Uncertainty in measurements

In physics, as in every other experimental science, one cannot make any measurement without having some degree of uncertainty. A proper experiment must report for each measured quantity both a “best” value and an uncertainty. Thus it is necessary to learn the techniques for estimating them. Although there are powerful formal tools for this, simple methods will suffice in this course. To a large extent, we emphasize a “common sense” approach based on asking ourselves just how much any measured quantity in our experiments could be “off”.

One could say that we occasionally use the concept of “best” value and its “uncertainty” in everyday speech, perhaps without even knowing it. Suppose a friend with a car at Stony Brook needs to pick up someone at JFK airport and doesn't know how far away it is or how long it will take to get there. You might have made this drive yourself (the “experiment”) and “measured” the distance and time, so you might respond, “Oh, it’s 50 miles give or take a few, and it will take you one and a half hours give or take a half-hour or so, unless the traffic is awful, and then who knows?” What you’ll learn to do in this course is to make such statements in a more precise form about real experimental data that you will collect and analyze.

Semantics: It is better (and easier) to do physics when everyone taking part has the same meaning for each word being used. Words often confused, even by practicing scientists, are “uncertainty” and “error”. We hope that these remarks will help to avoid sloppiness when discussing and reporting experimental uncertainties and the inevitable excuse, “Oh, you know what I mean (or meant).” that attends such sloppiness.

We rarely carry out an experiment by measuring only one quantity. Typically we measure two or more quantities and then “fold” them together in some equation(s), which may come from theory or even be assumed or guessed, to determine some other quantity(ies) that we believe to depend on them. Typically we compare measured result(s) with something – previous measurement(s) or theory(ies) or our assumption(s) or guess(es) – to find out if they do or do not agree. Since we never know exactly results being compared, we never obtain “exact agreement”. If the results being compared differ by less/more than the combined uncertainties (colloquially, the “sum” of their respective uncertainties), we say that they agree/disagree, but the dividing line is fuzzy. Without uncertainties, you can't say anything about agreement or disagreement, which is why uncertainties are so important in experimental science. We say that there is a “discrepancy” between two results when they “disagree” in the above sense.

Though we may assume that some quantity has an exact “true” result, we cannot know it; we can only estimate it. Now think this way about the agreement/disagreement comparison. If both compared values were known exactly, agreement would mean that the difference between them is zero. Since you don’t know them exactly, the actual compared difference is never exactly zero. We may summarize this by the simple statement, worth remembering, “You cannot measure zero.” What you can say is that if there is a difference between them, it’s less than such-and-such amount. If that amount is less than the combined uncertainty, then we say, “We do not find a discrepancy. The difference between them is consistent with zero.” The difference can never be exactly zero in a real experiment.

A frequent misconception is that the “experimental error” is the difference between our measurement and the accepted “official” value. (Who accepts it? Why? Not just because someone tells you without any evidence why it should be accepted.) What we mean by experimental uncertainty/error is the estimate of the range of values within which the true value of the quantity we’re trying to measure is likely to lie. This range is determined from what we know about our lab instruments and methods. It is conventional to choose the uncertainty/error range as that which would comprise 68% of the results if we were to repeat the measurement a very large number of times.

In fact, we seldom make enough repeated measurements to calculate the uncertainty/error precisely, so we are usually given an estimate for this range. Note, however, that the range is established to include most of the likely outcomes, but not all of them. You might think of the process: As a result of picking the range so that if you bet on the outcome being within this range, you will be right about 2/3 of the time. If you underestimate the uncertainty, you will eventually lose money after repeated bets. (Now that’s an error you probably don’t want to make!) If you overestimate the range, few will be willing to take your bet!

Error

Since nearly everyone refers to “Error Analysis” and not “Uncertainty Analysis” in measurement science, we bow to custom and will use “error” even if we really mean “uncertainty”. We say that there is a difference between two results when they “disagree” in the above sense.

