Error Analysis in Experiments

'The aim of science is not to open a door of infinite wisdom, but to set a limit to infinite error'- by Galileo in 'The Life of Galileo' written by 'Bertolt Brecht'

Why Error Analysis?

Physics is a quantitative science. This means that in the Physics laboratory we are concerned with making measurements which are both accurate and precise. In order to be able to make a meaningful interpretation of our results we have to have an idea of how reliable those results are. This is where the notion of experimental error comes into the picture. It is an honest expression of the uncertainty of the measurements, not an indication of mistake.

Types of Error

An ***uncertainty*** is a range, estimated by the experimenter, that is likely to contain the true value of whatever is being measured. For example, if you measure a distance with a meter stick you usually assign an uncertainty of ± 1mm to the result. Uncertainties can be expressed in ***absolute*** terms or ***relative*** terms, just as errors can. People often say "error" when they mean uncertainty, just because it doesn't take as long, but what is meant can usually be figured out from the context.

A ***systematic error*** results reproducibly from faulty calibration of equipment or from bias on part of the observer. These errors must be estimated from an analysis of the experimental conditions and techniques. In some cases corrections can be made to the data to compensate for systematic errors where the type and extent of error is known. In other cases, the uncertainties resulting from these errors must be estimated and combined with uncertainties from statistical fluctuations.

Random error is the fluctuations in observations which yield results that differ from experiment to experiment and that requires repeated experimentation to yield precise results. The problem of reducing random errors is essentially one of improving the experiment and refining the techniques as well as simply repeating the experiment. If the random errors result from instrumental uncertainties, they can be reduced by using more reliable and more precise measuring instruments. If the random errors result from statistical fluctuations of counting finite number of events, they can be reduced by counting more events.

Probable error is the magnitude of error which is estimated to have been made in determination of results. This does not mean that we expect our results to be wrong by this amount. It means, instead, that if our answer is wrong, it probably won't be wrong by more than the probable error. The probable error has another significance. If we repeat the experiment, making the measurements in as nearly identical a manner as possible but not necessarily obtaining the identical observations, we expect the new result to have the same probable error as the first. Since we expect both determinations to be approximately within the probable error of the "true" value, they will also probably be within some fraction of the probable error of each other. Thus, the probable error for the result is also a measure of the probable discrepancy between two results obtained under identical conditions.

The ***accuracy*** of a measurement is a way of talking about the total error in your final result. An accurate measurement is very close to the true value. Just because a measurement is accurate doesn't mean it's precise; an accurate value with a wide possible range isn't very useful.

The ***precision*** of a measurement is the total amount of random error present. A very precise measurement has small random errors, but just because a measurement is precise doesn't mean that it's accurate (see above); undiscovered systematic errors might skew your results drastically.

The accuracy of an experiment is generally dependant on how well we can control or compensate for systematic errors. The precision of an experiment is dependant on how well we can overcome or analyze random errors.

Significant Figures and Round off

The precision of an experimental result is implied by the way in which the result is written, though it should generally be quoted specifically as well. To indicate the precision, we write a number with as many digits as are significant. The number of significant figures in a result is defined as follows :

1. The leftmost nonzero digit is the most significant digit.
2. If there is no decimal point, the rightmost nonzero digit is the least significant digit.
3. If there is a decimal point, the rightmost digit is the least significant digit, even if it is a 0.
4. All digits between the least and most significant digits are counted as significant digits.

For example, the following numbers each have four significant digits: 1,234; 123,400; 123.4; 1,001, 1,000., 10.10, 0.0001010, 100.0. If there is no decimal point, there are ambiguities when the rightmost digit is a 0. For example, the number 1,010 is considered to have only three significant digits even though the last digit might be physically significant. To avoid this ambiguity, it is better to supply decimal points or write such numbers in exponent form as an argument in decimal notation times the appropriate power of 10. Thus, our example of 1,010 would be written as 1,010. or 1.010×10^3 if all four digits are significant.

When quoting results of an experiment, the number of significant figures given should be approximately one more than that dictated by the experimental precision. The reason for including the extra digit is that in computation one significant figure is sometimes lost. Errors introduced by insufficient precision in calculations are classified as *illegitimate error*. If an extra digit is specified for all numbers used on the computation, the original precision will be retained to a greater extent. For example, in the experiment if the absolute precision of the result is 10 mm, the third figure is known with an uncertainty of ± 1 and the fourth figure is not really known at all. We would be barely justified in specifying four figures for computation. If the precision is 2 mm, the third digit is known quite well and the fourth figure is known approximately. We are justified in quoting four figures, but probably not justified in quoting five figures since we cannot even have much confidence in the value of the fourth figure.

