TUTORIAL

Experimental Uncertainty and Error

1 Scientific Measurement

Essential to much of scientific experimentation is measuring values. Sometimes, these measurements are in the service of some larger project (e.g. testing a theoretical prediction or determining a functional relationship between two quantities). Other times, determining the value of some quantity is the goal itself, as in many classic experiments like Milliken's measurements of the charge of an electron and Planck's constant.

If some quantity has a well-defined value—which we might call the *true value*—we would expect repeated measurements of it to yield that same value every time. In real experiments, however, there is always some fluctuation. We need some way to reduce all of the measured values into a single, *best-estitmate value*.

Further, it is not enough to simply state a number (with units). It's important to know, also, how confidant we can be that our best-estimate reflects the true value. Comparisons like "the result almost agrees with theory" or "our results were different from Joe's" are not precise enough. To make experimental results useful, we need to *quantify* these statements. How to accomplish these tasks will be discussed here in brief[1] with the help of the following example.

Suppose we have a spring cannon, with a spring constant of $k = (4204 \pm 9)$ N/m. We launch from it a small ball of mass $m = (100 \pm 1)$ g, after compressing the spring by $d = (6.0 \pm 0.1)$ cm. We use a stopwatch to measure the time t for the ball to rise and fall back to its launch height, and from the time we determine the launch speed. We record the following times.

Times (s)									
2.51	2.57	2.46	2.64	2.53	2.47	2.63	2.54	2.58	2.59
2.52	2.41	2.42	2.62	2.67	2.49	2.52	2.54	2.48	2.54

1.1 The Best Estimate Value

Since we're doing the *same* experiment over and over, measuring the *same* thing, there should be a *single* value. Our first task will be to condense given data to one value, which represents our best estimate of the true value.

Mean If the fluctuations in the data are random and uncorrelated (i.e., not due to a defective or poorly designed experimental setup), they will be as often above the true value as below it, and the best candidate for the true value is usually the **average** or **arithmetic mean**. If we perform N measurements of some quantity x, resulting in a sequence of values $x = \{x_1, x_2, \ldots, x_N\}$, the mean $\bar{x}$ is given by

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

Median If a data set contains significant outliers, however, the set may be better represented by the **median**, which is the midpoint of the data set: there are as many values larger than it as there are smaller than it. To find the median, first sort the data. If the number of values is odd, the median is the one in the middle. If it is even, it is the average of the two values in the middle.

[1] For more information, or for derivations of some of the equations, consult [1] and/or a statistics textbook.

Cannon Example For the launched ball, the measured times were $t = \{2.51, 2.57, \dots, 2.54\}$ s. The mean value is

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_i = \frac{(2.51 + 2.57 + \cdots + 2.54)\ \mathrm{s}}{20} = 2.5365\ \mathrm{s}.$$

The sorted times are $t = \{2.41, 2.42, \dots, 2.53, 2.54, \dots, 2.64, 2.67\}$ s. Taking the average of the two values in the middle, we get a median value of $t_{\mathrm{med}} = 2.535$ s.

1.2 Experimental Uncertainty

Experimental uncertainty is an indication of how inexact a measurement is. Few measurements can provide results with *perfect* exactitude (one exception is when the measurement consists simply in counting discrete objects), so it is essential to report the uncertainty in every measurement.

One way to get a quick impression of the uncertainty in an experiment is to do some graphing. For our cannon example, Fig. 1 is a histogram showing the number of results in each 0.05 s interval, or "bin." It shows the spread, or scatter, in our data. The true value should lie somewhere near the peak of the histogram. Indeed, the average we calculated above, $\bar{t} = 2.5365$ s, lies within the highest bar, near the center of the graph.

Another way to visualize the spread in the data is simply to plot the measured values against the order in which the data were taken, as in Fig. 2. Beyond showing the spread in the data visually, a graph like this could reveal problems with the experimental setup. If the measurements show some systematic change when we expect random fluctuation, we would need to double-check our equipment.

Most of the data lie in a range from 0.1 s below the mean to 0.1 s above it, so 0.1 s is a decent first approximation of the experimental uncertainty.

