4

Introduction to uncertainty in measurement

In this chapter we describe how consistency and clarity may be brought to the calculation and expression of uncertainty in measurement.

The goal of any measurement is to establish a numerical value for the measurand. Depending on the accuracy that we wish to claim for the numerical value, the procedure that gives us the value may be relatively simple and direct, involving no more than a tape-measure, for example. In other situations the process may be more complicated, with several intermediate stages requiring the resources of a well-equipped laboratory. Thus, if the measurand is the width of a table, the tape-measure is all that is needed. On the other hand, if the measurand is the accurate mass of an object, we need to know the value of the buoyancy correction (since the weight of the object is less by an amount equal to the weight of the volume of air that it displaces). This in turn requires knowledge of the volume of the object and of the density of air (which is a function of temperature, pressure and composition) at the time of measurement.

There are three components of a measurement: the measurand itself; the measuring instrument (which can be a stand-alone instrument such as a thermometer, or a complex system that occupies a whole laboratory); and the environment (which includes the human operator). The environment will, in general, affect both the measurand and the measuring instrument.

4.1 Measurement and error

4.1.1 Specifying measurand and environment

Error is the difference between the measured value of a measurand and the true value of the measurand. The true value cannot be known; it is an unreachable ideal in an imperfect world. However, we can regard it as the value close to the value that we would obtain if we could specify both the measurand and its environment in

very great detail, and if we possessed a measuring instrument of very high accuracy that was traceable to international standards.

If we do not specify the measurand in sufficient detail, then it is not fully defined, and so two people, measuring what they think is the same measurand, may actually make measurements under slightly different conditions, obtaining different values for this reason alone. As an example, the task might be to measure the diameter of a cylindrical brass rod. Here the diameter is the measurand. Although the rod may look cylindrical to the eye, its diameter will actually vary slightly, because of imperfections in the lathe that was used to turn the rod. As part of the process of fully specifying the measurand, we would therefore need to specify *where* the diameter should be measured – say, at the mid-point of the rod.

Similarly, if we do not specify the environment in sufficient detail, we are in effect neglecting the possibility that the measurand may be sensitive to the environment. In the above example, since brass expands or contracts with a rise or drop in temperature, we would need to specify the temperature of the environment in order to specify the measurand. By contrast, we note that, in attempting to estimate the true value of the measurand, we should not have to specify the instrument to be used (for example, its type, manufacturer or model), except for demanding that the instrument should have a certified very high accuracy.

The true value is, then, the value we would obtain for a completely specified measurand if we could use an ideal instrument in a completely specified environment. So we expect an error when we measure the measurand in an imperfect but more practically realisable manner.

Next we recognise that errors come in two flavours: 'random' and 'systematic'.

4.1.2 Random errors

The distinction between random and systematic errors is best seen by considering the notion of 'repeating the measurement under unchanging conditions', or as closely as we can arrange such conditions. By 'unchanging conditions' we mean a well-defined measurand, a tightly controlled environment and the same measuring instrument. Often when we repeat the measurement in this way, we will obtain a different value.[1] The reason for this lack of perfect repeatability is that the instrument we use or the measurand, or both, will be affected by *uncontrollable and small* changes in the environment or within the measurand itself. Such changes may be due, for example, to electrical interference, mechanical vibration or changes in temperature. So if we make the measurement ten times, we are likely to get ten

[1] We may obtain exactly the same value simply as a result of the limited *resolution* of the instrument – for example, if a digital instrument displays only two or three digits.

Table 4.1. *Voltage values as displayed by a DMM and difference from the mean voltage of 2.889* μV

DMM indication (μV)	Differences from mean (μV)
2.87	−0.019
2.91	+0.021
2.89	+0.001
2.88	−0.009
2.87	−0.019
2.88	−0.009
2.86	−0.029
2.95	+0.061
2.88	−0.009
2.90	+0.011

values that, although similar, vary by a small amount. When our intention is to obtain a single value for the measurand, we interpret such variations as the effect of errors. The errors fluctuate, otherwise we would see no variation in our values. Errors that fluctuate, because of the variability in our measurements even under what we consider to be the same conditions, are called *random* errors. In brief, random errors arise because of our lack of total control over the environment or measurand.

The first column of table 4.1 contains ten values in microvolts (millionths of a volt, symbol μV) recorded by a digital multimeter (DMM) once a second in a temperature-controlled laboratory. The values were obtained during the calibration of a source of constant voltage of nominal value 1 V. The small values of voltage are the differences in voltage that the DMM indicates between the source and a known and very stable value of a voltage standard. The mean of these ten values is 2.889 μV. The second column shows the ten differences between the measured values and this mean value, which is the measurand.

