DEPARTMENT OF PHYSICS
AND ASTRONOMY

UCL

PHAS1240

Experimental Methods and Data Analysis

Prof N. Skipper and Dr P. Jones

(Revised by Dr P. Bartlett and Dr J Grozier)

(Revised by K. Dunnett)

Contents

Experimental Methods and Data Analysis

These notes are intended to support course PHAS1240: Practical Physics.

Physics or Astronomy experiments should always start with a specified *objective* and proceed via some *method* to make *measurements*, to which are applied some form of *data analysis* in order to arrive at *results*, from which we then draw *conclusions* relevant to the original objective. Our overall aim is that students will develop their practical, observational, investigative and data analysis skills to the level required of a professional physicist.

Assessment will be conducted via marking of your task books and formal reports, and also through two quizzes (Data Retrieval Tests).

Recommended text books

The following books are recommended for use in this course:

- *Measurements and their Uncertainties: A Practical Guide to Modern Error Analysis*, Ifan G. Hughes and Thomas P.A.Hase. Oxford University Press (2010), ISBN 0 191 57656 5

- *Experimental Methods - an introduction to the analysis and presentation of data*, Les Kirkup. John Wiley & Sons, ISBN 0 471 33579 7

- *Practical Physics*, G. L. Squires. Cambridge University Press, ISBN 0 521 77940 5

- *A practical guide to data analysis for physical science students*, Louis Lyons. Cambridge University Press, ISBN 0 521 46463 1

1 Experimental Uncertainties

1.1 What is an experimental uncertainty?

When performing experiments or making measurements we do not expect to measure the value of a quantity *exactly.* To do so would be unreasonably time-consuming or expensive, or even limited by fundamental physics. Instead we arrive at a number for our measurement, and another number that gives an indication of the probability a subsequent measurement will fall in a certain range around the measurement. This number is the **experimental uncertainty** It is sometimes referred to in older texts as the **experimental error** but it is better to express it as the *uncertainty* in the result - a limitation to the knowledge we have of the measured value. Some uses of the word 'error' have persisted, however, particularly in references to the *standard error* and the *error function*, the meaning of which will be explained below.

As an example, you may measure the acceleration due to the earth's gravity to be

$$g = (9.8 \pm 0.2)\ \mathrm{ms}^{-2},$$

that is, a measurement of $g = 9.8$ ms^{-2} with an accompanying associated experimental uncertainty of $\Delta g = \pm 0.2$ ms^{-2}. Writing it in this way tells us not that any subsequent measurements will all be in the range 9.6 ms^{-2} $< g <$ 10.0 ms^{-2}, but that we can expect a certain fraction (and we will see later what that fraction is) to do so.

The experimental uncertainty is just as important as the measured value as it gives us an idea of the *reliability* of the measurement that has been made. It also enables us to decide whether differences between measurements are significant, for instance when measured at the equator $g = 9.78$ ms^{-2}, compared to when measured in London $g = 9.81$ ms^{-2}. Is this difference significant? It depends on the magnitude of the experimental uncertainty in each case. If, say, $\Delta g = \pm 0.5$ ms^{-2} then the range of values around each measurement overlap somewhat, and it would be *probable to within experimental uncertainty* that the two measurements agreed. If the uncertainty were $\Delta g = \pm 0.1$ ms^{-2} then the ranges would not overlap, and we would be more confident in saying that the measurements were genuinely different.

For this reason **all** measurements made in the laboratory **must** be accompanied by the experimental uncertainty. **There are no exceptions to this rule**.

But where does the experimental uncertainty come from? We will divide sources of uncertainty into two kinds: **systematic** uncertainties and **random** uncertainties.

1.1.1 Systematic uncertainties

A systematic uncertainty is one that is constant throughout a set of readings. It may arise from, for example, poor calibration of the apparatus, or an incorrect assumption into the supporting theory. Making repeated measurements will not eliminate the systematic uncertainty as all measurements are affected in the same way.

A simple example of a systematic uncertainty would be trying to measure a length using a rule which, unnoticed by you, had the end missing and so started at 2 cm. All your measurements would be affected in the same way by this and so your end result would suffer from a systematic uncertainty. Your task would then be to discover the source and eliminate it.

1.1.2 Random uncertainties

Random, or statistical, uncertainties are those which vary between measurements, equally likely to increase or decrease the quantity being measured. At first sight these would seem harder to deal with than systematic uncertainties - how can we account for a quantity that fluctuates randomly in successive measurements? Fortunately there exist statistical methods for dealing with these fluctuations and the resulting *distribution* of values around the average. We will examine these methods in more detail when we have looked at the mathematics needed to describe these distributions. Obviously a small spread in measurements about the average is desirable.

1.2 Accuracy and precision

The aim of the experimental measurement is ultimately to achieve a result that is both *accurate* and *precise*. In the context of experimental uncertainties these concepts have different meanings, and can be related to the amount of systematic or random uncertainty present in the result.

An accurate result is one that is free from systematic uncertainty. A precise result is one for which random uncertainties are small. It is possible for a result to be accurate, or precise, both, or neither!

As an example, consider the defective rule that we discussed earlier. If this had millimetre markings it would be possible to use this to obtain a very precise measurement of a length, that is all the measurements would be tightly clustered around an average value. However, this measurement would still be *inaccurate* as it would differ from the true measurement by 2 cm (much greater that the spread of measurements about the average).

Now imagine trying to time the period of one oscillation of a pendulum using a hand-held stopwatch. If all the students taking this course tried this, each timing one oscillation only, then we would expect to obtain a broad spread in the measurements as each student will have slightly different reaction times and may start or stop the stopwatch too early or too late. The average value may well be accurate, in that it could be close to the true value, but the measurement is likely to be *imprecise* because of the large spread in the data.

Of course, we know that it is better to time several periods of oscillation to reduce the random uncertainty effects of reaction times by averaging, and in this way achieve a measurement that is both accurate **and** precise.

1.3 Quoting uncertainties

It is very tempting when using an electronic calculator to write down all the figures that appear on the screen, leaving a final answer with up to 10 significant figures. Before doing this **stop** and **think**. By doing this you are claiming that all of these figures are important, and known reliably enough to record for someone else to use. But we have just seen that any set of measurements will be distributed with some spread about an average value. This means that, depending on the spread, not all of the figures will be reliable enough to write down. How, then, should we decide which to include?

A good general rule is that your experimental uncertainty should be quoted to one significant figure only - usually the estimations made in arriving at the experimental uncertainty will preclude any more. The measured value you are reporting can then only be written to the **same** precision, as any further figures involve those deemed to be unreliable.

Consider again the acceleration due to the earth's gravity. Above we quoted the experimental uncertainty on this to be ± 0.2 ms^{-2}, and the value was therefore written to the same precision: one decimal place only, 9.8 ± 0.2 ms^{-2}.

In 2001, Nobel Prize winner Steve Chu's group at Stanford University[1] measured an average value for the acceleration due to gravity of $g = 9.799\,331\,58$ ms^{-2}. Can all those decimal places *really* be justified? If the experiment were simply measuring the oscillations of a pendulum that would be extremely hard! However the actual experiment is a rather more sophisticated *atom interferometer* using ultra-cold caesium atoms, and considerable effort was made to identify and account for all sources of systematic uncertainty (including the change in g produced by the presence of the student doing the experiment in the same room), and reduce the random uncertainty by a long period of averaging. The final reported measurement was then

$$g = (9.799\,331\,58 \pm 0.000\,000\,03)\ \text{ms}^{-2}.$$

Note that the uncertainty has one significant figure, and the value is quoted to the same precision.

All experimental uncertainties must be reported in the form (value ± uncertainty)unit. The value and uncertainty should have the same unit (obviously!) with the same prefix, that is for example (22.06 ± 0.04) kg, **not** 22.06 kg ± 40 g.

