

Chapter I

Preliminary Description of Error Analysis

Error analysis is the study and evaluation of uncertainty in measurement. Experience has shown that no measurement, however carefully made, can be completely free of uncertainties. Because the whole structure and application of science depends on measurements, the ability to evaluate these uncertainties and keep them to a minimum is crucially important.

This first chapter describes some simple measurements that illustrate the inevitable occurrence of experimental uncertainties and show the importance of knowing how large these uncertainties are. The chapter then describes how (in some simple cases, at least) the magnitude of the experimental uncertainties can be estimated realistically, often by means of little more than plain common sense.

1.1 Errors as Uncertainties

In science, the word *error* does not carry the usual connotations of the terms *mistake* or *blunder*. Error in a scientific measurement means the inevitable uncertainty that attends all measurements. As such, errors are not mistakes; you cannot eliminate them by being very careful. The best you can hope to do is to ensure that errors are as small as reasonably possible and to have a reliable estimate of how large they are. Most textbooks introduce additional definitions of error, and these are discussed later. For now, error is used exclusively in the sense of uncertainty, and the two words are used interchangeably.

1.2 Inevitability of Uncertainty

To illustrate the inevitable occurrence of uncertainties, we have only to examine any everyday measurement carefully. Consider, for example, a carpenter who must measure the height of a doorway before installing a door. As a first rough measurement, he might simply look at the doorway and estimate its height as 210 cm. This crude “measurement” is certainly subject to uncertainty. If pressed, the carpenter might express this uncertainty by admitting that the height could be anywhere between 205 cm and 215 cm.

If he wanted a more accurate measurement, he would use a tape measure and might find the height is 211.3 cm. This measurement is certainly more precise than his original estimate, but it is obviously still subject to some uncertainty, because it is impossible for him to know the height to be exactly 211.3000 cm rather than 211.3001 cm, for example.

This remaining uncertainty has many sources, several of which are discussed in this book. Some causes could be removed if the carpenter took enough trouble. For example, one source of uncertainty might be that poor lighting hampers reading of the tape; this problem could be corrected by improving the lighting.

On the other hand, some sources of uncertainty are intrinsic to the process of measurement and can never be removed entirely. For example, let us suppose the carpenter's tape is graduated in half-centimeters. The top of the door probably will not coincide precisely with one of the half-centimeter marks, and if it does not, the carpenter must *estimate* just where the top lies between two marks. Even if the top happens to coincide with one of the marks, the mark itself is perhaps a millimeter wide; so he must estimate just where the top lies within the mark. In either case, the carpenter ultimately must estimate where the top of the door lies relative to the markings on the tape, and this necessity causes some uncertainty in the measurement.

By buying a better tape with closer and finer markings, the carpenter can reduce his uncertainty but cannot eliminate it entirely. If he becomes obsessively determined to find the height of the door with the greatest precision technically possible, he could buy an expensive laser interferometer. But even the precision of an interferometer is limited to distances of the order of the wavelength of light (about 0.5×10^{-6} meters). Although the carpenter would now be able to measure the height with fantastic precision, he still would not know the height of the doorway *exactly*.

Furthermore, as our carpenter strives for greater precision, he will encounter an important problem of principle. He will certainly find that the height is different in different places. Even in one place, he will find that the height varies if the temperature and humidity vary, or even if he accidentally rubs off a thin layer of dirt. In other words, he will find that there is no such thing as *the* height of the doorway. This kind of problem is called a *problem of definition* (the height of the door is not a well-defined quantity) and plays an important role in many scientific measurements.

Our carpenter's experiences illustrate a point generally found to be true, that is, that no physical quantity (a length, time, or temperature, for example) can be measured with complete certainty. With care, we may be able to reduce the uncertainties until they are extremely small, but to eliminate them entirely is impossible.

In everyday measurements, we do not usually bother to discuss uncertainties. Sometimes the uncertainties simply are not interesting. If we say that the distance between home and school is 3 miles, whether this means "somewhere between 2.5 and 3.5 miles" or "somewhere between 2.99 and 3.01 miles" is usually unimportant. Often the uncertainties are important but can be allowed for instinctively and without explicit consideration. When our carpenter fits his door, he must know its height with an uncertainty that is less than 1 mm or so. As long as the uncertainty is this small, the door will (for all practical purposes) be a perfect fit, and his concern with error analysis is at an end.

1.3 Importance of Knowing the Uncertainties

Our example of the carpenter measuring a doorway illustrates how uncertainties are always present in measurements. Let us now consider an example that illustrates more clearly the crucial importance of knowing how big these uncertainties are.

Suppose we are faced with a problem like the one said to have been solved by Archimedes. We are asked to find out whether a crown is made of 18-karat gold, as claimed, or a cheaper alloy. Following Archimedes, we decide to test the crown's density ρ knowing that the densities of 18-karat gold and the suspected alloy are

$$\rho_{\text{gold}} = 15.5 \text{ gram/cm}^3$$

and

$$\rho_{\text{alloy}} = 13.8 \text{ gram/cm}^3.$$

If we can measure the density of the crown, we should be able (as Archimedes suggested) to decide whether the crown is really gold by comparing ρ with the known densities ρ_{gold} and ρ_{alloy} .

Suppose we summon two experts in the measurement of density. The first expert, George, might make a quick measurement of ρ and report that his best estimate for ρ is 15 and that it almost certainly lies between 13.5 and 16.5 gram/cm^3 . Our second expert, Martha, might take a little longer and then report a best estimate of 13.9 and a probable range from 13.7 to 14.1 gram/cm^3 . The findings of our two experts are summarized in Figure 1.1.

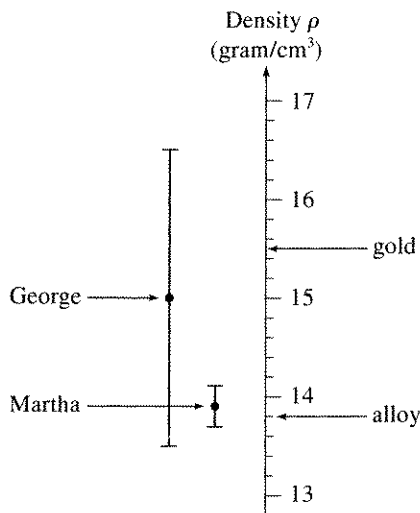


Figure 1.1. Two measurements of the density of a supposedly gold crown. The two black dots show George's and Martha's best estimates for the density; the two vertical error bars show their margins of error, the ranges within which they believe the density probably lies. George's uncertainty is so large that both gold and the suspected alloy fall within his margins of error; therefore, his measurement does not determine which metal was used. Martha's uncertainty is appreciably smaller, and her measurement shows clearly that the crown is not made of gold.

