Measurement and Uncertainty in Scientific Experiments

1. Introduction

1.1 Experimental Precision and Accuracy

There are two kinds of numeric values that are used in science: numeric values that are defined (physical constants) and numeric values that are measured. The essential difference is this: defined values are exact while measured values, or quantities, are always subject to uncertainty or a combination of uncertainties. If you count ten oranges in a fruit bowl or sixteen students in a classroom, then these numbers (ten and sixteen) are exact. But if you measure the weight of the ten oranges then the result would have some inherent uncertainty in it. The uncertainty in a measured value, of course, can never be stated exactly, because if that were the case, there would be no need for accounting for the uncertainty. In real experiments this is not the case, and the uncertainty gives us a range of estimated values.

Experimental uncertainty is introduced primarily because no measuring instrument can be made to measure a quantity with infinite precision.

We cannot make repeated measurements in the same way every time but might expect that the measured value would converge around some average value. **How closely we converge to that average value defines the *precision* of a measurement.**

We must also ask how much the **average value** of a set of multiple measurements deviates from the true value, if true value is known independently. **The deviation from the true value describes the *accuracy* of a measurement.**

Experimental uncertainty, often referred to as *experimental error*, is not to be confused with "human error" or errors introduced in a measurement due to carelessness or erratic measuring devices.

Uncertainty in a measurement is affected by the precision of the measuring scale employed. For example, suppose it is known to be exactly 1:55:30 pm according to a time standard, but you had to tell time using **Clock A** as shown in Figure 1. You could easily tell that the time is a little bit after *1:55 pm* (you know it is *pm* because it is daytime). You could even make a reasonable guess that the time is somewhere between *1:55 pm* and *1:56 pm*. If you were to read the time from another **Clock B**, where the dial is marked off in one-minute intervals (only the relevant region has been shown marked in minute intervals), you could say with greater certainty that the time is closer to *1:55:30*. There is less uncertainty in your read of the time using **Clock B** than using **Clock A**. Thus, **Clock B** provides a measurement with greater *precision* than **Clock A**. Now

consider **Clock C**. **Clock C** reads a time of *1:57.* Although **Clock C** can be read with the same precision as Clock **B,** both **Clock A** and **Clock B** are more *accurate* than **Clock C,** and **Clock C** is more precise than **Clock A**. Of the three, **Clock B** is most precise and accurate.

Thus, accuracy and precision in making measurements are different. ***Accuracy is a measure of how close your measurements are to an accepted value (real value), whereas precision is related to the smallest change that can be detected in the quantity that is being measured, and to how closely multiple measurements of the same quantity converge on an average value.***

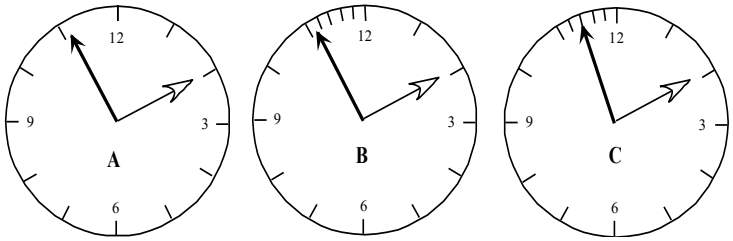
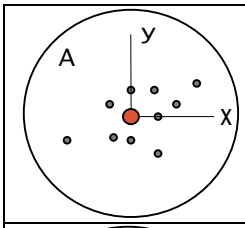
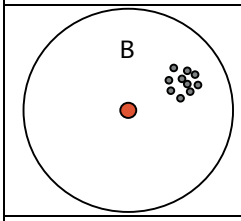
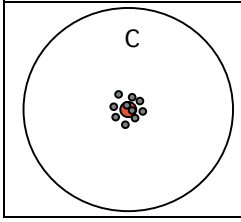


Figure 1

Consider another example to contrast accuracy and precision. Players A, B, and C play a game of darts. The aim, of course, is to hit the bull's eye.

