Introduction to Error Analysis

I. HOW TO REPORT EXPERIMENTAL DATA

Every time we perform a measurement, no matter how careful we are, we cannot produce results free of errors. Scientifically speaking, by errors we do not mean mistakes, but the fact that each experimental result is obtained with some uncertainty. Knowing possible sources of errors and how to eliminate or reduce them is very important for experimental physics.

Generally, we can distinguish between two types of error. One type of error is a bias or systematic error. A bias means that every time we make a measurement with a device, the measurement overestimates or underestimates the "true" value by some amount. An example of a bias would be a meter stick that was stretched so that the 100 cm mark on the stick was really 102 cm from the 0 cm mark. Any measurements made with this meter stick will be 2% too small. Another example is a clock that is running fast.

You can often eliminate biases from your measurements by calibrating the measuring device against some standard whose true value (you think) you know accurately. For example, if you used the stetched meter stick to measure some object that you think is really 100 cm long, you might discover the 2% bias. You could then scale every measurement made with the stick down by 2%, eliminating the bias. However, the difficulty with biases is that's generally all you can do. It is therefore important to carefully check that your instruments before beginning a measurement. There is usually no way to correct for, estimate, or account for biases through experimental technique during the measurement, or by analyzing your data after you perform the measurement. If you perform a measurement with biased equipment your result may simply be wrong with little more that can be said.

Any measuring device also has some limited precision; this means it cannot measure the parameter it was designed for to an arbitrary number of significant figures. These types of errors are called ambiguities. Ambiguities cannot be eliminated as we can with biases, they are inherent to experimental science. However, we can usually estimate them, and perhaps even limit them to some extent, by using good experimental and data analysis techniques.

Example of an ambiguity Consider measuring the length of a pencil using a measuring stick. You can use a standard ruler that is commercially available and has long ticks for cm and shorter ticks for mm. We set one end of the pencil on the zero tick of the meter stick, and estimate the value of the tick for the other end of the pencil. Often the end of the pencil falls in between two consecutive tick marks on the ruler, and one has to decide which mark on the meter stick gives the correct length of the pencil. Suppose, for example, we decide the end of the pencil is closest to the 125 mm mark. We are then saying the "true" length of the pencil is closer to 125 mm than it is to 126 mm or 124 mm, but we cannot say anything more precise than that, as the ruler does not have any finer markings. Our measurement therefore includes all values that round to 125 mm. We can conclude that the precision of our ruler is ± 0.5 mm, one half of the smallest increment shown on the ruler.

In other words, every time we want to measure the length of an object we cannot state the length of the object any more precisely than ± 0.5 mm. So, for example, we could never determine that the pencil was 125.03 mm in length with this ruler, even if that were the true value, as 0.03 mm is below the 0.5 mm precision of the ruler. We have to round our result to 125.0 mm, as this is the most precision our ruler is capable of.

Another source of ambiguity is related to how we actually perform the experiments. If we take consecutive measurements of the same event, most of the time we will end up with slightly different results. For example, if I time some event with a stopwatch, probably the watch reports slightly different durations each time I make a measurement, due to variations in my reactions starting and stopping the watch. Such ambiguities are called random errors or statistical errors. One can estimate and even reduce random errors by repeating the same measurement multiple times.

Example of random errors Consider measuring the period of a pendulum using a stop watch. You will need to start the stop watch when the pendulum reaches one end of its swing, and stop the watch once it returns to the same end. Your reaction time will be the main source of random errors, as you will need to start and then stop the stop watch exactly when the pendulum reaches the end point of the swing. Let's consider 5 measurements recorded by the stop watch:

$$4.9 \text{ s};\ \ 5.1 \text{ s};\ 4.9 \text{ s};\ 5 \text{ s};\ 5.1 \text{ s};\ . \tag{1}$$

Clearly, the differences between these values are not due the precision of the instrument we used, as all measurements vary by more than the 0.1 s increment of the watch (which is obvious since the watch can measure the variation).

