Appendix A

Measurement and Error Analysis

A.1 Measurement, True Values, and Errors[1]

In a typical experiment, we are usually interested in determining the value of one or more physical quantities: the width of a block of glass, the period of a pendulum, the mass of a body, and so on. Such measurements are always subject to influences and uncertainties that will affect the observations. It is the job of the experimenter to try to minimize these effects, but it is never possible to completely eliminate them. Therefore, we need a method to quantitatively handle the errors that creep into every experiment; i.e., we need to perform a statistical analysis of our data.

Case Study in Randomness: Fairness of a Coin

Consider the following seemingly trivial experiment. A student wants to decide if a coin is fair (or "unbiased"). If the coin is fair, then when she flips it, the number of times it lands heads-up should roughly equal the number of times it lands heads-down.

The student flips the coin 50 times and counts the number of heads n. She expects that if the coin is fair, heads should come up half the time, giving her a count $n = 25$. In this trial, she actually counts 28. To check that she didn't miscount, she decides to flip the coin another 50 times. This time, she counts $n = 19$ heads.

Intrigued, the student decides to repeat her experiment a large number of times N (e.g., $N = 100$). After finishing all 100 trials, she histograms the frequency of each head count (see Fig. A.1). The histogram indicates that over the course of 100 trials, the most common result[2] is 25 heads out of 50 tosses. However, there is significant spread in the data: six trials have a head count under 20; ten trials have a count greater than 30.

[1] Adapted from "Introduction to Experimental Error," Susan Cartwright, University of Sheffield, UK (2003).

[2] This is referred to as the *mode*.

The student also decides to calculate the average (or mean) head count $\bar{n}$ for the 100 trials, knowing that this number is related to the "most likely" outcome of the experiment. She expects 25 heads on average, but calculates

$$\bar{n} = \frac{1}{N}\sum_{i=1}^{N} n_i = 24.79$$

This result is close to 25, but not exactly. The difference is small, but she must ask: is this evidence that the coin is slightly biased? Or are random fluctuations that occur from trial to trial responsible for the discrepancy? How should she quantify these fluctuations?

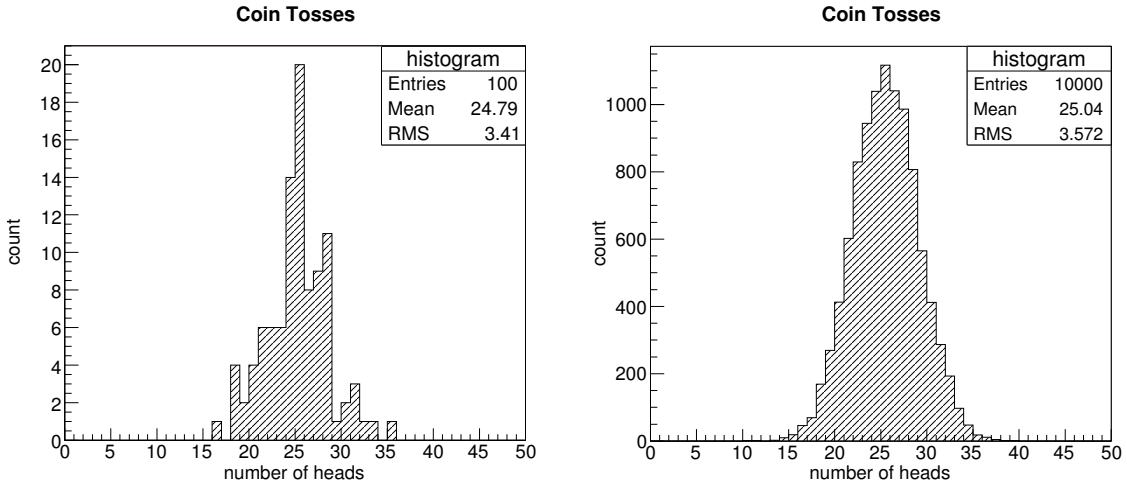

Figure A.1: LEFT: histogram of the number of head counts recorded in 100 trials when tossing a fair coin 50 times per trial. RIGHT: same histogram for 10,000 trials.

Case Study in Precision: Thickness of a Block of Glass

Random fluctuations are a source of uncertainty[3] in all experimental quantities. However, these fluctuations may sometimes be too small to be observed. In such cases, we remember that physical measurements of infinite precision are impossible. Hence, we can use the *resolution* of our measurement devices to estimate the uncertainties in our observations.

For example, suppose a student is asked to measure the width of a block of glass. He uses a plastic ruler accurate to the nearest millimeter, and finds that the block appears to be 10.0 mm thick. Since he is careful, he tries the same measurement at various points along the length of the block. Each time he gets the same result: 10.0 mm.

[3] Uncertainty is often referred to as "experimental error."

