Analysis Of Experimental Data

1. Error Analysis And Accuracy Of Measurement

1.1 Introduction

All physical measurements are subject to various types of errors. It is important to plan any experiment with an accuracy appropriate to its purpose and perform it in such a way that within the limitation of the experiment set up, the errors are reduced to a minimum. It is however, much more important to estimate and quote the error or uncertainty of the measurement, without which, the result of the measurement is worthless, since it is of little value to anybody who wants to make use of this result.

The errors of measurements are of three types: Blunders, Systematic errors and random errors.

Blunders: These occur when an experimenter makes a genuine mistake by reading an instrument wrongly or taking down a reading erroneously. If the experimenter is aware of what the approximate result should be, gross errors of this type can be avoided. It is helpful to plot the results on graph while the measurements are in progress, so as to spot blunders as they happen. In real life, where the true result is not known, nothing much can be done about blunders except to take a lot of readings around the area, where a discrepancy is observed.

Systematic Errors: Errors that are repeated through an entire set of measurements are termed systematic errors. These errors arise because the experimental arrangement often is different from that assumed in theory and the correction factor that takes account of this difference is ignored. For example, the resistance of leads in an electrical experiment and heat losses in a calorimetric experiment are sources of systematic errors. Another common source of systematic error is inaccurate apparatus such as one with wrong calibration or zero offset. Another source of symmetric error is the experimenter's bias, for example parallax error.

There are no clear cut ways to eliminate systematic errors, though in case of a faulty apparatus, it can be checked against a well established standard or in case of the experimenter's bias, it helps to have a second person perform the same experiment and see whether there are systematic errors. For the purpose of this course, systematic errors need not be considered in the estimated errors, though it will be useful for the students to be aware of sources of systematic errors (if any) in the measurements.

Random Errors: Random errors are always present in all experiments and arise due to the combined effect of random fluctuations in the system being measured and the limitations of the measuring instruments themselves. This error cannot be eliminated and must be estimated and quoted as the uncertainty of the final result. The presence of random errors can be seen if the same measurement is repeated several times. In the absence of systematic errors, presence of random errors cause successive readings to spread about the true value of the quantity (Fig. 1a). If in addition, a systematic error is also present, the readings spread not about the true value but about some displaced value (Fig. 1b).

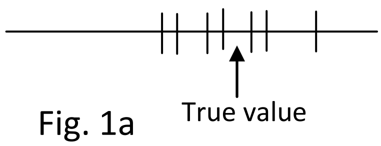

Fig. 1a

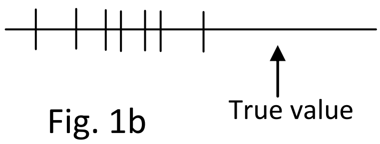

Fig. 1b

1.2 Estimate of Random Error in a Measurement

Let us assume that we are trying to measure the diameter of a ball bearing with an accurate micrometer gauge. No ball bearing is perfectly spherical. So we should take measurements in different directions and a series of values will result. We define the mean value of this series $x_i(i = 1,....,N)$ as

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (1)$$

The values of x_i will be distributed around the mean value $\bar{x}$. The standard deviation for the set x_i is defined to be

$$\sigma = \left[\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right]^{1/2} \qquad (2)$$

It can be shown that 68% of all the data is within the range $\bar{x}-\sigma$ to $\bar{x}+\sigma$ and 90% of the data is between the range $\bar{x}-1.6\sigma$ to $\bar{x}+1.6\sigma$. Thus the standard deviation gives an estimate of the random error or uncertainty in (Δx) the measurement of a certain quantity.

In practice, the uncertainty in the measurement of a particular quantity is often estimated from practical considerations. Suppose we made only two measurements instead of a large number of repeated measurements. Then the uncertainty in x in this case can be simply taken as $\Delta x = |x_1 - x_2|/2$, where x_1 and x_2 are the results of the two measurements. Another situation that may be encountered is when the random error in a measurement happens to be smaller than the least count (LC) of the measuring instrument. In such a case, repeated measurements cannot be used to estimate the random error and the quantity (LC/2) can be taken as an estimate of the uncertainty of the result. Also, in case of a single measurement, (LC)/2 serves as an estimate of the uncertainty.