When insignificant digits are dropped from a number, the last digit retained should be rounded off for the best accuracy. To round off a number to a smaller number of significant digits than are specified originally, truncate the number to the desired number of significant digits and treat the excess digits as a decimal fraction. Then

1. If the fraction is greater than ½ , increment the least significant digit.
2. If the fraction is less than ½, do not increment.
3. If the fraction equals ½, increment the least significant digit only if it is odd.

In this manner, the value of the final result is always within half the least significant digit of the original number. The reason for rule (3) is that in many cases the fraction equals either 0 or ½ and consistently incrementing the least significant digit for a fraction of ½ would lead to a systematic error. For example, 1.235 and 1.245 both become 1.24 when rounded off to three significant figures, but 1.2451 becomes 1.25.

Statistical Error Analysis

A. Gaussian Distribution

As you've no doubt seen in lab, every measurement is subject to a certain amount of random error. The roots of this problem lie deep in quantum mechanics, on Heisenberg's uncertainty principle.

Random errors can arise from minute vibrations in the apparatus, quantum uncertainties in the system being studied, and many other small but uncontrolled effects. Fortunately, almost all the random errors you encounter can be characterized by a Gaussian distribution, also known as a bell curve [Fig. 1]. This simple mathematical form describes the probability of encountering any given error.

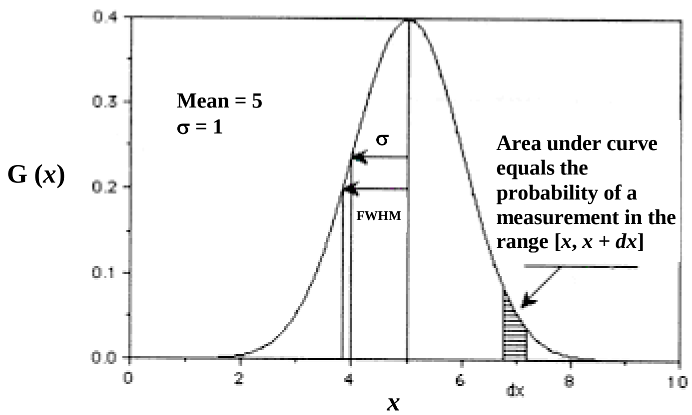


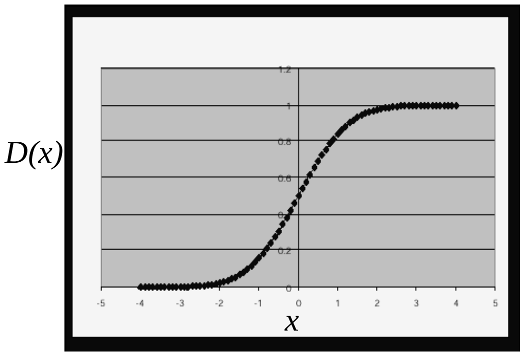
Fig. 1 Gaussian Distribution

The Gaussian distribution has two free parameters: the mean and the standard deviation. The probability of finding a measurement in the range $[x, x+dx]$ is equal to the area under the curve in that range. The curve is normalized to have a total area of 1, which is why its amplitude is not also a free parameter. Notice also that the distribution is symmetric; an error is equally likely to occur in either direction. The equation which describes this curve:

$$G(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[\frac{-(x-x_m)^2}{2\sigma^2}\right]$$

The standard deviation *(σ)* describes the width of the bell; a higher standard deviation means that you're more likely to find large errors. The mean *(x_m)* lies on the axis of symmetry of the bell. These two parameters completely determine the shape of the curve and are used to describe the results of your measurements. Another common way of describing the width of the bell is by using the "full width at half maximum", or FWHM, which is equal to 2.36σ and is easier to figure out from a plot. By integrating all or part of the Gaussian curve, we can make precise statements about how probable it is that our results are correct.

Cumulative Frequency distribution



In biology, for statistical analysis, one more commonly use cumulative distribution function, which gives the probability that a variate assumes a value $\leq x$, and is then the integral of the Gaussian function integrating from minus infinity to x.
Cumulative distribution function is given by

$$D(x)=\int_{-\infty}^{+\infty}P(x)=\frac{1}{2}\left[1+erf\left(\frac{x-x_m}{\sigma\sqrt{2}}\right)\right]$$

erf x is the error function.
The cumulative distribution is basically the answer to the question, "What is the probability that an instantaneous value of variate is less than *x*". Basically, the point of inflection of the cumulative distribution corresponds to maximum probability.

