Uncertainty Evaluation by means of a Monte Carlo Approach

Walter Bich, INRIM

BIPM Workshop 2 on CCRI (II) Activity
Uncertainties and Comparisons

Sèvres, 17-18 September 2008
• 1977-79 BIPM questionnaire on uncertainties
• 1980 Recommendation INC-1
• 1981 Establishment of WG3 on uncertainties under ISO
  TAG4: BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML
• 1981 Recommendation CI-1981
• 1986 Recommendation CI-1986
• 1993 Guide to the expression of uncertainty in measurement
• 1995 Reprint with minor corrections
• 1997 Establishment of the Joint Committee for Guides in Metrology JCGM – ILAC joins in 1998
Joint Committee for Guides in Metrology

Present Chair: the BIPM’s Director

The JCGM has two working groups (WGs)

• WG 1 “Expression of uncertainty in measurement”, has the task “to promote the use of the GUM and to prepare Supplements for its broad application”

• WG 2 “on International vocabulary of basic and general terms in metrology”, has the task “to revise and promote the use of the VIM”

See also www.bipm.org
Evaluation of measurement data

— An introduction to the “Guide to the expression of uncertainty in measurement” and related documents;

— Concepts and basic principles;

— Supplement 1 to the “Guide to the expression of uncertainty in measurement” — Propagation of distributions using a Monte Carlo method;

— Supplement 2 to the “Guide to the expression of uncertainty in measurement” — Models with any number of output quantities;
Evaluation of measurement data

— Supplement 3 to the “Guide to the expression of uncertainty in measurement” —Modelling;

— The role of measurement uncertainty in deciding conformance to specified requirements;

— Applications of the least-squares method.
Supplement 1 to the GUM
Propagation of distributions using a Monte Carlo method

AIM: to overcome some of the limitations of the GUM, especially when an interval of confidence with a stipulated coverage probability is needed.
Formulation stage
(common to both GUM and Suppl. 1)

You decide a model

\[ Y = f(X_1, X_2, \ldots, X_N) \]

or

\[ Y = f(X) \]
The same capital symbol is used with two meanings:
The (physical, chemical…) quantity entering in the “law”
The associated random variable used to describe the experimental outcome
The GUM Method

One seeks for the following items of information

- An \textit{estimate} $x_i$ for each input quantity $X_i$

- Its \textit{standard uncertainty} $u(x_i)$

(Guidance given in the GUM)
CAUTION

A lowercase symbol denotes a quantity estimate, that is, a realizations of the corresponding random variable.
You “propagate” $x_i$ by using

$$y = f(x_1, x_2, \ldots, x_N)$$
You propagate $u(x_i)$ by using a first-order Taylor expansion about $E(X)$

$$Y \approx f[E(X)] + \sum_{i=1}^{N} \left. \frac{\partial Y}{\partial X_i} \right|_{X=E(X)} [X_i - E(X_i)]$$
From which, by re-arranging terms, squaring and taking expectations, you have

\[
V(Y) \approx \sum_{i,j=1}^{N} \frac{\partial Y}{\partial X_i} \frac{\partial Y}{\partial X_j} \bigg|_{X_i, X_j = E(X_i, X_j)} \text{Cov}(X_i, X_j)
\]

or, in matrix terms
\[ V(Y) \approx J_X V_X J_X^T \]

where

\[ J_X = \left( \frac{\partial Y}{\partial X_1}, \ldots, \frac{\partial Y}{\partial X_N} \right) \]

and

\[ \{V_X\}_{i,j} = Cov(X_i, X_j) \]
This is a first-order approximation, good under the weak condition that the non-linearity of $f$ be negligible compared to the magnitude of the uncertainties.

$$E[f(X)] \approx f[E(X)]$$

It is distribution-free: no assumptions needed on the probability density functions (PDFs) of the random variables.
Comments II

In real world:

• Derivatives can be difficult (complicated models), and, more important

• Expectations, variances and covariances are unknown (to be further discussed), and only estimates are available for them
Further compromise...