The question is, *how does one report experimental results*? Let us consider the pendulum experiment. Our result indicates that the period of the pendulum ranges between 4.9 s and 5.1 s, meaning that our best estimate of the pendulum period is 5.0 s. We can report the result as

$$\text{pendulum period} = 5.0 \ \pm\ 0.1 \text{ s}\ . \tag{2}$$

or

$$\text{Result} \ =\ \text{Best Estimate} \ \pm\ \text{Uncertainty}\ .$$

TABLE I. Rules for error propagation.

Type of combination	Mathematical Expression	Uncertainty
Sums and Differences	$q = x + \cdots + y - (u + \cdots v)$	$\delta q = \sqrt{(\delta x)^2 + \cdots (\delta z)^2 + (\delta u)^2 \cdots (\delta v)^2}$
Products and Quatients	$q = \frac{x \times \cdots \times z}{u \times \cdots \times v}$	$\frac{\delta q}{\lvert q \rvert} = \sqrt{\left(\frac{\delta x}{\lvert x \rvert}\right)^2 + \cdots + \left(\frac{\delta z}{\lvert z \rvert}\right)^2 + \left(\frac{\delta u}{\lvert u \rvert}\right)^2 \cdots + \left(\frac{\delta v}{\lvert v \rvert}\right)^2}$
Multiplication by a constant a	$q = a \times x$	$\frac{\delta q}{\lvert q \rvert} = \frac{\delta x}{\lvert x \rvert}$
Power function	$q = x^n$	$\frac{\delta q}{\lvert q \rvert} = \lvert n \rvert \frac{\delta x}{\lvert x \rvert}$
Function of variables	$q \equiv q(x, \cdots, z)$	$\delta q = \sqrt{\left(\frac{\partial q}{\partial x}\delta x\right)^2 + \cdots + \left(\frac{\partial q}{\partial z}\delta z\right)^2}$

Scientifically, if we measure quantity x with uncertainty δx we report the result as

$$\text{(measured value of x)} \;=\; x_{best} \;\pm\; \delta x \;, \tag{3}$$

meaning that our best estimate for the result of the measurement is x_{best} and that we are pretty confident that any measurement will fall in a range $x_{best} - \delta x$ and $x_{best} + \delta x$.

Notice that when we reported our result, we reported δx to 1 significant figure, and x_{best} to the same number of decimal points as δx. The reason δx is only ever reported to 1 significant figure is uncertainties are never more than sensible estimates, so we report them with the least precision possible. The uncertainty then sets the number of significant digits for our result. In our example above, we are reporting that we are uncertain about the value of the "0" in "5.0", since this is the uncertainty limit. It would therefore be silly to try and report any additional digits in our result, as they could literally be anything.

There are various ways to report the uncertainty of a measurement. Note that the absolute uncertainty δx used in equation (3) has the same units as the quantity x that we are measuring. One other way to report errors is to use the so-called fractional uncertainty

$$\text{fractional uncertainty} \;=\; \frac{\delta x}{|x|} \;. \tag{4}$$

The advantage of the fractional uncertainty is that it is a dimensionless quantity, and it can be reported as a percentage. Using the fractional uncertainty we report the result of our measurement as

$$\text{(measured value of x)} \;=\; x_{best}\left(1 \pm \frac{\delta x}{|x|}\right) \;. \tag{5}$$

II. PROPAGATION OF UNCERTAINTIES

Most of the time, quantities of interest are expressed as a combination of other measurable quantities. For example, if one tries to measure the gravitational acceleration using a pendulum, they first measure the length of the pendulum l, its period of oscillation T, and finally estimate the gravitational acceleration using the formula $g = 4\pi^2 l/T^2$. The natural question that arises if we know the uncertainty for each measured quantity is *what will be the uncertainty for the combined quantities*?