	**Player A** throws the darts ***accurately but not precisely***. The throws are accurate because the average of all the positions measured from the bull's eye will be very close to the bull's eye. Player A does not throw the darts with great precision however because the distribution of the positions of the darts is very broad.
	**Player B** throws the darts ***precisely but not accurately***. The throws are precise because the distribution of throws is very narrow. It is not accurate because the average position of throws is far from the bull's eye.
	**Player C** throws the darts both ***precisely and accurately***. On average, the darts arrive very near the target, so Player **C** has an *accurate* throw. The darts all land very close to one another, so Player **C** has a *precise* throw as well.

1.2 Significant Figures

When you report results of an experiment or analysis, you must recognize that you are not only stating a numeric value but are also conveying how reliable your results are. In considering how precise a measurement is, we write our results with the appropriate number of significant figures. ***Significant figures*** **are the number of digits in a measurement to which the result is known with certainty *plus* one digit that represents a reasonable estimate.** (Note: This reasonable estimate refers to analog scales like a ruler. For digital instruments, like a digital timer, you need not make estimates but must consider which digits are relevant to report.)

For example, if the length of an object is reported as *13.5 cm*, then you are implying that the measurements were made with a ruler marked off in cm (0.5 cm is the estimated digit). Similarly, if the length is reported as *13.45 cm*, it would be proper to conclude that the scale was marked off in 1 *mm = 0. 1cm* gradations. To report the length as *13.45 cm* using only a *cm*-scale would be incorrect. On a digital stopwatch, if a time measurement is quoted as *12.3 s*, the reported value implies that the least count of the digital scale (presumably the smallest digit on the readout) is *0.1 s*.

There are some simple rules to remember when dealing with significant figures.

Rule 1: When there are no trailing zeros in the stated value, all digits are significant. For example, *1.23* has three significant figures and *4567* has four significant figures.

Rule 2 [a]: All zeros between significant figures are significant: *2.034* has four significant figures and *203* has three significant figures.

Rule 2 [b]: Zeros that are to the left of the first non-zero digit are not significant: *0.00123* has only three significant figures. Confusion about the number of significant figures can be avoided by expressing the number in scientific notation (a number between 1 and 10 multiplied by a power of ten). When expressed in this manner all digits are significant. For example, *0.00123* would be written as 1.23×10^{-3} with three significant digits. We write 0.0012 as 1.2×10^{-3} if the measurement only produces a precision for two significant digits. Note that if rather we had 0.00120, i.e. explicitly writing a trailing zero, we would write that as 1.20×10^{3} with three significant digits.

When a number does not contain a decimal point and ends in one or more zeros, the assignment of significant figures can be ambiguous. For example, the number of significant digits in 1760 is not clear. If it is a measured length, then the last digit (0) is not significant. However, if it represents the number of yards in a mile, then all four digits are significant**.** ***This ambiguity can always be avoided by writing the value in scientific notation.***

2. Measurement and Propagation of Error

Correlated errors

Suppose you are interested in determining some quantity Q which is a function of several measured parameters $x_1, x_2, x_3, \ldots x_n$. Obviously, there is uncertainty in the measurement of *each* of the parameters $x_1, x_2, x_3, \ldots x_n$ so we would expect that some combination of those uncertainties would affect the total uncertainty in Q. This is known as ***propagation of error.*** Remember, you never know the uncertainty in your measurements *exactly*, but there are several ways to *estimate* the uncertainty when the result is dependent on several parameters.

If $Q = Q(x_1, x_2, x_3, \ldots x_n)$, then we can write:

$$\Delta Q = \left(\frac{\partial Q}{\partial x_1}\right)\Delta x_1 + \left(\frac{\partial Q}{\partial x_2}\right)\Delta x_2 + \ldots\ldots\left(\frac{\partial Q}{\partial x_n}\right)\Delta x_n \qquad [1]$$

Where these are partial derivatives, i.e., only take the deriviative with respect to one variable at a time treating the other variables as constants.

EXAMPLE

The terms in the expression above have a very simple interpretation. The first term,

$$\Delta Q_1 = \left(\frac{\partial Q}{\partial x_1}\right)\Delta x_1$$

for example, is the uncertainty in Q due to the uncertainty in the measurement of x_1. Other terms have a similar interpretation. The partial derivatives are simply slopes of Q *vs* x curves. For example, Figure 2 (not to scale) shows the first term,

$\Delta Q_1 = \left(\frac{\partial Q}{\partial x_1}\right)\Delta x_1$ in Equation [1].

