Calculating and Expressing Uncertainty in Measurement

Les Kirkup, Department of Applied Physics, Faculty of Science, University of Technology, Sydney, New South Wales 2007, Australia.
email: Les.Kirkup@uts.edu.au

CONTENTS

Section 1: Introduction to error and uncertainty

1.1 Introduction

Recent exciting research raises the possibility that the speed of light in a vacuum, which until now has been assumed to be constant, has been changing slowly with time since the universe was created.

This controversial (and yet to be confirmed) finding, stems in part from work done on spectral lines of light emitted from metal atoms in quasars[1] situated at different distances from the earth. In particular, researchers have measured the fine structure constant[2], a, (which depends on the speed of light) and have concluded from their data that a was smaller in the distant past than it is today. The work has been published in the most prestigious physics journal[3]. If this finding *is* supported by the work of independent researchers, it is no exaggeration to say that physics books will need to be rewritten, so profound are the implications[4].

The researchers who carried out the work have carefully examined all aspects of their experimental technique in order to establish the size of the errors associated with their measurements. Their conclusion at this stage is that, even accounting for sources of error, they have strong evidence for an increase in the fine structure constant over time.

The nature of science (and the importance of this work in particular) is such that researchers around the world will be keen to repeat the experiments as well as subject the error analysis of the original researchers to close inspection. One overlooked (or incorrectly assessed) source of error could transform potential Nobel prize winning work into an amusing curiosity of little or no consequence.

Errors, and their identification, are so important in all of science that we devote this document to a discussion of errors and how their effects may be quantified and expressed as an *uncertainty* in values obtained through measurement. This document also includes worked examples and exercises.

---

[1] A quasar is an extremely bright (and distant) stellar object. For details refer to Longair M S *Our Evolving Universe* (1996) Cambridge University Press.
[2] Blatt F J *Modern Physics* (1992) McGraw-Hill, section 5.8.
[3] Webb J K, Murphy M T, Flambaum V V, Dzuba V A, Barrow J D, Churchill C W, Prochaska J X and Wolfe A M *Further Evidence for Cosmological Evolution of the Fine Structure Constant* (2001) Phys. Rev. Lett. **87**, (9) pp.091301/1-4
[4] For a popular account, see Brooks M *Blinding Flash* (2002) New Scientist **174**, 28

1.2 True value and experimental error

In the forthcoming sections we will deal with error and uncertainty in some detail. Here we introduce the ideas of measured value[5], true value and error.

It is often helpful to imagine that a quantity, such as the length of a wire, has a 'true' value and it is the true value that we seek through measurement[6]. If measurement procedures and instruments are perfect and no influences conspire to affect the value, then we should be able to determine the true value of a quantity to as many significant figures as we desire. Recognising that neither experimental methods nor instruments are perfect and that outside influences though they can be minimised or controlled, can never be completely eliminated, the best we can do is obtain a 'best' estimate of the true value. If we *could* know the true value of a quantity then, in an experiment to determine the length, l, of a copper wire we could perhaps write,

$$l = 32.1476569 \text{ cm}$$

The reality is that imperfect calibration of the measuring instrument, the influence of slight changes in temperature and/or humidity, and limited instrument resolution cause measured values to vary from the true value by such an amount as to render most of the figures after the decimal point in l above quite meaningless.

We refer to the difference between the measured value and the true value as the '*experimental error*'[7] (which is usually referred to simply as the '*error*').

Representing the true value of a quantity by the symbol, Y, the error in the ith measured value, dy_i, is given by,

$$dy_i = y_i - Y \qquad (1.1)$$

where y_i is the measured value, and is regarded as an estimate of Y.

As Y in equation 1.1 cannot be determined, dy_i is unknowable also. However, we are usually able to estimate the effect that various errors will have on the measured value. Once this has been done, we attach to y_i an *uncertainty* which indicates a range of values within which we expect Y to lie.

1.3 Errors affect all measurements

Even the most carefully designed and executed experiments using 'state of the art' instruments and which are performed in temperature and humidity controlled environments, yield values that are influenced by various sources of error. A key responsibility of an experimenter is to attempt to identify sources of errors that may affect the measurement process, and then quantify the likely extent of those errors.

---

[5] The 'measured value' is referred in some texts as the 'estimated value'.

[6] Some important documents, including the *Guide to the Expression of Uncertainty in Measurement* (which we consider in some detail beginning in section 2), argue that the word 'true' is superfluous, and refer simply to the 'value' of a quantity. In the interests of clarity, I will retain the word 'true'.

[7] By 'error' we are not considering a mistake such as misreading a scale or writing down the 'wrong' number. Such mistakes or 'blunders' cannot easily be accounted for in error analysis and we must rely on careful experimentation to reduce the probability of occurrence of such mistakes to a minimum.

This is not a trivial task, especially if an unfamiliar or novel experiment is to be performed (there are many examples of situations in which an error has been overlooked, or its effect has been underestimated).

Even what appears to be the simplest measurement has several sources of error that need to be considered. For example, consider the measurement of the length of a copper wire using a wooden metre rule. Focussing for a moment on the metre rule, factors that might affect the measured value include the,

- calibration of the rule (the rule can, at best, be only as good as the 'standard rule' against which it was calibrated)
- temperature of the rule (the rule would have been calibrated at a certain temperature - if the rule is now used at a different temperature, the change in dimensions of the rule that occurs will influence any measurement made).
- moisture content of the rule (a wooden rule will absorb or release moisture and this is dependent upon ambient humidity levels - changing moisture content will cause the wood to expand or contract)
- thickness of scale markings on the rule (this will affect how easy it is to line the rule up with the end of the wire)
- resolution limit of the rule (a metre rule is usually marked with a smallest scale division of 1 mm, it follows that lengths much less than 1 mm are difficult to measure).

We might also add to this list 'general wear and tear' due to the rule being moved, dropped, twisted, or flexed.

There are several sources of error associated with the wire itself, including,

- how the wire is supported or held (if the wire is suspended vertically, for example, it will increase in length slightly due to the gravitational force it experiences)
- any bending or twisting of the wire
- 'kinks' in the wire
- non uniform temperature distribution along the length of the wire (a copper wire will expand when its temperature increases). A long wire may be situated in such a position that a temperature gradient exists along the wire
- temperature variation of the wire that occurs during the course of the measurement (if the wire is handled thereby increasing its temperature, the wire would be expected to expand).

