GUM revision and its impact

Maurice Cox, NPL, Teddington, UK
Walter Bich, INRIM, Torino, Italy

BIPM Workshop on Measurement Uncertainty
15, 16 June 2015

Outline

- Quantities in measurement model represented by probability distributions based on experimenter's knowledge
- Used in current GUM only for Type B evaluations
- Type A evaluations and coverage intervals in revised GUM handled more rigorously
- Separate "examples" document
- Implications
- Concluding remarks

ISO 17025:2005 General requirements for the competence of testing and calibration laboratories

… ensure that the form of reporting of the result does not give a wrong impression of the uncertainty.

ISO 17025:2005 clause 5.4.6.2

Revised GUM

More attention paid to the reliability of uncertainty statements $\Longrightarrow$ Making fewer assumptions

Consistent probabilistic basis

Aim

National Physical Laboratory

Strive to obtain (at least) one correct significant decimal digit in an uncertainty statement.

If that cannot be achieved, that statement is meaningless.

Three major aspects

1. Knowledge of input quantities

Use all available knowledge

2. Propagation through measurement model

Does LPU work adequately?
When should distributions be used?

3. Coverage interval for measurand

Wisdom of always taking “$k = 2$”?

Advances

- Advances in all metrology sectors since 1993
- Measurement uncertainty evaluation has not kept pace
- Aim to ensure measurement uncertainties are reliable
- Take account of modern statistical methods, available computing power and software applications

To that end, GUM-S1 and -S2 [Monte Carlo (MC) propagation of distributions] produced

MC incorporated into Eurachem Guide, UKAS M 3003

Changes: 100 pages → 40 pages

What works well in current GUM?
What does not work so well?

Drivers for change

WHAT'S LEAVING?

- Unwarranted assumptions
- Standard uncertainties calculated inconsistently
- Reliance on the central limit theorem (CLT): 95 % coverage $\Longrightarrow$ "$k = 2$"
- Effective degrees of freedom

WHAT'S ENTERING?

- Knowledge-based PDFs for all quantities
- Option: distribution-free coverage intervals
- Emphasis on GUM-S1 for determining coverage intervals
- Separate "examples" document

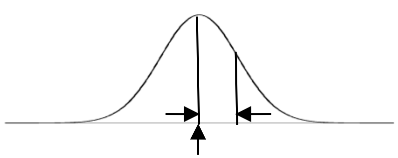
PDF: way to express experimenter's knowledge of a quantity

Knowledge-based PDFs

Knowledge of quantity

Application of principle

PDF for quantity



Expectation and standard deviation

Best estimate and standard uncertainty

Expanded uncertainty/coverage interval

Already intrinsic part of GUM-S1 and GUM-S2

Limits, repeated observations, …

PDFs provided for common scenarios

Treat uncertainty components in the same way

National Physical Laboratory

> … this *Guide* treats uncertainty components arising from random effects and from corrections for systematic effects in exactly the same way in the evaluation of the uncertainty of the result of a measurement.

Current GUM clause E.3

Standard deviation → standard uncertainty

… the [standard] uncertainty of the output quantity *z* … is taken equal to the standard deviation of the probability distribution of z.

Current GUM clause E.3.2

Probability as degree of belief

National Physical Laboratory

> Recommendation INC-1 (1980) upon which this *Guide* rests implicitly adopts such a viewpoint of probability [that based on degree of belief] since it views expressions such as Equation (E.6) [example of LPU] as the appropriate way to calculate the combined standard uncertainty of a result of a measurement.

Current GUM clause E.3.5

Probability as degree of belief

National Physical Laboratory

> … interpreting probability on the basis of degree of belief allows the uncertainty characterizing the effect to be evaluated from an *a priori* probability distribution … and to be included in the calculation of the combined standard uncertainty of the measurement result …

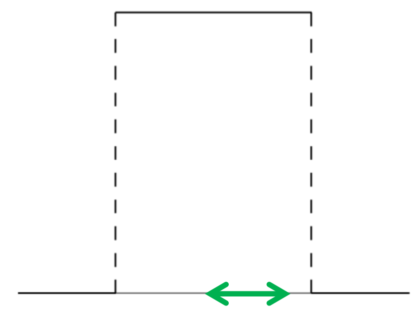
Current GUM clause E.4.4

Standard uncertainty (current GUM)

