DATA HANDLING 1

Contents

Aims and Objectives

The main aim is to enable you to convey the significance of your scientific observations in a quantitative manner. This involves development of skills to estimate the uncertainties attached to measurements of physical quantities, usually called ‘errors of observation’.

The objectives are:

- to explain types and categories of error: random error, systematic error, error in a piece of apparatus, error in resulting measurements
- to set out the ways of combining errors
- to describe the Gaussian model of errors and define the standard deviation
- to show how to estimate the error in the gradient of a straight line graph
- to demonstrate how best to plot a graph to extract the ‘physics’
- to enable you to interpret other peoples’ experimental results
- to enable you to present your own data with some reasonable indication of the confidence you place on them.
- to help you to judge, while you are doing an experiment, which of the measured quantities require most care in measurement
- to see ways to avoid unnecessary mathematics or long calculations.

Those who have previously taken a course in Statistics should be warned that there are differences in emphasis, but competent statisticians and intelligent physicists should agree on their conclusions.

1 Random Errors Add in Quadrature

1.1 What are random errors?

Suppose you are measuring the length of a rod. You will not line up the end with the end of the ruler exactly and your estimate of the exact fraction of a scale division at the other end will not be reproducible each time you repeat the measurement. These effects cause variations in successive readings of the length and are *random errors*. Some readings will be larger than the true value and some will be less. If, however, the metre rule has shrunk all the length measurements are too big; this is a *systematic error*.

1.2 Adding together quantities with errors

Suppose you are piling up bricks which are nominally $300 \times 120 \times 100\,\mathrm{mm}$. Due to variations in the manufacturing process each dimension can vary by up to $\pm 2\,\mathrm{mm}$. You place a brick with the $100\,\mathrm{mm}$ edge vertical and place on it one with the $300\,\mathrm{mm}$ edge vertical. What can we say about the total height? Clearly it lies between $396\,\mathrm{mm}$ and $404\,\mathrm{mm}$, but these extreme values are unlikely because they only occur when the variations in the size of both bricks are in the same direction. If we took lots of pairs of bricks and measured the total height we would more typically find variations of up to $\pm 3\,\mathrm{mm}$.

In this example the variations in size of the bricks were real but exactly the same discussion would hold if the variations were the result of errors in *measuring* the size of the bricks. If we measure lengths of $100 \pm 2\,\mathrm{mm}$ and $300 \pm 2\,\mathrm{mm}$ what can be say about the sum of these lengths. Despite Murphy's law, *random* errors do not conspire together always to produce the maximum possible error.

We need a rule for combining *independent* variations. It is:

Independent random errors add in quadrature.

What does 'add in quadrature' mean? It means that independent errors add like vectors at right-angles: the length of the resultant of the orthogonal vectors **a** and **b** is $\sqrt{a^2+b^2}$. Contrast this with the case where **a** and **b** are parallel when the length of the resultant is just $a+b$. In general, if we write δx and δy for the random errors in x and y:

$$(x \pm \delta x) + (y \pm \delta y) = x + y \pm \sqrt{(\delta x)^2 + (\delta y)^2}.$$

So the correct calculation of the error in the height of the two bricks is $\sqrt{2^2+2^2} = \sqrt{8} \simeq 3\,\mathrm{mm}$, and the calculation should be written $(100 \pm 2) + (300 \pm 2) = 400 \pm 3\,\mathrm{mm}$. At the moment the distinction between ± 3 and ± 4 may not seem very important. We have not even defined the meaning of 'a random error of $\pm 2\,\mathrm{mm}$' very carefully. But such numbers have a well-defined meaning and there are ways to obtain values for them, as we shall see.

Suppose you add two more bricks whose heights you have measured to be 119 ± 2 and $99 \pm 2\,\mathrm{mm}$. Note that although the bricks nominally have dimensions of

120 mm and 100 mm your measurements are not exact. The total height of the top surface is now:

$$(100 \pm 2) + (300 \pm 2) + (119 \pm 2) + (99 \pm 2)$$

$$= 618 \pm \sqrt{2^2 + 2^2 + 2^2 + 2^2} = 618 \pm 4\,\text{mm}.$$

You get 618 ± 4 mm whereas if you had assumed that all four errors acted in cooperation you would have had 618 ± 8 mm, clearly very different and also wrong.

The use of the word 'error' is rather unfriendly, to the layman 'error' means 'mistake'; but scientists use it in the sense of 'uncertainty'. Talking about the 'uncertainty' in a measurement is much less misleading, but 'error' is the normal scientific term.

1.3 Calculating Small Changes

Errors in measurements are usually small and physicists need to understand the effects of small changes. You will be asked to do the first set of questions *without using a calculator* and there is a good reason for this. The point of most of the calculations is not only to get the right answer but to find the right method. The questions can all be done using brute force but this does not lead to any understanding.

To see the effect of a small change on the power of a quantity it is useful to remember the Binomial expansion:

$$(1 + \delta x)^n = 1 + n\delta x + \frac{n(n-1)}{2!}\delta x^2 + \cdots + \frac{n!}{(n-r)!r!}\delta x^r + \cdots.$$

Here unity is being changed by the addition of δx. When $\delta x \ll 1$ the higher terms in the expansion diminish rapidly: $(1+\delta x)^n \simeq 1 + n\delta x$. Also, if $\delta x \ll 1$ and $\delta y \ll 1$ $(1+\delta x)(1+\delta y) \simeq 1 + \delta x + \delta y$.

Example Calculate $1/1.001$ to 3 decimal places.

$$\frac{1}{1.001} = (1 + 0.001)^{-1} = 1 + \frac{(-1)}{1} \times 0.001 + \frac{(-1)\times(-2)}{2 \times 1} \times (0.001)^2 \simeq 0.999$$

1.4 The error in a product or quotient

We have claimed that the error in $x + y$ is $\sqrt{\delta x^2 + \delta y^2}$, what about errors in more complicated expressions. Let us start with the case where we simply scale x. Suppose we measure $x \pm \delta x$ what is the error in $2x$? Since $2(x \pm \delta x) = 2x \pm 2\delta x$ it is clearly $\pm 2\delta x$.

Now suppose we measure $x \pm \delta x$ and $y \pm \delta y$. What about the error in $z = x \times y$? This can be found as follows:

$$(x \pm \delta x)(y \pm \delta y) = xy \pm y\,\delta x \pm x\,\delta y + \delta x\,\delta y \simeq xy \pm x\,\delta y \pm y\,\delta x$$

The product xy has two error terms added to it with values $x\,\delta y$ and $y\,\delta x$. These will add in quadrature giving the error in xy as $\sqrt{(y\,\delta x)^2 + (x\,\delta y)^2}$. This ungainly expression can be written simply as

$$\left(\frac{\delta z}{z}\right)^2 = \left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2,$$

that is *the fractional errors add in quadrature*.

What about the error in x/y? We can get this in two steps: first we work out the error in $1/y$ and then apply the product result. The error in $1/y$ follows from the binomial theorem:

$$\frac{1}{y \pm \delta y} = \frac{1}{y} \times \left(1 \pm \frac{\delta y}{y}\right)^{-1} \simeq \frac{1}{y} \times \left(1 \mp \frac{\delta y}{y} + \cdots\right) = \frac{1}{y} \mp \frac{\delta y}{y^2}.$$

The error in $1/y$ is therefore $\delta y/y^2$. Now applying the product result: if $z = x/y$

$$\left(\frac{\delta z}{z}\right)^2 = \left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta(1/y)}{(1/y)}\right)^2 = \left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2.$$

These results are easy to remember

> When calculating the sum of two quantities add the *absolute* errors in quadrature.
>
> When calculating the product or quotient add the *relative* errors in quadrature.