As the measurement is likely to be made by a person using the rule held close to the copper wire, there are possible errors introduced at this stage, including,

- positioning of the rule so that a scale marker coincides with one end of the wire (parallax error introduces an offset between scale marker and the end of the wire)
- determining the position of the other end of the wire against the scale on the rule (again parallax error could introduce an offset between scale marker and end of wire).

It is important to be aware that there is error introduced if the quantity to be measured is incompletely specified or defined. For example, if the length of a copper wire is to be determined, but no information is given as to the temperature at which the measurement is to be carried out, then two experimenters given the same piece of wire may obtain different values if their measurements are carried out at different temperatures.

The list given above is not intended to be complete, nor is there any indication of which errors are the most important and which are so small that they can be neglected. For example, consider a copper wire whose length is to be measured at a specified nominal temperature of 20 °C. The variation in length of a copper wire of 'about' 85 cm due to a temperature which varies by as much as 10 °C from the nominal value is less than 0.15 mm. It is quite possible that another source of error (such as that due to the limited resolution of the rule) will be more important and that the error due to expansion would be only a small contributor to the overall error and can be neglected.

From this example it may be seen that there are many sources of error that 'lie in wait' to trap the unwary when measuring the length of a wire and this may be generalised to all types of measurement. It is fair to say that experience can assist enormously when trying to identify sources of error. In addition, it is extremely valuable to perform a preliminary experiment which has as its primary focus the identification of influences which will affect a measurement.

1.3.1 Reality check

Perhaps it is reasonable to apply a 'reality check' at this point. Experimentation in science is essential, and careful measurement is of vital importance. However, all experiments must fit within a finite time scale, and this may preclude a detailed investigation of all the sources of error that have been identified. Recognising the most important sources of error, and quantifying the uncertainties they introduce in order to assure that the values obtained are fit for the purpose for which they were collected, should be primary goals of the experimenter.

1.4 Systematic and random error

Random errors cause measured values to lie above and below the true value and, due to the scatter they create, they are usually easy to recognise. So long as only random errors exist, the best estimate of the true value, which is often taken to be the mean of many measured values, tends towards the true value as the number of repeat measurements increases[8].

Another type of error is that which causes the measured values to be consistently above or consistently below the true value. This is termed a *systematic* error, also sometimes referred to as a *bias* error. An example of this is the zero error of a spring balance. Suppose with no mass attached to the spring, the balance reads 0.01 N. We can ‘correct’ for the zero error by,

- if the facility is available, adjusting the balance to read zero or,

[8] See chapter 5 of Data Analysis with Excel by Kirkup (in press with Cambridge University Press, 2002).

- subtracting 0.01 N from every value obtained using the balance.

No value, including an offset, can be known exactly, so the correction applied to a value will have its own error. As a consequence, we cannot assume that a correction applied to a value has completely eliminated the systematic error.

Some systematic errors are difficult to identify and have more to do with the way a measurement is made, rather than any obvious inadequacy or limitation in the measuring instrument. Systematic errors have no effect on the precision of values, but rather act to impair measurement accuracy[9].

1.5 Error and uncertainty

To this point we have focussed upon errors that may affect the measurement process. As we are unable to determine the error, it seems that we might be forced to say something to the effect:

The length of the copper wire at 20 °C is about 85.5 cm. The error in the length of the wire is unknown.

While these statements may be correct, they are of limited worth.

What we need is a 'number' related to the error that we *are* able to determine. This will allow us to express a value of a quantity obtained through measurement in such a way that we indicate,

i) the most likely true value (i.e. the 'best estimate') of the quantity
ii) the range of values that may reasonably be assigned to the true value, based on our 'imperfect' measurements and any other factors that we are aware of.

We would like to be able to say:

The best estimate of the true value of the length of the copper wire at 20 °C is 85.5 cm. There is a probability of 0.95 that the true value lies between (85.5 ± 0.6) cm.

This maybe abbreviated to,

$$l = (85.5 \pm 0.6)\ \text{cm}$$

where it assumed that 85.5 cm is the best estimate of the length of the wire. 0.6 cm is the *uncertainty* which is used to indicate an interval likely to contain the true value. By considering the best estimate and uncertainty, which are both quantifiable, we move away from the concepts of true value and error which are not quantifiable.

How are uncertainties in values calculated and expressed ? This important question has brought together some of the leading organisations in metrology around the world, including the Comite International des Poids et Mesures (CIPM) and the

[9] Accuracy and precision are qualitative terms. If a measured value (or the mean of many values) is accurate, it is close to the true value. If values are precise, the scatter of values is small (but nothing can be inferred about how close those values are to the true value).

International Organisation for Standardisation (ISO). A goal of these organisations has been to arrive at consensus regarding the expression of uncertainty. Through the publication of authoritative documents the intention is to recommend to scientists and engineers around the world how uncertainties should be dealt with, in effect encouraging a consistent world wide approach to the calculation and expression of uncertainty.

Section 2 Guide to the Expression of Uncertainty in Measurement (GUM)

2.1 Introduction

To encourage uniformity in the way uncertainty in measurement is expressed, the ISO prepared and published the document 'Guide to the Expression of Uncertainty in Measurement' (GUM) in 1993.

The work, which culminated in GUM, was begun in the late 1970's and drew together experts in metrology from around the world. As a starting point, an ISO working group was given the following terms of reference[10]:

To develop a guidance document based on the recommendation of the BIPM[11] Working group on the Statement of Uncertainties which provides rules on the expression of measurement uncertainty for use within standardisation, calibration, laboratory accreditation and metrology services.

Though the terms of reference appear to focus upon services and facilities provided to the scientific, engineering and other communities by standards and calibration laboratories, GUM also anticipates the general rules it conveys to be applicable to a broad range of measurements carried out in science and engineering. For example, the rules are intended to be applicable to measurements carried out for the following purposes[12]:

- maintaining quality control and quality assurance in production
- complying with and enforcing laws and regulations
- **conducting basic research, and applied research and development, in science and engineering** (My emphasis)
- calibrating standards and instruments and performing tests throughout a national measurement system in order to achieve traceability to national standards
- developing, maintaining, and comparing international and national physical reference standards, including reference materials.

