Theory of measurement uncertainty

1. Introduction

The complete statement of a measured value should include an estimate of the level of confidence associated with the value. Properly reporting an experimental result along with its uncertainty allows other people to make judgements about the quality of the experiment, and it facilitates meaningful comparisons with other similar values or a theoretical prediction. Without an uncertainty estimate, it is impossible to answer the basic scientific question: "Does my result agree with a theoretical prediction or results from other experiments?" This question is fundamental for deciding if a scientific hypothesis is confirmed or refuted.
When we make a measurement, we generally assume that some exact or **true value** exists based on how we define what is being measured. As we make measurements by different methods, or even when making multiple measurements using the *same* method, we may obtain slightly different results. So how do we report our findings for our best estimate of this elusive *true value*? The most common way to show the range of values that we believe includes the true value is:

experimental result = (best estimate ± uncertainty) [physical unit]

For our case, we should first define the terms **accuracy** and **precision**:

> **Accuracy** is the closeness of agreement between a measured value and a true or accepted value. Measurement error is the amount of inaccuracy.
>
> **Precision** is a measure of how well a result can be determined (without reference to a theoretical or true value). It is the degree of consistency and agreement among independent measurements of the same quantity; also the reliability or reproducibility of the result.

The statement of uncertainty associated with a measurement should include factors that affect both the accuracy and precision of the measurement.

2. Error versus uncertainty

It is important not to confuse the terms ‘error’ and ‘uncertainty’.

> **Error** is the difference between the measured value and the ‘true value’ of the thing being measured.
>
> **Uncertainty** is a quantification of the doubt about the measurement result.

Whenever possible we try to correct for any known errors: for example, by applying corrections from calibration certificates. But any error whose value we do not know is a source of uncertainty.

Absolute Error of a quantity X

An absolute error has the same units as the physical quantity itself. We denote it as κ_x.

Relative Error of a quantity X

We will also encounter relative error, defined as the ratio of the error to the best value of the quantity, so that

$$\kappa_{rx} = \frac{\kappa_x}{x}$$ (Note the additional subscript "*r*" meaning "relative")

Types of Errors

Measurement errors may be classified as either **random** or **systematic**, depending on how the measurement was obtained (an instrument could cause a random error in one situation and a systematic error in another).

> **Random errors** are statistical fluctuations (in either direction) in the measured data due to the precision limitations of the measurement device. Random errors can be evaluated through statistical analysis and can be reduced by averaging over a large number of observations (see standard error).
>
> **Systematic errors** are reproducible inaccuracies that are consistently in the same direction. These errors are difficult to detect and cannot be analysed statistically. If a systematic error is identified when calibrating against a standard, the bias can be reduced by applying a correction or correction factor to compensate for the effect. Unlike random errors, systematic errors cannot be detected or reduced by increasing the number of observations.

3. Common sources of errors in physics laboratory experiments

Incomplete definition (may be systematic or random) - One reason that it is impossible to make exact measurements is that the measurement is not always clearly defined. For example, if two different people measure the length of the same rope, they would probably get different results because each person may stretch the rope with a different tension. The best way to minimize definition errors is to carefully consider and specify the conditions that could affect the measurement.

Failure to account for a factor (usually systematic) – The most challenging part of designing an experiment is trying to control or account for all possible factors except the one independent variable that is being analyzed. For instance, you may inadvertently ignore air resistance when measuring free-fall acceleration, or you may fail to account for the effect of the Earth's magnetic field when measuring the field of a small magnet. The best way to account for these sources of error is to brainstorm with your peers about all the factors that could possibly affect your result. This brainstorm should be done before beginning the experiment so that arrangements can be made to account for the confounding factors before taking data. Sometimes a correction can be applied to a result after taking data to account for an error that was not detected.

Environmental factors (systematic or random) - Be aware of errors introduced by your immediate working environment. You may need to take account for or protect your experiment from vibrations, drafts, changes in temperature, electronic noise or other effects from nearby apparatus.

Instrument resolution (random) - All instruments have finite precision that limits the ability to resolve small measurement differences. For instance, a meter stick cannot distinguish distances to a precision much better than about half of its smallest scale division (0.5 mm in this case). One of the best ways to obtain more precise measurements is to use a null difference method instead of measuring a quantity directly. Null or balance methods involve using instrumentation to measure the difference between two similar quantities, one of which is known very accurately and is adjustable. The adjustable reference quantity is varied until the difference is reduced to zero. The two quantities are then balanced and the magnitude of the unknown quantity can be found by comparison with the reference sample. With this method, problems of source instability are eliminated, and the measuring instrument can be very sensitive and does not even need a scale.