The differences sum to exactly zero (as all differences from a mean value must do), so both plus and minus signs must be present. These differences are scattered over a 0.090-μV range extending from −0.029 μV to +0.061 μV. This scatter, or 'dispersion', creates an *uncertainty* in the value obtained for the measurand.

This lack of total control over the environment, creating random errors, also affects cases where we make intentional changes to the environment. For example, the electrical resistance of a conducting material varies with temperature. To measure its temperature coefficient of resistance, we measure the resistance at intentionally different temperatures. When the resistance is a very stable and accurately known resistance making up what is known as a 'standard resistor',[2] we require

[2] A standard resistor is an example of an artefact standard (see section 3.1.4 and figure 3.2).

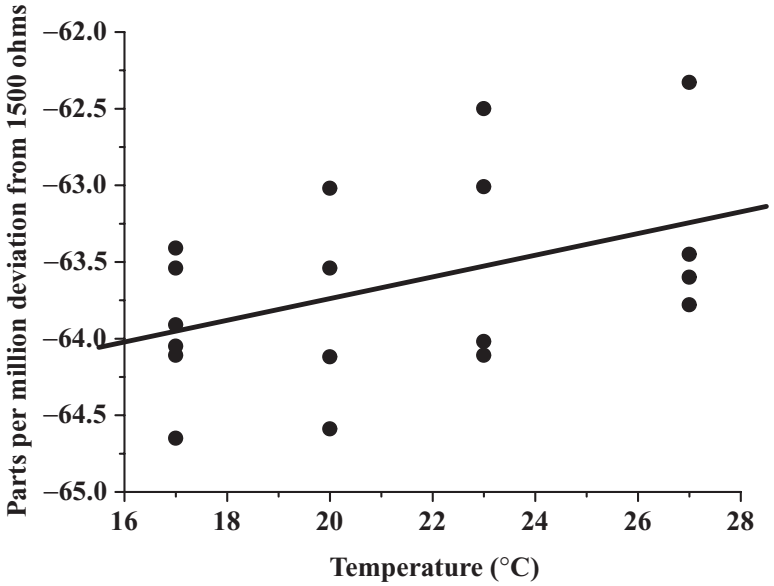

Figure 4.1. Random errors when measuring the temperature coefficient of a resistor (courtesy of the National Measurement Institute of Australia).

the distribution of temperature over its surface to be as uniform as possible. It is therefore immersed in a tank of stirred oil that can be set to various temperatures. We cannot *fully* control the temperature, however; nor its distribution over the body of the resistor. There may also be small fluctuations in the indication of the measuring instrument, possibly because the connecting wires pick up electromagnetic interference (from power-line and TV transmissions, for example). In brief, the environment has a basic randomness or 'noise' that we are unable to eliminate completely. So if we plot the measured resistance against temperature, as in figure 4.1, we are likely to observe a scatter of random errors around the 'line of best fit' that gives us the temperature coefficient of resistance. In this example, the resistor has a 'nominal' value of 1500 Ω and is wound from a special type of wire with a very low temperature coefficient. Several measurements have been taken at each of four selected temperatures. The temperature coefficient in figure 4.1, namely the slope of the line of best fit in the figure, is about $+0.071\ \mu\Omega/\Omega\ (^{\circ}\mathrm{C})^{-1}$ and, as will be discussed in section 5.2.3, the scatter of the points about this line can be given a quantitative value, namely $0.59\ \mu\Omega/\Omega$.

A sequence of reasonably stable measurements suggests a possible general way in which we might obtain the true value of a measurand. We make as many measurements as possible under the same conditions and calculate their mean. It is often correct that, in calculating the mean, the random errors will tend to cancel out, and their cancellation will yield a net error that we can claim with high confidence

to be very nearly zero if there is a very large number of measurements. Making many measurements is in fact generally preferable to making only a few – time and resources permitting. This applies also to cases where intentional changes are made, as in the values shown in figure 4.1; the greater the number of measurements, the more precisely we might expect to establish the value of the temperature coefficient. However, it is often pointless to take *very* many measurements to ascertain the true value of the measurand. The reason is the probable existence of the other flavour of error: a *systematic* error.

4.1.3 Systematic errors

During any measurement, there will probably be an error that remains constant when the measurement is repeated under the same conditions. An example of such an error is a constant offset in a measuring instrument. Unlike random errors, such *systematic* errors cannot be reduced by repeating the measurements and taking their mean; they resist statistical attack. The DMM in the above example (see table 4.1) might consistently – but unknown to us – have an offset, so that it indicates 1 μV too high no matter how many measurements we make. This systematic error will then be transferred to the value of voltage that we finally calculate for the voltage source. On the other hand, we might expect a measurement of temperature coefficient, as in figure 4.1, to be less susceptible to the effect of an offset. As may be checked, a constant offset in the temperature values or resistance values will shift the line in figure 4.1 left or right or up or down, but will not affect its slope.