A broad aim of the GUM document (and other documents based on GUM) is to encourage those involved with measurement to supply full details on how uncertainties are calculated, permitting the comparison of values obtained through measurement by workers around the world.

The success of any method of evaluating and expressing uncertainty depends on several factors including,

i) its applicability to all types of measurements

---

[10] See GUM page v.
[11] Bureau International des Poids et Mesures.
[12] See GUM page 1 section 1.1.

ii) the ease with which uncertainties derived from different sources (such as calibration certificates, or tables of values published in data books) can be combined to give a 'total uncertainty'
iii) the willingness of scientists and engineers to apply recommended methods of combining and expressing uncertainties.
iv) a 'critical mass' of scientists and engineers adopting any method
v) the effectiveness of the dissemination of the details of any new method.

In the same way that the SI system of units took a number of years to become established and widely adopted, it may be anticipated that any new method for calculating and combining uncertainties, such as that described in the GUM document, will take some time to become widely known and established beyond the 'standards and calibration community'.

2.2 Basis of GUM

The foundation upon which GUM is based is explained in this section (please note this is my interpretation of descriptions given in GUM. For specific and definitive wording, see GUM page viii).

1.
Uncertainties can be usefully categorised in two ways according to how they are estimated.

Type A
These are uncertainties evaluated by the statistical analysis of a series of measurements

Type B
These are uncertainties evaluated by means other than the statistical analysis of a series of measurements

Note that the traditional classification of uncertainties into two categories, one which classifies uncertainties as being due to 'random effects' and those due to 'systematic effects' is *not* recommended by GUM.

Any report of uncertainty should give a detailed and complete description of how the uncertainty was determined.

2.
Uncertainties classified as Type A are characterised by an estimate of the variance, s_i^2 and the number of degrees of freedom, n_i. Where the correlation between uncertainties is significant, the covariance should be given.

3.
Type B uncertainties are characterised by a variance, u_j^2, which is derived from such sources as calibration certificates or tolerance limits.

4.
Combined uncertainty is calculated by taking the squares root of the sum of the variances due to Type A and Type B uncertainties.

5.
In some situations, it may be necessary to quote an interval within which the 'true' value of the measurands is likely to lie. In such cases, a coverage factor, k, must be stated.

2.2.1 Terms used in GUM

Several terms appear in the GUM document which are unfamiliar and require some explanation.

Standard uncertainty
This is the uncertainty in the result of a measurement expressed as a standard deviation

Combined standard uncertainty, u_c
This is obtained by taking the positive square root of the sum of the variances of the standard uncertainties of all contributing quantities (whether these be Type A or Type B uncertainties).

Relative combined standard uncertainty

This is the ratio of the combined standard uncertainty, $u_c(y)$, to the magnitude, $|y|$ of the best estimate of the measurand, i.e.,

$$\text{The relative combined uncertainty} = u_c(y)/|y|$$

where $|y| \neq 0$.

Expanded uncertainty, U
This is an uncertainty which encompasses a fraction of values (often 95 %) that can reasonably be attributed to the measurand. The term *level of confidence* is used to express the probability that the true value will lie within U of the best estimate.

Coverage factor, k
This is used to obtain the expanded uncertainty, U, given the combined standard uncertainty, u_c. Specifically,

$$U = ku_c$$

For an expanded uncertainty where the level of confidence is 95 %, k usually lies between 2 and 3.

Section 3: Uncertainty

A quantity, such as temperature, time or electrical resistivity determined through measurement is referred to as a *measurand*. Temperature and time are quantities that can be measured more or less 'directly' using an instrument. By contrast, several quantities need to be brought together in order to determine the resistivity of a material[13]. For example, for a material prepared in the form of a cylinder, the resistivity, r, of the material can be written,

$$r = \frac{RA}{l} \tag{3.1}$$

where R is the electrical resistance between the two ends of the cylinder, A is the cross-sectional area of the cylinder, and l is the length of the cylinder. The quantities on the right hand side are susceptible to error (due, for example, to temperature effects) and these errors contribute to an uncertainty in the resistivity, r.

In general, any measurement (or series of measurements) will only yield an *estimate* of the true value of a measurand. In order to communicate this to others and to indicate the likely range in which the true value of the measurand lies, we quote an uncertainty, so that (for example) if the value of the resistivity of a wire is written as

$$r = (1.72 \pm 0.11) \times 10^{-8}\ \Omega\cdot\text{m}$$

We can infer that there is some probability (usually 0.95) that the true value lies between $1.61 \times 10^{-8}\ \Omega\cdot$m and $1.83 \times 10^{-8}\ \Omega\cdot$m. As far as possible, any *inference* required by the way a value is expressed is better replaced by detailed information regarding what the uncertainty 'really means'. This leaves a reader in no doubt as to meaning of the estimate of the true value and its associated uncertainty.

The GUM document suggests several possible sources of uncertainty when determining the value of a measurand, which include,[14]

a) incomplete definition of measurand (for example, if the resistivity of a metal is to be determined, the purity of the metal and the temperature at which the resistivity is required must be specified)
b) non-representative sampling (for example, if the cross-sectional area of a cylinder varies, but measurement of the diameter of the cylinder is carried out at only one position along the cylinder, the variation in the diameter is likely to be underestimated)
c) finite instrument resolution
d) inexact values of constants and other parameters obtained from external sources and used in the data-reduction algorithm (for example, if the value of the charge on the electron appears in an equation, it is most likely that a 'data-book' value for the charge will be entered into the equation. It follows that any calculations of uncertainty should include the 'data-book' uncertainty in the charge).

---

[13] Refer to any introductory Physics text such as University Physics 8th Edition by Young (1992) Addison Wesley p717.
[14] see pages 5 and 6 of GUM.

e) variation in repeated observations of the measurand under apparently identical conditions (for example, dropping a ball from a height of 1 m onto a metal surface a number of times and assessing the rebound height 'by eye' is likely to yield values which exhibit scatter or variability).