National Physical Laboratory

Type B evaluation

Standard deviation of knowledge-based probability distribution



Type A evaluation

Standard error of the mean of a number of repeated observations

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

$$u(x) = s/\sqrt{n}, \quad \nu = n - 1$$

degrees of freedom

Standard uncertainty (revised GUM)

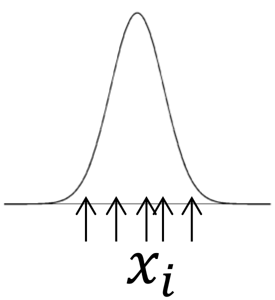
Type B evaluation

Take as standard deviation of knowledge-based probability distribution

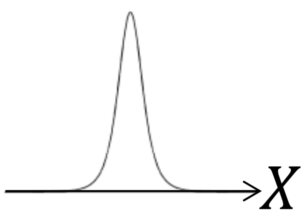
As in current GUM

Type A evaluation

Ditto



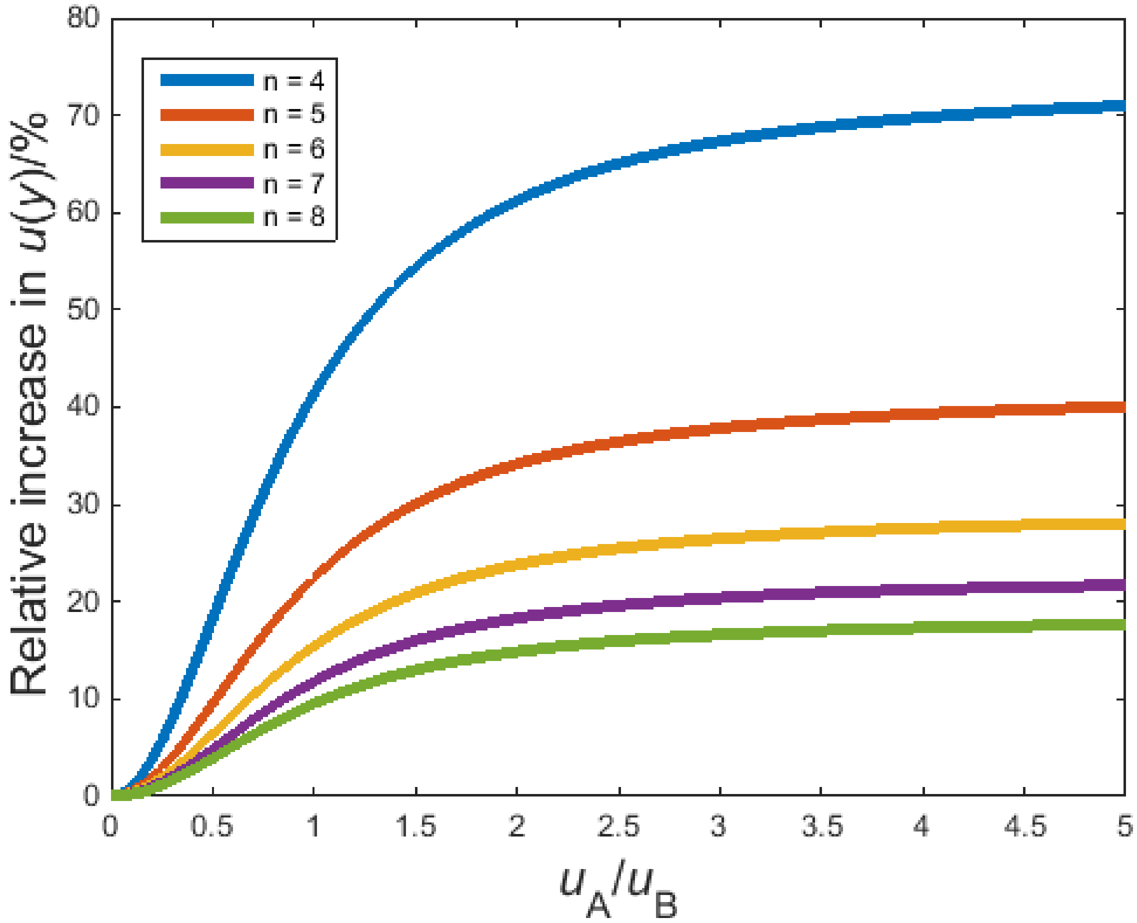
$x_i \sim \mathrm{N}(\mu, \sigma^2) \Longrightarrow$ t-distribution for X:
$t_\nu(\bar{x}, s^2/n)$, $\nu = n - 1$ degrees of freedom