Frequently you will want to compare your measured value with another source. In this case **do not** give the difference as a simple percentage, as this gives no information about the reliability of your measurement. Instead you should quote the difference as a multiple of your experimental uncertainty. As we shall see later, this gives a criterion as to whether the agreement can be considered 'good' or not.

1.4 Units and dimensions

All measured quantities **must** be reported with their associated unit. **There are no exceptions**.

A measurement is considered to be a product of two parts: a *number* and a *unit*. In the SI system there are seven *base units* for seven *base quantities* that are mutually independent. These are shown in table 1.

Quantity	Unit	Symbol	Dimension
length	metre	m	L
mass	kilogram	kg	M
time	second	s	T
electric current	ampere	A	I
thermodynamic temperature	kelvin	K	Θ
amount of substance	mole	mol	N
luminous intensity	candela	cd	J

Table 1: The SI base quantities and base units

Other units are derived from these, such as the unit for acceleration ms^{-2} we used above, and some of these have special names, such as the Newton for force, or the Joule for energy.

[1] A. Peters, K. Y. Chung & S. Chu, *Metrologia* **38** 25 (2001)

Furthermore each of these units has a dimension which tells us what 'sort' of quantity - a length (L), a time (T) etc - is being measured. The derived units also have dimensions that are combinations of the base unit dimensions. When quantites are multiplied the resulting dimension is the product of the individual dimensions, and similarly for division (the quotient). For example a velocity is a distance per unit time, and so has dimensions LT^{-1}. Similarly a force (measured in Newtons) is the product of a mass and an acceleration, so has dimensions MLT^{-2}. Obviously you can not add together units with different dimensions, just as you would never try to add a time to a distance - they are different sorts of quantity.

Note that it is possible to have a unit that is dimensionless. A common example of this is the unit of angle the radian. As the size of an angle in radians is the *length* of the arc subtending the angle divided by the *length* of the radius, the two dimensions cancel, leaving the final unit with no dimension.

You can use the dimensions of the quantities in an equation to check for uncertainties should you get an answer that you suspect is wrong. You can only add (or subtract) quantities with the same dimension, and the dimensions should match on each side of the equals sign.

The convention "quantity = number × unit" gives rise to the correct way to display, for instance, the axis labels on a graph or the heading of a table of data. The quantity you are measuring - say it's a length - is given a symbol, l which represents the number×unit product. However, what is recorded in a table or plotted on a graph is **just** the numerical part. This number is therefore your measurement with the unit divided out, i.e. quantity/unit. The correct way to label the graph axis or table column is therefore:

$$\text{length, } l/\mathrm{m}$$

which can be read as "length, l in metres".

2 Combination and propagation of uncertainties

When performing an experiment we will usually take direct measurements of several quantities A, B, C etc, each of which will have its associated experimental uncertainty $\Delta A, \Delta B, \Delta C$ etc. We will then often have to manipulate these measurements to obtain an indirect measure of another quantity, call it Z, which is therefore a function of A, B and C, i.e. $Z = Z(A, B, C)$. How do we go about finding the uncertainty on Z given A, B, C and $\Delta A, \Delta B, \Delta C$?

2.1 Functions of a single variable

Let us start with a simplified example where Z is a function of one variable only, i.e. $Z = Z(A)$. We want to find ΔZ, that is, by how much does Z vary given the relationship between Z and A and the amount that A can fluctuate? Look at figure 1 for a graphical interpretation.

If ΔA, the change in A about the value A_0, is not too large then Z varies linearly with

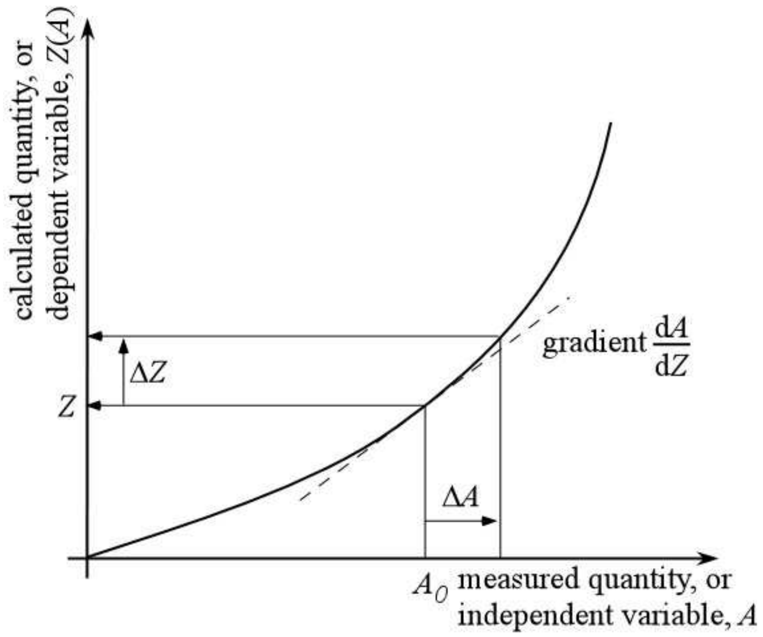

Figure 1: This figure shows how a change in the measured quantity A affects the derived quantity Z.

A and we can say that the ratio:

$$\frac{\Delta Z}{\Delta A} \rightarrow \left(\frac{dZ}{dA}\right)_{A\rightarrow A_0}$$

i.e. the gradient of the function $Z(A)$ evaluated at $A = A_0$. And so for a function of a single variable we have that

$$\Delta Z \approx \left(\frac{dZ}{dA}\right)_{A=A_0} \Delta A. \tag{1}$$

If ΔA is large, then this derivative approach is clearly inaccurate. In this case, the functional approach has to be used:

$$\Delta Z = |Z(A + \Delta A) - Z(A)| \tag{2}$$

2.2 Functions of more than one variable

But what of a function of several (independent) variables, e.g. $Z = Z(A, B)$ with known uncertainties $\Delta A, \Delta B$? There are two approaches that can be used, the first is valid for any function (known or unknown) and any size of uncertainty, the second is useful if the function and its derivative are known and the uncertainties are small.

A function Z of two variables, $Z(A, B)$ can be represented on a three-dimensional graph as the height above the $AB-$plane as shown in figure 2 for the function $Z = \sin(A)\cos(B)$. Obviously this function has gradients - the value of $Z(A, B)$ changes as we change A or B, and the gradients are different depending on whether we change A or B.

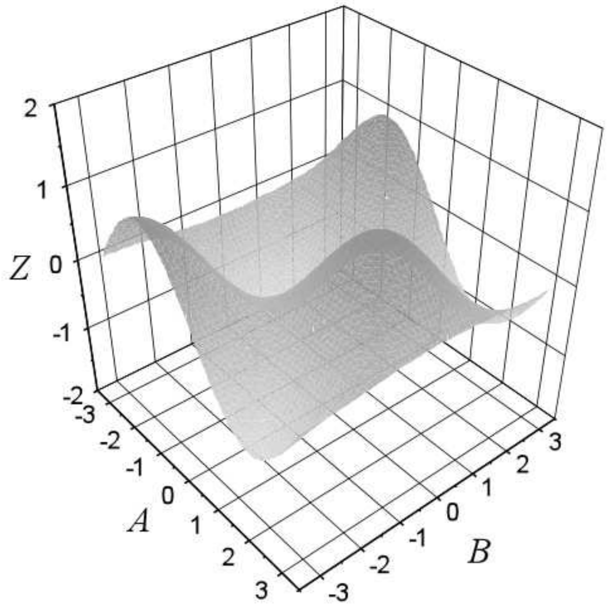

Figure 2: The function $Z = \sin(A)\cos(B)$ represented as the height of the surface above the AB plane.

2.2.1 Functional approach

If A is varied and B kept constant, then the change in Z is given by:

$$\Delta Z_A = Z(A + \Delta A, B) - Z(A, B)$$

and similarly for varying B while keeping A fixed.