There is no reason to classify uncertainties at all, as the identification and quantification of uncertainties are the primary considerations. Nevertheless, GUM recommends that uncertainties be divided into two categories, Type A and Type B, for the convenience of discussion. Type A and Type B uncertainties are based upon probability distributions. Type A uncertainties are calculated by the experimenter on the basis of repeat measurements carried out by the experimenter (very often assuming the normal or *t* distribution is valid for the variability in the mean of the values).

A Type B evaluation of uncertainty requires information from, as examples, data books, calibration certificates, 'historical records' of the variability in a quantity or known tolerance limits. By and large a Type B evaluation of uncertainty is obtained by *assuming* a particular probability distribution is valid (such as a rectangular or triangular distribution). It is a recognition that equipment, facilities, personnel or time is not available to determine the uncertainty of every measurand by a Type A evaluation that make Type B evaluations of uncertainty necessary.

Throughout the process of evaluating the uncertainty, the goal is to arrive at a 'realistic assessment' of uncertainty and not to unnecessarily overestimate the uncertainty in an attempt to 'play safe'.

Before moving onto consider calculations of Type A and Type B uncertainties, it is worth considering some advice contained in GUM[15]:

Although [*GUM*] *provides a framework for assessing uncertainty, it cannot substitute for critical thinking, intellectual honesty, and professional skill. The evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on the detailed knowledge of the nature of the measurand and the measurement. The quality and utility of the uncertainty quoted for the result of a measurement therefore ultimately depend on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value.*

3.1 Evaluating uncertainty

In the most general situation we write that some measurand, Y, depends on N quantities written as X_1 through to X_N, such that,

$$Y = f\left(X_1, X_2 .... X_N\right) \qquad (3.2)$$

The quantities $X_1.X_2... X_N$ which appear in equation 3.2 are themselves measurands and they may depend on other quantities. Note that sometimes the symbols $X_1.X_2... X_N$ are used to represent quantities. At other times the same symbols are used to represent the true value of the quantities.

[15] see GUM page 8.

In order to make some of these ideas more 'concrete' we continue with the example introduced in the last section in which the electrical resistivity, r, of a material is determined. The electrical resistivity, r, is given by,

$$r = \frac{RA}{l} \quad (3.3)$$

where R is the electrical resistance, A, cross-sectional area, and l the length of a cylinder made from the material.

The area¸ A, of cross-section of the cylinder depends on the diameter, d, through the equation,

$$A = \frac{pd^2}{4} \quad (3.4)$$

3.2 Type A evaluation of uncertainty

In many cases, the input quantities X_1, X_2... X_N appearing in equation 3.2 are determined through repeat measurements. For example, the area of cross-section of a cylinder, A, may be determined by measuring the diameter, d, of the cylinder at several, positions (chosen at random) along the cylinder using a micrometer.

We can denote the individual measured values as d_1, d_2.....d_n, where n is the number of repeat measurements made. The best estimate of the true value of the diameter (and in general any quantity determined by repeat measurement) is found by taking the arithmetic mean of the values. i.e.,

$$\bar{d} = \frac{\sum_{i=1}^{i=n} d_i}{n} \quad (3.5)$$

where $\bar{d}$ is the mean diameter.

Example 1
As part of an experiment to determine the circular cross-sectional area of a cylinder, the diameter of the cylinder is measured. The values obtained for the diameter are shown in table 3.1

d (mm)	1.25	1.27	1.25	1.29	1.26	1.26	1.21	1.20

Table 3.1: values of diameter of a cylinder

The best estimate of the true value is given by equation 3.5,

$$\bar{d} = 1.24875 \text{ mm}$$

As measured values of a quantity generally show some variability, we use the standard deviation, s, as a measure of this variability[16]. The equation for the sample standard deviation, where the individual measured values are denoted by d_1, $d_2 .....d_n$, is given by,

$$s = \left(\frac{\sum \left(d_i - \bar{d} \right)^2}{n-1} \right)^{1/2} \tag{3.6}$$

While s is a measure of the spread of the whole data, the standard deviation of the mean (often referred to as the *standard error* of the mean) is taken as a measure of the variability of the means (If an experiment consisting n measurements is repeated many times, then the distribution of means has its own standard deviation).

If we represent the standard deviation of the mean as $s\left(\bar{d}\right)$, then[17],

$$s\left(\bar{d}\right) = \frac{s}{\sqrt{n}} \tag{3.7}$$

The terms 'standard error' and 'standard deviation' are found throughout the statistics literature. GUM introduces a new term, *standard uncertainty*, represented by the symbol, u.

When the mean, $\bar{d}$, is the best estimate determined from repeat measurements, then the standard uncertainty is written

$$u\left(\bar{d}\right) = s\left(\bar{d}\right) \tag{3.8}$$

Using more general symbolism, if $\bar{X}$ is the best estimate of X, then the standard uncertainty is written as $u\left(\bar{X}\right)$.

When communicating the standard uncertainty from a Type A evaluation, the number of degrees of freedom, n, should be stated[18], as this information is required when calculating the expanded uncertainty, U.

Exercise 1
Ten measurements are made of the electrical resistance of a wire at 300 K. The values obtained are shown in table 3.2.

R(Ω)	0.257	0.253	0.259	0.250	0.251	0.251	0.257	0.258	0.255	0.252

Table 3.2: Values of resistance of a wire obtained by repeat measurements

[16] s is often referred to as the 'sample' standard deviation, because the values used to calculate s in equation 3.6 represent a sample taken from a much larger population (often the population can be regarded as infinite).
[17] see chapter 3 of Kirkup (2002).
[18] In situations in which n repeat measurements are made of the same quantity, $n = n$ -1.

Determine,

i) the best estimate of the true value of the resistance [$\overline{R} = 0.2543\,\Omega$]

ii) the standard deviation of the resistance values[19] [s = 0.003302 Ω]

iii) the standard uncertainty in the best estimate [$u\left(\overline{R}\right) = 0.001044\,\Omega$]

iv) the number of degrees of freedom [n = 9].

3.3 Type B evaluation of uncertainty

Where the best estimate of an input quantity X_i is not determined by repeat measurements, we cannot use the Type A evaluation methods to establish the standard uncertainty. In such a situation we must use previous measurement data, calibration certificates, manufacturers specifications, data tables and the like for both the best estimate of the input quantity and the uncertainty. An uncertainty determined in this manner is referred to as a 'Type B' uncertainty. In such cases much care must be taken to ensure that the uncertainty, as expressed in the manufacturers' tables, calibration certificate and so on, is correctly interpreted.