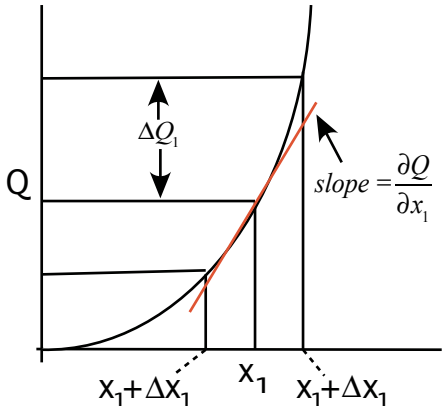


Fig. 2

If all terms, ΔQ_n, in eqn. [1] are taken as positive (absolute value) and their sum is ΔQ, then ΔQ is called the ***maximum absolute uncertainty*** in Q.

$$\Delta Q = |\Delta Q_1| + |\Delta Q_2| + |\Delta Q_3| + ...|\Delta Q_n| \qquad [2]$$

This represents a worst-case scenario giving an upper limit of the propagated uncertainty in the determination of *Q* using the measured inputs $x_1, x_2, x_3, ... x_n$. If the measured value of *Q* deviates from the accepted value of *Q* by more than ΔQ in eqn. [2], then there must be *systematic errors* present in the measurement. Systematic errors refer to reproducible errors due to limitations or poor calibration of equipment, for example, a meter stick which is actually longer than or shorter than a meter.

Uncorrelated errors

In practice, the terms in eqn. [1] are generally not correlated. For example, if x_1 represents a length and x_2 a time measurement then the errors in these two measurements are not correlated to each other. Simply adding them all does not give a realistic picture of what is likely to happen in an experiment. So, we devise another measure of estimating uncertainty. We add the terms in quadrature (it's like considering various terms as components of a vector in a multi-dimensional space) to express the uncertainty in *Q* as:

$$\Delta Q_s = \sqrt{\left[\left(\frac{\partial Q}{\partial x_1}\right)\Delta x_1\right]^2 + \left[\left(\frac{\partial Q}{\partial x_2}\right)\Delta x_2\right]^2 + ...\left[\left(\frac{\partial Q}{\partial x_n}\right)\Delta x_n\right]^2} \qquad [3]$$

ΔQ estimated in this manner is known as the ***standard uncertainty*** in the determination of Q. Sometimes it is also described as root of sum of squares (RSS) uncertainty in the determination of Q.

2.1 Examples

[i] Calculated of a quantity expressed by a power formula:

Suppose you calculate the area of a circle, $A = \pi r^2$ by measuring its radius, *r*.
The maximum uncertainty in the measurement of A is:

$$\Delta A = \frac{\partial A}{\partial r}\Delta r = 2\pi r \Delta r .$$

The **maximum relative uncertainty,** which in this case is equivalent to the **standard relative uncertainty,** is:

$$\frac{\Delta A}{A} = \frac{2\pi r \Delta r}{\pi r^2} = 2\frac{\Delta r}{r}$$

Relative uncertainty can be expressed as a **percent error** by simply multiplying by 100. For example, if your relative error is 0.05, the percent error is 5 percent, because 0.05 x 100 = 5.

[ii] Numeric calculation of uncorrelated uncertainty of a quantity expressed by a power formula:

Kinetic energy (KE), or the energy of a mass due to motion, is given as $KE = ½\ m\ v^2$, where *m* is the mass and *v* is the velocity. Suppose you were to determine the kinetic energy, *KE* of a toy car of mass *204 g* measured with a scale with a precision of *1 g*. Further, suppose that the velocity was determined by measuring the time taken by the toy car to travel between two sensors *50.8 cm* apart, and the time was measured to be *10.1 s*. Thus, the kinetic energy:

$$KE = \frac{1}{2}mv^2 = \frac{1}{2}(204 \times 10^{-3}\text{kg})\frac{50.8 \times 10^{-2}\text{m}}{10.1\text{s}} = 2.58 \times 10^{-4}\text{J}$$