If the uncertainties are uncorrelated and the variables are independent, the total uncertainty in Z can be calculated using Phythagoras' thorem (in generally N dimensions for N independent variables and associated uncertainties). Then, for a function of two variables

$$(\Delta Z)^2 = (\Delta Z_A)^2 + (\Delta Z_B)^2.$$

In its most general form:

$$\begin{aligned}(\Delta Z)^2 \quad = \quad & [Z(A + \Delta A, B, C, \ldots) - Z(A, B, C, \ldots)]^2 \\ & +[Z(A, B + \Delta B, C, \ldots) - Z(A, B, C, \ldots)]^2 \\ & +[Z(A, B, C + \Delta C, \ldots) - Z(A, B, C, \ldots)]^2 \\ & + \ldots \end{aligned} \tag{3}$$

If we take the example plotted in figure 2:

$$Z = \sin(A)\cos(B), \tag{4}$$

we find:

$$(\Delta Z)^2 = (\sin(A+\Delta A)\cos(B))^2 + (\sin(A)\cos(B+\Delta B))^2 \tag{5}$$

which can be easily entered into a computer.

(**HINT:** Always write out the version of eq. (5) relevant for your particular analysis by hand before attempting to enter it into a computer, you will spot your own mistakes.)

2.2.2 Calculus approximation

If all the uncertainties on A and B are small, then a calculus approximation, requiring partial differentiation, can be used.

So if we move along the surface $Z = \sin(A)\cos(B)$ along a line parallel to the A-axis we can calculate the gradient in this direction by differentiating with respect to A while treating B as though it were a constant (because along this line B does not change). This is the *partial derivative* with respect to A, and is represented by the curly ∂ symbol to distinguish it from a full derivative. Using the above example we can see that:

$$\frac{\partial Z}{\partial A} = \cos(A)\cos(B) \qquad \text{and} \qquad \frac{\partial Z}{\partial B} = -\sin(A)\sin(B).$$

By continuing the argument we made in the previous section we can see that ΔZ, the total uncertainty in Z arising from uncertainties in A and B, will depend on how much Z changes when A and B change by ΔA and ΔB. As before we can make the case that these contributions will be of the form

$$\frac{\partial Z}{\partial A}\Delta A \qquad \text{and} \qquad \frac{\partial Z}{\partial B}\Delta B.$$

so that, for small differences (linear approximation):

$$\Delta Z \approx \left(\frac{\partial Z}{\partial A}\right)_{A_0} \Delta A + \left(\frac{\partial Z}{\partial B}\right)_{B_0} \Delta_B. \tag{6}$$

Squaring both sides:

$$(\Delta Z)^2 \approx \left(\frac{\partial Z}{\partial A}\right)^2 (\Delta A)^2 + \left(\frac{\partial Z}{\partial B}\right)^2 (\Delta B)^2 + 2\left(\frac{\partial Z}{\partial A}\right)\left(\frac{\partial Z}{\partial B}\right)\Delta A \Delta B. \tag{7}$$

But this is $(\Delta Z)^2$ for a *single measurement*. In order to find the expected value of $(\Delta Z)^2$ we must average over many readings. The third term then becomes negligible, because A and B are independent variables, and ΔA and ΔB are equally likely to be positive or negative (assuming symmetric distributions as we have done throughout).

And so we arrive at a formula for the uncertainty on Z which can easily be extended to be a function of many variables $Z = Z(A, B, C\ldots)$:

$$(\Delta Z)^2 = \left(\frac{\partial Z}{\partial A}\right)^2 (\Delta A)^2 + \left(\frac{\partial Z}{\partial B}\right)^2 (\Delta B)^2 + \left(\frac{\partial Z}{\partial C}\right)^2 (\Delta C)^2 \ldots \tag{8}$$

To see how this works we can apply equation (8) to our function $Z = \sin(A)\cos(B)$ using the partial derivatives $\frac{\partial Z}{\partial A}$ and $\frac{\partial Z}{\partial B}$ already derived:

$$(\Delta Z)^2 = \cos^2(A)\cos^2(B)(\Delta A)^2 + \sin^2(A)\sin^2(B)(\Delta B)^2.$$

Often we can find an especially pleasing (simple) form if we then divide both sides by $Z^2 = \sin^2(A)\cos^2(B)$, giving an equation for the fractional uncertainty in Z:

$$\left(\frac{\Delta Z}{Z}\right)^2 = \cot^2(A)(\Delta A)^2 + \tan^2(B)(\Delta B)^2.$$

In this case the uncertainties on the angles ΔA and ΔB should be in radians (we shall see why in the next section).

2.3 Standard results

While the required formula for the propagation of uncertainties through any function can be derived from equation (8) , there are some standard results you should know. Table 2 gives some commonly used uncertainty propagation formulae, and you should check that you are able to derive these results from the above general formula. Some important points to note include the uncertainty in the difference in two measuremenst which is the sum of the squares of the *absolute* uncertainties. This can mean that, for example, when trying to measure a small change in displacement the resultant uncertainty can be very large compared to the actual value. Also the effect of raising a quantity to a power, which multiples the *fractional* uncertainty by the value of the exponent. This makes the uncertainty on the square or the cube of a measurement considerably larger.

Formula	Uncertainty combination
$Z = A \pm B \pm C$	$(\Delta Z)^2 = (\Delta A)^2 + (\Delta B)^2 + (\Delta C)^2$
$Z = A \times B$ or $Z = \frac{A}{B}$	$\left(\frac{\Delta Z}{Z}\right)^2 = \left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2$
$Z = A^n$	$\frac{\Delta Z}{Z} = \lvert n\rvert \frac{\Delta A}{A}$
$Z = \ln A$	$\Delta Z = \frac{\Delta A}{A}$
$Z = \exp A$	$\frac{\Delta Z}{Z} = \Delta A$

Table 2: Combination of uncertainties for simple functions.

Another commonly encoutered situation is the propagation of uncertainties through a trigonometric function. We will use this as a worked example to show where mistakes frequently occur. Say we have measured an angle θ, and want to find how an uncertainty in the measurement of θ affects the calculation of $Z = \sin(2\theta)$. We know that generally

$$\Delta Z = \frac{\partial Z}{\partial \theta}\Delta\theta$$

as we have a function of a single variable $Z = Z(\theta)$ only. Since we have our experimental uncertainty $\Delta\theta$ and want to calculate ΔZ we only have to calculate:

$$\frac{\partial Z}{\partial \theta} = 2\cos(2\theta)$$

in order to derive the required result:

$$\Delta Z = 2\cos(2\theta)\Delta\theta. \tag{9}$$

The proviso here is that the uncertainty $\Delta\theta$ must be in **radians**. Hopefully this will be obvious by considering the dimensions of each side of the equation. Since Z is the result of a trigonometric function it, and therefore ΔZ on the left hand side of equation (9), also must be *dimensionless* - the result of taking the sine of an angle is just a number. The right hand side of equation (9) then is also dimensionless. Since this is the product of a number (2, which has no dimensions), a trigonometric function ($\cos(2\theta)$, which has no dimensions) and $\Delta\theta$ which therefore **must** be **dimensionless**. This requires that it is measured in radians rather than degrees.

2.4 The weighted mean

When calculating the average of a set of data $(x_1, x_2, x_3 \ldots x_n)$ we have typically been using the (arithmetic) mean:

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

in which all the measurements x_i are assigned an equal importance, or equal *weight.* There are many situations in which this is not the case - one example is the calculation of course marks from lecture courses at the end of year assessment. Typically these are made up of a coursework mark, call it x_1, and an exam mark, call it x_2. A simple arithmetic mean could be calculated, but it is generally accepted that the exam mark x_2 is more important than the continuous assessment, and so it receives a higher *weighting* in the averaging process. The average is then calculated by:

$$\overline{x} = \frac{w_1 x_1 + w_2 x_2}{w_1 + w_2}$$

where w_1 and w_2 are the weights assigned to x_1 and x_2 respectively. The presence of $w_1 + w_2$ in the denominator ensures the correct normalisation. The weights can then be chosen to reflect the overall importance placed on the components of the assessment.[2]

It is straightforward to extend this to a set of n measurements as:

$$\overline{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}. \tag{10}$$

So given our set of data $(x_1, x_2, x_3 \ldots x_n)$, what should we choose for the weights, and why would we want to weight them differently anyway? We have already discussed how some measurements may be less reliable than others for a variety of reasons, and so it seems right that those less reliable measurements should contribute less to finding the average. We also have a way of quantifying the reliability with the standard error, so this would seems a

[2] Out of interest $w_1 = 0.1$ and $w_2 = 0.9$ for most courses

good way to weight the measurements to make those with a large experimental uncertainty less important.