An instrument may have a systematic error other than an offset. An offset, as commonly understood, is an additive (or subtractive) systematic error, as in the case of the DMM that reads 1 μV too high. A systematic error may also be multiplicative. In the case of a DMM, such an error is often called a 'gain error'; for example, the DMM may read three parts per million too low over a particular range of voltages, so that when it displays (for example) 2.000 000 V, the actual value of voltage is 2.000 006 V. In the case of the temperature-coefficient measurement in figure 4.1, such a multiplicative systematic error will affect the slope.

A systematic error may be revealed by one of two general methods. In the following discussion, we use the term 'device' to refer to either an instrument or an artefact. We may look up previously obtained information on the devices used in a measurement. This information may take the form of specifications by a manufacturer or supplier, or look-up tables of physical constants of materials, and previously reported measurements against higher-accuracy devices. We note especially the latter resource: any device, particularly if used in an accurate measurement, should have been calibrated recently. There are laboratories that perform calibrations and issue a *calibration report* for a specified device. The devices of higher accuracy used

in the calibration are themselves calibrated against devices of yet higher accuracy. In this manner, all devices are traceable to the 'top of the food chain' – the international primary standard for the particular quantity. We may call this general class of information 'specific information', since it is specific to the actual measurand that is of immediate concern. Any discrepancy between this specific information and the result of the present measurement suggests that there is a systematic error in the present measurement.

The other method of identifying a systematic error is by changing the experimental set-up. The change may be intentional in order to seek out any systematic error, or may occur for other reasons, with the systematic error being discovered 'by accident' as a result of the change. The change may also take place as a slow natural process, generating an increasing and significant systematic error, which, however, remains unsuspected for a prolonged period. In high-accuracy electrical measurements, the slow deterioration in the insulating property of materials, permitting increasing leakage currents, is such a process.

Here are four examples of intentional change that may uncover a systematic error.

1. In high-accuracy electrical measurements of voltage, swapping the electrical leads connecting a source of constant voltage to a high-accuracy DMM can reveal the systematic errors arising from the DMM's 'zero-offset' and from small thermal voltages caused by the Seebeck effect. The zero-offset error is a non-zero DMM reading when it should be exactly zero (as when a short-circuiting wire is connected across the input terminals), and is due to imperfections in the DMM's internal electronics. The Seebeck effect creates small voltages at junctions between different metals at different temperatures.[3]
2. Exchanging one instrument for another that is capable of the same accuracy and preferably made by a different manufacturer.
3. Having a different person perform the measurement. Thus the exact position of a marker on a scale or of a pointer on a dial will be read differently by different people (a case of so-called 'parallax' error, caused by differences in the positioning of the eye relative to an observed object). In high-accuracy length measurements, using gauge blocks of standard thicknesses, the blocks must often be wrung together to form a stack, and the wringing process, which will determine the overall length of the stack, varies with the operator.
4. An established method of measurement and a novel method that promises higher accuracy may give discrepant results, which will be interpreted as revealing a systematic error in the older method.

An example of such a novel method occurred in high-accuracy measurements of voltage in the early 1970s. Until then, a standard of voltage was provided by banks of

[3] These systematic errors are usually no larger than several microvolts when copper wiring is used for the electrical connections. In section 6.2 these errors are discussed in greater detail.

standard cells. These are electrochemical devices containing mercury and cadmium and their sulfates in sulfuric acid, which were developed by Edward Weston in the 1890s, and provide a stable voltage of about 1.018 V at room temperatures (Vinal 1950). However, in 1962 Brian Josephson predicted an effect in superconductors that radically changed the situation.[4] The prediction was that a constant ('direct-current' or 'dc') voltage, V, would exist across a very narrow gap of the order of nanometres (a 'Josephson junction') between two superconductors,[5] if the gap was irradiated by microwaves at frequency f. Crucially for electrical metrology, the relation $V = n[h/(2e)]f$ would be obeyed. The Planck constant, h, and the electron charge, e, are constants of nature, n is a known integer selected by the experimenter, and a frequency, f, can be measured with extremely high accuracy. So this was potentially, and turned out to be in practice, a much superior method of maintaining a standard of voltage compared with the use of standard cells. In many countries the as-maintained unit of voltage was changed as a result of the new method; in the case of Australia, this amounted to a change of about half a part per million introduced in January 1973. Later, in January 1990, all countries that based their voltage standards on the Josephson effect made a further and larger change of about eight parts per million, as a result of absolute measurements of voltage.[6]

In this fourth category we may also include measurements of fundamental constants where there can be no established method and where, because the measurand is a fundamental constant, any variation in results is attributed to experimental error with a strong systematic component. The extremely challenging measurements of the gravitational constant, G, constitute a prime example. Figure 4.2 summarises measurements (Quinn *et al*. 2001) made between 1997 and 2001 of G, which has an approximate value $6.68 \times 10^{-11}\ \mathrm{m^3 \cdot kg^{-1} \cdot s^{-2}}$. The horizontal 'error-bars' in figure 4.2, some of which do not overlap, indicate the difficulty of assigning an uncertainty to the measured value of G.

