Revision of the Guide to the expression of uncertainty in measurement — impact on national metrology institutes and industry

Maurice Cox/Peter Harris
National Physical Laboratory, Teddington, UK

CCRI
BIPM, 16 - 17 May 2013

Outline

- JCGM: current GUM and GUM Supplements
- GUM revision and the approach taken
- Why Bayesian?
- Technical aspects
- Consequences for NMIs and industry
- Summary
- Advertisement!

Joint Committee for Guides in Metrology

"The tasks of the JCGM are to maintain and promote the use of the Guide to the Expression of Uncertainty in Measurement (GUM) and the International Vocabulary of Basic and General Terms in Metrology (VIM)"

Working groups: JCGM–WG1: GUM JCGM–WG2: VIM

Member organizations of the JCGM:

BIPM	Bureau International des Poids et Mesures
IEC	International Electrotechnical Commission
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine
ILAC	International Laboratory Accreditation Cooperation
ISO	International Organization for Standardization
IUPAC	International Union of Pure and Applied Chemistry
IUPAP	International Union of Pure and Applied Physics
OIML	International Organization of Legal Metrology

JCGM documents

NPL National Physical Laboratory

JCGM 100:2008	GUM (essentially the original edition of 1993/5)
JCGM 101:2008	Propagation of distributions (Monte Carlo)
JCGM 102:2011	Multivariate version of JCGM 101
JCGM 103	Modelling
JCGM 104:2009	Introduction
JCGM 105	Underpinning concepts
JCGM 106:2012	Conformity assessment
JCGM 200:2012	VIM

Green: published – available from BIPM and Member Organization websites

Current GUM

The GUM has served, virtually unchanged, for 20 years, during which period its merits have largely outweighed its limitations

Main merit: provides conceptual framework allowing consistent treatment of uncertainties arising from both random and systematic effects

Two main limitations:

1. Lack of generality of procedure to obtain a coverage interval (an interval containing value of measurand with stipulated probability)

2. Little guidance for the case of more than one measurand

Specific guidance documents – GUM-S1 (JCGM 101), GUM-S2 (JCGM 102) – developed rather than carrying out extensive revision of GUM

Need for revision

National Physical Laboratory

In a Bayesian interpretation of probability, available knowledge about a quantity is described by a probability distribution

Current GUM:

- Bayesian view is adopted only as a way to treat non-statistical contributions to uncertainty, in the so-called Type B evaluations
- Estimates of variances of distributions are used and propagated, thus following a frequentist approach
- Degrees of freedom are attached to these estimates, legitimately, for Type A evaluations, or artificially, for Type B evaluations

Need for revision

National Physical Laboratory

Current GUM:

- Bayesian approach is introduced only to avoid a difficulty, but the frequentist view is dominant
- Approach involves a series of complications, especially in the construction of a coverage interval
- The given procedure suffers from lack of generality

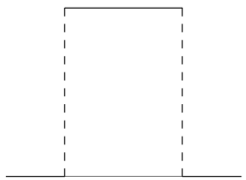
A Bayesian approach was used consistently in GUM-S1 and GUM-S2

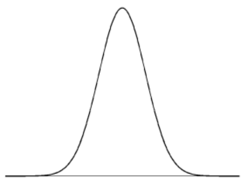
Consequence is that the GUM is no longer consistent with its Supplements $\Rightarrow$ principal motivation for decision to revise the GUM

Approach to revision

Simplest way to overcome difficulties and remove lack of generality: adopt Bayesian view

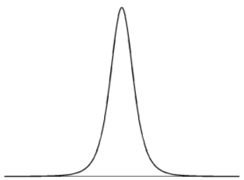
In GUM-S1 and GUM-S2, knowledge-based probability distributions assigned to the input quantities in a measurement model





Distributions propagated through model using well-established numerical procedure (Monte Carlo)

Outcome: (numerical approximation to) probability distribution for the measurand, or joint distribution for multivariate measurand