The first point to notice about these results is that although Martha's measurement is much more precise, George's measurement is probably also correct. Each expert states a range within which he or she is confident ρ lies, and these ranges overlap; so it is perfectly possible (and even probable) that both statements are correct.

Note next that the uncertainty in George's measurement is so large that his results are of no use. The densities of 18-karat gold and of the alloy both lie within his range, from 13.5 to 16.5 gram/cm³; so no conclusion can be drawn from George's measurements. On the other hand, Martha's measurements indicate clearly that the crown is not genuine; the density of the suspected alloy, 13.8, lies comfortably inside Martha's estimated range of 13.7 to 14.1, but that of 18-karat gold, 15.5, is far outside it. Evidently, if the measurements are to allow a conclusion, the experimental uncertainties must not be too large. The uncertainties do not need to be extremely small, however. In this respect, our example is typical of many scientific measurements, for which uncertainties have to be reasonably small (perhaps a few percent of the measured value) but for which extreme precision is often unnecessary.

Because our decision hinges on Martha's claim that ρ lies between 13.7 and 14.1 gram/cm³, she must give us sufficient reason to believe her claim. In other words, she must *justify* her stated range of values. This point is often overlooked by beginning students, who simply assert their uncertainties but omit any justification. Without a brief explanation of how the uncertainty was estimated, the assertion is almost useless.

The most important point about our two experts' measurements is this: Like most scientific measurements, they would both have been useless if they had not included reliable statements of their uncertainties. In fact, if we knew only the two best estimates (15 for George and 13.9 for Martha), not only would we have been unable to draw a valid conclusion, but we could actually have been misled, because George's result (15) seems to suggest the crown is genuine.

1.4 More Examples

The examples in the past two sections were chosen, not for their great importance, but to introduce some principal features of error analysis. Thus, you can be excused for thinking them a little contrived. It is easy, however, to think of examples of great importance in almost any branch of applied or basic science.

In the applied sciences, for example, the engineers designing a power plant must know the characteristics of the materials and fuels they plan to use. The manufacturer of a pocket calculator must know the properties of its various electronic components. In each case, somebody must measure the required parameters, and having measured them, must establish their reliability, which requires error analysis. Engineers concerned with the safety of airplanes, trains, or cars must understand the uncertainties in drivers' reaction times, in braking distances, and in a host of other variables; failure to carry out error analysis can lead to accidents such as that shown on the cover of this book. Even in a less scientific field, such as the manufacture of clothing, error analysis in the form of quality control plays a vital part.

In the basic sciences, error analysis has an even more fundamental role. When any new theory is proposed, it must be tested against older theories by means of one or more experiments for which the new and old theories predict different outcomes. In principle, a researcher simply performs the experiment and lets the outcome decide between the rival theories. In practice, however, the situation is complicated by the inevitable experimental uncertainties. These uncertainties must all be analyzed carefully and their effects reduced until the experiment singles out one acceptable theory. That is, the experimental results, with their uncertainties, must be *consistent* with the predictions of one theory and *inconsistent* with those of all known, reasonable alternatives. Obviously, the success of such a procedure depends critically on the scientist's understanding of error analysis and ability to convince others of this understanding.

A famous example of such a test of a scientific theory is the measurement of the bending of light as it passes near the sun. When Einstein published his general theory of relativity in 1916, he pointed out that the theory predicted that light from a star would be bent through an angle $\alpha = 1.8''$ as it passes near the sun. The simplest classical theory would predict no bending ($\alpha = 0$), and a more careful classical analysis would predict (as Einstein himself noted in 1911) bending through an angle $\alpha = 0.9''$. In principle, all that was necessary was to observe a star when it was aligned with the edge of the sun and to measure the angle of bending α . If the result were $\alpha = 1.8''$, general relativity would be vindicated (at least for this phenomenon); if α were found to be 0 or $0.9''$, general relativity would be wrong and one of the older theories right.

In practice, measuring the bending of light by the sun was extremely hard and was possible only during a solar eclipse. Nonetheless, in 1919 it was successfully measured by Dyson, Eddington, and Davidson, who reported their best estimate as $\alpha = 2''$, with 95% confidence that it lay between $1.7''$ and $2.3''$.¹ Obviously, this result was consistent with general relativity and inconsistent with either of the older predictions. Therefore, it gave strong support to Einstein's theory of general relativity.

At the time, this result was controversial. Many people suggested that the uncertainties had been badly underestimated and hence that the experiment was inconclusive. Subsequent experiments have tended to confirm Einstein's prediction and to vindicate the conclusion of Dyson, Eddington, and Davidson. The important point here is that the whole question hinged on the experimenters' ability to estimate reliably all their uncertainties and to convince everyone else they had done so.

Students in introductory physics laboratories are not usually able to conduct definitive tests of new theories. Often, however, they do perform experiments that test existing physical theories. For example, Newton's theory of gravity predicts that bodies fall with constant acceleration g (under the appropriate conditions), and students can conduct experiments to test whether this prediction is correct. At first, this kind of experiment may seem artificial and pointless because the theories have obvi-

¹This simplified account is based on the original paper of F. W. Dyson, A. S. Eddington, and C. Davidson (*Philosophical Transactions of the Royal Society*, **220A**, 1920, 291). I have converted the *probable error* originally quoted into the 95% confidence limits. The precise significance of such confidence limits will be established in Chapter 5.

ously been tested many times with much more precision than possible in a teaching laboratory. Nonetheless, if you understand the crucial role of error analysis and accept the challenge to make the most precise test possible with the available equipment, such experiments can be interesting and instructive exercises.