Random and systematic errors have contrasting natures. Random errors can be revealed when we repeat the measurement while trying to keep the conditions constant. Systematic errors can be revealed when we *vary* the conditions, whether

[4] Josephson's paper with this discovery is cited, and practical voltage standards based on the Josephson effect are described, in the paper by Hamilton *et al*. cited at the end of chapter 1.

[5] As used in voltage standards, these superconductors are metals (for example, niobium) cooled to temperatures near absolute zero. At low temperature these metals have zero electrical resistance and are therefore known as superconductors.

[6] 'Absolute' electrical measurements, which are invariably complex and demand major laboratory resources, are those made by direct reference to the 'mechanical' standards of mass, length and time. The mercury-electrometer project was such a measurement. It involved defining a voltage through the measurement of the small elevation (a fraction of a millimetre) of a liquid-mercury surface when attracted upwards by a high voltage. The density of mercury and the acceleration due to gravity needed to be accurately known, and a major engineering feat in this experiment was the successful isolation of the system from mechanical vibration (Clothier *et al*. 1989).

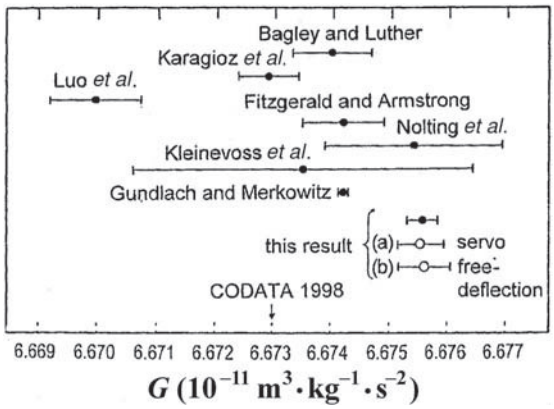

Figure 4.2. Measurements of G by various groups between 1997 and 2001 (courtesy T. J. Quinn, National Physical Laboratory, UK).

deliberately or unintentionally. Varying the conditions can be done in a relatively minor way, as in the lead-swapping example above, or it may amount to a major change in the experimental method and system. In general, the bigger the change, the greater the chance of uncovering systematic errors. Deliberately varying the conditions is more troublesome and time-consuming than simply repeating a measurement; this is one reason why systematic errors can remain hidden and unsuspected for prolonged periods.

We see that, even though the existence of *standards of measurement* is fundamental to metrology, *diversity of methods and procedures* is a powerful defence against systematic errors. Indeed, the richness of metrology derives in part from the continuing interplay of these two apparently discordant principles.

Both methods of revealing systematic errors – specific information and changes to the experimental set-up – require a good grasp of the science underlying the measurement. Since any attempt at accurate measurement is potentially or actually beset by systematic errors from many sources – awareness of this is part of the mental atmosphere of metrology – it is useful to have some familiarity with scientific areas apart from the area of immediate relevance to the measurement. As an example, the elaborate experiment mentioned above to measure the 'absolute volt', using a carefully designed mercury electrometer, demanded expertise not only in electricity and magnetism, but also in optics, the physics and chemistry of liquids, metallurgy and other disciplines.

No sharp distinction is to be made between the two ways in which systematic errors are revealed. Specific information can be obtained from the calibration report on an instrument, and the procedure of calibration itself involves a change in the experimental set-up. Nevertheless, the two-way classification serves as a useful reminder of the practical methods by which constant vigilance against systematic errors can be maintained.

After the existence and cause of a systematic error have been established, an experimental routine can often be developed that automatically takes it into account and eliminates it from the final result. Since the magnitude of the systematic error cannot be known exactly, this process of elimination must itself leave an uncertainty. An example of such an experimental routine, which readily lends itself to statistical analysis, concerns the systematic error caused by DMM offset and by a thermal voltage, as mentioned above and as will be described more fully in chapter 6. In other cases, we have to remove the systematic error through an actual calculation that *corrects* for it. For example, if scales consistently overestimate a nominal 1-kilogram mass by 9 grams, then a value of, say, 989 grams should be corrected to 980 grams.

The correction itself is likely to be known only approximately. Errors must therefore be associated with the correction, and we can regard them as random errors scattered around the correction. We note that looking up specific information is hardly a usefully repeatable exercise, and it is generally impracticable to vary the experimental conditions more often than, say, twice. However, just as in the previous case of usefully repeatable measurements with their 'visible' or *explicit* scatter, the *uncertainty* of the correction can be estimated as representing notionally the *implicit* scatter of its associated random errors. So, whether or not we have usefully repeatable measurements, the measurand is measured with an uncertainty that is described as follows.

