**Precision in Scientific Applications**

Lynn Robert Carter, Ph.D.

**Introduction**

The more we study, the more we realize there is to learn. As small children, we learned math, and the idea that there are right and wrong answers to simple arithmetic problems. I enjoyed math, because I liked being able to check my work to ensure I had the right answer. It wasn’t until much later that I came to understand that there are seldom simple right and wrong answers to the important questions in the real world.

This turns out to be true in so many things. In scientific applications, there are some cases where there are exact values. In most, however, there is some degree of uncertainty with whatever figure we use. For example, it is true that we can count the number of cars in a small parking lot and feel comfortable about the number being exact. As the parking lot becomes larger and larger, it becomes more difficult to be so certain. While you were in the middle of the count in one portion of the lot, did some people come back to their cars and leave from that portion of the lot you have already counted without you knowing it?

This paper provides an initial survey of key concepts from a most excellent book by John R. Taylor, entitled “An Introduction to Error Analysis; The study of uncertainties in physical measurements,” second edition.

**The writing and understanding of uncertainties**

Taylor does a fine job helping us understand the myriad reasons that produce the degree of uncertainty in just about every value we use, so I will not cover that here. We just know that nearly every value should be thought of as a pair, often expressed in the following form:

126.45 ± 0.01

The first number expresses the “best estimate” of the true value, while the second specifies the amount of uncertainty. What is often left unspoken is yet a third number, the level of confidence we have that the true value is between 126.44 and 126.46 (that is that the range of uncertainty is really ± 0.01). As we demand higher and higher levels of assurance that the expressed range indeed brackets the true value, the larger the range must be and, therefore, the larger the amount of uncertainty must become. If we really wanted to be precise, the best form for writing out a value would be:

126.45 ± 0.01 at 95%

This tells us that 95% of the time the true value will be between 126.44 and 126.46. If we need to increase our confidence to 99% or 99.9%, we need to go back to the source of the information and understand the source of the uncertainty to come up with a new value. For the purposes of this paper, we will assume that the level of confidence has been properly selected and we only need to deal with the pair of values.

The “error term” (in the example above, the “± 0.01” term) is usually rounded to just a single significant digit. The best way to come up with the error term is to think about the range of values. As mentioned above, there are many reasons why there might be a range of values as opposed to just a single value. Let’s assume that 100 people were asked to measure something, given exactly the same thing to measure and the same measuring tool. When all of the 100 people are done and we plot a histogram of their results, we expect to see something approaching a normal distribution (see Figure 1).

Macintosh HD:Users:LRCarter:Documents:  CMU-Q:Courses: 15-110 IIP:   Lectures - 2010 Fall:Graphics:Normal Curve.pdf

Figure 1

A Normal Distribution

The “*µ*” is the mean value of this set of measurements, and we usually consider this to be the “best estimate” for the measurement. As you can see, we expect to see values larger and smaller than this best estimate. Depending on the confidence that we want in our expression of this value, we must select the appropriate number of standard deviations (“σ”) above and below the mean. If we select 1σ, we will be correct roughly 68% of the time. If we select 2σ, we will be correct roughly 95% of the time. 3σ would mean we are correct roughly 99.8% of the time. The more standard deviations away from the mean, the greater the confidence you will have, but the larger the size of the uncertainty.

Be very careful. Not all situations produce a normal distribution. It is beyond the scope of this paper to address how to determine the error term in all cases. A good statistician in needed to evaluate the data and use the proper analysis method to compute what the appropriate error term should be. The reason for showing the normal case is to drive home that there is a distribution that must be considered and that statistical tools can be used to produce the error term.

Give the real data; we can compute the standard deviation. We can then use that figure with the confidence level we require, and the size of the uncertainty can be computed. It is convention to express the error term rounded to just a single significant digit[[1]](#footnote-1), because we want people to be able to accurately produce the range of values in their head. The only situation where we might provide two significant digits, is where that first error term digit is a “1” and the next digit is a “4”, “5”, or a “6”. If the deviation from the best estimate is “0.00144”, one might argue that using “0.001” would be a significant proportion reduction in the bounds of the real range and could lead to excluding too many values that might lead to problems. Selecting “0.002” would similarly lead to including too many values which could be equally bad.