For example, if a calibration certificate supplied with a standard resistor states that the nominal resistance at 20 °C is R_s = 100.05 Ω, with an uncertainty of 50 mΩ at the 2 standard deviation level, then the standard uncertainty $u(R_s)$ = (50/2) mΩ, ie 25 mΩ.

The quoted uncertainty may not be as simple to interpret as in this example. Often an uncertainty is quoted without additional information as to whether the uncertainty is a multiple of a standard deviation, or represents an interval with a certain 'level of confidence'. This is one reason why those who prepared GUM hope that their approach to quoting uncertainties becomes widely adopted!

In situations in which Type B evaluations are carried out, some assumption must be made regarding the distribution that underlies the variability in the value. This is not unlike the assumptions that are usually made when Type A evaluations of uncertainty are carried out. It is possible to establish the validity of the assumptions regarding the distribution used in a Type A evaluation, as the data gathered in an experiment can be compared with a 'theoretical' distribution. By their very nature, Type B evaluations do not permit such comparisons.

3.3.1 Type B evaluation: Estimate of true value

If a calibration certificate gives upper and lower limits for, say, the value of a standard mass, then we assume that the best estimate of the true value of the mass lies halfway between the limits. Denoting the upper limit as a_+ and the lower limit as a_-, then the best estimate of the true value is given by,

$$\text{best estimate} = \left(\frac{a_+ + a_-}{2}\right) \qquad (3.9)$$

If we are given no information regarding which values lying between the limits a_+ and a_- are likely to be more probable, we have little alternative but to assume that every value between those limits is equally probable. As a consequence, the probability

[19] It is customary to quote standard deviations to no more than two significant figures. More figures are retained here to allow for careful checking of calculations.

distribution representing the variability of values is taken to be rectangular as shown in figure 3.1.

Figure 3.1 shows the situation in which it is assumed that a rectangular distribution is applicable to the distribution of values. $f(x)$ is the probability density function, such that $f(x)\Delta x$ is the probability that the true value lies in an interval of width Δx.

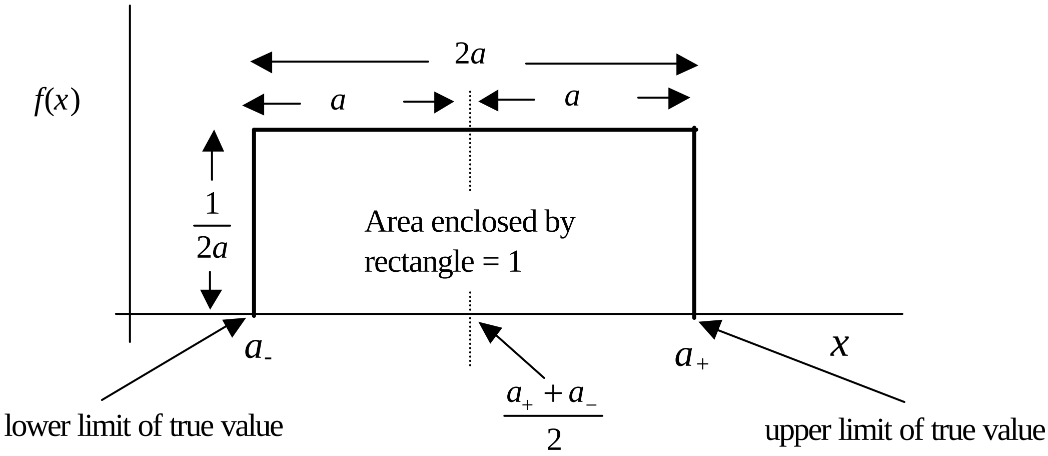

Figure 3.1: Rectangular probability distribution

For example, if we are told the maximum diameter of a wire is 1.46 mm and the minimum diameter is 1.42 mm, then the best estimate of the true value of the diameter is $\left(\frac{1.46+1.42}{2}\right)$= 1.44 mm. We can represent the distribution of diameters by redrawing figure 3.1 as shown in figure 3.2.

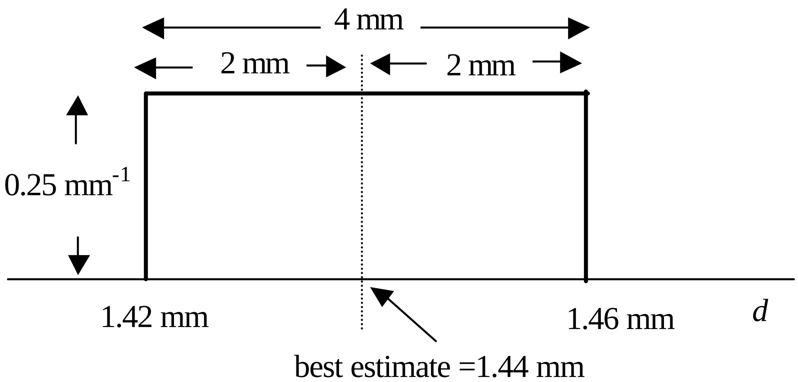

Figure 3.2: Assumed distribution of diameters when all that is known are the upper and lower bounds.

3.3.2 Type B evaluation: Standard uncertainty in best estimate of true value when underlying distribution is rectangular

In order to determine the standard deviation in the best estimate, when the underlying distribution is rectangular, we apply the relationship for the variance, s^2,

$$s^2 = \int_{-\infty}^{\infty} (x-m)^2 f(x)dx \qquad (3.10)$$

x is a continuous random variable and m is the population mean[20]. In the case of the rectangular distribution shown in figure 3.1, the best estimate of the variance, s^2, is given by,

$$s^2 = \frac{a^2}{3} \tag{3.11}$$

where a is the half width of the distribution[21].

The standard deviation (which we take to be the Type B standard uncertainty, u) of the distribution is given by,

$$u = s = \frac{a}{\sqrt{3}} \tag{3.12}$$

Example 2: If the true value for the diameter of a wire lies within a distribution of half width 0.02 mm, determine the standard uncertainty of the best estimate of the true value of the diameter.