4.2 Uncertainty is a parameter that characterises the dispersion of values

The dispersion of data is characterised numerically by a standard deviation (defined in section 4.3). From this standard deviation, it is common practice to obtain a '$\pm$' figure. This figure describes the range of values that is very likely to enclose the true value of the measurand. The number following the '$\pm$' is normally about twice the standard deviation of the measurand and can be loosely referred to as the 'uncertainty' attaching to the measurand. As will be discussed in chapter 10, this uncertainty is referred to in the GUM as the 'expanded' uncertainty, expressing the 'expansion' by that factor of about two from the standard deviation of the measurand.

If a value of a mass is given as (1.24 ± 0.13) kg, the actual value is asserted as very likely to be somewhere between 1.11 kg and 1.37 kg. The uncertainty is 0.13 kg and we note that uncertainty, like standard deviation, is a positive quantity. By contrast, an error may be positive or negative.

4.2.1 Type A and Type B categories of uncertainty

These do not differ in essence, but are given these names in order to convey the notion that *they are evaluated in different ways*.

4.2.1.1 Type A uncertainties are evaluated by statistical methods

In a common situation, a sequence of repeated measurements giving slightly different values (because of random errors) is analysed by calculating the mean and then considering individual differences from this mean. The scatter of these individual differences is a rough indication of the uncertainty of the measurement: the greater the scatter, the more uncertain the measurement.

The calculation of the mean, by summing the values and then dividing this sum by the number of values, is perhaps the simplest example of statistical analysis. The scatter around the mean contributes a Type A uncertainty to the uncertainty of the mean. In a more complicated example requiring statistical analysis, a quantity may change with time, so that the rate of change, commonly called 'drift', is of interest. Often this drift is partially or almost completely obscured by random scatter, as in the case of measurements of climate where a long-term change in temperature may be masked by day-to-day fluctuations. To tease out the value of temperature drift from this background 'noise' is a matter for statistical analysis. The numerical value of drift then has an uncertainty determined by the amount of scatter. This again will be a Type A uncertainty. In exactly the same way, the value of the temperature coefficient in figure 4.1 has an uncertainty determined by the scatter (about the best-fit line) of the 18 measurement points.

4.2.1.2 Type B uncertainties are evaluated by non-statistical methods

A Type B uncertainty may be determined by looking up specific information about a measurand such as that found in a calibration report or data book. When the specific information consists of the calibration report on a device, the value of the measurand is stated in the report – this is the 'calibrated value'. The calibration report also includes the estimated uncertainty in the value of the measurand. The calibrated value can tell us how much systematic error would exist if we ignored the calibration report, and obtaining this information is the primary purpose of calibrating a device. The uncertainty of the calibrated value is always Type B, from our point of view as reader and user of the report. The reason is that no statistical analysis can or needs to be done when reading the report; unlike in the typical case of Type A uncertainty discussed in section 4.2.1.1, reading the report several times will give exactly the same result!

The values summarised in the report were presumably obtained from repeated measurements with an associated Type A uncertainty. The calibration is likely to have entailed repeated measurements in order to cancel out as much as possible any random fluctuations and to check the stability of the instrument or artefact. In the calibration report the measurements are summarised and the uncertainty of the result is estimated using statistical methods. This uncertainty will therefore have a Type A component. Suppose that – unrealistically but as an illustration –

the measurements made by the calibrating laboratory have no Type B component of uncertainty. The uncertainty, which was wholly Type A as determined by the calibrating laboratory, is a Type B uncertainty from the point of view of the reader of the report. The act of writing the report 'fossilises' a Type A uncertainty into a Type B uncertainty. If the reader of the report now uses the reported value in some subsequent application of the instrument or artefact, the uncertainty that was stated in the report is Type B.

When the specific information consists of manufacturer's specifications, the contents of tables of physical properties or the like, it often happens that no associated uncertainties are stated. The information provided by these sources will remove the systematic error that would be present if we used only an approximate value. However, we then have to estimate the associated uncertainty ourselves, without benefit of either statistical analysis or a reported uncertainty. As discussed in section 2.3, this Type B uncertainty can often be estimated from the stated number of decimal places.

4.2.2 Combining Type A and Type B uncertainties

Specific information or changing the conditions of an experiment, whether deliberately or accidentally, may reveal an unsuspected systematic error. This error must itself have an associated uncertainty. After the error has been corrected for, this uncertainty may be Type A or Type B and is then combined with the Type A uncertainty evaluated from random errors. Depending on the particular circumstances, both the Type A and the Type B uncertainties may or may not be reported separately. However, what is always reported is the uncertainty formed from the combination of the Type A and Type B components. From the point of view of the user of the report, this combined uncertainty is wholly Type B.