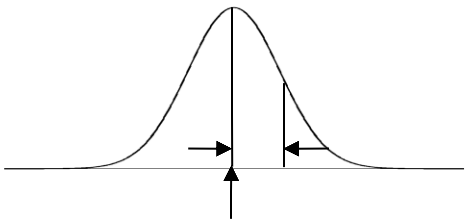


From that distribution all information needed readily obtained

Probability density functions (PDFs)

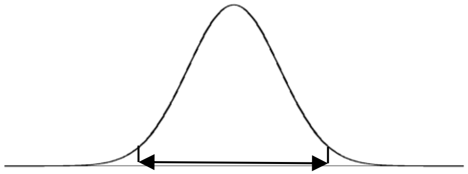
PDF: comprehensive way to express an experimenter's knowledge of the measurand

Summary data extracted from the PDF:



- expectation (mean) taken as an estimate of the true value of the measurand
- standard deviation, taken as the associated standard uncertainty

Coverage interval for any stipulated coverage probability can be obtained from the PDF



Greater attention paid to providing reliable results given the measurement model and PDFs for the input quantities established from available knowledge of those quantities

Reliability issues (current GUM)

Current GUM often interpreted in following manner:

1. Form estimate y of the measurand Y by evaluating measurement model
2. Evaluate associated standard uncertainty $u(y)$ using law of propagation of uncertainty (LPU)
3. For given coverage probability p, form coverage factor k
   - For p = 95 % take k = 2 (assumes normality)
4. Form expanded uncertainty $U_p = ku(y)$

Variant for finite degrees of freedom ν_{eff}

- For p = 95 % take k factor from t-distribution

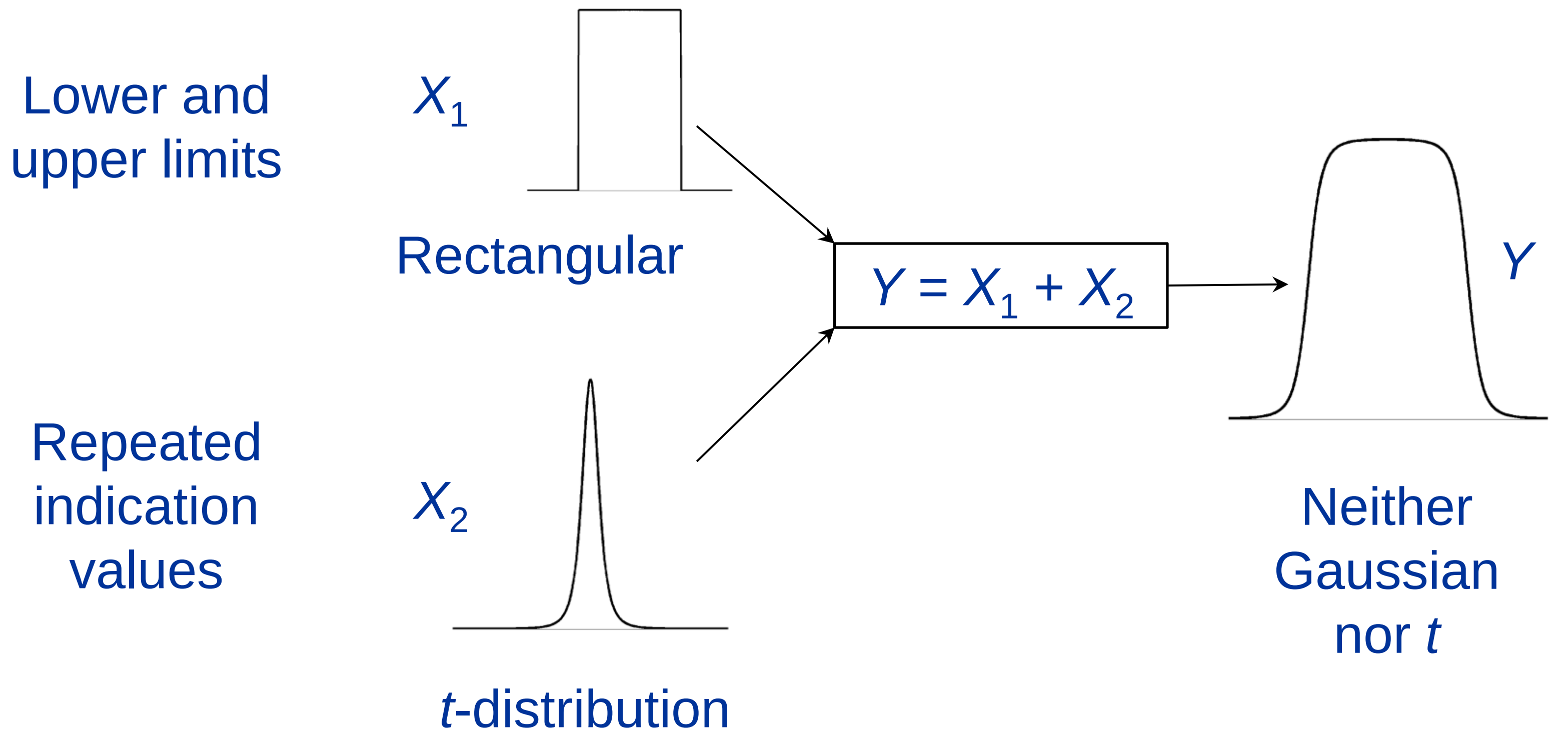
Steps 1 and 2 assume measurement model can safely be linearized