1.5 Estimating Uncertainties When Reading Scales

Thus far, we have considered several examples that illustrate why every measurement suffers from uncertainties and why their magnitude is important to know. We have not yet discussed how we can actually evaluate the magnitude of an uncertainty. Such evaluation can be fairly complicated and is the main topic of this book. Fortunately, reasonable estimates of the uncertainty of some simple measurements are easy to make, often using no more than common sense. Here and in Section 1.6, I discuss examples of such measurements. An understanding of these examples will allow you to begin using error analysis in your experiments and will form the basis for later discussions.

The first example is a measurement using a marked scale, such as the ruler in Figure 1.2 or the voltmeter in Figure 1.3. To measure the length of the pencil in

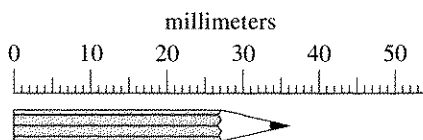


Figure 1.2. Measuring a length with a ruler.

Figure 1.2, we must first place the end of the pencil opposite the zero of the ruler and then decide where the tip comes to on the ruler's scale. To measure the voltage in Figure 1.3, we have to decide where the needle points on the voltmeter's scale. If we assume the ruler and voltmeter are reliable, then in each case the main prob-

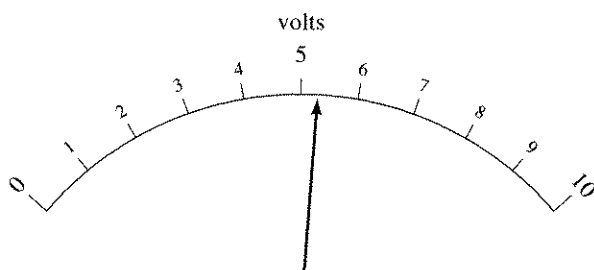


Figure 1.3. A reading on a voltmeter.

lem is to decide where a certain point lies in relation to the scale markings. (Of course, if there is any possibility the ruler and voltmeter are *not* reliable, we will have to take this uncertainty into account as well.)

The markings of the ruler in Figure 1.2 are fairly close together (1 mm apart). We might reasonably decide that the length shown is undoubtedly closer to 36 mm than it is to 35 or 37 mm but that no more precise reading is possible. In this case, we would state our conclusion as

$$\begin{aligned}\text{best estimate of length} &= 36 \text{ mm}, \\ \text{probable range: } &35.5 \text{ to } 36.5 \text{ mm}\end{aligned}\tag{1.1}$$

and would say that we have measured the length to the nearest millimeter.

This type of conclusion—that the quantity lies closer to a given mark than to either of its neighboring marks—is quite common. For this reason, many scientists introduce the convention that the statement “ $l = 36 \text{ mm}$ ” without any qualification is presumed to mean that l is closer to 36 than to 35 or 37; that is,

$$l = 36 \text{ mm}$$

means

$$35.5 \text{ mm} \leq l \leq 36.5 \text{ mm}.$$

In the same way, an answer such as $x = 1.27$ without any stated uncertainty would be presumed to mean that x lies between 1.265 and 1.275. In this book, I do not use this convention but instead always indicate uncertainties explicitly. Nevertheless, you need to understand the convention and know that it applies to any number stated without an uncertainty, especially in this age of pocket calculators, which display many digits. If you unthinkingly copy a number such as 123.456 from your calculator without any qualification, then your reader is entitled to assume the number is definitely correct to six significant figures, which is very unlikely.

The markings on the voltmeter shown in Figure 1.3 are more widely spaced than those on the ruler. Here, most observers would agree that you can do better than simply identify the mark to which the pointer is closest. Because the spacing is larger, you can realistically estimate where the pointer lies in the space between two marks. Thus, a reasonable conclusion for the voltage shown might be

$$\begin{aligned}\text{best estimate of voltage} &= 5.3 \text{ volts}, \\ \text{probable range: } &5.2 \text{ to } 5.4 \text{ volts}.\end{aligned}\tag{1.2}$$

The process of estimating positions between the scale markings is called *interpolation*. It is an important technique that can be improved with practice.

Different observers might not agree with the precise estimates given in Equations (1.1) and (1.2). You might well decide that you could interpolate for the length in Figure 1.2 and measure it with a smaller uncertainty than that given in Equation (1.1). Nevertheless, few people would deny that Equations (1.1) and (1.2) are *reasonable* estimates of the quantities concerned and of their probable uncertainties. Thus, we see that approximate estimation of uncertainties is fairly easy when the only problem is to locate a point on a marked scale.

1.6 Estimating Uncertainties in Repeatable Measurements

Many measurements involve uncertainties that are much harder to estimate than those connected with locating points on a scale. For example, when we measure a time interval using a stopwatch, the main source of uncertainty is not the difficulty of reading the dial but our own unknown reaction time in starting and stopping the watch. Sometimes these kinds of uncertainty can be estimated reliably, if we can repeat the measurement several times. Suppose, for example, we time the period of a pendulum once and get an answer of 2.3 seconds. From one measurement, we can't say much about the experimental uncertainty. But if we repeat the measurement and get 2.4 seconds, then we can immediately say that the uncertainty is probably of the order of 0.1 s. If a sequence of four timings gives the results (in seconds),

$$2.3, 2.4, 2.5, 2.4, \quad (1.3)$$

then we can begin to make some fairly realistic estimates.

First, a natural assumption is that the best estimate of the period is the *average*² value, 2.4 s.

Second, another reasonably safe assumption is that the correct period lies between the lowest value, 2.3, and the highest, 2.5. Thus, we might reasonably conclude that

$$\begin{aligned} \text{best estimate} &= \text{average} = 2.4 \text{ s}, \\ \text{probable range} &: 2.3 \text{ to } 2.5 \text{ s}. \end{aligned} \quad (1.4)$$

Whenever you can repeat the same measurement several times, the spread in your measured values gives a valuable indication of the uncertainty in your measurements. In Chapters 4 and 5, I discuss statistical methods for treating such repeated measurements. Under the right conditions, these statistical methods give a more accurate estimate of uncertainty than we have found in Equation (1.4) using just common sense. A proper statistical treatment also has the advantage of giving an objective value for the uncertainty, independent of the observer's individual judgment.³ Nevertheless, the estimate in statement (1.4) represents a simple, realistic conclusion to draw from the four measurements in (1.3).