Figure 4.3 illustrates the relationships among the errors and uncertainties.

4.3 Standard deviation as a basic measure of uncertainty

If there are n values of a quantity, $x_1, x_2, \ldots, x_n$, the standard deviation, s, of these n values is given by[7]

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}, \tag{4.1}$$

where $\bar{x}$ is the mean of the n measurements, defined as $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$.

[7] Strictly this is an approximate estimate of the standard deviation. This is considered in more detail in sections 5.1.3 and 5.1.4.

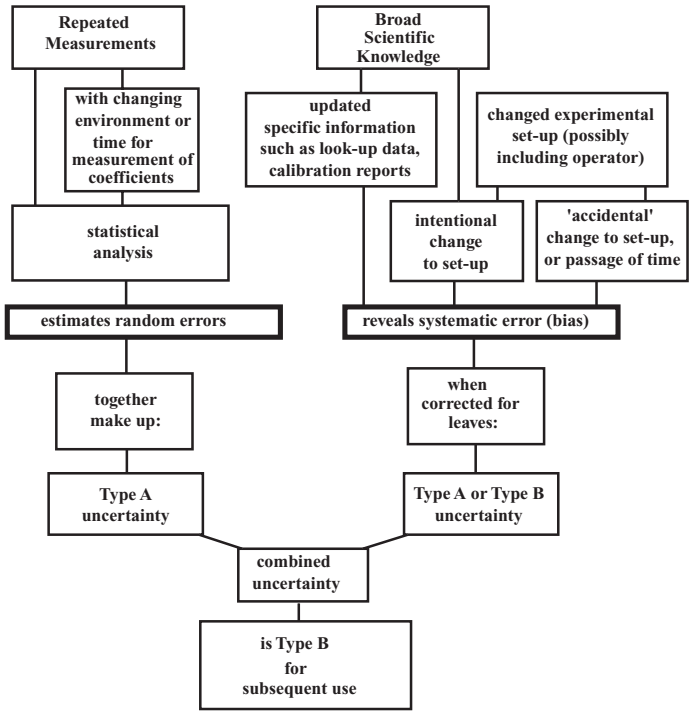

Figure 4.3. The relationship between Type A and Type B uncertainties.

The value of s in the example of small voltage differences in table 4.1 is about 0.026 μV. We note that this is substantially less than the overall range (0.090 μV) by a factor of roughly 3.5. The standard deviation is often less than the overall range of the values (or of the random errors) by a factor between 3 and 4.

The square of s, s^2, is known as the *unbiased variance* of the x_i ($i = 1, 2, \ldots, n$), or more exactly the unbiased estimate of the variance of the entire population of the x's of which our n values form a sample. The variance s^2 of the population is, then,[8]

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}. \tag{4.2}$$

The spread of values is a source of uncertainty in the final result. Since the standard deviation is a measure of the spread, the name given in metrology to the standard deviation is 'standard uncertainty'. The symbol frequently used for standard uncertainty is a lower-case u, so that $u(x)$ is the standard uncertainty of a quantity x. Similarly, $u^2(x)$ denotes the variance of x.

[8] Section 5.1.3 discusses the reason for the presence of $n - 1$ rather than n in the denominator of equation (4.2).

Table 4.2. *Number of airborne particles in a fixed volume in a cleanroom*

Number of particles
137
114
88
102
95
102

Example 1

Six successive measurements of the number of airborne particles within a fixed volume of air within a clean room are made. Table 4.2 show the values obtained. Use these data to calculate (a) the variance and (b) the standard uncertainty in the number of particles.

Answer

(a) Calculation of the variance using equation (4.2) is best accomplished using an electronic calculator or a spreadsheet package, such as Excel. Such a calculation gives $s^2 = 300.3 = u^2(x)$, where x represents the number of particles.

(b) Since $u^2(x) = 300.3$, the standard uncertainty in the number of particles is $u(x) = \sqrt{300.3} = 17.3$.

Exercise A

(1) Ten samples of an oxide of nominally the same mass are heated in an oxygen-rich atmosphere for 1 hour. The mass of each sample increases by an amount shown in table 4.3. Using the data in table 4.3, calculate the variance and the standard uncertainty of the mass gain.

(2) The thickness of an aluminium film deposited onto a glass slide is measured using a profilometer. The values obtained from six replicate measurements are shown in table 4.4. Using these data, calculate the variance and standard uncertainty in the film thickness.

The standard uncertainty, $u(\bar{x})$, of the mean $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$ may be expected to be less than s. This is correct if the values x_i $(i = 1, 2, \ldots, n)$ are uncorrelated.[9] If they are uncorrelated, then

$$u(\bar{x}) = s/\sqrt{n} \qquad \text{for uncorrelated measurements.} \tag{4.3}$$

[9] This is discussed more fully in chapters 5 and 7.

