APPENDIX D
ACCURACY OF MEASUREMENTS AND TREATMENT OF EXPERIMENTAL UNCERTAINTY

"A measurement whose accuracy is unknown has no use whatever. It is therefore necessary to know how to estimate the reliability of experimental data and how to convey this information to others." —E. Bright Wilson, Jr., *An Introduction to Scientific Research*

Our mental picture of a physical quantity is that there exists some unchanging, underlying value. It is through measurements we try to find this value. Experience has shown that the results of measurements deviate from these "true" values. The purpose of this Appendix is to address how to use measurements to best estimate the "true" values and how to estimate how close the measured value is likely to be to the "true" value. Our understanding of experimental uncertainty (i.e., errors) is based on the mathematical theory of probability and statistics, so the Appendix also includes some ideas from this subject. This Appendix also discusses the notation that scientists and engineers use to express the results of such measurements.

ACCURACY AND PRECISION

According to many dictionaries, "accuracy" and "precision" are synonyms. To scientists, however, they refer to two distinct (yet closely related) concepts. When we say that a measurement is "accurate", we mean that it is very near to the "true" value. When we say that a measurement is "precise", we mean that it is very reproducible. [Of course, we want to make accurate AND precise measurements.] Associated with each of these concepts is a type of error.

Systematic errors are due to problems with the technique or measuring instrument. For example, many of the rulers found in labs have worn ends. One can make very precise (reproducible) measurements that are quite inaccurate (far from the true value).

Random errors are caused by fluctuations in the very quantities that we are measuring. You could have a well calibrated pressure gauge, but if the pressure is fluctuating, your reading of the gauge, while perhaps accurate, would be imprecise (not very reproducible).

Through careful design and attention to detail, we can usually eliminate (or correct for) systematic errors. Using the worn ruler example above, we could either replace the ruler or we could carefully determine the "zero offset" and simply add it to our recorded measurements.

Random errors, on the other hand, are less easily eliminated or corrected. We usually have to rely upon the mathematical tools of *probability* and *statistics* to help us determine the "true" value that we seek. Using the fluctuating gauge example above, we could make a series of independent measurements of the pressure and take their average as our best estimate of the true value.

Measurements of physical quantities are expressed in numbers. The numbers we record are called *data*, and numbers we compute from our data are called *statistics*[*].

[*] A statistic is by definition a number we can compute from a set of data. An example is the av-

PROBABILITY

Scientists base their treatment of random errors on the theory of probability. We do not have room here for a lengthy survey of this fundamental subject, but can only touch on some highlights. Probability concerns random events (such as the measurements described above). To some events we can assign a theoretical, or *a priori*, probability. For instance, the probability of a "perfect" coin landing heads or tails is ½ for each of the two possible outcomes; the *a priori* probability of a "perfect" die* falling with a particular one of its six sides uppermost is $\frac{1}{6}$.

These examples illustrate four basic principles about probability:

1. The possible outcomes have to be mutually exclusive. If a coin lands heads, it does not land tails, and *vice versa*.

2. The list of outcomes has to exhaust all possibilities. In the example of the coin we implicitly assumed that the coin neither landed on its edge, nor could be evaporated by a lightning bolt while in the air, or any other improbable, but not impossible, potential outcome. (And ditto for the die.)

3. Probabilities are always numbers between zero and one, inclusive. A probability of one means the outcome always happens, while a probability of zero means the outcome never happens. There is no meaning to a probability larger than one, since something cannot happen more often than every possible time. Similarly, there is no meaning to a probability smaller than zero, since something cannot happen less often than never.

4. When all possible outcomes are included, the sum of the probabilities of each exclusive outcome is one. That is, the probability that *something* happens is one. So if we flip a fair coin, the probability that it lands heads *or* tails is $\frac{1}{2} + \frac{1}{2} = 1$. If we toss a fair die, the probability that it lands with 1, 2, 3, 4, 5, or 6 spots showing is $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$.