The correct number of figures When you have an estimate of the uncertainty in a measured quantity you should give the appropriate number of significant figures in both the result and its error. Quoting a length to a fraction of a millimetre would be inappropriate if the error was $\pm 10\,\text{mm}$. Conversely if you write $z = 14.5 \pm 0.002$ we don't know if you measured 14.500 and forgot to write the significant zeroes, or you measured 14.512, failed to realise that the error was so small, and approximated the reading. If you write 14.500 ± 0.002 this is made absolutely clear.

Questions on combining errors

Calculate to 3 decimal places without calculator, (show your working to prove how you did it):

Q1a $\frac{1}{1.003} =$

Q1b $\frac{1}{0.998} =$

Calculate to 4 significant figures (without a calculator):

Q1c $\frac{1}{2.001} =$

Q1d $(1.00006)(1.017)^2 =$

Given two measured quantities x and y whose random errors are $\pm\delta x$ and $\pm\delta y$ respectively, what is the error in

Q1e $(x+y)/3$

Q1f $x-y$

Q1g $x+2y$

If $x = 310 \pm 10\,\text{mm}$ and $y = 290 \pm 10\,\text{mm}$, write down the values, with errors and units, of

Q1h $x+y$ (Don't leave expressions like $\sqrt{2}$.)

Q1i $x-y$

Let x be measured with random $\pm\delta x$. Work out algebraically the values of

Q1j x^2

Q1k $1/x^2$

If $x = 100 \pm 4\,\text{mm}$ and $y = 200 \pm 6\,\text{mm}$, write down the values (as always, with uncertainties and units) of

Q1l xy

Q1m x^2

Q1n $y - x$

Q1o x/y

Q1p $(x - y)/y$ Think hard about this one!

2 Errors in the Apparatus, Errors in the Results

2.1 Assigning values to errors

Our next job is to find a reasonable way to estimate experimental uncertainties. Two students, Alf and Betty, have set up apparatus to measure the time of fall of a particular size of ball-bearing through oil. Alf takes one reading and says "9.4 seconds". Betty, using a different method of timing says "9.62 seconds". Should we believe Betty because she quotes two decimal places or Alf because he has an honest face? We need to know if Betty thinks 9.61 s, for example, is also credible and what Alf thinks about 9.5 s. Without more information we do not know what confidence to place on the quoted results. We must ask Alf and Betty to take more readings. After ten readings -

Alf:	9.4	9.6	9.5	9.5	9.4	9.3	9.1	9.8	9.3	9.5
Betty:	9.62	9.32	9.48	9.46	9.59	9.39	9.54	9.43	9.50	9.61

We can present these data also as *histograms*, that is, number of events N per 0.2 s interval versus observed time t. These are shown in figure 1

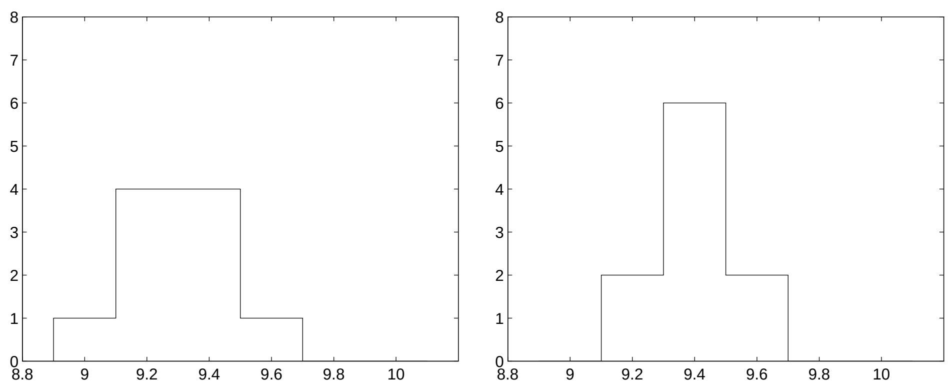

Figure 1: Alf's readings (left) and Betty's readings (right)

After ten readings it is already clear that Betty's apparatus is more precise than Alf's. What is the reason for these fluctuations? We can consider them due to numerous random factors in the sighting of the balls and the stopping and starting of the clocks. These add up to a random error which can be positive or negative.

In figure 2 we shown what Alf and Betty's results might look like after one thousand readings each; the difference is even more obvious. We can imagine that if Alf and Betty kept taking readings, their data would produce histograms which were smooth curves.

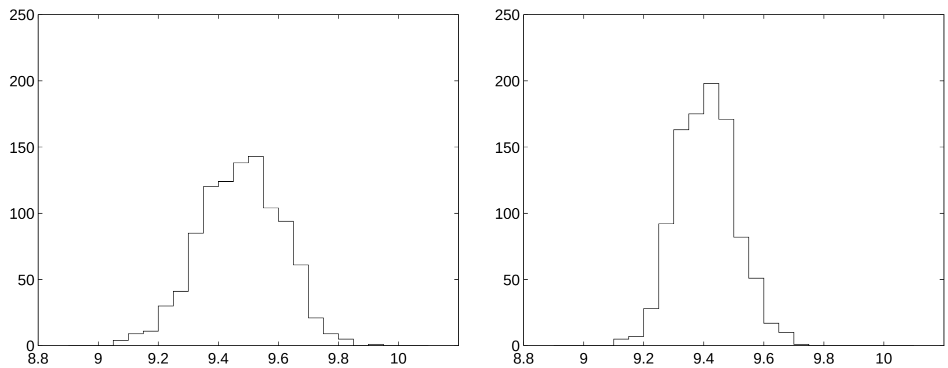

Figure 2: 1000 readings by Alf (left) and 1000 readings by Betty (right)

2.2 The Gaussian distribution

Experimentally we often find that such curves have a form called the Gaussian distribution; it is described by a function containing two parameters:

$$f(x) \propto \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right].$$

It is a peak centred at $x = \mu$ ('mu'), called the *mean* and with a 'width' σ ('sigma') called the *standard deviation* of the distribution. Figure 3 shows the curve for $\mu = 0$ and $\sigma = 1$.

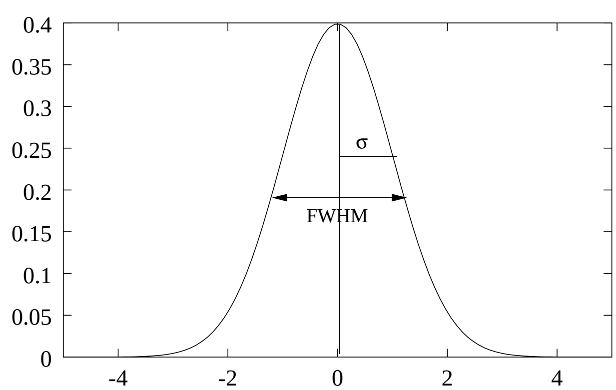

Figure 3: The Gaussian distribution

From the form of the function we can deduce some useful properties of the distribution.

- When $x = \mu \pm \sigma$, its height is about 0.6 of the maximum.

