Week 3: Uncertainties

for Northwestern University Undergraduate Laboratories

The fundamental idea of Physics is to understand the physical laws of nature. Inherent in this idea is observation of physical phenomena. And key to this is measurement. One assumes that certain physical quantities have values that can be determined, for example the Universal Gravitational Constant, or the mass of an electron. How well we can determine some of these quantities may have far reaching consequences. It was only after we determined to a great accuracy the mass of the protons, neutrons and electrons did we realize that for an atom composed of the particles, the whole did not equal the sum of the parts, were we able to harness the energy of the atom.

A physical quantity has a true value. Then there is the value that we as physicist determine by measurement. Due to the inherent limitations of measurement we can never be certain that the value we measure is exactly the same as the true value.

For a physicist knowing the value of a measured quantity is only half the measurement. Inherent in measuring a quantity is the limitation in our ability to measure the quantity to an infinite precision. Consequently every experimentally observed quantity should have associated with it an uncertainty or confidence that the actual value lays within that range of values.

In a typical undergraduate laboratory it is almost always the case that we will be measuring a quantity that has been measured before and is well know to many decimal places such as the acceleration of gravity *g*. It might seem obvious to determine how well we have performed the experiment to compare our result with the known value and to calculate a percent error. This would be a misleading measure of the quality of our work. We could have a value of 30% deviation, pretty awful. Yet it may be the best we can expect with the equipment we use, which might be capable of determining the value with confidence only to ±35%.

We would like to discuss how to determine the confidence of our measurement for a particular measurement. Determining a fitting value for the confidence limit is a complicated procedure that involves intricate math, but usually also includes some estimating and guesswork. One of the purposes of this lab is introduce you to the concept of uncertainty.

There are two situations in which we need to consider uncertainties: 1) when we want to decide whether two quantities are the same or different, and 2) when we want to make a quantitative prediction. Note: often people use the word “error” meaning a deviation of a measured or predicted value from the true value, but more and more, scientists avoid “error” and use “uncertainty.” (An “error” is when you get the wrong value whereas an “uncertainty” quantifies how precise a value is.)

There are two types of uncertainty, and they are fundamentally different. The first kind is <u>statistical uncertainty</u> and it relates to randomness[1] that sometimes plays a major role. For example, the number of fish in a pond will fluctuate randomly year-to-year even if nothing about the pond or the fish changes. The second kind is <u>systematic uncertainty</u>, which relates to some cause-and-effect phenomenon that impacts the way a measurement turns out. For example, weighing individual fish specimens could change year-by-year if the scale goes out of calibration because the internal spring wears out or rusts. In this exercise, you will investigate simple cases of statistical uncertainty and systematic uncertainty.

---

1 The professional term for “random” is <u>stochastic</u>. They key point is that the exact outcome fundamentally cannot be predicted. The result of tossing two dice is stochastic – i.e., random. There is no cause-and-effect, practically speaking.

Aside: there are subtle cases in which a systematic uncertainty originates in theory, such as the validity of a mathematical model. There are also cases where the systematic uncertainty is based on an auxiliary measurement which itself has a statistical uncertainty, so the given systematic uncertainty has the quality of a statistical uncertainty. These confusing cases will not arise in this course, however.

Statistical Uncertainties

Illustration: Two basketball teams, the Wildcats and the Ruffians, are matched against each other twice. The first time, the Wildcats win, 88-70. The second time, they win again, 90-68. The student newspaper touts the second win as clear signs of a stronger team. But is it? Is 90 significantly higher than 88, and is a 22 point difference significantly greater than 18, given all the random things that affect play in a college basketball game?

Exercise: A "fair" coin is one which has an equal probability to land heads up or tails up. Assuming that the coin never lands on its edge, this means 50% chance of getting a heads when you toss the coin. It is meaningless, of course, to *predict* that you will get heads when you toss the coin once, because getting a heads or a tails is completely random. Suppose you toss the coin 16 times in a row. You could predict that you will get 8 heads and 8 tails, but that is only an expected outcome, ie, what should happen *on average*. In other words, if 200 students perform the same experiment and toss fair coins 16 times each, then the average outcome would be 8 heads. Your actual observed outcome, however, could well be different[2].

Table 1: Records of coin toss experiments and statistics.

Series	N_0	$D = N_O - N_E$	D^2
1	9	1	1
2	6	-2	4
3	10	2	4
4	12	4	16
5	7	-1	1
6	9	1	1
7	3	-5	25
8	7	-1	1
9	8	0	0
10	6	-2	4
Averages	7.7	**0**	5.7

Find an ordinary coin and use it for the following exercise. What type of coin is it? A quarter

Toss the coin 16 times and count the number of heads. How many heads did you observe? How many heads would you expect to observe, on average? What is the difference (N observed – N expected)? Record your observations in Table 1.