Table 4.3. *Mass gained by samples of oxide*

Mass gain (mg)
12.5
11.2
11.8
11.8
12.1
11.5
11.0
12.1
11.7
12.8

Table 4.4. *Thickness of aluminium film*

Thickness (nm)
420
460
400
390
410
460

In the case of the voltage differences in table 4.1, where $s = 0.026$ μV, the overall correlation among the ten measurements may be shown to be low and the value of $u(\bar{x})$ is $0.026/\sqrt{10}$ μV ~ 0.008 μV.

$u(\bar{x})$ is sometimes called the experimental standard deviation of the mean (ESDM). (In some books this is referred to as the 'standard error' of the mean.) The ESDM when obtained using equation (4.3), with a divisor $\sqrt{n}$, should be used with caution. Equation (4.3) is valid only for uncorrelated values; if, for example, the values exhibit a steady drift in time, then this high correlation implies that the ESDM is not significantly less than s and in fact is closely equal to it. This topic will be discussed further in section 7.2.2.

Example 2

Calculate the mean, $\bar{x}$, and standard uncertainty in the mean, $u(\bar{x})$, for the values in table 4.2.

Answer

The mean of the values of the number of airborne particles in table 4.2 is $\bar{x} = 106.3$. $u(\bar{x})$ is given by equation (4.3). Here, $s = 17.3$ and the number of values is $n = 6$, so $u(\bar{x}) = 17.3/\sqrt{6} \simeq 7.1$.

Exercise B

(a) Using the data in table 4.3, calculate the mean mass gain and standard uncertainty in the mean.

(b) Using the data in table 4.4, calculate the mean thickness of the aluminium film and the standard uncertainty in the mean.

We can see intuitively why, when the values are uncorrelated, the standard uncertainty, $u(\bar{x})$, of the mean is less than the standard deviation, s, of the scatter. When the mean is calculated, the random errors tend to cancel out. This follows from the fact that the measured values are summed for calculating the mean, and the random errors come with both positive and negative signs.[10] However, the cancellation is itself an uncertain process; this is why the reduction from s to $u(\bar{x})$ is not by a factor of n, but only by a factor of $\sqrt{n}$. So, if we go to the trouble of taking not 10 but 100 uncorrelated measurements, the standard uncertainty of the mean will be reduced further only by a factor of about three.

4.4 The uncertainty in the estimate of uncertainty

If the standard uncertainty is denoted by s and its own uncertainty by $u(s)$, then[11]

$$u(s) \sim \frac{s}{\sqrt{2\nu}}. \tag{4.4}$$

where ν is the number of 'degrees of freedom'.[12] ν is equal to the number of values, n, minus the number of quantities determined using the values. In the case where the mean is the only quantity determined using the values, $\nu = n - 1$.

Expressed as a percentage uncertainty, we can write equation (4.4) as

$$\frac{u(s)}{s} \times 100\% \sim \frac{1}{\sqrt{2\nu}} \times 100\%. \tag{4.5}$$

If ν is as low as 4, then $u(s)/s$ is high at about 35%, and ν has to reach 50 to give a 10% uncertainty in the uncertainty.

Equation (4.5) is particularly useful for the Type B category of uncertainty. Owing to the tentative nature of the estimation of Type B uncertainties, it is good

[10] As shown in the voltage-measurement example in section 4.1.2.
[11] Equation (4.4) is discussed further in section 9.3.1.
[12] We will consider degrees of freedom more fully in section 5.1.5.

Table 4.5. *Minimum force required to move a glass block*

Force (N)
5.6
5.7
5.2
5.5
5.8
5.7
5.4

to have some numerical indicator of the reliability which we think attaches to a Type B estimate. Often the value of ν for a Type B uncertainty will not exceed 10, implying a reliability in the estimated uncertainty of no better than about 20%.

Equation (4.5) may be used to determine the uncertainty in a Type A uncertainty. For example, for the data in table 4.1, $s = 0.026$ μV, and $\nu = n - 1 = 9$. On substituting these values into equation (4.5) we find that the percentage uncertainty in s is surprisingly high at almost 25%. The percentage uncertainty in the standard uncertainty in the mean, $u(\bar{x})$, is also 25%. Since $u(\bar{x}) = 0.008$ μV, the uncertainty in $u(\bar{x})$ is about 0.002 μV.

Exercise C

Table 4.5 shows repeat measurements of the minimum force required to cause a glass block to move when it is resting on a smooth metal plate. Using these data, determine

(a) the mean minimum force to move the glass block,
(b) the standard uncertainty in the mean and
(c) the uncertainty in the standard uncertainty.