- The **F**ull **W**idth of the curve at **H**alf the **M**aximum height, usually known as FWHM, has the value 2.35σ.
- Roughly two-thirds of the readings fall within the range $x = \mu \pm \sigma$.

It can be shown theoretically that random errors will have approximately this distribution if the error is the sum of many small contributions, as is often the case in physical measurements.

Note that the smaller the spread of readings the more precise the measurement, so high precision goes with small σ, and Betty's apparatus has a σ smaller than Alf's.

We now realise that when we hesitated originally over believing Alf and Betty, the first thing we wanted to know was: how precise is the measuring apparatus? That is, how precise is any single reading made using it? Now we know a way to answer it:

> The precision of the measuring apparatus is measured by σ which is determined from the distribution of readings.

2.3 Determining the standard deviation

In principle to determine the standard deviation we could take a lot of independent repeated readings of the same quantity and find the width of the Gaussian curve which fits the histogram. But to obtain anything like smooth curves requires a very large number of readings; therefore in practice we want to be able to *estimate* the width from a smaller group of readings such as the sets of ten that Alf and Betty obtained in what they considered a reasonable time. A convenient measure of the spread of a set of n observations x_i, around their average $\bar{x}$, is the root-mean-square (rms) deviation, s, given by

$$s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n}}.$$

We first calculate the average of all the readings $\bar{x}$; and then the deviation of each of the repeated readings x_i from the average: $x_i - \bar{x}$. We then square these deviations and add up the squares. The summation sign (Σ) means take the sum of the squares of the deviations from the average of all the n repeated measurements of the same quantity, x_i, from $i = 1$ to $i = n$. We then divide by the number of readings n to get the mean-square-deviation, and take the square root to get the root-mean-square deviation.

The rms deviation, s, is derived from our set of readings. It can be shown that as $n \to \infty$, s becomes closer and closer to σ. We shall use s as an estimate of the quantity σ in the Gaussian distribution that one imagines describing the spread of readings taken with the apparatus. Thus s can be used to estimate the precision of the apparatus.

When Alf calculates s from his 10 readings he gets 0.18 s, while Betty gets just half, 0.09 s.

2.4 What to conclude from a set of repeated readings

In addition to the precision of the apparatus given by the rms deviation s, the other information we wanted from Alf and Betty was: what should be quoted as the result of their measurements? In other words:

- What is the best value of the time of fall that can be obtained, with a particular apparatus from an experiment consisting of n readings?
- What range of values can be accepted as credible?

Most scientists believe, without justification, that they know the answer and that the best value is $\bar{x}$ the average of the n readings. Provided the distribution of errors is Gaussian, the average is indeed the best value. But there is nothing magic about the arithmetic average and there are circumstances where it is not the best value. If you meet one of these you need to consult an expert.

What is the range of plausible values? It can be shown to be approximately $\bar{x} \pm s/\sqrt{n}$ where s is the uncertainty in a single reading. You can see that this is reasonable by working out the error in the expression for the mean

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

where each x has an error $\pm\delta x$ given by s as estimated above. Thus we have the best value of x and its uncertainty:

> The best value of a quantity obtained from a set of n repeated readings is $\bar{x} \pm s/\sqrt{n}$.

The formulas, precision $\sigma \simeq s$ and result $\simeq \bar{x} \pm s/\sqrt{n}$, are correct if the number of measurements n is large. You will see in books other formulas involving $n - 1$. If your calculator has a standard deviation function it may give slightly different results for small n. We point out that in practice the difference should not matter. If n is so small that the distinction is important, you should be doing a better experiment!

2.5 Quick Calculations of Mean and Standard Deviation

The next section shows how to simplify the calculation of the mean and standard deviation. Choose a round number (called a 'working mean') and subtract it from all the data values so that the number of digits to be handled in each reading is reduced. Compute the mean of these modified readings and then add the 'working mean' again to get the mean of the original data.

Example Four independent readings are $1340, 1345, 1348, 1343\,\mathrm{mm}$; what is the average and standard deviation?

Take the 'working mean' as 1343. The deviations from this are: -3, $+2$, $+5$, and 0. The sum of these is $+4$, so the average relative to the working mean is $+1$, and the average of the original readings $1343+1=1344$.

It is also worth knowing that subracting a constant from all the readings does not change the value of s, so you can work with the values after the working mean has been subtracted. We calculate s by subtracting the average $(+1)$ from each reading, giving -4, $+1$, $+4$, and -1, summing the squares of these, which can be done in your head, giving 34, dividing by 4 and taking the square root giving, in our head, a bit less than 3. If you have a calculator handy you can get $s=2.92$, but the extra accuracy is unecessary.

This figure describes the variability of individual readings. We are going to condense these into a single figure and its uncertainty which will be $2.92/\sqrt{4}=1.46$. The deduction from these four readings should therefore be that the value of the quantity is 1344.0 ± 1.5.

Rough values for the error A simple method of getting a *very rough* estimate of the standard deviation of a distribution of readings is $s \sim \text{range}/\sqrt{n}$ where the range is highest value minus the lowest value. This is a poor way to get s but is very useful to see if you have made a serious error, like a factor of 10!

Another way to get a rough estimate of s when the number of readings is large uses the fact that approximately $2/3$ of the readings should lie in within $\mu \pm s$. Again this is most valuable for quickly spotting mistakes.

2.6 How Many Significant Figures?

After working on their sets of ten readings Alf and Betty come back from their calculations and present their results. Alf writes '9.44 ± 0.0569', then adds quickly, 'seconds'.

Betty writes '9.494 ± 0.0295364 s'. After a quiet word from a friend, she changes it to '9.494 ± 0.0295 s'.

They have each written down initially what they read on their calculators and are not sure how many figures to give. What guidance can we give them? In fact things are worse than they think. The uncertainty in using s as an estimate for the true error σ depends on the number of readings. It can be shown that $s \simeq \sigma \pm \sigma/\sqrt{2n}$. (Statisticians would write $\sigma/\sqrt{2n-2}$). Thus for 10 readings the 'error in the error' is more than 20%! For 50 readings it falls only to about 10%. This warns us about spurious precision and the futility of elaborate calculations.

Remember the 'error-in-the-error' is merely a guide to the number of figures to quote. Do not use it as an excuse to increase your estimate of the error. In any ordinary experiment calculating two figures in the error is always enough and we suggest you do this as a matter of routine even though the second figure is often not justified. Keeping a second figure means that any further calculations involving errors do not suffer a loss of accuracy from repeated rounding. However, a second figure in the error quoted in a *final result* may be meaningless. Our conclusion on what Alf and Betty should give as their results is therefore:

If this is not the last word on the measurement of the time of fall, Alf and Betty should record:

$$9.440 \pm 0.057\,\text{s} \text{ and } 9.494 \pm 0.030\,\text{s}$$

If this is their final report on the subject then common sense would demand that Alf and Betty should tell the outside world:

$$9.44 \pm 0.06\,\text{s} \text{ and } 9.49 \pm 0.03\,\text{s}$$

2.7 What do the Results of Alf and Betty Mean?

What do Alf's and Betty's numbers tell us about the value of the time of fall? What do their statements really mean?

