www.iop.org/journals/physed

Teaching uncertainties

Ian Duerdoth

School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK

E-mail: ian.duerdoth@manchester.ac.uk

Abstract

The subject of uncertainties (sometimes called errors) is traditionally taught (to first-year science undergraduates) towards the end of a course on statistics that defines probability as the limit of many trials, and discusses probability distribution functions and the Gaussian distribution. We show how to introduce students to the concepts of uncertainty based on the idea of degree of belief. This enables them to move on more quickly to the important problems actually met in laboratory work, namely the estimation of uncertainties and their propagation. We also consider the mean and the weighted mean of several results and the method of least squares.

Introduction

Many students find the subject of measurement errors (more meaningfully called uncertainties) difficult and confusing. We suggest why this is so and make some suggestions.

Errors and uncertainty

First of all, the word *error* is itself confusing because it has at least three meanings. Often it means a *mistake*, as in 'I made an error; I omitted the factor 2π to convert from frequency to angular frequency'. Secondly, it often means a *difference*. Having measured g as 9.814 ± 0.003 m s^{-2} and found a book value to be 9.816, one says 'The error of my result is 0.002 m s^{-2}'. Lastly, it often means *uncertainty*. 'The error on my value of g is $\pm$ 0.003'. So the first step in reducing the confusion, and helping students to learn, is to not use the word error, but to be more explicit. In the literature there seems to have been a slow drift to using the word uncertainty where previously the word error (meaning uncertainty) had been used. The statement that $g = 9.814 \pm 0.003$ m s^{-2} means that the uncertainty is $\pm$ 0.003 and that we are 68% sure that the true value lies between 9.811 and 9.817. This article is concerned with uncertainties.

Traditional teaching

Many students are introduced to the notions of uncertainty that are needed for laboratory work by first being presented with a course on statistics. This typically covers probability as the limit of many trials, then the distributions of measurements, that these are Gaussian, evaluation of the mean, and evaluation of the standard deviation and the standard deviation of the mean. Laboratory measurements are seen as a small sample of an underlying Gaussian distribution and the spread is in some way caused by a host of small uncontrollable random effects. The uncertainty may be estimated from the spread of the measurements. Some workers argue that the spread, the uncertainty of the measurement, is intrinsic to the equipment, and/or to the technique, just as the true value is there to be measured. Others point out that the uncertainty of the measurement should also include variations due to the skill of the experimenter. Faced with this build up, students shy away from uncertainties, thinking they are about nasty summation signs (with infinite limits) and probability distribution functions (pdfs). This approach has several other weaknesses.

 0031-9120/09/020138+07$30.00

It would be helpful if students could be shown the distribution of a large set of measurements of the same thing taken by the same experimenter under (supposedly) identical conditions that illustrated the Gaussian distribution. How about 100 measurements of the mass of a marble or of the time of 10 swings of a pendulum? In many cases the distribution does not look at all Gaussian. Consider reading the temperature from a digital thermometer and getting 25.2 °C time and time again with no change. For better or worse, text books and lecture courses very rarely show real data that illustrate the Gaussian distribution. We take here the practical point of view that the distributions are often not Gaussian at all. After all, digital instruments may give distributions containing only one or two values; counting experiments, such as rates from radioactive sources, give Poisson distributions; and dice throwing experiments give binomial distributions.

A search of the literature threw up few examples. In one, see [1] (in fact one that comes close to presenting convincing data), a histogram is presented of 174 measurements of the outer radius of a cylindrical annulus—taken by 174 students. The distribution is consistent with a Gaussian, but other distributions cannot be excluded. Surely the differing skills of the experimenters contributes to the spread, and as no details are given one suspects that the annulus may be elliptical, and different radii are being measured. This hardly supports the classical picture of many small random effects combining to give the uncertainty.

In practice it is easier to obtain distributions that look Gaussian by evaluating a quantity that depends on several measurements. For instance, see [1], the distribution of the volume of the annulus (obtained from measurements of the inner radius, outer radius, and length) may appear Gaussian. Of course here the central limit theorem is coming into play.