To calculate the uncertainty, express v in terms of its basic variables x and t:

$$KE = \frac{1}{2}mv^2 \sim \frac{1}{2}mx^2t^{-2}$$

Now we can find the **relative maximum uncertainty** in the determination of KE:

$$\frac{\Delta KE}{KE} = \left|\frac{\Delta m}{m}\right| + \left|2\frac{\Delta x}{x}\right| + \left|-2\frac{\Delta t}{t}\right| = \frac{1}{204} + 2\cdot\frac{0.1}{50.8} + 2\cdot\frac{0.1}{10.1} = 0.027 \approx 0.03$$

(Note: uncertainty should always be expressed only to the first significant figure)

The **maximum relative error** in percent is thus 0.03 x 100 = 3%.

Using quadratures, the **relative standard uncertainty** is given by:

$$\frac{\Delta KE}{KE} = \sqrt{\left(\frac{1}{204}\right)^2 + \left(2 \cdot \frac{0.1}{50.8}\right)^2 + \left(-2 \cdot \frac{0.1}{10.1}\right)^2} = 1.95 \times 10^{-2} \approx 0.02$$

The examples above were an illustration of how to estimate uncertainty in a single measurement. In practice you will not be measuring *KE* or any other quantity just once but repeat your measurement many times. We generally do not report uncertainty in a single measurement. Often, but not always, the uncertainty (error) is reported for the average of a series of measurements. A generally accepted way of reporting your experimental results in physics/engineering is to use the following notation:

$$X = X_{best} \pm \Delta X \qquad [7]$$

Here, X is the quantity for which you are reporting a measurement and X_{best} is typically the average of a series of measurement. ΔX is the uncertainty in the value of X. In the case above, $\Delta KE = 0.027 \times 2.58 \times 10^{-4} J = 6.96 \times 10^{-6} J$, and you will report your result as $KE = (2.58 \pm 0.07) \times 10^{-4} J$.

In casual conversation, you might say that your results are good (precise) to within 3%. But properly, you should also specify in your reported results whether your ΔX is the maximum uncertainty or the standard uncertainty.

2.3 Estimation of Statistical Uncertainty in a Mean

There are many ways in which uncertainty can be introduced in measured quantities in an experiment. Such sources of error are broadly of two types - systematic uncertainty and statistical uncertainty. In a simple sense, sources of statistical uncertainty tend to produce a random distribution of data points about the mean when repeated measurements are made of the same quantity. As such, statistical uncertainty affects the precision of measured values, but not the accuracy.

Systematic uncertainty, on the other hand, affects all measured data points in the same way and therefore does not contribute to the precision of a measurement. Think of a meter stick worn out on one end, so that all measurements made with it are slightly low. When you repeatedly measure the length of an object using this meter stick, every reading will be affected in the same way. Systematic uncertainty thus affects the accuracy of a measurement, but not the precision of the measured value. If you need to compare your measured value to a known true value (for instance, perhaps in a measurement of *g*), you will need to consider systematic effects,

particularly if you find that your measurement was **inaccurate** (the true value did not fall within your uncertainty of your measured best value).

To summarize, you will encounter two main types of statistical uncertainty in introductory physics labs - those due to finite instrumental resolution and those due to “noise" or random fluctuations sources that cause repeated measurements of the same quantity to fluctuate around an average value.

The most obvious method to estimate the statistical uncertainty in a measurement is to make several measurements and look at the spread in the measured values. Suppose you make N measurements of the same quantity, X. Even if you do not have a good estimate of the uncertainty in each X_i, it is still possible to determine, statistically, a best value and uncertainty in X. In our data analyses, we will assume the best value is the average or mean of our measured values:

$$X_{best} = \bar{X}$$

As we will see shortly, the average value of a measured quantity by itself may be misleading and not tell us much about the quality or reliability of our measured experimental parameter. A commonly used method that provides insight into the “spread” in the values of the measured parameter is to use the **standard deviation** of the collection of measured values of X. The standard deviation is defined as:

$$\sigma_x = \sqrt{\frac{1}{N}\sum_i^N (X_i - \bar{X})^2} \qquad [8]$$

Here N is the number of measurements in the data set.