Sample standard deviation To more rigorously quantify the spread in data, we can calculate the **sample standard deviation**, sometimes called the "uncertainty in a single measurement" or simply the "standard deviation," which is essentially the average *distance* between the individual measurements and the mean. If we have measured N values of some quantity $x = \{x_1, x_2, \dots, x_N\}$, with an average of $\bar{x}$, the sample standard deviation is given by

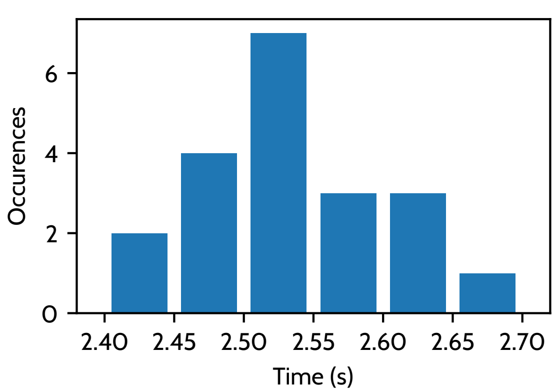

Figure 1: A histogram indicating the number of readings in each 0.05 s interval. The average value is 2.54 s, the standard deviation is 0.071 s, and the uncertainty in the mean is 0.016 s.

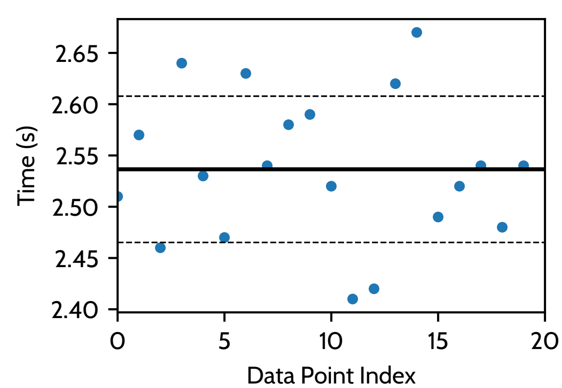

Figure 2: The times from the measurement plotted against position in the table. The solid line is the mean $\bar{t}$, and the dashed lines represent $\bar{t} + \sigma_t$ and $\bar{t} - \sigma_t$.

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}. \tag{2}$$

This quantity is a measure of the variation from one measurement to another. We would typically expect that roughly 2/3 of the measurements will lie between $\bar{x} - \sigma_x$ and $\bar{x} + \sigma_x$.

Uncertainty in the mean As you measure some quantity more and more times, the mean and the standard deviation tend toward constant values, and the mean approaches the true value, as the random fluctuations cancel out. We can capture this effect by calculating the **uncertainty in the mean**, usually just called the **uncertainty**[2]. For N measurements of a quantity x, with standard deviation σ_x, the uncertainty is

$$\Delta x = \frac{\sigma_x}{\sqrt{N}}. \tag{3}$$

It quantifies how close we can reasonably expect our mean, best-estimate value to be to the true value.

It is always smaller than the standard deviation, and we can always decrease it—improving the precision of our experiment—by making more measurements. As a general rule, you should take as many as time and money allow.

Cannon Example For the cannon, the sample standard deviation is[3],

$$\sigma_t = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(t_i - \bar{t})^2} = \sqrt{\frac{(2.51-2.5365)^2 + (2.57-2.5365)^2 + \cdots + (2.54-2.5365)^2}{19}\ \mathrm{s}^2}$$
$$= 0.0713\ \mathrm{s}.$$

The range $(\bar{t} - \sigma_t, \bar{t} + \sigma_t)$ is indicated in Fig. 2 by dashed lines; 65% of the data—close to the expected 2/3—lie in the range.

The uncertainty in the mean is

$$\Delta t = \frac{\sigma_t}{\sqrt{N}} = \frac{0.0713\ \mathrm{s}}{\sqrt{20}} = 0.0159\ \mathrm{s}.$$

1.3 Reporting Results and Significant Digits

The standard form for reporting the value of some measured or calculated quantity x is

$$x = \bar{x} \pm \Delta x, \tag{4}$$

where $\bar{x}$ is the best estimate (which is usually the mean, so we use the bar notation).

When we quote our result, we only keep the digits which are *significant*. Significant figures are determined by the value of Δx. Only one digit is kept, unless the first digit is a 1, in which case either one or two digits can be retained. Then $\bar{x}$ should be reported such that the least significant digit (or two digits, if that is allowed) represents the same power of 10 as the least significant digit of Δx.

Returning to our example, we found that $\bar{t} = 2.5365$ s and $\Delta t = 0.0159$ s. Given that the first significant digit in Δt is a 1, we can keep either one or two digits.

[2] Uncertainty is called by most statistics books the *standard error*. Terms involving the word "error" can be misleading in the context of experiment design.

[3] Normally, you will do the calculation of standard deviation using a computer. In Python, using `numpy`, the command is `numpy.std`, with the option `ddof=1`. In Excel, the function you want is `=STDEV.S`.