1.3. Combination of Errors

So far we have only discussed errors in a single measurement. Frequently, we measure several different physical quantities, e.g., x, y, ω etc. And combine them together to obtain the final result for a quantity, say x, will have an uncertainty (Δx) associated with it. How do these uncertainties combine to give the uncertainty Δz of the quantity z in the final result?

(a) Addition/Subtraction: For both $z = x \pm y$

$$\Delta z = \Delta x + \Delta y \tag{3}$$

where $\Delta x, \Delta y$ and Δz are the absolute errors or uncertainties in x, y and z respectively.

(b) Multiplication/Division: For both $z = xy$ and $z = x / y$

$$\left(\frac{\Delta z}{z}\right) = \left(\frac{\Delta x}{x}\right) + \left(\frac{\Delta y}{y}\right) \tag{4}$$

where the quantity $\left(\frac{\Delta x}{x}\right)$ is called the fractional error in x and so on. For the repeated measurements of a certain quantity x, we either directly make an estimate Δx by one of the methods described above and take $\left(\frac{\Delta x}{x}\right)$ as an estimate of $\left(\frac{\Delta x}{x}\right)$. In case of a single measurement $\left(\frac{LC/2}{x}\right)$ can be taken as an estimate of $\left(\frac{\Delta x}{x}\right)$.

(c) For a more general case such as, $z = C \cdot \frac{x^m y^n}{\omega^p}$ (where C is a constant), the fractional error in z is given by

$$\left(\frac{\Delta z}{z}\right) = m\left(\frac{\Delta x}{x}\right) + n\left(\frac{\Delta y}{y}\right) + p\left(\frac{\Delta w}{w}\right) \tag{5}$$

For all the experiments in this course, specific suggestions have been made in this manual regarding the estimation of absolute/fractional error for each measurement. Once the combined fractional error $\Delta z / z$ is estimated using the above procedure, the absolute error or uncertainty Δz of the final result 'z' can be obtained as $\Delta z = \left(\frac{\Delta z}{z}\right)\bar{z}$, where $\bar{z}$ represents the average value of z obtained through repeated measurement of *x, y, w* ...etc.

Often the overall uncertainty in a result is dominated by the uncertainty of the measured quantity which has the most significant error. In such cases, the uncertainties in the other measured quantities may be neglected. This will become obvious when we discuss below the termination of decimal places in a result.

1.4. Termination of Decimal Places in the final Result and the Quoted Absolute Error

The uncertainty (or absolute error) in a result is essentially a probability statement which suggests a range of measured values that are likely to be obtained in case of repeated measurements and hence reflects the precision or accuracy of the measurement. This statement of accuracy is made by stating the final value of the measured quantity and its uncertainty upto the same number of decimal places, which should not be more than are meaningful. In general, this corresponds to the most significant decimal place (digit) in the value of the uncertainty. However, if the most significant decimal place (or digit) in the uncertainty value happens to be 1 or 2, the next significant decimal place (digit) should also be include in the uncertainty value as well as in the value of the measured quantity (see table below).

Let us illustrate this first, through a simple example. Suppose we are measuring the diameter (d) of a ball bearing with a scale having a least count of 1mm and a large no. of repeated measurement lead to measured values between 24-25 mm. This result can be expressed as $d = (24.5 \pm 0.5)$ mm, where the estimate for Δd has been taken as $LC/2 \approx 0.5$ mm.

Notice that an improvement in the precision of measurement which results in a decrease in the uncertainty by an order of magnitude permits the results to be quoted with an additional decimal place. Also note that in the first case, it would be meaningless to quote the result beyond the first place of decimal (eg. 24.53) when Δd is of the order of 0.5 mm. Thus it must be realized that a result quoted as 24.534.96 for either of the above situations is nothing short of absurd!

