1

Experimental errors

1.1 Why estimate errors?

When performing experiments at school, we usually considered that the job was over once we obtained a numerical value for the quantity we were trying to measure. At university, and even more so in everyday situations in the laboratory, we are concerned not only with the answer but also with its accuracy. This accuracy is expressed by quoting an experimental error on the quantity of interest. Thus a determination of the acceleration due to gravity in our laboratory might yield an answer

$$g = (9.70 \pm 0.15) \text{ m/s}^2.$$

In Section 1.4, we will say more specifically what we mean by the error of ±0.15. At this stage it is sufficient to state that the more accurate the experiment the smaller the error; and that the numerical value of the error gives an indication of how far from the true answer this particular experiment may be.

The reason we are so insistent on every measurement including an error estimate is as follows. Scientists are rarely interested in measurement for its own sake, but more often will use it to test a theory, to compare with other experiments measuring the same quantity, to use this parameter to help predict the result of a different experiment, and so on. Then the numerical value of the error becomes crucial in the interpretation of the result.

For example, maybe we measured the acceleration due to gravity in

order to compare it with the value of 9.81 m/s^2,* measured in another laboratory a few miles away last year. We could be doing this in order to see whether there had been some dramatic change in the gravitational constant G over the intervening period; to try to detect a large gold mine which could affect the gravitational field in our neighbourhood; to find out if the earth had stopped spinning (although there are easier ways of doing this); to discover the existence of a new force in nature which could make the period of a pendulum depend on the local topography, etc.

With a measurement of 9.70 m/s^2, do we have evidence for a discrepancy? There are essentially three possibilities.

Possibility 1

If as suggested above the experimental error is ±0.15, then our determination looks satisfactorily in agreement with the expected value,

i.e. 9.70 ± 0.15 is consistent with 9.81.

Possibility 2

If we had performed a much more accurate experiment and had succeeded in reducing the experimental error to ±0.01, then our measurement is inconsistent with the previous value. Hence, we should worry whether our experimental result and/or the error estimate are wrong. Alternatively, we may have made a world shattering discovery.

i.e. 9.70 ± 0.01 is inconsistent with 9.81.

Possibility 3

If we had been stupid enough to time only one swing of the pendulum, then the error on g could have been as large as ±5. Our result is now consistent with expectation, but the accuracy is so low that it would be incapable of detecting even quite significant differences.

i.e. 9.70 ± 5 is consistent with 9.81,
and with many other values too.

Thus for a given result of our experiment, our reaction – 'Our measurement is in good shape' OR 'We have made a world shattering discovery' OR 'We should find out how to do better experiments' – depends on the

---

* Since this is an experimental number, it too has an uncertainty, but we assume that it has been measured so well that we can effectively forget about it here.

numerical estimate of the accuracy of our experiment. Conversely, if we know only that the result of the experiment is that the value of g was determined as 9.70 m/s^2 (but do not know the value of the experimental error), then we are completely unable to judge the significance of this result.

The moral is clear. Whenever you determine a parameter, estimate the error or your experiment is useless.

A similar remark applies to 'null measurements'. These occur in situations where you investigate whether changing the conditions of an experiment affects its result. For example, if you increase the amplitude of swing of your pendulum, does the period change? If, to the accuracy with which you can make measurements, you see no effect, it is tempting to record that 'No change was seen'. However this in itself is not a helpful statement. It may become important at some later stage to know whether the period was constant to within 1%, or perhaps within 1 part in a million. Thus, for example, the period is expected to depend slightly on the amplitude of swing, and we may be interested to know whether our observations are consistent with the expected change. Alternatively we may need to know how accurate the pendulum is as a clock, given that its amplitude is sometimes 10° and at others 5°. With simply the statement 'No change was seen', we have no idea at all of what magnitude of variation of the period could be ruled out. It is thus essential in these situations to give an idea of the maximum change that we would have been capable of detecting. This could consist of a statement like 'No change was observed; the maximum possible change in period was less than 1 part in 300'.

It is worth remembering that null measurements, with sufficiently good limits on the possible change, have sometimes led to real progress. Thus, at the end of the last century, Michelson and Morley performed an experiment to measure the speed of the earth through the hypothesised aether. This would have produced shifts in the optical interference fringe pattern produced in their apparatus. They observed no such shift, and the limit they were able to place on the effect was sufficiently stringent that the idea of the aether was discarded. The absence of an aether was one of the cornerstones on which Einstein's Special Theory of Relativity was built.

