ERRORS AND THE TREATMENT OF EXPERIMENTAL RESULTS

FJ Lowes

Department *of* Physics
University of Newcastle upon Tyne

PREFACE

These notes were prepared for use in an undergraduate physics teaching laboratory. My aim was to provide a simple, *but correct*, introduction to the fundamentals of "error theory" appropriate to a physics context, and to encourage an *understanding* of the subject rather than a blind reliance on applying formulae, often incorrectly!

The notes were first produced in the days when any calculation was inevitably tedious (it had to be done by hand, by slide-rule, or by using logarithm tables - if anyone now knows what those things were!), so one had to consider carefully how much calculation was justified. However, even with all the facilities available now, such as spread sheets and other programs, data-entry is still a not insignificant chore, and it is still important that the user of the program understands what the number-crunching is actually doing. Just as with a laboratory "black-box" measuring instrument the physicist needs to understand its limitations such as input and output "impedances", frequency-response, etc., then so with a statistics/calculation program the user needs to understand the assumptions it makes about the input data, the limitations of its algorithm, and exactly what numbers it outputs. So when preparing the current version of these notes, I decided not to omit very much.

The data which are going to be analyzed come from measurements, and an integral part of our teaching was to help the students to understand enough about "experimentation" (measurement theory and experiment design) so that the results they obtained were as free from errors as reasonably possible. Much of this training has to be "hands-on", but the last section of the notes is an attempt to summarize those aspects most directly relevant in the present context.

I am grateful to very many colleagues over many years for much helpful discussion, without which my present understanding of the subject would not have been possible.

F.J. Lowes

April 2001

CONTENTS

Errors and the Treatment of Experimental Results

1.   INTRODUCTION

All measurements are subject to error, and it is important to know the effect of these errors on the final result. The statement that $F = 30$ is, by itself, of little value. It may mean that F lies between 25 and 35, or between 29.5 and 30.5. When a result is given it is necessary to give also an estimate of its accuracy.

The errors present in a measurement may be divided into three classes, Careless, Systematic and Random.

a)   **Careless errors**: Due to mistakes in reading scales, careless setting of crosswires, etc. They will usually be obvious when the reading is repeated. Every measurement, however simple, should therefore be made at least twice.

b)   **Systematic errors**: These effect every reading, made under the same conditions, to the same extent, and so produce no spread of readings on repetition.

They may for example, be due to

(i)   Instrumental causes. (Calibration assumed linear when it is not. Scale divisions non-uniform. Scale incorrect because of change of condition, for example scale expands as temperature rises. Instrument zero varies with time.)

(ii)   Personal errors. (In signalling the coincidence of a crosswire and a moving image the signal may be made consistently too early or too late. When using a stop-watch the delay in starting the watch may be different from the delay in stopping it. The colour response curve of individual eyes is not the same as the standard curve.)

(iii)   Specimen not ideal as assumed in theory. (Specimen may be non-isotropic; for example the refractive index may vary with direction. Size may not be constant; for example the particles of lycopodium powder have a range of diameters. The shape may not be perfect; for example a "sphere" will not be perfectly spherical.) Appropriate measurements on the specimen (for example diameters of powder) will show a spread of values. This spread is real, and should not be confused with any additional spread due to random errors of measurement. If the experiment introduces some averaging (for example diffraction haloes of lycopodium, determination of viscosity by a sphere falling through a liquid) then it may be possible to use some average value. The correct average is *not* necessarily the *arithmetic* mean of the observed values, and if the wrong average is used a systematic error is introduced.

(iv)   Errors of definition. (The act of measurement may alter the quantity being measured; for example a moving coil voltmeter may alter the potential difference across the terminals to which it is connected because of the current it takes. The quantity measured may not be exactly that which it was intended to measure; for example a thermocouple in a furnace may be at a temperature different to that of the specimen.)

(v)   Errors of interpretation. (An inaccurate value of a constant may be used; for example insufficiently accurate value of π, the wavelength of the wrong spectrum line. The theory used may not be applicable; for example

the perfect gas law may be used in conditions where it is not sufficiently accurate; the phenomenon observed, or some part of it, may be due to some other, probably unsuspected, cause.)

(vi)   Systematic variation of conditions. (The zero of an ammeter may be changed by the stray magnetic field of an electromagnet. The temperature of the cold junction of a thermocouple may be increased by radiation from the furnace whose temperature is being measured.)

(vii)   Non-systematic variation of conditions. These are not strictly systematic errors according to the definition given above, as with completely constant conditions the same result would always be obtained. However, the dependence of some property of the apparatus on some external condition is often unsuspected, or forgotten, and it may not be realised that the relevant condition is varying. (For example in determining elastic constants by measuring deformations under load, then temperature variations may give rise to spurious deformations.) This may give a spread of results for what was thought to be constant conditions and it may not be valid to treat these as random values. If the dependence has been realised then an attempt can be made to keep the condition constant (for example use a "constant" temperature bath) and the small spread of results due to the inevitable small fluctuations of the condition may be treated as random errors.

These are only some of the many possible sources of systematic error, and the headings suggested are arbitrary and overlap to a certain extent.

c)   **Random errors**: These give a spread of readings on repetition with apparently the same conditions, and this spread can be used to estimate the inaccuracy of the observations due to random errors. They are such that the average value of the individual errors of observation approaches zero as the number of observations becomes very large.

They are due, for example, to the inability of the observer or apparatus to determine a quantity exactly (for example because of the finite width of graduations, the blurring of an image by lens aberrations or diffraction, the finite resolving power of the eye, the essentially digital character of such devices as stop watches), to the fact that the quantity being measured is not precisely defined (for example the finite width of a spectrum line), to the essentially random nature of the phenomenon (for example the random emission of alpha-particles), to random fluctuations of experimental conditions (for example vibration, small temperature fluctuations), and to fundamental limits of measurement (for example Brownian motion, the atomicity of matter and charge, and the Heisenberg uncertainty principle).

Careless errors should almost always be detected in careful experimentation.

Unfortunately it is never possible to be certain that systematic errors are not present, but much can be done to reduce their magnitude. Much of the work of experiment design and planning should in fact be directed to allowing the detection or compensation of systematic errors, and this aspect is discussed in section 5. Unfortunately, being systematic, and of unknown frequency distribution, there is no objective way of estimating the effect of individual contributions to the result of a calculation, or of combining them with the random errors also present; each case must be treated (subjectively) on its merit.

Fortunately, because random errors *are* random, it is possible to make various predictions about their *average* effects, and a large body of theory exists on this subject. Most of the rest of these notes is an attempt to introduce some of the simpler and more useful results of theory of random errors. It must be emphasised that the apparent neglect of systematic errors is *not* due to their being unimportant, but to their being unmanageable!

The following sections deal with some aspects of the theory of random errors and the method of least squares, some graphical and numerical methods of handling data, and some aspects of experimental design and procedure.

These and similar topics are treated at a comparable level in the books listed in the Bibliography.

2.   THEORY OF RANDOM ERRORS

In this section we discuss the theory of random errors and ignore the possible effects of systematic errors. It is probably impossible to give rigorous definition of randomness; one possible definition has been given above. Another equivalent definition is that the errors are random if the arithmetic mean of a very large number of the observations subject to error, carried out under nominally identical conditions, is the same as would be obtained if there were no errors; this is obviously not true for systematic errors.

2.1   Some Statistical Concepts

Experimental observations may be of a quantity which is essentially constant, for example the mass of a calorimeter, individual observations giving different values because of the errors of observation; alternatively the quantity itself may be different for each observation, for example the height of soldiers, the errors of measurement usually being negligible. In the physical sciences we are usually concerned with observation of the first type. Most of the older work on statistics has been based on the second type of observation; the methods can be applied to the first type of observation with only minor changes, but this different emphasis sometimes leads to confusion in the physical interpretation of the results of statistical procedures.

Let a set of n observations, for example heights of soldiers, mass of a calorimeter, have values x_1, x_2, x_3 .... x_i .... x_n. In either case we wish to have some simple way of characterising the distribution of observations without having to give the individual observations. This is usually done by specifying some mean value of the distribution, and some parameter which characterises the spread about the mean.

The mean value almost always used is the ARITHMETIC MEAN $\bar{x} \equiv \sum_{i=1}^{n} x_i / n$.

We can then define the deviation of the individual readings from the mean by $d_i \equiv x_i - \bar{x}$, and the spread parameter can be some mean of d_i. We cannot use the arithmetic mean of d_i as this is zero because of the definition of $\bar{x}$. The d_i must first be made all positive, either by taking the modulus, or by squaring.

Hence we can define   MEAN DEVIATION $\eta \equiv \sum |d_i| / n$

          and   ROOT MEAN SQUARE DEVIATION $\mu \equiv \left(\sum d_i^2 / n \right)^{1/2}$.

The r.m.s. approach is usually preferred, and μ is usually called the STANDARD DEVIATION. μ^2 is called the VARIANCE.

Very often our observations are only a comparatively small sample of a much larger "population" (only a finite number of observations were made on the calorimeter, not all soldiers were measured). In the case of observations of the second type, for example heights of soldiers, the population can itself be characterised by a (true) MEAN x_0, MEAN DEVIATION v, and STANDARD DEVIATION σ:

$$x_0 \equiv \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$v \equiv \frac{1}{N} \sum_{i=1}^{N} |x_i - x_0|$$

and
$$\sigma^2 \equiv \frac{1}{N} \sum_{i=1}^{N} (x_i - x_0)^2$$

where N is the total number in the population (often assumed to be infinite). In the case of the calorimeter the mass is a characteristic of the calorimeter, independent of the particular observations that are made. In this and similar cases of the first type we put

$$x_0 = \text{"True" value of observed quantity,}$$

and define
$$v \equiv \lim_{n \to \infty} \left[\sum |x_i - x_0| / n \right],$$

$$\sigma^2 \equiv \lim_{n \to \infty} \left[\sum (x_i - x_0)^2 / n \right].$$

It must be emphasised that the exact numerical values of these parameters are *never* known to the experimenter; one of the main aims of statistical theory is to provide methods by which approximations to these values can be estimated. In general the values $\bar{x}$ and μ determined from a particular sample of n will not be the same as x_0 and σ, and will be different for different samples. We now show how $\bar{x}$ and μ can be used to estimate x_0 and σ.

Estimation of x_0

We may define the deviation of an individual from the true mean or true mean or true value, whichever the case may be, as $e_i \equiv x_i - x_0$.

Again it must be emphasised that this deviation (sometimes called the error) can *never* be known for an individual observation; the best we can do is obtain an estimate of some mean (in practice the root mean square) value of it if averaged over a large number of observations.

For a particular sample of n
$$\bar{x} = \frac{1}{n} \sum x_i = \frac{1}{n} \sum (x_0 + e_i) = x_0 + \frac{1}{n} \sum e_i.$$

Now
$$\lim_{n \to \infty} \left[\frac{1}{n} \sum e_i \right] = 0 \qquad \text{for soldiers by definition}$$

and
$$\lim_{n \to \infty} \left[\frac{1}{n} \sum e_i \right] = 0 \qquad \text{for the calorimeter (provided there is no systematic error)}$$

so
$$\lim_{n \to \infty} \bar{x} = x_0 ;$$

if we take a large number of samples the average value of $\bar{x}$ is x_0. In fact it can be shown that for any random sample the best estimate we can make of x_0 is $\bar{x}$, that is $x_{0\,est} = \bar{x}$.

6

Uncertainty of the estimate of x_0.

Thus the result of an experiment to determine x_0 is a value of $\bar{x}$, and we would like to know how good an estimate of x_0 this is. Just as the observations themselves can be represented by a mean, $\bar{x}$, and a standard deviation, μ, so also can the values of $\bar{x}$ (obtained from successive samples of n observations, that is repetitions of nominally identical experiments in each of which the same number of observations were made) be represented by their own mean and standard deviation. We have already seen that the mean value of $\bar{x}$ is x, and we will now deduce the standard deviation of the values of $\bar{x}$ about this mean.