The discussions tend to concentrate on random uncertainties, but surely systematic effects must come into the description too. The inclusion of uncertainty due to any unknown systematic effects is often glossed over. How well was the voltmeter calibrated?

The subject of measurement uncertainties (then called errors) was developed in an era of analogue instruments: mercury in glass thermometers, moving coil voltmeters with needles and dials, metre rules, hand-held mechanical stop watches, cathode-ray tube displays, ... In such a world, the concept of many small unknown random contributions giving rise to a Gaussian distribution by way of the central limit theorem seemed quite sensible. However, technology has moved on, and instrumentation is now dominated by digital techniques. The experimenter is generally more remote from the raw measurement. He is more of an observer and sees a digital display. The data may have been digitally processed in sophisticated ways (for instance they may have already averaged over several samples). As far as the experimenter is concerned, the uncertainty of the measurement is not obtained by considering the basic technique, but is determined by the instrument manufacturer, and estimating uncertainty is a matter of reading the specification and having confidence in the equipment.

Uncertainty as degree of belief

The purpose of this article is to present an alternative approach to teaching uncertainties. Our approach is conceptually much simpler, less mathematical, and more general. It enables students to estimate uncertainties sensibly, overcome their reluctance to consider uncertainties, and to move on quickly to other important and essential experimental issues such as propagating the uncertainties.

We ignore all the statistics, means, standard deviations, Gaussians, etc, and give a simple description of what we mean by uncertainty. It would be too presumptuous to call it a definition.

The essential points are that we want to specify a range in which we think the true value lies, i.e. to quote $x \pm \sigma_x$. We could try for a range corresponding to complete certainty, but in scientific measurements we can never be absolutely sure, so that is not realistic. So we want a statement of the form—‘I am $S\%$ sure that the true value lies in the range $x - \sigma_x$ to $x + \sigma_x$’. The emphasis here is on being $S\%$ sure, implying that I think there is a $(100 - S)\%$ chance that the true value is outside the range. All this amounts to a ‘degree of belief’. We find that students have no problem with this approach. Indeed it is in line with the ISO guidelines [2].

We now ask what would be a sensible value for S, the degree of belief? The ISO guidelines imply that you can choose any value. They use the term *coverage*. Choose it according to the nature and details of the measurement. They recommend using the symbol u for the uncertainty. An obvious choice is to take $S = 50\%$, and this is indeed perfectly sensible. In the scientific community there is a strong preference for standardizing on the value $S \simeq 68\%$ (which in practice can be taken as two thirds) because with this particular value the uncertainty is the same as the root mean square (rms) deviation (in cases where there is a large sample). If the underlying distribution of (repeated) measurements is in fact Gaussian, then the uncertainty defined in this way is the rms width parameter of the Gaussian. This is also called the SD (standard deviation). Using $S \simeq 68\%$ does not imply that the distribution is in fact Gaussian. But there is an inevitable association in people's minds. Of course the concept of an underlying distribution is not strictly speaking logically consistent with the concept of a measurement that is done once. We should also point out that relating the range for 95% sure with that for 68% sure (see section 'Estimating an uncertainty') does make an assumption about the shape of the distribution. One might consider using u for uncertainties corresponding to an arbitrary range, and using σ for the specific case of 68%.

Concepts of probability

Students are quite happy with the basic concept of probability. The probability is 1/6 that a die falls displaying 4. The probability of drawing a king from a pack of cards is 4/52. An extension is to plot a histogram of the frequency or probability of drawing a ball of a given colour from a bag containing 12 red balls, 10 blue, 5 yellow, and 13 green. These cause no problems.

They are also content with statements such as 'The probability that it will rain tomorrow is 50%' and 'There is a 1/365 probability that it's Jack's birthday today'. (This even though the statement 'It is Jack's birthday today' is either true or false.)

When confronted with 'The result of your measurement is 4.3 V. How well do you believe your value?' they do not question the phrase 'believe your value'. They respond (perhaps after some thought): 'Well it could be out by say ± 0.2 V, so I would say there is a 50% chance (i.e. probability) that the true value is between 4.1 and 4.5 V'.