4.5 Combining standard uncertainties

A measurand may be measured indirectly, through the measurement of so-called 'input quantities'. If y is the measurand and $x_1, x_2, \ldots, x_n$ are the input quantities, then y is a function $y = f(x_1, x_2, \ldots, x_n)$ of the x's. The standard uncertainty, $u(y)$, in y resulting from standard uncertainties $u(x_1), u(x_2), \ldots, u(x_n)$ in the input quantities is calculated using the equation

$$u^2(y) = \left(\frac{\partial y}{\partial x_1}\right)^2 u^2(x_1) + \left(\frac{\partial y}{\partial x_2}\right)^2 u^2(x_2) + \left(\frac{\partial y}{\partial x_3}\right)^2 u^2(x_3) + \cdots + \left(\frac{\partial y}{\partial x_n}\right)^2 u^2(x_n). \tag{4.6}$$

Equation (4.6) may be written more compactly:

$$u^2(y) = \sum_{i=1}^{n} \left(\frac{\partial y}{\partial x_i} \right)^2 u^2(x_i). \tag{4.7}$$

Equation (4.6) is valid *only* if the x_i are mutually uncorrelated. A correlation exists if, for example, two or more of the x_i have been measured using the same instrument that has a systematic error with a significant associated uncertainty. The generalisation of equation (4.6) for the correlated case is discussed in section 7.2.

The standard uncertainties of the inputs, namely the u's on the right-hand side of equation (4.6), may be either Type A or Type B uncertainties. *No distinction between Type A and B is made when evaluating the standard uncertainty of the measurand, y.*

Example 3

The velocity of a wave, v, is written in terms of the frequency, f, and the wavelength, λ, as

$$v = f\lambda. \tag{4.8}$$

An ultrasonic wave has $f = 40.5$ kHz with a standard uncertainty of 0.15 kHz and $\lambda = 0.822$ cm with a standard uncertainty of 0.022 cm. Assuming that there is no correlation between errors in f and λ, calculate the velocity of the wave and its standard uncertainty.

Answer

The velocity $v = f\lambda$ is given by $v = (40.5 \times 10^3) \times (0.822 \times 10^{-2}) = 332.9$ m/s.

Writing equation (4.6) in terms of v, f and λ gives

$$u^2(v) = \left(\frac{\partial v}{\partial f} \right)^2 u^2(f) + \left(\frac{\partial v}{\partial \lambda} \right)^2 u^2(\lambda). \tag{4.9}$$

Using equation (4.8), we have

$$\frac{\partial v}{\partial f} = \lambda, \qquad \frac{\partial v}{\partial \lambda} = f. \tag{4.10}$$

Substituting in values gives

$$u^2(v) = (0.822 \times 10^{-2})^2 \times (0.15 \times 10^3)^2 + (40.5 \times 10^3)^2 \times (0.022 \times 10^{-2})^2 \,(\text{m/s})^2, \tag{4.11}$$

hence $u^2(v) = 80.9$ (m/s)^2, so that $u(v) = 9.0$ m/s.

Exercise D

(1) The flow rate of blood, Q, through an aorta is found to be 81.5 cm^3/s with a standard uncertainty of 1.5 cm^3/s. The cross-sectional area, A, of the aorta is 2.10 cm^2 with a standard uncertainty of 0.10 cm^2. Find the flow speed of the blood, v, and the standard uncertainty in the flow speed using the relationship[13]

$$Q = Av. \tag{4.12}$$

(2) The velocity, v, of a wave on a stretched string is given by

$$v = \sqrt{\frac{F}{\mu}}, \tag{4.13}$$

where F is the tension in the string and μ is the mass per unit length of the string. Given that $F = 18.5\,\text{N}$ with a standard uncertainty of 0.8 N and $\mu = 0.053\,\text{kg/m}$ with a standard uncertainty of 0.007 kg/m, calculate the velocity of the wave and its standard uncertainty.

Historical note. It used to be the common practice, before the introduction of the GUM, for measurement and testing laboratories to report uncertainties as so-called 'errors'. It was also common to report separately the random and systematic errors in the measurand. This often created the complication that, in any subsequent use of the report by others, a single number for the uncertainty, though desirable, was not immediately apparent. There was no consensus regarding the measure of uncertainty: whether this should be the standard deviation or a small multiple of this. Instead of the root-sum-square rule, errors and/or uncertainties were often simply summed linearly. This linear sum applies strictly to perfectly positively correlated input quantities, and if there is little or no correlation the linear sum gives a needlessly pessimistic estimate of the uncertainty in the measurand.

4.6 Review

While errors are conveniently categorised as random or systematic, the GUM introduces the new terms 'Type A' and 'Type B' to categorise uncertainties. Type A and Type B uncertainties are not related directly to random and systematic errors, but reflect the way in which uncertainties are evaluated. In the next chapter we will turn our attention to useful statistical methods that allow us to summarise key features of experimental data.

[13] Equation (4.12) is often referred to as the 'continuity equation'.