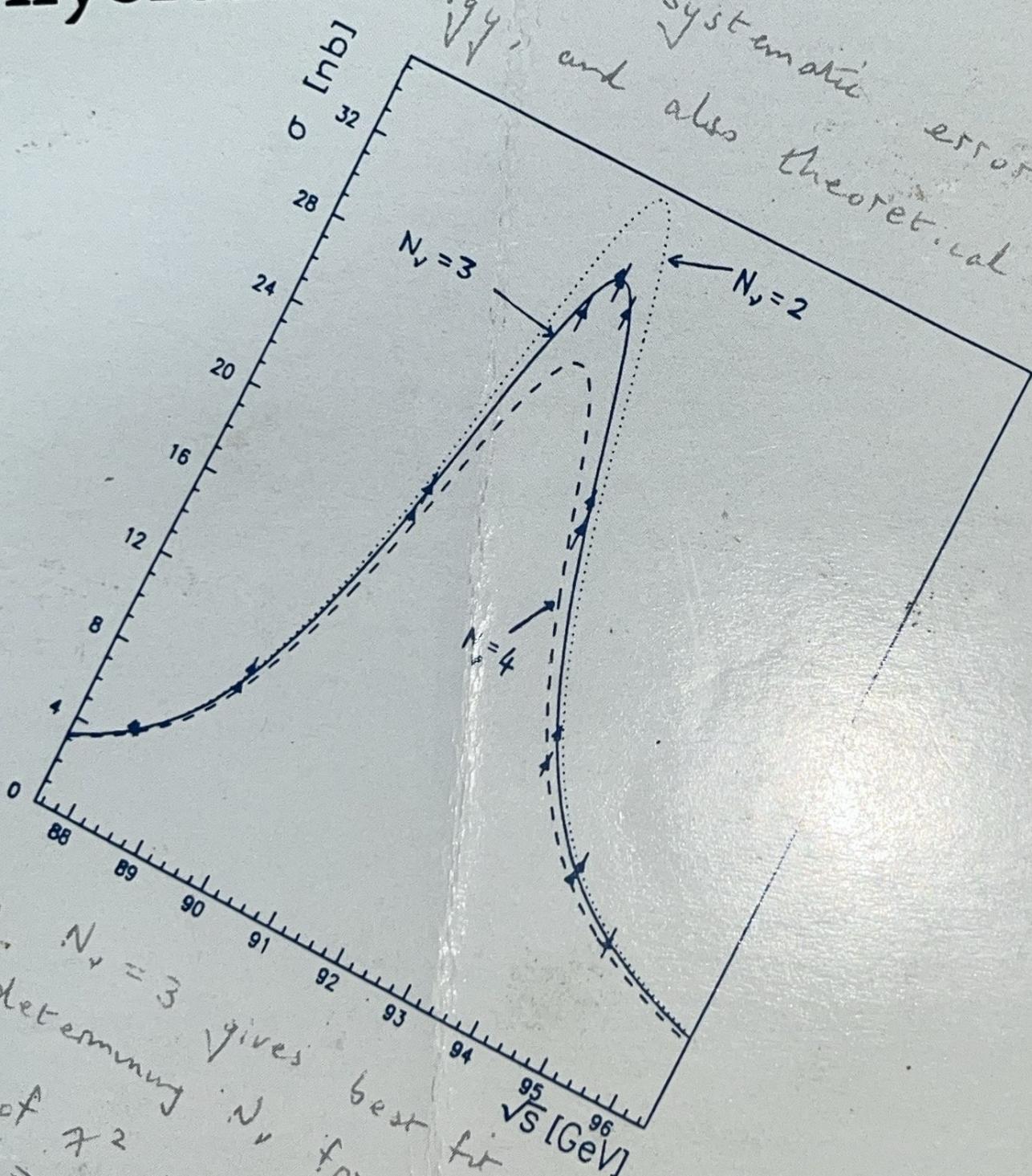


A PRACTICAL GUIDE TO Data Analysis for Physical Science Students



Louis Lyons

Contents

<i>Preface</i>	xii
<i>Glossary and Conventions</i>	xiii
1 Experimental errors	
1.1 Why estimate errors?	1
1.2 Random and systematic errors	4
1.2.1 What they are	4
1.2.2 Estimating random errors	7
1.2.3 Worrying about systematic errors	8
1.3 Distributions	9
1.4 Mean and variance	12
1.5 Gaussian distribution	16
1.6 The meaning of σ	18
1.7 Combining errors	21
1.7.1 Linear situations	22
1.7.2 Products	25
1.7.3 The general case	26
1.8 Systematic errors	27
1.9 An example including random and systematic errors	29
1.10 Combining results of different experiments	31
1.11 Worked examples	33
1.11.1 Mean and variance	33
1.11.2 Using a Gaussian distribution	34
1.11.3 Central Limit Theorem	34
1.11.4 Combining errors	35
1.11.5 Combining results	36

1.12 Does it feel right?

*Problems***2 Least squares fitting**

- 2.1 What are we trying to do?
 2.2 Weighted sum of squares
 2.3 Determining the parameters
 2.4 The error on the gradient and intercept
 2.4.1 Using the errors σ_i
 2.4.2 Using the scatter of points
 2.5 Other examples
 2.6 y^{obs} as numbers
 2.7 Parameter testing
 2.8 Distribution testing
 2.9 Worked example of straight line fit
 2.10 Summary of straight line fitting

*Problems**Appendix 1 Useful formulae**Appendix 2 Partial differentiation**Appendix 3 The binomial distribution**Appendix 4 The Poisson distribution**Appendix 5 Student's t distribution**Appendix 6 Statistical tables**Appendix 7 Random numbers**Index*

Preface

This short book is intended to be a practical guide, providing sets of rules that will help you to analyse the data you collect in your regular experimental sessions in the laboratory. Even more important, explanations and examples are provided to help you understand the ideas behind the formulae. Emphasis is also placed on thinking about the answers that you obtain, and on helping you get a feeling for whether they are sensible.

In contrast, this does not set out to be a text on statistics, and certainly not to be a complete course on the subject. Also, no attempt is made to provide rigorous mathematical proofs of many of the required formulae. These are important, and if required can be consulted in any standard textbook on the subject.

I believe that it will be necessary to read this material more than once. You really need to have understood the ideas involved before you do your first practical; but on the other hand, it would be much easier to absorb the material after you have actually done a couple of experiments and grappled with problems of trying to do the analysis yourself. Thus it is a good idea to read the book quickly, so that you at least discover what topics are covered and where to find them again when you need them. At this stage, you need not worry if not everything is entirely comprehensible. Then you take the book with you into your practicals, so that you can refer to it for help with each of your early calculations. As you become experienced, you will need to consult it less and less. However, it is a good idea to return to reading the whole book, this time aiming to understand it almost completely.

Since data analysis is a very practical subject, the only way to become proficient is to be involved in actually doing it. To help you achieve this, there are a few problems at the end of each chapter. You are strongly recommended to solve them. Omitting them is analogous to trying to learn to swim by reading a book on the subject without ever getting into the water.

This book is a slightly extended version of a short series of lectures I have been giving to first year Oxford Physics students. The material for these lectures was based on experience I have gained over several years by supervising and marking the weekly practical work of undergraduates here. It is clear that calculating the accuracy of an experimental result often becomes the most difficult part of the assignment. Many students regard the necessary calculations as obscure and complicated. The aim of this book is to dispel these ideas, to make it clear what the correct procedure is, to help you avoid excessively long calculations and to enable you to realise when your calculations have yielded ridiculous answers.

I hope that this book will be a useful companion that will assist you with your error calculations and data analysis, and to stop them being a chore. In this way, you should be able to enjoy your practical work, and to devote your energy to understanding the basic physical ideas involved.

I am grateful to the Numerical Algorithms Group (NAG) for permission to use routines from their program library for producing the Tables in Appendices 6 and 7.

Louis Lyons
Oxford, 1991

Glossary and Conventions

μ = true mean of a distribution (or of a large population)

\bar{x} = estimated mean for a sample

σ^2 = true variance of a distribution (or of a large population)

s^2 = variance as estimated from the spread of observations in a sample

u^2 = variance of the estimated mean

Experimental results are usually quoted in the form $y \pm \varepsilon$. The estimate y is said to be unbiassed provided that, were we to repeat our experiment N times, the average of the y values for the whole set of measurements would tend to the true (but generally unknown) value y_0 as N becomes larger and larger. The quantity ε is our estimate of the standard deviation (RMS deviation from the mean) of the distribution of results that we would expect if we repeated the experiment with similar apparatus many times. We refer to ε as 'the error' on the result. Occasionally results are given in the form $y \pm \varepsilon \pm \delta$. Here ε is the contribution from the random (or statistical) error, and δ that from systematic effects.

For a result like 20 ± 1 cm, we refer to the 1 cm as the 'absolute error', while the 'fractional error' is $1/20 = 5\%$.

The average of y is denoted by \bar{y} or $\langle y \rangle$.

The sign \sum means that we have to perform a summation. Thus

$$\sum_{i=1}^N y_i = y_1 + y_2 + \cdots + y_N.$$

Then the average $\bar{y} = \sum y_i / N$.

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

$$R_2 = \frac{V_2 - V_1}{V_1} R_1 \quad (1.1)$$

(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

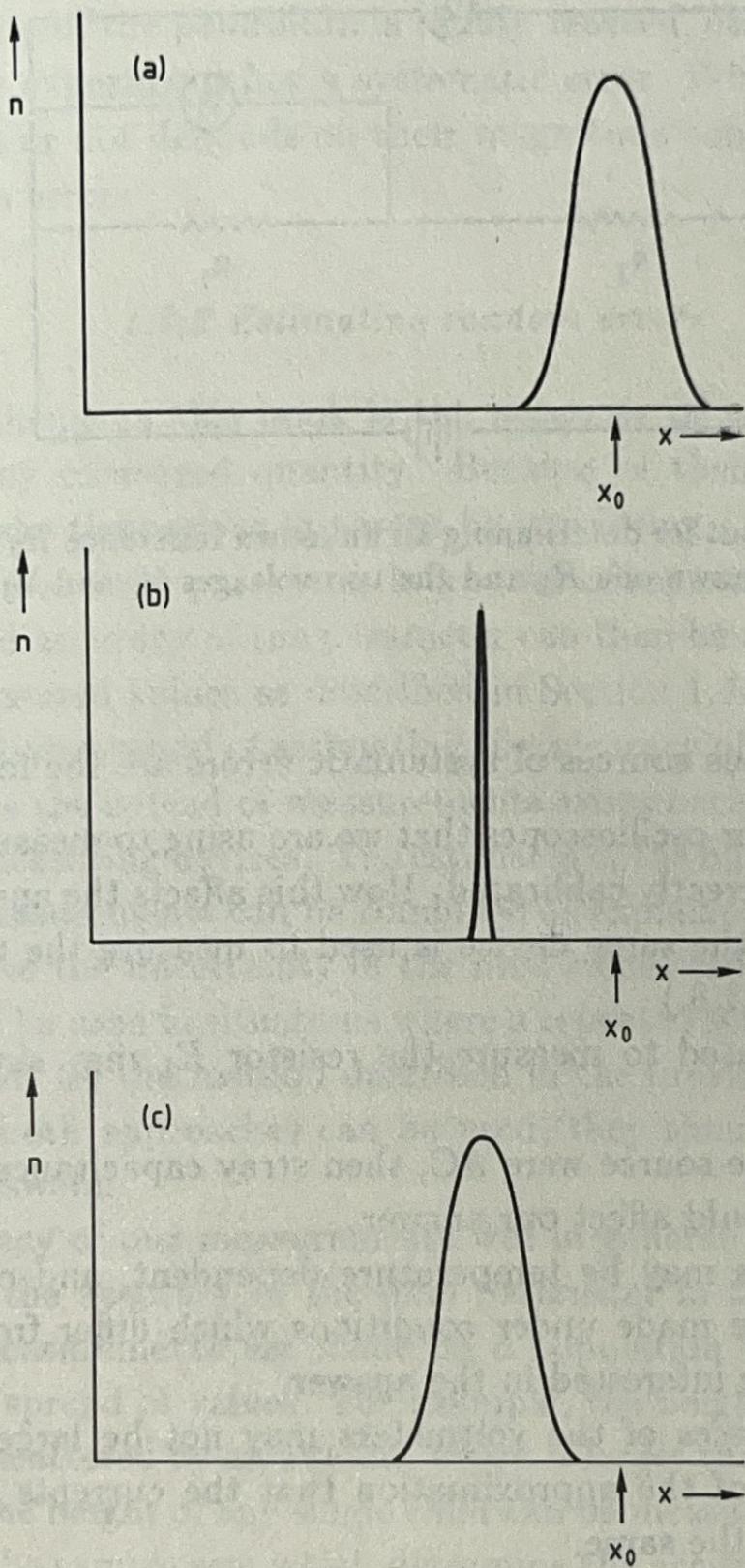


Fig. 1.1. Random and systematic errors. The figures show the results of repeated measurements of some quantity x whose true value is shown by the arrows. The effect of random errors is to produce a spread of measurements, centred on x_0 (see (a)). On the other hand, systematic effects (b) can shift the 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_0 .

other resistance R_1 . The random errors are those associated with the measurements of these quantities.