For the dataset $(x_1, x_2, x_3 \ldots x_n)$ with associated uncertainties $(\Delta x_1, \Delta x_2, \Delta x_3 \ldots \Delta x_n)$ we will define the weight of each as:

$$w_i = \frac{1}{(\Delta x_i)^2}. \tag{11}$$

From this definition, equation (10) becomes

$$\overline{x} = \frac{\sum_{i=1}^{n} \frac{x_i}{(\Delta x_i)^2}}{\sum_{i=1}^{n} \frac{1}{(\Delta x_i)^2}}. \tag{12}$$

The variance of $\overline{x}$ is:

$$\frac{1}{\sigma_{\overline{x}}^2} = \sum_{i=1}^{n} \frac{1}{(\Delta x_i)^2}, \tag{13}$$

and so we quote the final weighted mean result as $\overline{x} \pm \sigma_{\overline{x}}$ (with the appropriate units, of course!)

There is an important proviso surrounding use of the weighted mean, that the data you use must be independent and non-correlated, otherwise a bias may be introduced.

3 Standard deviation and standard error on the mean

The discussions in the previous sections have shown that a set of measurements of a physical quantity will be distributed about some average value. The extent of the distribution is in some way related to our experimental uncertainty. We should therefore be able to use the mathematical tools of statistical distributions to put a numerical value on the size of the uncertainty. We are then able to express a quantitative judgement on the reliability of the average value

3.1 Standard error in a single measurement, σ

Suppose we take a set of n measurements of a quantity: $x_1, x_2, x_3 \ldots x_n$, and that, given the limited time, money and patience available, n is not too large - probably 7 - 10 measurements. We already know that the measurements are distributed about some mean value:

$$\overline{x} = \frac{1}{n}\sum_{1}^{n} x_i, \tag{14}$$

and that in all probability $\overline{x}$ is close to, but not exactly equal to the true value, X. Right now you should be *demanding* to know the variance of the distribution. But what exactly is the distribution which we are trying to find the variance of?

A single measurement, value x differs from the true mean X by the error $e = x - X$. These measurements will be spread about the true value with a variance

$$\sigma^2 = \langle e^2 \rangle = \int_{-\infty}^{+\infty} (x - X)^2 y(x) \mathrm{d}x. \tag{15}$$

The variance σ^2 and standard deviation σ are those of *single measurements* about the true value.

Intuitively we probably expect that these might be quite widely spread, and that we can improve our experiments by taking several sets of measurements and averaging which after all is why we have taken several measurements in the (finite) dataset $x_1, x_2, x_3 \ldots x_n$. What we really want to find is the variance of the distribution of *mean values* derived from many independent repetitions of the experiment (independent datasets) about the true value. For this we require the *standard error on the mean.*

3.2 Standard error on the mean, σ_m

Considering our one dataset $x_1, x_2, x_3 \ldots x_n$, the error in the measurement x_i is $e_i = x_i - X$. Of course X here is the true value of the quantity which we do not know - all we have is the mean of our data set, $\overline{x}$. The difference of the dataset mean from the true value is given by E where

$$E = \overline{x} - X = \frac{1}{n}\sum_1^n x_i - X = \frac{1}{n}\sum_1^n (x_i - X) = \frac{1}{n}\sum_1^n e_i \tag{16}$$

and so by squaring:

$$E^2 = \frac{1}{n^2}\sum_1^n e_i^2 + \frac{1}{n^2}\sum_i \sum_{j \neq i} e_i e_j. \tag{17}$$

The double summation in the second term of equation (17) ensures that each term in the series multiplies each of the others in turn without double-counting. Remember that this is for our *one* dataset. If we take many sets of data and average over all the datasets we can get an average value $\langle E^2 \rangle$ for the (squares of) the uncertainties on the mean, which is nothing but the variance of the distribution of mean values about the true value σ_m^2 just as we wanted.

The average value of the first summation term is $n\langle e^2 \rangle$, and the average of the double summation is zero, and so we find that:

$$\langle E^2 \rangle = \frac{1}{n}\langle e^2 \rangle, \tag{18}$$

or, in terms of the standard deviations (standard errors):

$$\sigma_m = \frac{\sigma}{\sqrt{n}}. \tag{19}$$

This demonstrates that taking an average helps to improve precision, because it shows that the standard error in the mean of n readings is smaller than the standard error in a single reading by a factor $\frac{1}{\sqrt{n}}$.

We can draw two very useful lessons from this. The first is that averaging is good for you, it reduces the standard error in the mean and so improves the precision. The second is that the rate of improvement only varies as $\frac{1}{\sqrt{n}}$, so in order to improve your precision by a factor of, say, 10 you need 100 times as many data points. Trying to do too much averaging, while not necessarily bad for you, stops doing you so much good after a while. A good demonstration of the Law of Diminishing Returns!

3.3 How to estimate σ and σ_m

We have not yet solved the problem of how to calculate σ_m, as equation (19) merely relates it to σ, which relies on the variance around the as yet unknown true value. The only information we have are our measurements, $x_1, x_2, x_3 \ldots x_n$ (a *sample*), and their mean, $\overline{x}$.

What we can calculate is the mean square deviation (variance) from the mean of the sample. Each measurement differs from the sample mean by an amount $d_i = x_i - \overline{x}$, called the *residual.* The mean of the square of these differences (deviations) is then:

$$s^2 = \frac{1}{n}\sum_1^n d_i^2 = \frac{1}{n}\sum_1^n (x_i - \overline{x})^2 \tag{20}$$

that is, s^2 is the variance of the sample, and s the standard deviation of the sample.

But since the error on the measurement x_i is $e_i = x_i - X$, and the error on the mean is $E = \overline{x} - X$ we can substitute into equation (20):

$$s^2 = \frac{1}{n}\sum_i^n (e_i - E)^2 = \frac{1}{n}\sum_1^n e_i^2 - E^2 \tag{21}$$

which if we average over many sets of measurements as before gives us that:

$$\langle s^2 \rangle = \sigma^2 - \sigma_m^2. \tag{22}$$

So, by using equation (19) in equation (22) we find that:

$$\sigma^2 = \frac{n}{n-1}\langle s^2 \rangle \qquad \text{and} \qquad \sigma_m^2 = \frac{1}{n-1}\langle s^2 \rangle. \tag{23}$$

At last we have equations for σ and σ_m! **But** our problems are not over, for in actual fact we **do not know** the average of the sample variances over datasets $\langle s^2 \rangle$, but just the variance of a *single* dataset s^2. We expect that for enough measurements s^2 will approach $\langle s^2 \rangle$, but all we can do is take the square root in equations (23), and replace the ideal and thus unknown $\sqrt{\langle s^2 \rangle}$ with the known s from equation (20) to give a best estimate:

$$\sigma \approx \sqrt{\frac{n}{n-1}}s, \tag{24}$$

and

$$\sigma_m \approx \sqrt{\frac{1}{n-1}}s. \tag{25}$$

Just how good is this approximation? A full treatment is beyond the scope of this course, but we could calculate the difference between the sample variance of each dataset and the mean of the sample variances (i.e. $s^2 - \langle s^2 \rangle$) and then the mean of the squares of these differences - the variance of the distribution of sample variances about the mean sample variance! We can then find the standard deviation (take the square root) and express this as a fraction of the mean value. We then want the fractional standard deviation of s, rather than s^2. It turns out that for a Gaussian distribution this fractional standard deviation is (approximately) $\frac{1}{\sqrt{2n-2}}$ for a reasonably large number of readings, n.

