Errors and Analysis of Data

The sole purpose of performing an error analysis is to show to what extent your results can be believed. There are three different aspects to reporting errors: (1) keeping track of measurement uncertainties; (2) discussing sources of random and systematic error in your experiment; and (3) error propogation in the analysis of the data. The first two are by far the most important. The third has a rough-and-ready aspect as well as a rigorously mathematical aspect. Usually it **won't** be necessary to do a thorough error propogation. But it is important, when combining measured quantities in equations, to know **which** errors are going to have the **greatest** effect on the accuracy of the final result and have some idea to the latter, if only an estimate.

Measurement Uncertainties

The **precision** of any measuring instrument is limited, and when recording data you have an obligation to determine or at least estimate and report those limits. This prevents incorrect conclusions from being drawn from the measurements.

For example, suppose you are taking a voltage reading from the meter face shown in Figure 1. The small ticks are spaced 2 millivolts (**mV**) apart. Suppose that you measure a voltage of **60 mV**. You would be able to tell easily if the reading were 64 mV instead; on the other hand, if it changed from 60.0 to 60.2, you would probably not be able to see the difference. So your uncertainty in the measurement is less than **4 mV**, but more than **0.2 mV**. It's your job to make a *common-sense* estimate of a reasonable *maximum probable error* of
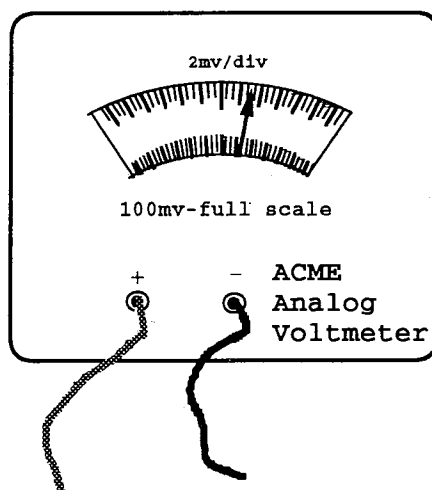
Figure 1: A voltmeter you might use to make a measurement.

your result. You might decide that the voltage is known to an accuracy of 1 mV: in other words, you think a change of 1 mV or more would be noticeable. For rulers, dials, and other measuring devices with tick marks, the error will probably be comparable to the **smallest** marked division. For a digital instrument, the error will probably be in the **last** digit of the display and typically ±1 unit. Other devices will usually have a certain percentage accuracy associated with them, for example ±1%, ±10%, ±20% and so on. Consult your TA if you cannot estimate this yourself.

Once you have decided upon the likely accuracy of the meter, you can report the appropriate uncertainty for your voltage measurements. When you report a result of (60±1) mV, you are saying you think that the voltage is probably between 59 and 61 mV. You are also saying that 58 or 62 mV would not be out of the question, but you are quite sure that the voltage is less than 64 mV and more than 56 mV.

There are a few standard ways of reporting such errors. If you read many measurements from one device, you might write a note such as *error in voltage* = **±1** mV in your lab notebook and formal report.

When the errors vary from measurement to measurement, you might present them each with their own error intervals, i.e., (60 ± 1) mV. One *common* mistake is to write *more* digits in the measurement than the precision allows: for example, it makes no sense to report (60.05 ± 1) mV, since 0.05 mV is less than the unknown accuracy. When data are graphed, the errors should be *indicated on the graph*, either as error bars or in a comment on the graph (for example, if the error is very small compared to the scale of the plot or the size of the symbols).

Random and Systemic Error

Every numerical measurement you will make in the Physics laboratory will have an uncertainty (or error) associated with it. As in the case of the voltmeter reading described above, the error may be due to the finite resolution of your measuring device, but it is **not** limited to this. The measuring device may have an unknown offset, or it may have been calibrated incorrectly. Generally, errors are grouped into two categories, *random* and *systematic*. *Random* error comes from things in your experiment which are slightly different each time you make a measurement, in a way you cannot predict or control. *Systematic* error comes from your experimental setup and is the same whether you make the measurement once or one hundred times.