For a sample of n the deviation of the observed mean $\bar{x}$ from the true value of x_0 is

$$E = (\bar{x} - x_0) = \Sigma e_i / n$$

and

$$E^2 = (x - x_0)^2 = (\Sigma e_i)^2 / n^2 = \left(\sum_i e_i^2 + 2 \sum_{i \neq j} e_i e_j \right) \Big/ n^2 .$$

The value of $2\Sigma e_i e_j$ will, with any particular sample of n, be small compared with Σe_i^2, and for different samples it will be positive and negative. Therefore, if we average over a randomly large number of samples, each of size n,

$$E^2_{average} = \left[\sum_{i=1}^n e_i^2 \Big/ n^2 \right]_{average}$$

$$= \sum_{i=1}^n \left(\left[e_i^2 \right]_{average} \right) \Big/ n^2 = \sum_{i=1}^n \sigma^2 \Big/ n^2 = \sigma^2 / n .$$

If the standard deviation of the observed means $\bar{x}$ about the "true" value x_0 is σ_m then

$$\sigma_m^2 = E^2_{average} = \sigma^2 / n$$

or

$$\sigma_m = \sigma / \sqrt{n} .$$

As n, the number of individual observations which are measured to give $\bar{x}$, is increased then σ_m decreases. This is because the individual observation are scattered about x_0 at random, and the more there are the more likely it is that the deviations from x_0 will tend to cancel.

In a given experiment we may have only one value of $\bar{x}$, but the relation between σ_m and σ enables us to estimate a typical value of the uncertainty of $\bar{x}$ provided we have an estimate of the value of σ. We may say that the result of our experiment is that x_0 is $\bar{x} \pm \sigma_M$. The $\pm \sigma_m$ is a notation implying that σ_m is the standard deviation of the x about x_0 (and hence also of x_0 about the $\bar{x}$). It does *not* imply that x_0 *must* lie in the range $\bar{x} - \sigma_m \leq x_0 \leq \bar{x} + \sigma_m$; as σ_m is a (root mean square) mean of the differences $(\bar{x} - x_0)$ then obviously an appreciable fraction of these differences must be *larger* than σ_m. We do not know if our particular $\bar{x}$ has a smaller or larger deviation than σ_m; we can give the probability of the deviation lying in a particular range only if we know the distribution function which the original observations obey (see later).

7

(Some authors call σ_m the STANDARD ERROR, but there are so many different notations that it is safer not to rely on the use of a name whose meaning may vary, but to call it what it is, the standard deviation of the mean.)

Estimation of σ

Although the average value of $\bar{x}$ is x_0, the average value of μ is **not** equal to σ. This is because μ is the r.m.s. deviation of the sample observations about the sample mean $\bar{x}$, while σ is essentially the r.m.s. deviation of the observations about the "true" mean x_0, and in general $\bar{x} \neq x_0$. As the r.m.s. deviation of a set of numbers is a minimum when taken about their own arithmetic mean (see under "Least Squares" below), μ will on average be less than σ.

If μ is obtained from a sample of n then it can be shown that the best estimate we can make of σ is

$$\sigma_{est}^2 = \frac{n}{n-1}\mu^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x - \bar{x})^2 .$$

(That the divisor is $(n-1)$ and not n can be thought of as due to the fact that while we started with n independent pieces of information (n degrees of freedom) one of these was used in determining $\bar{x}$, which appears in the definition of μ.) This relation between σ and μ will now be derived.

For observations of the second type (soldiers) the standard deviation σ of the population is defined by

$$\sigma^2 \equiv \sum_{i=1}^{N}(x_i - x_0)^2 \Big/ N$$

where N is the total number of the population; there is a similar definition in the case of observations of the first type (calorimeter). At any one time we have access only to a small sample, n, of the population, so we must average over all the possible samples to obtain the correct value of σ^2. We may put

$$n\sigma^2 = \left[\sum_{i=1}^{n}(x_i - x_0)^2\right]_{average}.$$

Now, for a particular sample,

$$\sum(x_i - x_0)^2 = \sum[(x_i - \bar{x}) + (\bar{x} - x_0)]^2$$

$$= \sum(x_i - \bar{x})^2 + 2(\bar{x} - x_0)\sum(x_i - \bar{x}) + n(\bar{x} - x_0)^2.$$

The first term on the right hand side is equal to $n\mu^2$. The second term is zero as $\sum(x_i - \bar{x}) = 0$ by the definition of $\bar{x}$, so

$$\sum_{i=1}^{n}(x_i - x_0)^2 = n\mu^2 + n(\bar{x} - x_0)^2 .$$

Therefore $\quad n\sigma^2 = \left[\sum_{i=1}^{n}(x_i - x_0)^2\right]_{average} = n\left[\mu^2\right]_{average} + n\left[(\bar{x} - x_0)^2_{average}\right]$

8

$$= n \lfloor \mu^2 \rfloor_{average} + n \sigma_m^2$$

$$= n \lfloor \mu^2 \rfloor_{average} + \sigma^2 \quad \text{from above}$$

and hence $\qquad \lfloor \mu^2 \rfloor_{average} = \left(\frac{n-1}{n} \right) \sigma^2 .$

It follows that for a particular sample of n observations, the best estimate we can make of σ is

$$\sigma_{est}^2 = \left(\frac{n}{n-1} \right) \mu^2 .$$

We can then use this estimated value of σ to estimate the value of σ_m from

$$\sigma_m^2 = \sigma^2 / n .$$

Uncertainty of the estimate of σ

The value $\sigma_{est}^2 = n \mu^2 / (n-1)$ is itself uncertain, and it is possible to predict by how much, on average, it will vary from one set of observations to another. This will not be done here as in Physics we usually want σ only as a guide to the uncertainty of $\bar{x}$, and are not interested in σ for its own sake. However it is important to realise that σ_{est} is itself uncertain, and there is no point in quoting it to a large number of significant figures, particularly when n is small, as it usually is.

Also the difference between $n \mu^2 / (n-1)$ and μ^2 is always less than the uncertainty of σ_{est}^2, so that in practice it rarely matters if the $n/(n-1)$ factor is omitted if we are using σ_{est} only as a guide to the uncertainty of $\bar{x}$.

These points are discussed again later.

Weighting of observations

Using the arithmetic mean $\bar{x}$ as the best estimate of x_0 implied that all the observations x_i were of equal accuracy (more strictly, that all the readings had the same standard deviation σ, or that they were all a sample drawn from the one population). However, there is sometimes reason to believe that the observations x_i are not all of equal accuracy, and in this case it is obviously desirable that the better observations should be given more weight than the poorer ones. If it is possible to assign a numerical "weight" w_i to each x_i, then the best estimate of x_0 is the weighted mean $\tilde{x}$ given by

$$\tilde{x} \equiv \sum_{i=1}^{n} w_i \, x_i \left/ \sum_{i=1}^{n} w_i \right. ,$$

and the (weighted) standard deviation $\tilde{\mu}$ of the observations about $\tilde{x}$ is

$$\tilde{\mu}^2 \equiv \sum w_i (x_i - \tilde{x})^2 / \sum w_i ,$$

whence $\qquad \tilde{\sigma}_{est}^2 = n\,\tilde{\mu}^2 / (n-1)$

and $\qquad \tilde{\sigma}_{\tilde{x}}^2 = \tilde{\sigma}^2 / n \cong \Sigma\, w_i (x_i - \tilde{x})^2 / (n-1)\Sigma\, w_i$.

The major difficulty in applying these formulae is the choice of weights w_i , and this will now be discussed for some typical situations.

The values x_i are sometimes themselves the arithmetic means of several observations. (Perhaps an experiment was repeated several times on one day, a different number of times on another day, and so on, and only the daily mean and numbers of repetitions are available.) If it is believed that each repetition resulted in an observation of equal accuracy then the obvious weight w_i to use is the number of observations n_i contributing to the value x_i . In fact taking the weighted mean $\tilde{x} \equiv \Sigma\, w_i x_i / \Sigma\, w_i = \Sigma\, n_i x_i / \Sigma\, n_i$ is exactly equivalent to taking the arithmetic mean of all the original observations. Thus we put w_i equal to (or proportional to) n_i .

If, as above, each x_i is the mean of n_i observations of standard deviation σ we would expect to find σ_i , the standard deviation of x_i , to be given by $\sigma_i^2 = \sigma^2 / n_i$. Putting the weights w_i proportional to $1/\sigma_1^2$ would then have the same effect as above. If, in fact, the x_i were not necessary the result of different numbers of observations, but did have different standard deviations σ_i , it is usual to put w_i equal to, or proportional to, $1/\sigma_i^2$. (Such a situation might arise if there were different amounts of disturbance on different days, or if different observers were used.)

There are also occasions when each of the x_i is essentially a single measurement, for which a standard deviation cannot be determined, but where there is other evidence that they were not of equal accuracy. The weights w_i must then be assigned on whatever evidence there is, but this is very difficult to do in an objective manner.

Validity of the theory

All the results of this section are valid for *any* distribution of *random* errors, and for any value of *n*. They are *not* restricted only to error distributions which are symmetrical, or to the Gaussian distribution, as is often implied by text-books.

2.2 Distribution Functions

As we have seen in the preceding section, several useful results can be obtained without any knowledge of how the errors are distributed, provided only that they are random. However a knowledge of the distribution can be useful.

Derivation of the distribution function

Let the complete set of possible observations have values $x_1, x_2 \dots x_i \dots x_N$.

As above $x_0 \equiv \sum_{i=1}^{N} x_i \Big/ N$ and $\sigma^2 \equiv \sum_{i=1}^{N} (x_i - x_0)^2 \Big/ N$

However some of the x_i may have the same value. If there are n_a observations with value x_a, n_b with x_b, and in general n_r with x_r, then we have the equivalent definitions

$$x_0 \equiv \sum_r n_r x_r \bigg/ N \quad \text{and} \quad \sigma^2 \equiv \sum_r n_r (x_r - x_0)^2 \bigg/ N$$

where $N = \sum n_r$ Even if there are no x_i exactly the same it may be convenient to lump together all those in a given range of values, so n_r would then be the number with values between x_r and $x_r + \delta x$.

When large numbers of observations are considered it is more convenient to use the *relative* frequency of occurrence $f_r \equiv n_r / N$. Then $x_0 \equiv \sum f_r x_r$ and $\sigma^2 \equiv \sum f_r (x_r - x_0)^2$ as $\sum f_r = 1$. (A small correction is necessary if the interval δx is not small compared with σ.)

In the limit as the number of possible observations tends to infinity, and the interval δx tends to zero, we may define the probability density function $f(x)$ such that the proportion of observations with values between x and $x + dx$ is given by $df = f(x)dx$.

(This is also the probability that any one observation will lie in this range.) It follows from this definition that

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

and that the proportion of observations with values between x_1 and x_2 is

$$\int_{x_1}^{x_2} f(x)dx.$$

Thus $x_0 \equiv \int_{-\infty}^{\infty} f(x)x\,dx$ and $\sigma^2 \equiv \int_{-\infty}^{\infty} f(x)(x - x_0)^2\,dx$.

Note that it is the *product* $f(x)dx$ which is dimensionless; the density function $f(x)$ therefore has the dimensions of $1/x$.

Gaussian or Normal distribution

This is a distribution which it is often convenient to use as an approximation to the distribution of experimental observations. It can be shown that if each observation of a certain fixed quantity is subject to a very large number of independent very small errors, each of them having the same magnitude and having equal probability of being positive or negative then the distribution of the observations about the true value is Gaussian.

In the original notation the form of the distribution is

$$f(x) = \left(h / \sqrt{\pi} \right) \exp\left[-h^2 (x - a)^2 \right]$$

where a is the true value, and h is a spread parameter; h large means that the observations are closely grouped about a.

The ARITHMETIC MEAN x_0 is given by $\qquad x_0 = \int f(x)x\,dx = a$.

The STANDARD DEVIATION is given by $\qquad \sigma^2 = \int f(x)(x-x_0)^2\,dx$.

Put $\qquad (x-x_0) = e$, $\quad$ so $\quad f(x-x_0) = f(e) = (h/\sqrt{\pi})\exp(-h^2 e^2)$

so $\qquad\qquad \sigma^2 = \int f(e)e^2\,de = 1/2h^2$

and $\qquad\qquad \sigma = 1/h\sqrt{2} \quad$ or $\quad h = 1/\sigma\sqrt{2}$.

We can therefore rewrite the Gauss distribution as

$$f(x) = (1/\sigma\sqrt{2\pi})\exp\left[-(x-x_0)^2/2\sigma^2\right]$$
$$= (1/\sigma\sqrt{2\pi})\exp\left[-e^2/2\sigma^2\right],$$

and this is the form in which it is usually quoted nowadays.

