MEASUREMENT

In preceding Chapters we discussed the tactical problems associated with *describing* quantitative measurements, reminding ourselves that the tools we use (numbers, dimensions, units) are almost perfectly arbitrary in isolation but embody a functional or *relational* truth in the "grammar" of their use. Accepting these tools provisionally, we turn now to the far messier problem of actually *performing* measurements.

5.1 Tolerance

(Advertising Your Uncertainty)

Virtually all [I could follow the consensus and say *all*, but I feel like hedging] "scientific" procedures involve *measurement* of experimental parameters such as distance, time, velocity, mass, energy, temperature, … *etc.* Virtually all measurements are subject to *error*; that is, they may be *inaccurate* (wrong) by some unknown amount due to effects ranging from errors in recording ["I said 3.32, not 3.23!"] to miscalibrated instruments ["I thought these tic marks were *centimetres*!"]. Such "systematic errors" are embarrassing to the experimenter, as they imply poor technique, and are always hard to estimate; but we are honour-bound to try. An entirely different source of error that conveys *no* negative connotations on the experimenter is the fact that all measurements have limited *precision* or "tolerance" — limited by the "marks" on the generalized "ruler" used for measuring-by-comparison. (*E.g.*, the distance your measure with a micrometer is more precisely known than the distance you measure with a cloth tape measure.)

Knowing this, most scientists and virtually all physicists have an æsthetic about measured values of things: they are *never* to be reported without an explicit estimation of their *uncertainty*. That is, measurements must always be reported in the form

$$(\text{VALUE} \pm \text{UNCERTAINTY}) \text{ UNITS}$$

or equivalent notation (sometimes a shorthand version), such as 3.1416(12) radians, meaning (3.1416 ± 0.0012) radians. [The (12) means the uncertainty in the last two digits is $\pm$ 12.] This shorthand form is convenient for long strings of digits with only the last 1 or 2 digits uncertain, but the explicit form with the $\pm$ is more pleasing to the æsthetic mentioned above.

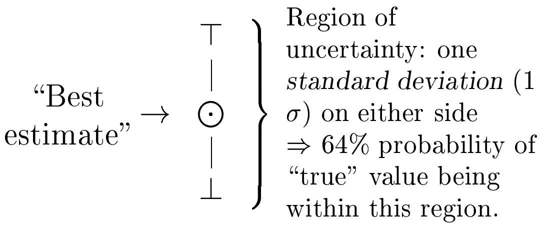
When, as in some elementary particle physics experiments lasting many years and costing millions of dollars, a great deal of effort has gone into measuring a single number, it is common practice to make a clear distinction between "statistical errors" (the *precision* of our instrumentation) and suspected "systematic errors" (mistakes). In most situations, however, both are lumped together or "added in quadrature" (the total uncertainty is the square root of the sum of the squares of the uncertainties due to all the independent sources of error).[1] It is considered poor form to cavalierly overestimate one's uncertainty to reduce the significance of deviations from expectations.

To write a measured value without its *tolerance* (uncertainty, "possible error," *etc.*) is as bad form as leaving out the *units* of the measurement. The significance of your measurement is lost. To do this in the presence of physicists is like ordering Ripple with your meal at Maxim's. Sadly, values are slipping throughout society, and otherwise respectable scientists can often be heard to quote numbers without specifying uncertainties. The best we can do is to be sure we do not contribute to this decay.

[1] More on this later. . . .

5.1.1 Graphs and Error Bars

When plotting points on a *graph*, the uncertainty is included in the form of "error bars" which look like this:



5.1.2 Vector Tolerance

Allow me to slip into something a little more formal....

Usually this topic would be called "Error Propagation in Functions of Several Variables" or something like that; I have used the term "vector tolerance" because (a) the word "error" has these perjorative connotations for most people, whereas "tolerance" is usually considered a *good* thing;[2] (b) when our final result is calculated in terms of several other quantities, each of which is uncertain by some amount, and when those uncertainties are *independent* of each other, we get a situation much like trying to define the overall length of a *vector* with several independent perpendicular components. Each contribution to the overall uncertainty can be positive or negative, and on average you would not expect them to all add up; that would be like assuming that if one were positive they all must be. So we *square* each contribution, add the squares and take the square root of the sum, just as we would do to find the length of a vector from its components.