Consider now a quantity q that is a combination of quantities x, y, ..., each with an absolute uncertainty δx, δy, ... Table I summarizes the rules for calculating the uncertainty in quantity q for various combinations of quantities x, y, ... To better understand the use of these general formulas let us consider few examples.

Sums and Differences The water volume in a big cylinder is measured using two different smaller graduated cylinders. The first measurement gives $V_1 = 200 \pm 3$ ml and the second $V_2 = 500 \pm 4$ ml. Assuming that the total water volume is given by $V = V_1 + V_2$ we can calculate the total absolute error as

$$\delta V = \sqrt{(\delta V_1)^2 + (\delta V_2)^2} = \sqrt{3^2 + 4^2} = 5 \text{ ml}$$

and report the final result as $V = 700 \pm 5$ ml.

Products and Quatients Let us return to our previous example for the estimation of the gravitational acceleration using a pendulum. Assume that the pendulum length was measured as $l = 99.20 \pm 0.02$ cm and the period of oscillation as $T = 2.00 \pm 0.01$ s. The calculated best value for the gravitational acceleration is

$$g = \frac{4\pi^2 l}{T^2} = 979 \text{ cm/s}^2$$

and the fractional uncertainty

$$\frac{\delta g}{g} = \sqrt{\left(\frac{\delta l}{|l|}\right)^2 + \left(\frac{\delta T^2}{|T^2|}\right)^2} = 1\% \ .$$

Finally we can report the value of the gravitational acceleration as $g = 980 \pm 10$ cm/s^2.

Function of Variables We would like to calculate the acceleration of a cart down an incline. Assume that the angle of the incline was measured as $\theta = 20 \pm 3^0$ and we know the gravitational acceleration from the previous example as $g = 9.8 \pm 0.1$ m/s^2. According to the general formula, neglecting friction on the incline, $a = g \sin\theta$. Our best estimate for the acceleration down the incline is $a = 3.35$ m/s^2. The corresponding uncertainty

$$\begin{aligned}\delta a &= \sqrt{\left(\frac{\partial a}{\partial g}\delta g\right)^2 + \left(\frac{\partial \sin\theta}{\partial\theta}\delta\theta\right)^2} \\ &= \sqrt{\left(\sin\theta\delta g\right)^2 + \left(g\cos\theta\delta\theta\right)^2} \\ &= \sqrt{(0.03)^2 + (0.48)^2} = 0.48 \text{ m/s}^2 \ .\end{aligned}$$

The cart acceleration down the incline can be reported as $a = 3.35 \pm 0.48$ m/s^2. In the calculation of the acceleration uncertainty $\delta\theta$ has to be expressed in radians.

III. STATISTICAL ANALYSIS OF RANDOM ERRORS

Random uncertainties can be estimated using statistical analysis. Consider that we were tasked with measuring a certain quantity x, and after we repeated the experimental procedure we ended with a set of measured values $x_1, x_2, \ldots, x_N$. Reasonably, our best estimate for x will be the average of the measured values

$$x_{best} = \bar{x} = \frac{x_1 + x_2 + \cdots x_N}{N} \ .$$

We can introduce the *sample standard deviation* of the set $x_1, x_2, \ldots, x_N$ as

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} \ .$$

If the number of measured values in the set, N, is large enough, and all the sources of uncertainty in the experiment are small and random, then the measured values in the set will be distributed around the true value x_{true} in accordance with a normal, bell-shaped curve. In particular, about 68% of the results in the set will be within a distance σ_x on either side of x_{true}. This suggests that we can adopt σ_x as a measure of the random uncertainty in our experiment. However, our best estimation of the outcome of measuring x is the average value of set $\bar{x}$, and a better measure of the uncertainty in the measurement is given by the so-called *standard deviation of the mean* defined as

$$\delta x = \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}} \ .$$

We will report our experiment outcome as

$$(\text{measured value of } x) = \bar{x} + \sigma_{\bar{x}} \ . \tag{6}$$

Notice that as $\sigma_{\bar{x}} \propto 1/\sqrt{N}$ the uncertainty in our experiment decreases as we increase N, as we might expect. So you can always reduce the random error in your data by simply measuring additional data points.

