

## APPENDIX D

# ACCURACY OF MEASUREMENTS AND TREATMENT OF EXPERIMENTAL UNCERTAINTY

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 measurements deviate from these “true” values. We address here how to use measurements to best estimate the “true” values and how to estimate possible uncertainties/errors in our measurements. Our understanding of experimental uncertainty (i.e., errors) is based on the mathematical theory of probability and statistics.

### **Accuracy and Precision; Random and Systematic Errors**

In common usage, “accuracy” and “precision” are synonyms. To scientists and engineers, 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. We want to make accurate AND precise measurements.

*Systematic* errors are due to problems with the technique or measuring instrument. For example, as many of the rulers found in labs have worn ends, length measurements could be wrong. One can make very precise (reproducible) measurements that are quite inaccurate (far from the true value). We address systematic errors through careful experimental design and attention to detail. 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 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). We address random errors by using 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.

### **Probability**

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  $\frac{1}{2}$  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  $\frac{1}{6}$ .

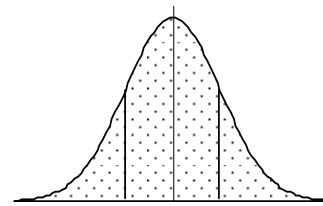
The previous examples illustrate four basic principles about probability:

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 did not land on its edge.
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.
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 coin, the probability that it lands heads *or* tails is  $\frac{1}{2} + \frac{1}{2} = 1$ . If we toss a 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 only estimate through observation. The most common probability distribution encountered in the lab is the *Gaussian* distribution. The Gaussian distribution is also known as the *normal* distribution or a *bell curve* (it is shaped somewhat like a fancy bell) when applied to grade distributions. The mathematical form of the Gaussian distribution is:



Gaussian Distribution

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

The Gaussian distribution is ubiquitous because it is the end result you obtain if you have a number of processes, each with their own probability distribution, that “mix together” to yield a final result. A key aspect of this distribution is that the probability of obtaining a result (a value for  $d$ ) between  $-\sigma$  and  $\sigma$  is approximately two out of three.

## Statistics

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. Perhaps the single most important statistic is the *mean* or *average*. A “bar” over a variable (e.g.,  $\bar{x}$ ) or “angle brackets” (e.g.,  $\langle x \rangle$ ) are used to indicate that it is an average. The mean of the distribution will be our best estimate of the “true” value:

$$x \approx \bar{x} \equiv \langle x \rangle = (x_1 + x_2 + \dots + x_N) / N = \frac{1}{N} \sum_{i=1}^N x_i \quad (2)$$

A histogram of our measured values will usually yield a Gaussian distribution. Our best estimate of the standard deviation of this distribution is given by the *sample standard deviation*:

$$s_x = \sqrt{\left(\frac{1}{N-1}\right) \sum_{i=1}^N (\bar{x} - x_i)^2} \quad (3)$$

The sample standard deviation is defined as the average amount by which values in a distribution differ from the mean (ignoring the sign of the difference). The standard deviation is useful because it gives us a measure of the spread or statistical uncertainty in the measurements. Modern spreadsheets (such as MS Excel™) as well as some calculators (such as HP and TI) also have built-in statistical functions. For example, **AVERAGE** (Excel) and  $\bar{x}$  (calculator) calculate the average of a range of values; whereas **STDEV** (Excel) and  $s_x$  (calculator) calculate the sample standard deviation.

We expect the average or mean to give us a good representation of the true value of our measurement or the parameter of which we seek. But how well does the mean value represent the true value? The sample standard deviation gives us the likely error that the next measurement or value will represent the true value. However, the mean or average value should have a significantly lower error, because it represents many measurements.

Through straightforward application of error propagation methodology, we find that our best estimate of the *standard error in the mean* is given by

$$\sigma_{\bar{x}} = s_x / \sqrt{N} \quad (4)$$

### Standard Error

Whenever we report a result, we want to specify a standard error in such a way as to indicate that we think that there is roughly a 2/3 probability that the “true” value is within the range of values between our result minus the standard error to our result plus the standard 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 standard error in  $\bar{x}$ , then there is a 2/3 probability that:

$$\bar{x} - \sigma_{\bar{x}} \leq x \leq \bar{x} + \sigma_{\bar{x}} \quad (5)$$

We use the notation  $\bar{x} \pm \sigma_{\bar{x}}$  to report results. For example, the electron mass is given by  $m_e = (9.1093826 \pm 0.0000016) \times 10^{-31}$  kg. By this we mean that the electron mass lies between  $9.1093810 \times 10^{-31}$  kg and  $9.1093842 \times 10^{-31}$  kg, with a probability of roughly 2/3.