Repeated measurements such as those in (1.3) cannot always be relied on to reveal the uncertainties. First, we must be sure that the quantity measured is really the *same* quantity each time. Suppose, for example, we measure the breaking strength of two supposedly identical wires by breaking them (something we can't do more than once with each wire). If we get two different answers, this difference *may* indicate that our measurements were uncertain *or* that the two wires were not really identical. By itself, the difference between the two answers sheds no light on the reliability of our measurements.

²I will prove in Chapter 5 that the best estimate based on several measurements of a quantity is almost always the average of the measurements.

³Also, a proper statistical treatment usually gives a *smaller* uncertainty than the full range from the lowest to the highest observed value. Thus, upon looking at the four timings in (1.3), we have judged that the period is "probably" somewhere between 2.3 and 2.5 s. The statistical methods of Chapters 4 and 5 let us state with 70% confidence that the period lies in the smaller range of 2.36 to 2.44 s.

Even when we can be sure we are measuring the same quantity each time, repeated measurements do not always reveal uncertainties. For example, suppose the clock used for the timings in (1.3) was running consistently 5% fast. Then, all timings made with it will be 5% too long, and no amount of repeating (with the same clock) will reveal this deficiency. Errors of this sort, which affect all measurements in the same way, are called *systematic* errors and can be hard to detect, as discussed in Chapter 4. In this example, the remedy is to check the clock against a more reliable one. More generally, if the reliability of any measuring device is in doubt, it should clearly be checked against a device known to be more reliable.

The examples discussed in this and the previous section show that experimental uncertainties sometimes can be estimated easily. On the other hand, many measurements have uncertainties that are *not* so easily evaluated. Also, we ultimately want more precise values for the uncertainties than the simple estimates just discussed. These topics will occupy us from Chapter 3 onward. In Chapter 2, I assume temporarily that you know how to estimate the uncertainties in all quantities of interest, so that we can discuss how the uncertainties are best reported and how they are used in drawing an experimental conclusion.

Chapter 2

How to Report and Use Uncertainties

Having read Chapter 1, you should now have some idea of the importance of experimental uncertainties and how they arise. You should also understand how uncertainties can be estimated in a few simple situations. In this chapter, you will learn some basic notations and rules of error analysis and study examples of their use in typical experiments in a physics laboratory. The aim is to familiarize you with the basic vocabulary of error analysis and its use in the introductory laboratory. Chapter 3 begins a systematic study of how uncertainties are actually evaluated.

Sections 2.1 to 2.3 define several basic concepts in error analysis and discuss general rules for stating uncertainties. Sections 2.4 to 2.6 discuss how these ideas could be used in typical experiments in an introductory physics laboratory. Finally, Sections 2.7 to 2.9 introduce fractional uncertainty and discuss its significance.

2.1 Best Estimate \pm Uncertainty

We have seen that the correct way to state the result of measurement is to give a best estimate of the quantity and the range within which you are confident the quantity lies. For example, the result of the timings discussed in Section 1.6 was reported as

$$\begin{aligned}\text{best estimate of time} &= 2.4 \text{ s}, \\ \text{probable range: } &2.3 \text{ to } 2.5 \text{ s}.\end{aligned}\tag{2.1}$$

Here, the best estimate, 2.4 s, lies at the midpoint of the estimated range of probable values, 2.3 to 2.5 s, as it has in all the examples. This relationship is obviously natural and pertains in most measurements. It allows the results of the measurement to be expressed in compact form. For example, the measurement of the time recorded in (2.1) is usually stated as follows:

$$\text{measured value of time} = 2.4 \pm 0.1 \text{ s}.\tag{2.2}$$

This single equation is equivalent to the two statements in (2.1).

In general, the result of any measurement of a quantity x is stated as

$(\text{measured value of } x) = x_{\text{best}} \pm \delta x.$

(2.3)

This statement means, first, that the experimenter's best estimate for the quantity concerned is the number x_{best} , and second, that he or she is reasonably confident the quantity lies somewhere between $x_{\text{best}} - \delta x$ and $x_{\text{best}} + \delta x$. The number δx is called the *uncertainty*, or *error*, or *margin of error* in the measurement of x . For convenience, the uncertainty δx is always defined to be positive, so that $x_{\text{best}} + \delta x$ is always the *highest* probable value of the measured quantity and $x_{\text{best}} - \delta x$ the *lowest*.

I have intentionally left the meaning of the range $x_{\text{best}} - \delta x$ to $x_{\text{best}} + \delta x$ somewhat vague, but it can sometimes be made more precise. In a simple measurement such as that of the height of a doorway, we can easily state a range $x_{\text{best}} - \delta x$ to $x_{\text{best}} + \delta x$ within which we are *absolutely* certain the measured quantity lies. Unfortunately, in most scientific measurements, such a statement is hard to make. In particular, to be *completely* certain that the measured quantity lies between $x_{\text{best}} - \delta x$ and $x_{\text{best}} + \delta x$, we usually have to choose a value for δx that is too large to be useful. To avoid this situation, we can sometimes choose a value for δx that lets us state with a certain percent confidence that the actual quantity lies within the range $x_{\text{best}} \pm \delta x$. For instance, the public opinion polls conducted during elections are traditionally stated with margins of error that represent 95% confidence limits. The statement that 60% of the electorate favor Candidate A, with a margin of error of 3 percentage points (60 ± 3), means that the pollsters are 95% confident that the percent of voters favoring Candidate A is between 57 and 63; in other words, after many elections, we should expect the correct answer to have been *inside* the stated margins of error 95% of the times and *outside* these margins only 5% of the times.

Obviously, we cannot state a percent confidence in our margins of error until we understand the statistical laws that govern the process of measurement. I return to this point in Chapter 4. For now, let us be content with defining the uncertainty δx so that we are "reasonably certain" the measured quantity lies between $x_{\text{best}} - \delta x$ and $x_{\text{best}} + \delta x$.

Quick Check¹ 2.1. (a) A student measures the length of a simple pendulum and reports his best estimate as 110 mm and the range in which the length probably lies as 108 to 112 mm. Rewrite this result in the standard form (2.3). (b) If another student reports her measurement of a current as $I = 3.05 \pm 0.03$ amps, what is the range within which I probably lies?