The way to do this is easily prescribed if we use a little calculus notation: suppose the "answer" A is a function of several variables, say x and y. We write $A(x, y)$. So what happens to A when x changes by some amount δx?[3] Simple, we just write $\delta A_x \approx (\partial A/\partial x)\ \delta x$ where the $_x$ subscript on δA_x reminds us that this is just the contribution to the change in A from that little change in x, not from any changes in y; the $\approx$ sign acknowledges that this doesn't get exact until $\delta x \to dx$, which is *really* small; and the ∂ symbols are like derivatives except they remind us that *we are treating y as if it were a constant* when we take this derivative.

The same trick works for changes in y, of course, so then we have two "orthogonal" shifts of the result to combine into one *uncertainty* in A. I have already given the prescription for this above. The formula reads

$$(\delta A)^2 \approx \left(\frac{\partial A}{\partial x}\, \delta x \right)^2 + \left(\frac{\partial A}{\partial y}\, \delta y \right)^2 \qquad (1)$$

This can be extended to a function of N variables $\{x_1, x_2, \cdots x_i \cdots x_N\}$:

$$(\delta A)^2 \approx \sum_{i=1}^{N} \left(\frac{\partial A}{\partial x_i}\, \delta x_i \right)^2 \qquad (2)$$

where the $\sum$ symbol means "sum over all terms of this form, with the index i running from 1 to N."

The treatment above is a little too "advanced" mathematically for some people (or for anyone on a bad day), so here are a few special cases that the enthusiast may wish to derive from the general form in Eq. (2):

- **Uncertainty in a Sum:** If $A(x, y) = a\,x + b\,y$, with constants a and b, then

  $$(\delta A)^2 \approx (a\ \delta x)^2 + (b\ \delta y)^2. \qquad (3)$$

  That is, just *add the uncertainties in quadrature.*

[2] "Uncertainty" is somewhere in between.

[3] Notational convention: we use Δx to denote "a change in x, not necessarily tiny" whereas δx usually means "a little bitty change in x, but definitely finite!" and dx means "a change in x that is so teensy that it can be neglected relative to anything else but another really teensy thing." That last one (dx) is called a "differential" — Mathematicians don't like it much but Physicists use it all the time.

- **Uncertainty in a Product:** If $A(x,y) = a\,x\,y$, with constant a, then

$$\left(\frac{\delta A}{A}\right)^2 \approx \left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2. \quad (4)$$

  That is, just *add the fractional uncertainties in quadrature.*

- **Uncertainty in a Quotient:** If $A(x,y) = a\,x/y$, with constant a, then

$$\left(\frac{\delta A}{A}\right)^2 \approx \left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2. \quad (5)$$

  That is, just *add the fractional uncertainties in quadrature*, just like for a *product*.

- **Uncertainty in a Product of Power Laws:** If $A(x,y) = a\,x^p\,y^q$, with constant a, p and q, then

$$\left(\frac{\delta A}{A}\right)^2 \approx \left(p\,\frac{\delta x}{x}\right)^2 + \left(q\,\frac{\delta y}{y}\right)^2 \quad (6)$$

  which includes simple products and quotients.

These should get you through almost anything, if applied wisely.

5.2 Statistical Analysis

It's all very well to say that one should always report the results of measurements with uncertainties (or "errors" as they are often misleadingly called) specified; but this places a burden of judgement on the experimenter, who must estimate uncertainties in a manner fraught with individual idiosyncracies. Wouldn't it be nice if there were a way to *measure* one's uncertainty in a rigourous fashion?

Well, there is. It is a little tedious and complicated, but easily understood: one must make a *large number of repeated measurements of the same thing* and analyze the "scatter" of the answers!

Suppose we are trying to determine the "true" value of the quantity x. (We usually refer to unspecified things as "x" in this business.) It could be your pulse rate or some other simple physical observable.

We make N *independent* measurements x_i ($i = 1, 2, 3, \ldots, N$) under as close to identical conditions as we can manage. Each measurment, we suspect, is not terribly precise; but we don't know just how imprecise. (It could be largely due to some factor beyond our control; pulse rates, for instance, fluctuate for many reasons.)

Now, the x_i will "scatter" around the "true" x in a *distribution* that will put some x_i smaller than the true x and others larger. We *assume* that whatever the cause of the scatter, it is basically *random* — *i.e.* the exact value of one measurement x_{i+1} is not directly influenced by the value x_i obtained on the previous measurement. (Actually, perfect randomness is not only hard to define, but rather difficult to arrange in practice; it is sufficient that *most* fluctuations are random *enough* to justify the treatment being described here.) It is intuitively obvious (and can even be rigorously proved in most cases) that our *best estimate* for the "true" x is the average or *mean* value, $\bar{x}$, given by:[4]