The mapping of a probability to each possible outcome is called a *probability distribution*. Just as our mental picture of there being a "true" value that we can only estimate, we also envision a "true" probability distribution that we can, again, only estimate through observation. Using the coin flip example to illustrate, if we flip the coin four times, we should not be too surprised to get heads only once. Our estimate of the probability distribution would then be $\frac{1}{4}$ for heads and $\frac{3}{4}$ for tails. We do expect that our estimate would improve as the number of flips† gets "large". In fact, it is only in the limit of an infinite number of flips that we can expect to approach the theoretical, "true" probability distribution.

The probability distributions we've discussed so far have been for discrete possible outcomes (coin flips and die tosses). When we measure quantities that are not necessarily discrete (such as pressure read from an analog gauge), our probability distributions become more correctly termed

---

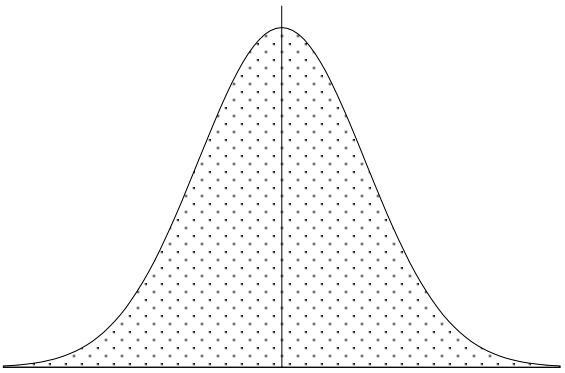
erage, or *mean*. Another is the *variance*, which we shall define below.

* …one of a pair of dice.

† Each flip is, in the language of statistics, called a *trial*. A scientist or engineer would probably say that it is a *measurement* or *observation*.

probability density function (although you often see "probability distribution" used indiscriminately). The defining property of a probability distribution is that its sum (*integral*) over a range of possible measured values tells us the probability of a measurement yielding a value within the range.

The most common probability distribution encountered in the lab is the *Gaussian* distribution. The Gaussian distribution is also known as the *normal* distribution. You may have heard it called the *bell* curve (because it is shaped somewhat like a fancy bell) when applied to grade distributions.



Gaussian Distribution

The mathematical form of the Gaussian distribution is:

$$P_G(d) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-d^2/2\sigma^2} \qquad (1)$$

The Gaussian distribution is ubiquitous because it is the end result you get if you have a number of processes, each with their own probability distribution, that "mix together" to yield a final result. We will come back to probability distributions after we've discussed some statistics.

STATISTICS

Perhaps the single most important statistic is the *mean* or *average*. Often we will use a "bar" over a variable (e.g., $\overline{x}$) or "angle brackets" (e.g., $\langle x \rangle$) to indicate that it is an average. So, if we have N measurements x_i (i.e., x_1, x_2, ..., x_N), the average is given by:

$$\overline{x} = \langle x \rangle = (x_1 + x_2 + ... + x_N)/N = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (2)$$

In the lab, the average of a set of measurements is usually our best estimate of the "true" value*:

$$\overline{x} \approx x \qquad (3)$$

In general, a given measurement will differ from the "true" value by some amount. That amount is called a *deviation*. Denoting a deviation by d, we then get:

$$d_i = x - x_i \approx \overline{x} - x_i \qquad (4)$$

Clearly, the average deviation is zero (to see this, take the average of both sides).