- If we keep only one digit, we have $\Delta t = 0.02$ s. The most significant digit is in the 10^{-2} place, so we have to round $\bar{t}$ to 2.54 s. Then, in standard form, $t = (2.54 \pm 0.02)$ s.
- If, instead, we keep two digits, $\Delta t = 0.016$ s. The most significant digit is now in the 10^{-3} place, so now $\bar{t} = 2.537$ s. Then, our final result is $t = (2.537 \pm 0.016)$ s.

2 Uncertainty in Individiual Measurements

2.1 Sources of Uncertainty

What are the sources of uncertainty in individual measurements? The answer, of course, depends on the experiment. Here are some possible factors to consider:

Resolution The **resolution** of an instrument is the smallest interval between two readings that the instrument can meaningfully say are different. The stopwatch in our example had a resolution of 0.01 s. The practical resolution of an instrument may also be smaller than the intrinsic resolution, if external circumstances (e.g. air currents above a scale or electrical fluctuations in wires) introduce extra noise.

Human factors Though the resolution of the stopwatch above was 0.01 s, the sample standard deviation was nearly an order of magnitude larger—0.07 s. This is because human reaction time is significantly longer than 0.01 s (typical values are between 0.15 s and 0.30 s [2]).

Intrinsic roughness Sometimes our precision is limited by the thing we're trying to measure itself. For example, determining the diameter of a cotton ball to within a fraction of a millimeter would be impossible—a cotton ball is far too fuzzy.

All of these sources of uncertainty will be present in any measurement. However, often, one of them will be so much larger than the others that only it needs to be considered. In the case of the cannon, the resolution of the stopwatch was irrelevant because the uncertainty due to human reaction time was so much larger.

2.2 Determining Uncertainty

In most cases, we simply *estimate* uncertainties in directly measured quantities.

For example, a typical meter stick is marked in 1 mm increments. Since it is often possible—as long as what we're measuring has clear, sharp edges—to interpolate between the markings, we might estimate the uncertainty to be 0.5 mm or even 0.2 mm if we have good vision.

We can also estimate the effects of human factors (e.g. response time) and roughness. For example, suppose we want to measure the distance between two lenses on a horizontal track using a meter stick built into the track. It is often difficult to tell precisely where the center of each lens is. Also, if the lenses are raised vertically off of the track, measuring the positions will require looking from above, and any deviation from perfect vertical alignment will make the measurement less accurate. Both of these factors should increase our estimate of the uncertainty in the distance. In such a case, 0.5 cm or 1 cm might be more reasonable than the numbers given in the previous paragraph.

3 Propagation of Uncertainty

Often, what we want to know is not measured directly but is instead calculated from quantities that are measured. The process for determining the uncertainty in a calculated quantities from the uncertainties in the measured values is called **propagation of uncertainty**.

Table 1: Rules for the propagation of uncertainty. The numbers K, ℓ, m, and n must have no uncertainty.

Operation	Rule
1. Addition and subtraction	
$Q = x \pm y \pm z \pm \cdots$	$\Delta Q = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 + \cdots}$
2. Multiplication, division, and raising to powers	
$Q = K\dfrac{x^\ell y^m \dots}{z^n \dots}$	$\Delta Q = \lvert Q\rvert \sqrt{\left(\dfrac{\ell \Delta x}{x}\right)^2 + \left(\dfrac{m\Delta y}{y}\right)^2 + \left(\dfrac{n\Delta z}{z}\right)^2 + \cdots}$
3. Function of one variable	
$Q = Q(x)$	$\Delta Q = \left\lvert \dfrac{dQ}{dx}\Delta x\right\rvert$
4. Function of multiple variables	
$Q = Q(x, y, z, \dots)$	$\Delta Q = \sqrt{\left(\dfrac{\partial Q}{\partial x}\Delta x\right)^2 + \left(\dfrac{\partial Q}{\partial y}\Delta y\right)^2 + \left(\dfrac{\partial Q}{\partial z}\Delta z\right)^2 + \cdots}$

To discuss, let us assume that we have some quantity Q which depends on other quantities $x, y, z, \dots$, and we know the uncertainties in these to be $\Delta x, \Delta y, \Delta z, \dots$. The rules for how to calculate the uncertainty in Q from the uncertainties in the quantities on which it depends are listed in Table 1. Rule 4 is the most general, though the others are simpler in many cases. However, if the same independent variable appears more than once in the expression, then Rule 4 *must* be used.