That said, there is another implication of the $1/\sqrt{N}$ scaling. Suppose we have taken n measurements and we estimate our random error as $\sigma_{\bar{x}}$. Say we wanted to halve our uncertainty to improve our measurement. Simply taking more data points, we will need to perform $4n$ as many measurements. If we wanted to reduce the uncertainty by a factor of 10 (the amount needed to claim an additional significant figure in our result), we have to perform $100n$ measurements! Be careful not to think that you can always reduce random errors to any arbitrarily low value just by taking more data. You can do that, but the amount of work required quickly gets very large to achieve substantive reductions. It is usually better to reduce random errors by improving the design of your experiment, say by using better equipment or techniques, if possible.

Another factor to consider is that the total uncertainty in your data usually includes other amibiguities beyond random variations, such as the instrumental precision. Even if you do reduce your random uncertainty to a very low value by taking a huge number of data points, your total uncertainty will eventually become limited by other sources.

Example As an example of statistical analysis of random errors we will consider again the case of a pendulum that allows us to estimate the value of the gravitational acceleration from the relation $g = 4\pi^2 l/T^2$. In the experiment, we measure the length of the pendulum l and its period of oscillation T. To better estimate the oscillation period, instead of measuring the time for a single swing of the pendulum, we will let the pendulum swing multiple times ($n = 5$ in our case) and find the period as the total measured time divided by the number of swings $T = T_n/n$. Accordingly, the gravitational acceleration can be calculated as $g = 4\pi^2 n^2 l/T_n^2$. Table II presents our experimental results for T_n.

According to the analysis presented in the table, the result of measuring the oscillation time can be reported as $T_n = 10.39 \pm 0.021$ s. The precision of the stop watch we used to measure the oscillations time will introduce another source of ambiguity, in our case $(\delta T_n)_{instr} = 0.005$ s. Finally one can present the result of measuring the oscillation time as $T_n = 10.39 \pm 0.021 \pm 0.005$ s, or if we combine the two errors $T_n = 10.39 \pm 0.0217$ s. Additionally, the length of the pendulum was measured to $l = 1.077 \pm 0.002$ m. (Note that for the pendulum length

TABLE II. Experimental data for the pendulum oscillation time.

Trial	$T_n(s)$	Average T_n	σ_x	$\sigma_{\bar{x}}$
1	10.35			
2	10.37			
3	10.41	10.39	0.047	0.021
4	10.35			
5	10.46			

we do not have statistical error, the only error is introduced by the ruler we used $(\delta l)_{instr} = 0.005$ m.) Using the measured length and oscillation time, we can estimate the best value of the gravitational acceleration as $g_{best} = 9.850$ m/s^2. Finally, we can evaluate the ambiguity in the gravitational acceleration as

$$\delta g = g_{best}\sqrt{\left(\frac{\delta l}{|l|}\right)^2 + \left(2\frac{\delta T_n}{|T_n|}\right)^2} = 0.04 \text{ m/s}^2 \ .$$

We report our result as $g = 9.85 \pm 0.04$ m/s^2.

IV. GRAPHICAL ANALYSIS

During the course of your physics labs, you will perform experiments to check the relation between different physical quantities. Assume that you are measuring two quantities x and y. When you try to relate the values obtained for the two quantities there are two possibilities. One possibility is that the two quantities are not related, i.e., if we change x there will be no change in y and vice versa. Another possibility is that if we change x there will be a change in y and vice versa. In this case we do have a relation between the two quantities x and y, and we can use graphical analysis to predict what type of relation we have between the two quantities.