Let us take another example. In a repeated measurement of 'g' by a certain method, if the average value turns out to be 9.833618 m/s^2, then the presentation of final result will depend on the value of Δg (obtained through the above described procedure) in the manner shown below:

Uncertainty, Δg(m/sec^2)	Final Result
0.000436	(9.8336 ± 0.0004) m/s^2
0.007325	(9.834 ± 0.007) m/s^2
0.012825	(9.83 ± 0.01) m/s^2 or (9.834 ± 0.013) m/s^2
0.381506	(9.8 ± 0.4) m/s^2
1.1358	(10 ± 1) m/s^2 or (9.8 ± 1.1) m/s^2

Thus, usually it is the most significant decimal place in the value of the uncertainty which decides the number of decimal places that can be legitimately quoted in the final result. The final result with appropriately terminated decimal places must always be presented as

(AVERAGE VALUE ± UNCERTAINTY) UNITS

2. LINEAR FIT OF DATA BETWEEN TWO VARIABLES

One often comes across a situation in which the slope or intercept of a straight line representing the relationship between two variables is a quantity of physical interest. The least squares fitting of a straight line is a standard method to obtain the slope and intercept as well their uncertainties.

2.1. Least Squares Fit of a Straight Line

Suppose we have N experimental data points (x_i, y_i, i = 1, ..., N). We would like to plot them and draw a straight line which fits best to this data. We know that the equation of a straight line is $y = mx + c$. Therefore, we have to find out the best values of m and c for which the error is minimum. We define a function

$$\mathrm{P(m, c)} = \sum_{i=1}^{N}(y_i - c - mx_i)^2 \tag{1}$$

We assume that the best fit straight line will minimize P(m, c). We consider m and c as parameters and vary them so that P(m, c) is minimised. We obtained

$$0 = \frac{\partial P}{\partial c} = -2\sum(y_i - c - mx_i) \tag{2}$$

$$0 = \frac{\partial P}{\partial m} = -2\sum x_i(y_i - c - mx_i) \tag{3}$$

Therefore we get the following two equations

$$\sum y_i = cN + m\sum x_i \tag{4}$$

$$\sum x_i y_i = c\sum x_i + m\sum x_i^{\ 2} \tag{5}$$

Multiplying equation (4) by $\sum x_i$ and subtracting it from equation (5) multiplied by N we get

$$m = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{N\sum x_i^2 - (\sum x_i)^2} = \frac{1}{b}\sum(x_i - \bar{x})y_i \tag{6}$$

Where $\bar{x} = \sum x_i / N$ and $D = \sum (x_i - \bar{x})^2$. Substituting m in the equation (4) we get

$$c = \bar{y} - m\bar{x}, \tag{7}$$

where, $\bar{y} = \sum y_i / N$

The estimates of uncertainty in m and c, respectively, are given by (see appendix at the end of the section)

$$(\Delta m)^2 \approx \left[\frac{1}{D} \frac{\sum d_1^2}{(N-2)} \right] \tag{8}$$

$$(\Delta c)^2 \approx \left[\frac{1}{N} + \frac{\bar{x}^2}{D} \right] \frac{\sum d_1^2}{N-2} \tag{9}$$

where, $d_i = y_i - mx_i - c$.

For a straight line passing through origin, the slope m is given by

$$m = \frac{\sum x_i y_i}{\sum x_i^2} \tag{10}$$

$$(\Delta m)^2 \approx \frac{1}{\sum x_i^2} \frac{\sum d_i^2}{(N-1)} \tag{11}$$

where $d_i = y_i - mx_i$.

2.2. Qualitative Best Fit Straight Line

In practice, a straight line graph and its uncertainty may also be determined by drawing the qualitative best fit line, which is drawn such that roughly as many points are above it as below it (Fig. 2a). The slope (S) of such a line is very close to the least squares fit slope. The uncertainty in the slope can be estimated by drawing the limiting lines for the data, as shown below. The limiting lines are drawn by essentially considering the data points near the two extreme ends of the set of data. Also, it is often possible to represent the uncertainty of one or both variables along with the data points (fig. 2b). In such a case, the limiting lines are drawn by enveloping the uncertainties at the two ends of the set of data points. Notice that in general, the two limiting lines may intersect qualitative best fit line at different points. In both the above situations, the slopes S_1 and S_2 of the two limiting lines can be used to estimate the uncertainty in the slope of the qualitative best fit line, which is given by $\Delta S = \frac{1}{2}|S_1 - S_2|$.

