A probabilistic approach to measurement and uncertainty

A quick guide to GUM

Andy Buffler
Department of Physics, University of Cape Town, South Africa

... with acknowledgments to Saalih Allie (UCT), Fred Lubben and Bob Campbell (U of York, UK)
Measurement ... ?
... you’re kidding ...

What’s there to worry about?

What’s wrong with the way we always do (did?) things?
From some recent CSIRO publications:

“... the standard deviation of the residuals is 0.2 wt % which represents the indicative error for the gauge ...”

“... the resulting standard deviation of 0.5 % ...”

“... the concentration could be determined with an accuracy of 0.3 % ...”

“... the rms errors between gauge measurements and chemical laboratory analyses was 2.3% ... “

... Huh ... ?
Some problems with the traditional approach to measurement and uncertainty:

1. “Traditional” data analysis based upon the “frequentist” approach ... fantasizing about how things might be if we had enough data. “What would happen if I did this (experiment) a very large number of times?” Measurement results are typically statements about the data (which are “in error”).

2. The theoretical framework is only really valid for large data sets anyway.

3. “Systematic” and “random” measurement errors describe deviations from the so-called “true value” ... which is not known ... ?
... more problems with the traditional approach to measurement and uncertainty:

4. "Systematic" and "random" errors are estimated through completely different approaches and cannot be combined.

5. There is no formal logical link between the data and the measurand.

6. The result is that most practicing scientists use an amalgam of rigorous computational methods and arcane rules of thumb (and software packages) when dealing with "errors" ...
... clearly a need to standardize metrology and methods of reporting measurement and uncertainty ...
The need for a consistent international language for evaluating and communicating measurement results prompted (in 1993) the ISO (International Organization for Standardization) to publish recommendations for reporting measurements and uncertainties based on the probabilistic interpretation of measurement.

All standards bodies have adopted these recommendations for reporting scientific measurements:

- BIPM: International Bureau of Weights and Measures
- IUPAC: International Union of Pure and Applied Chemistry
- OIML: International Organization of Legal Metrology
- NIST: (United States) National Institute of Standards and Technology
- NML: (Australian) National Measurement Laboratory

... and most journals.
A number of documents currently serve as international standards, including:

**VIM** (International Vocabulary of Basic and General Terms in Metrology)

**GUM** (Guide to the Expression of Uncertainty in Measurement)

... published by the ISO in 1993 and 1995.
What is the purpose of a measurement?

... to update / improve our state of knowledge about a measurand.

We assume that the quantity to be measured exists (like a physical object) before the measurement ... it is regarded as the cause of specific effects in the form of data from measuring apparatus.

The value of the measurand, however, does not pre-exist ... it is established by the measurement.
The assignment of the measurement value to the object measured is achieved through a comparison with a quantity whose value is known... the “reference standard”

Two questions then arise:

1. How well do we know the reference quantity with regard to some generally accepted system of units?
2. How well can we estimate the measurand using the reference quantity?

Leads to the many sources of measurement uncertainty
A key element of the ISO guidelines is how it views the measurement process.

The GUM states that ...

“In general, the result of a measurement is only an approximation or estimate of the value of the specific quantity subject to measurement, that is, the **measurand**, and thus the result is complete only when accompanied by a quantitative statement of its uncertainty.”

**Uncertainty** itself is defined as ...

"a parameter associated with a measurement result, that characterizes the dispersion of the values that could reasonably be attributed to the measurand".
The **GUM** approach

... has roots within the Bayesian approach to the inferences that can be drawn from incomplete knowledge.

... the short story ...

Starting from the principle of maximum entropy (!) the (more or less) complete knowledge of a quantity \( X \) can be described by a **probability distribution** of the values compatible with the knowledge.
The best estimate of the value $x$ is associated with the expectation value of the distribution,

$$x = E[X]$$

and the measurement uncertainty $u(x)$ associated with its standard deviation, is given by the positive square root of the variance:

$$u(x) = \sqrt{\text{Var}[X]}$$

$u(x)$ is related to the “mean width” of the distribution.
The “distributions” referred to here are, in fact, **probability density functions** (pdf’s) ... which describe how our knowledge about a particular measurand is distributed.

For example:

\[ \int p(M) \, dM = 1 \]

If the shaded area = 0.2 say, then \( p(30 < M < 40) = 0.2 \)

then we can make statements of the following type:

“The probability that the value of \( M \) lies between 30 kg and 40 kg is 0.2.”
There is no limitation on the shape of pdf to use ...

In each case, the best estimate of the value is given by the expectation value (most often the centre of the pdf) and the standard uncertainty is given by the square root of the variance, which is related to the second moment of the pdf.
Three useful pdf’s in measurement:

(a) the uniform (or rectangular) pdf

\[ p(x) = \frac{1}{a} \]

\[ u(x) = \frac{a}{2\sqrt{3}} \]

(b) the triangular pdf

\[ p(x) = \frac{2}{a} \]

\[ u(x) = \frac{a}{2\sqrt{6}} \]
(c) the Gaussian pdf

\[ p(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left[ -\frac{(x - \mu)^2}{2\sigma^2} \right] \]

From \( N \) data we estimate \( \mu \) by \( \bar{x} \) (the mean) and \( \sigma \) by \( s(\bar{x}) \) (the experimental standard deviation of the mean).

\[ u(x) = s(\bar{x}) = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^{N} (x_i - \bar{x})^2} \]
Evaluation of uncertainties

Two approaches: **Type A** and **Type B**

**Type A evaluation of uncertainty** ... uses statistical procedures to obtain the best estimate and standard uncertainty. It will be used if one or several input quantities are observed in the measurement several times under unchanged conditions and different values are determined.