The simplest possible relation is a linear dependence between the two quantities, in other words, if we have a Δx change in x, it will trigger a change in y so that $\Delta y = m\Delta x$, m being a constant. In this case, if we plot y vs x, we should obtain a linear dependence of the form

$$y = mx + p \ , \tag{7}$$

where m is the slope and p is the y-intercept of the line. In reality, if you consider the experimental values for both x and y and you try to put these values on a plot y vs x, almost certainly these points will not lie exactly on a straight line. The important question is if this situation is due to the ambiguities in our measurements or even from the possibility that y and x are not linearly dependent. To answer this question we will consider for simplicity that when we measure x, we have almost no ambiguity, but when we measure y we end up with an error $\delta y = \sigma_{\bar{y}}$. To account for the error, whenever we plot a value of y we will need to attach an error bar of size δy around each point to indicate the range in which it lies. If we can find a straight line that passes close or within all error bars, we can conclude that we have a linear dependence between x and y. The advantage of linear dependences is that all data analysis softwares (Excel, Google Analytics) have very good interpolation routines to find a straight line that fits our experimental points.

TABLE III. Experimental data for the pendulum period.

l (m)	T (s)	T^2 (s^2)	δT^2 (s^2)
0.2	0.90	0.81	0.022
0.25	0.99	0.98	0.024
0.3	1.07	1.14	0.026
0.35	1.13	1.28	0.027
0.4	1.20	1.44	0.029
0.45	1.33	1.77	0.032
0.5	1.38	1.90	0.034
0.55	1.49	2.22	0.036
0.6	1.55	2.40	0.038
0.65	1.60	2.56	0.039
0.7	1.68	2.82	0.041

Sometimes, the two considered quantities do not have a linear dependence, but they are connected through a more complicated relation, for example a power law $y = Ax^n$, or even an exponential law $y = A\exp[x]$ (A and n are constants). In cases like this, it is always preferably to bring the plot back to a linear dependence For example, in the case of a power law, instead of plotting y vs. x, it is more advantageous to plot y vs. x^n. For an exponential law, we can exploit the properties of the logarithmic function and plot a linear dependence in the form $\ln y = \ln A + x$.

Uncertainties in slope and y-intercept When we extract physical information from a fitted function, such as the slope and y-intercept of a straight line, we must recall that the measured slope and offset are themselves subject to some uncertainty. There are various ways we can estimate the uncertainty in fit parameters. We can use a simplified procedure and calculate the uncertainty in the slope (δm) and y-intercept (δp) as

$$\begin{aligned} \delta m &= \frac{\delta y_n + \delta y_0}{x_n - x_0} \\ \delta p &= x_n \delta m - \delta y_n \ , \end{aligned} \tag{8}$$

where the n and 0 subscripts denote the rightmost and leftmost data points, respectively.

Example We will use the pendulum experiment to exemplify how we can extract the gravitational acceleration value using graphical analysis. As we already mentioned, the oscillation period, T, of a pendulum of length, l, is given by $T = 2\pi\sqrt{l/g}$. Clearly, the relation between T and l is not linear, so we will have to linearize this relation. One choice is $T^2 = (4\pi^2/g)\cdot l$. If we plot T^2 vs. l, the slope of the curve will allow us to extract the value of the gravitational acceleration as $g = 4\pi^2/m$, with m being the slope of the line. In Table III we present experimental data for the period of the pendulum. The ambiguity in the length of the pendulum measurement

FIG. 1. T^2 vs. l. The slope of the line is $m = 4.105$ and the y-intercept $p = -0.091$.

is $\delta l = 0.005$ m and that in its oscillation period measurement is $\delta T = 0.012$ s. The ambiguity in the value of T^2 can be evaluated using the rules we introduced as $\delta T^2 = 2T\delta T$.

Figure 1 presents our plot for T^2 vs. l (we used Excel to graph the data). Notice that we include error bars to account for the ambiguity in our measurements.The line fit is given by $y = 4.105x - 0.091$, so the slope of the line is $m = 4.105$ and its y-intercept is $p = -0.091$. We also estimate $\delta m = 0.125$ and $\delta p = 0.047$. Finally, we report the value of the gravitational acceleration as $g = 9.6 \pm 0.3$ m/s^2.