Let us look again at the Gaussian model for distributions of errors. When we know the value of σ we know to expect that about two-thirds of the data fall within $\pm\sigma$ of the peak. Thus the value of σ gives us betting odds of about 2 to 1 that a value falls within this range. The same betting odds apply to the statement that a result is $X \pm \delta X$; it can be understood to mean that:

When the result of an experiment is $X \pm \delta X$ then the odds are 2 to 1 that the true value μ of X lies between $(X - \delta X)$ and $(X + \delta X)$

So Alf's statement gives odds of 2 to 1 that the fall-time is between 9.38 and 9.50 while Betty's result gives the same odds that the fall-time is between 9.46 and 9.52 s. Betting odds of 2 to 1; that is all. Don't be surprised if subsequent experiments produce a value which lies outside one or other's range. One repeat in three may lie outside.

It is useful to have an idea of the probabilities predicted by the Gaussian model for a result to lie within a certain number of σ of the peak. We quote here, for reference and possibly for memorising, the percentages of a Gaussian distribution which lie inside the bands $\pm 1\sigma, \pm 2\sigma$ and $\pm 3\sigma$. These are given with much greater precision than is useful. For convenience we give also the approximate fractions that lie outside these bands in both tails.

region	%inside	outside, in both tails
$\pm 1\sigma$	68.27%	one in three: 1/3
$\pm 2\sigma$	95.45%	one in twenty: 1/20
$\pm 3\sigma$	99.73%	one in four hundred: 1/400
$\pm 4\sigma$	99.994%	1/16000

2.8 Summary of What You Have Learned So Far

1. Independent random errors add in quadrature.

2. In the Gaussian model of typical random errors, σ is the parameter describing the spread of readings in an infinite experiment. σ is a measure of the precision of the apparatus and gives the uncertainty in a single reading.

3. As an estimate of σ, we take s, the rms deviation from the average, of a finite number of readings.

4. In a real experiment aiming to measure x, provided the n readings have approximately a Gaussian distribution, the best value of x is the mean $\bar{x}$ with uncertainty $s/\sqrt{n}$. Compared to taking just a single reading, taking n readings reduces the uncertainty by a factor $1/\sqrt{n}$.

5. The fractional uncertainty, about $\pm 1/\sqrt{2n}$, in the error estimate itself, (called the 'error-in-the-error') is a caution against excessive precision.

6. A further restraint is the recognition that the conventionally quoted 'one standard-deviation' errors have only about two-thirds probability of containing the true value.

7. We also gave two quick on-the-spot alternatives for making a rough estimate of s, (a) the 'range' method: and (b) the 'one-sixth above, one-sixth below' method for large n.

Questions on mean and standard deviation

Here is a question intended to convince you that in order to halve the uncertainty in a mean value you have to take four times as many readings. Suppose each of n independent readings x_i has the random error $\delta x = \pm s$ (plus or minus exactly the r.m.s error).

Q2a What is the error in the sum of these n readings?

Q2b What is the error in the average of these n readings?

The following readings of the length (in inches) of a plank were made:

61.2	61.4	61.5	61.1	61.3	61.3

Q2c What value would you give for the length and its uncertainty?

A physicist recorded that he made the following observations of a time period:

4.95	5.06	5.08	5.09	4.96	5.13	5.07	4.86	5.12	5.11	4.85	5.19
5.15	4.86	4.95	5.00	5.06	4.86	4.82	4.95	5.13	4.96	4.91	4.88

He quoted the best estimate of the period ($\bar{x} \pm s/\sqrt{n}$) as 5.00 ± 0.04 s.

Q2d Do you think he has made a mistake?

Q2e Use the 'range' formula to estimate the s.d. of Alf's readings of time of fall listed in section 2.1

Alf's limited data produced a value of 0.18 for the precision of his apparatus.

Q2f What is a plausible range for the true value of this quantity?

Rewrite the following more sensibly:

Q2g Frequency = $(176.34 \pm 2)\,\mathrm{Hz}$

Q2h Speed of sound = $(333.2 \pm 29.63)\,\mathrm{m\,s^{-1}}$

3 The Uncertainty in the Gradient of a Straight-Line Graph

3.1 Introduction

A very common way of processing data from physics experiments is to plot a graph which should be a straight line and the quantity we want is the slope of the line. For example we could plot the length of a spring against the weight attached to the end of it.

In some cases we might have to plot some function of the readings to get a straight line. An example would be an experiment to determine g from measurements of the time of swing T of a pendulum as the length l was varied. In this case we need to plot either T^2 against l, or T against $\sqrt{l}$, in order to get a straight line graph.

Another example is the 'log-plot' which would be appropriate for dealing with measurements on radioactive decay. The number of atoms of a sample remaining after time t is given by

$$N(t) = N(0)e^{-\lambda t}$$

where the constant λ is related to the half-life $t_{1/2}$ by $\lambda t_{1/2} = \ln 2$. A plot of $\ln N(t)$ against t is then a straight line.

Conventions about graphs Conventionally, if you fix x and measure $y = f(x)$, you plot x horizontally and y vertically. You measure the gradient by picking two points on the line and measuring their separations along x and y as Δx and Δy respectively. The gradient is then $\Delta y/\Delta x$. Sometimes the quantity you want is $\Delta x/\Delta y$. There is scope for confusion, so write down the definition of the slope you are measuring and include the units. For example, if you are determining a resistance by measuring the current through it for various applied voltages you will be plotting the current vertically and the voltage horizontally, then

$$\text{gradient} = \frac{\Delta I}{\Delta V} = \frac{145\,\text{mA}}{5\,\text{V}}$$

Which value of resistance is right? 29Ω? 0.029Ω? 34.5Ω? 0.0345Ω?.

3.2 What is the 'best' straight line?

One cannot answer this question without asking the subsidiary question "Why do not the points all lie on a perfect straight line?" One possibility is that the theory is only an approximation. Another is that there are systematic errors in the readings. An example of this would occur in measuring the length of a pendulum where the point to which the length should be measured is ill-defined. In either of these cases the line will not be straight. One hopes for the situation where the line appears straight but the individual points are scattered about it, presumably due to random errors in measuring one or both quantities. It is clear that one *must* plot the graph and look at it closely.

The 'ruler and eyeball' method This involves placing a transparent ruler over the points and drawing the line that looks best. Experiments (in this department) have shown that this gives very good results when compared with more mathematical approaches. It can still work where the line is not straight and points that are clearly wrong can be ignored. The real disadvantage, apart from not being very professional, is that it is very difficult to judge the uncertainty in the resulting slope.

Least-Squares Computer Programs This is the sophisticated end of the market; a computer program adjusts the line until it is the best fit. The criterion used is that the sum of the squares of the distances of the points from the line is minimised. You may see this described as the 'Principle of Least Squares' as if this were something fundamental in physics. This is misleading; there is no such principle. The correct statistical approach to data processing leads to least-squares in the special case of equal Gaussian errors.

The problem with computer programs is that you get a value for the slope, and its error, even if the line is not straight and even if you make a mistake typing the data. Obviously you need to examine the results carefully to see if they are sensible. The programs that you usually meet assume that x is measured accurately and you can specify an error for y. There is no difficulty if there are *known* errors in both x and y; the reason you rarely come across programs for this case is that it is believed by some statisticians to be ill-posed.

Points in pairs This method is convenient if you don't have a computer handy, and it will give more sensible results if the line is not quite straight or the deviations of the points from the line do not have a Gaussian distribution.