Another notable statistic is the *variance*, defined as the mean square deviation:

$$\mathrm{var}(x) \equiv (d_1^2 + d_2^2 + ... + d_N^2)/N = \frac{1}{N}\sum_{i=1}^{N} d_i^2 = \frac{1}{N}\sum_{i=1}^{N} (x - x_i)^2 \qquad (5)$$

The variance is useful because it gives us a measure of the spread or statistical uncertainty in the

* For these discussions, we will denote the "true" value as a variable without adornment (e.g., x).

measurements. You may have noticed a slight problem with the expression for the variance: We don't know the "true" value x, we have only an estimate, $\bar{x}$, from our measurements. It turns out that using $\bar{x}$ to instead of x in equation (5) systematically underestimates the variance. It can be shown that our best estimate of the "true" variance is given by *sample variance*:

$$\mathrm{var}_{sample}(x) = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{x} - x_i)^2 \tag{6}$$

A related statistic is the *standard deviation*, which is simply the square root of the variance:

$$\sigma_x = \sqrt{\mathrm{var}(x)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x - x_i)^2} \tag{7}$$

Note that the standard deviation has the same problem as does the variance in that we don't know x. Again we find that using $\bar{x}$ to instead of x systematically underestimates the standard deviation. We define the *sample standard deviation* to be the square root of the sample variance:

$$s_x = \sqrt{\mathrm{var}_{sample}(x)} = \sqrt{\left(\frac{1}{N-1}\right)\sum_{i=1}^{N}(\bar{x} - x_i)^2} \tag{8}$$

The sample standard deviation is our best estimate of the "true" standard deviation. [If, however, we have a situation where we can make *all possible* measurements, then we should use equation (7). Equation (7) defines a statistic which, for clarity, is often called the *population standard deviation*.]

To illustrate some of these points, consider the following: Suppose we want to know the average height and associated standard deviation of the entering class. We could measure every entering student (the entire population) and simply calculate the average. We would then simply calculate x and σ directly. Tracking down all of the entering students, however, would be very tedious. We could, instead, measure a representative* sample and calculate $\bar{x}$ and s_x as estimates of x and σ.

Modern spreadsheets such as MS Excel™ or Corel Quattro Pro™ also have built-in statistical functions. For example, **AVERAGE** (Excel) and **AVG** (Quattro) calculate the average of a range of cells; whereas **STDEV** (Excel) and **STDS** (Quattro) calculate the *sample* standard deviations. **STDEVP** (Excel) and **STD** (Quattro Pro) compute the *population* standard deviation.

PROBABLE ERROR

We now return to probability distributions. Consider equation (1), the expression for a Gaussian distribution. You should now have some idea as to why we wrote it in terms of d and σ. Most of the time we find that our measurements (x_i) deviate from the "true" value (x) and that these deviations (d_i) follow a Gaussian distribution with a standard deviation of σ. So, what is the significance of σ? Remember that the integral of a probability distribution over some range gives the probability of getting a result within that range. A straightforward calculation shows that the

* You have to be careful when choosing your sample. Measuring the students who have basketball scholarships would clearly bias your results. In the lab we must also take pains to ensure that our samples are unbiased.

integral of P_G from $-\sigma$ to $+\sigma$ is about 2/3. This means that there is probability of 2/3 for any *single*[*] measurement being within $\pm\sigma$ of the "true" value. It is in this sense that we introduce the concept of *probable error*.

Whenever we give a result, we also want to specify a probable error in such a way that we think that there is a 2/3 probability that the "true" value is within the range of values between our result minus the probable error to our result plus the probable error. In other words, if $\bar{x}$ is our best estimate of the "true" value x and $\sigma_{\bar{x}}$ is our best estimate of the probable error in $\bar{x}$, then there is a 2/3 probability that:

$$\bar{x} - \sigma_{\bar{x}} \le x \le \bar{x} + \sigma_{\bar{x}}$$

When we report results, we use the following notation:

$$\bar{x} \pm \sigma_{\bar{x}}$$

Thus, for example, the electron mass is given in data tables as

$$m_e = (9.109534 \pm 0.000047) \times 10^{-31} \text{ kg}.$$

By this we mean that the electron mass lies between 9.109487×10^{-31} kg and 9.109581×10^{-31} kg, with a *probability* of roughly 2/3.