If you are reading from a meter with a quivering needle, or a digital readout where the last digit is flickering, these will cause random error in the measurement. Since the needle is just as likely in this case to read slightly high as slightly low, you can improve your data by taking many measurements and averaging. This is an example of random error.

On the other hand, the calibration of the meter might be wrong. It might consistently give a reading 2 mV *too high*. This is an example of systematic error. Repeating the measurements in this case will not make any difference; all measurements will still be too high, but you could in principle detect the problem and account for it. (For example, you could check the meter with a constant voltage source known to better than 2 mV.) This is an example of systematic error.

To further illustrate the difference between random and systematic errors consider an example from target practice (Figure 2). The random
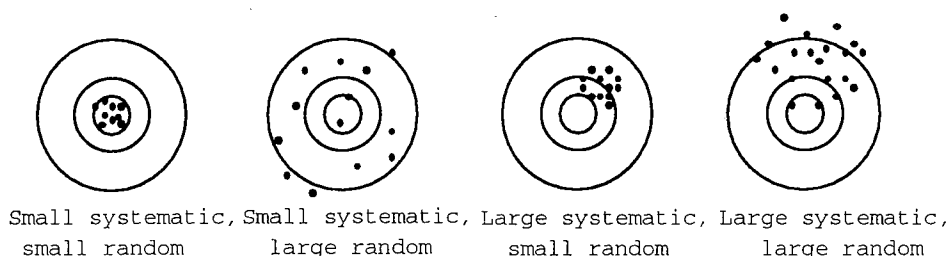
| Small systematic, | Small systematic, | Large systematic, | Large systematic, |
| small random | large random | small random | large random |

Figure 2: Target practice.

errors might be due to variations in the cartridges, jitter while aiming, etc. The systematic errors might be wind, misaligned sights, or a consistent bias in aiming. Note that random errors become smaller if the data are averaged over many tries or measurements, while systematic errors do not. In other words, we can decrease random errors by taking many measurements and averaging, but we must combat systematic errors in other ways.

When you report experimental results you will usually be comparing them to something — either an *accepted* value or, much more often, a value you measured in a different way to check for consistency. Since the values you are comparing will never (or very rarely) be *exactly* equal, it is important to know the uncertainty associated with a given measurement in order to correctly compare it to the expected value.

A discrepancy between a measured and expected value is signaled by the expected value falling outside the limits established on the measured value by its associated uncertainty. Because comparisons between expected and measured values allow us to draw meaningful results from experimental measurements it is essential to understand and characterize both the random and systematic uncertainties associated with a given measurement. Furthermore, experimental results are often derived from equations which make use of many separate measurements. In order to derive meaningful experimental results it is often necessary
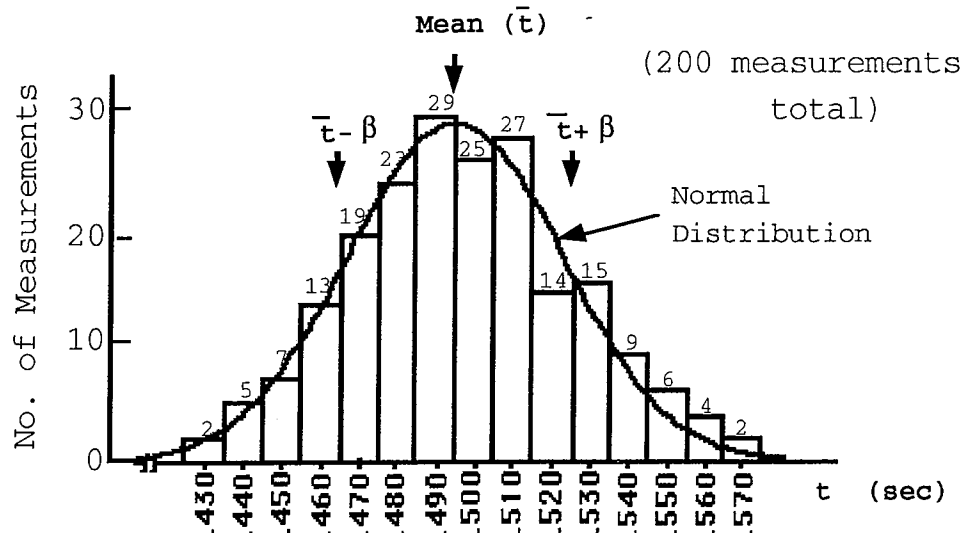
Figure 3: Gaussian distribution.

to propagate measurement uncertainties through these equations. This, along with other important aspects of uncertainty determination, are described in detail below.