### Significant Figures

In informal usage the *least 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  $\pm 0.5$  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  $\frac{1}{2}$  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 write  $U = 0.001325$  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 voltage  $U$  has thus been stated to four (4), not seven (7), significant figures. When we write it this way, we know its value to about  $\frac{1}{2}$  part in 1,000 (strictly,  $\frac{1}{2}$  part in 1,325 or one part in 2,650). We could bring this out more clearly by

writing either  $U = 1.325 \times 10^{-3}$  V, or  $U = 1.325$  mV. How many significant figures does the number 3,000 have? Most scientists would say 1. If there were a decimal at the end, we would usually say 4. But some textbook authors might explicitly state that 3,000 has 4 significant figures. You need to be aware of some differences like this.

## Propagation of Errors

We often 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 standard error in a particular calculated quantity given the standard errors in the input values?” Let’s discuss some new terms that are often used. The *relative error* of a quantity  $Q$  is simply its standard error,  $\sigma_Q$ , divided by the absolute value of  $Q$ . For example, if a length is known to  $49 \pm 4$  cm, we say it has a relative error of  $4/49 = 0.082$ . We might also refer to this as a *fractional error*. It is often useful to express such values in *percent*. In this case we would say that we had a relative error or a percent error of 8.2%.

On the last page of this Appendix we give the formulae for propagating errors [the derivations are a bit beyond the scope of this exposition]. These equations are an important part of this appendix, and you will refer to them often. We now use these formulae to find the standard error in the mean.

### 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. The distance of fall ( $D$ ), time of fall ( $t$ ) and  $g$  are related by  $D = \frac{1}{2}gt^2$ , so we have  $g = 2Dt^{-2}$ . We can determine  $g$  by simply measuring the time it takes an object to fall a known distance.

We use an electronic device to measure the time from when we release a ball to when it arrives at the photogate. We very carefully use a ruler to set the distance that the ball drops to  $D = 1.800$  m. *We estimate that we can read our ruler to within 1 mm.* We drop the ball ten times and obtain 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 0.605150 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  $9.809 \text{ m/s}^2$  by  $0.0215 \text{ m/s}^2$  (0.22% difference). We expect experimental uncertainties to cause our value to be different, but is this difference too large? To check this we must estimate our standard error.

**VERY IMPORTANT NOTE:** Do NOT round off intermediate results when making calculations. Keep full “machine precision” to minimize the effects of *round-off errors*. Only round off final results and use your error estimate to determine significant figures.

What are our experimental standard errors? We’ve estimated that our standard error in the distance ( $\sigma_D$ ) is 1 mm (hence a relative (or fractional) error,  $\sigma_D/|D|$ , of 0.000556 or 0.0556%). From our time measurements we calculate the sample standard deviation ( $s_t$ ) to be 0.000255 seconds. This means we have a 2/3 probability of finding that the next measurement will be within 0.000255 s of the mean value of 0.605150 s. But because we have taken several measurements, we want to find the **standard error in the mean**. We

divide  $s_t$  by the square root of the number of samples ( $N = 10$ ):

Standard error of the mean:  $\sigma_{\bar{t}} = s_t / \sqrt{10} = 0.000255 \text{ s} / \sqrt{10} = 0.0000806 \text{ s}$  and,  
 $\sigma_{\bar{t}} / |\bar{t}| = 0.0000806 \text{ s} / 0.605150 \text{ s} = 0.000133$ .