SIGNIFICANT FIGURES

In informal usage the last *significant digit* implies something about the precision of the measurement. For example, if we measure a rod to be 101.3 mm long but consider the result accurate to only ±0.5 mm, we round off and say, "The length is 101 mm." That is, we believe the length lies between 100 mm and 102 mm, and is closest to 101 mm. The implication, if no error is stated explicitly, is that the uncertainty is ½ of one digit, in the place following the last significant digit.

Zeros to the left of the first non-zero digit do not count in the tally of significant figures. If we say $U = 0.001325$ Volts, the zero to the left of the decimal point, and the two zeros between the decimal point and the digits 1325 merely locate the decimal point; they do not indicate precision. (The zero to the left of the decimal point is included because decimal points are small and hard to see. It is just a visual clue—and it is a good idea to provide this clue when you write down numerical results in a laboratory!) The voltage U has thus been stated to four (4), not seven (7), significant figures. When we write it this way, we say we know its value to about ½ part in 1000 (strictly, ½ part in 1325 or one part in 2650). We could bring this out more clearly by writing either $U = 1.325\times10^{-3}$ V, or $U = 1325\times10^{-3}$ mV.

PROPAGATION OF ERRORS

More often than not, we want to use our measured quantities in further calculations. The question that then arises is: How do the errors "propagate"? In other words: What is the probable error in the calculated quantity given the probable errors in the input quantities?

---

[*] We'll come back to the issue of the probable error in the mean.

Before we answer this question, we want to introduce two new terms:

The *relative error* of a quantity Q is simply its probable error, σ_Q, divided by the absolute value of Q.

When we say that quantities *add in quadrature*, we mean that first you square the individual quantities, then you sum squared quantities, and then you take the square root of the sum of the squared quantities.

We will simply give the results for propagating errors rather than attempt to derive the formulas as the derivations are a bit beyond the scope of this write-up.

1. If the functional form of the derived quantity (f) is simply the product of a constant (C) times a quantity with known probable error (x and σ_x), then the probable error in the derived quantity is the product of the absolute value of the constant and the probable error in the quantity:
$$f(x) = Cx \rightarrow \sigma_f = |C|\sigma_x$$

2. If the functional form of the derived quantity (f) is simply the sum or difference of two quantities with known probable error (x and σ_x and y and σ_y), then the probable error in the derived quantity is the quadrature sum of the errors:
$$f(x,y) = x + y \text{ or } f(x,y) = x - y \rightarrow \sigma_f = \sqrt{\sigma_x^2 + \sigma_y^2}$$

3. If the functional form of the derived quantity (f) is simply the product or ratio of two quantities with known probable error (x and σ_x and y and σ_y), then the relative probable error in the derived quantity is the quadrature sum of the errors:
$$f(x,y) = x \times y \text{ or } f(x,y) = x/y \rightarrow \sigma_f / |f| = \sqrt{(\sigma_x / x)^2 + (\sigma_y / y)^2}$$

4. If the functional form of the derived quantity (f) is a quantity with known probable error (x and σ_x) raised to some constant power (a), then the relative probable error in the derived quantity is the product of the absolute value of the constant and the relative probable error in the quantity:
$$f(x) = x^a \rightarrow \sigma_f / |f| = |a|\sigma_x / |x|$$

5. If the functional form of the derived quantity (f) is the log of a quantity with known probable error (x and σ_x), then the probable error in the derived quantity is the relative probable error in the quantity:
$$f(x) = \ln(x) \rightarrow \sigma_f = \sigma_x / |x|$$

6. If the functional form of the derived quantity (f) is the anti-log of a quantity with known probable error (x and σ_x), then the relative probable error in the derived quantity is the probable error in the quantity:
$$f(x) = e^x \rightarrow \sigma_f / |f| = \sigma_x$$