Answer: Applying equation 3.12 gives,

$$u = \frac{a}{\sqrt{3}} = \frac{0.02}{\sqrt{3}} = 0.012 \text{ mm}$$

Exercise 2: A data book indicates that the density of an alloy definitely lies between the limits 6472 kg/m^3 and 6522 kg/m^3. Using this information, determine,

a) the best estimate for the alloy density [6497 kg/m^3]
b) the Type B evaluation of standard uncertainty [u =14 kg/m^3, assuming a rectangular distribution is appropriate]

3.3.3 Type B evaluation: Standard uncertainty in best estimate of true value when underlying distribution is triangular

A rectangular distribution, in which the probability of a true value being close to the limits of the distribution is the same as the probability of the true value being close to the centre, is likely to be 'unphysical' in many situations.

An alternative to the rectangular distribution is to assume that the underlying distribution is triangular, as shown in figure 3.3.

[20] See chapter 3 of Kirkup (2002).

[21] A convenient way to show equation 3.11 is to consider a situation in which the expected value for x is equal to 0 (ie m =0). Given that the probability density function, $f(x)$, is equal to $1/2a$ between $-a$ and $+a$ (and zero outside these limits), integrate equation 3.10 (with m = 0) between $x = +a$ to $x = -a$.

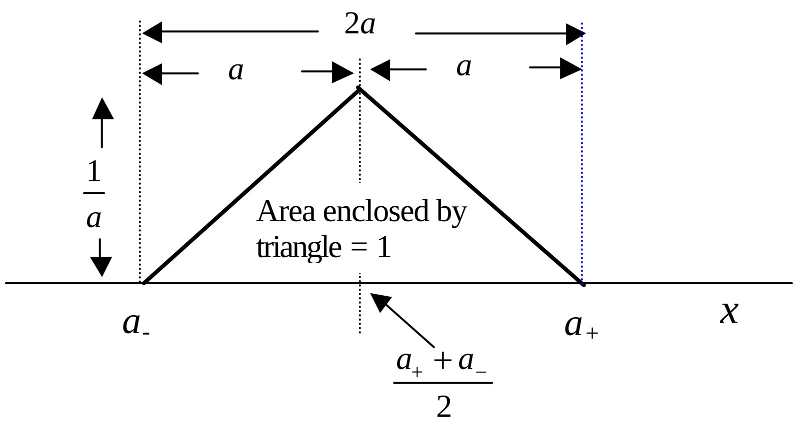

Figure 3.3: Triangular probability distribution

When the distribution is triangular, the best estimate of the true value is given by,

$$\text{best estimate} = \left(\frac{a_+ + a_-}{2}\right) \tag{3.13}$$

By contrast, the standard uncertainty in the best estimate when the distribution is triangular (which can be found by applying equation 3.10) is given by :

$$u = s = \frac{a}{\sqrt{6}} \tag{3.14}$$

Exercise 3: A calibration certificate indicates that a weight lies between 100.052 g and 100.074 g. Assuming that a <u>triangular</u> distribution is appropriate, determine,

a) the best estimate of the true value of the weight [100.063 g]
b) the standard uncertainty in the best estimate [0.0045 g]

Exercise 4
Show that the standard deviation of a triangular distribution of half width a, (as shown in figure 3.3) is given by $s = \frac{a}{\sqrt{6}}$.

Section 4: Combined standard uncertainty

If we assume that Y depends on quantities A, B, and C and so on, then we can combine best estimates of A, B and C, (which we will refer to as a, b and c) to find the best estimate of Y. If there is no correlation between the standard uncertainties in a, b and c, we can write the combined variance $u_c^2(y)$ (which takes into account the standard uncertainties in the best estimates a, b and c, written as $u(a)$, $u(b)$ and $u(c)$ respectively) as,

$$u_c^2(y) = \left(\frac{\partial f}{\partial a}u(a)\right)^2 + \left(\frac{\partial f}{\partial b}u(b)\right)^2 + \left(\frac{\partial f}{\partial c}u(c)\right)^2 \qquad (4.1)$$

where f is the function which relates Y to A, B and C. This may also be written:

$$u_c^2(y) = u_1^2(y) + u_2^2(y) + u_3^2(y) \qquad (4.2)$$

where $u_1^2(y) = \left(\frac{\partial f}{\partial a}u(a)\right)^2$, $u_2^2(y) = \left(\frac{\partial f}{\partial b}u(b)\right)^2$ etc. Equation 4.2 may be written more succinctly as,

$$u_c^2(y) = \sum_{i=1}^{i=N} u_i^2(y) \qquad (4.3)$$

The derivatives, $\frac{\partial f}{\partial a}, \frac{\partial f}{\partial b}$ etc. appearing in equation 4.1, are sometimes referred to as 'sensitivity coefficients'. These are equal to $\frac{\partial f}{\partial A}, \frac{\partial f}{\partial B}$ evaluated at the best estimates of the contributing quantities (which in the case of a Type A evaluation, would the mean of each quantity).

Example 3

The mass, m, of a wire is found to be 2.255 g with a standard uncertainty of 0.032 g. The length, l, of the wire is 0.2365 m with a standard uncertainty of 0.0035 m. The mass per unit length, $\mathtt{m}$, is given by:

$$\mathtt{m} = \frac{m}{l} \qquad (4.4)$$

Determine the,

a) best estimate of $\mathtt{m}$,

b) standard uncertainty in $\mathtt{m}$..

Answer:

a) Using equation 4.4, $\mathtt{m} = \frac{m}{l} = \frac{2.255}{0.2365} = 9.535$ g/m

b) Using equation 4.1,

$$u_c^2(m) = \left(\frac{\partial m}{\partial m}u(m)\right)^2 + \left(\frac{\partial m}{\partial l}u(l)\right)^2 \qquad (4.5)$$

From equation 4.4,

$$\frac{\partial m}{\partial m} = \frac{1}{l} = \frac{1}{0.2365} = 4.2283 \text{ m}^{-1},$$

$$\text{and } \frac{\partial m}{\partial l} = -\frac{m}{l^2} = -\frac{2.255}{0.2365^2} = -40.317 \text{ g/m}^2$$

Substituting values into equation 4.5 gives,

$$u_c^2(m) = (4.2283 \times 0.032)^2 + (-40.317 \times 0.0035)^2$$
$$= 0.01831 + 0.01991 = 0.03822 \text{ (g/m)}^2$$

therefore,

$u_c(m)$ = 0.1955 g/m

Exercise 5

For a given metal wire it is found that the,

best estimate of its length is 1.21 m with a standard uncertainty of 0.01 m
best estimate of its diameter is 0.24 mm with a standard uncertainty of 0.01 mm
best estimate of its resistance is 0.52 Ω with a standard uncertainty of 0.02 Ω.