So, in general, use one significant digit, except in the case where the significant digits are “14”, “15”, and “16”, where it is acceptable to use two digits, if they are warranted.

Once you know the error term, you can then figure out how to write out the best estimate. If the error term is “± 3”, it does not make sense to have a value expressed as “15.35 ± 3”. The error term specifies uncertainty about the best estimate. If the uncertainty is “± 3”, it is a waste of time and energy to show any more than that same precision in the best estimate value. Therefore, it should be shown as “15 ± 3”.

The general rule about writing out these values is to compute and round the error term to one, or at most two, significant figures, and then round the best estimate to the same number least significant digit. For example, if the error term ends up as “± 0.14” and the best estimate was shown as “12.72316”, the best estimate should be rounded to four significant decimal places, or “12.72” in this case and the value would be shown as “12.72 ± 0.14” (There are two decimal places in the error term, so round to two decimal places in the best estimate.)

Consider the situation where the error term is “± 200” and the best estimate was shown to be “672,316”. How should one write out the value? Again, we should round the best estimate to the same significant digit as the error term. Since the significant digit in the error term is in the 100s digit, we should round the best estimate value to the 100s. Therefore, the result should be written as “672,300 ± 200”.

**Accuracy and Precision**

It is unfortunate, but many people use the words “accuracy” and “precision” without really knowing what they mean or how they are different. The following table shows, in very graphical terms, the difference between accuracy and precisions. The left column shows two data sets. The upper is not accurate, for the set of measured values are not clustered around a mean, which is close to the true value. The lower left chart shows data that is more accurate, for the mean of the five values is close to the true value.

|  |  |
| --- | --- |
| Macintosh HD:Users:LRCarter:Documents:  CMU-Q:Courses: 15-110 IIP:   Lectures - 2010 Fall:Graphics:LessAccurate.pdf  Less Accurate | Macintosh HD:Users:LRCarter:Documents:  CMU-Q:Courses: 15-110 IIP:   Lectures - 2010 Fall:Graphics:LessPrecise.pdfLess Precise |
| Macintosh HD:Users:LRCarter:Documents:  CMU-Q:Courses: 15-110 IIP:   Lectures - 2010 Fall:Graphics:MoreAccurate.pdfMore Accurate | Macintosh HD:Users:LRCarter:Documents:  CMU-Q:Courses: 15-110 IIP:   Lectures - 2010 Fall:Graphics:MorePrecise.pdf  More Precise |

Table 1: Precision and Accuracy

The right column shows precision. The upper chart is less precise, as the measured values differ from one another and the difference is ± 4, which is slightly more than 10% of the value. The lower right chart is more precise, as the measured values are closer together than the upper chart, with an uncertainty of ± 2, which is just a bit more than 5% of the value.

Accuracy is a measure of the degree to which a set of measured values clusters around the actual value of something being measured. Precision is a measure of the amount of deviation among a set of measured values of a something being measured.

We scientists would like measurements than are both accurate and precise.

**Propagating uncertainties with addition and subtraction**

Reading and writing out uncertain values is just the start. It is usually the case that people want to do arithmetic with them. We will show you a formula later to generalize the result, but we’d like you to deeply understand what is going on, so you can always develop the formula for yourself. If you just memorize the formula, you may not really understand the “why” behind the formula and if you remember it wrong, you may not recognize the error in the result, before it is too late.

If we want to know if two boxes can fit into a space that is limited to roughly 2.5 meters in length, we know we need to add the length of the two boxes together. If one box is roughly 1 meter in length and the other is roughly 1.496 meters in length, we would assume that there would be no problem, since 1 + 1.496 = 2.496 < 2.5.

To be sure, let’s zoom in on the word “roughly”. As we have learned, all measured values are approximations. Not all boxes are shaped precisely, so the length on one side might not be precisely the same as the length of the other side. If we were to measure these dimensions and determine the uncertainty, we might discover that the available space might be better written as “2.500 meters ± 0.002 meters”, and the length of the two boxes as “1.000 ± 0.002” and “1.496 ± 0.002”. Now, things appear to be a bit more confusing.