B Mean value

Suppose an experiment were repeated many, say *N*, times to get,

$$x_1, x_2, \ldots\ldots x_i, \ldots., x_n,$$

N measurements of the same quantity, *x*. If the errors were random then the errors in these results would differ in sign and magnitude. So, if the average or mean value of our measurements were calculated,

$$x_m = \frac{x_1 + x_2 + ... + x_n}{N} = \frac{\left[\sum_{i=1}^{n} x_i\right]}{N}.$$

Some of the random variations could be expected to cancel out with others in the sum. This is the best that can be done to deal with random errors: repeat the measurement many times, varying as many "irrelevant" parameters as possible and use the average as the best estimate of the true value of *x*. (It should be pointed out that this estimate for a given *N* will differ from the limit as $N \to \propto$ the true mean value; though, of course, for larger *N* it will be closer to the limit).

Doing this should give a result with less error than any of the individual measurements. But it is obviously expensive, time consuming and tedious. So, eventually one must compromise and decide that the job is done. Nevertheless, repeating the experiment is the only way to gain confidence in and knowledge of its accuracy. In the process an estimate of the deviation of the measurements from the mean value can be obtained.

C Standard deviation

In terms of the mean, the standard deviation of any distribution is,

$$\sigma = \sqrt{\frac{\sum_i (x_i - x_m)^2}{N}}$$

The quantity σ^2 is called the ***variance***. The best estimate of the ***true standard deviation*** is,

$$\sigma_x = \sqrt{\frac{\sum_i (x_i - x_m)^2}{N-1}}$$

Proportional Errors

In this workbook, we will evaluate proportional errors in each experiment. Consider the measurement of a physical quantity *x*. We infer about the value of another quantity *y* which depends on *x* through the relation $y = f(x)$. If Δx is the error (or least count) for the *x* measurements, we would like to know what is the proportional error in *y*, i.e., what is $\frac{\Delta y}{y}$?

Since $y = f(x)$ and *x* is measured with precision Δx, we may assume that the inferred value of *y* is $y + \Delta y$. Thus, we can write,

$$y + \Delta y = f(x + \Delta x) \qquad (1)$$

Expand the RHS in a Taylor series up to linear order in Δx. This gives,

$$y + \Delta y = f(x) + \frac{\partial f}{\partial x}\Delta x \qquad (2)$$

which leads to

$$\Delta y = \frac{\partial f}{\partial x}\Delta x \qquad (3)$$

Note that the partial derivatives are evaluated at *x*.
Dividing both sides by *y*, we obtain

$$\frac{\Delta y}{y} = \frac{1}{f}\frac{\partial f}{\partial x}\Delta x = \left(\frac{\partial}{\partial x}\ln f\right)\Delta x \qquad (4)$$

The proportional error formula would be

$$\frac{\Delta y}{y} = \left|\frac{1}{f}\frac{\partial f}{\partial x}\right|\Delta x = \left|\left(\frac{\partial}{\partial x}\ln f\right)\right|\Delta x \qquad (5)$$

The absolute value is important (and physical) because, without it, one might end up getting a negative quantity, which is meaningless.
Hence, for any given *f(x)*, one can now calculate the proportional error in *y* using the formula above.
One may extend this to a function of several variables. For example, let us take the formula for the volume of a cylinder with inner radius *r* and outer radius *R* and height *h*. This is given as

$$V = \pi(R^2 - r^2)h = \frac{\pi}{4}(D^2 - d^2)h \qquad (6)$$

where $D = 2R$ and $d = 2r$.

In the above formula, *V* is a function of three variables (*R*, *r*, *h*) or (*D*, *d*, *h*). For the assigned work in the lab (first day), it is better to use an error formula for (*D*, *d*, *h*) because we are measuring diameter directly. Let us now derive the formula for the proportional error in *V*.
We have the following from the single variable example,

$$V + \Delta V = f(D + \Delta D, d + \Delta d, h + \Delta h)$$

$$= f(D, d, h) + \frac{\partial f}{\partial D}\Delta D + \frac{\partial f}{\partial d}\Delta d + \frac{\partial f}{\partial h}\Delta h \qquad (7)$$

Therefore,

$$\frac{\Delta V}{V} = \frac{1}{f}\frac{\partial f}{\partial D}\Delta D + \frac{1}{f}\frac{\partial f}{\partial d}\Delta d + \frac{1}{f}\frac{\partial f}{\partial h}\Delta h \qquad (8)$$