2 In fact, the probability that you obtain 8 heads is not at all that high, though it is the most likely outcome – all other outcomes (7 heads, 9 heads, etc.) are less likely that 8 heads.

Repeat the experiment 10 times and fill in Table 1. (This means you will toss the coin a total of 160 times.) Definitions:

N_T = number of coin tosses (sometimes called "trials") in one experiment

N_O = number of heads observed

D = difference between N_O and the number expected, $N_E = 8$.

The "Average" in the last row is simply the sum of the 10 values in a given column divided by the number of experiments, $N_X = 10$.

$$\bar{X} = \frac{1}{N_X} \sum_{n=1}^{N_X} X_n , \tag{1}$$

How close is the average N_0 to the expected number, $N_E = 8$? It is only 0.3 away.

The square root of the average value for D^2 is called the root-mean-squared, or "rms" for short. It provides a numerical value for the spread of individual observed values N_O from the expected value N_E. The larger the rms, the more spread out they tend to be. A mathematical formula for the *standard deviation* that we calculated from Table 1 is

$$s_X = \sqrt{\frac{\sum_{n=1}^{N} (X_n - \bar{X})^2}{N-1}} . \tag{2}$$

Let p = 0.5 be the probability for getting a heads. The theoretical value for the mean of N_O is simply $p*N_T$. For the binomial distribution the theoretical value for the rms is

$$rms = \sqrt{p(1-p)N_T} . \tag{3}$$

What is this theoretical value, and is your calculated value for the rms reasonably close to this theoretical value? The theoretical value is 8 heads. The calculated rms is 1.58, so it is not.

Aside: As you do more and more experiments, your mean value for N_O would get closer and closer to the expected value. In fact, the deviation of your mean value for N_O from the expected value decreases in proportion to $\frac{1}{\sqrt{N_X}}$. Similarly, your observed value for the rms gets closer to the theoretical value as N_X increases. This behavior of getting closer to the true value as you collect more and more data is characteristic of statistical uncertainties. Repeating measurements *reduces* statistical uncertainties.

Another aside: the values of N_O in each experiment must fluctuate around N_E. If you got $N_O = 8$ every time, we would know that something was definitely wrong because it is exceedingly unlikely to obtain 8 heads in 10 different sets of 16 tosses. Financial institutions use the patterns of randomness to detect likely cases of fraud – and scientists do, too!

Systematic Uncertainties

Illustration: Natalie wants to buy a new, fuel-efficient car. She asks her friends Henrietta and Holgar for advice. Henrietta says she measured her car's gas mileage when she drove home to Nebraska last fall a year ago and got 31.4 miles per gallon (mpg). Holgar's car has a built-in mileage monitor. He lives in the city and uses his car for running errands in bad weather. Last January he was getting 29.6 mpg. Natalie might conclude that Henrietta's car is more fuel-efficient than Holgar's, but she recognizes that the two values, 31.4 mpg and 29.6 mpg, are not far apart and the conditions under which the measurements were made were very different. Perhaps Holgar's car would deliver a fuel efficiency higher than 31.4 mpg if he drove straight to Nebraska in the spring as opposed to driving in

Chicago traffic on a snowy day in January. The gas mileage value depends on the way it is measured. The fact that Henrietta and Holgar made measurements under very different conditions means that Natalie will have trouble comparing the values they obtained.

Exercise: Metal expands when it is heated to a higher temperature, and contracts when it is cooled. The effect is small, but it is not difficult to observe[3]. As a consequence, a metallic ruler is accurate only for a certain temperature. When it is warm, it tends to give values that are too low, and when it is cold, it gives values that are too high.

Download the SkewedRulers.pdf from the CANVAS website and display it with your Adobe® Acrobat Reader.© If you have a printer, you can print out the rulers. Otherwise, adjust the reader's window and View/Zoom magnification so that the rulers fit your entire screen. Find an item with a standard size (credit card, business card, note card, bank note, etc.) that can be measured on these three length scales. We might prefer that the resulting calibrated scale has cm units; however, different screen sizes among students' computers necessarily results in an uncalibrated measurement scale. Then let us invent an unused length measurement unit (i.e. “bw,” “as,” etc.) whose *definition* is the “Calibrated” scale. Once you start making measurements, don't change the scales' magnification; scrolling is ok but re-scaling is not. What object did you select?