2.2 Significant Figures

Several basic rules for stating uncertainties are worth emphasizing. First, because the quantity δx is an estimate of an uncertainty, obviously it should not be stated

¹These "Quick Checks" appear at intervals through the text to give you a chance to check your understanding of the concept just introduced. They are straightforward exercises, and many can be done in your head. I urge you to take a moment to make sure you can do them; if you cannot, you should reread the preceding few paragraphs.

with too much precision. If we measure the acceleration of gravity g , it would be absurd to state a result like

$$(\text{measured } g) = 9.82 \pm 0.02385 \text{ m/s}^2. \quad (2.4)$$

The uncertainty in the measurement cannot conceivably be known to four significant figures. In high-precision work, uncertainties are sometimes stated with two significant figures, but for our purposes we can state the following rule:

Rule for Stating Uncertainties

Experimental uncertainties should almost always be rounded to one significant figure. (2.5)

Thus, if some calculation yields the uncertainty $\delta g = 0.02385 \text{ m/s}^2$, this answer should be rounded to $\delta g = 0.02 \text{ m/s}^2$, and the conclusion (2.4) should be rewritten as

$$(\text{measured } g) = 9.82 \pm 0.02 \text{ m/s}^2. \quad (2.6)$$

An important practical consequence of this rule is that many error calculations can be carried out mentally without using a calculator or even pencil and paper.

The rule (2.5) has only one significant exception. If the leading digit in the uncertainty δx is a 1, then keeping two significant figures in δx may be better. For example, suppose that some calculation gave the uncertainty $\delta x = 0.14$. Rounding this number to $\delta x = 0.1$ would be a substantial proportionate reduction, so we could argue that retaining two figures might be less misleading, and quote $\delta x = 0.14$. The same argument could perhaps be applied if the leading digit is a 2 but certainly not if it is any larger.

Once the uncertainty in a measurement has been estimated, the significant figures in the measured value must be considered. A statement such as

$$\text{measured speed} = 6051.78 \pm 30 \text{ m/s} \quad (2.7)$$

is obviously ridiculous. The uncertainty of 30 means that the digit 5 might really be as small as 2 or as large as 8. Clearly the trailing digits 1, 7, and 8 have no significance at all and should be rounded. That is, the correct statement of (2.7) is

$$\text{measured speed} = 6050 \pm 30 \text{ m/s}. \quad (2.8)$$

The general rule is this:

Rule for Stating Answers

The last significant figure in any stated answer should usually be of the same order of magnitude (in the same decimal position) as the uncertainty. (2.9)

For example, the answer 92.81 with an uncertainty of 0.3 should be rounded as

$$92.8 \pm 0.3.$$

If its uncertainty is 3, then the same answer should be rounded as

$$93 \pm 3,$$

and if the uncertainty is 30, then the answer should be

$$90 \pm 30.$$

An important qualification to rules (2.5) and (2.9) is as follows: To reduce inaccuracies caused by rounding, *any numbers to be used in subsequent calculations should normally retain at least one significant figure more than is finally justified.* At the end of the calculations, the final answer should be rounded to remove these extra, insignificant figures. An electronic calculator will happily carry numbers with far more digits than are likely to be significant in any calculation you make in a laboratory. Obviously, these numbers do not need to be rounded in the middle of a calculation but certainly must be rounded appropriately for the final answers.²

Note that the uncertainty in any measured quantity has the same dimensions as the measured quantity itself. Therefore, writing the units (m/s², cm³, etc.) after both the answer *and* the uncertainty is clearer and more economical, as in Equations (2.6) and (2.8). By the same token, if a measured number is so large or small that it calls for scientific notation (the use of the form 3×10^3 instead of 3,000, for example), then it is simpler and clearer to put the answer and uncertainty in the same form. For example, the result

$$\text{measured charge} = (1.61 \pm 0.05) \times 10^{-19} \text{ coulombs}$$

is much easier to read and understand in this form than it would be in the form

$$\text{measured charge} = 1.61 \times 10^{-19} \pm 5 \times 10^{-21} \text{ coulombs.}$$

Quick Check 2.2. Rewrite each of the following measurements in its most appropriate form:

(a) $v = 8.123456 \pm 0.0312 \text{ m/s}$

(b) $x = 3.1234 \times 10^4 \pm 2 \text{ m}$

(c) $m = 5.6789 \times 10^{-7} \pm 3 \times 10^{-9} \text{ kg.}$

2.3 Discrepancy

Before I address the question of how to use uncertainties in experimental reports, a few important terms should be introduced and defined. First, if two measurements

²Rule (2.9) has one more small exception. If the leading digit in the uncertainty is small (a 1 or, perhaps, a 2), retaining one extra digit in the final answer may be appropriate. For example, an answer such as 3.6 ± 1 is quite acceptable because one could argue that rounding it to 4 ± 1 would waste information.

of the same quantity disagree, we say there is a *discrepancy*. Numerically, we define the discrepancy between two measurements as their difference:

$$\text{discrepancy} = \text{difference between two measured values of the same quantity.} \quad (2.10)$$

More specifically, each of the two measurements consists of a best estimate and an uncertainty, and we define the discrepancy as the difference between the two best estimates. For example, if two students measure the same resistance as follows

Student A: 15 ± 1 ohms

and

Student B: 25 ± 2 ohms,

their discrepancy is

$$\text{discrepancy} = 25 - 15 = 10 \text{ ohms.}$$

Recognize that a discrepancy may or may not be *significant*. The two measurements just discussed are illustrated in Figure 2.1(a), which shows clearly that the discrepancy of 10 ohms is *significant* because no single value of the resistance is compatible with both measurements. Obviously, at least one measurement is incorrect, and some careful checking is needed to find out what went wrong.