Does this set of observations suggest that the block is 10.0 mm thick, *exactly*? No; it only indicates that any fluctuations in the length of the block are smaller than what can be measured with a cheap ruler. In this situation, we quantify the uncertainty in the measurement as one half of the smallest division that appears on the scale of the instrument. (The rule also holds for instruments with digital readouts.) For a ruler accurate to the nearest mm, that means an uncertainty of ± 0.5 mm.

Method	*Typical Error*
cheap ruler	± 0.5 mm
draftsman's ruler	± 0.2 mm
calipers with vernier scale	± 0.05 mm
traveling microscope	± 0.005 mm
interferometer	± 0.00001 mm (if n known accurately)

Table A.1: Experimental error on the thickness of a glass slab.

The resolution lets us place an upper bound on the unobservable but present random fluctuations in the measured quantity. As Table A.1 indicates, we can improve the resolution by using more sensitive instruments.

Importance of Experimental Errors

The scientific approach to understanding the world is based on a number of fundamental assumptions and techniques. Two of the most important are:

- Experiments are reproducible. If you say experiment X produces result Y, I should be able to do experiment X myself and expect to get result Y.
- Theories are tested by experiment. If I assert that theory A is an improvement on theory B, I should be able to point out experimental results that are explained by A and not B, and also predict the results of new experiments yet to be performed.

Clearly, uncertainties are vital to our interpretation of experimental results. If two individuals perform an experiment, it is likely that they will not report exactly the same result due to random fluctuations. However, the relevant question is not whether their results are exactly the same, but whether they agree within the range set by experimental errors. If so, the results are said to be consistent. If not, the discrepancy may be due to an interesting physical effect that should be investigated.

Reporting Results

Hopefully you are now convinced that whenever something is measured, an uncertainty is always involved in that measurement. The complete specification of a physical quantity always includes this uncertainty,

as well the units in which the quantity was measured. For example:

1. $m = 9.0$ kg: WRONG — no uncertainty.
2. $m = 9.0 \pm 0.3$: WRONG — no units.
3. $m = 9.0 \pm 0.3$ kg: RIGHT.

A.2 Precision and Accuracy

Precision

The uncertainty (or "experimental error") reported above is perhaps more accurately described as the *precision* of the measurement. The uncertainty reflects the range of values in which we expect to measure a physical quantity, most of the time. In other words, it is the typical scatter that we see in data when we make repeat measurements, and it is related to the reproducibility of a measurement. A precise measurement is one in which the scatter of the data is small.

Accuracy

Independent of scatter, a value is said to be accurate if it is numerically close to some "true" value. We can define the accuracy (or "relative error") of an experimental result as

$$\boxed{\text{accuracy (in \%)} = 100 \times \frac{|\text{expt} - \text{true}|}{\text{true}}}$$

Note how precision and accuracy are far from the same thing. Values are precise when the scatter in the values about some mean is small. However, this does not imply that the values are close to the true value. Ideally, we want to achieve both accuracy and precision; i.e., we want to obtain results close to the true value with little scatter.

A.3 Types of Errors: Statistical and Systematic

There are two fundamentally different types of experimental error. *Statistical* errors are random in nature: repeated measurements will differ from each other and from the true value by amounts which are not individually predictable, although the average behavior over many repetitions can be predicted. For example, the spread in the number of heads n in our coin-flipping experiment, or the resolution uncertainties of an instrument, are statistical errors. There is an extensive mathematical literature dealing with statistical errors, and most of the rest of this note will be concerned with them.

Systematic errors arise from problems in the design of an experiment. They are not random, and tend to affect all measurements in some well-defined way. For example, suppose you are asked to read

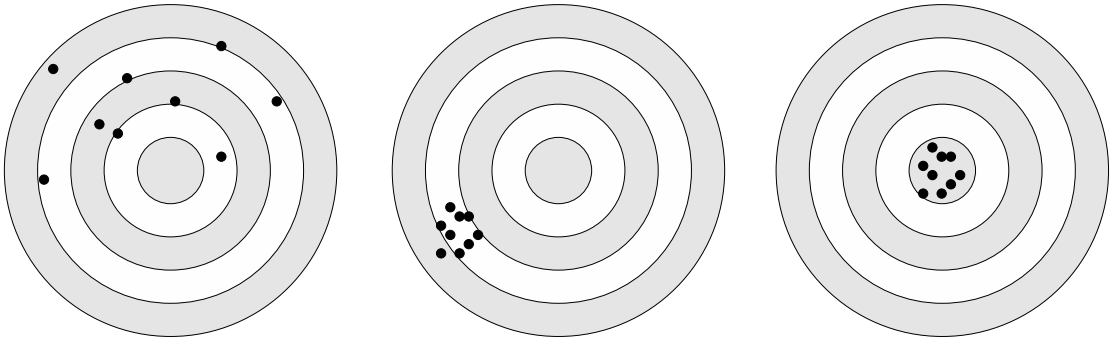

Figure A.2: Precision and accuracy visualized as a bullseye: imprecise and inaccurate results (left); precise but not accurate (center); precise and accurate (right).

the temperature from a mercury-in-glass thermometer. The position of your eye with respect to the scale on the glass introduces a parallax error — a reading error related to the relative position of the scale and the top of the mercury column.