A straw

Measure the item on each of the three scales and record your results in Table 2. Estimate *two* decimal places precision for each measurement by interpolating between the closest two graduated marks. Repeat these measurements to complete Table 2.

Table 2: Length measurements one item at three different temperatures.

Cold Scale (cm)	**Calibrated (cm)**	**Hot Scale (cm)**
21.29	21.00	20.79
21.28	21.01	20.80
21.30	21.02	20.79
21.29	21.01	20.80
21.28	21.01	20.78

Which is the primary uncertainty: the random variations between elements in the same column or the systematic variations between different columns? The systematic variations.

Since the primary uncertainty is systematic variation of length with temperature, we could measure the temperature and use the thermal expansion property of the ruler's metal to compensate our measurements for the thermal expansion and contraction of our ruler. Once this is achieved, all Table 2 entries will agree far better and this systematic uncertainty will be far less. Possibly this correction will make this uncertainty far less than whatever uncertainty was secondary before the correction. Otherwise, more careful temperature and/or thermal expansivity measurements must be made to

3 Bridges must account for thermal expansion or they can buckle in hot weather and tear open in cold. In northern parts of the US you can see expansion grooves which allow for changes in the size of the bridge’s span. They tend to be narrow in the summer and wide in the winter.

improve the experiment further. (Another improvement strategy might include controlling the temperature where the measurements are performed.)

If you did not know the temperature when you measured the dimension of your object, you might have to consider all three values to be equally plausible. In this case, a scientist takes the “normal” value (i.e., the one from the calibrated ruler) as her/his best guess for the true value and uses the range as an indication of the “margin of error” – better described as the systematic uncertainty. So, if you measured a bill’s length and found 13.0 cm, 13.1 cm, and 13.2 cm, then you should report this as 13.1 cm with a systematic uncertainty of 0.1 cm, or: L = 13.1 ± 0.1 cm.

L = (21.01 ± 0.2) cm

Aside: if you measured the same object several times under a fixed set of conditions, you would not be able to eliminate this systematic uncertainty because there is nothing random about the expansion or contraction of the ruler. If you needed to reduce this uncertainty, you would have to control the relevant (environmental) variable, which in this case is the temperature. For such a highly accurate measurement, it would be important to report, as auxiliary information, the temperature at which you made your measurement.

Real World Measurements

Real measurements typically have both stochastic and systematic influences. Usually it is not obvious a priory which character of measurement will dominate. Other times neither aspect dominates and we must combine the two to determine an uncertainty that agrees with observation.

Observations Containing Stochastic Influences

First, make yourself a paper airplane. If you don't know how to do this, click the link to view an instructional video. Take the airplane, a notepad, and a pen or pencil outside so you can measure how far it flies. We will measure this distance 10 times and enter them into Table 3. When you are ready, choose a direction with enough space to fly the airplane and loft it horizontally at moderate speed. Once the airplane stops, walk normally toward the airplane and count your steps until you reach the airplane. Estimate a left-over fraction of a step if an even integer is unnatural. To avoid systematic effects, return to your original location and loft the airplane again in the same direction. Repeat for a total of 10 measurements.

If you are familiar with spreadsheet programs like Microsoft® Excel,© you should enter your measurements there and allow the program to perform the statistics calculations for you. If you click a blank cell and type “=average(<range>)” and press enter, the computer will place the average of <range> into the cell. Instead of typing “<range>” drag your mouse across the 10 data cells. If you click another blank cell and type “=stdev(<range>)” and press enter, the computer will place the rms deviation or standard deviation into that cell. These two calculations are the purpose of Table 3, so you only need to do this by hand if you have no tools to do it quickly.

Since about half of our measurements are above the average and the other half (or so) is below the average, we expect that the average is more representative of the quantity we wish to measure than is any single measurement alone. The average “averages out” some of the measurement uncertainty. We should expect, then, that the uncertainty in the measured average is significantly less than the uncertainties in individual measurements. We will state without proof that the optimum uncertainty for the average is the *standard error* or the *deviation of the mean*

$$s_{\bar{X}} = \frac{s_X}{\sqrt{N}} . \tag{4}$$

Table 3: Measured distances for paper airplane flights.