You have learned how to find errors from repeated measurements of the same quantity. A set of points on a straight-line graph is really making repeated measurements of the slope, $\Delta y/\Delta x$, as you can see from figure 4. Each pair of readings can be considered as an independent measurement of slope, $\Delta y/\Delta x$. The pairings will have equal weight (equal significance) if each point has equal precision and if the points in each pairing are equally far apart. Pairing up the points like this you get, from eleven points, five independent measurements of the gradient; you can average these to get the gradient and estimate the uncertainty in the gradient from their scatter. Forget the middle point if there is one; it scarcely affects the measured slope. The lines are drawn in the figure simply to make the associations clear.

If, having calculated the gradient, you want to draw a line with this gradient through the points, it ought to pass through the point whose coordinates are $(\overline{x}, \overline{y})$. This method of obtaining the slope by pairing points has the advantage, over the ruler and eyeball method, of giving you a quantitative estimate of the uncertainty in the slope.

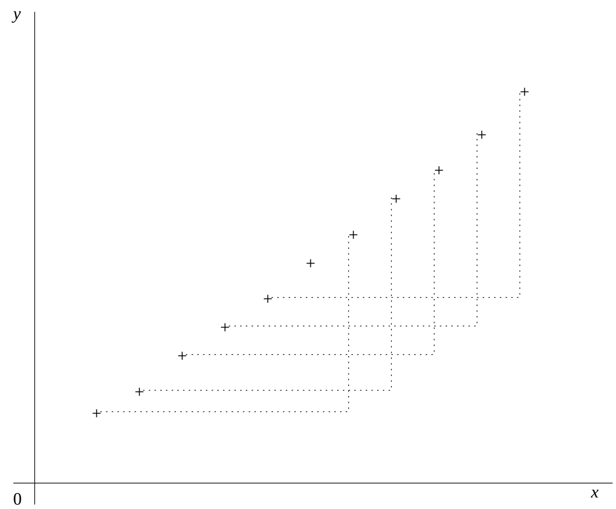

Figure 4: Pairs of points give independent values for the slope

Which data points affect the gradient? Awareness of the points-in-pairs method is valuable in appreciating which data points are important in the determination of a slope. Two wrong ways to group the points into pairs are illustrated in figure 5. On the left the slopes from the pairs are independent but have unequal weights.

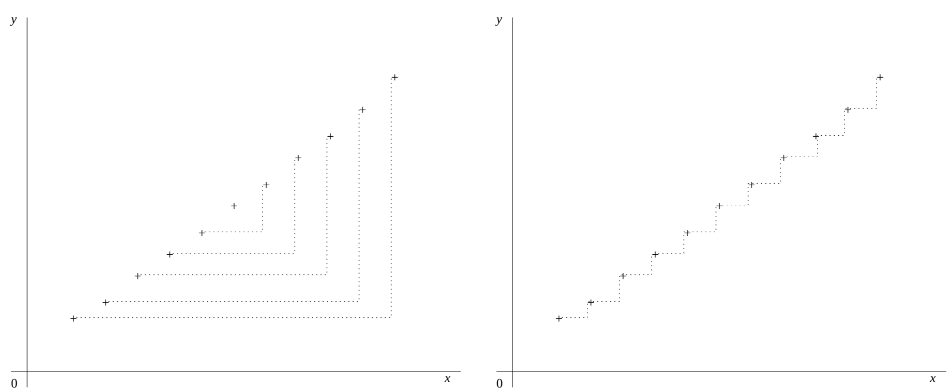

Figure 5: Examples of bad pairing of points. Unequal weight (left) and Not Independent (right)

The central pair of points gives a very poor estimate of the slope and this could have a disastrous effect on the average value. The other choice (right) has pairs with equal weights but they are not independent; in fact, if you think about it, only the end points are contributing to the average value of the gradient, the positions of all the rest cancel out in the calculation and their measurement is wasted! This situation commonly arises when measurements are taken in equal steps, for example measurements on interference-fringes.

3.3 New Forest Linefit

`New Forest Linefit` is a program designed for determining the parameters in linear relations between physical quantities. It is installed on SUCS computers under departmental software; if you want to install it on your own computer go to `www.phys.soton.ac.uk/teaching/year1.htm`, click on **`lab1`** and then **`downloadable software`**. Follow the instructions in the **`Readme`** file.

When the program is run the screen shows three panels; to get started click on the **`Example`** button on the left panel which creates a dataset with errors and displays it in the centre panel. To fit the line click on the **`Fit Line`** button on the right hand panel when the points, errorbars and line are shown on the graph together with the slope and intercept and their errors. Try selecting the **`errors from scatter`** button and refit the line. The third column in the dataset is now ignored and the scatter of points about the line is used to find the errors in slope and intercept. There are other models for the errors where you can assume all the points have the same error or that the error in **`y`** is a fixed fraction of **`y`**.

Usually you will have your data on a disk and this can be read into the program. Click on the **`Clear`** button to remove the previous data. The form of the datafile is one point per line, with either the values of **`x`** and **`y`** or **`x y`** and the error in **`y`**. Columns may be separated by either commas or spaces. You can edit your data when it is loaded, to remove any explanatory text for example, and you can save it again with the **`Save Data`** button. Be very careful using cut and paste to get data from other programs like spreadsheets; these usually embed invisible characters into the data which make it look illegal to the fitting program. If you want to get data out of a spreadsheet save it as an **`ascii`** file first.

Normally you will write down the values of the slope and intercept from the screen. You can get a printed copy of the graph using the **`Save Graph`** button which generates a **`postscript`** file of the graph 10cm square. This can be printed directly on postscript printers or via a package like **`Ghostview`**, or inserted into other documents.

An Aside on Correlation and Line Fitting Most calculators programmed to fit lines give also the 'correlation coefficient'. This is useful to statisticians, but not to scientists, because statisticians *are solving a different problem!*. Suppose we record the annual beer consumption and exam marks of 50 students and plot one against the other. One would not be surprised to find that the points were roughly spread about a straight line (with negative slope!). This is an example of a *statistical correlation* between two quantities but no one would suggest there is a *physical law* connecting them. It happens that the solution of the physicist's problem of obtaining parameters in physical laws is similar to the solution of the statistician's problem of decribing correlation. But just because they use the same mathematics does not mean the interpretation of the answers is the same.

Questions on getting physics from a graph

Q3a If you were determining g from measurements of T and l using a simple pendulum, would you plot T^2 against l or T against $\sqrt{l}$?

First Data Set You have been measuring the displacement, d, of one end of a steel spring under a load, W, and have obtained the data given below. You expect the relation between W and d to be $W = K \times d$ and you wish to find the value of the spring constant K. First plot the data from table 1 on figure 6 using the convention that the quantity you have selected (W) is horizontal and the dependent quantity (d) is vertical. We recommend that after positioning each (small) point, you draw a circle round it; this makes the points visible and improves the graph's appearance. You do not need to calculate the actual displacement from the zero load position. Note that in this case the zero load position is subject to the same uncertainty in its measurement as any other of the points.

Q3b Do the plotted data appear to justify the assumption that the force is a linear function of the displacement with no evidence for non-linearity?

Let us now compare two methods of determining the gradient and its error: the ruler-and- eyeball method and the quantitative method suggested above.

Place a transparent rule along the points and draw what you think is the best line through the points; measure the gradient and make an estimate of what you think its uncertainty might reasonably be.