The MEAN DEVIATION v is given by

$$v = \int_{-\infty}^{\infty} f(x)|x-x_0|\,dx = \int_{-\infty}^{\infty} f(e)|e|\,de = \sqrt{\frac{2}{\pi}}\,\sigma = 0.80\sigma \quad \text{very nearly.}$$

The probability of x lying in the range

$$(x_0 - s\sigma) \leq x \leq (x_0 + s\sigma)$$

is $\qquad P(s\sigma) = \displaystyle\int_{-s\sigma}^{s\sigma} f(e)\,de = 2\int_{0}^{s\sigma} f(e)\,de = 2\int_{0}^{s\sigma}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)\exp\left(\frac{-e^2}{2\sigma^2}\right)de$.

A function erf (z) was defined by

$$\text{erf}(z) = \left(\frac{2}{\sqrt{\pi}}\right)\int_{0}^{z}\exp(-t^2)\,dt ,$$

so that putting $\qquad t = e/\sigma\sqrt{2} \quad$ gave $\quad P(s\sigma) = \text{erf}\left(\dfrac{s}{\sqrt{2}}\right)$.

However most modern tables give $\quad P(s\sigma) = \sqrt{2/\pi}\displaystyle\int_{o}^{s}\exp\left(-\tfrac{1}{2}t^2\right)dt \quad$ directly, where now $t = e/\sigma$; the following tables

show how $P(s\sigma)$ and s are related for some useful simple cases.

s	$P(s\sigma)$	$1 - P(s\sigma)$
1	0.682	0.318
2	0.954	0.046
3	0.997	3.10^{-3}
4		7.10^{-5}
5		6.10^{-7}

$P(s\sigma)$	s
0.5	0.674
0.9	1.64
0.99	2.6
0.999	3.3

(When dealing with distributions of errors, a spread parameter sometimes used in the past was the PROBABLE ERROR. This was defined such that half of the population lay in the range $(x_0 - P.E.) \le x \le (x_0 + P.E.)$; therefore for a Gaussian distribution $P.E. = 0.674\ \sigma$. However the use of the "probable error" can easily lead to confusion, and it is much better to use only the standard deviation.)

Estimation of x_0 and σ for a Gaussian distribution

As for any distribution, the results obtained in section 2.1 are applicable. For n observations taken at random from a Gaussian distribution the best estimate of the "true" value which can be obtained is $\bar{x}$, and the uncertainty of this estimate can be expressed in terms of its standard deviation about x_0, that is $\sigma_m = \sqrt{n^{1/2}}$. We may write $x_0 = x \pm \sigma_m$.

It can be shown that for sets of observation drawn from a Gaussian distribution the distribution of $\bar{x}$ about x_0 is itself Gaussian (with a standard deviation σ_m), so the tables given above can be used. They show, for example, that the probability that x_0 lies in the range $\bar{x} - \sigma_m \le x_0 \le \bar{x} + \sigma_m$ is 68%, and that the probability that x_0 lies outside the range $\bar{x} - 2\sigma_m \le x_0 \le \bar{x} + 2\sigma_m$ is only 5%.

The best estimate of σ is given by

$$\sigma_{est}^2 = \mu^2 n/(n-1) = \sum (x - \bar{x})^2 /(n-1)$$

and this estimate is also uncertain. For sets of n observations drawn from a Gaussian distribution it can be shown that the standard deviation of σ_{est} is approximately $\sigma/\sqrt{2(n-1)}$ if $n \ge 5$. We may write as a result of our experiment

$$\sigma = \mu\sqrt{\frac{n}{n-1}}\left(1 \pm \frac{1}{\sqrt{2(n-1)}}\right).$$

13

The distribution of σ_{est} is *not* Gaussian so we cannot use the probability tables given above. However, for example, for 10 observations there is a 97% probability that σ_{est} lies within ±50% of the true σ, and for 50 observations there is a 95% probability that it lies within ± 20%.

In physics σ is used primarily to estimate σ_m, so the standard deviation of σ is not usually quoted as such, but it *is* necessary to specify the number of observations upon which the quoted x_0 and σ_m are based.

The best estimate of σ_m is then

$$\sigma_m^2 = \sigma^2/n \cong \Sigma \left(x - \bar{x}\right)^2 / n\left(n - 1\right).$$

The best estimate of the mean deviation ν is

$$\nu_{est} = \sqrt{n/(n-1)}\,\eta \cong \sqrt{1/n(n-1)} \sum_{i=1}^{n} \left|x - \bar{x}\right|.$$

With a large sample the hand computation of μ can be tedious. If the observations follow a Gaussian distribution it is sufficient to compute the mean deviation η and estimate the standard deviation σ by

$$\sigma = \sqrt{\pi/2} \quad \sqrt{n/(n-1)}\,\eta = \sqrt{\pi/2n(n-1)} \sum_{i=1}^{n} \left|x - \bar{x}\right|$$

or $\qquad \sigma \cong \left[1.25/\sqrt{n(n-1)}\right]\sum \left|x - \bar{x}\right|$

and similarly $\qquad \sigma_m \cong \left[1.25/n\sqrt{n-1}\right]\sum \left|x - \bar{x}\right|.$

These are known as Peters' formulae.

Poisson Distribution

For independent events occurring at random in time (or space), but with a definite average rate of occurrence, this distribution gives the probability that a given number of events will occur in a particular interval of time (or space). It is directly applicable, for example, to counting experiments in nuclear physics, and this discussion is based on events random in time; however the distribution is equally applicable to such problems as the probability of finding a given number of bacteria in a particular sample of water.

If, on the average, x_0 counts occur in a time interval of constant length (the average being taken over a large number of such intervals) the probability that r counts occur in such an interval is

$$P(r) = (x_0)^r \exp(-x_0)/r!$$

(This is not a continuous frequency distribution; $P(r)$ is defined only for integral values of $r \geq 0$, though of course x_0 need not to be integral.) As r cannot be less than zero the Poisson distribution is obviously skew (non-symmetrical about the maximum), and is particularly skew for small x_0.

By definition the mean value of r is x_0. It turns out that the standard deviation σ of r about x_0 is given by $\sigma^2 = x_0$; the Poisson distribution has only one parameter, x_0.

Estimation of x_0 and σ for a Poisson distribution

For a single experiment giving a count of r in the appropriate time interval the best estimate of x_0 is r, and the best estimate of σ^2 is also r. We may say that the result of our experiment is $x_0 = r \pm \sqrt{r} = r\left(1 \pm 1/\sqrt{r}\right)$.

The Poisson distribution is such that repeated counts over short intervals are equivalent to the same total count over the corresponding total interval; no extra information is gained by dividing the total time available into smaller intervals. (In practice such division may well be advisable in order to detect careless errors, and to confirm that the assumption of constant average counting rate is justified.) This property will now be demonstrated. Let n repeated experiments over a time interval for which the (unknown) average count is x_0 give counts $r_1, r_2 \dots r_i \dots r_n$. Treating this first as one experiment we have

$$nx_0 = \sum r \pm \sqrt{\sum r}$$
$$= n\bar{r} \pm \sqrt{n\bar{r}}$$

or $\qquad x_0 = \bar{r} \pm \sqrt{\bar{r}/n}$.

Alternatively, treating this as n separate experiments, as the Poisson distribution obeys our criterion for random errors we have $x_0 = \bar{r} \pm \sigma_m = \bar{r} \pm \sigma/\sqrt{n}$.

Now the individual estimates of σ^2 are r_i, so the best estimate of σ^2 is $\bar{r}$, giving $x_0 = \bar{r} \pm \sqrt{\bar{r}}/\sqrt{n}$ as above.

For large x_0 (say $x_0 > 20$) the Poisson distribution curve near its maximum closely resembles a Gauss distribution having mean x_0 and standard deviation $\sqrt{x_0}$; in this case the probability tables given can be used.

Statistics of small samples

For any distribution, if the estimates ($\bar{x}$ and μ) of x_0 and σ are based on only a small number of observations (say less than $5 - 10$) then caution is needed in using the ordinary tables of probability. Because the standard deviation μ will on occasion be much less than σ, then the probability of the difference $|\bar{x} - x_0|$ being larger than (say) $r\mu$ is bigger than might be expected. There are tables giving such probabilities for the means derived from a small number n of observations which obey a Gauss distribution (the 't' distribution).

15

Significance tests

There are tests available which enable one to say if it is reasonable to assume that a given set of observations were drawn at random from a specified type of distribution.

There are also tests which enable one to say if it is reasonable to assume that two or more sets of observations (for example from repeated experiments) were drawn from distributions (or specified type) with the same x_0, that is that the sets of observations were of the same quantity. (For a Gauss distribution this is the 't' test.) Similarly one can test if the standard deviations obtained in repetitions of the experiment are consistent. (For a Gauss distribution this is the 'χ^2' test.)

2.3 Combination of Errors

So far we have considered the observation of a single quantity when the observation is subject to random error. However in most experiments the end result is a derived quantity which is some known function of the observed quantities; for example in Poiseuille's experiment

$$\text{viscosity} = \frac{\pi \times \text{pressure drop} \times (\text{diameter of tube})^4 \times \text{time}}{128 \times \text{length of tube} \times \text{volume of liquid}} .$$

And in an experiment using Ohm's law

$$\text{resistance} = \frac{\text{potential difference}}{\text{current}} .$$

We now consider how the uncertainties of the observed quantities contribute to the uncertainty of the derived quantity. (In some books this subject is called the "Propagation of Errors".)

It is convenient to start by discussing the case of (known) systematic errors of observation, and then to extend this to (individually unknown) random errors.

Systematic Corrections

Suppose the required quantity F is some function of the observed quantities x, y, z ... , that is $F = F(x, y, z...)$. Then if x is altered by a small amount δx, y by δy, etc., the resultant change in F is, to first order in δx, δy etc.,

$$\delta F = \left(\frac{\partial F}{\partial x}\right)\delta x + \left(\frac{\partial F}{\partial y}\right)\delta y + \left(\frac{\partial F}{\partial z}\right)\delta z + \dots .$$

For example for $\quad F = ax + by - cz ,\qquad \delta F = a.\delta x + b.\delta y - c.\delta z$

and for $\quad F = \dfrac{x^a y^b}{z^c} ,\qquad \dfrac{\delta F}{F} = a.\dfrac{\delta x}{x} + b.\dfrac{\delta y}{y} - c.\dfrac{\delta z}{z}$

These are the types of expressions to use if we wish to correct *F* for some *known small* alteration in *x*, *y*, etc. For example if a calculation has been performed with a slightly wrong value of a constant it may be much quicker to use such a correction formula than to repeat the whole calculation.

Random Errors

The general expression for δF given above is of course true for any one set of *x*, *y*, *z* ... observations. However in general we do *not* know either the sign or magnitude of the *individual* values of $\delta x . \delta y$ ... , but we have only an estimate of their root mean square values, which we have called the standard deviations of *x*, *y*, ... For random errors therefore we never know the individual values of δF, and it is reasonable to specify only the root mean square value (which we call the standard deviation of *F*) which would be obtained from a large number of sets of observations.

For one observation

$$(\delta F)^2 = \left(\frac{\partial F}{\partial x}\right)^2 \delta x^2 + \left(\frac{\partial F}{\partial y}\right)^2 \delta y^2 + 2\left(\frac{\partial F}{\partial x}\right)\left(\frac{\partial F}{\partial y}\right)\delta x \delta y + 2\left(\frac{\partial F}{\partial x}\right)\left(\frac{\partial F}{\partial z}\right)\delta x \delta z + ... ;$$

the derivatives are to be evaluated at the "correct" values of *x*, *y*, *z* ... and can be considered to be simply numerical constants.

$$\therefore \left[(\delta F)^2\right]_{average} = \left(\frac{\partial F}{\partial x}\right)^2 \left[\delta x^2\right]_{average} + ... + 2\left(\frac{\partial F}{\partial x}\right)\left(\frac{\partial F}{\partial y}\right)\left[\delta x \delta y\right]_{average} + ...$$

or

$$\sigma_F{}^2 = \left(\frac{\partial F}{\partial x}\right)^2 \sigma_x{}^2 + ... + 2\left(\frac{\partial F}{\partial x}\right)\left(\frac{\partial F}{\partial y}\right)\left[\delta x \delta y\right]_{average} + ...$$

We have already assumed that the δx are random, the δy are random, etc. If we now also assume that the δx, δy, δz ... are *independent* or *uncorrelated* (that is in any one set of observations of *x*, *y*, *z* ... the value of δx does not depend in any way on the value of $\delta x . \delta z$, ...) then the average value of the products $\delta x . \delta y$, δy, δz, ... will tend to zero as the number of sets of observations increases.