If you consistently view the scale from below the top of the column, your temperature measurements will be consistently too low. If you consistently view the scale from above the top of the column, all of your temperature measurements will be too high. In either case, your data points will systematically shift away from the correct readings.

Systematic errors can be extremely subtle and difficult to diagnose. Repeating measurements usually won't help, but you can identify them by looking for some of the following common symptoms:

- Curved lines on a plot appear where straight lines were expected (especially on log plots);
- Nonzero values appear where zero was expected (e.g., nonzero intercepts in plots);
- Inability to reproduce results, even on the same equipment. This could indicate a dependence on ambient temperature or pressure, or on the running time of the apparatus (common with lasers).

In experiments in the undergraduate lab, systematic errors are often discovered by hindsight during the analysis phase of the experiment. Ideally, one would not want to do things this way; it is better to think about and record possible sources of systematic error *before* starting an experiment, and then attempt to eliminate these effects as measurements are conducted. In practice, this can be quite a challenge. The ability to identify systematics and neutralize their effects is a skill acquired through much practice.

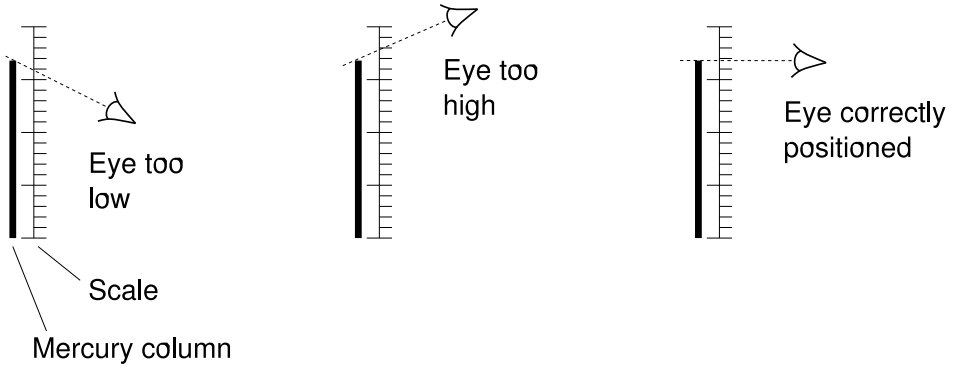

Figure A.3: Systematic errors introduced by parallax. From Les Kirkup, *Data Analysis with Excel: An Introduction for Physical Scientists*, Cambridge (2002).

A.4 Statistical Uncertainties in Measured Quantities

When we record data in the laboratory, our task is to provide a quantitative estimate of the uncertainties in our measurements. This task divides into two parts: first, we estimate the errors on directly measured quantities; second, we use these to calculate the resulting errors on derived quantities.

Basic Statistical Concepts

For single measurements, our only guide to evaluating the error is our knowledge of the experimental setup and our own capabilities. We already discussed using the resolution of an instrument to roughly estimate the uncertainty on a single measurement. This is unlikely to be very accurate, but it is still a reasonable guide to the precision of the final result.

Repeated measurements provide a more satisfactory estimate, because we can use the spread in the observed values to derive the characteristics of the underlying random error. Suppose we make N independent measurements of some quantity x. The average or mean of the N samples is the sum of the measured values divided by the total number N:

$$\boxed{\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i}$$

As mentioned earlier, the mean of a sample is an estimate of the "true" value of x. The uncertainty in this estimate is given by the standard deviation, also known as the root-mean-square (RMS) deviation s:

$$\boxed{s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}}$$

The quantity s is the best estimate of the error on an individual measurement x_i. However, our best estimate of the true value is the mean $\bar{x}$, not any individual measurement. The error on this is called the standard error of the mean:

$$\boxed{\bar{x} \pm \frac{s}{\sqrt{N}}}$$

That is, given a data sample $\{x_i\}$, we expect that the mean $\bar{x}$ — our estimate of the true value of x — should usually fall within the range given by $\pm s/\sqrt{N}$. Note that this range decreases in size as the sample size N goes up. As we increase the number of measurements, our estimate of the mean and the precision of our result improves.

The Gaussian Distribution

It is often convenient to assume that the distribution of measurements about the true value is given by a bell-shaped curved called the Gaussian probability distribution:

$$P(x)\ dx = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\ dx$$

In this expression, μ is the true value (or mean, or "expectation value") of the quantity we are trying to measure; σ, the standard deviation, is the typical spread expected for measurements of μ; and $P(x)\ dx$ is the probability that a given measurement will fall into the interval between x and $x + dx$. In the language of statistics, we say that x is a Gaussian random variable.