And, finally, we give the general form (you are not expected to know or use this equation; it is only given for "completeness"):

$$f(x,y,...) \rightarrow \sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + ... \qquad (9)$$

PROBABLE ERROR IN THE MEAN

Suppose that we make two independent measurements of some quantity: x_1 and x_2. Our best estimate of x, the "true" value, is given by the mean, $\bar{x} = \frac{1}{2}(x_1 + x_2)$, and our best estimate of the probable error in x_1 and in x_2 is given by the sample standard deviation, $\sigma_{x_1} = \sigma_{x_2} = s_x = \sqrt{\left(\frac{1}{2-1}\right)\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2\right]}$. Note that s_x is *not* our best estimate of $\sigma_{\bar{x}}$, the probable error in $\bar{x}$. We must use the propagation of errors formulas to get $\sigma_{\bar{x}}$. Now, $\bar{x}$ is not exactly in one of the simple forms where we have a propagation of errors formula. However, we can see that it is of the form of a constant, $\left(\frac{1}{2}\right)$, times something else, $(x_1 + x_2)$, and so:

$$\sigma_{\bar{x}} = \left|\frac{1}{2}\right| \sigma_{x_1 + x_2}$$

The "something else" is a simple sum of two quantities with known probable errors (s_x) and we do have a formula for that:

$$\sigma_{x_1 + x_2} = \sqrt{\sigma_{x_1}^2 + \sigma_{x2}^2} = \sqrt{s_x^2 + s_x^2} = \sqrt{2} s_x$$

So we get the desired result for two measurements:

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{2}} s_x$$

By taking a second measurement, we have reduced our probable error by a factor of $1/\sqrt{2}$. You can probably see now how you would go about showing that adding third, x_3, changes this factor to $1/\sqrt{3}$. The general result (for N measurements) for the probable error in the mean is:

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{N}} s_x \qquad (10)$$

EXAMPLE

We can measure the gravitational acceleration g near the Earth's surface by dropping a mass in a vertical tube from which the air has been removed. Since the distance of fall (D), time of fall (t) and g are related by $D = \frac{1}{2} g t^2$, we have $g = 2Dt^{-2}$. So we see that we can determine g by simply measuring the time it takes for an object to fall a known distance. We hook up some photo-gates* to a timer so that we measure the time from when we release a ball to when it gets to the photo-gate. We very carefully use a ruler to set the distance (D) that the ball is to fall to 1.800 m. We estimate that we can read our ruler to within 1 mm. We drop the ball ten times and get the following times (t_i): 0.6053, 0.6052, 0.6051, 0.6050, 0.6052, 0.6054, 0.6053, 0.6047, 0.6048, and 0.6055 seconds. The average of these times $(\bar{t})$ is 0.60515 seconds. Our best estimate of g is then $g_{\text{exp}} = 2D/\bar{t}^2 = 9.8305\ \text{m/s}^2$. This is larger than the "known" local value of 9.809 m/s^2 by 0.0215 m/s^2 (0.2%). We do expect experimental uncertainties to cause our value to be different,

* A device with a light source and detector that changes an output when something comes between the source and detector.

but the question is: Is our result consistent with the "known" value, within experimental errors? To check this we must estimate our probable error.

Our expression for *g* is, once again*, not precisely in one of the simple propagation of errors forms and so we must look at it piecemeal. This time we will not work it all out algebraically, but will instead substitute numbers as soon as we can so that we can take a look at their effects on the final probable error.