The Normal Distribution, Mean, and Standard Deviation

Suppose we consider a measurement whose result can take on a continuous range of values. To be concrete let us imagine a very simple experiment. We want to measure the time it takes a ball to fall 1.00 meter. To get an accurate value we use a good stopwatch and repeat the measurement 200 times. Figure 3 shows the results of our hypothetical experiment in the form of a histogram.

The vertical height of each rectangle or *bin* gives the number of measurements that lie within the range of the bin. For example, there were 29 measurements with fall times between 0.485 and 0.495 sec.

The distribution in Figure 3 is somewhat *idealized*, but is typical

of what *real* data from a well-designed and executed experiment might look like. The most *notable* features are:

1. The values are clustered about a well-defined mean value which is close to the most probable value (the value of t where the distribution has its maximum). The *arithmetic* mean of the t values in **gaussian** is approx. 0.497 s.

2. Values which are far from the mean are *very unlikely*.

3. The distribution is reasonably *symmetric* about the mean. There is no *obvious* skewing toward the high or low side.

If we took many, many measurements and made the bins very fine, our histogram might begin to look like the smooth, bell-shaped curve. This curve is the limiting case in an ideal situation. It is referred to as the *normal* or *Gaussian* distribution. Measurement errors which follow this distribution are said to be normally distributed. The mathematical form of the normal distribution is really not very important because, in a given experiment, you cannot prove that the measurements will follow a normal distribution. Nevertheless, a distribution resembling the normal distribution is usually found, and it is usually assumed that a normal distribution is appropriate.

The bell-shaped curve and, to a good approximation, the histogram of Figure 3 can be characterized by two quantities, the **mean value** and the **width**. The *mean* $\bar{t}$ of the measured times is just the arithmetic average of the data,

$$\bar{t} = \frac{1}{n}[t_1 + t_2 + \ldots + t_N] \equiv \frac{1}{n}\sum_{i=1}^{N} t_i \tag{1}$$

Here $t_1, t_2, \ldots$ are the measured times, the symbol $\sum$ stands for a sum, and N is the number of measurements. (For the normal distribution, the mean is defined in terms of an integral analogous to Equation 1.) The width of the distribution can be defined in various ways – for example, the *full width at half maximum*, the mean deviation $\sum |\bar{t} - t_i|/N$ and the **root mean square (rms)** deviation $\sigma = [\sum(\bar{t} - t_i)^2/N]^{\frac{1}{2}}$. The rms deviation turns out to be the most common, and we shall accept

it as our definition. It is also called the **standard deviation** of the measurement. We shall generally refer to it as the standard deviation and use the symbol σ. Thus

$$\sigma \equiv [\frac{1}{N} \sum_{i=1}^{N} (\bar{t} - t_i)^2]^{\frac{1}{2}} \qquad (2)$$

where $\bar{t}$ is the mean from Equation 1.

If the measurements follow a normal distribution then ideally 68.3% of the measurements lie within $\pm 1\sigma$ from the mean. Thus, from Figure 3, which contains 200 measurements, we can estimate σ by counting off 68 measurements in either direction from the mean. This includes a band of width approx. ± 0.030, so the standard deviation per measurement is approx. 0.03. This band is indicated in Fig. 1 by the arrows labelled +s and s. Note that we could get a more precise value of σ by numerical calculation from Eq. 2, but the increase in precision of σ is insignificant. In other words, we shouldn't feel obliged to estimate σ to very high accuracy. In practice, it is safest to use the histogram method for estimating σ when possible because it gives a chance to judge whether the data look *normally* distributed. We might be tempted to discard a measurement which lies many standard deviations from the mean. (We shall not discuss the correctness of this procedure; the point is that it is often done.)