The proportional error formula thus becomes

$$\frac{\Delta V}{V} = \left|\frac{1}{f}\frac{\partial f}{\partial D}\right|\Delta D + \left|\frac{1}{f}\frac{\partial f}{\partial d}\right|\Delta d + \left|\frac{1}{f}\frac{\partial f}{\partial h}\right|\Delta h \qquad (9)$$

where we have taken into account the absolute values (positive) for each term, in order to get the maximum proportional error.
Using the functional form for *V (D, d, h)*, it is now a straightforward exercise to show that

$$\frac{\Delta V}{V} = \frac{2D\Delta D + 2d\Delta d}{D^2 - d^2} + \frac{\Delta h}{h} \qquad (10)$$

Since *D* and *d* are measured with the same instrument, we get

$$\frac{\Delta V}{V} = \frac{2\Delta D}{D - d} + \frac{\Delta h}{h} \qquad (11)$$

If we had chosen to work with the radii *R* and *r* (instead of the diameters), then we would get,

$$\frac{\Delta V}{V} = \frac{2\Delta R}{R - r} + \frac{\Delta h}{h} \qquad (12)$$

An important issue is to know when and how we can use the radius-based formula while directly making diameter measurements. If we indeed intend to do so, then we must be careful because, by measuring diameter we will need to infer about the radius from $R = \frac{D}{2}$. This gives $\frac{\Delta R}{R} = \frac{\Delta D}{D}$, but $\Delta R = \frac{\Delta D}{2}$. If ΔD is the least count for measuring diameter, then the effective least count for the radius is $\Delta R = \frac{\Delta D}{2}$. Using this relation in the radius-based formula one can see that it matches with the diameter-based formula, as it should, as long as we measure diameter.

On the other hand, if we are able to measure the radii, then the ΔR must be directly taken as the least count of the measuring apparatus and the radius-based formula may be used to find the error. One will have to do the opposite of what we stated above in the previous paragraph, if we use the diameter-based formula to start with, in the case where radius is being measured directly.

E Removing systematic errors

To hunt for systematic errors one should go through this mental process, while designing an experiment:

1. What physical quantities (including environmental factors) is the measurement most sensitive to?
2. Are there any other sources of error in the quantity that is being measured?
3. If so how we isolate the experiment from these effects?
4. If we can not get rid of the systematic error, can we measure it and account for it later?

Of course, there always remains the possibility that a systematic error is present which we might not think of. To account for this one needs to **calibrate** the instruments used and if possible, the experiment itself. Calibrating means that we use our instrument to measure some known quantities and check whether the measured answer tallies with known results. One should be cautious when using this method to correct results outside the domain which we have calibrated. There is no way to know whether other effects would become important in the new region.

If we do not have any good way of producing known values, and think of a systematic error which we are not able to remove from the experiment, then the only way to correct it is by using Physics. We make an educated guess as to the exact nature of the error, and then use an established theory to figure out what impact it will have on the experiment.

Questions

1. Consider a set of plates produced by the same mechanism, which look squarish but might be rectangular because of minute differences in the lengths of the sides. We would like to measure the sides of these plates and decide about the squareness.

There are two ways to decide about squareness--(i) find the values of (length)-(breadth) (ii) find the values of (length)/(breadth). Which do you think is a better formula to use to accurately determine the `squareness'? Use error analysis to answer this question.

2. Suppose, in an experiment, you are measuring two lengths – one of the order of 1.0 m and another of the order of 0.1 m. Both are being measured by an instrument of least count 0.01 m? Which length is being measured more inaccurately (i.e. with larger error)? (Note that you are using the same instrument to measure both the lengths).

3. Determine the number of significant digits of each of these numbers : (a) .0002053 (b) 1.3456 (c) .010234 (d) 1001.23 (e) 100.00

4. If you are given a formula $y=f(x_1,x_2,x_3)$ say, a way to find the error is to take the logarithm and evaluate $(\delta y)/y$. Remove all minus signs in the resulting expression involving $(\delta x_1)/x_1$ etc., so that you get the maximum error. Try this method out for the following formulae (a) $y= ax^2z^3$ (b) $y= ax + bz^2$ (a and b are constants here).

5. One of your friends measures a quantity five times and writes the data as 100.01, 100.002, 100.0, 100.01, 100.00. He shows it to the instructor and gets a thorough lecture on `writing data properly'. What do you think the instructor said to your friend or where did your careless friend mess up?

Reference

1. An Introduction to Error Analysis by J.R. Taylor, University Science Book, 1962
2. Statistical Treatment of Experimental Data by H.D. Young, McGraw-Hill 1962.
3. Data Reduction and Error Analysis for the Physical Sciences by P.R. Bevington, McGraw-Hill 1969 519.8 BEV/D
4. Physics Part I, Textbook for Class XI, Chapter 2, National Council of Education Research and Training (NCERT), New Delhi (2017)