There is nothing wrong in using standard deviation as an estimate of statistical uncertainty in your experiments. However, this measure lacks one important feature - it does not get smaller as you take more readings to generate your data set. One would expect the statistical spread (the spread of measured values) to get narrower as you increase the number of readings. This expectation is satisfied by introducing a new measure:

$$\delta_x = \frac{\sigma_x}{\sqrt{N}} = \frac{1}{\sqrt{N}}\sqrt{\frac{1}{N}\sum_i^N (X_i - \bar{X})^2} = \frac{1}{N}\sqrt{\sum_i^N (X_i - \bar{X})^2} \qquad [9]$$

You may use the equation above to express an estimate of the statistical uncertainty in the mean of a collection of data points based only on the distribution of those points. This value is known as the ***standard error***.

In Introductory Physics, you will generally be asked to express your experimental results in one of the following ways:

1. $X = X_{best} \pm \Delta X$, when using the maximum error method.
2. $X = \overline{X} \pm \sigma_X$, when using the standard deviation method.
3. $X = X_{best} \pm \delta_X$, when using the standard error method.

We have the used term ***standard uncertainty*** before when we were looking for the uncertainty in the single measurement of a quantity. Here we are using the term ***standard error*** to characterize a distribution of measured values of a quantity.

A simple example is presented below to [a] illustrate the usefulness of standard deviation/standard error to estimate the statistical uncertainty in a data set, and [b] how to refine/improve an experimental procedure.

A SAMPLE EXPERIMENT: You must measure the diameter of a steel rod of length *10.0cm* using a digital Vernier Caliper. The claimed precision of the caliper is *0.1mm or 0.01cm.*

You can find out more about Vernier caliper and how to use one here:

https://www.youtube.com/watch?v=N61VhMxAcQQ

https://www.youtube.com/watch?v=4fBycE1dk0I

Data: The two columns A and B contain eight measurements of the diameter D of the steel rod made by two different students.

TABLE-I

	A	B
S. No.	D(cm)	D(cm)
1	2.02	2.06
2	2.02	1.98
3	2.03	2.10
4	2.03	1.96
5	2.02	1.94
6	2.03	2.11
7	2.04	2.06
8	2.01	1.99
Average	2.025	2.025
Std. dev, σ_D	0.0093	0.0655
Std. error, $\delta_D = \frac{\sigma_D}{\sqrt{8}}$	0.003	0.023

The two data sets have the same average. However, if you examine at the data in the two columns, you will see that the data in column B has a larger spread in the reported measured values. This fact is easily captured by the standard deviation and standard error values of for the two columns.

This is how student A should report the results:

$D = 2.025 \pm 0.009 cm$, if reporting using the standard deviation method.
$D = 2.025 \pm 0.003 cm$, if reporting using the standard error method.

Student B should report their results in the same manner.

We can ask: Whose results are more trustworthy?

How could this 'experiment' – measuring the diameter of a steel rod - be refined/improved?

1. The number of readings should have been increased, especially by student B. There is a large statistical spread in the data of Student B. Generally, when you increase the number of readings, the mean tends to be more reliable. Two of B's readings, 2.11 cm and 1.96 cm, differ by *1.5 mm,* or *15* times the precision of the caliper. Such readings can be misleading. When you increase the number of observations, the role of such outliers is diminished so your results are more reliable. Sometimes, however, these outliers exist for a physical reason, and increasing the number of observations can reveal that. Therefore, you should not exclude outliers from your report. Sometimes such 'odd' readings have led to important scientific discoveries. An increase in the number of readings may not affect the standard deviation as much as it would the standard error. (Do you know why? Examine the expressions for standard deviation and standard error.)

2. If you report that the diameter, *D* of the steel rod you measured, there is an expectation that it is the same *D* all along the length of the rod. But there is nothing in the reported experiment that shows that it is. Therefore, for the reported *D* to be meaningful, the measurements should have been made at a few (three or four) places along the length of the rod.

No two experiments will involve taking the same steps to conduct them or to refine them. Here, improving/refining simply means that the experiment be conducted and analyzed in such a manner that your intended audience can use your data and results and feel confident that the results presented are reliable and that all claims/objectives stated in the experiment are well supported by the data and the analysis presented in the experiment's report.