For certain types of error, it can be proven mathematically that this is the appropriate formula. In such cases, μ and σ can be predicted before performing any measurements. As an example, consider again the experiment in which a student flips a coin to determine its fairness. A well-known result from elementary statistics tells us that if the coin has a probability p of landing heads up[5], then for n tosses we expect

$$\begin{aligned} \mu &= np \\ \sigma &= \sqrt{np(1-p)} \end{aligned}$$

Look at the right panel of Fig. A.1. As the student increases the number of trials, the distribution of the head count starts to look more and more like a bell-shaped curve. In the limit as $N \to \infty$, this distribution will become a perfect Gaussian curve.

In practice, we do not measure μ and σ, because no one has time to let $N \to \infty$. Rather, we have to use a finite data sample to *estimate* their values. You probably already guessed that the best estimator for μ corresponds to the arithmetic mean $\bar{x}$, and the estimator for σ is the standard deviation s. Note that while s refers to the standard deviation of a finite data set and σ refers to the theoretical

[5] For a fair coin, $p = 0.5$.

spread in the data, the symbols σ and s tend to be used interchangeably. In a report, you would record your observation of μ as

$$\mu \approx \bar{x} \pm \sigma_{\bar{x}} = \bar{x} \pm \frac{\sigma}{\sqrt{N}} \approx \bar{x} \pm \frac{s}{\sqrt{N}}$$

For example, note how the estimate of μ improves in the coin toss experiment when the student increases N from 100 to 10,000:

$$\begin{aligned} \bar{x}_{100} &= 24.79 \pm \frac{3.41}{\sqrt{100}} = 24.79 \pm 0.34 \\ \bar{x}_{10000} &= 25.040 \pm \frac{3.572}{\sqrt{10000}} = 25.040 \pm 0.036 \end{aligned}$$

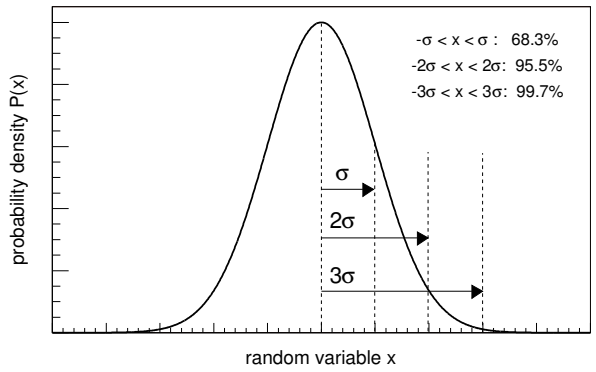

Figure A.4: Graphical representation of the Gaussian distribution. About 68.3% of the measurements of x should fall within one σ of μ; 95.5% should fall within 2σ of μ; and 99.7% should fall within 3σ of μ.

A.5 Errors on Derived Quantities

In most experiments, the desired quantity is not measured directly, but is calculated from other quantities which are measured. In this case we need to know how to deduce the error on the calculated result from the estimated errors on the measured values.

Error on $f(x)$ Due to an Error in x

Suppose we have measured x to be $X \pm \Delta X$. The quantity we want to determine is some function $f(x)$, and $f(X) = F$. What is the error on F, ΔF, corresponding to ΔX?

Common sense does work; we can evaluate $f(X + \Delta X)$ and $f(X - \Delta X)$ to get $F \pm \Delta F$. However, it is often convenient to use the fact that

$$\frac{df}{dx} = \lim_{\Delta X \to 0} \frac{\Delta F}{\Delta X}$$

If ΔX is small, this gives

$$\Delta F \approx \left| \frac{df}{dx} \right|_{x=X} \Delta X,$$

or in a more standard notation,

$$\boxed{\sigma_f = \left| \frac{df}{dx} \right| \sigma_x}$$

We take the absolute value because (by convention) we choose ΔF and ΔX (σ_f and σ_x) to be positive.

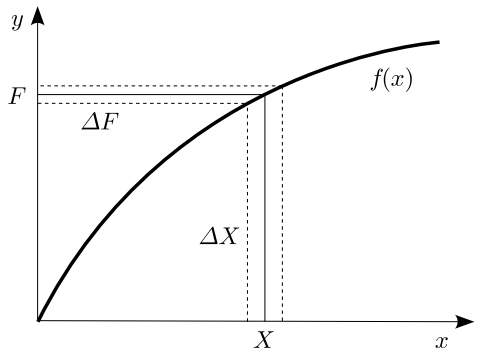

Figure A.5: Relationship between uncertainty of a function and slope.