Error such as that quoted above for the book length is called the absolute error; it has the same units as the quantity itself (cm in the example). Note that if the quantity $X$ is multiplied by a constant factor $a$, the absolute error of $(aX)$ is

$$
\Delta(aX) = a \Delta X
$$

(E.1)
Relative Error

We will also encounter **relative error**, defined as the ratio of the error to the best value of the quantity, so that the relative error of \( X \) is given by

\[
\text{relative error of } X = \frac{\Delta X}{X}
\]

Thus the relative error of the book length is \( \frac{\Delta l}{l} = \frac{|0.1|/25.1|}{25.1} = 0.004 \). (If a decimal number is in the range \(-1 < x < 1\), always write it with the "leading zero", e.g., 0.004 in the previous sentence.)

The relative error is dimensionless, and should be quoted with as many significant figures as are known for the absolute error. Note that if the quantity \( X \) is multiplied by a constant factor \( a \) the relative error of \((aX)\) is the same as the relative error of \( X \),

\[
\frac{\Delta (aX)}{aX} = \frac{\Delta X}{X}
\]

since the constant factor \( a \) cancels in the relative error of \((aX)\). Note that quantities with errors assumed to be negligible are treated as constants.

You are probably used to the **percentage error** from everyday life. The percentage error is the relative error multiplied by 100. In the example above, it is 0.004 = 0.4%.

**Changing from a relative to absolute error:**

Often in your experiments you have to change from a relative to an absolute error by multiplying the relative error by the best value,

\[
\Delta X = \frac{\Delta X}{X} \times X
\]

Random Error

Random error occurs because of small, uncorrelated variations in the measurement process. For example, measuring the period of a pendulum with a stopwatch will give different results in repeated trials for one or more reasons. One reason could be that the watch is defective, and its ticks don't come at regular intervals. Let's assume that you have a "good" stopwatch, and this isn't a problem. (How do "you know for certain" that it isn't a problem? Think about this!) A more likely reason would be small differences in your reaction time for hitting the stopwatch button when you start the measurement as the pendulum reaches the end point of its swing and stop the measurement at another end point of the swing. If this error in reaction time is random, the average period over the individual measurements would get closer to the correct value as the number of trials \( N \) is increased. The correct reported result would begin with the average for this best value,

\[
\bar{t} = \frac{\sum t_i}{N},
\]

and it would end with your estimate of the error (or uncertainty) in this best value. This usually taken as the standard deviation of the measurements. (In practice, because of time limitations we seldom make a very large number of measurements of a quantity in this lab course.)

An estimate of the random error for a single measurement \( t_i \) is

\[
\Delta t = \sqrt{\frac{\sum (t_i - \bar{t})^2}{N-1}},
\]

and an estimate for the error of the average \( \bar{t} \) is

\[
\Delta \bar{t} = \sqrt{\frac{\sum (t_i - \bar{t})^2}{N(N-1)}}.
\]

where the sum denoted by the \( \Sigma \) symbol is over the \( N \) measurements \( t_i \). Note in equation (E.5b) the "bar" over the letter \( t \) (\( \bar{t} \) is pronounced "tee bar") indicates that the error refers to the error in the average time \( \bar{t} \). (Each individual measurement \( t_i \) has its own error \( \Delta t_i \).

In the case that we only have one measurement but somehow know (from, say, a previous set of measurements) what the error of the average is, we can use this error of the average \( \Delta \bar{t} \), multiplied by \( \sqrt{N} \) as the error of this single measurement (which you see when you divide equation (E.5a) by equation (E.5b).) [this paragraph updated 9/19/12 because of update of Eq. (5a)]

If you don't have a value \( \Delta \bar{t} \) for the error of \( \bar{t} \), you must do something! Better than nothing is a "guesstimate" for the likely variation based on your experience with the equipment being used for the measurements. For example, for measurements of the book length with a meter stick marked off in millimeters, you might guess that the random error would be about the size of the smallest division on the meter stick (0.1 cm).

