The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States

EMUE
Examples of Measurement Uncertainty Evaluation

Views from the JCGM-WG1 Type A measurement uncertainty workshop

Maurice Cox, National Physical Laboratory, UK

EMUE M18 workshop LNE, January 2020
Draft 2020-01-17

Workshop scope and aims

The calculation of a mean and its associated uncertainty from a small number n of repeated observations is an essential part of the evaluation and expression of uncertainty in measurement

GUM and its supplements use approaches that provide different standard uncertainties, the difference being particularly marked for small n

The aim of the workshop is to work towards a harmonized view by finding answers to the following questions:

1. For what purpose is the standard uncertainty required?
2. What additional knowledge is available and how to account for it?
3. What is a reasonable way to proceed when n is small?

Personal selection of material presented followed by my view

Walter Bich (INRIM, JCGM-WG1 convenor)
Welcome address and introductory remarks

- Informal workshop
- JCGM-WG1 circulated a few years ago draft of proposed revised GUM
- Rationale: GUM inconsistent with its Supplements
- Main inconsistency in Type A evaluation in JCGM 100 (GUM) and JCGM 101 (Monte Carlo)
- Standard uncertainties provided by JCGM 100 and JCGM 101 are different
- Result of attempt to place proposed revised GUM on a coherent (Bayesian) footing
- Proposed revised GUM rejected by many
- No consensus in JCGM-WG1 as to how to take matter forward
- Views by workshop participants welcomed!

Carlo Carobbi (U Florence) Type A evaluation of standard uncertainty in EMC: IEC/TR 610001-6 and ANSI C63.23

- JCGM 100 (GUM): $\dfrac{s}{\sqrt{n}}$
- JCGM 101 (MC): $\left(\dfrac{n-1}{n-3}\right)^{1/2} \times \dfrac{s}{\sqrt{n}} \quad (n > 3)$ Normal prior
- Factor $[(n-1)/(n-3)]^{1/2} \approx$ ratio of percentile for 95 % coverage probability of t distribution with $n-1$ degrees of freedom to corresponding percentile of standard normal distribution

Carlo Carobbi (continued)

- Kacker-Jones: ad hoc extension to $n = 2$ and 3

n	2	3	4	5	6	7	8	9
K-J	6.48	2.20	1.62	1.42	1.31	1.25	1.21	1.18
Factor	—	—	1.73	1.41	1.29	1.22	1.18	1.15

- Carobbi introduced factor to standards EMC: IEC/TR 610001-6 and ANSI C63.23:
- Little opposition

Carlo Carobbi (continued)

- 'Welch-Satterthwaite (WS) formula and effective degrees of freedom (DoF) never used by calibration and testing laboratories accredited to ISO/IEC 17025'
- Anomaly associated with WS (Balllico, 2000)
- Some dissension on these points:
  - UKAS-accredited laboratories use WS and DoF
  - Ballico reasoning flawed: Hall & Willink address issue correctly
  - Many laboratories have problems with calculating effective degrees of freedom
  - In analytical chemistry, random variation often dominates the uncertainty budget
    $\Longrightarrow$ limited DoF an issue
  - Also in temperature measurement

Carlo Carobbi (continued)

- Emphasized greater use of prior information:
  - Earlier observations with same measuring system
  - General experience of performance of similar measuring systems
  - Suppliers' specifications
  - Other expert knowledge
- Together with current repeated observations, Bayesian estimate of standard uncertainty

Carlo Carobbi (continued)

- Combine
  - s as in GUM
  - previous estimate σ_0
  - value σ_M unlikely to be exceeded
- Gives standard uncertainty formula: spreadsheet implementation
- Mentioned related work: Cox & Shirono (Metrologia, 2018)
- [Also relevant: van der Veen (Metrologia, 2018)]

Antonio Possolo (NIST) What is your problem? . . .
We offer many solutions

- GUM should promote full information on how uncertainty should be evaluated
- Guidance should be understandable and practicable and conform with common sense
- Assuming normality is often unrealistic; biases may exist
- Guidance should be provided that does not require mathematicians and statisticians
- Assumptions should be clearly stated

Antonio Possolo (continued)

- Prior information (informative priors) should be elicited and used
- Non-informative priors should be avoided
- Use Bayes when appropriate — although never a panacea and costly
- Employ realistic, adequate models
- No general solution to forming mean and standard deviation
- Take half length of 68 % coverage interval as standard uncertainty

Discussion session

- **Rules.** 'Rules' for testing and calibration laboratories?