To add the lengths of the two boxes together, let’s be explicit about the range for the length of the two boxes and add these ranges together. The first box’s length is [0.998, 1.002][[2]](#footnote-2) and the second is [1.494, 1.498]. To compute the smallest value of the sum, we add the smallest possible value of each range. Similarly, to compute the largest possible sum, we add the largest from each range. In this case, the sum is [2.492, 2.500], a span of 0.008 or an uncertainty of “± 0.004”, and we could write it as “2.496 ± 0.004”. This looks good until we recall that the space for these boxes was “2.500 meters ± 0.002 meters” and the range of the space available could be [2.498, 2.5002]. Therefore, there are cases, where the two boxes will not fit.

From this example, we can see that the sum of two uncertain values appears to be the sum of the best estimates plus/minus the sum of the error terms. This is a reasonable working formula, but in some situations a more precise formula is possible, but it is beyond the scope of this paper to explore that.

Now lets explore how to deal with the situation where two values have very different error terms. Consider the case of adding “15.75 ± 0.05” with “17.62375 ± 0.00001”. Converting these values to ranges and then adding the ranges works well here, again. So the range values would be [15.70, 15.80] and [17.62374, 17.62376]. Again, computing the range requires us to add the two smallest and the two largest, which results in the value [34.32374, 34.42376]. The size of the range is “0.10002”. If we assume a normal distribution, the best estimate would be at the middle, “34.37375”, and the error term would have half above and half below this value, resulting in “34.37375 ± 0.05001”. Following our rule about error terms having just one significant digit, the appropriate error term would be “± 0.05”. Using this to round the best estimate, the sum should be listed at “34.37 ± 0.05” with a range of [34.32, 34.42][[3]](#footnote-3) which is accurate to two decimal places.

To add two uncertain values, add the two best estimates, add the error terms, round the error term to just one or maybe two significant digits, use the significance of the new error term to determine the significance of the sum, and write the rounded sum with its new error term.

In the case of subtracting two uncertain values, the same results apply, subtract the two best estimates (producing a difference), **add** the error terms, round this new error term to just one or maybe two significant digits, use the significance of the new error term to determine the significance of the difference, and write out the rounded difference with its new error term.

**Propagating uncertainties with multiplication and division**

Let’s assume we have a large room with a floor that must be tiled. In order to secure the right amount of materials, we need to know the area to be covered. Since the room is a nice and simple rectangle of roughly 6 meters by 12.5 meters, we can compute the area by multiplying width bye the length and find that we will be tiling 75 meters2 of floor. We are in luck, for that is exactly how much mastic comes in a container. Or are we?

Let’s assume that in reality, our measurements are “± 0.005”. This means that these two measurements could be as much as 5 millimeters longer than we thought. How significant is that error in this case? To find out, lets do the same range calculation we did for addition and see what the range of results would be.

The range of the “6 ± 0.005” meter value is [5.995, 6.005] meters. The range of the second value is [12.495, 12.505]. The smallest and largest area figures come from multiplying the two smallest and the two largest values. So the range of the results is [74.907525, 75.092525]. The size of this range is 0.185. Dividing this range by two (half above the mean value and half below) gives us 0.0925. Rounding this error term to just one digit gives us an error of “± 0.09”, which is much larger than the error of the two input values. Computing the mean of the range, we get 75.000025. Using the significance of the error term, we round the result to 75.00, producing “75 ± 0.09”. I will leave it to you to determine how you feel about gambling that the actual amount will be the low side of this figure as opposed to the high side when it comes to purchasing just one can. (Especially if you can return an unused can and get your money back.)

The formula for computing the error term in the product is obviously more complex that just the sum of the component error terms. Multiplication and division work like a magnifier and that can magnify errors as well as the product and quotient. We will not derive the formula here, but we will explain its use in the example we’ve just covered.