In many circumstances, assumptions justified — but not all

Greater reliability

National Physical Laboratory

Greater attention paid in revised GUM to providing reliable results given the measurement model and PDFs for the input quantities established from available knowledge of those quantities



Modus operandi

Revision along following lines:

- Clarity: reconcile scientific rigour and simplicity of presentation – not easy to achieve!
- Structure: as close as possible to that of the present GUM: produce a document that will appear as an evolution rather than a revolution
- Level of presentation: comparable to that of the present GUM
- Better specification of the conditions of applicability
- At same time strive for consistency with VIM as far as possible

Outcome

Revised GUM, while keeping the LPU as a central concept, will follow a Bayesian approach for both Type A and Type B evaluations, giving relevant hyper-links to GUM Supplements

Main differences from present GUM:

- Improved guidance on the evaluation of standard uncertainties associated with estimates of input quantities
- Bayesian approach extended to Type A evaluations of uncertainty – an important consequence: revised GUM will be easier to use
- Improved guidance on determining coverage intervals
- Effective degrees of freedom and Welch-Satterthwaite formula no longer needed
- More examples: applications taken from biology, chemistry, etc. — separate hyper-linked document, allowing future updating

Technical issues

National Physical Laboratory

Statement in GUM clause 2.2.3:

> Some uncertainty components can be evaluated from the statistical distribution of the results of series of measurements and be characterized by experimental standard deviations
>
> The other components, which also can be characterized by standard deviations, are evaluated from assumed probability distributions based on experience or other information

Thus different approaches used for Type A and Type B evaluations

Revised GUM: all uncertainty components evaluated on the same probabilistic basis by using available information to establish state-of-knowledge distributions

Type A evaluation

A standard uncertainty resulting from a Type A evaluation is no longer an estimate of a standard deviation, but a parameter of a state-of-knowledge PDF — as in other cases

$\Rightarrow$ Concept of an uncertainty having (effective) degrees of freedom not needed

Standard deviation of the quantity characterized by this PDF is used as the standard uncertainty — as for all quantities

This change not only simplifies uncertainty calculations involving such quantities, but places the GUM on a much more consistent foundation

Type A evaluation (current GUM)

National Physical Laboratory

Estimate $\bar{x}$ taken as average of a set of n repeated indication values x_i

$n = 6$

Standard uncertainty associated with $\bar{x}$ taken as standard deviation of that average with a corresponding degrees of freedom

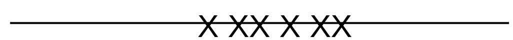
$$s^2(\bar{x}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2, \qquad \nu = n - 1$$

Standard uncertainty is not the standard deviation obtained from a state-of-knowledge distribution