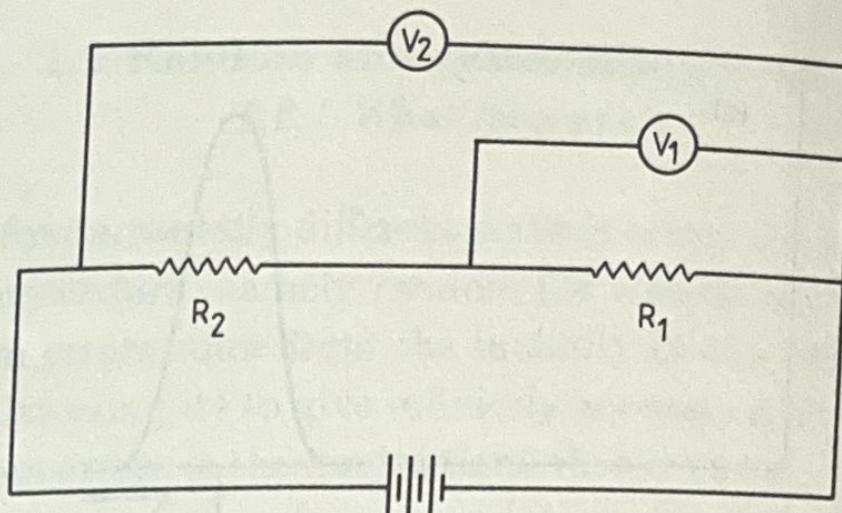


Fig. 1.2. Circuit for determining an unknown resistance R_2 in terms of a known one R_1 and the two voltages V_1 and V_2 .

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 inductances could affect our answer.
- (iv) The resistors may be temperature dependent, and our measurement 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

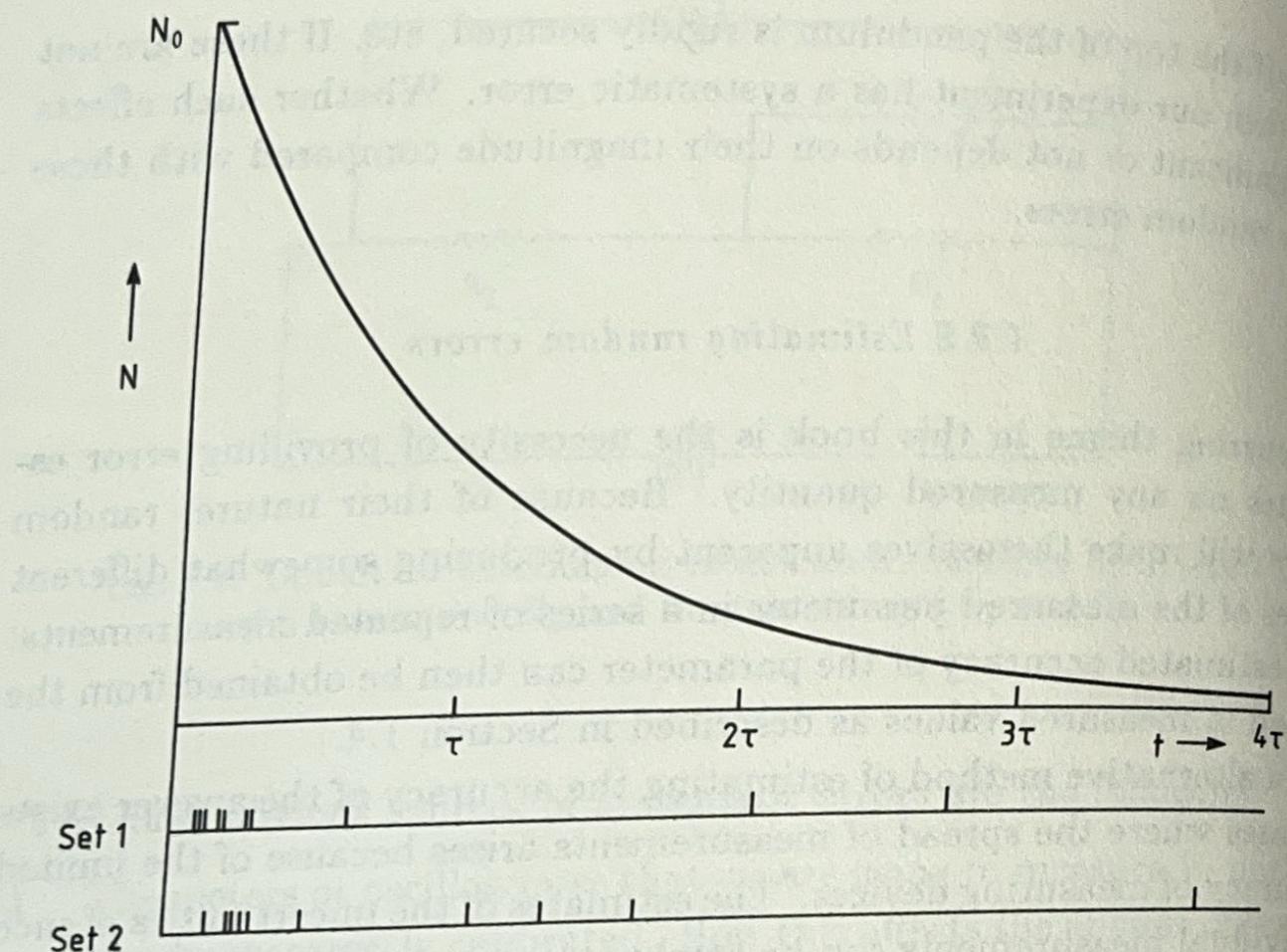


Fig. 1.3. An exponential $N = N_0 e^{-t/\tau}$, for the expected distribution of decay times t of radioactive disintegrations of a source of mean lifetime τ . The bars below the t axis give two possible sets of observed decay times in experiments where each detected ten decays. The means of these times 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 ~ 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
	$-13.6 \text{ eV} \rightarrow 0$	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 measurement x is observed
Continuous	$0 \rightarrow \infty$	Time it takes to solve all problems in this book	No. of readers taking time x
	$0 \rightarrow 24 \text{ hours}$	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

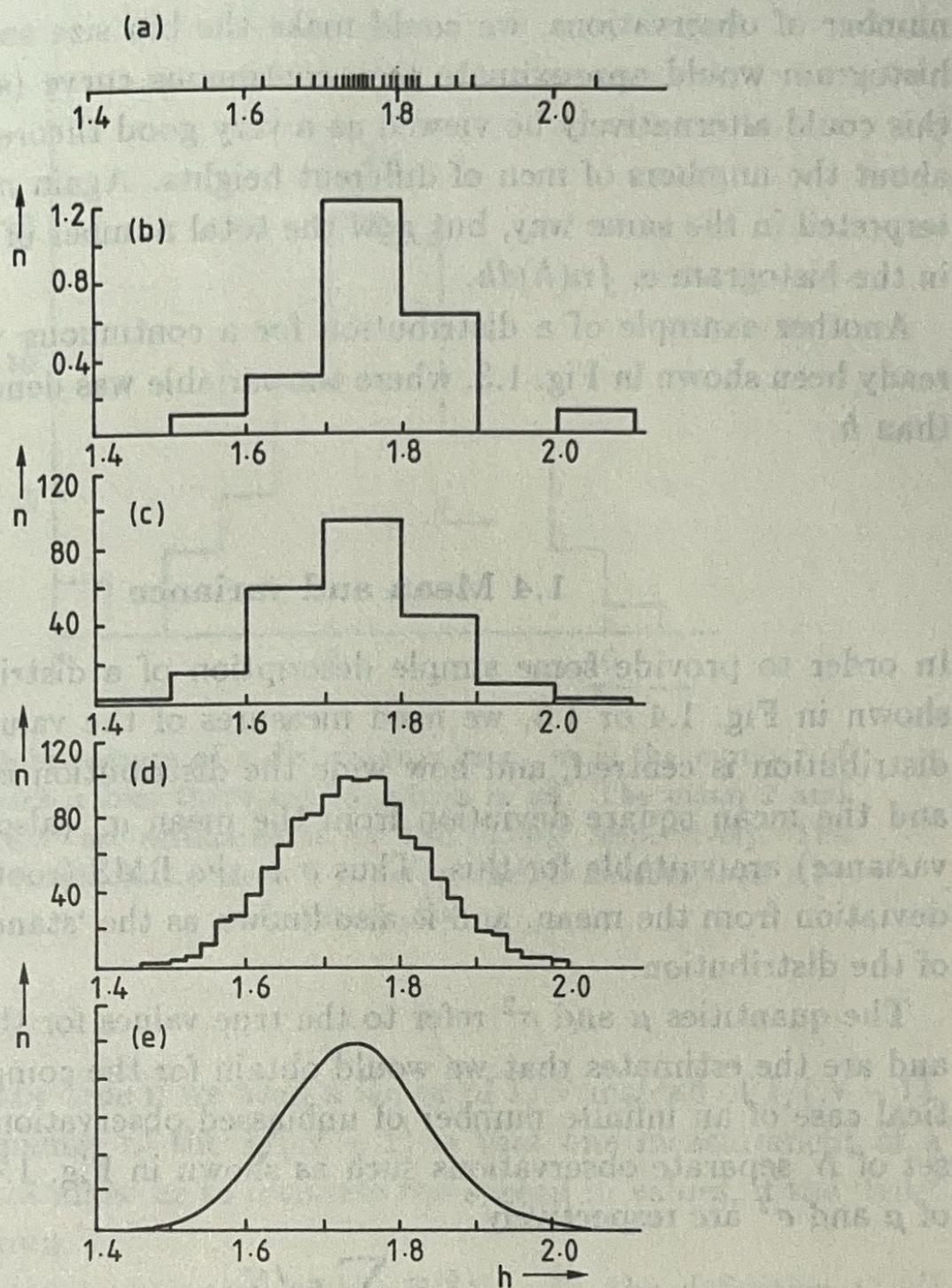


Fig. 1.4. Some examples of distributions of, say, the heights h (in metres) of 30-year-old men. (a) With only a few observations, each one is represented as a bar at the relevant position along the h axis. (b) The data of (a) could alternatively be drawn as a histogram, where n is the number of men per centimetre interval of h , even though the bin size in h is 10 cm. (c) A histogram as in (b), but with 100 times more observations. (d) The same data as in (c), but drawn with smaller bins of h . n is still the number of men per centimetre interval of h . (e) For an even larger number of observations and with smaller bin size, the histogram of (d) approaches a continuous distribution.

we plot $n(h)$ as the number of observations per centimetre, irrespective of bin size, the overall height of the histogram does not change much when we change the bin size (see Fig. 1.4(d)). Finally, for an even larger

number of observations, we could make the bin size so small that the histogram would approximate to a continuous curve (see Fig. 1.4(e)); this could alternatively be viewed as a very good theoretical prediction about the numbers of men of different heights. Again $n(h)$ is to be interpreted in the same way, but now the total number of men appearing in the histogram is $\int n(h)dh$.

Another example of a distribution for a continuous variable has already been shown in Fig. 1.3, where the variable was denoted by t rather than h .

1.4 Mean and variance

In order to provide some simple description of a distribution such as shown in Fig. 1.4 or 1.5, we need measures of the value at which the distribution is centred, and how wide the distribution is. The mean μ and the mean square deviation from the mean σ^2 (also known as the variance) are suitable for this. Thus σ is the RMS (root mean square) deviation from the mean, and is also known as the 'standard deviation' of the distribution.