Examples

Consider a measured quantity x with uncertainty σ_x. The resulting uncertainty in $f(x)$ is

$f(x)$	σ_f
x^2	$2x\sigma_x$
x^n	$nx^{n-1}\sigma_x$
$\sin x$	$(\cos x)\sigma_x$
$\ln x$	$\frac{1}{x}\sigma_x$

NOTE: when propagating uncertainties in angles, the angles must always be expressed in radians. Also note that the formula for the error on a logarithm applies only to natural logarithms — the derivative of $\log_{10} x$ is not $1/x$.

Functions of More than One Variable

Suppose we measure two quantities x and y, and the value that we need is a function of both, $f(x, y)$. What is the uncertainty in f given errors in x and y? If x and y are independent, and the uncertainties are relatively small, then we add the errors *in quadrature* [6]:

$$\boxed{\sigma_f^2 = \left(\frac{\partial f}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial f}{\partial y}\sigma_y\right)^2}$$

As in the single variable case, all of the partial derivatives are evaluated at the measured values X and Y. Note that this equation is easily extended to the case where f is a function of N variables:

$$\sigma_f^2 = \sum_{i=1}^{N}\left(\frac{\partial f}{\partial x_i}\sigma_{x_i}\right)^2$$

Examples

The following formulas are derived from the expression above:

$f(x)$	σ_f^2
$x + y$	$\sigma_x^2 + \sigma_y^2$
xy	$(y\sigma_x)^2 + (x\sigma_y)^2$
x/y	$(\frac{\sigma_x}{y})^2 + (\frac{x\sigma_y}{y^2})^2$

A.6 Combining Data: Fits and Averages

We have already seen how repeated measurements of a quantity can be averaged to obtain an improved estimate of its true value. This is a simple case of combining data. We shall often meet more complex cases: for example, we want to combine two measurements 1.5 ± 0.2 and 1.70 ± 0.05. It is intuitively obvious that the true value is likely to be between 1.5 and 1.70, but probably closer to the latter (the more precise value).

The Weighted Mean

Suppose you have measured the same quantity using several different techniques, which naturally give different errors. How do you combine the results to get the best possible estimate of the true value?

Clearly, you want to somehow weight your data so that the more precise values have more influence on the final answer. For Gaussian variables with independent errors, the appropriate weight is $1/\sigma_i^2$. In

[6] For a derivation of this expression, see Stuart L. Meyer, *Data Analysis for Scientists and Engineers*, Wiley (1975).

other words, we define the weighted mean by

$$\boxed{\bar{x}_w = \frac{\sum_i x_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}}$$

and the standard error of the weighted mean by

$$\boxed{\sigma_{\bar{x}_w} = \left(\frac{1}{\sum_i \frac{1}{\sigma_i^2}}\right)^{\frac{1}{2}}}$$

Both of these expressions reduce to the usual values when the $\sigma_i = \sigma$ (the errors are all the same):

$$\bar{x}_w = \frac{\sum_i x_i/\sigma^2}{\sum_i 1/\sigma^2} = \frac{1/\sigma^2 \sum_i x_i}{1/\sigma^2 \sum_i (1)} = \frac{\sum_i x_i}{N} = \bar{x}$$

and

$$\sigma_{\bar{x}_w} = \left(\frac{1}{\sum_i \frac{1}{\sigma^2}}\right)^{1/2} = \left(\frac{1}{\frac{1}{\sigma^2}\sum_i (1)}\right)^{1/2} = \left(\frac{\sigma^2}{\sum_i (1)}\right)^{1/2} = \left(\frac{\sigma^2}{N}\right)^{1/2} = \frac{\sigma}{\sqrt{N}} = \sigma_{\bar{x}}$$

In the weighted average, data points with large uncertainties are guaranteed to contribute almost nothing to the overall mean. Provided that N is reasonably large, the weighted and unweighted means should be roughly the same. If they give drastically different values, it is likely that your error estimates are off.

The Unweighted Linear Least Squares Fit

A common situation is the case where two variables or suitable functions of two variables are linearly related:

$$y = ax + b,$$

where a and b are unknown constants. Typically, the data consist of N pairs of observations (x_i, y_i) and the desired physical quantity is a or b or both. The problem is to derive the values of a and b which best fit the observations, and the corresponding uncertainties σ_a and σ_b.

To find the best line through x-y data, we need to decide on a measure of the goodness of fit of the line to the data. Fig. A.6 graphically depicts what we want to do. At each point (x_i, y_i), there is a fitted value of y $\hat{y}_i = ax_i + b$. The best fit line should somehow minimize the differences $\Delta y_i = y_i - \hat{y}_i$ between the fitted values and the data points, known as the *residuals* of the fit.