ProductErrorTerm Value1ErrorTerm Value2ErrorTerm

= +

⎮Product⎮ ⎮Value1⎮ ⎮Value2⎮

To compute the error term (ProductErrorTerm) we first need to compute the result:

Product = Value1 × Value2

Compute the relative fraction of Value1 that is uncertain (Value1errorFraction):

Value1ErrorFraction = Value1ErrorTerm / ⎢ Value1 ⎢

Compute the relative fraction of Value2 that that is uncertain (Value2ErrorFraction):

Value2ErrorFraction = Value2ErrorTerm / ⎢ Value2 ⎢

Compute the Product’s error (ProductErrorTerm):

ProductErrorTerm = (Value1ErrorFraction + Value2ErrorFraction) × ⎢ Product ⎢

Round the Product’s error term to one or at most two significant digits.

Given the error term, determine how many significant decimal places the result should have, round the product to that many places, and produce the final result.

The formula is like an average of the proportion of the two input value’s uncertainties. So, let’s try this formula to see how it compares with our range method value.

Value1ErrorFraction = Value1ErrorTerm / ⎢Value1 ⎢ (1)

Value1ErrorFraction = 0.005 / ⎢6 ⎢ = 0.00083 (2)

Value2ErrorFraction = Value2ErrorTerm / ⎢Value2 ⎢ (3)

Value2ErrorFraction = 0.005 / ⎢12.5 ⎢ = 0.0004 (4)

Product = 6 × 12.5 = 75 (5)

ProductErrorTerm = (Value1ErrorFraction + Value2ErrorFraction) × ⎢ Product ⎢ (6)

ProductErrorTerm = (0.000833 + 0.0004) × ⎢ 75 ⎢ = 0.092475 = 0.09 (7)

Product = 75.00 meter2  (8)

Result = 75.00 meter2 ± 0.09 meter2 (9)

Please notice that the error computed on the previous page (0.0925) agrees to three decimal places with the error computed in line (7) above prior to rounding. This should bring some measure of comfort that this formula does what we claim it does.

The process for determining the uncertainty for division is precisely the same as the uncertainty for multiplication. It is common for some to think that step (6) on the previous page should be changed to be a division, but this in not true. Exactly the same algorithm is performed, including the multiplication by the quotient, as shown below:

QuotientErrorTerm Value1ErrorTerm Value2ErrorTerm

= +

⎮Quotient⎮ ⎮Value1⎮ ⎮Value2⎮

To compute the error term (QuotientErrorTerm) we first need to compute the result:

Quotient = Value1 / Value2

Compute the relative fraction of Value1 that is uncertain (Value1errorFraction):

Value1ErrorFraction = Value1ErrorTerm / ⎢ Value1 ⎢

Compute the relative fraction of Value2 that that is uncertain (Value2ErrorFraction):

Value2ErrorFraction = Value2ErrorTerm / ⎢ Value2 ⎢

Compute the Quotient’s error (QuotientErrorTerm):

QuotientErrorTerm = (Value1ErrorFraction + Value2ErrorFraction) × ⎢ Quotient ⎢

Round the Quotient’s error term to one or at most two significant digits.

Given the error term, determine how many significant decimal places the result should have and round the quotient to that many places, and produce the final result.

The formula is the same as that for multiplication. The following is the recommended algorithm:

Value1ErrorFraction = Value1ErrorTerm / ⎢Value1 ⎢ (10)

Value2ErrorFraction = Value2ErrorTerm / ⎢Value2 ⎢ (11)

Quotient = Value1 / Value2 (12)

QuotientErrorTerm = (Value1ErrorFraction + Value2ErrorFraction) × ⎢ Quotient ⎢ (13)

Result = Quotient ± QuotientErrorTerm (14)

1. The value 0.001 has just one significant digit. The leading zeros do not count. The easiest way to think about this is to express the value in scientific notation, where the mantissa is less than 1 and greater than or equal to .1. In this situation, the value would be .1 × 10-2. Here, then, it is clear that there is only one digit in the mantissa (the “.1”). [↑](#footnote-ref-1)
2. The first value is the smallest value in the range while the second value is the largest. The use of the “[“ and “]” characters signifies that the named value is inside of the range. (0, 1.5] would specify a range bounded below by zero and 1.5 above. Here the “(“ specifies that zero is outside of the range, while 1.5 is in the range. [↑](#footnote-ref-2)
3. Notice that the resulting range is shifted slightly toward zero. The amount of this shift is relatively insignificant, given the amount of error in the first operand. [↑](#footnote-ref-3)