V. SIGNIFICANCE AND DISCREPANCY

The basic point of experimental science is always to decide if our measurements are consistent with what we expect from theory. When we compare our results with what we expected, usually we will find that our measurements are not *exactly* what we though we will get. We therefore are often tasked with deciding if the difference, or “discrepancy” between our measurements and expectation is “significant”. Significant means that the difference cannot be explained by the limitationis of the measurement. If the discrepancy is significant, that result might indicate some flaw in the theory (or a flaw in the experiment). If the discrepancy is not significant, we might conclude that our result is consistent with what we expected from theory. In turn, that observation may cause us to increase our confidence in the soundness of the theory. In our example above, the period of 5 swings of a pendulum was measured to be 10.39 s, whereas theory predicted we should expect to get 10.41 s. Does this difference matter?

In this case, the discrepancy between our measured and expected value was 0.02 s. The uncertainty in the measurement was also 0.02 s. When a discrepancy is similar to or less than the uncertainty in the experiment, this means that the difference between our measurement and expected value can be explained by the limitations of the measurement. In this case, the discrepancy is not significant, and we might conclude that the result of our measurement is consistent with what theory predicts.

However, suppose we had actually measured a value of $T_n = 10.36 \pm 0.02$ s. In this case, the discrepancy with theory would be 0.05s. Since the uncertainty in our measurement only explains 0.02s, in this case the discrepancy with theory would be significant. When we have a significant discrepancy, this means we are entitled to conlude that the experiment and theory are showing different results. In that case, we have probably demonstrated that either the theory or experiment is flawed, and we might have to decide which. In a case like this, the theory for the period of a pendulum is very will accepted, so we would probably be inclined to suspect that we did something wrong in our measurement. However, professional experimental scientists spend most of their time measuring the predictions of theories that are not well established (the point of the measurement is generally to test if what the theory predicts is accurate). Professional scientists typically defer to the results of measurements and look for flaws in new theories when experiments show significant discrepancies with the theory.

VI. PROBLEMS

Problem 1. A student studying the motion of a cart on an air track measures its position, velocity, and acceleration at one instant, with the results shown in Table IV. Rewrite these results in the standard form $x_{best} \pm \delta x$.

TABLE IV. Experimental data for Problem 1.

Variable	Best Estimate	Probable Range
Position, x	53.3	53.1 to 53.5 (cm)
Velocity, v	-13.5	-14.0 to -13.0 (cm/s)
Acceleration, a	93	90 to 96 (cm/s^2)

Problem 2. A student makes the following measurements: $a = 5 \pm 1$ cm, $b = 18 \pm 2$ cm, $c = 12 \pm 1$ cm, $t = 3.0 \pm 0.5$ cm, and $m = 18 \pm 1$ g. Using the general rules, compute the following quantities with their uncertainties and percentage uncertainties: **(a)** $a + b + c$, **(b)** $a + b - c$, **(c)** ct, and **(d)** mb/t.

Problem 3. Calculate the mean, standard deviation, and standard deviation of the mean for the following 30 measurements of time t (in seconds):

8.16, 8.14, 8.12, 8.16, 8.18, 8.10, 8.18, 8.18, 8.18, 8.24
8.16, 8.14, 8.17, 8.18, 8.21, 8.12, 8.12, 8.17, 8.06, 8.10
8.12, 8.10, 8.14, 8.09, 8.16, 8.16, 8.21, 8.14, 8.16, 8.13

How will you report the time from this experiment.

[1] John R. Taylor, *An Introduction to Error Analysis*, Second Edition, University Science Books, Sausalito, California (1997).
[2] Frank Rice, *Physical Data Analysis*, Caltech, Pasadena

(2018).