The quantities μ and σ^2 refer to the true values for the distribution, and are the estimates that we would obtain for the completely impractical case of an infinite number of unbiased observations. For a finite set of N separate observations such as shown in Fig. 1.4(a), estimates of μ and σ^2 are respectively

$$\bar{x} = \sum x_i/N \quad (1.2)$$

and

$$s^2 = \sum (x_i - \mu)^2/N. \quad (1.3)$$

where the \sum signs mean that we must add up the specified variables for all the N members of the distribution. In general, the true mean μ is not known, and so eqn (1.3) cannot in fact be used to estimate the variance. Instead it is replaced by

$$s^2 = \frac{1}{N-1} \sum (x_i - \bar{x})^2, \quad (1.3')$$

where the factor $1/(N-1)$ is required in order to make s^2 an unbiased estimator of the variance (as can be proved fairly readily). This means that, if we repeat our procedure of taking samples lots and lots of times (say L times in all), then we expect that the average of s^2 from all these samples gets closer and closer to the correct value σ^2 as L increases.

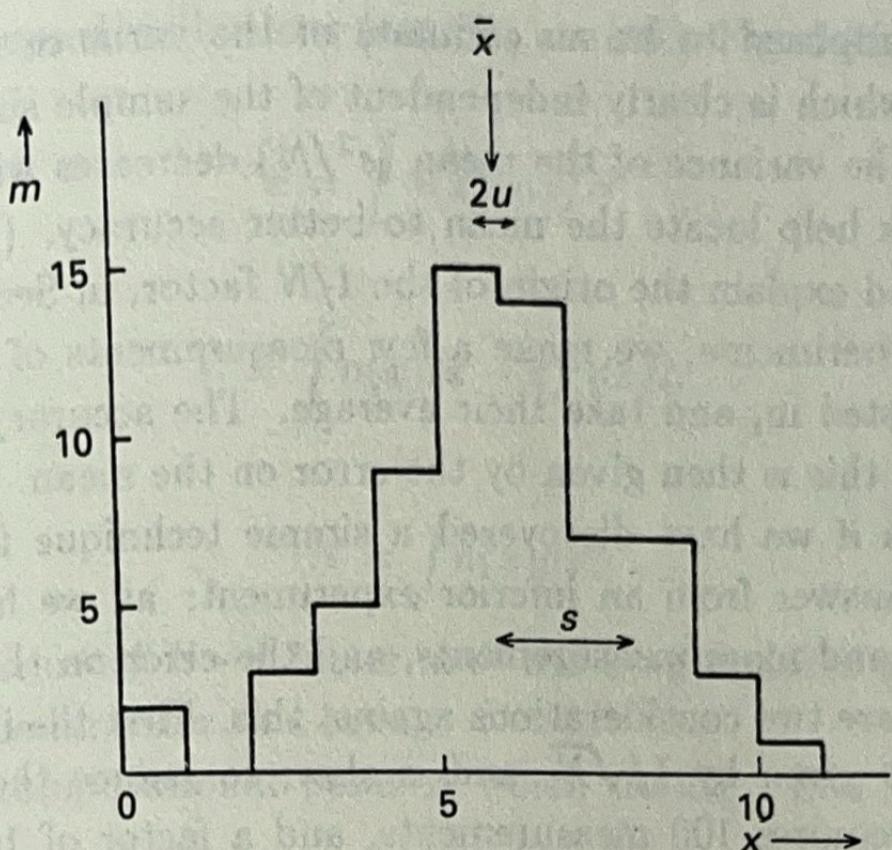


Fig. 1.5. A histogram of a distribution in x . m is the number of entries in each x bin; there are 66 entries in all. The mean \bar{x} and the variance s^2 are estimated as 5.9 and $(2.05)^2$ respectively. The accuracy u to which the mean \bar{x} is determined is smaller than s by a factor of $\sqrt{66}$.

This would not be true if we used a factor of $1/N$ instead of $1/(N - 1)$. Another consequence of the $1/(N - 1)$ is that one measurement of a quantity does not allow us to estimate the spread in values, if the ‘true’ value is not known.

Even if we accept the need for the $1/(N - 1)$, the definition (1.3') for s^2 still looks unnecessarily complicated. Since the deviation of an individual measurement from the estimated mean is $(x_i - \bar{x})$, we might have thought that $y = \sum(x_i - \bar{x})/(N - 1)$ would have given a simpler estimate of the width of the distribution. The trouble with this is that the definition of \bar{x} is such that $\sum(x_i - \bar{x})$ is guaranteed to be zero, and so y is useless as an estimate of anything. We could instead have used $\sum|x_i - \bar{x}|/(N - 1)$, but the modulus sign is rather messy from a mathematical viewpoint, so the definition (1.3') is adopted.

It is most important to realise that s is the measure of how spread out the distribution is, and is not the accuracy to which the mean \bar{x} is determined. This is known to an accuracy better by a factor of \sqrt{N} . Thus by taking more and more observations of x , the variance s^2 will not change (apart from fluctuations) since the numerator and denominator of eqn (1.3) or (1.3') grow more or less proportionally; this is sensible

since s^2 is supposed to be an estimate of the variance of the overall population, which is clearly independent of the sample size N . On the other hand, the variance of the mean (s^2/N) decreases with increasing N ; more data help locate the mean to better accuracy. (We return to this point, and explain the origin of the $1/N$ factor, in Section 1.7.1).

In some experiments, we make a few measurements of the quantity we are interested in, and take their average. The accuracy with which we determine this is then given by the error on the mean.

It seems as if we have discovered a simple technique for obtaining an accurate answer from an inferior experiment: all we have to do is to take more and more measurements, and the error on the mean goes down. There are two considerations against this. First the improvement in accuracy is given by $1/\sqrt{N}$, and is slow; to reduce the error by a factor of 10 requires 100 measurements, and a factor of 1000 needs a million repetitions, which in most circumstances is completely impractical. Secondly, it is true that, provided nothing significant changes during the course of this tedious procedure, the statistical error on the mean does decrease as specified. However, all experiments are in danger of having systematic errors as well as random ones. In a well-designed experiment, the systematic error is usually smaller than the random one. Now a repeated set of measurements reduces the statistical error but not the systematic one.* Thus as N increases we reach the point where the systematic error dominates the random error on the mean, and then further repetition of the measurements is of little value. Similarly undetected systematic errors can produce a bias, which again will usually not decrease as the number of measurements increases.

Sometimes the measurements are grouped together so that at the value x_j there are m_j observations (equivalent to $n_j \Delta h$ in Fig. 1.4(d)). Then simple extensions of eqns (1.2) and (1.3') are

$$\bar{x} = \sum m_j x_j / \sum m_j \quad (1.4)$$

and

$$s^2 = \sum m_j (x_j - \bar{x})^2 / (\sum m_j - 1), \quad (1.5)$$

where the summation now runs over the j bins of the grouped histogram. As usual, s^2 is our estimate of the variance of the distribution; the variance u^2 on the mean is $s^2 / \sum m_j$ (see Fig. 1.5).

* Of course, with extra data, it may be possible to look at potential systematic effects, and to discover how to reduce the error from these sources too, but that would require a lot more thought and work.

For continuous distributions (see, for example Fig. 1.4(e)), these become

$$\bar{x} = \int n(x)xdx/N \quad (1.6)$$

and

$$s^2 = \int n(x)(x - \bar{x})^2 dx/N, \quad (1.7)$$

where

$$N = \int n(x)dx,$$

and where the usual $N - 1$ factor in s^2 has been replaced by N which is assumed to be large for this case.

A minor computational point is worth noting. Eqn (1.3') can be written

$$s^2 = \frac{N}{N-1}(\bar{x^2} - \bar{x}^2), \quad (1.8)$$

where $\bar{x^2}$ is defined in analogy with eqn (1.2) as

$$\bar{x^2} = \sum x_i^2/N. \quad (1.9)$$

Thus if someone were reading out the data to you (or if you were accepting the data on a computer), it is not necessary for this to be done twice, first for you to calculate \bar{x} , and then to obtain s^2 from eqn (1.3'). Instead $\bar{x^2}$ and \bar{x} can be calculated in a single pass over the data, and then s^2 calculated from eqn (1.8) at the end. If your pocket calculator has the ability to compute standard deviations, it is likely that it does it this way.

However, in using eqn (1.8) for cases where s^2 is small compared with $\bar{x^2}$ and \bar{x}^2 , it is vital to keep enough significant figures in the numerical calculation, or the obtained value of s^2 can be meaningless. The x values of 9500, 9501 and 9502 provide an example of this.

A final but important point is that, if we attempt to determine the width of a distribution from only a few measurements, our estimate s^2 from eqn (1.3') will not be very accurate. This is because the particular small sample that we take may have the individual measurements accidentally closer together than might have been expected, or they may be unusually far apart; as the sample size increases, such effects become less likely. In fact, for a set of results that are Gaussian distributed, the fractional error on s is $1/\sqrt{2n-2}$. Thus, for example, with 9 measurements our error estimate is known to only 25%, and there is no sense in quoting more than one significant figure. (See Problem 1.10).

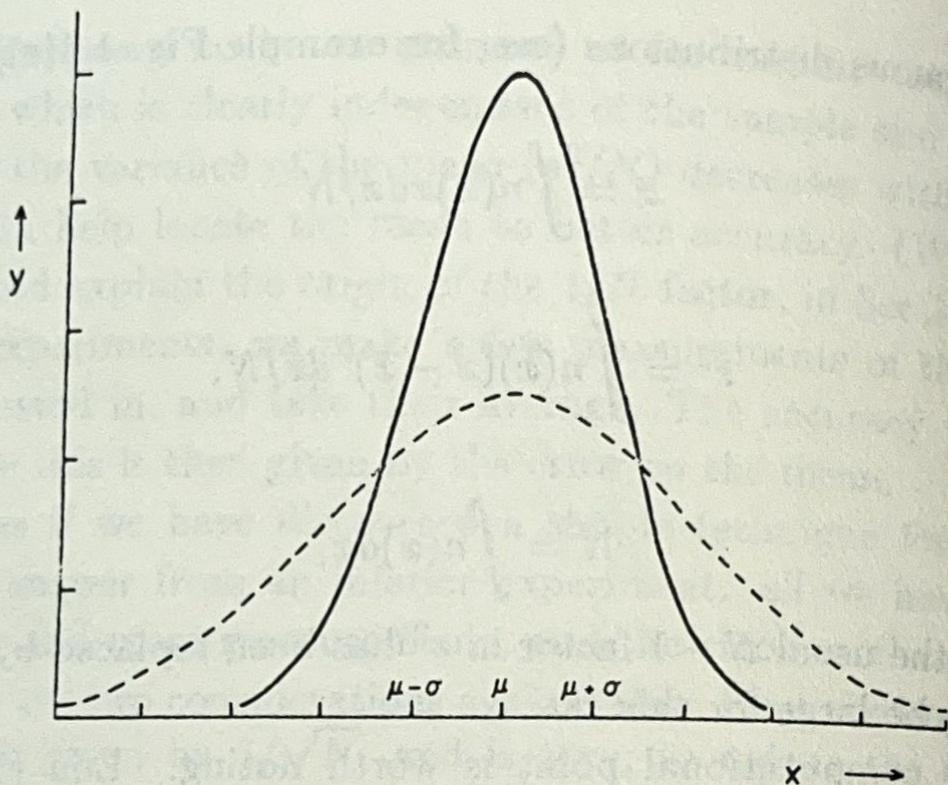


Fig. 1.6. The solid curve is the Gaussian distribution of eqn (1.10). The distribution peaks at the mean μ , and its width is characterised by the parameter σ . The dashed curve is another Gaussian distribution with the same value of μ , but with σ twice as large as the solid curve. Because the normalisation condition (1.11) ensures that the areas under the curves are the same, the height of the dashed curve is only half that of the solid curve at the maxima. The scale on the x axis refers to the solid curve.

1.5 Gaussian distribution

As the Gaussian distribution is of such fundamental importance in the treatment of errors, we now consider some of its properties.

The general form of the Gaussian distribution in one variable x is

$$y = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-(x-\mu)^2/2\sigma^2\}. \quad (1.10)$$

The curve of y as a function of x is symmetric about the value of $x = \mu$, at which point y has its maximum value. (See Fig. 1.6.) The parameter σ characterises the width of the distribution, while the factor $(\sqrt{2\pi}\sigma)^{-1}$ ensures that the distribution is normalised to have unit area underneath the whole curve, i.e.

$$\int_{-\infty}^{+\infty} y dx = 1. \quad (1.11)$$

The parameter μ is the mean of the distribution, while σ has the following properties.