Failure to calibrate or check zero of instrument (systematic) - Whenever possible, the calibration of an instrument should be checked before taking data. If a calibration standard is not available, the accuracy of the instrument should be checked by comparing with another instrument that is at least as precise, or by consulting the technical data provided by the manufacturer. When making a measurement with a micrometer, electronic balance, or an electrical meter, always check the zero reading first. Re-zero the instrument if possible, or measure the displacement of the zero reading from the true zero and correct any measurements accordingly. It is a good idea to check the zero reading throughout the experiment.

Physical variations (random) - It is always wise to obtain multiple measurements over the entire range being investigated. Doing so often reveals variations that might otherwise go undetected. These variations may call for closer examination, or they may be combined to find an average value.

Parallax (systematic or random) - This error can occur whenever there is some distance between the measuring scale and the indicator used to obtain a measurement. If the observer's eye is not squarely aligned with the pointer and scale, the reading may be too high or low (some analog meters have mirrors to help with this alignment).

Instrument drift (systematic) - Most electronic instruments have readings that drift over time. The amount of drift is generally not a concern, but occasionally this source of error can be significant and should be considered.

Lag time and hysteresis (systematic) - Some measuring devices require time to reach equilibrium, and taking a measurement before the instrument is stable will result in a measurement that is generally too low. The most common example is taking temperature readings with a thermometer that has not reached thermal equilibrium with its environment. A similar effect is hysteresis where the instrument readings lag behind and appear to have a "memory" effect as data are taken sequentially moving up or down through a range of values. Hysteresis is most commonly associated with materials that become magnetized when a changing magnetic field is applied.

Personal errors come from carelessness, poor technique, or bias on the part of the experimenter. The experimenter may measure incorrectly, or may use poor technique in taking a measurement, or may introduce a bias into measurements by expecting (and inadvertently forcing) the results to agree with the expected outcome.

Gross personal errors, sometimes called mistakes or blunders, should be avoided and corrected if discovered. As a rule, gross personal errors are excluded from the error analysis discussion because it is generally assumed that the experimental result was obtained by following correct procedures. The term human error should also be avoided in error analysis discussions because it is too general to be useful.

4. Estimating Experimental Error

Single measurements
Any measurement you make will have some error associated with it, no matter how precise your measuring tool is. How do you actually determine the error, and once you know it, how do you report it?

The error of a single measurement is limited by the precision and accuracy of the measuring instrument, along with any other factors that might affect the ability of the experimenter to make the measurement.

Measurement of dimensions
For a dimension measurement, one has to select an appropriate measuring tool. The standard measurement error m_x of basic tools certified by the producers are as follows (the value of the standard error m_x equals to the absolute error κ_x):

Tape measure: $m_x = 1$ mm
Calliper: $m_x = 0.1$ mm
Micrometre: $m_x = 0.01$ mm

Measurement of time

The measurement of time (periods, swings, etc.) is typically carried out using stop-watch. There are two basic measurement error that should be taken into account - the precision of the stop-watch (typically 0.01 s) and the human reaction time (typically 0.5 s) that could easily overlap the precision of the measuring tool.

Measurement of electrical quantities - analogue instruments

The maximum measurement error of an analogue (pointer-type) instrument measuring electrical quantity is typically expressed by so called accuracy classes. The accuracy class T_p means a maximum relative error that could be reached within the selected measuring range $x_{\max}$. Thus, the instrument error m_x could be calculated as follows

$$m_x = \frac{1}{100} T_p x_{\max}$$

Measurement of electrical quantities - digital instruments

The specific approach for error calculation could be found in the user's manual of particular instrument. For instance, the error calculation could be specified as follows:

m_x = 0,01% rdg. + 2 digits

The error consist of a base error expressed as a percentage of the reading (0.01% - see above) and a number of digits. One digit correspond to the instrument resolution - it is the least significant change of the value on the instrument display (last digital place typically). The resolution is dependent on the selected instrument range, thus, the real decimal place of the digit should be specified. For instance, if the reading is 12,778 V, the instrument resolution (one digit) is 0.001 V.