Thus 'null observations' can be far from useless, provided you specify what the maximum possible value of the effect could have been.

1.2 Random and systematic errors
1.2.1 What they are

There are two fundamentally different sorts of errors associated with any measurement procedure, namely random (or statistical) and systematic errors. Random errors come from the inability of any measuring device (and the scientist using it) to give infinitely accurate answers.* Another source of random errors is the fluctuations that occur in observations on a small sample drawn from a large population. On the other hand, systematic errors result in measurements that for one reason or another are simply wrong. Thus when we make a series of repeated measurements, the effect of random errors is to produce a spread of answers scattered around the true value. In contrast, systematic errors can cause the measurements to be offset from the correct value, even though the individual results can be consistent with each other. (See Fig. 1.1.)

Thus, for example, suppose someone asks you the exact time. You look at your watch, which has only hour and minute hands, but no second hand. So when you try to estimate the time, you will have a random error of something of the order of a minute. You certainly would have extreme difficulty trying to be precise to the nearest second. In addition to this random error, there may well be systematic errors too. For example, your watch may be running slow, so that it is wrong by an amount that you are not aware of but may in fact be 10 minutes. Again, you may recently have come back home to England from Switzerland, and forgotten to reset your watch, so that it is out by 1 hour. As is apparent from this example, the random error is easier to estimate, but there is the danger that if you are not careful you may be completely unaware of the more important systematic effects.

As a more laboratory oriented example, we now consider an experiment designed to measure the value of an unknown resistor, whose resistance R_2 is determined as

$$R_2 = \frac{V_2 - V_1}{V_1} R_1 \tag{1.1}$$

(see Fig. 1.2). Thus we have to measure the voltages V_1 and V_2, and the

---

* Except possibly for the situation where we are measuring something that is integral (e.g., the number of cosmic rays passing through a small detector during one minute). See, however, the next sentence of the text, and the remarks about Poisson distributions in Section 1.2.2.
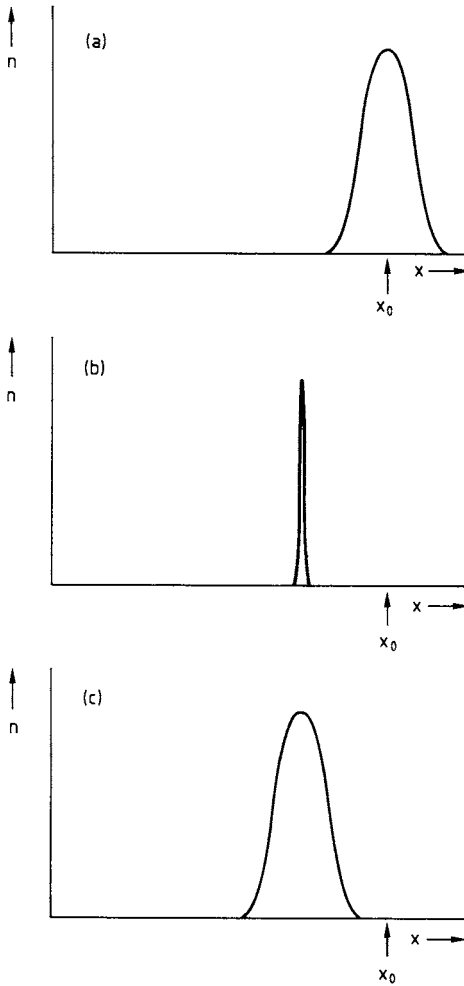
Fig. 1.1. Random and systematic errors. The figures show the results of repeated measurements of some quantity x whose true value is shown by the arrows. The effect of random errors is to produce a spread of measurements, centred on x_o (see (a)). On the other hand, systematic effects (b) can shift the results, while not necessarily producing a spread. Finally, the effect of random and systematic errors, shown in (c), is to produce a distribution of answers, centred away from x_o.

other resistance R_1. The random errors are those associated with the measurements of these quantities.
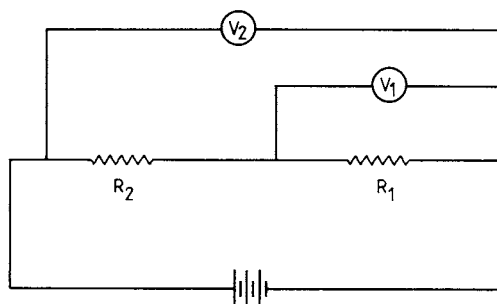
Fig. 1.2. Circuit for determining an unknown resistance R_2 in terms
of a known one R_1 and the two voltages V_1 and V_2.