We see that the relative error in the distance measurement (0.000556) is quite a bit larger than the relative error in the time measurement (0.000133) and so we might assume that we could ignore the time error. 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. Our measurement of  $g$  (which we denote by  $g_{\text{exp}}$ ) is of the form of a constant (2) times something else ( $D/\bar{t}^2$ ) and so:

$$\sigma_{g_{\text{exp}}} = |2| \sigma_{D/\bar{t}^2} \quad (6)$$

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

$$\frac{\sigma_{g_{\text{exp}}}}{g_{\text{exp}}} = \frac{|2| \sigma_{D/\bar{t}^2}}{2|D/\bar{t}^2|} = \frac{\sigma_{D/\bar{t}^2}}{|D/\bar{t}^2|} = \sqrt{\left(\frac{\sigma_D}{D}\right)^2 + \left(\frac{\sigma_{\bar{t}^2}}{\bar{t}^2}\right)^2} \quad (7)$$

We already found that  $\sigma_D/|D| = 0.000556$ . We know  $\sigma_{\bar{t}}/|\bar{t}|$ , but not  $\sigma_{\bar{t}^2}/\bar{t}^2$ .  $\bar{t}^2$  is of the form of a quantity raised to a constant power and so:

$$\sigma_{\bar{t}^2}/\bar{t}^2 = |2| \sigma_{\bar{t}}/|\bar{t}| = 2 \times 0.000133 = 0.000266$$

Now we can see the effect of squaring  $\bar{t}$ : Its contribution to the standard error is doubled. Consider the two terms under the square root in Eq. (7):

$$\left(\sigma_D/D\right)^2 = 3.09 \times 10^{-7} \quad \text{and} \quad \left(\sigma_{\bar{t}^2}/\bar{t}^2\right)^2 = 7.10 \times 10^{-8}$$

Now we can see that, even though the time enters as a square, we would have been justified in ignoring its contribution to the standard error in  $g$ . Inserting these values into Eq. (7) finally gives

$$\frac{\sigma_{g_{\text{exp}}}}{g_{\text{exp}}} = \frac{|2| \sigma_{D/\bar{t}^2}}{2|D/\bar{t}^2|} = \frac{\sigma_{D/\bar{t}^2}}{|D/\bar{t}^2|} = \sqrt{\left(\sigma_D/D\right)^2 + \left(\sigma_{\bar{t}^2}/\bar{t}^2\right)^2} = \sqrt{3.09 \times 10^{-7} + 7.10 \times 10^{-8}} = 0.000617$$

We have  $\sigma_{g_{\text{exp}}} = 0.000617 \times g_{\text{exp}} = (0.000617)(9.83 \text{ m/s}^2) = 0.0061 \text{ m/s}^2$ . Our measured value becomes  $g_{\text{exp}} = 9.8305 \pm 0.0061 \text{ m/s}^2$  or more properly,  $g_{\text{exp}} = 9.830 \pm 0.006 \text{ m/s}^2$ . Our experimental difference with the known value of  $0.0215 \text{ m/s}^2$  is 3.5 standard deviations. While not totally out of the question, such a difference is unlikely and we should look for a systematic error. In this case we find that the ruler is one of those with a worn end, and it is 5.0 mm short. Subtracting this changes  $D$  to 1.795 m and  $g_{\text{exp}}$  to  $9.803 \pm 0.006 \text{ m/s}^2$ , a value consistent with the known value within our experimental error.

## Summary of Statistical Formulae

**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 Formulae

**Functional form**

**Error propagation formula**

1.  $f(x) = Cx$

$$\sigma_f = |C| \sigma_x$$

2.  $f(x, y) = x \pm y$

$$\sigma_f = \sqrt{\sigma_x^2 + \sigma_y^2}$$

3.  $f(x, y) = x \times y$  or  $x/y$

$$\sigma_f / |f| = \sqrt{(\sigma_x / x)^2 + (\sigma_y / y)^2}$$

4.  $f(x) = x^a$

$$\sigma_f / |f| = |a| \sigma_x / |x|$$

5.  $f(x) = \ln(x)$

$$\sigma_f = \sigma_x / |x|$$

6.  $f(x) = e^x$

$$\sigma_f / |f| = \sigma_x$$

7.  $f(x) = Cx^a y^b$

$$\sigma_f / |f| = \sqrt{\left(\frac{a\sigma_x}{x}\right)^2 + \left(\frac{b\sigma_y}{y}\right)^2}$$

and the general form:

8.  $f(x, y, \dots)$

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