$$u(x) = \left(\frac{n-1}{n-3}\right)^{1/2} \frac{s}{\sqrt{n}}$$

When n < 4, recommendation to use prior information (already in current GUM)

Effect of new Type A evaluation



Model $Y = X_1 + X_2$

u_A and u_B: Type A and Type B standard uncertainty contributions

Figure shows relative increase in $u(y)$ as function of ratio u_A/u_B

Law of propagation of uncertainty

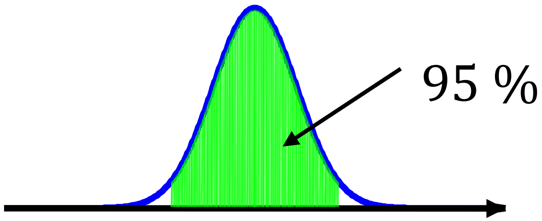
National Physical Laboratory

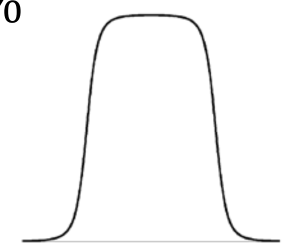
Applied using best estimates and associated standard uncertainties calculated in terms of knowledge-based PDFs for input quantities

- Integral part of revised GUM – to be fully retained
- Generally works very well even for non-linear models
  - There are exceptions!
- Always exact for linear models

Coverage interval

- Current GUM: measurand as normal or t
  - Huge assumption, since CLT often does not apply– especially for small number of input quantities
- Example: model $Y = X_1 + X_2$
  - X_1 rectangular, modelling a systematic effect
  - X_2 Gaussian or t, modelling a random effect
- Y often taken as Gaussian or t
  - Fine when X_2 dominates
  - Not so good when X_1 dominates





Revised GUM encourages use of GUM-S1 in determining coverage interval

Coverage intervals: historical

NPL National Physical Laboratory

> If the relationship between *Y* [the output quantity] and its input quantities is nonlinear, or if the [input quantities] are characterized by probability distributions …, the distribution of *Y* cannot be expressed as a convolution. In this case, numerical methods (such as Monte Carlo calculations) will generally be required and the evaluation is computationally more difficult.

ISO/IEC/OIML/BIPM draft (First Edition), June 1992, produced by ISO/TAG 4/WG 3, Clause G.1.5:

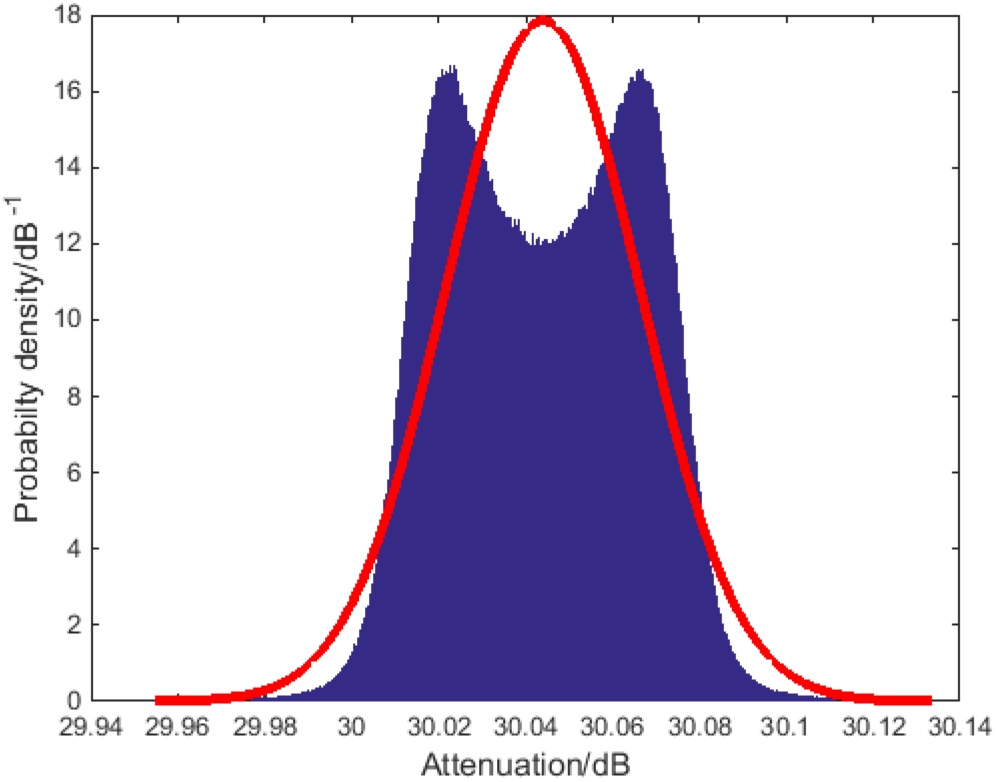
Specimen uncertainty budget
Source: EA/4-02M:2013 example S6