It is important to realize that σ is a measure of the *probable* uncertainty of **one** measurement i.e., if we make one measurement it has a 68% *probability* of being within 1σ of the mean value. The uncertainty in the mean is much smaller than σ because we have made many measurements. For N measurements the **standard deviation of the mean** is

$$\sigma_m = \sigma/\sqrt{N} \qquad (3)$$

This assumes that the measurements are *independent* and *uncorrelated*. In the example of measuring the fall time of a ball, if we started and stopped two clocks with the same switches the measurements of the two clocks would be strongly correlated; the amount of correlation would depend on how good the clocks were. (The better they are, the stronger the correlation.)

The result of a series of measurements of a quantity A and its error or uncertainty are usually written in the form $A \pm \sigma$. In the example above, the *mean* time was 0.497 sec and the standard deviation of the mean would be $0.03/\sqrt{200} \cong .002$, so we would write

$$\bar{t} = 0.497 \pm .002 \quad \text{sec} \tag{4}$$

Ideally this means that the "true" value of t has a 68.3% chance of lying between 0.495 and 0.499 sec. Two results are considered to be consistent with each other if they are within 1 or 2 standard deviations of each other. Obviously some judgement is required. The above discussion assumes that all of the measurements in an experiment are of equal intrinsic accuracy. If some measurements are better than others, the better ones should have a higher weight in computing the mean. The calculation of the weighted mean and probable errors in this kind of a situation is discussed in many references.

Experiments Whose Outcome Is an Integer: The Square Root Rule

Often the result of an experiment or measurement is an integerfor example, the number of mice out of an initial sample of 100 that die within ,one year or the number of radioactive nuclei out of a sample that decay in one second. The standard deviation of the number of such "events" (deaths, decays, or whatever) can be estimated by the **square root rule**. If N is the number of events, the *standard deviation in N* is

$$\sigma_m = \sqrt{N} \tag{5}$$

For this to be an accurate estimate, the following conditions must be satisfied. (The better they are satisfied, the better the estimate of σ.)

1. The number of events N must be *large*. [Some people might consider $N > 10$ to be large enough.]

2. The probability that any member of the initial sample dies or decays (or whatever) must be *small*. If, for example, we did an

experiment to see how many of 100 mice would die within 100 years, the answer would be 100 ± 0. The probability of death is 100%, surely not small. On the other hand, if we start out with 10^8 radioactive nuclei and they decay at the rate of 10^3 per sec, in a 10 sec "experiment" the number of nuclei which decay would be $N = 10000 \pm 100$. The *square root rule* should work very well because $N \gg 1$ and the probability of a given nucleus decaying is 10^{-4} during the experiment.

Error Propogation

For most experiments, you will not present just the raw data, but some quantity derived by combining the results in equations. The uncertainties in the measurements will all contribute to the uncertainty in the final result, and there is a specific way to handle this.

Even when you aren't going to do a full-fledged error propogation, there are some things you must keep in mind. First, give some thought to **how many digits** you write down when you record the data. It's a good habit to write down all the digits you can when you make the measurement, and **round off** later when you report the result with its uncertainty. This allows you to see what kind of fluctuations (random error) you have in your data, and also keeps roundoff error from building up and changing your results.

You may sometimes make measurements which have the same units and need to be added together, but which have different uncertainties. Remember that your precision is only as good as that of your **least accurate** measurement. For example: you measure a bunch of voltages with a digital voltmeter. For the small voltages, you use the smallest scale setting, which goes down to 0.001 V. For the larger voltages, you use a higher scale that goes down to 0.1 V. Your data look like this: 0.568, 0.450, 4.2, 3.8, with the error in the last reported digit of each number. Now suppose you need to sum them to find a total voltage. It would be incorrect to report the sum equal to 9.018, since you don't have that much accuracy in all of your measurements. You need to round to 0.1 V, since that is your largest uncertainty, and report 9.0 V as your final result.