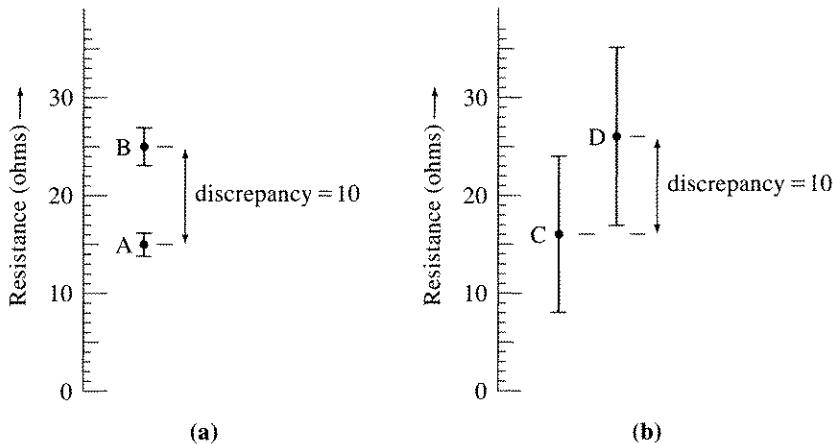


Figure 2.1. (a) Two measurements of the same resistance. Each measurement includes a best estimate, shown by a block dot, and a range of probable values, shown by a vertical error bar. The discrepancy (difference between the two best estimates) is 10 ohms and is *significant* because it is much larger than the combined uncertainty in the two measurements. Almost certainly, at least one of the experimenters made a mistake. (b) Two different measurements of the same resistance. The discrepancy is again 10 ohms, but in this case it is *insignificant* because the stated margins of error overlap. There is no reason to doubt either measurement (although they could be criticized for being rather imprecise).

Suppose, on the other hand, two other students had reported these results:

Student C: 16 ± 8 ohms

and

Student D: 26 ± 9 ohms.

Here again, the discrepancy is 10 ohms, but in this case the discrepancy is *insignificant* because, as shown in Figure 2.1(b), the two students' margins of error overlap comfortably and both measurements could well be correct. The discrepancy between two measurements of the same quantity should be assessed not just by its size but, more importantly, by how big it is *compared with the uncertainties in the measurements*.

In the teaching laboratory, you may be asked to measure a quantity that has been measured carefully many times before, and for which an accurate *accepted value* is known and published, for example, the electron's charge or the universal gas constant. This accepted value is not exact, of course; it is the result of measurements and, like all measurements, has some uncertainty. Nonetheless, in many cases the accepted value is much more accurate than you could possibly achieve yourself. For example, the currently accepted value of the universal gas constant R is

$$(\text{accepted } R) = 8.31451 \pm 0.00007 \text{ J/(mol} \cdot \text{K)}. \quad (2.11)$$

As expected, this value *is* uncertain, but the uncertainty is extremely small by the standards of most teaching laboratories. Thus, when you compare your measured value of such a constant with the accepted value, you can usually treat the accepted value as exact.³

Although many experiments call for measurement of a quantity whose accepted value is known, few require measurement of a quantity whose *true* value is known.⁴ In fact, the true value of a measured quantity can almost *never* be known exactly and is, in fact, hard to define. Nevertheless, discussing the difference between a measured value and the corresponding true value is sometimes useful. Some authors call this difference the *true error*.

2.4 Comparison of Measured and Accepted Values

Performing an experiment without drawing some sort of conclusion has little merit. A few experiments may have mainly qualitative results—the appearance of an interference pattern on a ripple tank or the color of light transmitted by some optical system—but the vast majority of experiments lead to *quantitative* conclusions, that is, to a statement of numerical results. It is important to recognize that the statement of *a single measured number is completely uninteresting*. Statements that the density

³This is not always so. For example, if you look up the refractive index of glass, you find values ranging from 1.5 to 1.9, depending on the composition of the glass. In an experiment to measure the refractive index of a piece of glass whose composition is unknown, the accepted value is therefore no more than a rough guide to the expected answer.

⁴Here is an example: If you measure the ratio of a circle's circumference to its diameter, the true answer is exactly π . (Obviously such an experiment is rather contrived.)

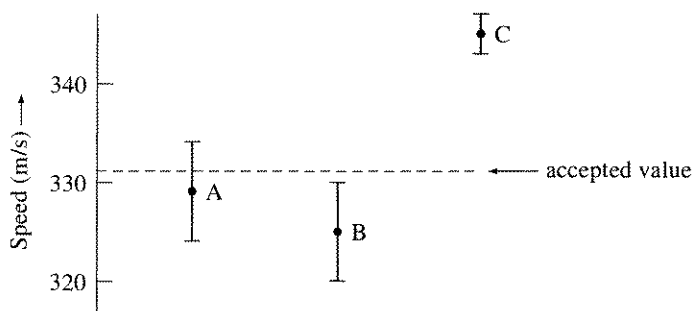


Figure 2.2. Three measurements of the speed of sound at standard temperature and pressure. Because the accepted value (331 m/s) is within Student A's margins of error, her result is satisfactory. The accepted value is just outside Student B's margin of error, but his measurement is nevertheless acceptable. The accepted value is *far outside* Student C's stated margins, and his measurement is definitely unsatisfactory.

of some metal was measured as 9.3 ± 0.2 gram/cm³ or that the momentum of a cart was measured as 0.051 ± 0.004 kg·m/s are, by themselves, of no interest. An interesting conclusion must *compare two or more numbers*: a measurement with the accepted value, a measurement with a theoretically predicted value, or several measurements, to show that they are related to one another in accordance with some physical law. It is in such comparison of numbers that error analysis is so important. This and the next two sections discuss three typical experiments to illustrate how the estimated uncertainties are used to draw a conclusion.

Perhaps the simplest type of experiment is a measurement of a quantity whose accepted value is known. As discussed, this exercise is a somewhat artificial experiment peculiar to the teaching laboratory. The procedure is to measure the quantity, estimate the experimental uncertainty, and compare these values with the accepted value. Thus, in an experiment to measure the speed of sound in air (at standard temperature and pressure), Student A might arrive at the conclusion

$$\text{A's measured speed} = 329 \pm 5 \text{ m/s}, \quad (2.12)$$

compared with the

$$\text{accepted speed} = 331 \text{ m/s}. \quad (2.13)$$

Student A might choose to display this result graphically as in Figure 2.2. She should certainly include in her report both Equations (2.12) and (2.13) next to each other, so her readers can clearly appreciate her result. She should probably add an explicit statement that because the accepted value lies inside her margins of error, her measurement seems satisfactory.