The most obvious sources of systematic errors are the following.

(i)   The meters or oscilloscopes that we are using to measure V_1 and V_2
       may be incorrectly calibrated. How this affects the answer depends
       on whether the same device is used to measure the two voltages.
       (See section 1.8.)
(ii)   The meter used to measure the resistor R_1 may similarly be in
       error.
(iii) If our voltage source were AC, then stray capacitances and/or in-
       ductances could affect our answer.
(iv)   The resistors may be temperature dependent, and our measure-
       ment may be made under conditions which differ from those for
       which we are interested in the answer.
(v)   The impedances of the voltmeters may not be large enough for
       the validity of the approximation that the currents through the
       resistors are the same.
(vi)   Electrical pick-up could affect the readings of the voltmeters.

Systematic errors can thus arise on any of the actual measurements
that are required in order to calculate the final answer (e.g. points (i) and
(ii) above). Alternatively, they can be due to more indirect causes; thus
effects (iii)–(vi) are produced not by our instruments being incorrect,
but more by the fact that we are not measuring exactly what we are
supposed to.

In other situations it might be that there are implicit assumptions in
the derivation of the equation on which we are relying for obtaining our
answer. For example, the period of a pendulum of length l is $2\pi\sqrt{l/g}$
only if the amplitude of oscillations is small, if we can neglect air resis-

tance, if the top of the pendulum is rigidly secured, etc. If these are not true, then our experiment has a systematic error. Whether such effects are significant or not depends on their magnitude compared with those of the random errors.

1.2.2 Estimating random errors

A recurring theme in this book is the necessity of providing error estimates on any measured quantity. Because of their nature, random errors will make themselves apparent by producing somewhat different values of the measured parameter in a series of repeated measurements. The estimated accuracy of the parameter can then be obtained from the spread in measured values as described in Section 1.4.

An alternative method of estimating the accuracy of the answer exists in cases where the spread of measurements arises because of the limited accuracy of measuring devices. The estimates of the uncertainties of such individual measurements can be combined as explained in Section 1.7 in order to derive the uncertainty of the final calculated parameter. This approach can be used in situations where a repeated set of measurements is not available for the method described in the previous paragraph. In cases where both approaches can be used, they should of course yield consistent answers.

The accuracy of our measurements will in general play little part in determining the accuracy of the final parameter in those situations in which the measurements are made on a population which exhibits its own natural spread of values. For example, the heights of ten-year-old children are scattered by an amount which is larger than the uncertainty with which the height of any single child can be measured. It is then this scatter and the sample size which determine the accuracy of the answer.

A similar situation arises where the observation consists in counting independent random events in a given interval. The spread of values will usually be larger than the accuracy of counting (which may well be exact); for an expected number of observations n, the spread is $\sqrt{n}$. This can be derived from the properties of the Poisson distribution, which is discussed in Appendix 4.

Another example is provided by the measurement of the mean lifetime τ of a radioactive element. This we can do by finding the average of the observed decay times of a sample of the disintegrations. The nature of radioactivity is such that not all decays occur at the identical time τ, but in fact a large number would follow an exponential distribution (see
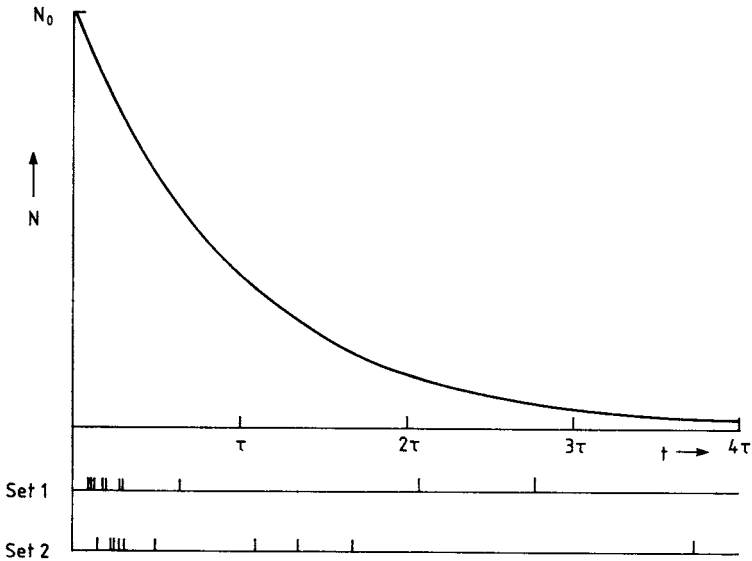
Fig. 1.3. An exponential $N = N_o\, e^{-t/\tau}$, for the expected distribution of decay times t of radioactive disintegrations of a source of mean lifetime τ. The bars below the t axis give two possible sets of observed decay times in experiments where each detected ten decays. The means of these times for the two samples are 0.68τ and 0.96τ. They differ from τ because of the statistical fluctuations associated with small samples.

