about them are dependent.
- **coverage probability:** From Wikipedia: **In statistics, the coverage probability of a confidence interval is the proportion of the time that the interval contains the true value of interest.**
  [Y. Dodge, (2003) The Oxford Dictionary of Statistical Terms.]

Are the differences in view becoming evident and understood?
Are the terms and symbols being chosen for compatibility?
Are the difficulties acknowledged? ■