So even for a set of nine measurements, which will probably take a reasonable amount of time in the lab, the best estimate of the uncertainty is only good to around 25%. Given this level of confidence you can see why it doesn't make sense to quote an uncertainty to more than one significant figure - it cannot be known so precisely.

You would be forgiven for thinking that physics, which had previously seemed to be all about precision and measuring things exactly and knowing about the Universe, has degraded into vague statements about not knowing what we have just measured and "uncertainty". In fact the opposite is true. The power of physics (and science as a whole) is that we can say what we know *and how reliably we know it.* We can sensibly quantify just how good (or how poor) our knowledge is, and therefore judge just how much importance we should place on it.

3.4 The error function

From the above discussion we now have a way of quantifying uncertainties in our experimental measurements, and indicating the expected spread of measurements about an average value. So given a particular experimental measurement, and its associated uncertainty, can we now say what the *probability* of any subsequent measurement agreeing with it is? After all, this situation is commonly encountered in the laboratory when you want to compare your measured value of some quantity with a reported value in a text book or data book.

If we remember that our measured value, with an uncertainty, gives us a Gaussian probability distribution for the outcome of subsequent measurements, we can work out the probability of a measurement being less than some value ξ by integrating the Gaussian:

$$P(x < \xi) = \frac{1}{\sqrt{2\pi}\sigma_m} \int_{-\xi}^{+\xi} \exp(-x^2/2\sigma_m^2)\mathrm{d}x = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_m} \int_{0}^{+\xi} \exp(-x^2/2\sigma_m^2)\mathrm{d}x. \tag{26}$$

This function must be evaluated numerically, but if we make the substitution $t = x/\sigma_m$ and $z = \xi/\sigma_m$ then this probability becomes:

$$\mathrm{erf}(z) = \sqrt{\frac{2}{\pi}} \int_{0}^{z} \exp(-t^2/2)\mathrm{d}t \tag{27}$$

known as the *error function.* We are particularly interested in the values of the error function for $z = 1, 2, 3$ etc, that is $\xi = \sigma_m, 2\sigma_m, 3\sigma_m$ etc, in other words the probability of another measurement being less than one standard error away from our measurement, or two, or

ξ	z	$\mathrm{erf}(z)$	$1-\mathrm{erf}(z)$
0	0	0	1
σ_m	1	0.683	0.317
$2\sigma_m$	2	0.954	0.046
$3\sigma_m$	3	0.997 3	0.002 7
$4\sigma_m$	4	0.999 94	0.000 06

Table 3: Tabulated values of the error function, $\mathrm{erf}(z)$

three etc. These are shown in table 3, along with values of $1 - \mathrm{erf}(z)$, i.e. the probability of a measurement being greater than $\xi = \sigma_m, 2\sigma_m, 3\sigma_m$ etc.

As you can see, the probability of obtaining a result that is different by more than two or three times your standard error becomes very small indeed. This, then, is the best way to compare a measured value of a quantity against an accepted, published value: you can say that the values differ by some multiple of your standard error. You would expect to find agreement within one standard error ("one sigma") 68.3% of the time, within two standard errors ("two sigma") 95.4% of the time, and so on. You can see that large deviations are, therefore, very unlikely, and if you find this is the case you should check your calculations and method very carefully. Usually a large deviation of several sigma is indicative of a mistake in calculating your value, or in calculating the uncertainty, or perhaps in not accounting for a significant systematic uncertainty.[3]

4 Least squares fitting

When performing an experiment we very often are measuring how one quantity varies as a function of another. We set a value of one parameter (call it x, the independent variable), and measure another (call it y, the dependent variable). We then want to know exactly how y depends on x, that is we want to know $y(x)$.

Obviously to see this dependence we plot a graph of $y(x)$ as we do the experiment. The shape of the graph can show the general form of the dependence, and we can measure parameters from it to extract more useful information. For instance we will usually be dealing with straight line graphs

$$y(x) = mx + c \tag{28}$$

where we simply plot a graph of y against x and then measure the gradient m and y-intercept c. Of course we might not be so fortunate that the quantity we are measuring has a simple linear dependence. In this case we can often rearrange the expected form to find quantities that do vary linearly. An example of this would be the energy of photoelectrons released from a metal surface illuminated with ultra-violet light. The maximum kinetic energy, E_{max}, of the photoelectrons varies with wavelength, λ, as:

$$E_{\mathrm{max}} = \frac{hc}{\lambda} - \Phi \tag{29}$$

[3] Of course, it could also be that you have discovered some new physics, but this, alas, is unlikely in the undergraduate laboratories.

where h is Planck's constant, and c is the speed of light. The quantity Φ is called the *work function* of the metal and is the amount of energy required to *just* release an electron - obviously an important quantity to measure!

If we examine equation (29) more closely we can see that:

$$\underbrace{E_{\text{max}}}_{y} = \underbrace{hc}_{m}\underbrace{\left(\frac{1}{\lambda}\right)}_{x}\underbrace{-\Phi}_{+c}, \tag{30}$$

that is the maximum photoelectron kinetic energy varies linearly with λ^{-1}. So if we were to do the experiment and measure E_{max} for several different values of λ but then plot a graph of E_{max} against $\frac{1}{\lambda}$ we would expect a straight line. The gradient of this will be equal to hc, and the intercept on the y-axis equal to $-\Phi$.

So all we need to do is plot a straight line and read off the gradient and y-intercept? Not quite. We know now that all our experimentally measured quantities have an uncertainty. On the graph the points won't all lie nicely on a straight line where you want them, but will be spread about, both above and below, some *line of best fit.* Previously you have probably drawn a line of best fit by inspecting to see which line goes through most of the data points, while ignoring any that lie too far away from this line. Now using a statistical method called the *method of least squares* we can calculate what the best values of m and c are from our data.

4.1 The method of least squares

Suppose that we have a set of n pairs of readings from an experiment $(x_1, y_1), (x_2, y_2) \dots$ up to (x_n, y_n), where x_i represents the parameter we set, and y_i is the corresponding measured quantity. For now we will assume that the largest source of uncertainty is in the measured values y_i. For a particular choice of m and c we can calculate $mx_i + c$ (what we would *expect* to measure), and compare it to y_i (what we actually did measure), that is we find the deviation $y_i - mx_i - c$. This is illustrated in figure 3 where the deviation of a point (x_i, y_i) from a straight line with gradient m and y-intercept c is shown.

The measured points will lie both above (positive deviation) and below (negative deviation) the line so if we sum the squares of the deviations:

$$S = \sum (y_i - mx_i - c)^2 \tag{31}$$

we will come up with a (positive) number to describe the total deviation. Summing the squares of the deviations ensures that the individual positive and negative deviations don't cancel out. The line of best fit is then the line $y = mx + c$ with values of m and c that make S a **minimum**.