Example 1 Suppose we have a rectangular sheet for which we have measured the length and width to be $L = (1.25 \pm 0.08)$ m and $W = (0.205 \pm 0.005)$ m. We can calculate the area to be

$$A = LW = (1.25\ \text{m})(0.205\ \text{m}) = 0.256\,25\ \text{m}^2.$$

Since we calculated a product, we can calculate the uncertainty using Rule 2 in Table 1 with $Q = A$, $x = L$, $y = W$, $K = 1$, and all powers equal to 1:

$$\Delta A = \lvert A\rvert \sqrt{\left(\frac{\Delta L}{L}\right)^2 + \left(\frac{\Delta W}{W}\right)^2} = (0.256\,25\ \text{m}) \sqrt{\left(\frac{0.08}{1.25}\right)^2 + \left(\frac{0.005}{0.205}\right)^2} = 0.017\,555\ \text{m}^2$$

If the first significant digit in the propagated uncertainty is a 1, it is acceptable to keep two digits, so we may conclude that the area can be reported as either $A = (0.26 \pm 0.02)\ \text{m}^2$ or $A = (0.256 \pm 0.018)\ \text{m}^2$.

Example 2 (a) A car starts with initial velocity $v_0 = (3.2 \pm 0.8)$ m/s and accelerates at a constant rate $a = (0.6 \pm 0.2)\ \text{m/s}^2$ over time $t = (5.36 \pm 0.03)$ s. What is its final velocity?

For constant acceleration, final velocity is

$$v = v_0 + at = (3.2\ \text{m/s}) + (0.6\ \text{m/s}^2)(5.36\ \text{s}) = 6.416\ \text{m/s}.$$

To find Δv without calculus, we will have to break the problem into parts. We can use Rule 2 to calculate the uncertainty in the quantity at and then use Rule 1 to get the final result. Rule 2 gives

$$\Delta(at) = at\sqrt{\left(\frac{\Delta a}{a}\right)^2 + \left(\frac{\Delta t}{t}\right)^2} = (0.6\ \text{m/s}^2)(5.32\ \text{s})\sqrt{\left(\frac{0.2}{0.6}\right)^2 + \left(\frac{0.03}{5.36}\right)^2} = 1.07\ \text{m/s}.$$

Now apply Rule 1:

$$\Delta v = \sqrt{(\Delta v_0)^2 + [\Delta(at)]^2} = \sqrt{(0.8\ \text{m/s})^2 + (1.07\ \text{m/s})^2} = 1.336\ \text{m/s}.$$

In standard form, the final velocity of the car is $v = (6.4 \pm 1.3)$ m/s or $v = (6 \pm 1)$ m/s.

Example 2(b) We could consider the same problem in one step using the (seemingly) more complicated Rule 4. The final velocity v is a function of three variables v_0, a, and t, each with a corresponding uncertainty. Thus, Rule 4 gives

$$\Delta v = \sqrt{\left(\frac{\partial v}{\partial v_0}\Delta v_0\right)^2 + \left(\frac{\partial v}{\partial a}\Delta a\right)^2 + \left(\frac{\partial v}{\partial t}\Delta t\right)^2} = \sqrt{[(1)\Delta v_0]^2 + [(t)\Delta a]^2 + [(a)\Delta t]^2}$$
$$= \sqrt{(0.8\ \text{m/s})^2 + \left[(5.36\ \text{s})(0.2\ \text{m/s}^2)\right]^2 + \left[(0.6\ \text{m/s}^2)(0.03\ \text{s})\right]^2} = 1.336\ \text{m/s},$$

as above.

Cannon Example From kinematics, we can show that the launch speed for the ball is

$$v_0 = \frac{1}{2}gt,$$

where t is the total time the ball is in the air and $g = 9.802\,46\ \text{m/s}^2$ (which we will take to have no uncertainty) is the acceleration due to gravity [3]. Using our earlier results, ($\bar{t} = 2.5365$ s and $\Delta t = 0.0159$ s,

$$v_0 = \frac{1}{2}\left(9.802\,46\ \text{m/s}^2\right)(2.5365\ \text{s}) = 12.431\ \text{m/s}.$$

We are treating g as a precise constant, so t is the only uncertain quantity. We can use Rule 2:

$$\Delta v_0 = \left|\frac{1}{2}\left(9.802\,46\ \text{m/s}^2\right)(0.016\ \text{s})\right| = 0.0781\ \text{m/s}.$$

In standard form, then, the launch velocity is

$$v_0 = (12.43 \pm 0.08)\ \text{m/s}. \tag{5}$$

4 Comparing Values with Uncertainties

To say that $x = \bar{x} \pm \Delta x$ means that we can only claim that the true value of x is in the range

$$\bar{x} - \Delta x \le x \le \bar{x} + \Delta x. \tag{6}$$

Now suppose you have measured some quantity two different ways, that two people performed the same procedure, or some similar situation. The results of the two experiments are $x_1 = \bar{x}_1 + \Delta x_1$ and $x_2 = \bar{x}_2 + \Delta x_2$. The results are consistent with each other if the ranges $(\bar{x}_1 - \Delta x_1, \bar{x}_1 + \Delta x_1)$ and $(\bar{x}_2 - \Delta x_2, \bar{x}_2 + \Delta x_2)$ overlap.