$$\bar{x} \equiv \frac{1}{N} \sum_{i=1}^{N} x_i. \quad (7)$$

But what is the *uncertainty* in $\bar{x}$? Let's call it

[4]The symbol $\sum_{i=1}^{N}$ represents an *operator* called "summation" — it means that {the stuff to the right of the Σ}, which will always have a subscript i in one or more places, is to be thought of as the "i^{th} term" and all such terms with i values running from 1 to N are to be *added together* to form the desired result. So, for instance, $\sum_{i=1}^{N} x_i$ means $\{x_1 + x_2 + x_3 + \ldots + x_{N-1} + x_N\}$, or (to be more specific) if $N = 3$, just $\{x_1 + x_2 + x_3\}$. This may seem a little arcane, but it is actually a very handy compact notation for the rather common *summation* operation.

$\bar{\sigma}_x$.

How can we *find* $\bar{\sigma}_x$ mathematically from the data? Well, if we assume that each individual measurement x_i has the same *single-measurement uncertainty* σ_x, then the distribution of x_i should look like a "bell-shaped curve" or *gaussian distribution*:

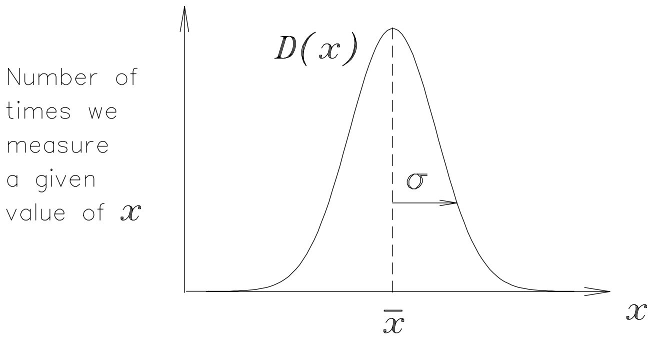


Figure 5.1 A typical graph of $\mathcal{D}(x)$, the *distribution* of x, defined as the relative frequency of occurence of different values of x from successive measurements. The "centre" of the distribution is at $\bar{x}$, the average or *mean* of x. The "width" of the distribution is 2σ (one σ on either side of the mean.

Obviously, $\Delta x_i \equiv x_i - \bar{x}$ is a measure of the "error" in the i^{th} measurement, but we cannot just find the average of Δx_i, since by definition the sum of all Δx_i is zero (there are just as many negative errors as positive errors). The way out of this dilemma is always to take the average of the *squares* of Δx_i, which are all positive. This "mean square" error is called the *variance*, s_x^2:

$$s_x^2 \equiv \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{8}$$

and its square root, the "root mean square error", is called the *standard deviation* — which can be shown (rigorously, in many cases, although not without a good deal of math) to be the best possible estimate for the single-measurement uncertainty σ_x.

So we actually have a way of "calculating" our uncertainty *directly from the data*! This is quite remarkable. But wait. We have not just measured x once; we have measured it N times. Our instincts (?) insist that our *final best estimate* of x, namely the *mean*, $\bar{x}$, is determined more precisely than we would get from just a single measurement. This is indeed the case. The uncertainty in the mean, $\bar{\sigma}_x$, is smaller than σ_x. By how much? Well, it takes a bit of math to *derive* the answer, but you will probably not find it implausible to accept the result that $\bar{\sigma}_x^2$ is smaller than σ_x^2 by a factor of $1/N$. That is,

$$\bar{\sigma}_x = \frac{\sigma_x}{\sqrt{N}}. \tag{9}$$

Thus 4 measurements give an average that is twice as precise as a single measurement, 9 give an improvement of 3, 100 give an improvement of 10, and so on. This is an extremely useful principle to remember, and it is worth thinking about its implications for a while.

COMMENT:

The above analysis of statistical uncertainties explains how to find the best estimate (the *mean*) from a number N of independent measurements with unknown but similar individual uncertainties. Sometimes we can estimate the uncertainty σ_{x_i} in each measurement x_i by some independent means like "common sense" (watch out for that one!). If this is the case, and if the measurements are not all equally precise (as, for instance, in combining all the world's best measurements of some esoteric parameter in elementary particle physics), then it is wrong to give each measurement equal *weight* in the average. There is then a better way to define the average, namely the "*weighted mean*":

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

where $w_i \equiv 1/\sigma_{x_i}^2$. If the reader is interested in the proper way to estimate the *uncertainty* $\bar{\sigma}_x$ in the mean under these circumstances, it is time to consult a statistics text; the answer is not difficult, but it needs some explanation that is beyond the scope of this *HyperReference*.