Systematic Error

Some sources of uncertainty are not random. For example, if the meter stick that you used to measure the book was warped or stretched, you would never get an accurate value with that instrument. More subtly, the length of your meter stick might vary with temperature and thus be good at the temperature for which it was calibrated, but not others. When using electronic instruments such as voltmeters and ammeters, you obviously rely on the proper calibration of these devices. But if the student before you dropped the meter and neglected to tell anyone, there could well be a systematic error for someone unlucky enough to be the one using it the next time. Estimating possible errors due to such systematic effects really depends on your understanding of your apparatus and the skill you have developed for thinking about possible problems. For example if you suspect a meter stick may be miscalibrated, you could compare your instrument with a 'standard' meter, but, of course, you have to think of this possibility yourself and take the trouble to do
the duty of the Federal Government "To coin Money, regulate the Value thereof, and of foreign Coin, and fix the Standard of Weights and

Institute of Standards and Technology, or by similar organizations in other countries.

to acquire and maintain.

standard, etc.

To achieve high experimental accuracy requires that all measuring instruments

are with respect to the bull's eye. But this tells you nothing about how accurate they are. To aim properly, an archer needs to know

where the bull's eye is, the archer's shots could still cluster tightly but there's no way of the archer knowing

whether they are correct or not. This is much better than having other scientists publicly question the validity of published results done by others that they have reason

to believe are wrong. Occasionally, if authors realize that their work in a published paper was "completely" wrong, they may ask the

journal editors to publish a "retraction" of their paper. When scientific fraud is discovered, journal editors can even decide on their own

to publish a retraction of fraudulent paper(s) previously published by the journal they edit. This does happen, and in this way "science

corrects itself."

Visual Comparison of Types of Error

A figure like the one below is often used to make a visual comparison of types of error, and it allows us to introduce additional
terminology that is often used (incorrectly!) when discussing measurements. You want to be sure you understand the terminology and
use it correctly.

Think of the round object as an archery target. The archer shoots some number of arrows at it, and each dot shows where one landed.

Now think of the "bull's eye" – the larger black dot in the center – as the "true" value of some quantity that's being measured, and think

of each arrow-dot as a measurement of that quantity. The problem is that the one doing the measurements does not know the "true"

value of the quantity; s/he's trying to determine it experimentally, and this means there must be uncertainty associated with the

experimentally determined value. Note that each archery target – we'll call them 1,2,3,4 from left to right – shows a different

distribution of arrow-hit/measurements.

In number 1 the measurements cluster pretty tightly: we say that the statistical (random) error is small, and the terminology we

introduce for that is, "These measurements are precise." However, the center of their distribution is far from the bull's eye: we say that

there is a large systematic error, and the terminology we introduce for that is, "These measurements are not accurate." In a few words,

"These measurements are precise but inaccurate."

In number 2 the measurements do not cluster tightly, but one can see that the center of their distribution is not far from the bull's eye.

These measurements have a large statistical error but a small systematic error. In a few words, "These measurements are imprecise

but accurate."

In number 3 the measurements do not cluster tightly, and one can see that the center of their distribution is not close to the bull's eye.

These measurements have a large statistical error and a large systematic error. In a few words, "These measurements are imprecise

and inaccurate."

In number 4 the measurements cluster tightly, and one can see that the center of their distribution is very close to the bull's eye. These

measurements have a small statistical error and a small systematic error. In a few words, "These measurements are precise and

accurate."

Here is a crucial point: You can always know your measurements achieve a high level of precision if they cluster tightly, and you can

quantify "how precise" they are. But this tells you nothing about how accurate they are. To aim properly, an archer needs to know

where the bull's eye is, but suppose, in our analogy, a white sheet is put up to block view of the target. Not knowing where the bull's

eye is, the archer's shots could still cluster tightly but there's no way of the archer knowing whether they are correct or not. The accuracy is unknown.

To achieve high experimental accuracy requires that all measuring instruments and all measurement procedures need to be thoroughly

understood and calibrated, quantitatively, against relevant "standards", e.g., the length standard, the time standard, the voltage

standard, etc. The average laboratory, and certainly our undergraduate teaching laboratories, lack such standards. They are expensive
to acquire and maintain. Periodically they should be compared with "the" (primary) standards maintained, say, by NIST, the National

Institute of Standards and Technology, or by similar organizations in other countries. Section 8 of the U.S. Constitution specifies that

is the duty of the Federal Government "To coin Money, regulate the Value thereof, and of foreign Coin, and fix the Standard of Weights and
Propagation of Errors

Often in the lab, you need to combine two or more measured quantities, each of which has an error, to get a derived quantity. For example, if you wanted to know the perimeter of a rectangular field and measured the length \( l \) and width \( w \) with a tape measure, you would then have to calculate the perimeter, \( p = 2(l + w) \), and would need to get the error of \( p \) from the errors you estimated for \( l \) and \( w \), \( \Delta L \) and \( \Delta w \). Similarly, if you wanted to calculate the area of the field, \( A = lw \), you would need to know how to do this using \( \Delta L \) and \( \Delta w \).