The standard best-fit algorithm used by most statistics packages is the Method of Least Squares, which involves minimizing the quantity

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2},$$

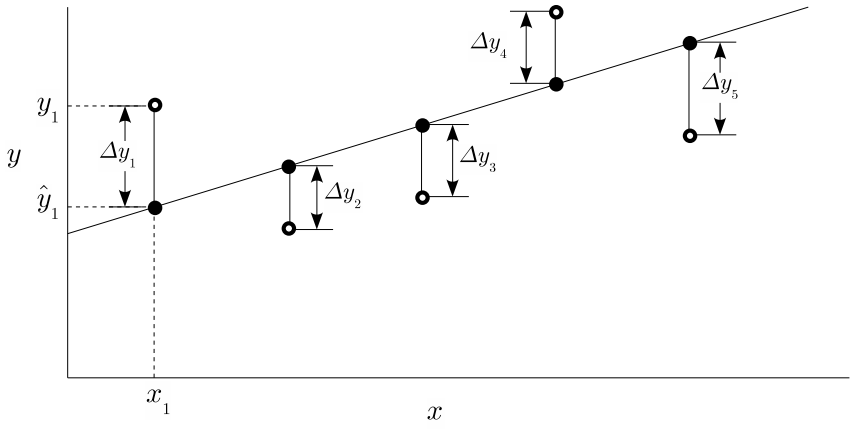

Figure A.6: x-y plot showing residuals. From Kirkup (2002).

where σ_i is the uncertainty in the data point y_i. The expression χ^2 (chi-square) is the weighted sum of squares of the residuals. For simplicity, let's assume for now that $\sigma_i = \sigma$, i.e., all the errors are the same. To find the best fit line, we write χ^2 in terms of the fit parameters a and b and differentiate with respect to a and b:

$$\begin{aligned}
\chi^2 &= \sum_i \left(\frac{y_i - ax_i - b}{\sigma} \right)^2 \\
\frac{\partial \chi^2}{\partial a} &= 0 = -\frac{2}{\sigma^2} \sum x_i (y_i - ax_i - b) \\
\frac{\partial \chi^2}{\partial b} &= 0 = -\frac{2}{\sigma^2} \sum (y_i - ax_i - b)
\end{aligned}$$

Manipulating these equations to solve for a and b, we find

$$\begin{aligned}
a &= \frac{N \sum x_i y_i - \sum x_i \sum y_i}{D} \\
b &= \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{D}
\end{aligned}$$

where

$$D = N \sum x_i^2 - \left(\sum x_i \right)^2$$

The uncertainties in the fit parameters are tedious to derive; if all of the uncertainties are equal ($\sigma_i = \sigma$), you can show that

$$\begin{aligned}
\sigma_a &= \sigma \left(\frac{N}{D} \right)^{1/2} \\
\sigma_b &= \sigma \left(\frac{\sum x_i^2}{D} \right)^{1/2}
\end{aligned}$$

These expressions look intimidating, but take comfort in two things: they are actually rather easy to calculate using a spreadsheet; and most statistics packages contain functions that automatically estimate a, b, σ_a, and σ_b for you[7].

The Weighted Linear Least Squares Fit

On a few rare occasions (to be discussed momentarily), you will want to perform a weighted least squares fit to your data that accounts for different σ_i. In this case, the expressions for the fit parameters and their standard errors change slightly:

$$\begin{aligned} a_w &= \frac{\sum \frac{1}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2}}{E} \\ b_w &= \frac{\sum \frac{x_i^2}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2}}{E} \end{aligned}$$

and

$$\begin{aligned} \sigma_{a,w} &= \left(\frac{\sum 1/\sigma_i^2}{E} \right)^{1/2} \\ \sigma_{b,w} &= \left(\frac{\sum x_i^2/\sigma_i^2}{E} \right)^{1/2} \end{aligned}$$

where the denominator E is given by

$$E = \sum \frac{1}{\sigma_i^2} \sum \frac{x_i^2}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2} \right)^2$$

The bad news is that few statistics packages automate these functions. The good news is that you will need to use them rarely, if ever.

Checking the Fit with the Residuals

Once you have completed the fit, a very revealing technique to test its goodness is to examine the distribution of residuals, Δy. Specifically, one usually plots the residual $\Delta y = y - \hat{y}$ on the vertical axis against x. If the fit is good and only random errors exist to obscure the true value of y at any x, then residuals should be scattered randomly about the $\Delta y = 0$ axis.

If we can discern a pattern in the residuals, then one or more factors may be affecting the data:

1. We have chosen the wrong equation to fit to the data, or a systematic error is present.

2. We should be using a weighted fit, because the uncertainties in the y values are growing or shrinking as a function of x.

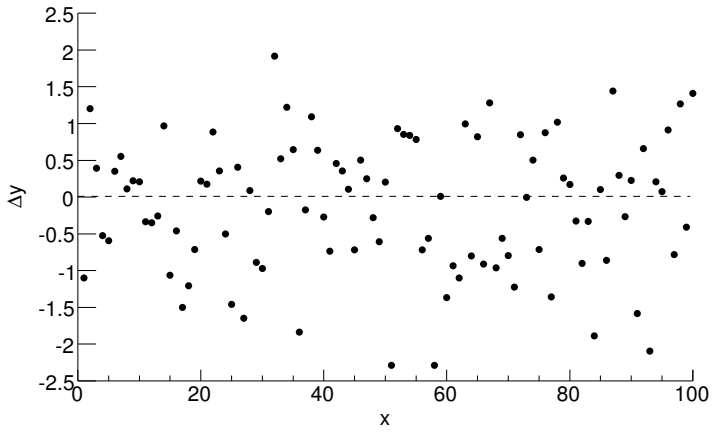

Figure A.7: Ideal distribution of residuals, with no discernable pattern. From Kirkup (2002).