Type A evaluation does not follow the GUM principle that both types of evaluation (Type A and Type B) are based on probability distributions

Type A evaluation (revised GUM)

(Only) available information is set of *n* repeated indication values taken as drawn from Gaussian distribution with unknown expectation and variance



$n = 6$

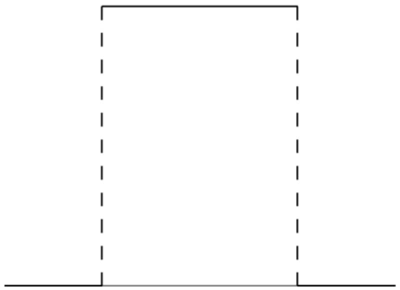
Standard uncertainty $u(\bar{x})$ associated with estimate $\bar{x}$ taken as standard deviation of *t*-distribution, which is the state-of-knowledge distribution for the quantity

$$u^2(\bar{x}) = \frac{n-1}{n-3} s^2(\bar{x})$$

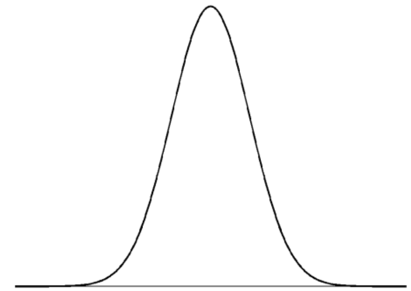
Standard uncertainty is larger than that given in the current GUM — by a factor of as much as $\sqrt{3} = 1.7$ (when $n = 4$)

$\Rightarrow$ Standard uncertainty associated with estimate of measurand also understated (especially when the “Type A contribution” dominates)

Type B evaluation



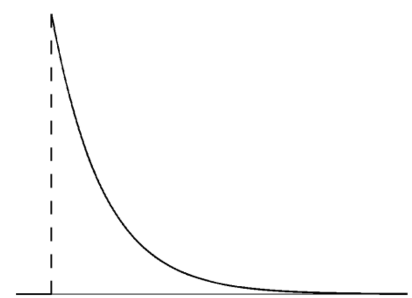
Key:
Available knowledge
Assigned distribution



Lower and upper limits: rectangular distribution

Estimate and associated standard uncertainty: Gaussian distribution

In all cases, standard uncertainty taken as standard deviation of assigned distribution



Estimate and non-negative: exponential distribution

Treat all input quantities similarly

NPL National Physical Laboratory

Guidance in GUM-S1

Uncertainty budgets typically state probability distribution for each quantity

Obtain knowledge of quantity

Apply appropriate rule

Obtain probability distribution for quantity

Compute expectation (mean)

Compute standard deviation

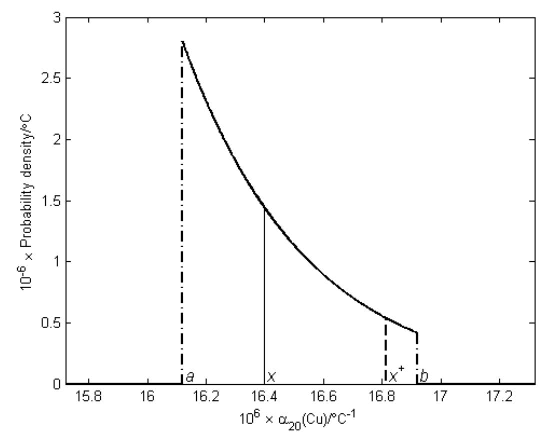
Use mean as estimate of quantity

Use standard deviation as standard uncertainty associated with estimate

Quantity between lower and upper limits and single measured value

Principle of maximum entropy

Doubly truncated exponential function



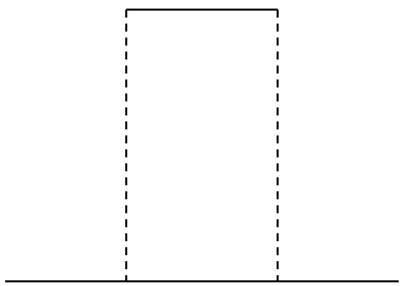
Type B evaluation (example)