Therefore, when the *x*, *y*, *z* ... have *independent random* errors we have

$$\sigma_F^2 = \left(\frac{\partial F}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial F}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial F}{\partial z}\right)^2 \sigma_z^2 + ...$$

For example, for

$$F = ax + by - oz, \quad \text{we have} \quad \sigma_F{}^2 = a^2\sigma_x{}^2 + b^2\sigma_y{}^2 + o^2\sigma_z{}^2$$

and for

$$F = \frac{x^a y^b}{z^0}, \quad \text{we have} \quad \left(\frac{\sigma_F}{F}\right)^2 = a^2\left(\frac{\sigma_x}{x}\right)^2 + b^2\left(\frac{\sigma_y}{y}\right)^2 + o^2\left(\frac{\sigma_z}{z}\right)^2 .$$

Provided the conditions are satisfied, these results are true regardless of the particular distribution functions obeyed by the errors in *x*, *y*, *z* .... . (If the *x*, *y*, *z* ... are *not* independent, then the value of σ_F can be found only if the mean values of the products (the "covariances") are known.)

If each of the observed quantities x, y, z ... has errors which obey a Gaussian (or Poissonian) distribution, and if F is a linear function of x, y, z..., then the δF also follow a Gauss (or Poisson) distribution. In general however the distribution of δF is not simple, and tables of probability will not be available.

Application of the expression for σ_F

In using the expression for σ_F, the values of σ_x, σ_y ... to be used are those applicable to the values of $x, y,$ ... which are inserted in the expression for F. Thus if the x, y ... are each single observations then the σ_x, σ_y must be known independently (for example from a preliminary experiment). Then

$$\sigma_F^{\,2} = (\delta F / \delta x)^2 \sigma_x^{\,2} + (\delta F / \delta y)^2 \sigma_y^{\,2} + ...$$

Often, however, the value of x which is used to calculate F is the mean of several observations of x, the value of y which is used is the mean of a series of observations of y which were not connected with the observations of x, and similarly for z ... . The observations can be said to be of the form $(x_1, x_2, x_3 ...)$, $(y_1, y_2, y_3 ...)$ ... . In this case $F = F(\bar{x}, \bar{y}, \bar{z}...)$, and the appropriate standard deviations are those of the means, that is $\sigma_{mx}, \sigma_{my},...,$ so

$$\sigma_{*F} = (\delta F / \delta x)^2 \sigma_{mx}^{\,2} + (\delta F / \delta y /)^2 \sigma_{my}^{\,2} + ... $$

(The symbol σ_{*F} has been used to remind us that the F is based on more than one observation each of x, y ... .) For example, in the Poiseuille experiment several observations of flow will be used to obtain a mean flow, several observations of length to obtain a mean length, and several observations of diameter to obtain a mean diameter[†].

Sometimes the repeated observations of x, y, z ... are linked in sets; that is the observations are of the form $(x_1, y_1, z_1 ...)$, $(x_2, y_2, z_2 ...)$ ... . For example, in an Ohm's law experiment in which the current through the resistance is varying, essentially simultaneous measurements of potential difference and current would be made, and the δV and δI would *not* be independent. In this case the correct approach is to calculate

$$F_i = F(x_i, y_i, z_i...) \qquad i = 1, 2, 3 ... n$$

and

$$F = \Sigma F_i / n ,$$

$$\sigma_{mF}^2 \cong \Sigma (F_i - F)^2 / n(n-1)$$

where σ_{mF} is the standard deviation of the F (about the "true" F).

If the observations happen to be of the type (x_i, y_i, z_i), even if the $\delta x, \delta y$ ... are independent, the observations can also be treated as of the form $(x_1, x_2 ...)$, $(y_1, y_2 ...)$ ... . Then both σ_{*F} and σ_{mF} can be calculated, and they should be in reasonable agreement. If they are not then one or more of the assumptions is not valid.

In observations of the type (x_i, y_i, z_i), even if the $\delta x, \delta y$ .... are not independent, there may be independent evidence for the values of σ_x, σ_y ... . For example, in Ohm's Law experiment previous experience in the use of the voltmeter and ammeter may have given estimates of σ_V, σ_I, and it may be reasonable to assume that the time variation does not

---

[†] Strictly, whether one should use for example the mean of (diameter)^4, or (arithmetic mean diameter)^4 depends on the exact conditions of measurement and the theory which gave the function F; in many cases it may not be possible to decide which is correct. However if the difference between the two means is significant then the expression for σ_F is not valid - the difference is of second order in $\delta x/x$ etc and the derivation of the expression for σ_F is correct to first order only.

increase the uncertainty of reading. Then σ_F^2 can be calculated by the first method of this section, and σ_F^2/n compared with σ_{mF}^2. Again a significant difference indicates the incorrectness of one or more of the estimates or assumptions.

The Weighted Mean

We have already seen that a weighted mean $\tilde{x}$ can be defined by $\tilde{x} = \sum w_i \, x_i / \sum w_i$, where w_i is the weight to be attached to the value x_i. Thus $\bar{x}$ is a linear function of the values x_i, the w_i being known constants. If the standard deviations σ_i of the x_i are known (or estimated) we can use the expression for σ_F derived above, putting $F = \tilde{x}$ and $x = x_1$, $y = x_2$ etc. We have

$$\left(\partial F / \partial x_i\right) = w_i / \sum w_i \, ,$$

so
$$\sigma_F{}^2 = \sum w_i{}^2 \, \sigma_i{}^2 \Big/ \left(\sum w_i\right)^2 .$$

If in fact the σ_i are known, it is usual to put $w_i \; \alpha \; 1/\sigma_i{}^2$, giving

$$\sigma_F{}^2 = 1 \Big/ \sum \frac{1}{\sigma_i{}^2} .$$

If there are a reasonable number of the x_i it is also possible to calculate

$$\tilde{\mu}^2 = \sum w_i \left(x_i - \tilde{x}\right)^2 \Big/ \sum w_i$$

and hence

$$\sigma_{\tilde{x}}{}^2 \approx \tilde{\mu}^2 \Big/ (n-1) = \sum w_i \left(x_i - \tilde{x}\right)^2 \Big/ \sum w_i \, ;$$

σ_F and $\sigma_{\tilde{x}}$ are obviously two different estimates of the uncertainty of $\tilde{x}$.

The value of $\sigma_{\tilde{x}}$ depends on what is often called the 'external' consistency of the x_i, whereas the value of σ_F depends only on their 'internal' consistency, that is on evidence internal to the experiment which produced each x_i. If all the assumptions about randomness, independence, and smallness of errors are valid then both σ_F and $\sigma_{\tilde{x}}$ should be estimates of the same quantity, and should therefore have reasonably similar values. If the values of σ_F and $\sigma_{\tilde{x}}$ are significantly different then one or more of the assumptions is not valid.

The more usual situation is when $\sigma_{\tilde{x}}$ is appreciably larger than σ_F, that is the values of x_i are more scattered than would be expected from the evidence of their σ_i. This is usually due to the presence of systematic errors which were constant (or nearly so) during each experiment giving one of the x_i, but which changed between the experiments. Such detection of the presence of systematic errors is a valuable reason for making the comparison.

The converse situation, $\sigma_{\tilde{x}} < \sigma_F$, usually means that the σ_i have been overestimated, probably due to faulty interpretation of the data. (For example the standard deviation of the individual rather than of the mean may have been used by mistake.)

2.4   Limitations of the Use of Error Theory

The quantities x_0, v, σ and $\bar{x}$, η, μ are defined for *any* distribution.

Estimation of x_0 and σ

The relations

$$x_{0_{est}} = \bar{x}$$

$$\sigma_{est}^2 = \mu^2 n/(n-1)$$

and

$$\sigma_m^2 = \sigma^2/n$$

are valid for *any* distribution of *random* errors.

For repetitions of an observation under nominally identical conditions (when any systematic error should be constant) these relations can be used to estimate and reduce the effect of random errors; the value of x_{0est} will of course still contain any systematic error that was present in the observations.

If there is a systematic error which is a function of the experimental conditions then observations made under differing conditions will contain varying amounts of systematic error as well as random error. If the contribution to the spread of readings from the systematic error is small enough so that it can be considered approximately as due to another (virtual) source of random error, then the above relations can be used to estimate a value of x_0 relevant to some average condition, and a value of σ (which will be larger than that due to random error alone) which will probably be a reasonable estimate of the uncertainty. If the contribution from systematic error is large the above relations are not valid, and an attempt must be made to estimate and remove the systematic contribution before applying the relations.

Combination of Errors

The relation $\sigma_F^2 = (\partial F/\partial x)^2 \sigma_x^2 + (\partial F/\partial y)^2 \sigma_y^2 + ...$ is valid for x, y ... having *any* distribution of *random* errors (with standard deviations which are not too large) provided only that the x, y ... are statistically *independent* (uncorrelated).

If the x, y ... are systematically different from their correct values, then so will be F, but the formula can still be used to estimate that part of the uncertainty of F due to random errors.

In laboratory experiments in Physics it is usually fairly obvious whether observations are dependent or not, so if there is no reason to believe that any of the x, y ... are dependent it is usually safe to use the formula. This is not necessary true in other cases.

If any of the x, y ... are known to be dependent, then either the function F must be rewritten in terms of other, independent, quantities, or some other method used. (In practice if the dependence is slight the formula will not be seriously in error.)

Assumption of Particular Distribution Function

The relation $v^2 = 2\sigma^2/\pi$, and the various tables of probability, are valid *only* for the Gauss distribution.

Therefore if one wishes to apply any of these results one should, strictly speaking, confirm that the observations one is using are in fact Gaussian. If there are a large number of observations this is possible, if tedious, but with the comparatively small number of observations usually available it is not possible. The assumption usually has to be made therefore that the distribution is approximately Gaussian. Fortunately this is usually (but not always) a valid assumption.

The basis of the Gauss distribution (a large number of independent small contributions to the error) is not necessary present in experimental work. Also, the Gauss Distribution predicts a finite, if small, proportion or readings with very large errors, and with negative values, the occurrence of which in an experiment would be absurd. However, investigations of several cases have shown that, while the experimental data did not conform exactly to a Gauss distribution, this distribution was a good approximation.

The Gauss distribution is the only one in which the means of sets of observations drawn from the distribution have the same type of distribution as the observations themselves (with of course a smaller standard deviation). For any other distribution the means of sets of observations are always distributed in a way which is more nearly Gaussian than was the distribution of observations. As in Physics we are more usually concerned with the properties of the mean rather than of the individual observations, this is another reason why in many cases we are justified in assuming an approximation to a Gauss distribution.

Fortunately also, the various relations and tests based on the Gauss distribution do in fact give reasonable results even when the actual distribution departs appreciably from the Gaussian one. Thus, unless there is evidence that the distribution of observations is appreciably non-Gaussian, it is usually safe to use results based on the Gauss distribution.

The relation $\sigma^2 = x_0$ is valid *only* for a Poisson distribution. Fortunately, in most applications in nuclear physics the assumption that the counts follow a Poisson distribution is almost always justified, but there are other situations where apparently random events do not obey this distribution.

3.   THEORY OF LEAST SQUARES FITTING

The method of least squares is one way in which an objective estimate of the values of unknown parameters can be obtained from a set of observational data subject to random error.

It is objective in that, if used properly, the values obtained for the parameters and their uncertainties depend only on the data, and not on the whim of the person handling the data. However it does not necessary give the "best" values of the parameters, and its arithmetical complexity is often not justified by the data available.

3.1   The Principle of Least Squares

The principle of Least Squares is most simply illustrated by considering the case of repeated observations (subject to random error) of a fixed quantity. The principle may then be stated as "the best estimate that can be obtained for the value of the quantity is that value which makes the sum of squares of the deviations of the observations from this value a minimum". Using our previous notation, this states that the best estimate y_{0est} of the "true" value y_0 which can be obtained from the set of observations $y_1, y_2 ... y_i ... y_n$ is such that

$$S \equiv \sum_{i=1}^{n} \left(y_i - y_{0est} \right)^2$$

is a minimum. It is easily shown this criterion is in fact satisfied when $y_{0est} = \bar{y}$, the arithmetic mean, a result we have already used.