Q3c Estimate of gradient, and its error:

Calculate the gradient and its error using the method of 'points-in-pairs'. Pair up the points in the table taking care that each pair is independent and that all pairs span roughly equal ranges and so carry equal weight. For these pairs, enter corresponding values of ΔW and Δd in the first two empty columns in the table. In the next column write the calculated value of gradient $\Delta d/\Delta W$. From this set of numbers make an estimate of the gradient and its uncertainty. Then derive the spring-constant K (spring-force/extension) and its uncertainty.

Q3d Gradient = $\Delta d/\Delta W$

W kg	Scale reading μm	ΔW kg	Δd μm	$\Delta d / \Delta W$ μm/kg
0.0	1716			
0.5	1557			
1.0	1374			
1.5	1214			
2.0	1022			
2.5	855			
3.0	664			
3.5	500			
4.0	337			
4.5	151			

Table 1: Displacement of the end of a spring with various weights on the end.

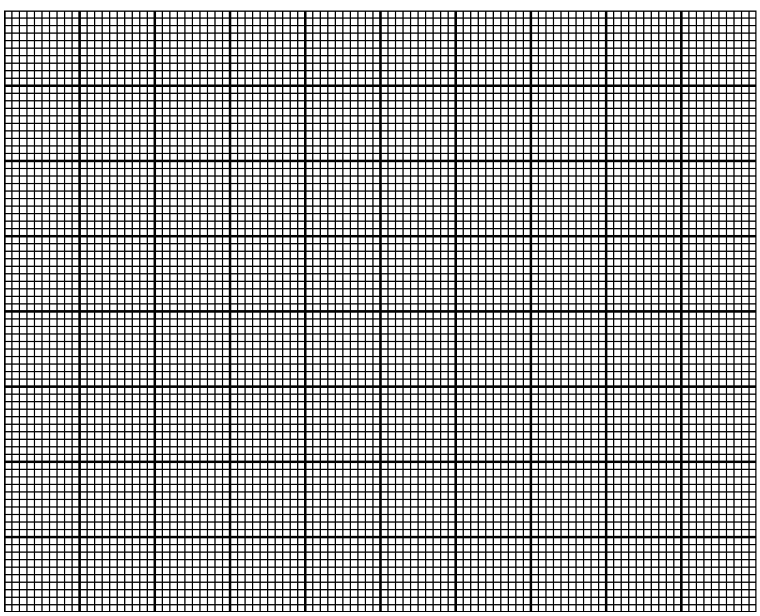

Figure 6: Graph of length against weight

Q3e Spring constant, K =

Second Data Set The data in table 2 are based on Reynolds' original paper in which he was beginning his quantitative measurements on turbulent flow of water through a tube. Without drawing a graph it is difficult to tell that at low pressure the rate of flow is proportional to the pressure gradient, but that above some pressure gradient a change occurs: the flow ceases to be streamline and becomes turbulent.

Plot a graph in figure 7 and make an estimate of the pressure gradient at which streamline flow changes to turbulent flow in Reynolds' apparatus. Try to give an estimate of the uncertainty also.

Q3f Critical pressure gradient =

Using the same methods as in the previous question, calculate the gradient of the linear portion of the graph which corresponds to streamline, or laminar, flow.

Q3g gradient =

Pressure gradient $N\,m^{-3}$	Average velocity $mm\,s^{-1}$			
7.8	35			
15.6	65			
23.4	78			
31.3	126			
39.0	142			
46.9	171			
54.7	194			
62.6	226			
78.3	245			
86.0	258			
87.6	258			
93.9	271			
101.6	277			
109.6	284			
118.0	290			

Table 2: Flow of water through a tube under various pressure gradients

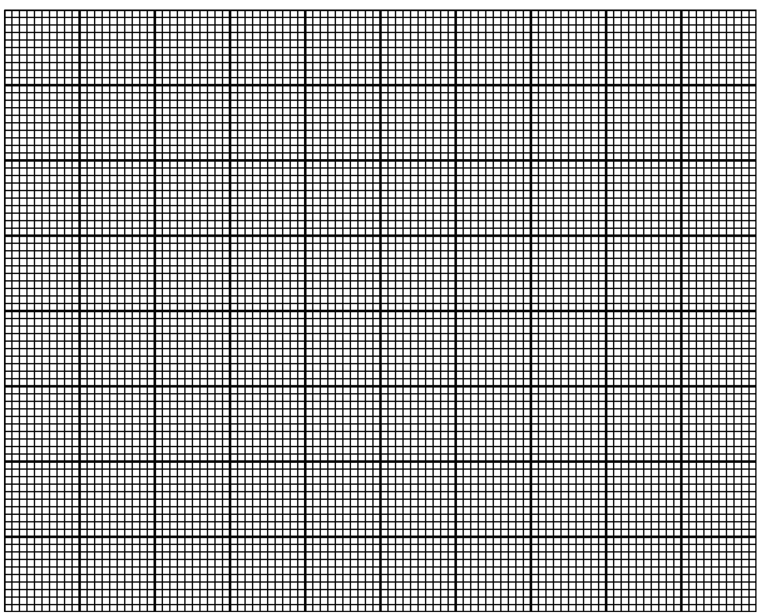

Figure 7: Graph of flow against pressure gradient

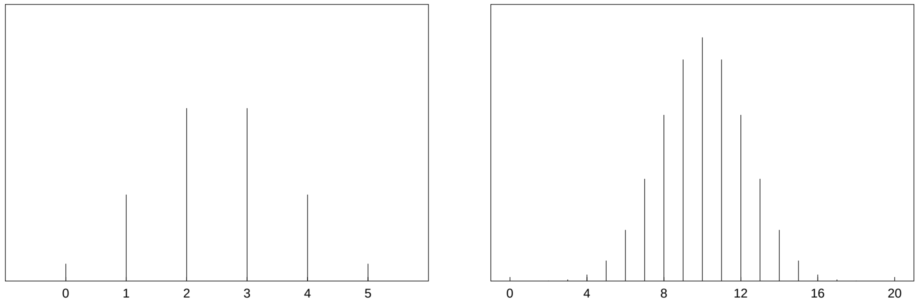

Figure 8: Binomial Distribution $N = 5$ (left) and $N = 20$ (right) with $q = 0.5$

4 Coin tossing and the Binomial distribution

The binomial distribution describes a number of independent identical trials each of which has two possible outcomes. The simplest example is coin tossing with outcomes *heads* or *tails*. Suppose a coin is tossed 4 times, then there are 16 possible outcomes

HHHH	HHHT	HHTH	HHTT
HTHH	HTHT	HTTH	HTTT
THHH	THHT	THTH	THTT
TTHH	TTHT	TTTH	TTTT

and the essential feature of binomial trials is that all these 16 possibilities are equally likely. The fact that the first three tosses fell *heads* does not make it any more likely that the fourth one will be *tails*. The general public have great difficulty in accepting this. As physicists you know that the motion of the coin is controlled by the laws of dynamics and there is no way that the results of the earlier tosses can affect the motion of the coin on the current toss.

Having accepted that for N tosses the 2^N possible outcomes are equally likely it is simply a matter of counting to compute the probability of getting r *heads*. If *heads* and *tails* are not equally likely, so the probability of *success* is q and *failure* $1-q$ then probability of getting exactly r *successes* in N trials is

$$P_r = \frac{N!}{r!(N-r)!} q^r (1-q)^{N-r}.$$

There are two technical results that physicists need to know about binomial distributions. The mean number of *successes* is (obviously) Nq and (not obviously) the standard deviation is $\sqrt{Nq(1-q)}$. Figures 8 to 10 show examples of the binomial probabilities.