- (i) The mean square deviation of the distribution from its mean is σ^2 . (This is the reason that the curious factor of 2 appears within the

exponent in eqn (1.10). Otherwise the root mean square deviation from the mean would have been $\sigma/\sqrt{2}$, which is unaesthetic.)

- (ii) The height of the curve at $x = \mu \pm \sigma$ is $1/\sqrt{e}$ of the maximum value. Since

$$1/\sqrt{e} = 0.607,$$

σ is very roughly the half width at half height of the distribution.

- (iii) The fractional area underneath the curve and with

$$\mu - \sigma \leq x \leq \mu + \sigma \quad (1.12)$$

(i.e. within $\pm\sigma$ of the mean μ) is 0.68.

- (iv) The height of the distribution at its maximum is $(\sqrt{2\pi}\sigma)^{-1}$. As σ decreases the distribution becomes narrower, and hence, to maintain the normalisation condition eqn (1.11), also higher at the peak.

By a suitable change of variable to

$$x' = (x - \mu)/\sigma \quad (1.13)$$

any normal distribution can be transformed into a standardised form

$$y = \frac{1}{\sqrt{2\pi}} \exp(-x'^2/2), \quad (1.14)$$

with mean zero and unit variance.

A feature which helps to make the Gaussian distribution of such widespread relevance is the Central Limit Theorem. One statement of this is as follows. Consider a set of n independent variables x_i , taken at random from a population with mean μ and variance σ^2 , and then calculate the mean \bar{x} of these n values. If we repeat this procedure many times, since the individual x_i are random, then \bar{x} itself will have some distribution. The surprising fact is that, for large n , the distribution of \bar{x} tends to a Gaussian (of mean μ and variance σ^2/n). The distribution of the x_i themselves is irrelevant. The only important feature is that the variance σ^2 should be finite. If the x_i are already Gaussian distributed, then \bar{x} is also Gaussian for all values of n from 1 upwards. But even if x_i is, say, uniformly distributed over a finite range, then the sum of a few x_i will already look Gaussian. Thus whatever the initial distribution, a linear combination of a few variables almost always approximates to a Gaussian distribution.

An example of the Central Limit Theorem is given below in Section 1.7.1.

1.6 The meaning of σ

Having concluded our mathematical digressions, we now return to our consideration of the treatment of errors.

For a large variety of situations, the result of repeating an experiment many times produces a spread of answers whose distribution is approximately Gaussian; the approximation is likely to be good especially if the individual errors that contribute to the final answer are small. When this is true, it is meaningless to speak of a 'maximum possible error' of a given experiment since the curve in Fig. 1.6 remains non-zero for all values of x ; the 'maximum possible error' would be infinite, and although this would make it easy to calculate the 'error' on any experiment, it would not distinguish a precision experiment from an inaccurate one.

It is thus customary to quote σ as the accuracy of a measurement. Since σ is not the maximum possible error, we should not get too upset if our measurement is more than σ away from the expected value. Indeed, we should expect this to happen with about $\frac{1}{3}$ of our experimental results. Since, however, the fractional areas beyond $\pm 2\sigma$ and beyond $\pm 3\sigma$ are only 5% and 0.3% respectively, we should expect such deviations to occur much less frequently.

The properties of Gaussian distributions are commonly used in interpreting the significance of experimental results. This is illustrated by the following example.

We measure the lifetime of the neutron in an experiment as 950 ± 20 seconds. A certain theory predicts that the lifetime is 910 s. To what extent are these numbers in agreement?

We consult Fig. 1.7, which is a graph showing the fractional area under the Gaussian curve with

$$|f| > r, \quad (1.15)$$

where

$$f = \frac{x - \mu}{\sigma}, \quad (1.16)$$

i.e. it gives (on the right hand vertical scale) the area in the tails of the Gaussian beyond any value r of the parameter f , which is plotted on the horizontal axis. In our example of the neutron lifetime, $f = 2$ and the corresponding probability is 4.6%. Thus if 1000 experiments of the same precision as ours were performed to measure the neutron lifetime, and if our theory is correct, and if the experiments are bias-free, then we expect about 46 of them to differ from the predicted value by at least as much as ours does. Of course, we still have to make up our mind

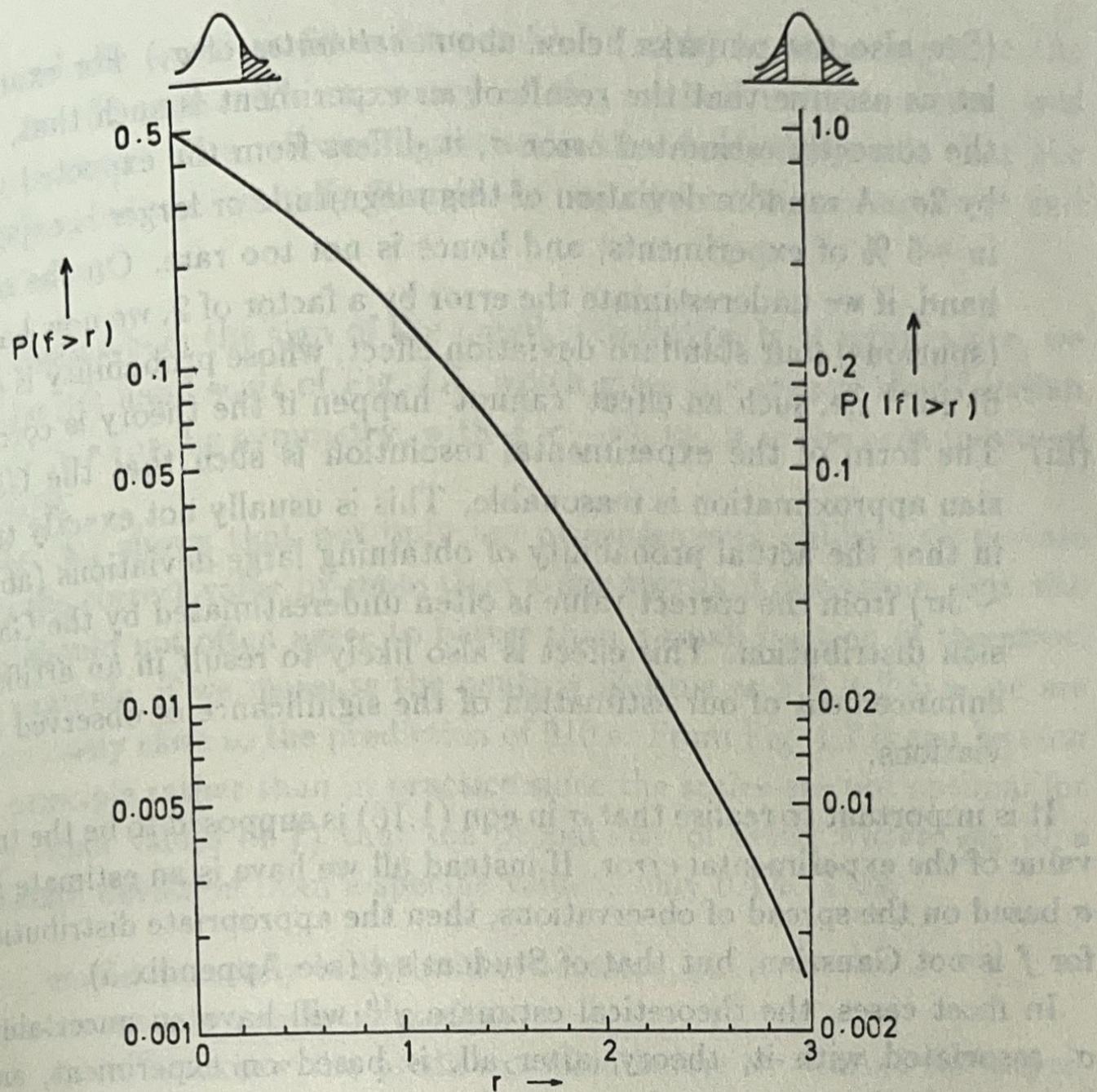


Fig. 1.7. The fractional area in the tails of a Gaussian distribution, i.e. the area with f greater than some specified value r , where f is the distance from the mean, measured in units of the standard deviation. The scale on the left hand vertical axis refers to the one-sided tail, while the right hand one is for $|f|$ larger than r . Thus for $r = 0$, the probabilities are $\frac{1}{2}$ and 1 respectively.

whether we regard the theory (and the experiment) as satisfactory or not, but at least we have a number on which to base our judgement.

In order to rely on the numerical value of this probability, it is essential to ensure that the following assumptions are satisfied.

- (i) The value of the quantity of interest has been correctly calculated (e.g. there are no important systematic biasses).
- (ii) The magnitude of the error has been correctly calculated. This is particularly important, in that an incorrect estimate of the accuracy of the experiment could have a very large effect on the calculated significance of our result and hence on our conclusions.

- (See also the remarks below about *estimates* of σ .) For example, let us assume that the result of an experiment is such that, with the correctly estimated error σ , it differs from the expected value by 2σ . A random deviation of this magnitude or larger is expected in $\sim 5\%$ of experiments, and hence is not too rare. On the other hand, if we underestimate the error by a factor of 2, we now have a (spurious) four standard deviation effect, whose probability is only $6 \cdot 10^{-5}$ i.e. such an effect 'cannot' happen if the theory is correct.
- (iii) The form of the experimental resolution is such that the Gaussian approximation is reasonable. This is usually not exactly true, in that the actual probability of obtaining large deviations (above $\sim 3\sigma$) from the correct value is often underestimated by the Gaussian distribution. This effect is also likely to result in an artificial enhancement of our estimation of the significance of observed deviations.

It is important to realise that σ in eqn (1.16) is supposed to be the true value of the experimental error. If instead all we have is an estimate for σ based on the spread of observations, then the appropriate distribution for f is not Gaussian, but that of Student's t (see Appendix 5).

In most cases, the theoretical estimate y^{th} will have an uncertainty σ' associated with it; theory, after all, is based on experiment, and hence predictions in general are calculated from other measured quantities which of course have their own experimental errors. In that case, we repeat the above procedure of consulting Fig. 1.7, but we redefine f for this case as

$$f = \frac{y^{obs} - y^{th}}{\sqrt{\sigma^2 + \sigma'^2}}, \quad (1.17)$$

where our measured value is $y^{obs} \pm \sigma$. The denominator of (1.17) arises because it is the error on the numerator, assuming that the errors on y^{obs} and y^{th} are uncorrelated (see Section 1.7.1).

Sometimes we are interested in the sign of possible deviations from predicted values.

Example (i)

Is there any evidence for other processes, which we have not allowed for in our theory, contributing to the neutron decay? In other words, is the observed lifetime for neutron decay *smaller* than the predicted value?

Example (ii)

A motorist is accused of speeding by a policeman, who claims that he

measured the speed as 38 miles per hour (mph) in an area where the limit was 30 mph. The accuracy of the method used, however, was only ± 5 mph. The relevant question is thus how likely is it that the driver's speed was actually 30 mph *or less*, given the measurement and its accuracy.

In cases where the sign of the possible deviation is of significance, we use the left hand scale of Fig. 1.7, which gives the area of the Gaussian with $f > r$ (or, by symmetry, with $f < -r$), i.e. it is the area in one of the tails.

Fig. 1.7 shows that not only are measurements unlikely to deviate from the correct value by more than a few standard deviations, but also they should not often agree to better than a small fraction of the error. For example, if we measure the neutron lifetime as 909 ± 200 s, we are suspiciously close to the prediction of 910 s. From Fig. 1.7 it can be seen (in principle rather than in practice since the scales are not optimal for very small values of f) that the probability of being within $\frac{1}{200}$ of a standard deviation from a specific value is only 0.4%. Thus

- (i) we are unusually lucky on this occasion; or
- (ii) our error is over-estimated; or
- (iii) we in fact knew the predicted value before we started the experiment, and (perhaps unconsciously) adjusted our measurement to get close to the 'right' answer.

The discussion in this section has been an example of what is known as 'Hypothesis Testing'. We return to a more detailed discussion of this subject in Chapter 2.

1.7 Combining errors

We are frequently confronted with a situation where the result of an experiment is given in terms of two (or more) measurements. Then we want to know what is the error on the final answer in terms of the errors on the individual measurements. We first consider in detail the case where the answer is a linear combination of the measurements. Then we go on to consider products and quotients, and finally we deal with the general case.