Attenuator calibration: additive model $L_X = L_S + \delta L_S + \cdots + \delta L_{0b}$

quantity X_i	estimate x_i	standard uncertainty $u(x_i)$	probability distribution	sensitivity coefficient c_i	uncertainty contribution $u_i(y)$
L_S	30,040 dB	0,0090 dB	normal	1,0	0,0090 dB
δL_S	0,003 dB	0,0025 dB	rectangular	1,0	0,0025 dB
δL_D	0 dB	0,0011 dB	U-shaped	1,0	0,0011 dB
δL_M	0 dB	0,0200 dB	U-shaped	1,0	0,0200 dB
δL_K	0 dB	0,0017 dB	U-shaped	1,0	0,0017 dB
δL_{ia}	0 dB	0,0003 dB	U-shaped	-1,0	-0,0003 dB
δL_{ib}	0 dB	0,0003 dB	rectangular	1,0	0,0019 dB
δL_{0a}	0 dB	0,0020 dB	rectangular	-1,0	0,0020 dB
δL_{0b}	0 dB	0,0020 dB	normal	1,0	-0,0020 dB
L_X	30,043 dB				0,0224 dB

EA/4-02 defines $u_i(y) = c_i u(x_i)$

Actual (GUM-S1) and (red) Gaussian PDF



Expanded uncertainty

Current GUM: U = 0.044 dB (k = 1.96)

Revised GUM: U = 0.039 dB (k = 1.75)

Revised GUM
(Not so) poor man's coverage intervals

Use conservative (distribution-free) coverage interval

- Gauss inequality (k = 3.0) for a symmetric distribution
- Chebyshev inequality (k = 4.5) for asymmetric distribution

If resulting interval not fit for purpose, apply GUM-S1

Taking "k = 2"
may be unsafe

Procedural implications for NMIs and industry

NPL National Physical Laboratory

- Different Type A evaluation

May require use of historical data

- No calculation of effective degrees of freedom using Welch-Satterthwaite formula

- Need to justify the calculation of a coverage interval

whether based on CLT, distribution-free or GUM-S1, …

Calibration certificates, CMCs, …

Consequences

- Measurement uncertainties evaluated reliably

Traceable to knowledge used

- Reliable decisions based on those uncertainties

Conformance decisions, etc.

- Re-expression of coverage interval for asymmetric measurand (as in GUM-S1)

Endpoints of interval

Transition

- No requirement to update calculated uncertainties retrospectively

Calibration certificates, CMCs, …

- Use of information from previous calibration certificates

Clause in revised GUM

- Results obtained in accordance with JCGM 100:201X

Previously in accordance with JCGM 100:2008

Concluding remarks

- Committee draft circulated for review in December 2014 to JCGM Member Organizations and NMIs
- JCGM Member Organizations and NMIs made collated comments to JCGM – JCGM-WG1 to respond
- Understand impact on NMIs, calibration and testing laboratories and accreditation organizations
- Software proposed to support the revised GUM on the BIPM website

Thank you

National Physical Laboratory