A more general case is when the observed quantity y is some known function of the observed quantity x, the function involving one or more parameters $a, b ...$, that is $y = f\left(x, a, b...\right)$. If the observations are free from error, and if there are as many pairs of observations (x_i, y_i) as there are parameters $a, b ...$ then we can solve the equations $y_i = f(x_i; a, b ...)$ to obtain exact values of the parameters; more observations would not add to our knowledge. In practice, however, there *are* errors of observation, and y_i does *not* equal $f(x_i; a, b ...)$. We can then only *estimate* the values of the parameters $a, b ...$, and the more observations we have the more accurate (on average) will be our estimate. Also, provided there are more than the minimum number of observations, we can estimate the uncertainties of our estimates of $a, b ...$. The principle of least squares is one way in which these estimates can be made.

In this general case the principle states that the best estimates of the parameters are those that make the sum of squares of the deviations of the experiment points (x_i, y_i) from the fitted curve $y = f(x; a_{est}, b_{est} ...)$ a minimum.

If both x_i and y_i are subject to error the application of the principle is possible but difficult. For example, if the scales of an (x, y) graph are chosen so that the standard deviations of x and y are represented by equal distances, then the appropriate deviations to be squared are the lengths of the perpendiculars from the points (x_i, y_i) to the curve $y = f\left(x; a_{est}, b_{est}\right)$.

Fortunately, in many cases it is reasonable to assume that the errors in the x_i are much smaller than the errors in the y_i, so that the former can be neglected. In these cases the appropriate deviations are the distances in the y direction between the observed y_i and the fitted curve; the estimate $a_{est}, b_{est}, ...$ are to be chosen so as to minimise the sum

$S \equiv \sum\limits_{i=1}^{n} \left[y_i - f\left(x_i; a_{est}, b_{est}, ...\right)\right]^2$. In all that follows it will be assumed that the observations of x are *not* subject to error, and we will use this expression for S.

The appropriate values of the parameters *can* be found simply by trial and error, but usually it is easier to use the algebraic method which follows.

If the form of the function $f(x; a, b ...)$ is known analytically then the sum S can be also expressed analytically, and the condition that it shall be a minimum is the set of simultaneous equations

$$\left(\partial S / \partial a\right) = 0$$

$$\left(\partial S / \partial b\right) = 0 \quad \text{etc.,}$$

where the observations x_i, y_i are now treated as known numerical constants, and the parameters a, b ... as the variables of differentiation. This set of equations, often called the 'normal equations', can then, at least in principle, be solved to give estimates of the parameters a, b ... .

The normal equations are simplest when the function $f(x; a, b)$ is a linear function of x, and this is the only case that will be considered.

3.2 Least Squares Fitting of Straight Line

It is assumed that the observations of x_i are free from error, and that the observations y_i are subject only to random error.

Straight Line with Two Parameter

The general straight line has two parameters. These can be either

(a)   the two values of y corresponding to two given values of x (defining two points through which the line passes) or

(b)   one value of y corresponding to one given value of x (defining one point through which the line passes) and the gradient of the line.

In practice (b) is normally used, giving $\qquad y - Y = b\left(x - X\right)$

$$\text{or} \qquad y = Y + b\left(x - X\right)$$

where Y is the (as yet unknown) y co-ordinate at $x = X$, and b is the (as yet unknown) gradient.

Two common choices of X are

(i) $X = 0$, giving $y = a + bx$, where a is the intercept on the y axis and

(ii) $X = \bar{x}$, the arithmetic mean of the observed x_i, giving $y = c + b\left(x - \bar{x}\right)$

where c is the y corresponding to $x = \bar{x}$.

Taking choice (i) for example, the parameters a and b are to be chosen so as to minimise S, where

$$S = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \, ,$$

that is to satisfy

$$\partial S/\partial a = -2\sum (y_i - a - bx_i) = 0$$

and

$$\partial S/\partial b = 2\sum x_i(y_i - a - bx_i) = 0 \, ,$$

or

$$a.n + b\sum x_i = \sum y_i$$

and

$$a\sum x_i + b\sum x_i^2 = \sum x_i y_i \, ,$$

giving

$$a_{est} = \frac{\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - (\sum x)^2} = \frac{\overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - (\overline{x})^2}$$

and

$$b_{est} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x^2)} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - (\overline{x})^2} = \frac{\sum y(x - \overline{x})}{\sum(x - \overline{x})^2}$$

as the best estimates of the a and b in the equation $y = a + bx$. (For simplicity the subscript i has been omitted; the alternative expressions in each line are algebraic identities.)

Similarly, for choice (ii) we would obtain

$$b_{est} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - (\overline{x})^2} = \frac{\sum y(x - \overline{x})}{\sum(x - \overline{x})^2} \quad \text{(as before)}$$

and

$$c_{est} = \frac{\sum y}{n} = \overline{y}$$

as the best estimates of the parameters b and c in the equation $(y - c) = b(x - \overline{x})$. We see that the least squares' straight line passes through the 'centre of mass' $(\overline{x}, \overline{y})$ of the observations, and we may write its equation as $y - \overline{y} = b(x - \overline{x})$.

Straight Line with One Parameter

There may be theoretical reasons which determine one of the parameters, leaving only one to be found experimentally. For example it might be demanded that the line pass through the point (X, Y), so that $(y - Y) = d(x - X)$, where X and Y are fixed constants. In this case the only parameter to be determined experimentally is the slope d, and there is only one normal equation; the estimate of d turns out to be

$$d_{est} = \frac{\sum(x - X)(y - Y)}{\sum(x - X)^2}$$

If the point (X, Y) is the origin $(X = 0, Y = 0)$ then this reduces to $d_{est} = \dfrac{\sum xy}{\sum x^2} = \dfrac{\overline{xy}}{\overline{x^2}}$.

24

Uncertainty of the Estimates of a, b, c and d.

The formulae derived above give only estimates of the 'true' values of a, b, c and d; we now derive an estimate of their uncertainty.

In general the least squares' method gives formulae in which the estimates of the parameters are expressed as functions (often complicated) of the observations x_i (which can be considered simply as numerical constant as they are error-free) and of the observations y_i (which are subject to error). We have therefore, for example, $a_{est} = f(y_1, y_2 .... y_n)$, and can express the uncertainty of a in terms of the uncertainty of the y_i by

$$\sigma^2_{a\,est} = \left(\frac{\partial\, a_{est}}{\partial\, y_1}\right)^2 \sigma^2_{y_1} + \left(\frac{\partial\, a_{est}}{\partial\, y_2}\right)^2 \sigma^2_{y_2} + ... + \left(\frac{\partial\, a_{est}}{\partial\, y_n}\right)^2 \sigma^2_{y_n}$$

where σ_{y_i} is the standard deviation we would obtain for repeated observation of y_i at a fixed x_i.

The straight line is a particularly simple case because the least squares' criterion gives expressions for the parameters which involve only linear combinations of the y_i, and each of the partial derivatives is a function of the x_i only.

There is a further simplification if all the σ_{yi} are the same, σ_y; that is if the uncertainty in measuring y does not depend on x. We then find

$$\sigma^2_{a\,est} = \frac{\sum x^2}{n\sum x^2 - (\sum x)^2}\sigma^2_y = \frac{\overline{x^2}}{n\left(\overline{x^2} - \overline{x}^2\right)}\sigma^2_y = \frac{\sum x^2 / n}{\sum(x - \overline{x})^2}\sigma^2_y$$

$$\sigma^2_{best} = \frac{n}{n\sum x^2 - (\sum x)^2}\sigma_y^2 = \frac{1}{n\left(\overline{x^2} - \overline{x}^2\right)}\sigma_y^2 = \frac{1}{\sum(x - \overline{x})^2}\sigma_y^2$$

$$\sigma^2_{cest} = \frac{1}{n}\sigma_y^2$$

$$\sigma^2_{dest} = \frac{1}{\sum(x - X)^2}\sigma_y^2$$

for $\qquad y = a + bx, \quad (y - c) = b(x - \overline{x}) \text{ or } (y - Y) = d(x - X).$

These standard deviations can then be taken as a measure of the uncertainty of our estimates of the parameters. If the y_i are themselves Gaussian (or nearly so) then these estimates (which are linear combinations of the y_i) are also Gaussian (or more nearly so) about the true values, and the Gaussian tables of probability can be used. Even if the observations are not Gaussian, the estimates are essentially averages of the observations, and the estimates will be more nearly Gaussian than were the y_i, so that in practice the tables of probability can usually be used without much error.

It is obvious from the expressions above that the standard deviations of our estimates increases as σ_y increases, as is to be expected. It is instructive to write the expressions in a slightly different form by putting $\sum(x - x)^2 \equiv n\mu_x^2$, where μ_x, a "standard deviation of the x's", is a (root mean square) measure of the range of x over which observations are available. We then have

$$\sigma_{a_{est}}^2 = \frac{\overline{x^2}}{\mu_x^2} \; \frac{\sigma_y^2}{n} \, ,$$

$$\sigma_{b_{est}}^2 = \frac{1}{\mu_x^2} \; \frac{\sigma_y^2}{n} \, ,$$

$$\sigma_{c_{est}}^2 = \frac{\sigma_y^2}{n} \, ,$$

$$\sigma_{d_{est}}^2 = \frac{1}{\mu_x^2 + (\overline{x} - X)^2} \; \frac{\sigma_y^2}{n} \, .$$

It is then obvious that, for a given distribution of observation ($\overline{x}^2$ and μ_x^2 constant), the standard deviations decrease as $1/\sqrt{n}$, as is also to be expected. The effect of the distribution of the x_i can now also be seen. The smaller the range of observation (the smaller μ_x) the larger are the standard deviations of the slopes b and d and intercept a; in addition the uncertainty of the intercept a increases as $\overline{x^2}$ increases, that is as the region of observation moves away from the y axis, and the uncertainty of the slope d decreases as $(\overline{x} - X)^2$ increases, that is as the region of observation moves away from the fixed point (X, Y).

So far we have assumed that σ_y, the standard deviation of repeated y observations, is known. If there are in fact repeated observations at one or more of the x_i, then σ_y can be estimated directly from these repetitions. Usually however this is not the case, and we have to estimate σ_y from the deviations of the y_i from the least squares' line. In this case it can be shown that the best estimate of σ_y is given by:

For the two parameter line $\qquad (\sigma_y^2)_{est} = \dfrac{S}{n-2} = \dfrac{\sum d_i^2}{n-2}$

where $d_i = y_i - a - bx_i$ or $d_i = c - b(x_i - x)$, and $S = \sum d_i^2$ is in fact the quantity we have minimised. (Two pieces of information (2 degrees of freedom) have been used to determine the parameters a and b, or b and c.)

For the one parameter line $\quad (\sigma_y^2)_{est} = \dfrac{S}{n-1} = \dfrac{\sum d_i^2}{n-1} \quad$ where $\quad d_i = y_i - Y - d(x_i - x)$.

(Only one degree of freedom has been lost.)

Use of the Least Squares' Line for Prediction

Having obtained an estimate of the best line (in the form $y = a + bx$ or $y = \overline{y} + b(x - \overline{x})$ or $y = Y + d(x - X)$) we may want to use it to predict the value of y corresponding to some particular x. This prediction will be subject to uncertainty; the uncertainty will be less if this x lies in the region of the observations, and will rapidly become larger as x moves away from this region.

In general there will be a contribution to the uncertainty of the predicted y from the uncertainty of each parameter (a and b, or $c (= \overline{y})$ and b). We can use our expression for the combination of errors, *provided* that the errors of a and b, or b and c, are *independent*. All three estimates are in fact functions of the y_i, and it turns out that while b and

$c\left(=\bar{y}\right)$ are statistically independent, b and a are not. Therefore we can apply our expression for the combination of errors to the form $y = \bar{y} + b\left(x - \bar{x}\right)$, but *not* to the form $y = a + bx$, unless by chance $\bar{x} = 0$. (Mathematical textbooks on statistics almost always use the form $\left(y - \bar{y}\right) = b\left(x - \bar{x}\right)$ for this reason; unfortunately physics textbooks often use the form $y = a + bx$ without mentioning its disadvantage.)