The above are different examples of the concept of probability—the examples of the die, cards, and balls might be termed propensity, and the others Bayesian. There are other concepts of probability including a formal mathematical definition and the classical (also called frequentist) concept (see below). Much has been written and discussions continue on the different concepts of probability and on their merits and difficulties. It is not the purpose of this article to advocate any of the concepts over another, but to observe that students are familiar with the ideas mentioned above, even though they may not be altogether logically consistent. We also note that some people do take strong positions and favour one particular concept.

The classical or frequentist definition of probability is given by the fraction of trials that are deemed successful, in the limit of an infinite number of trials:

$$P = \lim_{N\to\infty} \frac{N_s}{N}, \qquad (1)$$

where N is the number of trials and N_s is the number that are successful.

Strictly speaking, the true value has a fixed value in the frequentist viewpoint. So the notion that there is a probability that the true value lies between some limits is illogical. However, few people balk at such statements—for instance the statement at the end of the section 'Errors and uncertainty'.

This concept is often taught in a course on statistics and precedes the introduction of uncertainties, their estimation, and propagation. We suggest that uncertainty (as needed for laboratory work) does *not* need this material. It is better to build on the more general concept of probability that students already have, namely degree of belief, and to move on to the more relevant topics of estimation and propagation.

Estimating an uncertainty

We consider raw measurements. The traditional approach encourages students to estimate uncertainties by imagining what would happen if the measurement were to be repeated many times. They expect the measurements to have a spread, for this to be Gaussian, and the root mean square width to be the required uncertainty. This is quite a challenge! It emphasizes random uncertainties and tends to ignore systematic ones. Relating the uncertainty to a degree of belief opens up other more fruitful ways. A simple and effective technique is to zoom in and out. Suppose the length of a sheet of paper is measured as 296.7 mm. Ask 'Do you believe this to ±10 mm, ±5, ±2, ±1, ±0.5, ±0.2, ±0.1, ±0.05, ±0.02, ±0.01, ...?' Usually this pins down the uncertainty to a factor of about 2 or so, and this is often all that is needed. It is implicit that we are talking about a 2/3 certainty (68%) for the range. Another useful method is to ask 'Over what range are you almost certain the true value must lie?' Taking 'almost certain' to correspond to 95% certainty, the range corresponds to four times the uncertainty. (This correspondence makes an assumption about the probability distribution function.) If other students are doing similar experiments, a useful way to assess the degree of belief is to ask (for instance) 'If they got 297.2 mm, would you accept that as being within the claimed accuracy or would you challenge them?'

Notice that these techniques encourage the inclusion of systematic effects (as well as random ones) right from the beginning. They also demonstrate that knowing the value of an uncertainty to a factor 2 is already very useful. Indeed, it is rarely meaningful to give an uncertainty to better than 10% (of itself), and there are few instances where the uncertainty is meaningful to better than 1% (of itself). So quoting an uncertainty to more than two or three significant figures shows a lack of understanding.

Propagation of uncertainties

Students are mystified by the propagation of uncertainties, even after attending a course on the topic. Often, perhaps in desperation, they quote the general expression for propagation (see equation (6) in the next section) and try to differentiate and substitute, and all too often end up with a nonsensical result. Because of their lack of understanding they often do not realize that their result is obviously wrong and may be surprised that a demonstrator sees this at a glance. We propose a two-stage approach that is conceptually easy. Step one is the propagation of a single quantity and step two is combination in quadrature.

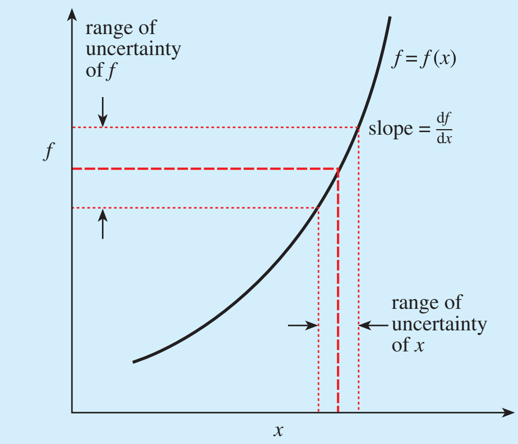

Figure 1. Propagation of the uncertainty from x to f for an arbitrary function $f = f(x)$. The range of the uncertainty for x maps onto the range for f. To good approximation the ratio is the slope, $|\frac{df}{dx}|$. Usually, the ranges are chosen to correspond to a coverage of 68%.