1.7.1 Linear situations

As a very simple illustration, consider

$$a = b - c.$$

To find the error on a , we first differentiate

$$\delta a = \delta b - \delta c.$$

If we were talking about maximum possible errors, then we would simply add the magnitudes of δb and δc to get the maximum possible δa . But we have already decided that it is more sensible to consider the root mean square deviations. Then, provided that the errors on b and c are *uncorrelated*,* the rule is that we add the contributions δb and $-\delta c$ in quadrature:

$$\sigma_a^2 = \sigma_b^2 + \sigma_c^2.$$

Two points are worth noting.

- (i) If in a particular experiment we know that the measurements of b and c were incorrect by specific amounts δb and δc , then the answer would be incorrect by an amount δa , given in terms of δb and δc by eqn (1.19). But the whole point is that in any given measurement we do not know the exact values of δb and δc (or else we would simply correct for them, and get the answer for a exactly), but only know their mean square values σ^2 over a series of measurements. It is for these statistical errors that eqn (1.20) applies.
- (ii) For linear combinations like eqn (1.18), it is the errors themselves which occur in eqn (1.20); percentage errors, which are useful for products (see Section 1.7.2) are here completely irrelevant. Thus if you wish to determine your height by making independent measurements of the distances of your head and your feet from the centre of the earth, each to 1% accuracy, the final answer will not in general be within 1% of the correct answer; in fact, you may well get a result of -40 miles for your height.

Next we discuss why we use quadrature for combining these statistical errors. We look at this in several ways.

(a) Mnemonic non-proof

The errors on b and on $-c$ can be 'in phase' with each other to give

* The meaning of 'uncorrelated' becomes clearer later in this section.

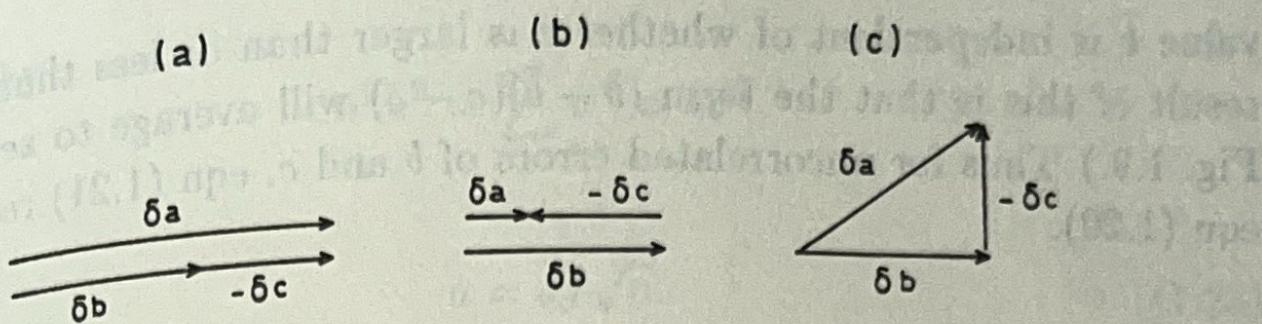


Fig. 1.8. Diagram illustrating the non-proof of formula (1.20). In (a) the contributions from δb and from $-\delta c$ are 'in phase', in (b) they are 'out of phase', while in (c) they appear to be in quadrature.

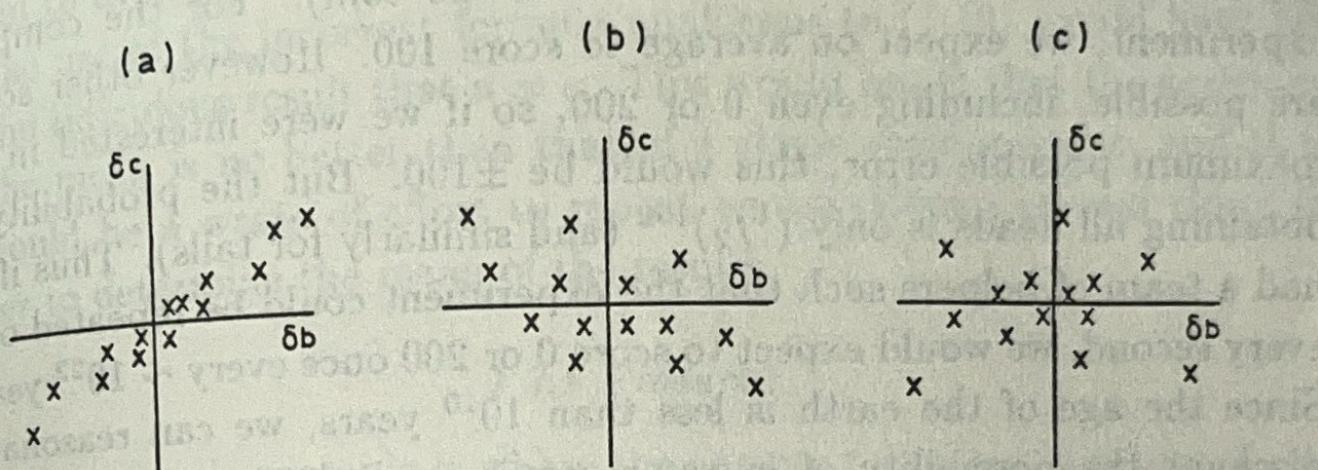


Fig. 1.9. The crosses represent the values of δb and δc for a repeated series of measurements. In (a), these errors are strongly correlated, with $\langle(b - \bar{b})(c - \bar{c})\rangle$ – known as the covariance – being large and positive. The correlation in (b) is less pronounced and slightly negative. In (c) there is almost no correlation, and the covariance is almost zero.

contributions which add up in δa ; or they can be 'out of phase', so that they partially cancel in δa . So perhaps on average they are 'orthogonal' to each other and hence Pythagoras' Theorem should be used for obtaining σ_a^2 . (See Fig. 1.8.)

We stress that this is not a proof; in particular there is no obvious second dimension in which δb and δc can achieve orthogonality.

(b) Formal proof

$$\begin{aligned}\sigma_a^2 &= \langle [a - \bar{a}]^2 \rangle \\ &= \langle [(b - c) - (\bar{b} - \bar{c})]^2 \rangle \\ &= \langle (b - \bar{b})^2 \rangle + \langle (c - \bar{c})^2 \rangle - 2\langle (b - \bar{b})(c - \bar{c}) \rangle.\end{aligned}\quad (1.21)$$

In the above line, the first two terms are σ_b^2 and σ_c^2 respectively. The last term depends on whether the errors on b and c are correlated. In most situations that we shall be considering, such correlations are absent. In that case, whether b is measured as being above or below its average

value \bar{b} is independent of whether c is larger than or less than \bar{c} . The result of this is that the term $(b - \bar{b})(c - \bar{c})$ will average to zero. (See Fig. 1.9.) Thus for uncorrelated errors of b and c , eqn (1.21) reduces to eqn (1.20).

(c) The infinitesimal probability argument

We perform an experiment which consists of tossing an unbiased coin 100 times. We score 0 for each heads and 2 for each tails (i.e. the expectation is 1 ± 1 each time we toss the coin). For the complete experiment, we expect on average to score 100. However other scores are possible, including even 0 or 200, so if we were interested in the maximum possible error, this would be ± 100 . But the probability of obtaining all heads is only $(1/2)^{100}$ (and similarly for tails). Thus if we had a team of helpers such that the experiment could be repeated once every second, we would expect to score 0 or 200 once every $\sim 10^{22}$ years. Since the age of the earth is less than 10^{10} years, we can reasonably discount the possibility of extreme scores, and thus consider instead what are the likely results.

The expected distribution for the final score follows the binomial distribution (see Appendix 3). For 100 tosses, this is very like the Gaussian distribution, with mean 100 and $\sigma \sim 10$. We thus have an example of the Central Limit Theorem mentioned in Section 1.5; by combining a large number N of variables, we end up with something very similar to a Gaussian distribution,* the width of which increases only like \sqrt{N} .

(d) Averaging is good for you

We know intuitively that it is better to take the average of several independent measurements of a single quantity than just to make do with a single observation. This follows from the correct formula (1.20), but not from simply adding the errors.

The average \bar{q} of n measurements q_i each of accuracy σ is given by

$$n\bar{q} = \sum_i q_i. \quad (1.22)$$

Then using (1.20) we deduce that the statistical error u on the mean

* Provided, of course, that we don't look at it with too great a resolution, since this distribution is defined only for integral values, whereas the Gaussian is continuous.

given by

$$n^2 u^2 = \sum_i \sigma_i^2 = n\sigma^2,$$

whence

$$u = \sigma / \sqrt{n}. \quad (1.23)$$

Thus we have obtained the result quoted in Section 1.4 that the error on the mean is known more accurately than the error characterising the distribution by a factor \sqrt{n} ; this justifies our intuitive feeling that it is useful to average.

The use of the incorrect formula analogous to (1.19) would have led to the ridiculous result that $u = \sigma$. This would imply that the accuracy of the mean is no better than that of a single measurement, and that it would be a waste of effort to repeat an experiment several times in order to determine the mean of the results.

1.7.2 Products

The next simple example is where the answer f is given in terms of the measurements x and y by

$$f = x^a y^b \quad (1.24)$$

where the powers a and b can be positive, negative or fractional. Thus this formula includes simple products, ratios, cubes, square roots, etc. etc.

As in the linear situation, we are going to differentiate this formula, but it is slightly simpler if we first take its logarithm. Then

$$\frac{\delta f}{f} = a \frac{\delta x}{x} + b \frac{\delta y}{y}. \quad (1.25)$$

Again in analogy with our earlier example, after we square and average, provided x and y are uncorrelated, we obtain

$$\left(\frac{\sigma_f}{f} \right)^2 = a^2 \left(\frac{\sigma_x}{x} \right)^2 + b^2 \left(\frac{\sigma_y}{y} \right)^2. \quad (1.26)$$

That is, the *fractional* error on f is simply related to the *fractional* errors on x and y . This contrasts with the linear case, where *absolute* errors were relevant.

The functions

$$f = xy$$

and $f = x/y$

are so common that it is worth writing the answer explicitly for them as

$$\left(\frac{\sigma_f}{f} \right)^2 = \left(\frac{\sigma_x}{x} \right)^2 + \left(\frac{\sigma_y}{y} \right)^2. \quad (1.27)$$

Thus a 3% error in x and a 4% error in y , assumed as usual to be uncorrelated, would combine to give a 5% error in f .

Because eqn (1.24) is in general not linear in x and y , eqn (1.26) will be accurate only if the fractional errors are small.

1.7.3 The General Case

There are two approaches that can be applied for a general formula

$$f = f(x_1, x_2, \dots, x_n) \quad (1.28)$$

which defines our answer f in terms of measured quantities x_i each with its own error σ_i . Again we assume the errors on the x_i are uncorrelated.

In the first, we differentiate and collect the terms in each independent variable x_i . This gives us*

$$\delta f = \frac{\partial f}{\partial x_1} \delta x_1 + \frac{\partial f}{\partial x_2} \delta x_2 + \dots + \frac{\partial f}{\partial x_n} \delta x_n. \quad (1.29)$$

As in our earlier examples, we then square and average over a whole series of measurements, at which point all the cross terms like $\overline{\delta x_1 \delta x_2}$ vanish because the different x_i are uncorrelated. We finally obtain

$$\sigma_f^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2. \quad (1.30)$$

This gives us the answer σ_f in terms of the known measurement errors σ_i . As with products and quotients, if f is non-linear in the x_i , this formula requires the errors σ_i to be small. (See problem 1.6.)

The alternative approach is applicable for any size errors. It consists of the following steps.

- (i) Calculate f_o as the value of f when all of the x_i are set equal to their measured values.
- (ii) Calculate the n values f_i , which are defined by

$$f_i = f(x_1, x_2, \dots, x_i + \sigma_i, \dots, x_n),$$

i.e. where the i th variable is increased from its measured value by its error.