You take:

- Estimates in place of expectations and

- Experimental standard deviations (Type A) or values coming from “different” knowledge (Type B) in place of variances (better, their square roots…)
...and further constraint

These approximations place a further constraint

The PDFs have to be reasonably symmetric; in the negative, you have to "symmetrize" (cosine error)
Comments III: The GUM is frequentist

• In terms of PDFs:

• Type A: you estimate parameters of a supposed (Gaussian?) frequency distribution characterizing the “population” of “indications” (new VIM III term). Therefore you have \( s \) instead of \( \sigma \), with \( \nu = n-1 \) degrees of freedom – Classical, frequentist view of probability
Comments IV:
The GUM is Bayesian

• In terms of PDFs:

• Type B: you assign a PDF describing your knowledge about a quantity value not obtained from a sample of indications - Subjective view of probability
No trouble…

…as far as the quantity of interest is standard uncertainty $u(y)$

Unfortunately….

…standard uncertainty $u(y)$ is of limited interest - mainly for further propagation with $y$ as input quantity to a new measurement model (MRA supersedes CI-1986)
Expanded uncertainty

In many cases (typically end users) one needs $U_p(y)$, where $p$ is a prescribed coverage probability.

In these cases, more knowledge is necessary than simply first ($x_i$) and second ($u^2(x_i)$) moments of $X_i$, namely

The pdf of $Y$
Troubles arise…

The Central Limit Theorem helps, under rather strong conditions (GUM, G.2).

If the conditions are met (GUM optimistic, JCGM-WG1 much less), then
Troubles grow up...

...the random variable

\[ \frac{(y - Y)}{u(y)} \]

behaves like a Student’s t distribution with degrees of freedom, where \( \nu_{\text{eff}} \) is determined with the Welch-Satterthwaite formula (GUM, G.4).
The Welch-Satterthwaite formula

\[ \nu_{\text{eff}} = \frac{u^4(y)}{\left( \frac{\partial Y}{\partial X_i} u_{x_i} \right)^4} \sum_{i=1}^{n} \frac{\nu_i}{\nu_i} \]
Each input uncertainty in the Welch-Satterthwaite formula has a degrees of freedom. Dof is a natural measure for Type A evaluations, artificial for Type B.

Little, and little persuasive, guidance in the GUM on this issue.
Consequences

Subjective, typically very large, dof are attached to Type B evaluations.

In many experiments Type B components tend to dominate, therefore effective dof is large (a seemingly, but only seemingly good thing)
Confidence intervals and intervals of confidence (!)

When determining an expanded uncertainty, the GUM extends the frequentist attitude also to Type B evaluations.

“Confidence interval” is a specific statistical term (Type A only), therefore let us invent “Interval of confidence” – Try to render the distinction in your language!
Problems with effective dof

Attaching a degrees of freedom to standard uncertainty conveys the concept of

Uncertainty of the uncertainty

Do you like this concept? (no shame if you do)

We do not…
Further limitations

If the conditions of applicability of the Central Limit Theorem are not met:

• The PDF for \( y \) is no longer a scaled-and-shifted \( t \)-distribution

• The Welch-Satterthwaite formula does not apply

• other ways (e.g., analytical) must be explored.

(Little guidance is given in the GUM, only for a particular case - dominant uniform)
The approach of Supplement 1

Instead of propagating only first and second moments of the input quantities $X_i$, their pdfs are propagated through the model.