So much for the rough-and-ready error analysis. You should now be able to answer the questions at the end of this appendix. Unfortunately, there is still more for you to learn before you can call yourself an expert and before you can successfully complete the analysis for some of the experiments, so read on. If you need or want to be completely rigorous in the error analysis, the following techniques must be used.

Notation: If the measured quantity is x, the symbol δx (delta x) will denote the experimental precision. For example, if the measured data is (60 ± 1) mV then $x = 60$ while $\delta x = 1$. Likewise if some number A is calculated from measured numbers, δA is the uncertainty in A. The following rules tell you how to find δA from the errors in the experimental quantities.

Rule 1: addition and subtraction. Suppose that you make several measurements, $x \pm \delta x$, $y \pm \delta y$, and $z \pm \delta z$, and the quantity A is given by

$$A = k_1 x + k_2 y + k_3 z$$

where the k's are all constants (positive or negative). Then

$$\delta A = \sqrt{(k_1 \delta x)^2 + (k_2 \delta y)^2 + (k_3 \delta z)^2}$$

and you report the result $A \pm \delta A$. (For more than three measured variables there would be additional terms in the square root sign.)

Rule 2: multiplication and division. Again let x, y, z be the measured variables, k constant, and A given by multiplication or division of these. Multiplication and division are treated the same, so A can be:

$$A = kxyz \quad \text{or} \quad A = kxy/z \quad \text{or} \quad A = k/xyz$$

or any similar variant. Then

$$\frac{\delta A}{A} = \sqrt{\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2 + \left(\frac{\delta z}{z}\right)^2}$$

Note that this expression has the same form as that for addition, but the fractional or percent errors are used instead of the absolute errors. That

is, $(\delta x/x)$ instead of δx is the important quantity. Hence, if $(\delta x/x)$ are $\pm 1\%$, and $(\delta z/z)$ is $\pm 5\%$ then it is mainly $(\delta z/z)$ that determines $\delta A/A$ and the others can be neglected. Also, then $(\delta A/A) \approx (\delta z/z) = \pm 5\%$.

Errors can be *approximated* since you only need one digit or so of accuracy. Therefore they usually need **not** be calculated to high precision, particularly just to get an estimate of the error and its propagation (as in the above example).

Rule 3: power function. When your measured variable x is taken to a power n, that is,

$$A = kx^n,$$

where k is constant, then

$$\delta A = A|n|\left(\frac{\delta x}{x}\right).$$

Note: n could be fractional since $\sqrt{x} = x^{\frac{1}{2}}$; also $\log_{10} x$ and $ln|x|$ correspond to $n < 1$. Here is an example to show you how a more complicated function can be handled in terms of these rules. The measured variables are x, y and z; say k and n are constants; and A is the calculated quantity. Other capital variables B, C, ... will be used as dummy variables for the algebra. Suppose your calculation is

$$A = k/x + yz^n.$$

This is of the form $A = B + C$ where we set $B = k/x$ and $C = yz^n$. So from **Rule 1** we have

$$\delta A \cdot = \sqrt{(\delta B)^2 + (\delta C)^2}.$$

Now δB and δC can be found separately.

$$B = k/x, \quad \text{so} \quad \delta B = B\sqrt{\left(\frac{\delta x}{x}\right)^2} = (k/x)(\delta x/x) = k\delta x/x^2.$$

Now,

$$C = yz^n;$$

to deal with this term, set $D = z^n$, so $C = yD$. Now you can find δC from the multiplication rule:

$$\delta C = C\sqrt{\left(\frac{\delta y}{y}\right)^2 + \left(\frac{\delta D}{D}\right)^2}.$$

Now of course we need to find δD:

$$\delta D = D|n|\frac{\delta z}{z} = (yz^n)|n|\frac{\delta z}{z} = |n|yz^{n-1}\delta z.$$

Now we can go back and plug all these numbers in to find δA. It's actually a lot of trouble. A much better idea is to look at the three rules again and note the following:

1. When quantities are added, it is the total error that matters.

2. When quantities are multiplied, it is the fractional or percentage error $(\delta x/x)$ which matters.

3. When a quantity comes into an equation raised to a certain power $(+ \text{ or } -)$, it is the power times the percentage error $(n\delta x/x)$ which is significant.