- (iii) Finally obtain σ_f from

$$\sigma_f^2 = \sum (f_i - f_o)^2, \quad (1.31)$$

* The curly letter ∂ 's in eqn (1.29) (and later in (1.34)) mean that we should differentiate *partially* with respect to the relevant variable. Appendix 2 contains a brief explanation of partial differentiation.

i.e. we combine in quadrature all the individual deviations caused by moving each variable (one at a time) by its error. To the extent that the errors are small, this approach should give the same answer as the previous one. For larger errors, the numerical method will give more realistic estimates of the errors. Furthermore, by moving each variable in turn both upwards and downwards by its error, we can deduce upper and lower error estimates for f which need not be identical. Thus, if

$$f = \tan x$$

$$x = 88 \pm 1^\circ$$

$$f = 29^{+29}_{-10}$$

$$f = 29 \pm 14$$

from using eqn (1.30).

When the errors are asymmetric, it is a clear indication that the distribution of f is not Gaussian. Then we should be careful about how we calculate the significance of being, say, two or more errors away from some specified value.

1.8 Systematic errors

In Section 1.9, we shall consider the measurement of a resistance by the method discussed earlier in Section 1.2. We assume that the experiment produced the following results:

$$\left. \begin{aligned} R_1 &= (2.0 \pm 0.1 \text{ k}\Omega) \pm 1\%, \\ V_1 &= (1.00 \pm 0.02 \text{ volts}) \pm 10\%, \\ V_2 &= (1.30 \pm 0.02 \text{ volts}) \pm 10\%, \end{aligned} \right\} \quad (1.32)$$

where in each case the first errors are the random reading ones, and the second are the possible systematic errors in the various meters.

Although random and systematic errors are different in nature, we may want the overall error estimate as a single figure, rather than expressed separately as above. Then we should add them in quadrature, since reading and calibration errors are uncorrelated. This yields

$\pm 0.1 \text{ k}\Omega$, and $\pm 0.10 \text{ V}$ and $\pm 0.13 \text{ V}$ respectively.* When given in this way, however, we lose the distinction between the random and systematic components, which is important, as we shall immediately see.

Next let us consider how accurately we know $V_2 - V_1$. The answer depends on whether the same voltmeter was used to measure both V_1 and V_2 , or whether separate meters were employed. In the latter case, presumably there is no correlation between the two $\pm 10\%$ errors on the two readings, and so the linear combination

$$\begin{aligned} V_2 - V_1 &= (1.30 \pm 0.13) - (1.00 \pm 0.10) \\ &= 0.30 \pm 0.16 \text{ V}. \end{aligned} \quad (1.33)$$

In contrast, if the same meter is used, then it is clearly incorrect to assume that the two systematic errors are uncorrelated, since if the first measurement suffers from, say, a -7% calibration error, then so does the second. In this situation, a $\pm 10\%$ systematic uncertainty on each measurement will produce a $\pm 10\%$ systematic error on the result, i.e.

$$\begin{aligned} V_2 - V_1 &= [(1.30 \pm 0.02) - (1.00 \pm 0.02)] \pm 10\% \\ &= 0.30 \pm 0.03 \pm 10\% \\ &= 0.30 \pm 0.04 \text{ V}. \end{aligned} \quad (1.33')$$

It is perhaps a little surprising that for $V_2 - V_1$, which is a linear combination, we are considering *fractional* errors. This is because the uncertainty in the calibration is most simply expressed in terms of a scale factor $f = 1.00 \pm 10\%$. Then the true voltage V^t is given in terms of the measured voltage V^m as

$$V^t = V^m f$$

and the voltage difference

$$V_2^t - V_1^t = (V_2^m - V_1^m) f.$$

Thus we are in fact concerned with a product, which explains why fractional errors are relevant.

Alternatively, if the meter had some systematic zero error (which was the same for V_2 and V_1), its effect would exactly cancel in $V_2 - V_1$, and hence its magnitude would be completely irrelevant.

Finally we consider the voltage ratio V_2/V_1 . In the case where the same meter was used for V_1 and V_2 , its possible scale error of $\pm 10\%$ is irrelevant for V_2/V_1 as it cancels out. In fact it would not matter if

* Throughout this discussion, the magnitudes of errors are rounded to one or two significant figures. In fact only very rarely will it be worth quoting errors to more significant figures. (See Section 1.12).

Table 1.2.

For two measured voltages V_1 and V_2 , the table shows the effect of
 (i) a random reading error σ_1 or σ_2 ;
 (ii) a common scale error;
 (iii) a common zero error
 on the voltage difference and the voltage ratio.

Quantity	Random error		Systematic error	
	Reading error	Scale error	Zero error	
		$1 \pm \delta f$		δz
$V_2 - V_1$	$\sigma^2 = \sigma_1^2 + \sigma_2^2$	$\sigma = (V_2 - V_1) \delta f$	$\sigma = 0$	
$r = V_2/V_1$	$(\frac{\sigma}{r})^2 = (\frac{\sigma_1}{V_1})^2 + (\frac{\sigma_2}{V_2})^2$	$\sigma = 0$		$\sigma = \frac{V_2 + \delta z}{V_1 + \delta z} - \frac{V_2}{V_1}$

we were mistaken as to whether the scale was calibrated in millivolts or in volts; or we could read off the voltage in terms of the deflection in centimetres of an oscilloscope beam, without even considering the voltage sensitivity. The voltage ratio V_2/V_1 would not, however, be independent of any zero error (see Table 1.2).

1.9 An example including random and systematic errors

We now work through the case of determining the error on a resistance, whose value is calculated from eqn (1.1), with the measured values as given in eqn (1.32). In fact, the formula for R_2 is best rewritten as

$$R_2 = (V_2/V_1 - 1) R_1. \quad (1.1')$$

The necessary steps in calculating the error on R_2 are then as follows.

- (i) Since R_2 is the product of R_1 and $(V_2/V_1 - 1)$, the fractional error on R_2 is determined from the fractional errors on these quantities.
- (ii) The absolute error on $V_2/V_1 - 1$ is equal to the absolute error on V_2/V_1 .
- (iii) The fractional error on V_2/V_1 is determined from the fractional errors on V_1 and V_2 .

We now work through these steps in the reverse order, assuming that the same meter is used to measure V_1 and V_2 .

- (iii) First we need the error on V_2/V_1 . Since the systematic scale error cancels, we need consider only the random reading errors on V_1 and V_2 , and hence

$$\begin{aligned}\frac{V_2}{V_1} &= (1.30 \pm 0.02)/(1.00 \pm 0.02) \\ &= (1.30 \pm 2\%)/(1.00 \pm 2\%) \\ &= 1.30 \pm 3\% \\ &= 1.30 \pm 0.03.\end{aligned}$$

- (ii) The value of $(V_2/V_1 - 1)$ is 0.30 ± 0.03 . Thus although this is the same absolute error as that of V_2/V_1 , its fractional error is larger.
- (i) The value of R_2 is $0.60 \pm 0.07 \text{ k}\Omega$.

Had the two voltmeters been different, the systematic effects on V_1 and V_2 no longer cancel, and indeed dominate the errors on these voltages. Thus

$$\begin{aligned}V_2/V_1 &= (1.30 \pm 10\%)/(1.00 \pm 10\%) \\ &= 1.30 \pm 0.18.\end{aligned}$$

Then

$$V_2/V_1 - 1 = 0.30 \pm 0.18$$

and

$$R_2 = 0.60 \pm 0.36 \text{ k}\Omega.$$

This is of considerably lower accuracy, because we cannot ignore the calibration uncertainties of the voltmeters.

Instead of working through the combinations of linear functions and of products and quotients, we could have used our derivative formula (1.30) to obtain

$$\sigma^2(R_2) = \left(\frac{\partial R_2}{\partial V_1}\right)^2 \sigma^2(V_1) + \left(\frac{\partial R_2}{\partial V_2}\right)^2 \sigma^2(V_2) + \left(\frac{\partial R_2}{\partial R_1}\right)^2 \sigma^2(R_1) \quad (1.34)$$

where the partial derivatives are evaluated from the equation defining R_2 (eqn (1.1)). Again we would have to be careful about whether the calibration errors on V_1 and V_2 cancelled or not.

We might wonder whether we could have obtained the error on R_2 from eqn (1.1), which consists of the three factors $V_2 - V_1$, R_1 and $1/V_1$. The error on $V_2 - V_1$ we obtained in Section 1.8 earlier; those on R_1 and V_1 are known. Then why cannot we combine the three fractional errors to obtain the fractional error on R_2 ? The reason is that, regardless of the question of whether V_1 and V_2 are measured by the same meter or

not, the error on V_1 cannot be uncorrelated with that of $V_2 - V_1$, since the same measurement V_1 occurs in both. Thus it would be incorrect to use any of the formulae for combining errors which assume the separate components have uncorrelated errors. This is a very important general point: if the same measurement occurs more than once in a formula, it is wrong to assume that they can be treated as having independent errors. Both the previous methods of this section avoid this problem.

1.10 Combining results of different experiments

When several experiments measure the same physical quantity and give a set of answers a_i with different errors σ_i , then the best estimates of a and its accuracy σ are given by

$$a = \sum(a_i/\sigma_i^2) / \sum(1/\sigma_i^2) \quad (1.35)$$

and

$$1/\sigma^2 = \sum(1/\sigma_i^2). \quad (1.36)$$

Thus each experiment is to be weighted by a factor $1/\sigma_i^2$. In some sense, $1/\sigma_i^2$ gives a measure of the information content of that particular experiment. The proof of eqns (1.35) and (1.36) is left as an exercise for the reader (see problem 2.1).

We now give some examples of the use of these formulae.

Example (i)

The simplest case is when all the errors σ_i are equal. Then the best combined value a from eqn (1.35) becomes the ordinary average of the individual measurements a_i , and the error σ on a is σ_i/\sqrt{N} , where N is the number of measurements. This is all very sensible.

Now there is an entirely different way of estimating the error on the average of the set of results. This consists of simply using the spread of the separate determinations to calculate s (their root mean square deviation from the mean – see eqn (1.3')), and then the required error is s/\sqrt{N} . While this approach ignores the accuracies σ_i with which the individual measurements have been made, eqn (1.36) pays no regard to the degree to which these determinations are mutually consistent. Thus σ_i/\sqrt{N} can be regarded as the theoretical error that we expect on the basis of the accuracies of each measurement, while s/\sqrt{N} is an experimental measurement of how spread out are the separate values of a_i .

Which method should we use for determining the error on the mean? If the errors σ_i are estimated correctly and the measurements a_i are unbiased, then σ_i/\sqrt{N} and s/\sqrt{N} should agree with each other satisfactorily. The problem with s is that, especially for small values of N , fluctuations can have a serious effect on its value. Thus some people adopt the strategy of choosing the larger of σ_i/\sqrt{N} and s/\sqrt{N} for the error on the mean. My preference is to use σ_i/\sqrt{N} provided the two values are reasonably in agreement, and to worry about the consistency of the measurements if s is significantly larger than σ_i .

Discussion of how well the two estimates of the error of the mean should agree, and also the extension of the above example to the situation where the individual errors are unequal, is best dealt with by the χ^2 technique (see Chapter 2).

Example (ii)

The eqns (1.35) and (1.36) are in agreement with common sense when the separate experiments have achieved different accuracies σ_i by using the same apparatus but by averaging different numbers n_i of repeated measurements. In this case the σ_i are proportional to $1/\sqrt{n_i}$ (see Example (d) in Section 1.7.1).

Then the eqns become

$$a = \sum n_i a_i / \sum n_i \quad (1.37)$$

and

$$N = \sum n_i. \quad (1.38)$$

The first of these says that each of the original measurements of equal accuracy is to be weighted equally; the second is self-evident. (See problem 1.5.)

Example (iii)

We are trying to determine the number of married people in a country, using the following estimates:

$$\begin{aligned} \text{Number of married men} &= 10.0 \pm 0.5 \text{ million}, \\ \text{Number of married women} &= 8 \pm 3 \text{ million}. \end{aligned}$$

Then the total is 18 ± 3 million, where we have obtained the error on the sum as described in Section 1.7.1.

If, however, we assume that the numbers of married men and women are equal, each provides an estimate of half the required answer. Then

use eqns (1.35) and (1.36) to determine the married population as ± 1 million. Thus we see that calculating the sum or the average of experimental quantities are not equivalent to each other, as the latter implicitly assumes that the quantities we are averaging are supposed to be equal. This extra information results in improved accuracy for the answer.

1.11 Worked examples

We now present some simple worked examples that illustrate the earlier sections of this chapter. The confident student can proceed directly to Section 1.12.

1.11.1 Mean and variance

The following are measurements in grams of the mass of an insect, all with equal accuracy:

5.0 5.3 5.9 5.3 5.2 5.7 5.4 5.1 4.8 5.3

What is our best estimate of its mass, and how accurate is this?