Fig. 1.3). Thus the observed times for a small number of decays could fluctuate significantly if we repeated the experiment. This variation is a random effect, and is not connected with the accuracy with which we can measure individual decay times, which could be very much better than τ.

1.2.3 Worrying about systematic errors

For systematic errors, the 'repeated measurement' approach will not work; if our ohmeter is reading in kilohms while we think it is in ohms, the resistance will come out too small by a factor of $\sim$1000 each time we repeat the experiment, and yet everything will look consistent.

Ideally, of course, all systematic effects should be absent. But if it is thought that such a distortion may be present, then at least some attempt can be made to estimate its importance and to correct for it. Thus if we suspect a systematic error on the ohmeter, we can check it by

measuring some known resistors. Alternatively, if we are worried that the amplitude of our pendulum is too large, we can measure the period for different initial displacements, and then extrapolate our answer to the limit of a very small amplitude. In effect, we are then converting what was previously a systematic error into what is hopefully only a random one.

One possible check that can sometimes be helpful is to use constraints that may be relevant to the particular problem. For example, we may want to know whether a certain protractor has been correctly calibrated. One possible test is to use this protractor to measure the sum of the angles of a triangle. If our answer differs significantly from 180°, our protractor may be in error.

In general, there are no simple rules or prescriptions for eliminating systematic errors. To a large extent it requires common sense plus experience to know what are the possible dangerous sources of errors of this type.

Random errors are usually more amenable to methodical study, and the rest of this chapter is largely devoted to them. Nevertheless, it is important to remember that in many situations the accuracy of a measurement is dominated by the possible systematic error of the instrument, rather than by the precision with which you can actually make the reading.

Finally we assert that a good experimentalist is one who minimises and realistically estimates the random errors of his apparatus, while reducing the effect of systematic errors to a much smaller level.

1.3 Distributions

In Section 1.6 we are going to consider in more detail what is meant by the error σ on a measurement. However, since this is related to the concept of the spread of values obtained from a set of repeated measurements, whose distribution will often resemble a Gaussian (or normal) distribution, we will first have three mathematical digressions into the subjects of (a) distributions in general, (b) the mean and variance of a distribution, and (c) the Gaussian distribution.

A distribution $n(x)$ will describe how often a value of the variable x occurs in a defined sample. The variable x could be continuous or discrete, and its values could be confined to a finite range (e.g. 0–1) or

Table 1.1. *Examples of distributions*

Character	Limits	x variable	$n(x)$
	$1 \rightarrow \infty$	Integer x	No. of times you have produced a completely debugged computer program after x compilations
Discrete	$1 \rightarrow 7$	Day of week	No. of marriages on day x
	-13.6 eV $\rightarrow 0$	Energies of ground and excited states of hydrogen atoms	No. of atoms with electrons in state of energy x in atomic hydrogen at $30000°$
	$-\infty \rightarrow \infty$	Measured value of parameter	No. of times measurment x is observed
Continuous	$0 \rightarrow \infty$	Time it takes to solve all problems in this book	No. of readers taking time x
	$0 \rightarrow 24$ hours	Hours sleep each night	No. of people sleeping for time x

could extend to $\pm\infty$ (or could occupy a semi-infinite range, e.g. positive values only). Some examples are given in Table 1.1.

As an example, Fig. 1.4 shows possible distributions of a continuous variable, the height h of 30-year-old men. If only a few values are available, the data can be presented by marking a bar along the h axis for each measurement (see Fig. 1.4(a)). In Fig. 1.4(b), the same data is shown as a histogram, where a fairly wide bin size for h is used and the vertical axis is labelled as n, the number of observations per centimetre interval of h, despite the fact that the bin size Δh used is 10 cm. The actual number of men corresponding to a given bin is $n\Delta h$, and the total number of men appearing in the histogram is $\sum n\Delta h$. If 100 times more measurements were available, the number of entries in each bin of the histogram would increase by a large factor (Fig. 1.4(c)), but it would now become sensible to draw the histogram with smaller bins, in order to display the shape of the distribution with better resolution. Because