  The attitude 'horses for courses' was liked but a default solution was suggested for 'normal' laboratories

  Idea attractive but risk in providing default solution: many practitioners would use default without considering its appropriateness
- **Subjectivity.** Scientists make many judgments that are subjective; science strives to make good judgments, starting with selecting the equipment for making the measurement
- **Guidance.** JCGM-WG1 produces high-level guidance, which should be sound

  GUM New Perspective has broader scope including testing laboratories, clinical laboratories.
- **Elicitation.** JCGM-WG1 believes in elicitation and tries to convince metrologists about its usefulness, but so far has been unsuccessful

Tony O'Hagan (Sheffield University) Up a GUM tree: a solid foundation for the GUM and Type A uncertainty

- GUM: one of the most-read statistical texts but has shaky foundations
- Proposed revision an effort to give it a Bayesian footing
- The GUM is about using a measurement model and estimating the measurand given what we know about the input quantities
- Fundamental question: how to express uncertainty, both about the input and output quantities
- Only proper way is through a probability distribution

Tony O'Hagan (continued)

- GUM uses different estimates:
  - Sample mean (unbiased in frequentist terms)
  - Statistical estimate (Type A)
  - Expert estimate (Type B) (GUM)
  - Probability distribution (GUM-S1)
- Mean and mode unhelpful (in the case of skewed distributions)
- Median most useful choice: always exists and has intuitive meaning
- Invariant in transformations

Tony O'Hagan (continued)

- Standard uncertainty: (estimate of) sampling standard deviation (GUM) or standard deviation of the probability density function (GUM-S1)
- Both unhelpful and may not exist
- Define $\boxed{\text{cover uncertainty}}$ $c(X)$ to be such that $m(X) \pm 2c(X)$ has 95 % probability that X lies in the interval
- Take half length of 95 % coverage interval as standard uncertainty
- $c(X)$ always exists.

Tony O'Hagan (continued)

- On certificates, probability distribution should be stated, plus metrics such as median $m(X)$, cover uncertainty $c(X)$ to aid interpretation
- Measurement result is to be defined as a probability distribution
- Both GUM and GUM-S1 give same coverage interval

Tony O'Hagan (continued)

- GUM (JCGM 100): $u_0(X) = s/\sqrt{n}$
- GUM Supplement 1 (JCGM 101): $[(n-1)/(n-3)]^{1/2} u_0(X)$
- Coverage interval: $x \pm k(n-1)/2 \times u_0(X)$ same for both
- Problem was that the revision redefined 'standard uncertainty'

Tony O'Hagan (continued)

- New definition has $c(X) = u_0(X) \times k(n-1)/2$
- Immediate intuitive meaning
- Interestingly, rather close to $u(X)$
- Controversial factor $[(n-1)/(n-3)]^{1/2}$ not so dumb after all!

n	2	3	4	5	6	7	8	9
K-J	6.48	2.20	1.62	1.42	1.31	1.25	1.21	1.18
$k(n-1)/2$	—	—	1.73	1.41	1.29	1.22	1.18	1.15
t (ToH)	6.35	2.15	1.59	1.39	1.29	1.22	1.18	1.15

Tony O'Hagan (continued)

- When propagating MU, can always use Monte Carlo (MC) method
- GUM provides another mechanism: LPU
- Can we apply LPU using $m(X)$ and $c(X)$?
- Model $Y = X_1 + X_2$, t distributions with $\nu = 4$ and $\nu = 8$
- Coverage uncertainty very close to that from LPU
- Intervals using WS systematically shorter than those using MC
- Approximation good if assessment done properly

Tony O'Hagan (continued)

- Problems with uninformative priors
- Genuine prior information always exists
- Unscientific to ignore it: a form of cheating
- To make it workable, propose informative prior on σ^2 based on belief that

  $$\sigma_0^2/3 \le \sigma^2 \le 3\sigma_0^2$$

  ($\sigma_0 =$ prior estimate)
- Informative prior $\implies$ shorter coverage intervals

Tony O'Hagan (continued)

- Bayesian method achieves on average coverage probability of 95 %
- Stronger information (priors) require more validation and justification
- Approach readily implementable by laboratories
- Acceptance first requires consensus about this way forward
- Frequentist methods cannot combine results from Type A and Type B evaluations, and cannot deal with Type B at all

Personal view in conclusion

- Need to respect views of current GUM users
- Also, promote benefits of working with probability distributions – the world moves on – or at least some of it
- Fits with objectives of GUM New Perspective – wide scope
- Re-emphasize the placing of Type A and Type B on a similar footing
  - First, consider expert knowledge, historical knowledge, etc. for both
  - Then, for Type A, also take account of available repeated observations — workshop presentations stressed this point

Personal view (continued)

- Strengthening of
  VIM 2.9 measurement result
  set of quantity values being attributed to a measurand together with any other available relevant information
  NOTE 1 . . . may be expressed in the form of a probability density function. . .

- Software culture: traditional GUM and Bayesian Type A evaluation
  — simple (spreadsheet) calculations preferred