Propagation of uncertainties: I. A single parameter

Suppose a quantity x is measured and we have assigned an uncertainty, σ_x, to it. We are S% sure that the true value lies between $x_o - \sigma_x$ and $x_o + \sigma_x$. We might have chosen S to be 68% (or any other value). We want a quantity f that depends on x and we know the function $f(x)$. For instance, we have measured the speed v and want the energy E, where $E = \frac{1}{2}mv^2$. It is straightforward to evaluate the value, f_o, of $f(x)$ that corresponds to the value x_o. Now we want the range of values of f in which we are S% sure that its true value lies. This is easy: evaluate $f(x)$ at $x = x_o - \sigma_x$ and at $x = x_o + \sigma_x$, giving

$$f_+ = f(x_o + \sigma_x) \qquad f_o = f(x_o)$$
$$f_- = f(x_o - \sigma_x).$$

So it must be that we are S% sure that the true value of f lies in the range f_- to f_+. This is illustrated in figure 1. A very good way of determining f_- and f_+ is to evaluate them numerically (for instance by using a calculator). In many cases we find that $(f_+ - f_o)$ is approximately equal to $(f_o - f_-)$, so calling (the modulus of) this σ_{fx}, the result for f is $f_o \pm \sigma_{fx}$. If they are not quite equal, it is usually acceptable to take the average. They are equal if σ_x is so small that $f(x)$ may be considered a straight line over the range $x_o - \sigma_x$ to $x_o + \sigma_x$. Treating them as equal corresponds to making a linear approximation. Most calculators have a precision of ten decimal places, so there is usually no problem in evaluating the difference between two nearly equal numbers.

Examination of figure 1 shows that, provided the linear approximation is good, the ratio of the two ranges is just the slope (the differential) of the function $f(x)$, i.e.

$$\sigma_{fx} \simeq \left|\frac{\mathrm{d}f}{\mathrm{d}x}\right|\sigma_x. \tag{2}$$

We have used the symbol σ_{fx} with the subscript fx to stress that it is the uncertainty in f due to the uncertainty in x. It is necessary to take the modulus of the differential because it might be negative. By definition all uncertainties are positive.

Many of the functions met with in the laboratory are simple and easily differentiated. It is then sensible to use equation (2). An important special case is that, if $f = Ax^n$, then

$$\left(\frac{\sigma_{fx}}{f}\right)^2 \simeq n^2\left(\frac{\sigma_x}{x}\right)^2. \tag{3}$$

The specific cases of $n = \pm 1$, i.e. $f = Ax$ and $f = A/x$, are essential, as are $n = \pm\frac{1}{2}$, etc.

If the function $f(x)$ is too complicated to easily differentiate there is no problem. Instead of persevering with the differentiation, either propagate the uncertainty by direct calculation or look for an intermediate quantity. Direct calculation, numerically using a calculator, was not a realistic technique before the development of low-cost digital equipment (see section 'Traditional teaching'). Intermediate quantities (sometimes called step-by-step) are best illustrated by an example. Suppose $f(x) = \frac{A}{x+B}$, where A and B are constants. Instead of differentiating (not so very difficult in this simple example), let $y = x + B$ and let $f = A/y$. It is now easy to propagate the uncertainty from x to y and then propagate from y to f.

Note that propagation is not dependent on the underlying distributions taking any particular form (for instance Gaussian), nor on the concept of the spread of many measurements. So the teaching of probability distribution functions, the Gaussian, and rms spread can sensibly be delayed until after propagation has been mastered. Our procedure emphasizes that propagation using equation (2) entails a linear approximation.