4. $\log x$ and $\ln x$ terms correspond to $n \ll 1$ so uncertainties in them are usually small i.e., $\log(x \pm \delta) \approx \log(x)$ for δ small.

5. Since high accuracy is not needed, you can make approximations to simplify calculations.

So look at the expression for A just one more time:

$$A = \frac{k}{x} + yz^n.$$

Looking at this, we can simply say the following. If $|n|$ is greater than one, then the percentage error in z will be more important than the percentage error in y. Also, if the error in one term of A is much **greater** than the error in the other term, you will only have to worry about the term with the *larger error*. If they are about the same, then we should use the square root rule to calculate their total contribution

to δA. but otherwise you can approximate and keep only the **dominant** term. Learning to make this kind of judgement will often save us the trouble of going through a long painstaking analysis, and is a **useful** intuition to develop. During a lab, you want to concentrate your efforts on the **key** items needing accurate measurement.

References on Errors and Treatment of Data

There are two handy paperbacks: *Theory of Errors* by Yardley Beers (Addison-Wesley Pub. Co.); and *Statistical Treatment of Data* by Hugh D. Young,(Mc Graw-Hill Pub. Co.). On the lighter side, there is *How to Lie with Statistics* by Darrel Huff (Norton Pub. Co.). A readable but more advanced book is P.R. Bevington, *Data Reduction and Error Analysis*. Another useful book is J.R. Taylor, *An Introduction to Error Analysis*. These books are available in the Science Library which in the Undergraduate Library.

Questions

1. Twenty measurements of the time for a ball to fall 1.0 meter are made with the results: .461, .512, .496, .470, .537, .501, .495, .449, .468, .444, .472, .483, .441, .467, .463, .557, .482, .430, .477, .450 sec. Histogram the data. Calculate the mean and estimate the standard deviation. Do these data seem to follow a normal (Gaussian) distribution? Discuss. Are these data consistent with the data in Figure 3?

2. From the data in Figure 3 calculate the acceleration of gravity from $g = 2y/t^2$ if the ball fell a distance $y = (1.102 \pm .004)m$. Estimate the standard deviation in the "measured" value of g. Compare it to the accepted value of $g = (9.81 \pm .01)m/s^2$. Is the measured g consistent with this? Are the errors in the measurement purely random? Give three examples of systematic errors

that might occur in this measurement and discuss how they might be minimized.

3. You make 10 measurements of the number of radioactive nuclei that decay in 1 sec. The results are: 60, 69, 58, 65, 71, 60, 62, 74, 68, 66. Are these data consistent? Calculate the mean number of decays per sec. and estimate its standard deviation. Would you have gotten a more or less accurate value if you just took one measurement for 10 sec and divided by 10? Discuss advantages and disadvantages of either strategy.

4. If you made a histogram of the grade point averages of all U. of M. students, would you expect it to follow a normal distribution? What about the distribution of the number of spectators per home game in the football stadium for the past 5 years? Suppose you randomly select students to make 300 groups, each containing 100 students. You then calculate the mean height for students in each of the 300 groups. Would you expect the 300 mean heights to follow a normal distribution? Discuss each example briefly and , sketch what you might expect the histogram to look like.

5. Joe believes a pair of dice is "loaded." To check this possibility, he takes one of them, throws it 100 times, and tallies how many times each number comes up. He finds 10 ones, 17 twos, 20 threes, 24 fours, 13 fives and 16 sixes. Can he conclude the die is loaded? Discuss.

6. A study was made of 85000 Vietnam veterans, who had been exposed to chemical defoliants, to determine whether they showed abnormally high rates for a variety of cancers. Eleven were found to have soft tissue sarcomas (a rare form of cancer) compared to a predicted number of 16 based on rates in similar groups of American males. Twenty compared to a predicted 15 had developed lymphomas. Can these differences between the observed and predicted rates be considered significant? Explain.