The meaning of the uncertainty δx is that the correct value of x *probably* lies between $x_{\text{best}} - \delta x$ and $x_{\text{best}} + \delta x$; it is certainly *possible* that the correct value lies slightly outside this range. Therefore, a measurement can be regarded as satisfactory even if the accepted value lies slightly outside the estimated range of the measured

value. For example, if Student B found the value

$$B's \text{ measured speed} = 325 \pm 5 \text{ m/s},$$

he could certainly claim that his measurement is consistent with the accepted value of 331 m/s.

On the other hand, if the accepted value is well outside the margins of error (the discrepancy is appreciably more than twice the uncertainty, say), there is reason to think something has gone wrong. For example, suppose the unlucky Student C finds

$$C's \text{ measured speed} = 345 \pm 2 \text{ m/s} \quad (2.14)$$

compared with the

$$\text{accepted speed} = 331 \text{ m/s}. \quad (2.15)$$

Student C's discrepancy is 14 m/s, which is seven times bigger than his stated uncertainty (see Figure 2.2). He will need to check his measurements and calculations to find out what has gone wrong.

Unfortunately, the tracing of C's mistake may be a tedious business because of the numerous possibilities. He may have made a mistake in the measurements or calculations that led to the answer 345 m/s. He may have estimated his uncertainty incorrectly. (The answer 345 ± 15 m/s would have been acceptable.) He also might be comparing his measurement with the wrong accepted value. For example, the accepted value 331 m/s is the speed of sound at standard temperature and pressure. Because standard temperature is 0°C, there is a good chance the measured speed in (2.14) was *not* taken at standard temperature. In fact, if the measurement was made at 20°C (that is, normal room temperature), the correct accepted value for the speed of sound is 343 m/s, and the measurement would be entirely acceptable.

Finally, and perhaps most likely, a discrepancy such as that between (2.14) and (2.15) may indicate some undetected source of systematic error (such as a clock that runs consistently slow, as discussed in Chapter 1). Detection of such systematic errors (ones that consistently push the result in one direction) requires careful checking of the calibration of all instruments and detailed review of all procedures.

2.5 Comparison of Two Measured Numbers

Many experiments involve measuring two numbers that theory predicts should be equal. For example, the law of conservation of momentum states that the total momentum of an isolated system is constant. To test it, we might perform a series of experiments with two carts that collide as they move along a frictionless track. We could measure the total momentum of the two carts before (p) and after (q) they collide and check whether $p = q$ within experimental uncertainties. For a single pair of measurements, our results could be

$$\text{initial momentum } p = 1.49 \pm 0.03 \text{ kg}\cdot\text{m/s}$$

and

$$\text{final momentum } q = 1.56 \pm 0.06 \text{ kg}\cdot\text{m/s}.$$

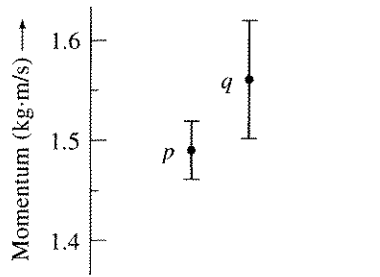


Figure 2.3. Measured values of the total momentum of two carts before (p) and after (q) a collision. Because the margins of error for p and q overlap, these measurements are certainly consistent with conservation of momentum (which implies that p and q should be equal).

Here, the range in which p probably lies (1.46 to 1.52) *overlaps* the range in which q probably lies (1.50 to 1.62). (See Figure 2.3.) Therefore, these measurements are consistent with conservation of momentum. If, on the other hand, the two probable ranges were not even close to overlapping, the measurements would be inconsistent with conservation of momentum, and we would have to check for mistakes in our measurements or calculations, for possible systematic errors, and for the possibility that some external forces (such as gravity or friction) are causing the momentum of the system to change.

If we repeat similar pairs of measurements several times, what is the best way to display our results? First, using a table to record a sequence of similar measurements is usually better than listing the results as several distinct statements. Second, the uncertainties often differ little from one measurement to the next. For example, we might convince ourselves that the uncertainties in all measurements of the initial momentum p are about $\delta p \approx 0.03$ kg·m/s and that the uncertainties in the final q are all about $\delta q \approx 0.06$ kg·m/s. If so, a good way to display our measurements would be as shown in Table 2.1.

Table 2.1. Measured momenta (kg·m/s).

Trial number	Initial momentum p (all ± 0.03)	Final momentum q (all ± 0.06)
1	1.49	1.56
2	3.10	3.12
3	2.16	2.05
etc.		

For each pair of measurements, the probable range of values for p overlaps (or nearly overlaps) the range of values for q . If this overlap continues for all measurements, our results can be pronounced consistent with conservation of momentum. Note that our experiment does not *prove* conservation of momentum; no experiment can. The best you can hope for is to conduct many more trials with progressively

smaller uncertainties and that all the results are *consistent* with conservation of momentum.

In a real experiment, Table 2.1 might contain a dozen or more entries, and checking that each final momentum q is consistent with the corresponding initial momentum p could be tedious. A better way to display the results would be to add a fourth column that lists the differences $p - q$. If momentum is conserved, these values should be consistent with zero. The only difficulty with this method is that we must now compute the uncertainty in the difference $p - q$. This computation is performed as follows. Suppose we have made measurements

$$(\text{measured } p) = p_{\text{best}} \pm \delta p$$

and

$$(\text{measured } q) = q_{\text{best}} \pm \delta q.$$

The numbers p_{best} and q_{best} are our best estimates for p and q . Therefore, the best estimate for the difference $(p - q)$ is $(p_{\text{best}} - q_{\text{best}})$. To find the uncertainty in $(p - q)$, we must decide on the highest and lowest probable values of $(p - q)$. The highest value for $(p - q)$ would result if p had its *largest* probable value, $p_{\text{best}} + \delta p$, at the same time that q had its *smallest* value $q_{\text{best}} - \delta q$. Thus, the highest probable value for $p - q$ is

$$\text{highest probable value} = (p_{\text{best}} - q_{\text{best}}) + (\delta p + \delta q). \tag{2.16}$$

Similarly, the lowest probable value arises when p is smallest ($p_{\text{best}} - \delta p$), but q is largest ($q_{\text{best}} + \delta q$). Thus,

$$\text{lowest probable value} = (p_{\text{best}} - q_{\text{best}}) - (\delta p + \delta q). \tag{2.17}$$

Combining Equations (2.16) and (2.17), we see that the *uncertainty in the difference* $(p - q)$ is the *sum* $\delta p + \delta q$ of the *original uncertainties*. For example, if

$$p = 1.49 \pm 0.03 \text{ kg}\cdot\text{m/s}$$

and

$$q = 1.56 \pm 0.06 \text{ kg}\cdot\text{m/s},$$

then

$$p - q = -0.07 \pm 0.09 \text{ kg}\cdot\text{m/s}.$$

We can now add an extra column for $p - q$ to Table 2.1 and arrive at Table 2.2.