Typical plots of residuals which illustrate patterns that can emerge are shown in Figs. A.7 to A.9.

Note that if the errors on y are Gaussian, you should expect to see about $2/3$ of the residuals (68%) lying on the line $\Delta y = 0$ within their 1σ error bars. If too many points lie off the line, you have underestimated your errors, or there is a systematic effect. If too few lie off the line, you have overestimated your errors, or the errors are not really all independent. If the errors are truly Gaussian, it is highly unlikely (1 in 2000 chance) that all the Δy_i should be within 1σ of zero. If this happens, there must be something wrong!

A.7 Quoting Errors

Once we have analyzed our data, we must present and correctly interpret our results. The aim here is to provide the information in a manner that is easily grasped by a reader who is not necessarily familiar with the details of the experiment.

Quoting the Numbers

Error estimates should not be quoted to unjustified precision. Usually one or two significant figures is enough: e.g., 14.2 ± 0.6 m or 14.18 ± 0.17 m are acceptable. If you are going to use these numbers in subsequent calculations, beware of round-off errors that result from throwing away too many significant figures too early. (This is typically not a problem if you do all your calculations using a spreadsheet.)

Once you have decided how many significant figures to use in your uncertainty, round the main value to the same number of decimal places. If using scientific notation, always use the same exponent for

[7] Example: the `LINEST` function in Excel and OpenOffice.

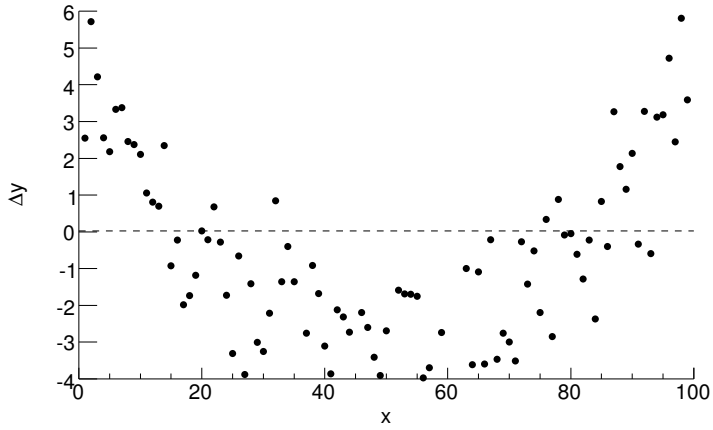

Figure A.8: Residuals revealing an incorrect equation fitted to the data. From Kirkup (2002).

the main figure and its uncertainty:

$h = (6.62606876 \pm 0.00000052) \times 10^{-34}$ J s, not $h = 6.62606876 \times 10^{-34} \pm 5.2 \times 10^{-41}$ J s

When reporting numbers, favor scientific notation or metric prefixes over presenting numbers with long strings of zeros. For example,

$$25 \pm 1 \ \mu\text{m} \qquad \text{and} \qquad 2.5 \pm 0.1 \times 10^{-5} \text{ m}$$

are much easier to read and understand than

$$0.000025 \pm 0.000001 \text{ m}$$

In summary:

- Unless instructed otherwise, do not quote errors to more than two significant figures.
- Quote values such that they have the same number of decimal places as the error.
- When using scientific notation, quote the value and error with the same exponent.
- When reporting a numerical value, never forget the error or the units (unless the value is dimensionless).

What to Say About Errors in Primary Quantities

As you measure quantities, try to work out errors as you go along. Do not wait until you have a final answer and then think about the uncertainties. This will often help improve your results. If, during the

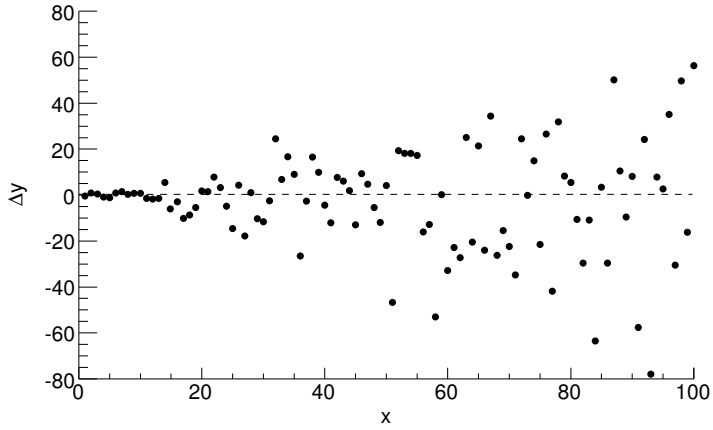

Figure A.9: Pattern in residuals suggesting weighted fit should be used. used. From Kirkup (2002).

course of the experiment, you find that some effect is contributing to the statistical errors in your data, you can take steps to mitigate the effect and improve your results.