What are our experimental probable errors? We've estimated that our probable error in the distance (σ_D) is 1 mm (hence a relative error, $\sigma_D/|D|$, of 0.00006 or 0.06%). From our time data we calculate the *sample* standard deviation (s_t) to be 0.00025 seconds. Recall that this is *not* the probable error in the mean (our best estimate of the "true" time for the ball to fall), it is the probable error in any *single one* of the time measurements (t_i). The probable error in the mean is s_t divided by the square root of the number of samples (10): $\sigma_{\bar{t}} = s_t/\sqrt{10} = 0.00008$ seconds (for a relative error, $\sigma_{\bar{t}}/|\bar{t}|$, of 0.00008 or 0.008%). We see that the relative error in the distance measurement is quite a bit larger than the relative error in the time measurement and so we might assume that we could ignore the time error (essentially treating the time as a constant). However, the time enters into *g* as a square and we expect that that makes a bigger contribution than otherwise. So we don't (yet) make any such simplifying assumptions. We see that our estimate of *g* (which we denote by g_{exp}) is of the form of a constant (2) times something else ($D/\bar{t}^2$) and so:

$$\sigma_{g_{\exp}} = |2|\,\sigma_{D/\bar{t}^2}$$

$D/\bar{t}^2$ is of the form of a simple product of two quantities (D and $\bar{t}^2$) and so:

$$\sigma_{D/\bar{t}^2}/\left|D/\bar{t}^2\right| = \sqrt{\left(\sigma_D/D\right)^2 + \left(\sigma_{\bar{t}^2}/\bar{t}^2\right)^2}$$

Now we are getting somewhere as we have $\sigma_D/|D|$ (0.06%). We need only find $\sigma_{\bar{t}^2}/\bar{t}^2$. $\bar{t}^2$ is of the form of a quantity raised to a constant power and so:

$$\sigma_{\bar{t}^2}/\bar{t}^2 = |2|\,\sigma_{\bar{t}}/|\bar{t}|$$

Now we can see the effect of squaring $\bar{t}$: Its contribution to the probable error is doubled. Consider the two terms under the square root:

$$\left(\sigma_D/D\right)^2 = 3.6\times10^{-7}$$

$$\left(\sigma_{\bar{t}^2}/\bar{t}^2\right)^2 = 6.4\times10^{-9}$$

Now we can see that, even though the time enters as a square, we would have been justified in ignoring its contribution to the probable error in *g*. Plugging the numbers back in, we finally get $\sigma_{g_{\exp}} = 0.0061\ \mathrm{m/s^2}$.

We see that our result is 3.5 standard deviations larger than the "known" value. While not totally out of the question, it is still very unlikely and so we need to look for the source of the problem. In this case we find that the ruler is one of those with a worn end. We carefully measure the offset and find it to be 5.0 mm too short. Subtracting this changes *D* to 1.795 m and g_{exp} to

* Refer to the discussion of the probable error in the mean.

9.803 m/s^2, well within our experimental error*.

* The implied error in our measurement of the offset (0.1 mm) is much smaller than the error in the original *D* and so we can afford to ignore its contribution to the probable error in g_{exp}.

SUMMARY OF STATISTICAL FORMULAS

Sample mean:

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

(best estimate of the "true" value of x, using N measurements)

Sample standard deviation:

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

(best estimate of error in any single measurement, x_i)

Standard error of the mean:

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{N}} s_x$$

(best estimate of error in determining the population mean, $\bar{x}$)

SUMMARY OF ERROR PROPAGATION FORMULAS

	Functional form	**Error propagation formula**
1.	$f(x) = Cx$	$\sigma_f = \lvert C \rvert \sigma_x$
2.	$f(x,y) = x \pm y$	$\sigma_f = \sqrt{\sigma_x^2 + \sigma_y^2}$
3.	$f(x,y) = x \times y$ or x/y	$\sigma_f / \lvert f \rvert = \sqrt{(\sigma_x / x)^2 + (\sigma_y / y)^2}$
4.	$f(x) = x^a$	$\sigma_f / \lvert f \rvert = \lvert a \rvert \sigma_x / \lvert x \rvert$
5.	$f(x) = \ln(x)$	$\sigma_f = \sigma_x / \lvert x \rvert$
6.	$f(x) = e^x$	$\sigma_f / \lvert f \rvert = \sigma_x$

and the general form:

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