Table 2.2. Measured momenta (kg·m/s).

Trial number	Initial p (all ± 0.03)	Final q (all ± 0.06)	Difference $p - q$ (all ± 0.09)
1	1.49	1.56	-0.07
2	3.10	3.12	-0.02
3	2.16	2.05	0.11
etc.			

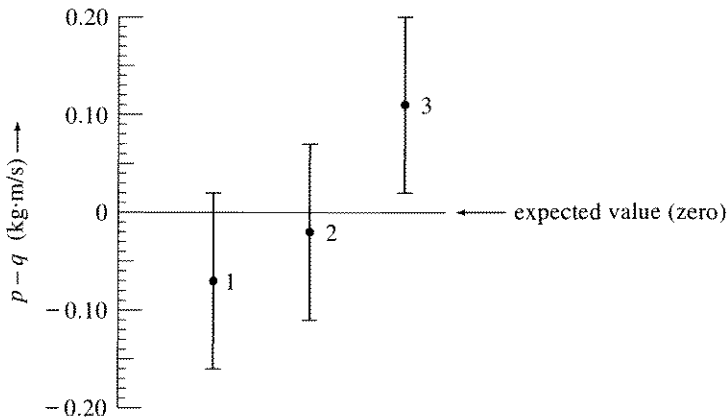


Figure 2.4. Three trials in a test of the conservation of momentum. The student has measured the total momentum of two carts before and after they collide (p and q , respectively). If momentum is conserved, the differences $p - q$ should all be zero. The plot shows the value of $p - q$ with its error bar for each trial. The expected value 0 is inside the margins of error in trials 1 and 2 and only slightly outside in trial 3. Therefore, these results are consistent with the conservation of momentum.

Whether our results are consistent with conservation of momentum can now be seen at a glance by checking whether the numbers in the final column are consistent with zero (that is, are less than, or comparable with, the uncertainty 0.09). Alternatively, and perhaps even better, we could plot the results as in Figure 2.4 and check visually. Yet another way to achieve the same effect would be to calculate the *ratios* q/p , which should all be consistent with the expected value $q/p = 1$. (Here, we would need to calculate the uncertainty in q/p , a problem discussed in Chapter 3.)

Our discussion of the uncertainty in $p - q$ applies to the difference of any two measured numbers. If we had measured any two numbers x and y and used our measured values to compute the difference $x - y$, by the argument just given, the resulting uncertainty in the difference would be the *sum* of the separate uncertainties in x and y . We have, therefore, established the following provisional rule:

**Uncertainty in a Difference
(Provisional Rule)**

If two quantities x and y are measured with uncertainties δx and δy , and if the measured values x and y are used to calculate the difference $q = x - y$, the *uncertainty in q* is the *sum of the uncertainties in x and y* :

$$\delta q \approx \delta x + \delta y. \quad (2.18)$$

I call this rule “provisional” because we will find in Chapter 3 that the uncertainty in the quantity $q = x - y$ is often somewhat smaller than that given by Equation

(2.18). Thus, we will be replacing the provisional rule (2.18) by an “improved” rule—in which the uncertainty in $q = x - y$ is given by the so-called quadratic sum of δx and δy , as defined in Equation (3.13). Because this improved rule gives a somewhat smaller uncertainty for q , you will want to use it when appropriate. For now, however, let us be content with the provisional rule (2.18) for three reasons: (1) The rule (2.18) is easy to understand—much more so than the improved rule of Chapter 3. (2) In most cases, the difference between the two rules is small. (3) The rule (2.18) always gives an upper bound on the uncertainty in $q = x - y$; thus, we know at least that the uncertainty in $x - y$ is never worse than the answer given in (2.18).

The result (2.18) is the first in a series of rules for the *propagation of errors*. To calculate a quantity q in terms of measured quantities x and y , we need to know how the uncertainties in x and y “propagate” to cause uncertainty in q . A complete discussion of error propagation appears in Chapter 3.

Quick Check 2.3. In an experiment to measure the latent heat of ice, a student adds a chunk of ice to water in a styrofoam cup and observes the change in temperature as the ice melts. To determine the mass of ice added, she weighs the cup of water before and after she adds the ice and then takes the difference. If her two measurements were

$$(\text{mass of cup \& water}) = m_1 = 203 \pm 2 \text{ grams}$$

and

$$(\text{mass of cup, water, \& ice}) = m_2 = 246 \pm 3 \text{ grams,}$$

find her answer for the mass of ice, $m_2 - m_1$, with its uncertainty, as given by the provisional rule (2.18).

2.6 Checking Relationships with a Graph

Many physical laws imply that one quantity should be proportional to another. For example, Hooke’s law states that the extension of a spring is proportional to the force stretching it, and Newton’s law says that the acceleration of a body is proportional to the total applied force. Many experiments in a teaching laboratory are designed to check this kind of proportionality.

If one quantity y is proportional to some other quantity x , a graph of y against x is a straight line through the origin. Thus, to test whether y is proportional to x , you can plot the measured values of y against those of x and note whether the resulting points do lie on a straight line through the origin. Because a straight line is so easily recognizable, this method is a simple, effective way to check for proportionality.

To illustrate this use of graphs, let us imagine an experiment to test Hooke’s law. This law, usually written as $F = kx$, asserts that the extension x of a spring is proportional to the force F stretching it, so $x = F/k$, where k is the “force constant”