Weighting

If the points $\left(x_i, y_i\right)$ which are to be fitted are believed to be of unequal accuracy, then weights w_i can be assigned, and the sum which is to be minimised is, for example, $\tilde{S} = \sum w_i\left(y_i - a - bx_i\right)^2$.

This leads to expressions for a, b, c and d which involve the w_i, as do the expressions for their standard deviations.

Such complication is very rarely justified, one case in which it may be necessary is when the y_i which are to be fitted have widely different uncertainties because they are derived by a non-linear algebraic transformation (for example the taking of logarithms) from the experimental observations.

3.3 Other Uses of Principle of Least Squares

The extension to the case where y is a function of several variables x, that is $y = f\left(x_1, x_2...; a, b...\right)$, is straightforward, and again the algebra is particularly simple if the function is linear in the x. The polynomial $y = \sum a_r x^r$, which can be thought of as a special case of this with $x_1 = x$, $x_2 = x^2$ etc., also gives reasonably simple algebra. Non-linear forms can be handled, though only at the expense of much more complicated algebra, and it is usually well worth while changing the variables to give a linear form. If reasonable first approximations to the parameters are already available a linear approximation to the function can be used to estimate corrections to be applied to the first approximations, and iteration continued until sufficient accuracy is obtained.

The principle of least squares can be applied to any situation where the number of equations involving experimental observations is greater than the number of unknown constant. Such situations occur, for example, in the reduction of a topographical survey based on triangulation (where the triangles are often overdetermined), and the selection of the best values of the fundamental constants c, e, m, h etc., using the results of experiments measuring different functions of them.

3.4 Justification of the Use of Method of Least Squares

In this section an attempt is made to explain why the Principle of Least Squares has such a wide application. In order to do this we have to introduce concepts which are not needed elsewhere in these notes; this section can be omitted at a first reading, if desired.

The basic problem is, given the functional relationship $y = f\left(x; a, b\ ...\right)$, the observations $x_1, x_2 ... x_n$ which are not subject to error, and the observations $y_1, y_2...y_n$ which are subject to error, to make the best use of these observations to obtain estimates of the parameters a, b ... . An estimate will be some function of the y_i (and also of the x_i which are essentially just numerical constants), and we wish to choose that estimate (or set of estimates) which, in some sense not as yet defined, is 'best'.

There are several possible criteria for the choice of the meaning of 'best', and the subject is controversial. Depending on which criterion we adopt we will obtain different methods for arriving at our estimates. One common, but not essential, criterion is that the estimate shall be *unbiased*, that is that the arithmetic mean of the estimates obtained from many repeated sets of observations should be equal to the 'true' value of the parameter. (For example, we have seen that the standard deviation of a set of observations about their mean is *not* an unbiased estimate of the standard deviation of the 'population' of observations about the 'true' value; in this particular case the amount of bias is known and can easily be allowed for.) Another, independent, criterion is that our estimate should be '*efficient*', meaning that the estimates obtained from repeated sets of observations should have a small r.m.s deviation (standard deviation) from the 'true' value. (It can be shown that there is a minimum value of this r.m.s deviation which no estimate can better.) A third criterion is that of *maximum likelihood*. The likelihood (or probability) that a particular set of observations will be observed is a function both of the frequency distribution of the errors and of the values of the parameters. On the Principle of Maximum Likelihood that set of values of the parameters is chosen which makes the likelihood of the actual set of observations occurring a maximum; obviously this principle can be applied only if the distribution function of the errors is known.

If the error distribution is Gaussian then the Principle of Maximum likelihood becomes the principle of Least Squares, and produces estimates which, in a very wide sense, are the best that can be obtained. In particular, if y is a linear function of the parameters then the Least Squares estimates are unbiased and have the smallest possible standard deviation.

In practice many actual error distributions are sufficiently nearly Gaussian that the difference does not matter. However if the error distribution is definitely not Gaussian then the Principle of Maximum Likelihood could lead to the minimisation of some other function of the deviations. The method of Least Squares still has some advantages however. In particular, if y is a linear function of the parameters, then the Least Squares' estimate is again unbiased, and has a smaller standard deviation than any other unbiased estimate which is also a linear function of the observations y_i. (This is why the arithmetic mean $\bar{y}$ is almost always used as the best estimate of the 'true' value y_o.)

Thus, in the vast majority of cases, the method of Least Squares does in fact give estimates which, if not perhaps the 'best' in some sense, are not appreciably inferior to those given by other possible methods. In addition there is the strong argument that, if the situation does justify the complication of an objective estimate, the method of Least Squares does give much simpler arithmetic (particularly for linear functions) than most other optimisation procedures.

4 HANDLING OF RESULTS

The previous sections have outlined the theory of random errors. In this section suggestions are made as to how this theory can be applied to experimental results and how the resulting arithmetic can be simplified. It also discusses the problem of how rigorous a procedure is justified by the data.

4.1 General

Before embarking on any statistical treatment of results make sure as far as possible that the results are reasonable and justify such treatment. A little time spent in devising a suitable method of tabulation is often worthwhile; if the values which are meant to be constant or vary smoothly are brought together in an orderly fashion a visual inspection of the rows and columns will often pick out any discontinuities. For dependent variables which have been obtained at constant increments of the independent variable a table of first differences will show up any marked irregularities. Of course in many cases a suitable graph is the best way of detecting any peculiar features - there is no point in attempting to fit a set of data by a straight line when the graph is obviously curved, or composed of two intersecting straight lines.

Such preliminary investigation of the data often reveals unsuspected systematic errors. For example a supposedly constant value may be found to change slowly with time (contamination?) or to vary with the time of day (temperature dependence?). As far as possible, sets of observations taken separately, or under different conditions, should be handled separately, and combined to form grand means only when the sub-means have proved to be not significantly different.

The nature of the calculation to be performed should also be looked at. Is it the *arithmetic* mean which is wanted? Is it the standard deviation of the observations themselves or of their mean which has to be inserted? In finding the standard deviation of a complex expression, are the different terms statistically independent? If not, does it matter? How much computation is justified by this particular set of data? (This is discussed later.)

To avoid possible loss of accuracy in involved calculations keep one or two (but not more) figures than are significant. (Fixed point programs give extra non-significant figures on multiplication, and can lose significant figures on division.) Use a method appropriate to the accuracy, for example mental arithmetic or calculator. Approximations can be just as valid in arithmetic as in algebra; where some quantities in an expression are small compared with others a binomial expansion keeping only first order terms can often save much time. Such an approach is particularly useful in applying small corrections.

It is usually convenient to perform the calculation of "errors" step by step with the main calculation, so that the loss of significant figures, or the retention of too many figures, is quickly noticed.

Similarly the specification of units should proceed step by step throughout the calculation. This will show up physical absurdities, and should show up any omission to convert all observations to a common system of units.

If it is found that accuracy is being lost because of the subtraction of two nearly equal numbers, check to see whether this is purely accidental, or is a consequence of the method of experimentation or calculation.

In a complicated calculation it is often worthwhile trying different sequences and layout of results; a good layout of data and intermediate results, and a logical sequence of calculation can reduce labour and decrease the chance of

mistakes. It may be worth dividing a calculation into smaller stages, particularly if there is some physical significance of the intermediate quantity.

Take as much care with the calculation as with the experiment, and check every calculation (including 'error' calculations), no matter how simple, preferably by some independent method. (For example with tabular data use row and column partial sums; in equation-solving insert the roots back into the equation to make sure that they *are* a solution.)

Make sure that the answers are reasonable; does the mean lie somewhere in the middle of the observed range? Is the standard deviation consistent with the spread of observations? Use an order of magnitude calculation to check the power of ten. Check that the result is physically plausible, a knowledge of the order of magnitude of physical quantities is particularly useful here; if the units are unfamiliar convert to some in which the order of magnitude of the quantity is known. (A repetition of a calculation using another set of units is often a good way of checking the arithmetic.)

4.2   Rejection of Observations

At some stage in the handling of results it may become apparent that one or two observations lie outside the general grouping, and we have the problem of what to do with them.

Most frequency distributions predict a finite, if very small, proportion of observations with very large deviations from the mean. If n is very large we must expect a few large deviations, and they will not unduly affect the mean. However, even when n is small there are sometimes one or two observations with large deviations, which appreciably alter the mean. They may be due to large random errors occurring by chance, or they may be due to a mistake for example an accidental error in reading a scale, setting crosswires on the wrong line etc. If the latter is the case we obviously do not want to include such a value, and if a very large random error occurred by chance it may affect the mean so much that a better estimate would be obtained if it were rejected. There is considerable controversy about rejection criteria. One suggestion is that an observation for which $|x - \bar{x}| \geq 3.5\mu$ (a probability of about 0.001) should be rejected, and that for one which $|x - x| \geq 2.5\mu$ ($P \cong 0.02$) should be examined, and rejected if the circumstances appear sufficient; the μ to use is the standard deviation of all the observations (*before* rejection) about the mean, and *not* the standard deviation of the mean.

Observations should *never* be rejected just because they disagree with some preconceived idea of what the result should be, or because they give a spread of observations larger than desired. Nor should observations be rejected at the experimental stage unless a mistake has obviously been made. *All* observations should be included in the report, and if any are rejected and not used in the final result this should be stated.

4.3   Computation of Mean and Standard deviation

(Some of the following tips are not relevant when using a calculator/computer, but who knows when you might be without access to either!)

The mean and standard deviation are defined by $\qquad \bar{x} = \sum x_i / n$,

$$\mu^2 = \sum (x_i - \bar{x})^2 / n .$$

The calculation of μ using this expression can be tedious because $\bar{x}$ is usually an inconvenient number which may have one more significant digit than the observations themselves. It may be simpler to use the result

$$\mu^2 = \sum x_i^2 / n - \bar{x}^2,$$

or $\qquad \mu^2 = \sum x_i^2 / n - \left(\sum x_i\right)^2 / n^2 ,$

which is what is used by calculator/computer programs.

However either or both calculations can often be simplified by changing the origin from $x = 0$ to $x = X$. Then

$$\bar{x} = \Sigma(x_i - X)/n + X$$

$$\mu^2 = \Sigma(x_i - X)^2 / n - (\bar{x} - X)^2 ,$$

and the value of X is chosen to make the subtractions $(x_i - X)$ easy. X, called a "false mean", does not in fact have to be near the true mean $\bar{x}$. In some case it may be convenient to take X appreciably less than $\bar{x}$ so as to keep all the $(x_i - X)$ positive, in other cases it may be convenient to take X as the nearest round figure to $\bar{x}$, with the advantage that $(X - \bar{x})^2$ is then probably negligible.

Even if this device is not used in the calculation of $\bar{x}$ it is usually worthwhile in the calculation of μ, and there is then the added advantage that $\Sigma(x_i - X)/n + X$ can be used as a check on the value of $\bar{x}$.

If there are many observations which appear to be roughly Gaussian then it may be simpler to compute the mean deviation $\Sigma|x_i - \bar{x}|/n$ and use Peters' formulae to convert this to a standard deviation; the device of a false mean cannot be used in this case.

Combination of Observations

The expression

$$\sigma_F^2 = (\partial F/\partial x)^2 \sigma_x^2 + (\partial F/\partial y)^2 \sigma_y^2 + ...$$

leads to simple algebra only if the function $F(x, y, ...)$ is a simple sum or product. With more complicated functions it is usually easier to break down the calculation of F into simpler steps, and to apply this expression step by step. Alternatively, if you already have a program which calculates $F(x, y ...)$ the derivatives can be estimated numerically as

$$\partial F/\partial x \cong [F(x + \delta x, y,...) - F(x, y, ...)]/\delta x .$$

Unfortunately these techniques should, strictly, not be used if the same quantity, x say, appears in more than one step of the calculation, because then the condition of statistical independence no longer holds.

Other Estimation of Standard Deviation

When a experiment involves the taking of a set of readings in similar conditions (for example when only one independent variable is being varied) it is often sufficient to estimate by repetition the standard deviation of a few

typical readings covering the change of interest: provided these typical standard deviations do not differ widely then the standard deviations of the other readings can be estimated, by rough interpolation if necessary, from them.