Given endpoints *a* and *b* with *a* and *b* imprecisely known, e.g., *a* = 9, *b* = 11 correctly rounded

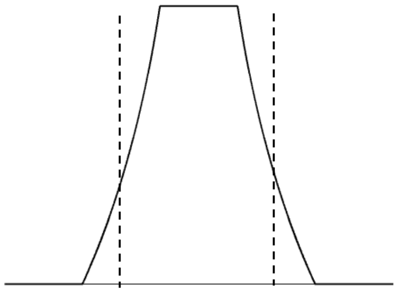
Endpoint quantities *A* and *B*:

$$8.5 \le A \le 9.5, \qquad 10.5 \le B \le 11.5$$

Current GUM: calculate standard uncertainty as standard deviation of rectangular PDF with fixed endpoints *a* = 9, *b* = 11 and assign degrees of freedom to reflect lack of knowledge in endpoints



Revised GUM: assign PDF based on available knowledge and then use PDF to form standard uncertainty (no degrees of freedom)



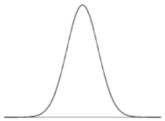
Standard uncertainty in this example again larger than that given in current GUM

Coverage intervals obtained reliably

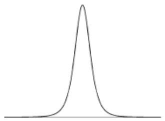
NPL National Physical Laboratory

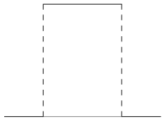
Calculated on the basis of the distribution for the measurand *Y*

When the measurement model can safely be linearized and the central limit theorem applies, *Y* taken as Gaussian — consequence: coverage factor of $k = 2.0$ will provide 95 % coverage



For non-Gaussian dominant input quantity (*t*-distribution, rectangular, …), *Y* particularly influenced by distribution for that quantity





In other circumstances, for a symmetric distribution for *Y*, apply Gauss inequality ($k = 3.0$) or for asymmetric case Chebyshev inequality ($k = 4.5$)

Or, if resulting interval too conservative, apply GUM-S1 to obtain an approximation (in a controlled way) to the distribution for *Y*

Consequences for NMIs and industry

Type A uncertainty evaluation: modified formula for standard uncertainty (already given in GUM-S1) — essentially no change for sufficiently large n

At least $n = 4$ repeated indication values required
$\equiv$ good measurement practice

When unavailable, use historical (or other) information

In Bayesian view adopted in the revised GUM (already given in GUM-S1), concepts such as “uncertainty of uncertainty” unnecessary

Consequences for NMIs and industry

Degrees of freedom in a Type A evaluation of uncertainty no longer viewed as a measure of reliability

Degrees of freedom in a Type B evaluation does not exist

Calculation of coverage interval no longer depends on (effective) degrees of freedom evaluated using the Welch-Satterthwaite formula

Coverage interval for measurand Y (for 95 % coverage, say) calculated reliably in terms of knowledge of the distribution for Y:

Concluding comments

- Standard uncertainties likely to be larger than those obtained by following current GUM — may assist in some key comparison work!

- Aim: first committee draft circulated for review by end 2014
  - To JCGM Member Organizations, NMIs, invited recipients

- Online survey in 2012 to solicit comments and opinions of the GUM
  - Results of the survey analyzed and duly considered
  - GUM criticized by some as difficult, and by others as simple and old-fashioned — seen as the best sign of reasonable balance

- In the same vein, this presentation intended to stimulate discussion so as to acquire as much feedback as possible

Reference: Walter Bich, Maurice G Cox, et al. Revision of the '*Guide to the expression of uncertainty in measurement*'. Metrologia 49 702–705, 2012

Advertisement

Guide to the Expression of Uncertainty in Measurement:
Past, Present and Future

A conference in celebration of the twentieth anniversary of the GUM

7 and 8 November 2013

NPL

Speakers by invitation

preceded by

Workshop on Key Comparison Data Evaluation

6 November 2013

NPL

Contributions welcomed

Further information on NPL web site

http://www.npl.co.uk/events/7-8-nov-2013-gum-past-present-and-future