When discussing error estimates in your lab notebook or final report, you should be as specific as possible. That is,

- Carefully state the source or sources of uncertainty, how you estimated the quoted values, and whether the resulting error is statistical or systematic.
- If the quoted value is a mean of N observations, state N, and make clear that the quoted error is the standard deviation (i.e., the error on an individual measurement) or the standard error of the mean.
- Combine different sources of statistical error by adding in quadrature: $\sigma_{\text{tot}}^2 = \sigma_1^2 + \sigma_2^2 + \ldots$.
- If you are able to quantify systematic errors, report them separately, e.g.,

$$1.57 \pm 0.09(\text{stat}) \pm 0.05(\text{sys}) \text{ kg}$$

What to Say About Errors in Derived Quantities

When reporting your results to the outside world, you do not need to derive or quote the basic error formulas used to propagate your uncertainties. Your intended readers will be trained scientists who expect you to do things right. While you should not go through lengthy derivations in your main report, in this course you should attach scratch work and derivations to the back of your write-up each week.

In the report itself, you need to:

1. Make sure that you have explained how the derived quantity is related to the measured quantities (i.e., give the equation).

2. Carefully explain any systematic errors and how you have dealt with them.

How to Compare Experimental and Predicted/Accepted Values

In most lab experiments, the result you obtain can be directly compared with a theoretical expectation or a "book value." In making a comparison, it is essential to take the errors into account. The basic idea is to subtract the true value from your result, calculate the error on the difference, and observe how far the result is from zero. You can express this difference in terms of the error using the expression

$$\text{difference} = \frac{|\text{expt} - \text{true}|}{\sigma}$$

For example, suppose you measure the landing position of a projectile. You expect it to land at $x = 0$, but you find that it actually lands at $x = 1.3$ m. Deciding whether or not this is a reasonable result depends on the uncertainty of the measurement. Roughly speaking,

- If the difference between expectations and measurments is within 1σ of zero, e.g., 1.3 ± 1.4 m, the results are in *good agreement*. For Gaussian data, 68% (or two-thirds) of the random fluctuations about the true value should occur in this region.

- If the difference is between 1σ and 2σ, e.g., 1.3 ± 0.8 m, the results are *consistent*. There is about one chance in three ($\sim 100\% - 68\%$) that the discrepancy is due to a random fluctuation.

- If the difference is between 2σ and 3σ, e.g., 1.3 ± 0.5 m, there is about one chance in 20 that this is a random fluctuation away from the true value. A fluctuation this large probably (but not definitely) indicates a statistically significant fluctuation.

- If the difference is bigger than 3σ, e.g., 1.3 ± 0.3 m, then your discrepancy is almost certainly statistically significant. The chance probability of a random fluctuation this big is about 1%.

If you find that the difference between your result and the accepted value is statistically significant, do not ignore it. For example, it is not acceptable to say, "The difference between my result of 1094.2 ± 0.3 and the accepted value of 1095.5 is 1.3 ± 0.3, but this is less than 0.1%, so my value is very close." Remember, for a result to be truly accurate, your answer must agree with the accepted value within the precision of your measurements. In this particular case, the experimental result is 4.3σ away from the expected value, a large discrepancy. Either the error estimate is significantly off, or there is a systemic effect at work, or you have discovered some new physics!

Note that it is not acceptable to simply say, "The discrepancy is due to inaccurate measurements." The inaccuracy of your measurements is already known from your error estimate; your job is to determine its source. Acceptable explanations for a statistically significant discrepancy might include:

- Different conditions. For example, the book value is quoted at 0° C, but you worked at room temperature.
- Dubious approximations. For instance, a theoretical prediction assumes a small angle approximation, but you carried out your measurements at 10° (0.175 rad).
- Possible systematic effects, such as a device that you could not calibrate accurately. **NOTE:** it is not sufficient for you to just say, "Our instrument was not calibrated." You need to estimate how large the miscalibration would have to be to cause the observed discrepancy. In this way, you can determine if a suggested systematic error is reasonable.

Finally, when quoting values from a text, always specify the source. Do not write in your report, "The book value of Planck's constant is..." Instead, you should include the source in the report, either as a footnote, as a cited reference at the end, or as an inline reference: "Planck's constant is $(6.62606876 \pm 0.00000052) \times 10^{-34}$ J s (source: PDG, Phys. Rev. **D66** (2002) 010001)."