If the ranges do *not* overlap, we can estimate how confident we are that the results are **inconsistent**. It can be shown that if two results are separated by more than two standard deviations,

$$|\bar{x}_1 - \bar{x}_2| > 2\max(\Delta x_1, \Delta x_2), \tag{7}$$

then we can be 95% confident that the differences between the two results are *not* due to chance—that the difference is **statistically significant**.

If neither condition obtains, the experiment is generally considered inconclusive.

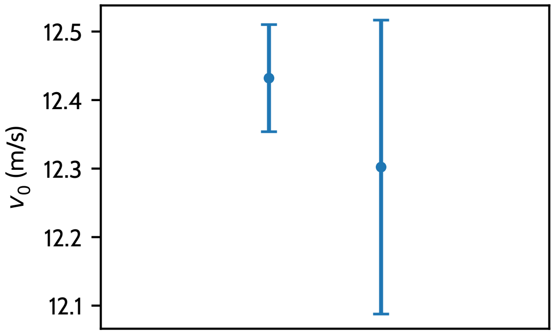

Figure 3: A graph showing the ranges of velocities consistent with two different calculation methods. The ranges are clearly overlapping.

Cannon Example When the cannon example was first described, enough information was given to determine the launch speed another way. From conservation of energy,

$$\frac{1}{2}kd^2 = \frac{1}{2}mv_0^2 \qquad \Rightarrow \qquad v_0 = \sqrt{\frac{kd^2}{m}} = \sqrt{\frac{(4204\ \mathrm{N/m})(0.060\ \mathrm{m})^2}{0.100\ \mathrm{kg}}} = 12.302\ \mathrm{m/s}.$$

Using Rule 2 from Table 1 (and distributing the square-root), we can determine the uncertainty.

$$\Delta v_0 = v_0\sqrt{\left(\frac{\Delta k}{2k}\right)^2 + \left(\frac{\Delta d}{d}\right)^2 + \left(\frac{\Delta m}{2m}\right)^2} = (12.302\ \mathrm{m/s})\sqrt{\left(\frac{9}{2(4204)}\right)^2 + \left(\frac{0.001}{0.060}\right)^2 + \left(\frac{0.001}{2(0.100)}\right)^2}$$
$$= 0.2145\ \mathrm{m/s}.$$

This method of determining the launch speed then yields

$$v_0 = (12.3 \pm 0.2)\ \mathrm{m/s}. \tag{8}$$

The ranges expressed in Eq. (5) and Eq. (8) do clearly overlap, and we can visualize this using a graph like Fig. 3. Therefore, we can conclude that the two results are in agreement.

5 Uncertainty vs. Error

As we've been using the term, uncertainties are not due to mistakes in measurement or experimental setup. They tend to be random, meaning that if we make repeated measurements, the results will be sometimes larger and sometimes smaller. Uncertainty is essentially a quantification of this random fluctuation.

On the other hand, *error* is a sign of a poorly designed experiment. It is systematic. For example, if our stopwatch ran slow, so that it ticked off 0.9 s for every 1.0 s that actually elapsed, our data would all be off by 10%. An important goal in experiment design is to eliminate as much systematic error as possible.

Some scientists will use the terms uncertainty and error interchangeably but prefix them with the adjectives "random" (for what we are calling "uncertainty") or "systematic" (for what we call "error").

References

[1] J. R. Taylor, *Error analysis: the study of uncertainties in physical measurements*, 2nd ed. (University Science Books, Sausalito, California, 1997).

[2] D. Yuhas, *Speedy science: how fast can you react*, Scientific American, (May 24, 2012) `https://www.scientificamerican.com/article/bring-science-home-reaction-time/` (visited on 01/03/2018).

[3]N. G. Survey, *Surface gravity prediction*, (The form was filled out with a latitude of 42°07’08”, longitude of 79°59’21”, and altitude of 299 m—a location in the Behrend Science Complex parking lot. The result was 980246 milligals +/- 2 milligals.), `https://geodesy.noaa.gov/cgi-bin/grav_pdx.prl` (visited on 07/30/2019).