Using this information and equations 3.3, 3.4 and 4.1, determine the,

a) best estimate of the cross-sectional area, A, of the wire [4.524×10^{-8} m^2]
b) standard uncertainty in A [3.77×10^{-9} m^2]
c) best estimate of the electrical resistivity, r, [1.944×10^{-8} Ω·m]
d) standard uncertainty in r [1.8×10^{-9} Ω·m].

4.1 Expanded uncertainty

The standard uncertainty, u_c, is the basic measure of uncertainty in the GUM document. However, there are situations in which it is preferable to quote an interval within which there is a specified probability (usually 0.95) that the true value will lie. Specifically, we can write,

$$y - U \le Y \le y + U \qquad (4.6)$$

Y is the true value of the quantity, y is the best estimate of the true value, and U is the *expanded uncertainty.*

Equation 4.6 is often written,

$$Y = y \pm U \qquad (4.7)$$

The expanded uncertainty, U, is related to the combined standard uncertainty, u_c, by the equation:

$$U = ku_c \qquad (4.8)$$

k is referred to as the *coverage factor*.

4.1.1 Coverage factor and Type A evaluation of uncertainty

If u_c is determined through a Type A evaluation of uncertainty, then it is usual to assume that the t distribution[22] may be applied when determining the coverage factor, k.

When the level of confidence[23] is 0.95 (i.e. the probability that the true value lies within a specified interval is 0.95), then table 4.1 gives the coverage factor for various degree of freedom, n. For values of n > 10, k tends towards a value of close to 2. When n > 10, experimenters often use 2 as the coverage factor when the level of confidence required is 0.95 (when the level of confidence is 0.99, the corresponding value of k is 3).

n	1	2	3	4	5	6	7	8	9	10
k	12.706	4.303	3.182	2.776	2.571	2.447	2.365	2.306	2.262	2.228

Table 4.1: Coverage factors in Type A evaluations for n degrees of freedom when the level of confidence is 0.95.

Example 4
In an experiment to calibrate a 1 mL pipette, the mass of water dispensed by the pipette was measured using an electronic balance. 10 measurements were made of the mass of water dispensed by the pipette. The mean mass of water was found to be 0.9567 g, with a standard error in the mean of 0.0035 g. Using this information, determine the,

a) best estimate of the true value of the mass of water dispensed by the pipette
b) standard uncertainty in the mass, u
c) coverage factor, k, for a 0.95 level of confidence
d) expanded uncertainty, U, for a 0.95 level of confidence.

Answer
a) The best estimate is taken to be the mean of values obtained through repeat measurements, which in this example is 0.9567 g.
b) When Type A evaluations are carried out, the standard uncertainty is equal to the standard error of the mean i.e, u = 0.0035 g.

[22] See Kirkup (2002), chapter 3.
[23] GUM avoids the term 'confidence interval' often used in statistics texts to designate an interval which contains the true value with a given probability. This is because such terminology should strictly only be used when Type A evaluations of uncertainty are carried out and the use of the terms *confidence interval* or *confidence level* should not be broadened to include Type B evaluations of uncertainty.

c) In this example, the number of degrees of freedom, n = 10 - 1 = 9. Using table 4.1, the corresponding value for k is 2.262.
d) $U = ku = 2.262 \times 0.0035 = 0.0079$ g

Exercise 6
The Ca content of five samples of powdered mineral is determined. The mean of the five values is (expressed as percent composition) 0.02725 with a standard uncertainty (obtained through a Type A evaluation) of 0.00012. Determine the expanded uncertainty at the 0.95 level of confidence. [U = 0.00033]

4.1.2 More on coverage factor

The combined uncertainty will, in general, be determined from Type A and Type B evaluations of uncertainty.

Let us focus on Type A evaluations of uncertainty. It is possible that the best estimate of each quantity is obtained from a different number of repeat measurements. For example, the resistivity of a metal is obtained from knowledge of the resistance, length and diameter of a wire made from the metal. We may imagine that five repeat measurements are made of the resistance of the wire, seven of its diameter and four of its length. We are able to determine the combined uncertainty using equation 4.1, but how are we able to calculate the coverage factor for the expanded uncertainty? In this situation we calculate the *effective degrees of freedom*, n_{eff}, using the Welch-Satterthwaite formula[24], which can be expressed:

$$n_{\mathrm{eff}} = \frac{u_c^4(y)}{\sum_{i=1}^{N} \frac{u_i^4(y)}{n_i}} \qquad (4.9)$$

Example 5
The optical density (o.d.) of a fluid is given by:

$$\text{o.d.} = eCl \qquad (4.10)$$

e is the extinction coefficient, C is the concentration coefficient of the absorbing species in the liquid, and l is the path length of the light. The best estimate of each quantity, standard uncertainty and degrees of freedom are as follows:

e = 14.9 L·mol^{-1}·mm^{-1}; $u(e)$ = 1.2 L·mol^{-1}·mm^{-1}; n_e = 5

C = 0.042 mol·L^{-1}; $u(C)$ = 0.003 mol·L^{-1}; n_C = 7

l = 1.42 mm; $u(l)$ = 0.21 mm; n_l = 8

Determine,
a) the best estimate of the optical density, o.d.
b) the combined standard uncertainty

---

[24] See Dietrich C F *Uncertainty, Calibration and Probability 2nd edition* (1991) Adam Hilger, section 7.15.

c) the effective degrees of freedom, n_{eff}
d) the coverage factor n_{eff} calculated in part c) (assume that the level of confidence is 0.95).