In some situations it may not be sensible to obtain an estimate of the uncertainty of an observation by making repeated observations, either because a repetition is impossible, or because, to the accuracy to which readings are made, repetitions would give the same reading; in such situations an estimate of the r.m.s. error of measurement must be made by a consideration of the type of measurement involved. An example of the first case would be the observation of the effects of a hydrogen bomb; other, calibration, experiments (on a smaller scale!) would have to be used to obtain an estimate of the accuracy and reproducibility of the instrument. An example of the second could be the use of a quartz crystal oscillator together with a digital counter as a clock; provided no bias is introduced by the 'clock' then the maximum error is $\pm\frac{1}{2}$ count, and it can easily be shown that the r.m.s. error is about $\frac{1}{4}$ count.

4.4 Mean Separation of Successive Readings

Sometimes the quantity Z which is required is observed as the difference of successive readings y_1, y_2 ... y_i .. y_n, for example fringe spacing, pendulum timing. Successive differences are

$$D_1 = (y_2 - y_1), \quad D_2 = (y_3 - y_2)... D_{n-1} = (y_n - y_{n-1})$$

but the mean $\overline{D}$ of these is *not* a good estimate of Z because it uses only the two end points:

$$\overline{D} = \Sigma D_i /(n-1) = (y_n - y_1)/(n-1).$$

(The n observations cover a range $n - 1$.)

Basically the problem is that of finding the gradient of the straight line best fitting the points $(1, y_1)$, $(2, y_2)$ ... (n, y_n), where $x_i = i$ and is completely free from error, though in practice of course the graph would not be drawn. One obvious way of determining a good estimate of the gradient is the method of least squares; however, because of the particularly simple nature of the x_i, another method can be used. This is known as the Gauss method, and gives an estimate of Z which is almost, but not quite, as good as the method of least squares (it gives a slightly larger uncertainty).

In the Gauss method the points are divided into pairs, and an estimate of the gradient is obtained from each pair; thus each observation is used once and once only. If all the observations have the same standard deviation it can be shown that the choice of pairs which makes best use of the observations is to divide the observations into first half and second half and to take the differences between corresponding observations in each half.

$$y_1 \quad y_2 \quad y_3 \quad \cdots \quad y_k \quad y_{k+1} \quad y_{k+2} \quad \cdots \quad y_n$$
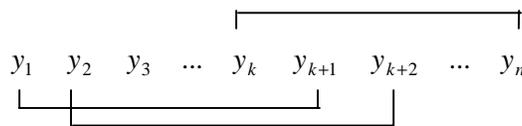
The individual estimates of Z are $D_i = (y_{i+k} - y_i)/k$, giving $\overline{D} = \sum\limits_{i=1}^{m}(y_{i+k} - y_i)\Big/km$ where there are m estimates.

For n even $k = m = n/2$. For n odd the middle observation is omitted giving $k = (n+1)/2$, $m = k - 1$.

This value of $\overline{D}$ is then our estimate of Z, and the derivations of the individual values of D_i from $\overline{D}$ enable us to estimate the standard deviation of the D_i and hence of Z.

Even though successive differences should not be used to estimate Z it may well be worth while to tabulate them as a quick check on the consistency of the results.

4.5   Determination of Parameters

An experiment often involves measuring corresponding values y_i, x_i of the quantities y and x, when there is a theoretical relation $y = f(x, a, b)$ between y and x involving the parameters a, b, and it is desired to find the values of a, b, which best fit the observations.

This is most easily done if the relation is linear, for example $y = a + bx$.

If the relation is not linear it is often worth while to transform to new variables, say u, v, (which are algebraic functions of the observed variables) which are linearly related; we then have $u(x, y) = A(a, b) + B(a, b) v(x, y)$, and the straight line through the (u_i, v_i) gives A and B, from which a and b can be deduced. For example, if $y = a$ exp (bx) we have 1n $(y) = $ 1n $(a) + bx$, and we would plot 1n (y) against 1n (x). (In practice we could use the logarithms to the base 10 and allow for the numerical difference afterwards.)

The determination of the parameters a and b can be done subjectively by eye, using a plotted graph, or objectively by some numerical method such as the Gauss Method, or the method of Least Squares.

Visual Estimate of Straight Line Parameters

If there are a small number of accurate points then a good estimate of the best straight line can be obtained by eye from a plot of the points.

Choose scales and the origin so that the graph has a geometrical slope of about 45° and fills most of the sheet, but do not expand the scales so much as to give a wrong impression of the accuracy of the points. Indicate on the graph the standard deviation of the points, if it is known. If the different points have different standard deviations (for example because they are each the mean of several observations, or the result of a transformation, such as taking logarithms, which changes the relative uncertainties) these must be indicated at each point. If the observational data have too many significant figures to be plotted with sufficient accuracy, then before plotting subtract from the observed y_i accurately known approximations y_i' corresponding to a straight line with convenient approximate constants. For example, when calibrating a scale which may be systematically too short or too long, as well as having errors at the individual graduations, plot (reading-true value) and not the reading itself.

In looking for the best line use a stretched thread, or a line on the bottom face of a transparent ruler, but *not* the edge of an opaque ruler. When taken in moderate-sized groups, about one half of the points in each group should lie on each side of the line. Do not be unduly influenced by the end points, they often indicate the limit of suitability of the apparatus used, in which case they should be given less weight than the other points.

If the standard deviation has been marked at each point remember that, on average, the best straight line should lie outside ±1 s.d. for about one third of the points. If the best straight line is obviously much better or much worse than this the reason should be looked for. A better fit than expected usually means that the random errors of the points have been overestimated (sometimes due to the incorporation of systematic errors). A poorer fit can be due to

an underestimate of the random errors of the points, the presence of systematic errors which are different at different points, or a straight line not being a reasonable approximation to the actual functional relationship.

The intercept which is determined from the fitted line does not need to be the intercept with the $x = 0$ axis. If the observed points are all some distance from this axis then the need for a reasonable plotting scale may mean that this axis is not on the graph paper, and for the reasons outline in section 3.2 it may well be better to take the y value corresponding to an $x = X$ somewhere in the middle of the observations. If the $x = 0$ intercept really is needed it can be found by numerical extrapolation. (Graphical extrapolation using large sheets of graph paper should be done with caution; many cheaper graph papers are surprisingly inaccurate.)

Even if it is only the gradient of the line which is wanted it is usually worth while to determine the intercept to make sure that it is reasonable. A physically unrealistic value may indicate some unsuspected systematic error.

If the origin is an experimental point (it very rarely is) give it the same weight as the other points. Even if the relation is thought to be $y = bx$ it is often safer to assume $y = a + bx$. Then if a turns out to be significantly different from zero this probably indicates some systematic error, which will have to be allowed for; it is surprising how often "end effects" occur. If the origin is not an experimental point it is even more necessary to check the validity of assuming that the *observed* phenomenon is represented by a line through the origin.

An attempt must be made to estimate uncertainties of the parameters a and b. This is usually done by drawing two or more other lines representing some estimate of the limit of lines which could reasonably be said to represent the data. Unfortunately it is difficult to know whether the consequential uncertainties of the parameters represent 1 s.d., 2 s.d. et. uncertainties.

It is surprising how good visual fitting of straight lines can be in suitable cases, and in many experiments nothing more complicated is justified. However it has several drawbacks:

> It is a subjective process, and different people will fit slightly different lines to the same points.

> It is difficult to obtain a reliable estimate of the uncertainties of the parameters a and b.

> It is not suitable if there are a large number of points having low precision.

Thus, in most cases we now prefer to use an objective, numerical, method of obtaining estimates of the parameters and their uncertainties. However such methods can involve a large amount of work, and very often the experimental data do *not* justify this. (Even though a computer program may be used to do the algebra, the numbers have to be entered.)

Numerical Estimate of Straight Line Parameters: Gauss Method

For observations of equal weight, and particularly for observations spaced equally in x, this is a simple method giving estimates not much poorer than those given by the, more tedious, method of Least Squares.

It uses the Least Squares result that the best line passes through the centre of gravity $(\bar{x}, \bar{y})$, and estimates the gradient by the Gauss method described in section 4.4. (The points are divided near the centre, and individual estimates of slope determined by taking corresponding pairs of points. The standard deviation of the mean slope is then determined from the spread of the individual slopes about the mean slope.) To determine the standard deviation of $\bar{y}$ the deviations of the points from the lines must be found.

The Gauss method is most useful when the points have approximately equal precision and are equally spaced in the x direction; only the differences $(y_{i+k} - y_i)$ need then be calculated.

If the points are not equally spaced the slopes $(y_{i+k} - y_i)/(x_{i+k} - x_i)$ must be determined for each pair of points, and if the $(x_{i+k} - x_i)$ are appreciably different a weighted mean should be used. Unequal weights must also be used, both in the determination of slopes and of the centre of gravity, if the points have different precision. In this case it is probably simpler to use a least square' method.

Numerical Estimate of Straight Line Parameters; Method of Least Squares

In a laboratory experiment with comparatively few observations it is rare for the rigour, and labour, of a Least Squares estimation to be justified. If it *is* justified there are several ways in which the arithmetic can be simplified.

It is usually most convenient to use the fact that the best straight line goes through the "centre of gravity" $(\bar{x}, \bar{y})$. This involves the computation of $\sum x$ and $\sum y$. To find the slope b (and the intercept a if required) the sums $\sum xy$ and $\sum^2$ must also be computed.

There are two ways in which the magnitudes of the quantities to be summed can be reduced. The first is a change of origin (in both x and y) to some convenient value, usually just below, or in the middle, of the points. The second is to subtract an accurately know approximation using convenient values for the parameters. (This second step of course includes the first. The two steps are the same as are used to produce a convenient plot for visual estimation.)

Hand computation is obviously simplified if the x_i are simple values. If the x_i differ from convenient values only by amounts which are small compared with their spacing, then linear interpolation can be used to obtain the value $(y_i + \delta y_i)$ of y corresponding to the nearby convenient value $(x_i + \delta y_i)$ of x. Provided δx_i is small this will not introduce any significant error, and the computation is then performed using the more convenient $(x_i + \delta x_i, y_i + \delta y_i)$. (Although Least Squares can be applied to the original data, programs such as Fourier analysis do demand equally spaced points.)

Whether any of these devices is worth using depends on personal preference and on the computing facilities available. However, even when using a calculator/computer, it is usually simpler, and less liable to error, to work with only a few significant figures. Using a false origin can significantly reduce the number of figures to be entered, and hence the possibility of error.

The calculation of the standard deviations requires an estimate of σ_y^2, usually obtained from $\sigma_y^2 \cong S/(n-2)$, where $S = \sum(y_i - a - bx_i)^2$ is the sum which has been minimised by the choice of a and b. The computation of S

using its basic definition is often the most tedious part of the Least Squares fitting, but it has the advantage that $\Sigma(y_i - a - bx_i) = 0$ is a very good check that the values of a and b are correct.

Alternatively S can be calculated from the algebraically equivalent expression

$$S = \left[\Sigma y^2 - (\Sigma y)^2/n\right] - \frac{(\Sigma xy - \Sigma x \Sigma y/n)^2}{\Sigma x^2 - (\Sigma x)^2/n}$$

which needs only one more summation, Σy^2 ; this is what calculator/computer programs do. (If this method is used manually it is essential that the correctness of the parameters be checked either graphically or numerically).

Do *not* use a calculator/computer least squares' program until you are *sure* you know what it is doing; entering numbers and pushing a button does *not* ensure that the result will be meaningful! For example, many calculators having a least-squares' program for the arithmetic mean have both "σ_n" and "σ_{n-1}" buttons. The latter will give you the correct estimate of σ, using $(\text{sum of squared deviations})/(n-1)$, i.e. allowing for the loss of one degree of freedom, but the former will give the root-mean-square scatter of the points about the mean, what was called μ above. And few (if any) calculators apply the extra $1/\sqrt{n}$ factor needed to go from the (estimated) standard deviation of the observations to the (estimated) standard deviation of their mean.

Whatever *you* might have called your variables, programs will assume that the variable *it* calls x is the *independent* variable, having zero (or trivial) errors, and that y is the *dependent* variable having values subject to error.