Repeated Measurements

Random error occurs because of small, uncorrelated variations in the measurement process. For example, measuring the period of a pendulum with a stopwatch will give different results in repeated trials for one or more reasons. One reason could be that the watch is defective, and its ticks don't come at regular intervals. Let's assume that you have a “good” stopwatch, and this isn't a problem. A more likely reason would be small differences in your reaction time for hitting the stopwatch button when you start the measurement as the pendulum reaches the end point of its swing and stop the measurement at another end point of the swing. If this error in reaction time is random, the average period over the individual measurements would get closer to the correct value as the number of trials n is increased. This statistical error behaviour is called Gaussian error distribution.

The correct reported result would begin with the average for this best value:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

and it would end with your estimate of the error in this best value. This usually taken as the standard deviation of the measurements. An estimate of the random error for a single measurement x_i is

$$s_x = \sqrt{\frac{\sum_{i=1}^{n} \left(\overline{x} - x_i\right)^2}{n-1}}$$

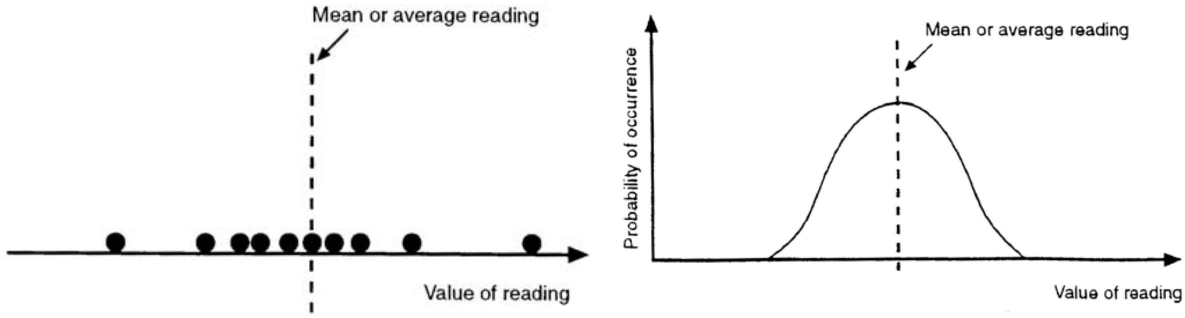
For our purpose, the standard deviation of the average value $\overline{x}$ is commonly used:

$$s_{\overline{x}} = \sqrt{\frac{\sum_{i=1}^{n} \left(\overline{x} - x_i\right)^2}{n\left(n-1\right)}}$$

5. Distribution - the 'shape' of the errors

The spread of a set of values can take different forms, or probability distributions. Typically, there are two most common distributions of errors - normal (Gaussian) distribution and rectangular (uniform) distribution.

Normal distribution
In a set of readings, sometimes the values are more likely to fall near the average than further away. This is typical of a normal or Gaussian distribution. For instance, you might see this type of distribution if you examined the heights of individuals in a large group of men. Most men are close to average height; few are extremely tall or short.
Following figure shows a set of 10 'random' values in an approximately normal distribution. A sketch of a normal distribution is shown as well.

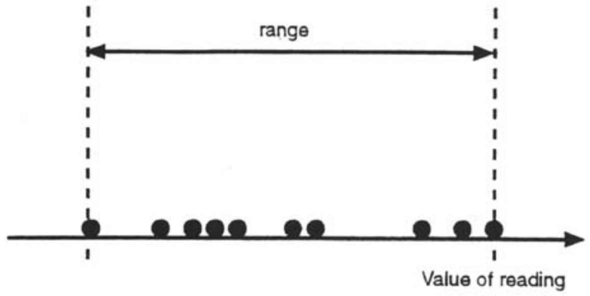


a) A set of values lying in a normal distribution *b) Sketch of a 'normal' distribution*

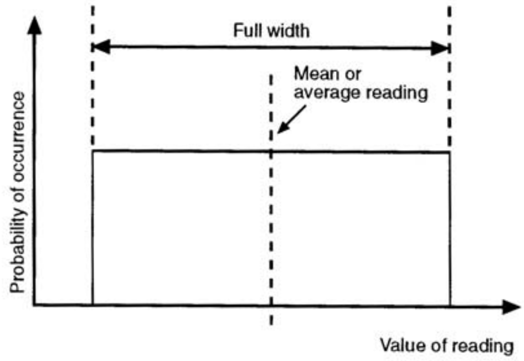
Uniform or rectangular distribution

When the measurements are quite evenly spread between the highest and the lowest values, a rectangular or uniform distribution is produced. This would be seen if you examined how rain drops fall on a thin, straight telephone wire, for example. They would be as likely to fall on any one part as on another.