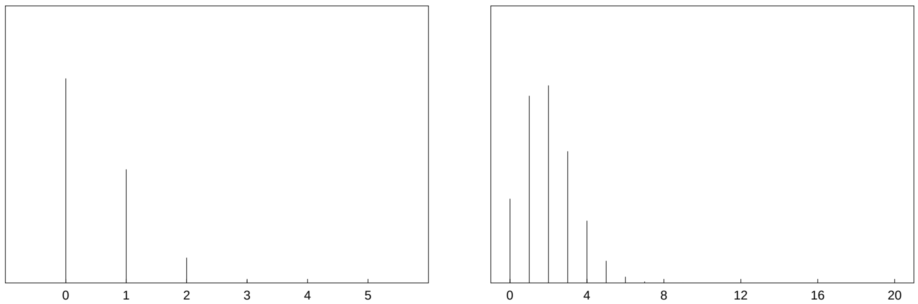

Figure 9: Binomial Distribution $N = 5$ (left) and $N = 20$ (right) with $q = 0.1$

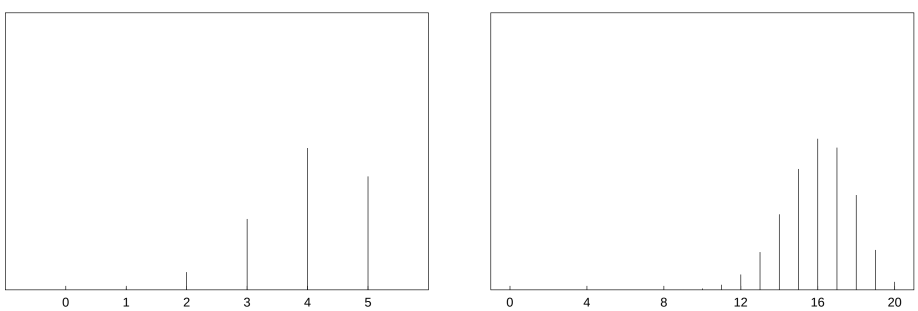

Figure 10: Binomial Distribution $N = 5$ (left) and $N = 20$ (right) with $q = 0.8$

Example: Detector Efficiency You are trying to record the tracks of cosmic ray muons using a bank of spark chambers. These are parallel plate electrodes with gas between; sparks are triggered where ionisation has been left by a fast charged particle. Your chambers are 90% efficient. Common sense suggests that you should require at least three points in order to define a track. How efficient at detecting tracks would a stack of three chambers be?

The probability of all three chambers firing is P_3 with $N = 3$ and $q = 0.9$, which is $0.9^3 = 0.729$, so this system would be 72.9% efficient. Would 4 or 5 chambers give a useful improvement?

The probabilty of *at least* three chambers firing in a bank of 4 is:

$$P_3 + P_4 = \frac{4!}{3!1!} (0.9)^3 (0.1)^1 + 0.9^4 = 0.292 + 0.656 = 0.948,$$

so 94.8% represents a useful improvement.

Questions on the binomial distribution

An astronomer has booked the use of a telescope for a period of 7 nights. The probability of a cloudless night at this time of year is 0.1.

Q4a What is the probability that she as *at least* 1 clear night's viewing? Hint: Calculate the probability of 0 nights.

Q4b What is the chance of exactly 2 nights?

Q4c What is the chance of exactly 3 nights?

Q4d What is the chance of 4 or more clear nights?

A defence system is 99.5% at intercepting ballistic missiles.

Q4e What is the probability that it will intercept all of 100 missiles?

Q4f How many missiles must an aggressor launch to have a better than evens chance of penetrating the defences?

Complete the study of the efficiency of detecting muon tracks in a five chamber stack by showing that the efficiency of recording at least three hits would reach 99% if five chambers were used.

Q4g

In a multiple choice test, students are given a choice of 5 answers on each of 6 questions.

Q4h What is the probability of getting more than half right by chance?

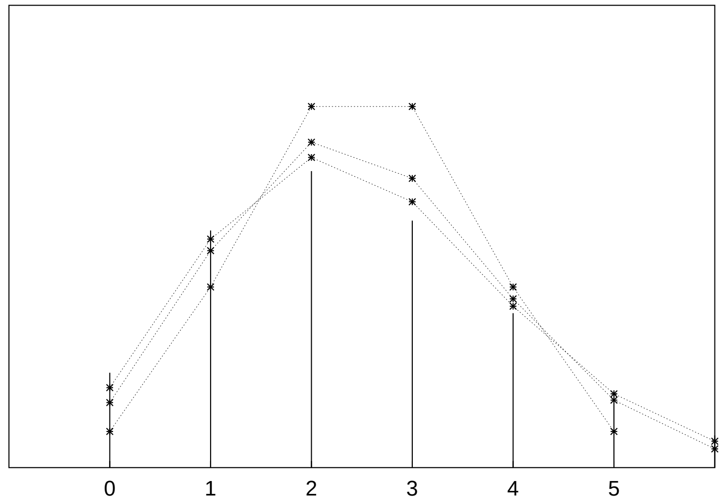

Figure 11: The binomial distribution as $q \to 0$ and $N \to \infty$

5 Random Events and the Poisson Distribution

We are all being bombarded, night and day, by cosmic ray muons. The rate is practically constant at about one per horizontal square centimetre per minute. If we use a detector to count them we find that the number we count in a minute has a definite average but fluctuates up and down around this average, not because of uncertainty in time measurement, but because of the intrinsically random nature of the arrivals. Can we say anything about the magnitude of these fluctuations? If we are detecting say 10 per minute averaged over a few hours, we are not surprised if we detect 11 in a minute, or 12, but should we consider a count of 15 as something out of the ordinary?

We want to know the probability of observing a certain number of events in a certain time interval when we know that the average rate of events is constant. This situation is simlar to the case of the binomial probability distribution, discussed above, in that the outcome is one of two results: a count or no count. But only the successes are counted while we are observing continuously, so how many trials have we made? .

The way round this difficulty notes that when the probability of success in any trial is q the average number of successes in N trials is Nq. By observing continuously we are making a very large number of trials but the probability of observing an event in any one trial is very small. So consider what happens to the binomial distribution when $N \to \infty$ and $q \to 0$ but the product $Nq = \nu$ remains fixed. Figure 11 shows the binomial distribution for a series of values of N and q: $N = 5, q = 0.5$ $N = 10, q = 0.25$ $N = 20, q = 0.125$ together with the limiting distribution, with $\nu = 2.5$ which is shown as vertical bars.

It can be shown that in this limit the binomial distribution becomes

$$P_r = \frac{\nu^r}{r!} e^{-\nu}.$$

and this result is called the Poisson distribution. Figures 5 to 5 shows some examples for $\nu = 0.1,\ \nu = 1.0,\ \nu = 2.0,\ \nu = 5.0,\ \nu = 10.0$ and $\nu = 50$.