So we need to minimise S with respect to both m and c. The values at which S is minimum are found from the (partial) derivatives:

$$\frac{\partial S}{\partial m} = -2\sum x_i(y_i - mx_i - c) = 0 \qquad \text{and} \qquad \frac{\partial \mathrm{S}}{\partial \mathrm{c}} = -2\sum(\mathrm{y_i} - \mathrm{mx_i} - \mathrm{c}) = 0 \tag{32}$$

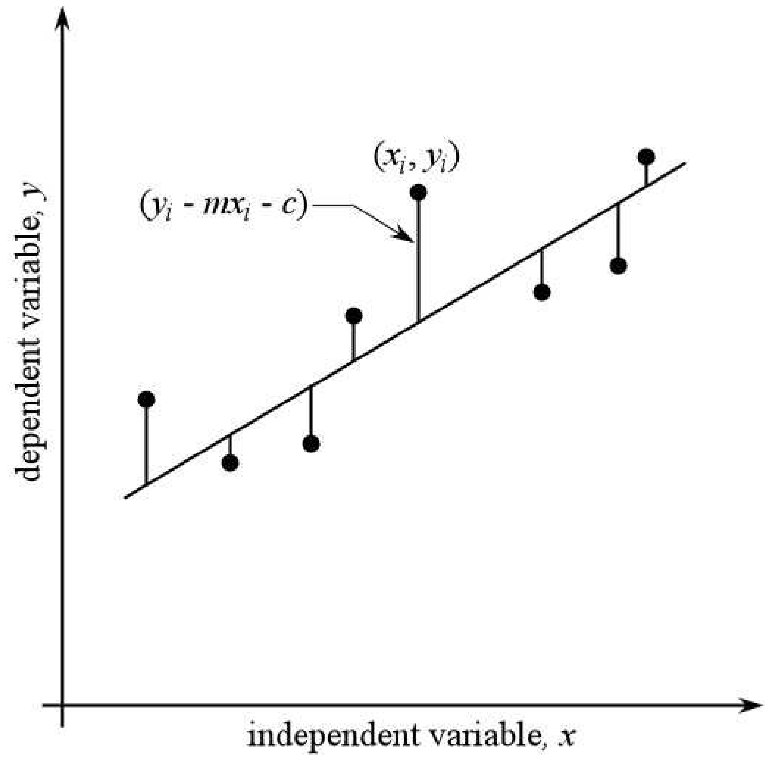

Figure 3: Illustration of the method of least squares. The best fit lines is where $\sum(y_i - mx_i - c)^2$ is minimum.

so that we have a pair of simultaneous equations:

$$m \sum x_i^2 + c \sum x_i = \sum x_i y_i \tag{33}$$

and

$$m \sum x_i + nc = \sum y_i \tag{34}$$

from which we can find m and c. If we divide equation (34) throughout by n, the number of measurements we can see that $\overline{y} = m\overline{x} + c$, that is the best fit line goes through the point that corresponds to the average values of x and y, called the centre of gravity of the points.

If we solve equations (33) and (34) for m and c we find that the best values are:

$$m = \frac{\sum(x_i - \overline{x})y_i}{\sum(x_i - \overline{x})^2} \tag{35}$$

and

$$c = \overline{y} - m\overline{x}. \tag{36}$$

Of course, now that we are familiar with distributions, means and variances we know that if we were to repeat the experiment several times we would expect to find a distribution of best fit values of m and c and some mean (of the sample) value of m and c. This ties in with what we know about all measured quantities having an experimental uncertainty.

Any parameter we determine from the gradient and y-intercept of the graph via the best fit method must have an uncertainty, so we need to find the standard error on the mean values of m and c.

We will not prove these results, but state here, that if we use the best fit values of m and c to define the residual:

$$d_i = y_i - mx_i - c$$

and also set

$$D = \sum (x_i - \overline{x})^2$$

then the standard errors in m and c are:

$$\Delta m \approx \sqrt{\frac{1}{D}\frac{\sum d_i^2}{(n-2)}}, \tag{37}$$

and

$$\Delta c \approx \sqrt{\left(\frac{1}{n} + \frac{\overline{x}^2}{D}\right)\frac{\sum d_i^2}{(n-2)}}. \tag{38}$$

Qualitatively you can understand this. If the sum of the squares of the residuals, $\sum d_i^2$, is big then the points are scattered far from the best fit line, and we would expect a large uncertainty in m and c. Also note that if D is large then Δm and Δc are small, that is we should aim to take measurements over a wide range about the average value. Again this is intuitively obvious.

4.2 Method of least squares with unequal weights

The discussion above has assumed that all the data points in the graph are equally important, that is, they all have equal *weighting.* We already know, however that this is not the case - when we draw a 'best fit' line by eye with a ruler we often neglect outlying points that lie far from a straight line, essentially ascribing them a zero weighting and making them unimportant to the fit.

We have also already seen how not all experimentally measured data points are of equal importance, and reduced the significance of those with a large uncertainty with an appropriate weighting when calculating the weighted mean. Now we will apply the above method of least squares for finding the best fit to data that requires unequal weighting because of different experimental uncertainties on each point. For simplicity we will assume that only the uncertainties in the measured quantities, i.e. Δy_i are significant.

In this case we calculate the *weighted* sum of the squares of the deviations of the points from a line with gradient m and intercept c:

$$S_w = \sum_{i=1}^{n} w_i (y_i - mx_i - c)^2. \tag{39}$$

Go back and compare equation (39) with equation (31) now. The weighting we have applied in equation (39) of course comes from the experimental uncertainty in y, i.e. $w_i = \frac{1}{(\Delta y_i)^2}$.

As before we want to minimise with respect to m and c to obtain a pair of simultaneous equations:

$$m \sum w_i x_i^2 + c \sum w_i x_i = \sum w_i x_i y_i \tag{40}$$

and

$$m \sum w_i x_i + c \sum w_i = \sum w_i y_i. \tag{41}$$

Again you should go back to equations (33) and (34) to check the similarity with (40) and (41). Note now that the best fit line goes through the *weighted* average values of x and y.[4]

Solving first for the gradient we find:

$$m = \frac{\sum w_i \sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i}{\sum w_i \sum w_i x_i^2 - \left(\sum w_i x_i\right)^2}, \tag{42}$$

and then obtain the y-intercept through

$$c = \sum w_i y_i - m \sum w_i x_i. \tag{43}$$

5 Distributions

5.1 Distribution functions

We know that almost all measurable quantites that we commonly encounter - the price of houses in London, the number of runs per innings scored by England batsmen, the number of UCAS points obtained by first year undergraduate physics students - show a variation. By this we mean that, to use the above examples, not all houses in London cost the same, some batsmen average higher than others, and different first year undergraduates achieve different marks for their A-level exams.

5.1.1 What is a distribution?

A distribution, $y(x)$, of a variable x tells us how often a value of x occurs within a sample. In order to provide useful information about $y(x)$ we need a measure of the value around which the other values are distributed, and also a measure of how widely they are distributed. For this we will use the **mean** and **variance** of the distribution[5].

5.1.2 Mean and variance of distributions

If a distribution consists of a finite number N of measurements then we already know that the mean of the distribution is:

$$\overline{x} = \frac{1}{N} \sum x_i. \tag{44}$$

[4] Although as an aside we note that the x_i are being weighted by values derived from the uncertainty in y_i as we assumed the uncertainties in x_i were insignificant.

[5] At this stage we will not consider higher order parameters such as *skew* and *kurtosis* that give additional information about the shape of the distribution

But this is obviously not the complete picture. We need to find a measure of how spread out around the mean value the distribution is. For this we use the variance:

$$\sigma^2 = \frac{1}{N}\sum(x_i - \overline{x})^2. \tag{45}$$

To find the variance of the distribution we calculate the *residual*, $d_i = x_i - \overline{x}$ of each measurement, that is the difference between it and the mean value, square it and divide by the number of measurements. For this reason it is sometimes called the mean square deviation.

For example, consider the distribution of ages of first year undergraduates taking course PHAS1240. It would be fairly reasonable to guess that the mean age is somewhere between 18 and 19 years old - most students come to university after completing A-levels aged 18, although some may have had a gap year which would raise the mean slightly. But what is the variation in this distribution? For the same reasons it would be not too unreasonable to suppose that the variation in ages is not large (probably only a few months) and so we would expect the distribution to be narrowly spread about the mean - a distribution with a small variance.

As another example, according to the Land Registry[6] the average price of a house in Camden is £538,321. It is very unlikely that all houses in Camden cost exactly this much, and we can probably expect the variation in prices to be quite broad (a large variance) as it includes everything from small flats to large mansions.

Furthermore, the average price in another borough (Hammersmith and Fulham) is apparently almost the same at £536,402. But we cannot make a meaningful comparison between the two areas without knowing the variances of the two distributions. It may be that the variance in one borough is much larger than the other, meaning that there will be more houses priced far from the mean, both much higher and much lower. The trouble is that without the variance of the distribution we cannot know for sure.