The method is more demanding in terms of amount of knowledge. One has to assign a pdf to each input quantity, based on the experimental data or other knowledge.
Assignment of input pdfs

According to the available information, it is based on

• Bayes’ theorem, typically when a series of indications is available

• The principle of maximum entropy – you maximize a functional $S$, the “information entropy”, under constraints given by the information
Assignment of input pdfs

• Luckily, extensive literature and guidance in Supplement 1 do most of the job

A dozen pdfs for the various cases are suggested Table 1

FOCUS: if a sample of indications is available, you assign a Student’s \( t \)-distribution. Departure from the GUM!
Evaluation of measurement data — Supplement 1 to the “Guide to the expression of uncertainty in measurement” — Propagation of distributions using a Monte Carlo method

Évaluation des données de mesure – Supplément 1 du “Guide pour l'expression de l'incertitude de mesure” — Propagation de distributions par une méthode de Monte Carlo
Table 1 — Available information and the PDF assigned on the basis of that information (6.4.1, C.1.2)

<table>
<thead>
<tr>
<th>Available information</th>
<th>Assigned PDF and illustration (not to scale)</th>
<th>Subclause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower and upper limits (a, b)</td>
<td>Rectangular: (R(a, b))</td>
<td>6.4.2</td>
</tr>
<tr>
<td>Inexact lower and upper limits (a \pm d), (b \pm d)</td>
<td>Curvilinear trapezoid: (CTrap(a, b, d))</td>
<td>6.4.3</td>
</tr>
<tr>
<td>Sum of two quantities assigned rectangular distributions with lower and upper limits (a_1, b_1) and (a_2, b_2)</td>
<td>Trapezoidal: (\text{Trap}(a, b, \beta)) with (a = a_1 + a_2), (b = b_1 + b_2), (\beta =</td>
<td>(b_1 - a_1) - (b_2 - a_2)</td>
</tr>
<tr>
<td>Sum of two quantities assigned rectangular distributions with lower and upper limits (a_1, b_1) and (a_2, b_2) and the same semi-width ((b_1 - a_1 = b_2 - a_2))</td>
<td>Triangular: (T(a, b)) with (a = a_1 + a_2), (b = b_1 + b_2)</td>
<td>6.4.5</td>
</tr>
<tr>
<td>Sinusoidal cycling between lower and upper limits (a, b)</td>
<td>Arc sine (U-shaped): (U(a, b))</td>
<td>6.4.6</td>
</tr>
<tr>
<td>Best estimate (x) and associated standard uncertainty (u(x))</td>
<td>Gaussian: (N(x, u^2(x)))</td>
<td>6.4.7</td>
</tr>
<tr>
<td>Best estimate (x) of vector quantity and associated uncertainty matrix (U_x)</td>
<td>Multivariate Gaussian: (N(x, U_x))</td>
<td>6.4.8</td>
</tr>
<tr>
<td>Series of indications (x_1, \ldots, x_n) sampled independently from a quantity having a Gaussian distribution, with unknown expectation and unknown variance</td>
<td>Scaled and shifted (t): (t_{n-1}(\bar{x}, s^2/n)) with (\bar{x} = \sum_{i=1}^{n} x_i/n), (s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1))</td>
<td>6.4.9.2</td>
</tr>
<tr>
<td>Best estimate (x), expanded uncertainty (U_p), coverage factor (k_p) and effective degrees of freedom (\nu_{eff})</td>
<td>Scaled and shifted (t): (t_{\nu_{eff}}(x, (U_p/k_p)^2))</td>
<td>6.4.9.7</td>
</tr>
<tr>
<td>Best estimate (x) of non-negative quantity</td>
<td>Exponential: (\text{Ex}(1/x))</td>
<td>6.4.10</td>
</tr>
<tr>
<td>Number (q) of objects counted</td>
<td>Gamma: (G(q + 1, 1))</td>
<td>6.4.11</td>
</tr>
</tbody>
</table>
Problem

Given the joint pdf of the \( N \) input quantities \( \mathbf{X} \), find the pdf of the output quantity \( y \).

• Using the Jacobian method (any textbook on mathematical statistics)

\[
g_Y(\eta) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} g_X(\xi) \delta(\eta - f(\xi)) \, d\xi_N \ldots d\xi_1
\]

\( \delta(\cdot) \) is the Dirac function
Solutions

• Analytical: A closed-form solution can be found only in the most simple (and therefore uninteresting) cases. In general, the integrals involved in this solution must be solved numerically. Not viable.