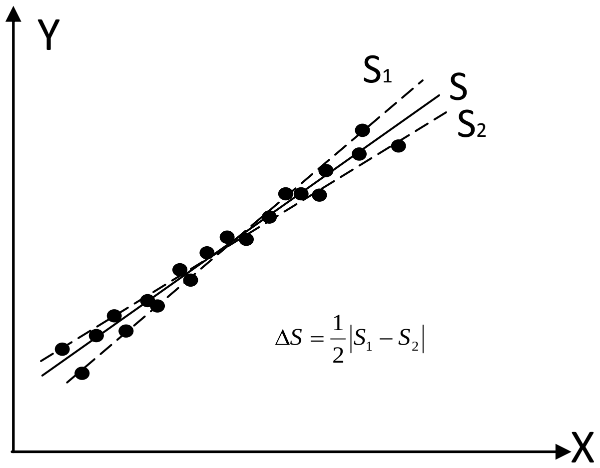

Fig. 2a

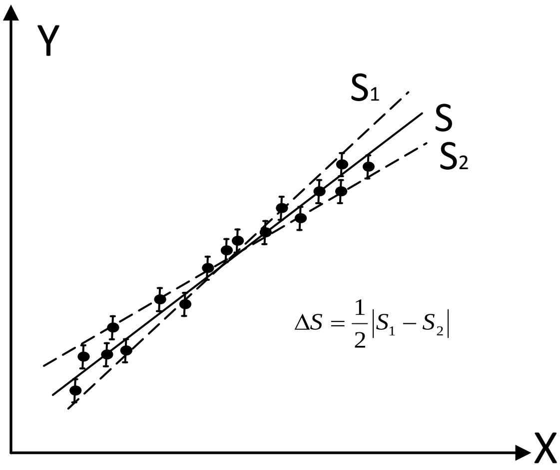

Fig. 2b

3. Concluding Remarks

As mentioned earlier, the uncertainty of a measurement reflects the accuracy or precision with which it is performed and is a vital information to be provided for any meaningful use of a measurement. As there could be several ways to estimate the uncertainty (some of the simple ones are discussed above), the choice is essentially determine by the purpose of the experiment and the constraints under which it is performed. In this manual, for different experiments, specific suggestions of a wide variety have been made with regards to error estimate in all the experiments. The purpose is to familiarize the student with some typical solutions and also make them appreciate the inherent flexibility of approach. It is hoped that with this experience, the students will be able to decide upon a course of action in any given situation.

REFERENCE: G.L. Squires, Practical Physics.

APPENDIX: Uncertainty in '*m*' [derivation of Eqn. (8)]

The possible deviation from the least squares value for '*m*' can be estimated in the following way. For a given pair of observations, the '*m*' value is

$$m_{kl} = \frac{y_k - y_1}{x_k - x_1} \;,$$

So that,

$$(\Delta m)^2 = \frac{1}{N(N-1)} \sum_{k \neq 1} \left[\frac{y_k - y_1}{x_k - x_1} - m \right]^2$$

$$= \frac{1}{N(N-1)} \sum_{k \neq 1} \left[\frac{y_k - mx_k - c - (y_1 - mx_1 - c)}{x_k - x_1} \right]^2$$

Neglecting the cross terms in the numerator (which will oscillate in sign, and hence there is a tendency for cancellation) and using the approximation,

$$(x_k - x_1)^2 = (x_k - \bar{x} - x_1 + \bar{x})^2$$

$$\approx 2 < (x_k - \bar{x})^2 > \qquad \text{[<> represents average]}$$

We get,

$$(\Delta m)^2 \approx \frac{N \sum_k (y_k - mx_k - c)^2}{N(N-1) < (x_k - \bar{x})^2 >}$$

$$\approx \frac{\sum d_i^2}{(N-1)D}$$

This is an approximate way of arriving at the expression for *Δm*. A more rigorous approach (see reference) leads to a factor of (N-2) in the denominator (Eqn. (8)).