To summarise, a distribution is characterised by **two** parameters, the **mean** and the **variance**, and **both** should be quoted in order to describe the distribution meaningfully. You should be very suspicious of sources that give a mean without an indication of the spread, and consider it to show that the author either does not understand statistics, or is trying to hide something.

5.2 Gaussian distribution

5.2.1 What is a Gaussian Distribution?

The general mathematical form for a Gaussian (in one dimension) is:

$$y(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-(x-x_0)^2/2\sigma^2). \tag{46}$$

[6] Land Registry figures, August 2007

What does this function look like? We can already deduce something about this from the above equation: $y(x)$ is maximum when $(x - x_0) = 0$, and decreases both for $(x - x_0) > 0$ and $(x - x_0) < 0$. So this function must be **peaked** about the value $x = x_0$, and at the peak the function is:

$$y(x = x_0) = \frac{1}{\sqrt{2\pi}\sigma}. \tag{47}$$

What else? The parameter σ is obviously important. To see why, look at what happens when $(x - x_0) = \pm\sigma$:

$$y(x = x_0 \pm \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-1/2) = 0.607 \times \frac{1}{\sqrt{2\pi}\sigma} \tag{48}$$

Compare the value of the function here $(x = x_0 \pm \sigma)$ to the value at the peak $(x = x_0)$ above, and we see that the function has decreased to 0.607 (or $\frac{1}{\sqrt{e}}$) of the peak value.

The parameter σ, therefore, gives us a measure of the **width** of the peaked function - the amount we have to move along the x-axis away from x_0 (the position of the peak) in order for $y(x)$ to decrease to $\frac{1}{\sqrt{e}}$ of the peak value.

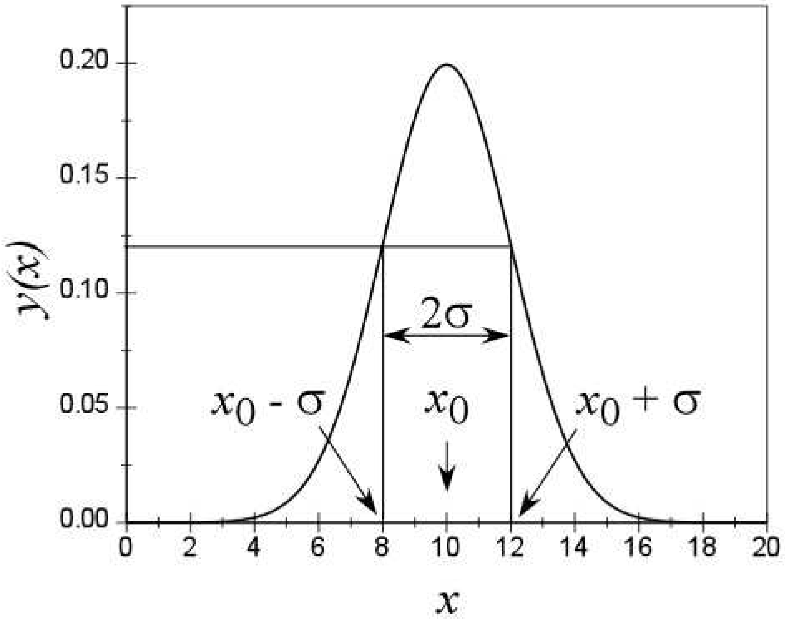

Figure 4: An example of a Gaussian function.

An example of a Gaussian function is shown in figure 4. You can see it has the characteristic peaked shape, sometimes called a bell-curve. In this example $x_0 = 10$, and $\sigma = 2$.

You can see that the Gaussian has a maximum value of $(\sqrt{2\pi} \times 2)^{-1} = 0.2$, that occurs when $x = x_0 = 10$. You can also see that when $x = x_0 \pm \sigma$, i.e. when $x = 12$ and $x = 8$, the value of the function is 0.12, and that this is a fraction $0.12/0.2 = 0.6$ of the maximum value.

Figure 5 shows two Gaussian curves that have the same mean value, $x_0 = 20$, but different widths, $\sigma = 1$ and $\sigma = 5$. Obviously the peak value of one curve is smaller than the other, which arises from the area under the curves being normalised to 1, making $y(x = x_0) = \frac{1}{\sqrt{2\pi}\sigma}$. But you can also see that the Gaussian with the larger value of σ is broader, that is, it spreads further away from the mean.

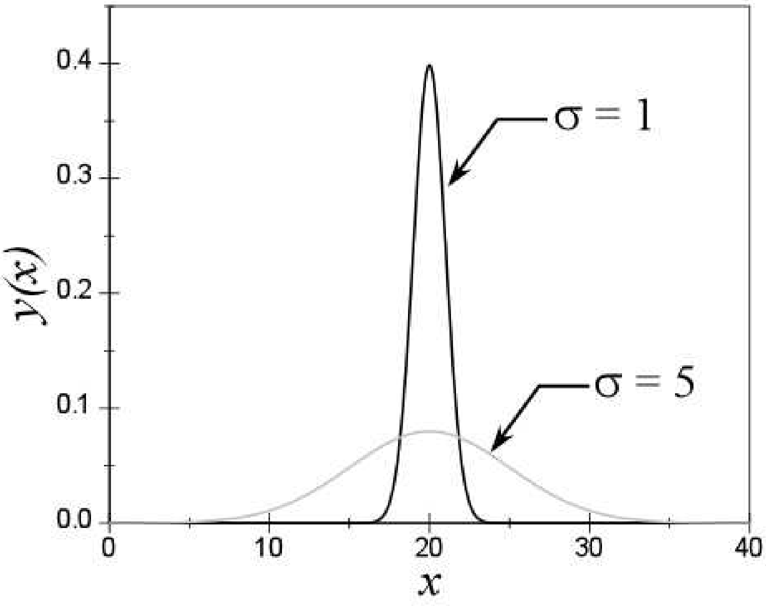

Figure 5: Comparison of Gaussian functions with different widths.

5.2.2 Properties of a Gaussian Distribution

There are several properties of the Gaussian function that make it relevant to the analysis of experimental data, and that you should be familiar with:

Normalisation: The function $y(x)$ in equation (46) is *normalised* such that the area under the curve is unity, that is:

$$\int_{-\infty}^{+\infty} y(x)dx = 1. \tag{49}$$

Mean: The parameter x_0 is the *mean* of the distribution, that is the average value of all x:

$$x_0 = \int_{-\infty}^{+\infty} xy(x)dx \tag{50}$$

if $y(x)$ is normalised as above. The integral in equation (50) is called the first moment of x. Be aware that for a Gaussian, x_0 is also the mode and the median value.

Compare equation (50) which is the mean of the *continuous* Gaussian distribution with equation (44), the mean for the distribution consisting of a *discrete* distribution with a finite number of measurements. You can see how these both describe the same process for finding the mean: each value is multiplied by the frequency with which it occurs in the distribution, these are summed (integrated in the limit of a continuous distribution), and divided by the number of values. For the continuous Gaussian distribution this division is by $\int_{-\infty}^{+\infty} y(x)dx = 1$.

Variance: It turns out that when written in the form of equation (46) the square of the parameter σ that we said describes the width of the distribution, is exactly the *variance* of the Gaussian distribution.

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - x_0)^2 y(x)dx \tag{51}$$

which is the second moment of the deviations, if $y(x)$ normalised as above. Again, by comparing equations (45) and (51) you will see that they both describe the same process for the discrete and continuous distributions respectively: the square of the deviation is weighted by the frequency, these are summed and averaged.

Standard deviation: The square root of the variance, σ, is called the root-mean-square or *standard deviation* from the mean.