We determine the average of the measurements from eqn (1.2). We first calculate $\sum m_i = 53.0$ g, where m_i are the individual measurements. Then

$$\bar{m} = \sum m_i / 10 = 5.3 \text{ g}$$

Next we use eqn (1.3') to estimate s^2 , the variance of the distribution of the measurements. The individual deviations δ_i from the estimated mean of 5.3 g are

-0.3 0.0 0.6 0.0 -0.1 0.4 0.1 -0.2 -0.5 0.0

(As a check, we note that the sum of these deviations is zero, as it must be.) Then

$$s^2 = \frac{1}{9} \sum \delta_i^2 = 0.91/9 = 0.10 \text{ g}^2$$

Thus our estimate of the width of the distribution is $\sqrt{s^2}$, which is 0.3 g. Numerically, this seems not unreasonable when we look at the individual deviations from the mean. Finally the error on the mean is better than this by a factor of $\sqrt{10}$, so we quote our answer as

$$5.3 \pm 0.1 \text{ g.}$$

1.11.2 Using a Gaussian distribution

An experiment measures the current gain of a transistor as 201 ± 19 . The expected value is 177 ± 9 . Assuming that both these are Gaussian distributed, how consistent are the measurements?

The difference between the measured and expected values is 24, and the error on the difference is $\sqrt{12^2 + 9^2} = 15$. Thus we want to know how likely it is that, if the true difference were zero and we performed an experiment with an accuracy of ± 15 , the observed difference would deviate from zero by at least 24. This discrepancy is 1.6 standard deviations, and from tables of the Gaussian distribution we find that the fractional area beyond 1.6σ is 10%. Thus in about 1 experiment in 10, we would expect a random fluctuation to give us a deviation of at least this magnitude, assuming the two values really are perfectly consistent. This probability does not seem too small, so we are likely to be happy to accept the two measurements as agreeing with each other. (We really should have decided beforehand on the cut-off level for an acceptable probability.)

1.11.3 Central Limit Theorem

Illustrate the Central Limit Theorem as follows. Add together four random numbers r_i , each distributed uniformly and independently in the range 0 to 1 (taken from Appendix 7 at the end of the book, or from your pocket calculator), to obtain a new variable $z = \sum r_i$. Repeat the procedure 50 times in all, to obtain a set of 50 z_j values. Plot these as a histogram, and compare it with the appropriate Gaussian distribution.

Using the first 200 random numbers of Appendix 7, we obtain the following z_j values:

1.759 2.161 2.150 2.896..... 2.792 0.834

These are drawn as a histogram in Fig. 1.10. Also shown there is the curve

$$y = \frac{50}{\sqrt{2\pi}\sigma} \exp \left\{ -(z - \mu)^2 / 2\sigma^2 \right\} \Delta z \quad (1.39)$$

where y is the number of entries per bin of width Δz , and the mean μ and width σ are chosen according to the paragraph below eqn (1.14) as 2 and $\sqrt{4/12}$ ($\sqrt{1/12}$ being the RMS of the uniform x distribution of width 1 – see problem 1.2(c) – and 4 being the number of x values added to construct z). Agreement between the histogram and the curve would be expected only if the number of x values added together were

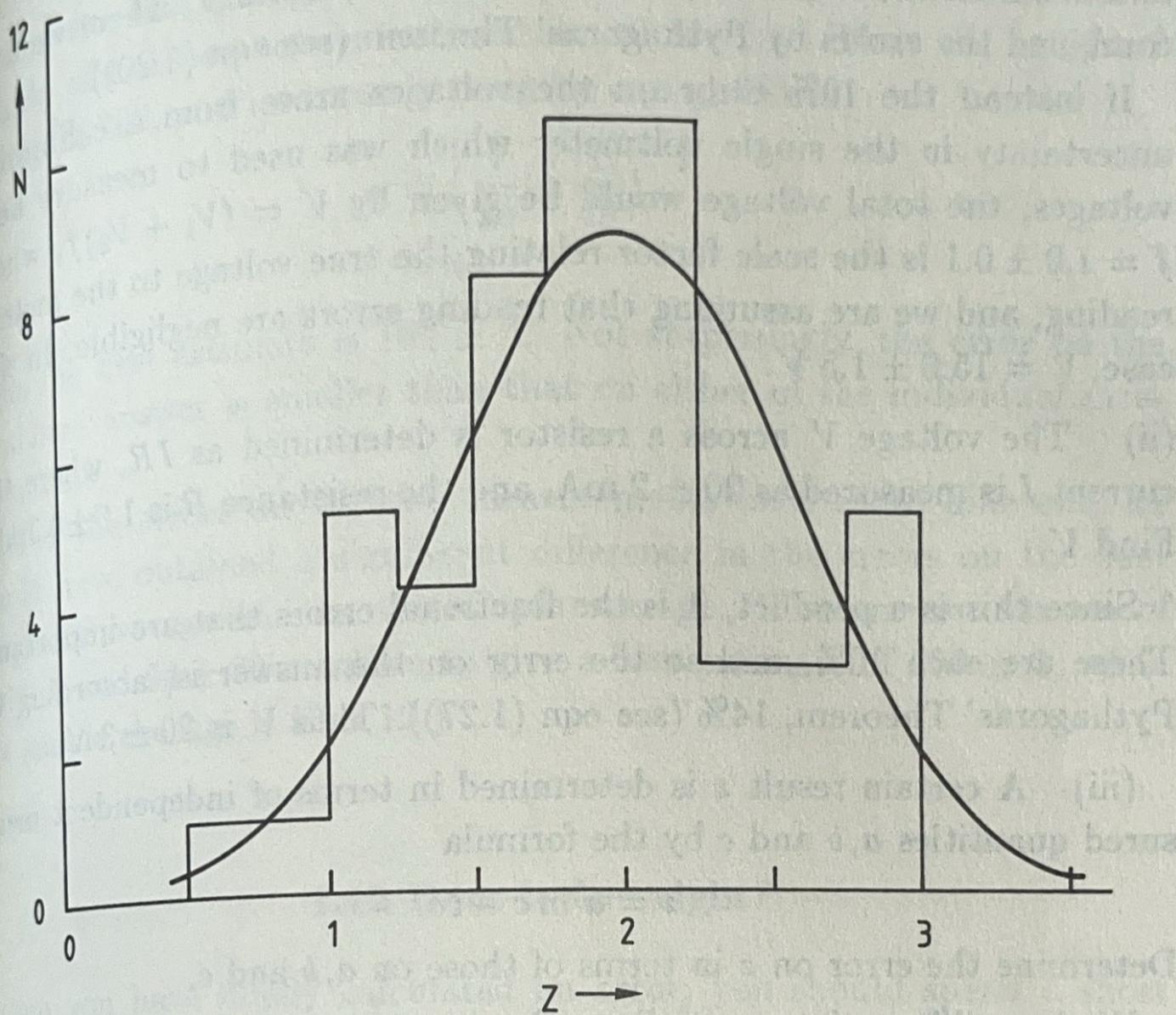


Fig. 1.10. The histogram is that of 50 z values, each of which was obtained by adding four consecutive random numbers from the Table A7.1. The curve is the Gaussian, given by eqn (1.39) of the text. According to the Central Limit Theorem, the distribution of the sum of n random numbers should approximate to a Gaussian if n is large. Even with $n = 4$, and with only 50 entries, the histogram is reasonably consistent with the curve.

large and if there were many entries in the histogram. We see that even with only four x 's added, and with their distribution being uniform, the resulting z distribution is reasonably consistent with a Gaussian, and of the expected parameters.

1.11.4 Combining errors

- (i) Find the total voltage across two resistors in series, when the voltages across the individual resistors are 10.0 V and 5.0 V, each measured with 10% random error.

Since the total voltage $V = V_1 + V_2$, we need to combine absolute errors rather than fractional ones. We thus express the voltages as 10.0 ± 1.0

and 5.0 ± 0.5 V, and obtain their sum as 15.0 ± 1.2 V, where we have combined the errors by Pythagoras' Theorem (see eqn (1.20)).

If instead the 10% error on the voltages arose from a calibration uncertainty in the single voltmeter which was used to measure both voltages, the total voltage would be given by $V = (V_1 + V_2)f$, where $f = 1.0 \pm 0.1$ is the scale factor relating the true voltage to the meter's reading, and we are assuming that reading errors are negligible. In this case, $V = 15.0 \pm 1.5$ V.

(ii) The voltage V across a resistor is determined as IR , where the current I is measured as 20 ± 2 mA, and the resistance R is 1.0 ± 0.1 k Ω . Find V .

Since this is a product, it is the fractional errors that are important. These are each 10%, and so the error on the answer is, according to Pythagoras' Theorem, 14% (see eqn (1.27)). Thus V is 20 ± 3 V.

(iii) A certain result z is determined in terms of independent measured quantities a, b and c by the formula

$$z = a \ln c - bc.$$

Determine the error on z in terms of those on a, b and c .

We first differentiate partially with respect to each variable:

$$\frac{\partial z}{\partial a} = \ln c, \quad \frac{\partial z}{\partial b} = -c, \quad \frac{\partial z}{\partial c} = \frac{a}{c} - b.$$

Then we use eqn (1.30) to obtain

$$\sigma_z^2 = (\ln c)^2 \sigma_a^2 + c^2 \sigma_b^2 + \left(\frac{a}{c} - b\right)^2 \sigma_c^2.$$

(The only part that could give some difficulty is the contribution from the error in c . It is important to combine the two contributions $\frac{a}{c}\delta c$ and $-b\delta c$ before squaring; thus a term $\{(a/c)^2 + b^2\}\sigma_c^2$ appearing in the answer would be incorrect. In fact, provided a/c and b have the same sign, we get some cancellation between the two contributions, and the error is somewhat smaller than we might at first have thought. If we are still not convinced, we can substitute specific numerical values for a, b and c to obtain z , and then recalculate z for a slightly different value of c , in order to see explicitly the contribution to the change in z arising from a change in c . Provided such changes are small, we should find that $\delta z \sim (\frac{a}{c} - b)\delta c$.)

1.11.5 Combining results

Two independent measurements of the same quantity are 100 ± 5 and 106 ± 7 . What is the best estimate?

From the first paragraph of Section 1.10, we find that the relative weights of the two determinations are $49:25 \sim 2:1$. Thus the weighted average is 102, and its error is given by eqn (1.36) as

$$\sigma = \left(\frac{1}{25} + \frac{1}{49} \right)^{-\frac{1}{2}}$$

$\sim 4.$

Thus the best estimate is 102 ± 4 . Not surprisingly, the error on the combined answer is smaller than that on either of the individual measurements.

(Had the errors on the two measurements been more different, we would have obtained a significant difference in the errors on the best weighted average and on the simple average. Thus measurements of 100 ± 5 and 94 ± 20 combine to give a best estimate of 100 ± 5 , while the simple average is 97 ± 10 .)

1.12 Does it feel right?

When you have finally calculated an error, you should spend a short while thinking whether it is sensible.* An important check is that your expression for the error should have the same dimensions as the quantity itself. If not, something terrible has gone wrong with your formula for the error.

Next you should see whether the magnitude of your error agrees with your intuitive feeling about the reliability of your result. Thus if your measurement of the resistance of a coil yields $5 \pm 4 \Omega$, your error is comparable in magnitude to the quantity itself. This should reflect the fact that your measurement is not significantly different from a value of zero for the resistance. If your feeling is that this experiment really did determine the resistance with reasonable accuracy, then you should go back and look at your calculation of the error again. Incidentally, if the error estimate is correct, it does not mean that your measurement is completely useless. For example, perhaps some behaviour of a circuit incorporating this coil could be understood if its resistance were 25Ω ; your measurement is inconsistent with this.

At the other extreme, an error that is very small compared with the measurement (for example, 1 part in 10^4) suggests a very accurate ex-

* Of course, similar considerations apply to the quantity itself.

periment, and you should check that your results indeed justify such confidence in your answer.*

In a similar way, there is usually little point in quoting your error to a large number of significant figures. There is no need to go as far as calculating the error on the error, but often the input to an error calculation consists of statements like 'I think I can read the deflection of the oscilloscope spot to about half a millimetre'. Now this does not imply that the error on the deflection is 0.5 mm rather than 0.4 or 0.6 mm. Indeed on another day you might have decided that this accuracy was 1/4 mm or 1 mm. Clearly with this degree of arbitrariness concerning the basic accuracies, the error on the answer is unlikely to justify more than one or at most two significant figures. Alternatively, if the error is estimated from the spread of the individual results, we need a large number of repetitions in order to make our error estimate accurate (see last paragraph of Section 1.4).