There are simple rules for calculating errors of such combined, or derived, quantities. Suppose that you have made primary measurements of quantities \( A \) and \( B \), and want to get the best value and error for some derived quantity \( S \).

**Case 1**: For addition or subtraction of measured quantities the absolute error of the sum or difference is the 'addition in quadrature' of the absolute errors of the measured quantities; if \( S = A \pm B \)

\[
\Delta S = \sqrt{\Delta A^2 + \Delta B^2}.
\]

(E.6)

This rule, rather than the simple linear addition of the individual absolute errors, incorporates the fact that random errors (equally likely to be positive or negative) partly cancel each other in the error \( \Delta S \).

**Case 2**: For multiplication or division of measured quantities the relative error of the product or quotient is the 'addition in quadrature' of the relative errors of the measured quantities; if \( S = A \times B \) or \( \frac{A}{B} \)

\[
\frac{\Delta S}{S} = \sqrt{\left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2}.
\]

(E.7)

Due to the quadratic addition in (E.6) and (E.7) one can often neglect the smaller of two errors. For example, if the error of \( A \) is 2 (in arbitrary units) and the error of \( B \) is 1, then

the error of \( S = A + B \) is \( \Delta S = \sqrt{(\Delta A)^2 + (\Delta B)^2} = \sqrt{2^2 + 1^2} = \sqrt{5} = 2.23 \).

Thus, if you don't want to be more precise in your error estimate than \( \sim 12\% \) (which in most cases is sufficient, since errors are an estimate and not a precise calculation) you can simply neglect the error in \( B \), although it is 1/2 of the error of \( A \).

**Case 3**: When you're interested in a measured quantity \( A \) that must be raised to the \( n \)-th power in a formula (\( n \) doesn't have to be an integer, and it can be positive or negative), viz., you're interested in \( A^n \), the relative error of the quantity \( A^n \) is the relative error of \( A \) multiplied by the magnitude of the exponent \( n \):

\[
\frac{\Delta A}{A} = |n| \times \frac{\Delta A}{A}.
\]

(E.8)

As an example for the application of (E.8) to an actual physics problem, let's take the formula relating the period \( T \) and length \( L \) of a pendulum:

\[
T = 2\pi \sqrt{\frac{L}{g}}
\]

(E.9a)

where \( g = 9.81 \text{ m/s}^2 \) is the constant acceleration of gravity. We rewrite (E.9a) as

\[
T = \left(\frac{2\pi}{\sqrt{g}}\right)^{1/2} L^{1/2}
\]

(E.9b)

to put all the constants between the parentheses. We now identify \( S \) in (E.8) with \( T \) and identify \( A^n \) with \( L^{1/2} \). Therefore, we identify \( A \) with \( L \) and see that \( n = +\frac{1}{2} \) for our example. Since \( |n| \) appears in (E.8) [the vertical bars around \( n \) mean "absolute value"], only the magnitude of \( n \) is important, so we don't have to worry about the sign of \( n \): we get the same result whether the exponent \( n \) is positive or negative, as long as it's \( \frac{1}{2} \) in our example. If we're interested in evaluating \( \frac{\Delta T}{T} \), we see from (E.3) that the constant \( \alpha \), which in our case equals \( \frac{2\pi}{\sqrt{g}} \), "drops out".

Therefore, we find that \( \frac{\Delta T}{T} = \frac{1}{2} \left(\frac{\Delta L}{L}\right) \). This example should help you apply (E.8) to cases having values of the exponent \( n \) different from the particular value used in this example.