Following figure shows a set of 10 'random' values in an approximately rectangular distribution. A sketch of a rectangular distribution is shown as well.



a) A set of values lying in a rectangular distribution



b) Sketch of a 'normal' distribution

6. How to calculate uncertainty of measurement

To calculate the uncertainty of a measurement, firstly you must identify the sources of uncertainty in the measurement. Then you must estimate the size of the uncertainty from each source. Finally the individual uncertainties are combined to give an overall figure.

There are clear rules for assessing the contribution from each uncertainty, and for combining these together.

The two ways to estimate uncertainties

No matter what are the sources of your uncertainties, there are two approaches to estimating them: 'Type A' and 'Type B' evaluations. The characters "A" and "B" are used as subscripts of the quantities to show the prevailing type of error source. In most measurement situations, uncertainty evaluations of both types are needed.

> ***Type A evaluations*** - uncertainty estimates using statistics (usually from repeated readings)
>
> ***Type B evaluations*** - uncertainty estimates from any other information. This could be information from past experience of the measurements, from calibration certificates, manufacturer's specifications, from calculations, from published information, and from common sense.

There is a temptation to think of ‘Type A’ as ‘random’ and ‘Type B’ as ‘systematic’, but this is not necessarily true.

Standard uncertainty

All contributing uncertainties should be expressed at the same confidence level, by converting them into standard uncertainties. A standard uncertainty is a margin whose size can be thought of as ‘plus or minus one standard deviation’. The standard uncertainty tells us about the uncertainty of an average (not just about the spread of values). A standard uncertainty is usually shown by the symbol u. Two basic uncertainties could be defined:

Absolute uncertainty of a quantity X

An absolute error has the same units as the physical quantity itself. We denote it as u_x.

Relative uncertainty of a quantity X

We will also encounter relative error, defined as the ratio of the error to the best value of the quantity, so that

$$u_{rx} = \frac{u_x}{X}$$ (Note the additional subscript "r" meaning "relative")

Calculating standard uncertainty for a Type A evaluation

When a set of several repeated readings has been taken (for a Type A estimate of uncertainty), the mean $\overline{x}$ and estimated standard deviation of the average value s_x can be calculated for the set. Note, that the standard deviation of the average represent the Type A evaluation of the uncertainty directly, thus

$$u_{xA} = s_{\overline{x}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Calculating standard uncertainty for a Type B evaluation

Where the information is more scarce (in some Type B estimates), you might only be able to estimate the upper and lower limits of uncertainty. You may then have to assume the value is equally likely to fall anywhere in between, i.e. a rectangular or uniform distribution. The standard uncertainty for a rectangular distribution is found from:

$$u_{zB} = \frac{\Delta z_{\max}}{\chi}$$

where $\Delta Z_{\max}$ is the maximum of all detectable systematic errors in the particular value measurement. In most of the cases, we will use the expression of the standard error m_x for our measurements (see above). The variable χ count the effect of the error distribution towards the calculated uncertainty. For the uniform distribution,

$\chi = \sqrt{3}.$

Rectangular or uniform distributions occur quite commonly, but if you have good reason to expect some other distribution, then you should base your calculation on that. For example, you can usually assume that uncertainties ‘imported’ from the calibration certificate for a measuring instrument are normally distributed.

7. Combining standard uncertainties of A and B Type

Individual standard uncertainties calculated by Type A or Type B evaluations can be combined validly by ‘summation in quadrature’ (also known as ‘root sum of the squares’). The result of individual measurement of the quantity x ("direct" value measurement) is called the *combined standard uncertainty*, shown by

$$u_x = \sqrt{u_{xA}^2 + u_{xB}^2}$$

If the result has to be calculated from specific formula and the estimation requires recording of several sub values, we call this approach as "indirect" measurement. For instance, determination of a rectangle area A requires measurement of two dimensions - sides a and b - and appropriate uncertainties u_a and u_b estimation. The area is calculated by $A = ab$. How to find a correct estimation of the area uncertainty u_A? Use the following procedure.