Propagation of uncertainties: II. Quadrature

Having established how to propagate an uncertainty for just one parameter, $f(x)$, we now consider the important, indeed essential, step of several parameters, $f(x, y, \ldots)$. This is often needed in laboratory work, and the prescription is simple. The uncertainties combine in *quadrature*. In terms of the uncertainties for single parameters the required relation is

$$\sigma_f^2 \simeq \sigma_{fx}^2 + \sigma_{fy}^2 + \cdots. \tag{4}$$

Here σ_f is the uncertainty in f arising from all the contributions. (The derivative in equation (2) should now be regarded as a partial.) It is straightforward to evaluate, given the separate contributions. This approach encourages students to tabulate the separate contributions and thus to become aware of those which are the most significant. It needs to be stressed that it is the contributions, σ_{fx}, that combine in quadrature (not the individual uncertainties, σ_x.)

In a first introduction, quadrature can be presented as a prescription to be justified later. A plausibility argument considers the possibilities that the differences (deviations), $\Delta x = x - x_{\mathrm{T}}$ and $\Delta y = y - y_{\mathrm{T}}$, may both be positive, both negative, or have opposite signs. Thus the contributions to the value of f sometimes augment each other and sometimes cancel. Here $x_{\mathrm{T}} \ldots$ are the true values.

A slightly more rigorous justification for combining in quadrature is given in the standard texts; see for instance [3]. Essentially, one considers the difference, Δx, between the measured and true values for x, and for y, etc, and also for f. A linear approximation is made for Δf, and then one considers what happens on average if the experiment were repeated many many times. The quantity $(\Delta f)^2$ has a contribution from the products $\Delta x \Delta y$. This is zero on average, provided the measurements are independent (i.e. uncorrelated). Identifying $\langle(\Delta x)^2\rangle$ as proportional to σ_x^2, etc, gives equation (6). The angle brackets $\langle\ \rangle$, indicate the average over infinitely many experiments. The derivation again entails approximations, so equation (4) is not exact.

Note that the justification of equation (4) does not depend on the deviations having any particular distribution (probability distribution function), let alone it being Gaussian. However, it does assume that all the uncertainties correspond to the same coverage (usually taken to be 68%).

Referring to equation (3), it is clear that, if say $\sigma_{fx} > 5\sigma_{fy}$, then σ_{fy} should be ignored, as when squared it makes a negligible contribution.

An important special case is $f = Ax^n y^m / z^l$; for this, equations (2) and (4) give

$$\left(\frac{\sigma_f}{f}\right)^2 \simeq n^2\left(\frac{\sigma_x}{x}\right)^2 + m^2\left(\frac{\sigma_y}{y}\right)^2 + l^2\left(\frac{\sigma_z}{z}\right)^2. \tag{5}$$

We stress this relationship because it is so important to move students on from *absolute* uncertainties to thinking in terms of *fractional* uncertainties. That fractional uncertainties combine in quadrature for this case is possibly the most important relationship for students to learn.

Let us contrast this with the more traditional approach. In this students are typically presented with the Gaussian distribution, rms deviations, etc, before meeting the general formula for the propagation of uncertainties:

$$\sigma_f^2 \simeq \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial f}{\partial z}\right)^2 \sigma_z^2 + \cdots. \tag{6}$$

Of course, this is just equations (2) and (4) combined. Unfortunately, students seem to latch onto this equation—after all it is valid for all functions. They differentiate the algebraic function (whether simple or complicated) and substitute a host of numbers and get a result. There is a lack of appreciation of what is happening, and because of this lack of transparency there is no way of knowing if the result is sensible or not. All too often it is crazy! The proposed two-step approach enables the propagation to be done numerically or by using intermediate parameters (as well as by differentiating). It is easier to check that the final result is sensible—it must be bigger than, but not much bigger than, the largest single contribution.