You can see that for small ν it is asymmetric, but becomes symmetric for large ν. In fact it becomes approximately Gaussian for large ν. A key result is the standard deviation. For the binomial $\sigma = \sqrt{Nq(1-q)}$, and as $q \to 0$ this approaches $\sigma = \sqrt{Nq} = \sqrt{\nu}$. *This is by far the most important thing to remember about Poisson statistics.*

It is easy in physics problems to get confused between *number* of events and the *rate* at which they are occuring. Suppose events occur at an average rate of λ events per second, and we observe for T seconds, we expect to count $\nu = \lambda T$ events. The problem is usually the other way round; suppose we have counted n events in time T, what can we say about the rate λ? It is pretty obvious that we should estimate the rate by $\lambda = n/T$, but what about its uncertainty? If we counted for another T seconds we would expect to get about n events again, but not exactly the same number because of the intrinsic random nature of the events. The question is how much do we expect the number to fluctuate?

This is got by the following argument. If you get confused it is because the argument is fundamentally muddled, *but remember the results!* If we expected ν events we would not be surprised to get n anywhere in the range $\nu \pm \sqrt{\nu}$. So if we got n events we should take the expected number as $\nu = n \pm \sqrt{n}$.

If you count n events anticipate fluctuations of $\pm\sqrt{n}$.

Now the rate and its error can be computed as

$$\lambda = \frac{n}{T} \pm \frac{\sqrt{n}}{T}.$$

Example Suppose we have counted 100 random events *from a constant source* in one minute, the best estimate of the rate, and its uncertainty is: $100 \pm \sqrt{100} = 100 \pm 10$ counts per minute. If we had observed for four minutes and observed 400 events the estimate of the rate is: $(400 \pm \sqrt{400})/4 = 100 \pm 5$ counts per minute. Note that by observing for longer the uncertainty is reduced.

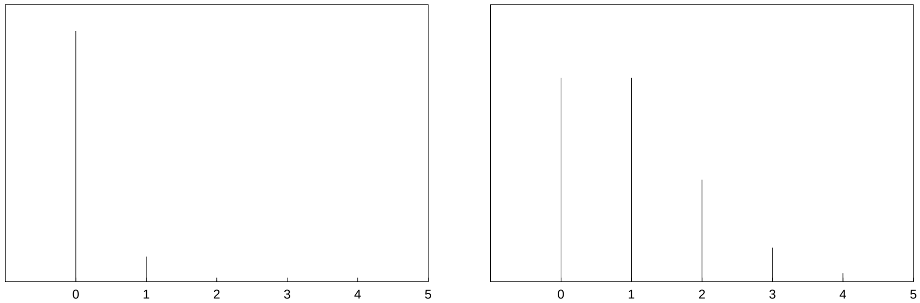

Figure 12: Poisson Distribution $\nu = 0.1$ (left) and $\nu = 1.00$ (right)

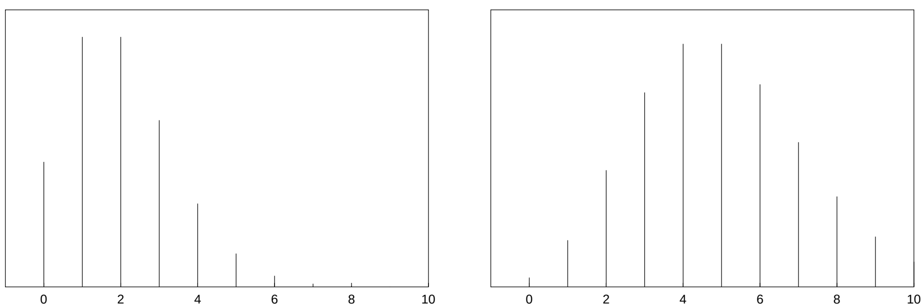

Figure 13: Poisson Distribution $\nu = 2.0$ (left) and $\nu = 5.00$ (right)

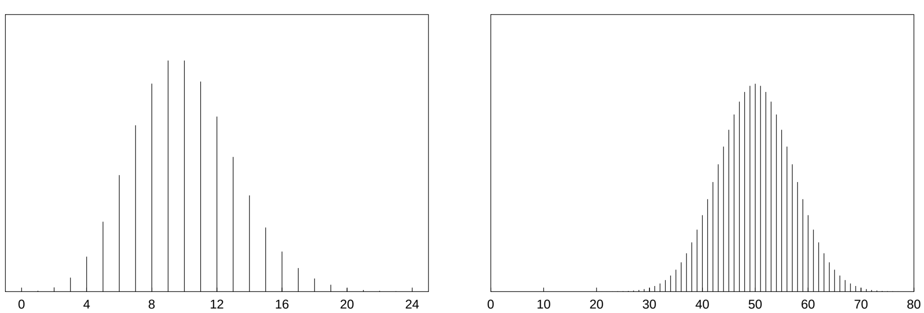

Figure 14: Poisson Distribution $\nu = 10.0$ (left) and $\nu = 50.0$ (right)

Are four readings better than one? Suppose in the above experiment we record the numbers of events in four separate periods of one minute and average the results. Let the numbers be n_1, n_2, n_3 and n_4. The individual estimates of the rate are $n_1 \pm \sqrt{n_1}$, $n_2 \pm \sqrt{n_2} \cdots$ and the result of combining these estimates, adding the errors in quadrature is

$$\frac{n_1+n_2+n_3+n_4}{4} \pm \frac{\sqrt{(\sqrt{n_1})^2+(\sqrt{n_2})^2+(\sqrt{n_3})^2+(\sqrt{n_4})^2}}{4} = \frac{N}{4} \pm \frac{\sqrt{N}}{4}$$

where N is the total number of events recorded. So we get exactly the same answer by regarding the experiment as one measurement over four minutes; the indivdual readings every minute give us no more information.

Questions on Poisson statistics

A detector counts a steady background of cosmic ray muons: in 10 hours 359992 counts are recorded.

Q5a What is the average background rate (in counts per minute) and its uncertainty?

.

In a five minute experiment to measure an artificial radiation source a detector counts both particles from the source *and* cosmic rays. The average background rate determined above will be used to subtract an estimate of the number of counts that are due to the background.

Q5b How many background counts do you expect during this experiment?

Q5c When subtracting this background what uncertainty should you use in the number of background counts: 55, 5 or $\sqrt{5}$?(think hard!)

Q5d If ν the mean number per bin is an integer, are the probabilities for $n = \nu - 1$ and $n = \nu$ always the same?

An experiment involves detecting all the decays from 1 g of ^{147}Sm, Samarium 147, which has a half-life of about 3×10^{18} s.

Q5e How many decays do I need to count to determine the half life to within 0.1%?

Avogadro's number = 6.022×10^{23} mole^{-1}. Atomic weight of ^{147}Sm = 146.91 g mole^{-1}
See section 3.1 for the definition of half life.

Q5f How long do I need to count for?

The following pair of questions really tests if you understand binomial and poisson statistics. The answers are only slightly different so quote them to 3 figures. Read the hints before attempting the questions.

A student is trying to hitch a lift. Cars pass at random intervals, at an average rate of 1 per minute. The probability of a car giving a lift is 1%. What is the probability that the student will still be waiting

Q5g after 60 cars have passed?

Q5h after 1 hour?

Q5i Explain why the answers are different

Hints: One way of looking at this situation is as a series of trials with outcome *success*, getting a lift, or *failure*. Another way of looking at it is as a series of events: getting offered a lift. In order that more than one event occurs you have to assume that the student does not accept the lifts, but this does not affect the calculation.