**Type B evaluation of uncertainty** ... when a statistical evaluation cannot be made but profound metrological knowledge is available (data from previous measurements, knowledge of the behavior of the instrument in question, manufacturer’s specifications, statements in calibrating certificates, etc.)
A well-founded, realistic estimate of the measuring situation (Type B evaluations) can furnish values for the standard uncertainty which are not inferior to those obtained by a Type A evaluation

... especially when the Type A evaluation can be carried out only for a very small number of observations, or in principle cannot be used.
Think about the reading on a digital meter, for example.

Due to the finite scale interval displayed, we do not have perfect knowledge of the measurand, even in the absence of fluctuations.
Random and systematic “errors” do not correspond to Type A and Type B evaluations!

If measurand $X = \bar{X} \pm x_{\text{correction}}$ where $\bar{X}$ is obtained from many observations and $x_{\text{correction}}$ is some correction term then the uncertainty $(u_1)$ associated with $\bar{X}$ will be estimated from a Type A evaluation while the uncertainty associated with the correction $(u_2)$ will come from a Type B evaluation.

The combined uncertainty for $X$ is then calculated from

$$u(X)^2 = (u_1)^2 + (u_2)^2$$
A probabilistic model for measurement

all prior information + new data

Observations from apparatus

inference engine

Probability density functions

Inferences about the measurand

best estimate

standard uncertainty

coverage probability

A practical example...

Say we are measuring a voltage across a component using an analogue voltmeter.

Our master equation might be of the form:

\[ V = V_{\text{component}} - V_{\text{zero}} + V_{\text{rating}} + V_{\text{contact}} \]

where:

- \( V_{\text{component}} \): reading with the component connected
- \( V_{\text{zero}} \): zero reading (no component)
- \( V_{\text{rating}} \): correction due to the rating of the instrument
- \( V_{\text{contact}} \): correction due to contact resistance
We think about each in turn.

Say that we see the following when the component is in:

We might decide that the best estimate of $V_{\text{component}}$ is 1.55 V. What about the standard uncertainty in reading the scale?

$$u(V_{\text{component}}) = \frac{0.08}{2\sqrt{6}} = 0.016 \text{ V}$$
Even if we observe 0.00 V when the component is out, i.e. for $V_{\text{zero}}$, we still need to estimate $u(V_{\text{zero}})$, which might also be 0.016 V.

The manufacturers might say that the voltmeter’s rating is a standard uncertainty of 1% of the value displayed.

In this case $V_{\text{rating}}$ is zero

and $u(V_{\text{rating}})$ is taken as $(0.01)(1.55 \text{ V}) = 0.015 \text{ V}$.
Finally, from experience, you might know to increase the value shown by 2% due to the contact resistance on the probes.

In this case $V_{\text{contact}}$ is $(0.02)(1.55 \, \text{V}) = 0.031 \, \text{V}$.

and $u(V_{\text{contact}})$ is estimated using a rectangular pdf:

$$u(V_{\text{contact}}) = \frac{0.04}{2\sqrt{3}} = 0.012 \, \text{V}$$
The uncertainty contributions are summarized in an uncertainty budget.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
<th>Standard uncertainty</th>
<th>Type of evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{\text{component}}$</td>
<td>1.55 V</td>
<td>0.016 V</td>
<td>B</td>
</tr>
<tr>
<td>$V_{\text{zero}}$</td>
<td>0</td>
<td>0.016 V</td>
<td>B</td>
</tr>
<tr>
<td>$V_{\text{rating}}$</td>
<td>0</td>
<td>0.015 V</td>
<td>B</td>
</tr>
<tr>
<td>$V_{\text{contact}}$</td>
<td>0.031 V</td>
<td>0.012 V</td>
<td>B</td>
</tr>
<tr>
<td>$V$</td>
<td>1.581 V</td>
<td>0.030 V</td>
<td></td>
</tr>
</tbody>
</table>
The final result may then be expressed in the following way:

The best approximation of the voltage is 1.581 V with a combined standard uncertainty of 0.030 V, using a Gaussian pdf.

(This implies that there is a 68% probability of finding the value of the voltage between 1.551 V and 1.611 V.)

Here, 68% is called the coverage probability or level of confidence.

The statements above summarize the final pdf which models all your present knowledge of the voltage.
Situations will arise within which it is appropriate to increase the level of confidence by introducing an expanded uncertainty $U$ where

$$U = k \, u_c$$

expanded uncertainty

coverage factor

combined uncertainty

For example, if a Gaussian pdf is used, then

a coverage factor of 2

... increases the coverage probability to 0.95

and

a coverage factor of 3

... increases the coverage probability to more than 0.99
A more complete statement of the final result may be:

The voltage \( V = (1.581 \pm 0.060) \) V, where the number following the symbol \( \pm \) is the numerical value of an expanded uncertainty \( U = k u_c \), with \( U \) determined from a combined standard uncertainty \( u_c = 0.030 \) V and a coverage factor \( k = 2 \).

Since it can be assumed that the possible estimated values of the voltage are approximately normally distributed with approximate standard deviation \( u_c \), the unknown value of the voltage is believed to lie in the interval defined by \( U \) with a level of confidence of approximately 95 percent.
There are a number of useful software tools available.

The best I know of is the “Gum Workbench” … from Metrodata.

There are also a growing number of web resources, such as … at NIST … http://physics.nist.gov
General procedure

Model equation for measurement

Identify all of sources uncertainty

Evaluate each contribution
(Type A or Type B evaluation)

Several standard uncertainties.

Combined standard uncertainty $u_c$
Conclusion

The probabilistic formalism for metrology offers a logical and consistent framework for data analysis, naturally incorporating the limiting cases of a single reading and a large number of dispersed data.

The approach also leads to unambiguous communication of measurement results.