8. Summation in quadrature for arbitrary functions

Let's consider that the value Y is a arbitrary combination of several sub values X_1, X_2, ... X_n.

$$Y = f(X_1, X_1, \dots . X_n)$$

The law of the uncertainties summation should be used as follows:

$$u_Y = \sqrt{\left(\frac{\partial Y}{\partial X_1}\right)^2 u_{X_1}^2 + \left(\frac{\partial Y}{\partial X_2}\right)^2 u_{X_2}^2 + \cdots + \left(\frac{\partial Y}{\partial X_n}\right)^2 u_{X_n}^2}$$

The standard procedure of the uncertainty u_Y calculation requires the estimation of all sub uncertainties, both A Type and B Type. Thus, a following set of uncertainty values should be available:

$$u_{X_1A},\ u_{X_2A} \ .... \ u_{X_nA};\ \ u_{X_1B},\ u_{X_2B} \ .... \ u_{X_nB}$$

The law of summation in quadrature is valid for both uncertainty A Type and B Type:

$$u_{YA} = \sqrt{\left(\frac{\partial Y}{\partial X_1}\right)^2 u_{X_1A}^2 + \left(\frac{\partial Y}{\partial X_2}\right)^2 u_{X_2A}^2 + \cdots + \left(\frac{\partial Y}{\partial X_n}\right)^2 u_{X_nA}^2}$$

$$u_{YB} = \sqrt{\left(\frac{\partial Y}{\partial X_1}\right)^2 u_{X_1B}^2 + \left(\frac{\partial Y}{\partial X_2}\right)^2 u_{X_2B}^2 + \cdots + \left(\frac{\partial Y}{\partial X_n}\right)^2 u_{X_nB}^2}$$

Finally, the combined standard uncertainty will be estimated as

$$u_Y = \sqrt{u_{YA}^2 + u_{YB}^2}$$

In many cases, the calculation of the quantity Y from two sub values X_1 and X_2 is simple - it is a direct addition, subtraction, multiplication or division. In most of the cases, the partial derivative components lead to very simple expressions and the resulting calculation formula is short. Let's see some examples in the following table (the values *a* and *b* are real constants; *m* and *n* are natural indices).

Operation	Formula	Uncertainty calculation
Addition or subtraction	$Y = X_1 \pm X_2$	$u_Y = \sqrt{u_{X_1}^2 + u_{X_2}^2}$
	$Y = aX_1 \pm bX_2$	$u_Y = \sqrt{a^2u_{X_1}^2 + b^2u_{X_2}^2}$
multiplication or division	$Y = X_1X_2$	$u_{rY} = \sqrt{u_{rX_1}^2 + u_{rX_2}^2}$ (relative !!)
	$Y = aX_1^mX_1^n$	$u_{rY} = \sqrt{m^2u_{rX_1}^2 + n^2u_{rX_2}^2}$ (relative !!)

9. Expression of the result

It is important to express the answer so that a reader can use the information. The most common way of expression requires writing the measurement result together with the uncertainty figure:

$$Y \pm u_Y$$

The uncertainty value should be rounded up in all the cases; the typical number of significant figures is two. The result *Y* should obtain the same number of decimal places as the uncertainty. There are some examples of correct result expressions (check the significant figures, decimal places and expression of the order of magnitude):

$$\lambda = (632.8 \pm 1.3)\text{nm}$$
$$T = (1.235 \pm 0.032)\text{s}$$
$$E = (2.05 \pm 0.12).10^{11}\ \text{Pa}$$

10. References

[1] Stephanie Bell, Measurement Good Practice Guide No. 11 (Issue 2): A Beginner's Guide to Uncertainty of Measurement, 1999, ISSN 1368-6550

[2] Introduction to Measurement Uncertainty and Error Analysis, online: https://www.physics.ryerson.ca/sites/default/files/u23/2017/08/Intro-Measurement-Uncertainty-general.pdf, cited on 01.10.2018

[3] Introduction to Measurements & Error Analysis, online: https://users.physics.unc.edu/~deardorf/uncertainty/UNCguide.html, cited on 01.10.2018

[4] Uncertainty, Error and Graphs, Stony Brook Physics Laboratory Guides, online: http://skipper.physics.sunysb.edu/~physlab/doku.php?id=phy124:error_and_uncertainty, cited on 01.10.2018