• Numerical 1: by numerical integration of the formal expression
Supplement 1 approach: MCM

Numerical 2 (adopted in Supplement 1):

- Numerical simulation.
- Method selected: Monte Carlo (MCM).
- Tools: suitable random number generators for the various pdfs, reasonable computing power.
- Outcome: a numerical approximation for the output distribution $G_Y(\eta)$ (in various possible forms).
Output of MCM

From the numerical approximation for the output distribution, the required statistics, such as the best estimate for the measurand, its standard uncertainty, and the endpoints of a prescribed coverage interval can be obtained.
The method in a nutshell

• From each input pdf draw at random a value $x_i$ for the random variable $X_i$.

• Use the resulting vector $\mathbf{x}_r \ (r = 1, \ldots, M)$ to evaluate the model, thus obtaining a corresponding value $y_r$. The latter is a possible value for the measurand $Y$.

• Iterate $M$ times the preceding two steps, to obtain $M$ values $y_r$ for $Y$. 
Representations of the probability distribution for $y$

- **Discrete ($G$):**
  - Sort the $M$ values $y_r$ for $Y$ in non-decreasing order.
  - Take $G$ as the set $y_{(r)}$, $r = 1, \ldots, M$

- Sufficient for most applications
Representations of the probability distribution for $y$

- Continuous: $\tilde{G}_Y(\eta)$

In the form of a piecewise-linear function, suitably obtained from $G$ (details in the Supplement).

Useful, e.g. for further samplings
Representations of the pdf for $y$

- Assemble the $(M) y_r$ values into a histogram with suitable cell widths (subjective!) and normalize to one.

  Useful for visual inspection of the pdf, or

When the sorting time ($M$ large and simple model) is excessive
Coverage interval(s)

The novelty with respect to the GUM is that, since the pdf is usually asymmetric, there is more than one coverage interval (for a given coverage probability $p$)

• Shortest (contains the mode)

• Probabilistically symmetric $\alpha = (1 - p)/2$

Endpoints $\alpha$ and $(p + \alpha)$ quantiles
A comparison of the two approaches

• The Monte Carlo approach works in a broader class of problems than the GUM approach. In this sense, it is more general, therefore

• It can be used to validate the results provided by the GUM uncertainty framework, however

• It is based on the same principles underlying the GUM

• It descends naturally from the GUM

• It is to be used in conjunction with the GUM
An example
(from Supplement 1)

- Red dotted: GUM
- Solid blue: S1
A comparison of the two approaches (continued)

- The main output is a coverage interval, not the standard uncertainty.
- The best estimate (as the expectation of the numerical approximation for the output distribution) does not necessarily coincide with that provided by the GUM.
- Also the standard uncertainties do not coincide. The GUM value may be smaller. This is a consequence of the pdf recommended for sampled data (Student).
A comparison of the two approaches (continued)

• Type A and B do not apply to pdfs (luckily…)

• There is no longer need for degrees of freedom. No longer uncertainty of the uncertainty!

• Actually, the approach of Supplement 1 is intrinsically Bayesian.
Internal (in)consistency of the GUM

Acceptable, as far as the issue is standard uncertainty

Poor, if an interval of confidence is required

Reason(s):

• presence of two different views of probability, and

• frequentist choice concerning expanded uncertainty
(In)consistency between the GUM and Supplement 1 (its first creature…) 

Resulting measurand values and associated standard uncertainties are different, but also:

- standard uncertainty:
  - is uncertain with the GUM,
  - has no uncertainty with Supplement1
Remedy

Revise the GUM, by

adopting a Bayesian evaluation of standard uncertainty also for sampled data (Type A), for example

\[ u(x_i)_{\text{GUM2}} = \sqrt{\frac{n-1}{n-3}} u(x_i)_{\text{GUM1}} \]
To go back to reality

If your experiment needs statistics, you ought to perform a better experiment

(Lord Rutherford)
Or, if you prefer:

the only statistics you can trust are those you falsified yourself

(Churchill)