5. EXPERIMENTAL PROCEDURE

This section is an attempt to indicate some (and by no means all!) of the ways in which proper planning and execution of an experiment can reduce the effect of systematic and random errors. It emphasises that much more is involved than simply connecting apparatus and taking readings - these aspects can be likened to the visible one tenth of an iceberg, with the more important, but invisible nine tenths being the planning, recording, and scrutiny!

Of course not all measurements of every quantity demand the formal carrying out of the full investigation implied here. An *experienced* experimenter designing a not very different experiment may feel justified in taking some short cuts, but only too often the experience has been obtained the hard way; every practising physicist has his own horror stories of time wasted, accuracy lost, or results spoilt because he ignored some aspect of experiment planning or execution.

The following division into 'before', 'during', and 'after' the 'experiment' is arbitrary, and is just a convenient way of giving a rough time sequence. 'Experiment' is used here in the very restricted sense of the more routine use of the apparatus to obtain 'results'.

5.1 Before the Experiment

One of the most important aspects of experimentation is the planning of an experiment so as to reduce systematic and random errors, and to enable a reasonable estimate of the remaining errors to be made. This planning includes not only the choice of technique, the scale of the experiment, and the choice of apparatus, but also the layout of apparatus, and the order in which observations will be made.

If it is possible, a direct comparison of the observed quantity with a standard whose magnitude can be varied is probably always preferable to an 'absolute' measurement of the quantity, and will much reduce the possibility of systematic error. (Compare the measurement of resistance by a Wheatstone bridge with the measurement of the ratio potential difference/current.) If only fixed standards are available, the measurement to a reasonable accuracy of a small difference, or a ratio near unity, between unknown and standard can give a much higher accuracy for the full value of the unknown.

The size of the apparatus may effect the accuracy of measurement, and the consideration of the factors influencing the optimum size often indicates possible sources of error. For example, in calorimetry a larger container (with less surface area/unit volume) will be affected relatively less by heat losses than a smaller one, but more energy input will be needed for a given temperature rise, it may take longer to reach equilibrium, and a larger sample is needed.

Although individual pieces of apparatus may work well in isolation, there may be difficulties when used in conjunction. For example, a voltage amplifier may have its expected gain when observed by a (high input impedance) electronic voltmeter, but have a much lower gain when feeding a (low impedance) moving-coil voltmeter; many measuring instruments have the equivalent of input and output impedance.

The layout of apparatus is important not only in avoiding interaction between different parts of it (for example the effects of the heat or magnetic field of a rheostat) but also in making manipulation of the apparatus and reading of scales safer (for you *and* the apparatus) and more convenient; this not only saves time but also leads to fewer reading errors.

The order in which observations are performed can reduce (or increase!) the effect of systematic errors, particularly those due to slow drifts of instrument calibration. A properly chosen sequence of observations can not only minimise the effect of such a drift, but also point out its presence.

The Importance of a Laboratory Notebook

It is obvious that most experimentation is useless unless a written record is kept of the numerical results obtained. It is less obvious that it is just as necessary to record many other details of the experiment, and that this record should begin long before numerical results are obtained. In fact probably the first thing to be done when given an experimental project is to obtain a notebook.

At the beginning this book will be used to record the discussion of, and decisions about, the factors influencing the choice of technique and apparatus. Then, as apparatus is collected and tested it should become more like a diary, or log book, with a record of the trials and tribulations, of the modifications made, and their effects. There should be sufficient description of the apparatus so that it could be reassembled if necessary. As trials give way to productive experiment there will be a transition to more orderly tables of observations, but throughout *all* stages of the experiment as full a record as possible should be kept of all that pertains to the experiment. Only too often is a disturbing effect found eventually to be due to some completely unsuspected cause - Rutherford discovered radon by observing that the results of an experiment depended on whether the laboratory door was open or shut!

The laboratory notebook should be a *book*, and not scraps of paper. It need only contain *notes*, and not a grammatical, well thought out, account, but it should contain sufficient information to enable the experiment to be written up some months later (when most of the details will have been forgotten).

It is convenient if the notebook is of the type with graph paper on alternate pages. This encourages the rough plotting of results as they are being obtained, which often avoids waste of time by showing that the spacing of readings is too small or too large, and is often a quick indication of something going wrong. The graph paper is also useful for tabulation of results.

Reduction of Systematic Errors

Systematic errors are characteristic of the experimental apparatus (or of bias consistently caused in the performance of the experiment, either by the surroundings or by the observer). They will affect every similar reading to the same extent, and therefore cannot be detected, on their effect reduced, by simple repetition of readings. As far as possible they must therefore be *recognised in advance* by means of the survey discussed above.

Once their presence has been recognised it should be possible to eliminate some systematic errors (or reduce them to a negligible value) by proper choice and adjustment of the apparatus. The effect of others can be eliminated by some cancellation process, or corrected for after their determination by some subsidiary experiment. By proper design of the experiment others (for example errors of scale graduation) can be converted to (pseudo-) random errors.

However it is still necessary to try to detect unpredicted systematic errors by varying the conditions of the experiment and of the apparatus over as wide a range as possible. It is impossible to be certain that all systematic errors have been eliminated, and the only certain test is to measure the required quantity by a completely different method using different apparatus.

Reduction of Random Errors

Decide which of the measured quantities will contribute most error to the final result (remember that if $F = x^a$ appears, $\sigma_F = a\sigma_x$) and reduce the error of this quantity as far as necessary, either by modification of the apparatus or of the experimental method, or by taking a larger number of readings. It is a waste of time to make the measurements of some quantities very accurately if other measurements are necessarily much less accurate; a useful guide is that it is pointless to reduce random errors which are already less then about one fifth of the largest error.

Perform a preliminary experiment (see below) to find to what accuracy the various measurements can in fact be made in the conditions of the experiment. Use it to determine if the apparatus needs, for example, a more rigid mounting (or proper anti-vibration support), better temperature control, or more efficient electromagnetic shielding.

It is often necessary to compromise, and to use the available, imperfect, apparatus, rather than wait 6 months for the ideal piece of equipment. Also, when the time available is limited, you must decide whether to obtain only a few, very accurate, results, or a larger number of less accurate results covering a wider range of the phenomenon under investigation.

Accuracy

Decide beforehand what accuracy is desired of the final result. Sometimes it is sufficient if a result is obtained to 10%, at other times an accuracy of 1 part in 10^6 may be needed. Do not waste time in refining the apparatus to produce a result to $\pm 0.1\%$ if $\pm 1\%$ is sufficient.

Make sure that the various measuring instruments or combination of instruments are sufficiently accurate for your purpose. If you are not certain about a particular instrument, find out; many laboratory instruments are far less accurate than is often thought. Remember that sensitivity (the ability to detect small signal, or a small change in a signal) is not the same as accuracy (the ability to measure with only a small error).

Sign Conventions

To avoid confusion later it is advisable to decide early, and to record unambiguously in the notebook, your choice of the conventional positive sense of every relevant physical quantity. Then stick to it! (Even if it proves to have been an unfortunate choice (for example that the electronic charge is negative) to change it may well lead to difficulty.) Then, on diagrams, always draw the arrows representing vector quantities (for example mechanical force, electric current) in the conventionally positive sense.

It is often also convenient to specify a sign convention or labelling scheme for use in describing apparatus; for example it would be wise to (physically) label differently the two ends of a nominally symmetrical piece of apparatus.

Preliminary experiment

Before using the apparatus for the desired measurements, if at all possible make some trial runs. Besides making you more familiar with the apparatus they should confirm (or otherwise) that the desired accuracy is likely to be

attained, and the range of operation in which results will be useful. They will almost certainly bring out some unsuspected difficulties!

Use these trials also to develop a convenient layout for the tabulation of observations.

5.2   During the Experiment

To avoid careless errors always make each measurement, *however simple*, at least twice. (If the measurement forms part of a closely spaced series it may be an acceptable alternative to plot a graph of the readings as they are taken, so that any mistakes become apparent while there is still the opportunity to correct them by repetition.)

If a subjective setting has to be made (for example setting of cross-wires) always repeat several times and try to avoid bias (for example approach the setting from both sides, do not look at the scale while making the setting).

Record ALL readings, irrespective of whether they are the same, nearly the same, or totally different. What may seem a trivial difference at the time may become important afterwards. Record the readings as *observed*, and perform any arithmetic (for example subtraction of zero reading) *after* recording them.

Record as many details of experimental conditions as possible, whether you think them relevant or not. If, for example, an unsuspected systematic error is discovered later it may then be possible to make appropriate corrections, and avoid the waste of a large amount of experimental time. Ideally a complete specification would include for example the ambient temperature and pressure and air velocity, the electric, magnetic, electromagnetic and gravitational fields, and the relative and absolute positions of the parts of the apparatus! The date and time should ALWAYS be recorded.

Remember that no amount of error theory can compensate for poor and slipshod experimentation.

5.3   After the Experiment

Do not dismantle apparatus until you are sure that the experimental results are adequate.

Before doing any large amount of analysis of the observations look at them critically. With practice, and if they have been well tabulated, it should be fairly easy to pick out possible discontinuities and other abnormalities. The origin of any obvious defects in the data should be investigated *before* carrying out numerical or graphical analysis (which may have been made invalid by the defects), and other queries noted for a closer look later. Remember that statistical analysis is complementary to, and not a substitute for, an intelligent and critical personal assessment of your results.

It is not always necessary to perform a complete statistical analysis for all quantities, in many cases it is sufficient to use an estimated error, particularly in contributions to small correction terms. However do give an indication of the accuracy of all the quantities which are combined to give the final result.

If you weight observations unequally, or reject some observations, give your grounds for doing so.

The determination of a calibration constant is usually subject to random error, but, as far as the main part of the experiment is concerned, any error in the value assigned to this constant is effectively a systematic error, as it is common to all results using that constant. Therefore such calibration constants should be introduced into calculations at as late a stage as possible, so that the effect of other, random, errors can be seen. For example, if the controlled variable is the magnetic field produced by a coil system, it is usually preferable to plot results against coil current (and then afterwards add alongside the current scale a scale in terms of magnetic field) rather than against calculated field values. This approach also reduces the amount of arithmetic and the possibility of arithmetical mistakes.

If an experiment gives several values for the same quantity then the spread of these values can be used to indicate the accuracy of the mean. However, a standard deviation should still be estimated for each of the individual values. Any discrepancy between the estimated and actual spreads should then be investigated.

When quoting the uncertainty of a particular quantity, x say, consider whether it is the absolute standard deviation σ_x or the relative standard deviation σ_x/x (usually expressed as a percentage $100\ \sigma_x/x$), which is more appropriate; it is not necessarily the one which it was convenient to use in calculation! In some cases the decision will depend on the instrument used; some, for example potentiometers, usually have accuracies quoted as a percentage of the actual reading, while others, for example moving coil meters, have accuracies quoted as a percentage of full-scale deflection, that is as an absolute value independent of the actual reading. In other cases the decision will depend on the quantity concerned, and on the context; for example it would be pointless to quote the uncertainty of a temperature in °C, or of the time of occurrence of an event, as a percentage, but might be appropriate to use a percentage figure for a temperature difference or a time interval. Usually it is the absolute value which is relevant.

During calculations keep sufficient digits to avoid loss of accuracy.

Do not quote a result to more significant figures than are justified by the observations. It is usual to omit digits which are uncertain by more than about 10 units; remember that standard deviations are rarely known to better than one significant figure. When rounding off an exact 5 round to the nearest *even* integer in the next decimal - this prevents rounding errors becoming systematic. Use 10^n notation; for example $(18.5 \pm 0.3) \times 10^3$ is preferable to and less ambiguous then 18500 ± 300.

When plotting graphs it is conventional to plot positive quantities to the right and upwards, and to use the horizontal axis for cause and the vertical one for effect (where these can be distinguished). If you do not use these conventions you may make interpretation of the graphs difficult for yourself and others.

In your account specify what standards have been used where this is relevant. Give particulars of make, type, and serial number so that they can be readily identified. The 'traceability' of standards, through a chain which goes back to standards (held by, for example the National Physical Laboratory) based on the fundamental definitions, is increasingly important in science and industry.

Bibliography (as at 2001)

An introduction to error analysis, JR Taylor, University Science Books, OUP, 1982

Statistics in Physical Science, WC Hamilton, Ronald Press, New York, 1964

Probability and Statistics for Engineers and Scientists, RE Walpole and RH Myers (5th Edn.) Macmillan, New York, 1993)