Flight	X (paces)	$X-\bar{X}$ (paces)	$(X-\bar{X})^2$ (paces^2)
1	3.7	0.06	.0036
2	4.1	0.46	.2116
3	3.0	-0.64	.4096
4	2.5	-1.14	1.2996
5	4.2	0.56	.3136
6	4.5	0.86	.7396
7	4.3	0.66	.4356
8	3.1	-0.54	.2916
9	2.5	-1.14	1.2996
10	4.5	0.86	.7396
Averages	3.64	**0**	0.574

If we perform this experiment twice to get two averages (or means) and two standard errors, $\bar{X} \pm s_{\bar{X}}$ and $\bar{Y} \pm s_{\bar{Y}}$, then these two means will agree within their uncertainties, $|\bar{X}-\bar{Y}| < \sqrt{s_{\bar{X}}^2 + s_{\bar{Y}}^2}$, 68% of the time.

Divide your standard deviation by $\sqrt{10} = 3.1$ and **round off to *two* significant figures**. Count the number of decimal places in this result and round off your average to the same number of decimal places. Report your results here.

$X = ($ 3.64 $\pm$ 0.240 $)$ paces

Since the uncertainties tell us how well we know our measurements, **the pair should always have the same number of decimal places**. The uncertainties tell us which decimal places are "fuzzy." Less significant decimal places are mere noise so we discard them. And rounding to more significant decimal places makes the round off error as big as (or bigger than) the specified uncertainty. In that case the correct uncertainty would be the combined error (or the round off error) instead of the measurement uncertainty. Since we spend money to achieve our uncertainty, throwing it away is silly and wrong.

How uncertainties are important for predictions

Scenario: Suppose your uncle owns a small sports arena and for business purposes he needs to make projections of attendance for future events. On a beautiful Saturday in June he counts 12,100 fans. He wants your input on predicting the number of fans the following Saturday and he wants your prediction to be as accurate and complete as possible.

Unless you have a sophisticated model that enables you to take environmental factors into account, you should take 12,100 as your prediction. The question is: how large could deviations of the actual number of fans be from the number you predict?

1. In situations like this in which you are simply counting (i.e., counting fans), the rms is just the square root of the count (i.e., the predicted number of fans). How big is that?

   110 fans

2. The weather turns out to be rainy, which from past experience can reduce attendance by a third. Since you can't control the weather, you have to take this as a kind of systematic uncertainty on your prediction. How big would that be?

   4033.33 fans

Which effect is larger, quantitatively speaking – the statistical uncertainty or the systematic uncertainty?

Systematic uncertainty

Notice that the statistical uncertainty and the systematic uncertainty have nothing to do with each other. They are independent sources of uncertainty. Often, scientists quote a "total" uncertainty for a measurement. The rule for combining the statistical and systematic uncertainty is:

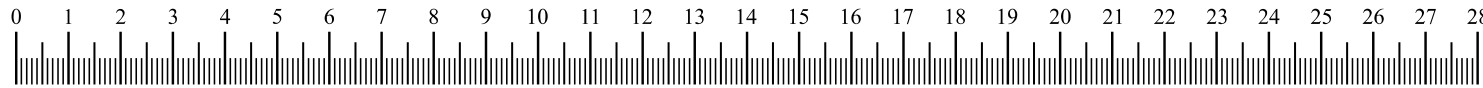
$$(total\ uncertainty) = \sqrt{(statistical\ uncertainty)^2 + (systematic\ uncertainty)^2} \quad (5)$$

i.e., they are "added together in quadrature". If you have several independent sources of systematic uncertainty, then they are added together in quadrature, too. If two sources are not independent, then their correlation must be taken into account, which is complicated and beyond the scope of this course.

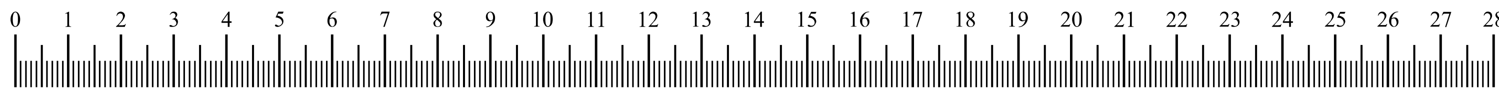
Summary

- Uncertainties are needed for comparing two measured values and for making predictions.
- There are two types of uncertainties:
  - Statistical, which pertain to random effects; and,
  - Systematic, which pertain to uncontrolled environmental factors.
- Repeating the measurement multiple times can reduce statistical uncertainties. Systematic uncertainties can only be reduced by controlling environmental factors.
- In some cases, the statistical uncertainty is larger than the systematic uncertainty, but in other cases it is the other way around. This means that improving some experiments means taking more data, while for others it means exerting greater control over the conditions of the experiment.

Cold Cold Cold Cold Cold Cold



Calibrated Calibrated Calibrated Calibrated Calibrated



Hot Hot Hot Hot Hot Hot