Answer
a) o.d. = eCl = 14.9 × 0.042 × 1.42 = 0.889
b)

$$u_c^2(y) = \left(\frac{\partial f}{\partial e}u(e)\right)^2 + \left(\frac{\partial f}{\partial C}u(C)\right)^2 + \left(\frac{\partial f}{\partial l}u(l)\right)^2$$

$$\frac{\partial f}{\partial e} = Cl, \qquad \frac{\partial f}{\partial C} = el, \qquad \frac{\partial f}{\partial l} = eC$$

Now we write,

$$u_1^2(y) = \left(\frac{\partial f}{\partial e}u(e)\right)^2 = (Clu(e))^2 = (0.042 \times 1.42 \times 1.2)^2 = 0.05122$$

$$u_2^2(y) = \left(\frac{\partial f}{\partial C}u(C)\right)^2 = (elu(C))^2 = (14.9 \times 1.42 \times 0.003)^2 = 0.004029$$

$$u_3^2(y) = \left(\frac{\partial f}{\partial l}u(l)\right)^2 = (eCu(l))^2 = (14.9 \times 0.042 \times 0.21)^2 = 0.01727$$

Using equation 4.2, this gives,

$$u_c^2(y) = u_1^2(y) + u_2^2(y) + u_3^2(y) = 0.05122 + 0.004029 + 0.01727 = 0.02642$$

It follows that u_c = 0.1625

c) To find the effective degrees of freedom, n_{eff}, we use equation 4.9, such that,

$$n_{\text{eff}} = \frac{u_c^4(y)}{\sum_{i=1}^{N}\frac{u_i^4(y)}{n_i}} = \frac{(0.02642)^2}{\frac{(0.05122)^2}{5} + \frac{(0.004029)^2}{7} + \frac{(0.01727)^2}{8}} = 15.56 \text{ (round to 15)}$$

d) When the number of degrees of freedom is 15, the coverage factor for a level of confidence of 0.95 is 2.131 (see table 2 in appendix 1 of Kirkup (2002) for table of critical values for the *t* distribution)

4.1.3 n_{eff} and Type B uncertainties

What number of degrees of freedom do we assign to a standard uncertainty that is determined by a Type B evaluation based on an assumed probability distribution, such as the triangular or rectangular distribution? A rectangular distribution of half width a, has a standard uncertainty given by,

$$u = a/\sqrt{3} \qquad (4.11)$$

The standard uncertainty given by equation 4.11 has no uncertainty, i.e., $\Delta u = 0$.

It is possible to show that the number of degrees of freedom can be determined from knowledge of the standard uncertainty, u and standard deviation of the standard uncertainty, Δu, (i.e. the uncertainty in the uncertainty!)

If the uncertainty in u is Δu, then the number of degrees of freedom, n, can be written[25]:

$$n = \frac{1}{2}\left[\frac{\Delta u}{u}\right]^{-2} = \frac{1}{2}\left[\frac{u}{\Delta u}\right]^{2} \qquad (4.12)$$

If $\Delta u = 0$ as is the case for an assumed rectangular distribution, then using equation 4.12, it follows that $n = \infty$. The fact that $n = \infty$ does not cause a problem when determining the effective number of degrees of freedom using equation 4.9, as any term in the equation with an infinite number of degrees of freedom is zero.

[25] See GUM section E.4.3.

Section 5: The reporting of uncertainty

5.1 General considerations

Work done in establishing best estimates of quantities may demand that time be spent in the laboratory devising ways to minimise sources of error or perhaps require the detailed evaluation of reports, calibration certificates or specifications prepared by others. So that the time is well spent, it is important that best values and uncertainties are communicated in a way that is clear, concise and useful to others.

The GUM document recommends[26] that methods used to calculate best estimates and uncertainties in best estimates be clearly described. In addition, it is useful to,

- provide a full list of the uncertainty components and how they were evaluated
- present the steps in the data analysis so that the calculation of the combined standard uncertainty can be repeated
- describe any corrections made to values
- give any constants used in the analysis as well as the source of the constants

5.2 Detailed guidelines

When reporting the best estimate of a quantity and the combined standard uncertainty, we need to,

- fully define the measurand. For example, if the resistance of a metal wire is required, the temperature at which the resistance is measured is an essential piece of information.
- give the best estimate, y, of the measurand and the combined standard uncertainty, $u_c(y)$. The unit of measurement must be clearly stated
- if required, report the 'relative combined standard uncertainty', given by $\frac{u_c(y)}{|y|}$
- Give the expanded uncertainty, U, along with the coverage factor, k, used on the calculation of U.
- State clearly the level of confidence, associated with the interval, $y \pm U$.

5.2.1 Example of reporting uncertainty

The resistance, R, of a resistor of nominal resistance, 2.2 kΩ is measured at a temperature of 300 K. The best estimate of the resistance is found to be 2.215 kΩ with a combined standard uncertainty, u_c, of 0.022 kΩ. The expanded uncertainty, $U = ku_c$, where the coverage factor $k = 2.26$ is based on the t distribution with 9 degrees of freedom. We can therefore write,

$$R = (2.215 \pm 2.26 \times 0.022)\ \text{k}\Omega \quad \text{or}$$
$$R = (2.215 \pm 0.050)\ \text{k}\Omega$$

[26] See GUM section 7.

Example 6

The following is a short extract from a report to determine the velocity of sound in air. Indicate shortcomings in the report and suggest ways in which the reporting of the best value and the uncertainty could be improved.

Using the data gathered in this experiment, we determine the velocity of sound in air, $v_{S,}$ *to be 326.0 m/s with a combined standard uncertainty of 1.2. The expanded uncertainty, U, is 2.5 m/s.*

Solution

The report of the best value and uncertainty has the following shortcomings:

1) The measurand (the velocity of sound) is not well enough specified. Specifically, information is required regarding the,

a) composition of the air (for example, give the mole fractions of N_2, O_2, Ar and CO_2)
b) water content of the air
c) temperature of the air
d) air pressure.

2) The unit for the combined standard uncertainty is omitted (the unit should be m/s)
3) For the expanded uncertainty, *U*, there is no indication of the level of confidence adopted.
4) No mention is made of the value of the coverage factor, *k*, or the number of degrees of freedom, ***n***