It is also very important to remember that statistics can provide you with a set of formulae to use, but in an actual practical situation it is not simply a case of choosing the correct formula and applying it blindly to your data. Rather you have to make specific judgements. For example, you may have made several measurements over a period of time, and want to combine them. Then it is necessary to decide whether all the measurements should be included in the average or whether some of them should be discarded because of possible bias; whether all the results have the same accuracy; what are the possible systematic effects; whether there might be a possible time dependence of the answer; etc. Thus although problems in books on errors may have unique answers, real life situations are more interesting and we have to use our experience and intelligence.

Finally, in order to demonstrate that error calculations do present problems even for experienced scientists, Fig. 1.11 shows the way in which the combined best value of the mass of the proton (as obtained from all the different experiments that have measured it) has behaved as a function of time. Assuming that the proton's mass really has not varied, we would expect all these values to be consistent with each other. The fact that they clearly are not demonstrates that either some or all

* A scientist who quoted his error as 1 part in a thousand was asked what the three significant figures represented. He replied 'Faith, Hope and Charity.'

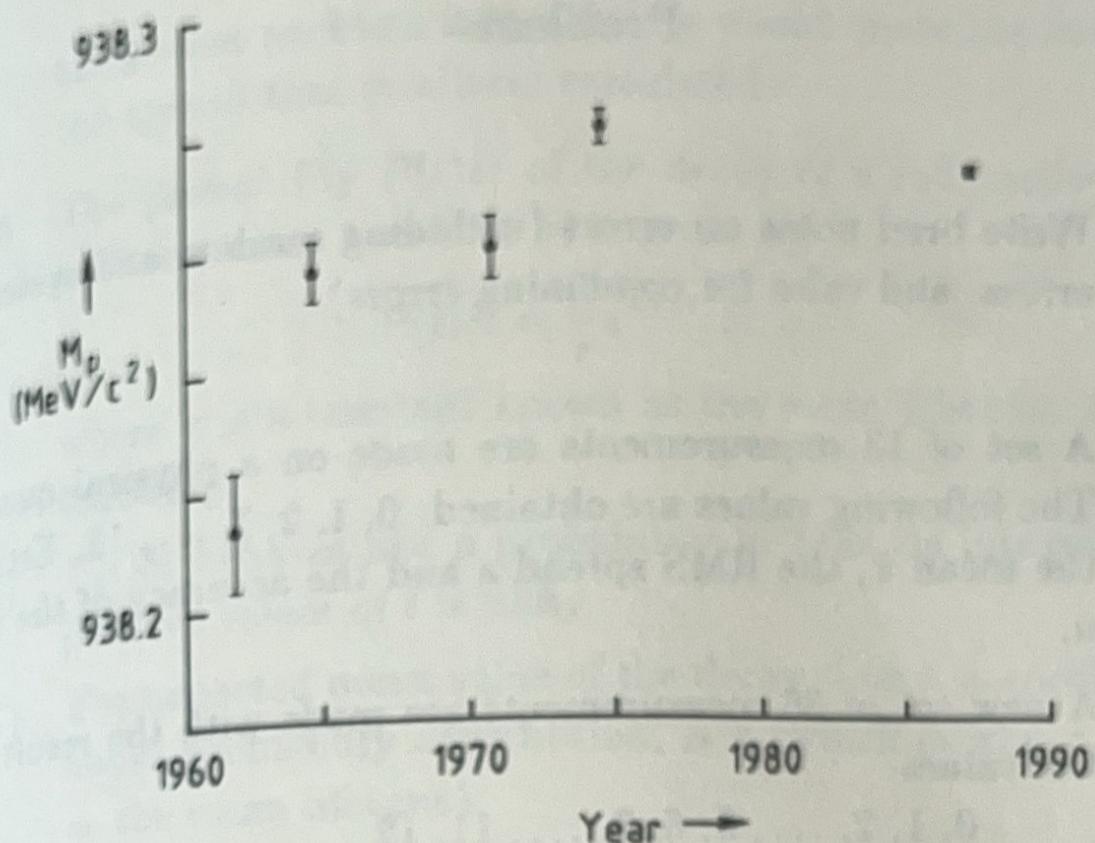


Fig. 1.11. The world average value of the proton mass M_p , as a function of time. The mass is quoted in MeV/c^2 . In these units, the electron mass is $0.5109991 \text{ MeV}/c^2$, with an error of 2 in the last decimal place. (Based on information from the Particle Data Group.)

of the errors have been underestimated. Thus, maybe there were biasses present in the experiments which were not allowed for in the quoted errors, or else the statistical errors were for some reason wrong.

Problems

1.1 Write brief notes on errors (including random and systematic errors, and rules for combining errors).

1.2

(i) A set of 13 measurements are made on a physical quantity. The following values are obtained: 0, 1, 2, 3, ..., 12. Estimate the mean \bar{x} , the RMS spread s and the accuracy of the mean u .

(ii) A new set of 36 measurements are made with the result that the values

$$0, 1, 2, \dots, 5, 6, 7, \dots, 11, 12$$

occur 0, 1, 2, ..., 5, 6, 5, ..., 1, 0 times respectively. Estimate \bar{x}, s and u .

(iii) The function $n(x)$ is defined as

$$n = \begin{cases} 1/L & \text{for } 0 \leq x \leq L, \\ 0 & \text{otherwise.} \end{cases}$$

Find the average value of x , and the spread s for this distribution.

(iv) Repeat the problem of (iii) above, but for the function

$$n = \begin{cases} 4x/L^2 & \text{for } 0 \leq x \leq L/2, \\ 4(L-x)/L^2 & \text{for } L/2 \leq x \leq L, \\ 0 & \text{otherwise.} \end{cases}$$

(v) Compare the answers for (i) and (iii), and for (ii) and (iv). (You should find that, for a sensible choice of L , the results in (i) and (iii) are approximately the same, and similarly for (ii) and (iv). You should also find that the value of s is smaller for (ii) and (iv), since the measurements are more concentrated near the mean, than are those in (i) and (iii).)

The situation described in (iii) is very relevant for a nuclear physics scintillator, which detects charged particles that pass through it. If the device is of width L in the x direction, all we know when it gives a signal is that a particle passed through it somewhere between $x = 0$ and $x = L$. If we want to specify the x coordinate of the particle (for example, for linking together with other measurements in order to find the direction of the

track - see problem 2.4), then we would quote the average and the spread that you have calculated.

- 1.3 The probability $P(t)\delta t$ of the decay of a radioactive particle between times t and $t + \delta t$ is given by

$$P(t)\delta t = \frac{1}{\tau} e^{-t/\tau} \delta t$$

where τ is a constant known as the mean lifetime. Prove the following.

- (i) $P(t)\delta t$ behaves like a probability in that its integral over all positive values of t is unity.
- (ii) The expected mean value of the decay time t , according to the above probability distribution, is τ (which is why τ is known as the mean lifetime).
- (iii) The expected root mean square deviation of decay times about the mean lifetime (i.e. $\sqrt{\langle(t - \tau)^2\rangle}$) is τ .

Several observations are made of the radioactive decay of a charmed meson. The measured decay times, in units of 10^{-12} seconds, are 0.28, 0.02, 0.09, 0.17, 0.10, 0.62, 0.48, 0.06, 0.85 and 0.08. Use the result (ii) above to obtain an estimate of the lifetime τ of this particle. Given that you know from (iii) above that each individual decay time has an error τ to be assigned to it, what is the error on the estimate of the lifetime that you have just obtained? As an alternative, use the observed scatter of the individual decay times in order to calculate the error on the mean lifetime.

- 1.4 By measuring yourself with four different rulers, you obtain the following estimates of your height: 165.6 ± 0.3 , 165.1 ± 0.4 , 166.4 ± 1.0 and 166.1 ± 0.8 cm. What is the best estimate of your height, and how accurate is it? What would have been the best estimate if you had neglected the accuracies of the individual measurements?

- 1.5 Three schoolchildren A , B and C perform a pendulum experiment with the same apparatus in order to determine the acceleration due to gravity g . An individual measurement consists of timing 100 swings of the pendulum, and this is what A does. However, B does this twice and averages the two values to obtain an improved answer, while C takes the average of ten sets of swings. If A 's answer has an uncertainty σ_a , what

are the expected accuracies of B 's and of C 's determinations?
 (Assume that the dominant error is the random one associated with timing the swings.)

The teacher now takes the three students' determinations ($a \pm \sigma_a$, $b \pm \sigma_b$ and $c \pm \sigma_c$) and uses the prescription (1.35) and (1.36) to obtain his estimate of g and its error. Show that these are identical with what the teacher would have obtained by taking all 13 individual measurements and averaging them, without regard to which student had performed which determination.

- 1.6** We wish to determine the ratio f of the strengths of two radioactive sources. For the first we observe 400 ± 20 decays in a minute, and for the second 4 ± 2 in the same time. According to eqn (1.27), the value of f is 100 ± 50 . Is this realistic, or is there a way of quoting f and its error which gives a better idea of our knowledge of the ratio?
- 1.7** For $f = x - 2y + 3z$ (with x , y and z having uncorrelated errors), prove from first principles that

$$\sigma_f^2 = \sigma_x^2 + 4\sigma_y^2 + 9\sigma_z^2.$$

- 1.8** In each of the following cases, determine the answer and its error, assuming that the errors on the relevant quantities involved in the calculation are uncorrelated.
- (i) Determine the distance between the points $(0.0 \pm 0.2, 0.0 \pm 0.3)$ and $(3.0 \pm 0.3, 4.0 \pm 0.2)$, and the angle that the line joining them makes with the x axis.
 - (ii) The number N of particles surviving a distance x in a medium is given by $N_o \exp(-x/\lambda)$, where N_o is the number at $x = 0$, and λ is the mean free path. What is N if $N_o = (1000 \pm 5) \cdot 10^6$, $x = 1.00 \pm 0.01$ m and $\lambda = 0.25 \pm 0.06$ m?
 - (iii) A particle travels along a straight line trajectory given by $y = a + bx$. If $a = 3.5 \pm 0.3$ cm and $b = (5.0 \pm 0.1) \cdot 10^{-2}$, what is the value of y at (a) $x = 4$ m and (b) $x = 4.0 \pm 0.1$ m?
 - (iv) The molar specific heat c of a metal at low temperature T is given by $c = aT + bT^3$. If $a = 1.35 \pm 0.05$ mJ mol $^{-1}$ K $^{-2}$, $b = 0.021 \pm 0.001$ mJ mol $^{-1}$ K $^{-4}$, and $T = 5.0 \pm 0.5$ K, what is the value of c ?
- 1.9** A man lives in a rectangular room for which he wants to buy

carpet and wallpaper. The required quantities of these will be proportional to the floor's area and perimeter respectively. He thus measures the floor, and finds that its dimensions are $l \pm \sigma_l$ and $b \pm \sigma_b$, with the errors being uncorrelated. Find the errors on the area and on the perimeter, and show that they are correlated.

This illustrates a general way in which correlations can arise: we make two or more uncorrelated measurements, and then derive new variables which are functions of the original measurements. Other examples include (i) measuring the x and y coordinates of a point, and then calculating the polar variables r and θ ; (ii) measuring x and y , and then rotating the coordinate system to obtain x' and y' ; and (iii) deducing the intercept and gradient of a straight line fit to several (x, y) sets of data (see Chapter 2, especially Fig. 2.4).

- 1.10 A measurement with some apparatus produces an answer x that is equally likely to be anywhere in the range 10 to 11. We would say that the likely result μ was 10.5 with an RMS spread σ of $1/\sqrt{12}$ (see problem 1.2(iii)).

Now imagine taking three measurements with this apparatus. You can simulate this by using three random numbers in the range 0 to 1 (which you can obtain either from your calculator, or from a table of random numbers such as is given in Appendix 7), and adding 10 to each. Then calculate \bar{x} and s^2 , the estimates of the mean and the variance, from eqns (1.2) and (1.3'). Repeat this procedure several times, and make a list of the \bar{x} and s^2 values. Note that \bar{x} and s^2 scatter about their true values μ and σ^2 respectively. (Compare comments at the end of Section 1.4.)