5.2.3 Relevance of the Gaussian distribution

Why is the Gaussian distribution so important? A feature that makes it so useful is the Central Limit Theorem, which relates to making a number of measurements of a quantity, each of which is subject to a random uncertainty. The Central Limit Theorem for a large number of measurements shows that the distribution of these measurements tends towards a Gaussian. We will return to the problem of sampling a number of measurements from a large population, the variance of the sample, the variance of the population and how this relates to experimental uncertainties in the laboratory later in the course.

5.3 Poisson Distribution

In most cases a distribution of measurements made in the laboratory will be indistinguishable from a Gaussian, however there are a few special cases. One of these is the Poisson distribution.

5.3.1 What is the Poisson distribution?

The Poisson distribution arises when a random process gives rise to discrete measured values, or when we are counting the number of independent events that occur within a fixed time interval. A good example of this is the counting of particles emitted in a radioactive decay.

Radioactive decay is a random process. Each decay event is independent of the others, and any one atom in a sample has a fixed probability of decaying in a time interval. The decay products (particles) are then emitted at some average rate, λ, and we want to know the probability that a particular decay rate r will be measured.

This probability, P_r, is given by the Poisson distribution:

$$P_r = \frac{\lambda^r}{r!}\exp(-\lambda). \tag{52}$$

5.3.2 Properties of the Poisson distribution

We can calculate the mean and variance of the Poisson distribution as follows. The mean value of r is

$$\overline{r} = \sum rP_r = \lambda, \tag{53}$$

and the variance is

$$\mathrm{Var}(r) = \sum (r-\lambda)^2 P_r = \lambda. \tag{54}$$

So for the Poisson distribution the mean and variance are equal.

Examples of the Poisson distribution for mean values $\lambda = 1, 2, 3$ are shown in figure 6. You can see that the maximum of the distribution moves to higher values of r as the mean increases, and also that the spread of the distribution gets wider.

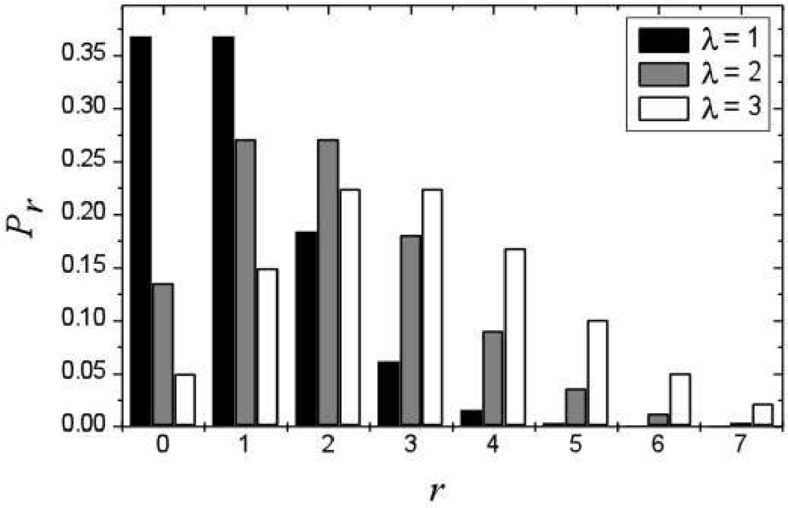

Figure 6: Examples of the Poisson distribution for mean values $\lambda = 1, 2, 3$.

Finally, for large λ the Poisson distribution is very like a Gaussian of mean λ and variance λ.

5.3.3 Relevance of the Poisson distribution

The Poisson distribution is important when a rare event is being monitored. It's most significant feature is that it is asymmetric, especially when there is a low mean value. If the mean is less than about 35, the asymmetry is pronounced and a Poission distribution should be used.

6 Glossary

Combining Uncertainties For a function $Z = Z(A, B, C \ldots)$ with small uncertainties $\Delta A, \Delta B, \Delta C \ldots$ the uncertainty on Z is:

$$(\Delta Z)^2 = \left(\frac{\partial Z}{\partial A}\right)^2 (\Delta A)^2 + \left(\frac{\partial Z}{\partial B}\right)^2 (\Delta B)^2 + \left(\frac{\partial Z}{\partial C}\right)^2 (\Delta C)^2 \ldots.$$

If the uncertainties on $\Delta A, \Delta B, \Delta C \ldots$ are large, then the funtional approach must be used:

$$\begin{aligned}(\Delta Z)^2 &= [Z(A+\Delta A, B, C, \ldots) - Z(A, B, C, \ldots)]^2 \\ &\quad +[Z(A, B+\Delta B, C, \ldots) - Z(A, B, C, \ldots)]^2 \\ &\quad +[Z(A, B, C+\Delta C, \ldots) - Z(A, B, C, \ldots)]^2 \\ &\quad +\ldots\end{aligned}$$

Mean. The mean of a continuous distribution $f(x)$ is

$$X = \frac{\int_{-\infty}^{+\infty} x f(x)\mathrm{d}x}{\int_{-\infty}^{+\infty} f(x)\mathrm{d}x}.$$

The presence of $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x$ in the denominator takes account of the function $f(x)$ not being normalised to equal 1. For the definition of the Gaussian in equation (46) this integral is equal to one.

For a finite number of discrete samples:

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

Variance. The variance of a continuous distribution $f(x)$ is:

$$\sigma^2 = \frac{\int_{-\infty}^{+\infty} (x-X)^2 f(x)\mathrm{d}x}{\int_{-\infty}^{+\infty} f(x)\mathrm{d}x}.$$

Again, the integral in the denominator applies for the case of a function that is not normalised to unity.

For a finite number of discrete samples the variance *about the sample mean* is:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} d_i^2$$

where $d_i = x_i - \overline{x}$ is the ith residual.

Residual. The residual of a single measurement is the difference between it and the mean

$$d_i = x_i - \overline{x}$$

or between it and the fit used:

$$d_i = y_i - y(x_i).$$

If there is structure in a plot of the residuals, this indicates that there are terms missing in the fit.

Standard deviation. The standard deviation is the square root of the variance.

Standard error on the mean. The standard error on the mean is related to the variance of the population distribution by:

$$\sigma_m = \frac{\sigma}{\sqrt{n}}$$

Best estimate of standard error. For a finite number of samples with standard deviation s about the sample mean $\overline{x}$, the best estimate we can make of σ_m is:

$$\sigma_m \approx \sqrt{\frac{1}{n-1}}s.$$

Weighted mean. For a dataset x_i with associated uncertainties Δx_i the weighted mean of the data is:

$$\overline{x} = \frac{\sum w_i x_i}{\sum w_i}$$

with the weights:

$$w_i = \frac{1}{(\Delta x_i)^2}$$

and standard error on the weighted mean given by:

$$\frac{1}{\sigma_{\overline{x}}^2} = \sum \frac{1}{(\Delta x_i)^2}.$$

Method of least squares. For a straight line $y = mx + c$ the gradient of the best fit to the data (x_i, y_i) with equal weights is

$$m = \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2}$$

and the y-intercept is

$$c = \overline{y} - m\overline{x}.$$

The uncertainties on these quantities are

$$\Delta m \approx \sqrt{\frac{1}{D} \frac{\sum d_i^2}{n-2}},$$

and

$$\Delta c \approx \sqrt{\left(\frac{1}{n} + \frac{\overline{x}^2}{D}\right) \frac{\sum d_i^2}{n-2}}$$

where $d_i = y_i - mx_i - c$ and $D = \sum(x_i - \overline{x})^2$.

If the weightings $y_i = \frac{1}{(\Delta y_i)^2}$ are not equal, then the best values are

$$m = \frac{\sum w_i \sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i}{\sum w_i \sum w_i x_i^2 - \left(\sum w_i x_i\right)^2}$$

and

$$c = \sum w_i y_i - m \sum w_i x_i.$$

and the uncertainties are:

$$\Delta m = \sqrt{\frac{1}{n-2} \frac{\sum w_i d_i^2}{D}}$$

and

$$\Delta c = \sqrt{\frac{1}{n-2} \left(\frac{D}{\sum w_i} + \overline{x}^2 \right) \frac{\sum d_i^2}{D}}$$