Uncertainty of the mean of several measurements

It is obvious that the uncertainty of the mean of several measurements is less than the uncertainty of each individual measurement, and it is usually good practice to repeat a measurement a few times and take the average, and students often need to do this in their laboratory work. The question that now arises is how well to believe the mean. We assume that the uncertainty on the individual measurements has been estimated (see for instance section ‘Estimating an uncertainty’) and now point out that it is easy to evaluate the uncertainty of the mean, without reference to any statistical arguments. The mean of N measurements, $x_1, x_2, \ldots, x_N$, is defined by

$$m = \frac{x_1 + x_2 + \cdots x_N}{N}. \tag{7}$$

The uncertainty σ_m of the mean is given by propagation of uncertainties, equations (2) and (4). We assume that the measurements are independent and that each of the x_i has the same uncertainty, σ. The uncertainty in the mean, equation (7), due to each individual measurement is

$$\sigma_{mx} = \frac{1}{N}\sigma, \tag{8}$$

so combining in quadrature (for N measurements) we get

$$\sigma_m = \frac{\sigma}{\sqrt{N}}. \tag{9}$$

The fact that uncertainties decrease as the square root of the number of measurements is a useful result, which is quite general, and which is often needed in laboratory work. The derivation shows that it applies only to the random uncertainties.

Uncertainty from the spread of several measurements

An alternative way to estimate the uncertainty of individual measurements is to evaluate the spread. The best estimate of σ^2 is

$$s^2 = \frac{1}{N-1}\sum (x_i - m)^2. \tag{10}$$

In an introductory course it can be stated that this is for a coverage of 68%, and this can be justified in a subsequent course on statistics. But notice that there is no mention of limits of infinite sums or of probability distributions. Of course, the uncertainty estimated in this way ignores all systematic effects. If there is only one data point ($N = 1$), then there is no estimate of the spread, so the factor $N - 1$ has some plausibility.

When students first meet the method of least squares for fitting a straight line to data points, it is usually done without reference to the Gaussian distribution even though it is related to the justification of the method. We note here that the uncertainty of the data points (using conventional notation) is that the best estimate of σ^2 is given by

$$s^2 = \frac{1}{N-1}\sum(ax_i - y_i)^2 \qquad \text{for } y = ax \tag{11}$$

and

$$s^2 = \frac{1}{N-2}\sum(ax_i + b - y_i)^2 \quad \text{for } y = ax + b. \tag{12}$$

These have the same structure as equation (10).

The formulae for the weighted mean of several measurements and its associated uncertainty can also be derived using least squares minimization and propagation of uncertainties, without reference to the Gaussian distribution.

Summary

In this article we have made the following suggestions. Avoid the word error. Use mistake, difference, or uncertainty, as appropriate.

Postpone the introduction of statistical concepts, in particular the Gaussian distribution, until the following have been mastered.

Introduce the concept of uncertainty as 'degree of belief', because this is compatible with students' concepts, includes systematic as well as random effects and the skill of experimenter, and is easy to evaluate.

Teach propagation of uncertainties in two steps: single parameter followed by quadrature, stressing that the results must be independent. (Do not give the general form, equation (6), because it is never needed.)

Present how the uncertainty of the mean of several results falls with N, equation (9), by using the propagation of uncertainties. Present the estimation of (random) uncertainty from the spread of several results by quoting expression (10), giving only plausibility arguments (for the factor $N - 1$). Likewise, give expressions for estimating uncertainties from the spread of data points from fitted lines using the least squares method.

Only later, introduce the statistical concepts of a probability distribution, the Gaussian distribution, the standard deviation, the central limit theorem, and uncertainties as (possibly) being the result of many small unknowable contributions. This is also the place to explain the origin of quadrature and the uncertainty of the mean.

Received 15 July 2008, in final form 20 August 2008
doi:10.1088/0031-9120/44/2/004

References

[1] Court G R, Donald R A and Fry J R 1976 Teaching practical physics. II *Phys. Educ.* **11** 488–93

[2] ISO 1993 *Guide to the Expression of Uncertainties in Measurement* (Geneva: International Organization of Standardization) Corrected and reprinted 1995

Taylor B N and Kuyatt C E 1994 Guidelines for evaluating and expressing the uncertainty of NIST measurement results *NIST Technical Note 1297* (Washington, DC: US Government Printing Office) See also http://physics.nist.gov/cuu

[3] Barlow R J 1989 *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences* (New York: Wiley)

Ian Duerdoth recently retired from the School of Physics and Astronomy, at The University of Manchester, UK. Besides developing and building detectors for the Atlas experiment at the LHC at CERN, one of his main interests remains teaching physics in small groups (tutorials and laboratory).