Uncertainties

Random or Statistical Uncertainties

Random or statistical uncertainties correspond to random variations in the results of repeated measurements. These random variations are sometimes due to limitations of the measuring device. For example, electronic noise and air currents lead to random fluctuation in motion detector readings. These fluctuations occur even when the motion detector is measuring the distance to a stationary object. Random variations in repeated measurements can also be a characteristic of the system being measured. For example, if we use a meter stick to measure the landing positions of a series of projectiles shot from a spring-loaded launcher, we see significant random variations which clearly do not arise from the limitations of the meter stick. Instead, we suspect that the launch velocity given to projectiles by the launcher is subject to small random variations.

Truly random variations average to zero, and so the way to remove them is to average several measurements,

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

The **average value**, or **mean value**, $\bar{x}$ approaches the "true value" as the number of measurements in the average approaches infinity. Finding the "true value" is impractical, so we must settle for the "best value" given by the average of a finite set of measurements. When working with more than just a few measurements, it is a good idea to use the `average()` function of your spreadsheet (or the `numpy.average()` function if you are using Python).

Random variations are described by the normal distribution, or Gaussian distribution, or "bell curve." The uncertainty in the "best value" of a large collection of normally distributed measurements can be calculated using the **standard deviation**

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

which describes the width of the distribution. More precisely, about 68% of a normal distribution falls within $\pm\sigma$ of the average value. The standard deviation is the uncertainty in a single measurement in the distribution. Rather than doing this calculation by hand, use the `stdev()` function of your spreadsheet (or the `numpy.std()` function if you are using Python).

The uncertainty in the average of a collection of measurements is less than σ. This follows from the idea that the more measurements we make, the closer the average value comes to the "true value." The **standard deviation of the mean** is given by

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \tag{3}$$

We report this as the uncertainty in $\bar{x}$. (The `sqrt()` function of your spreadsheet or `numpy` computes square roots.)

Systematic Errors

Systematic errors are are due to a defect in the equipment or methods used to make measurements. For example, a motion sensor can be poorly calibrated so that it gives distance readings which are only 90% of the true values. It has a systematic uncertainty (10%) that is much greater in magnitude than the statistical uncertainty in its readings. Systematic errors are often difficult to detect, because they do not show up as variations in the results of repeated measurements. It is important to think about possible sources of systematic errors and to try to correct them or rule them out, for example by

- checking calibrations
- comparing results obtained using independent means (comparing measurements with meter sticks and motion detectors)
- comparing results with accepted values (when available)

Estimating Uncertainties

We often do not have the luxury of a large collection of normally distributed measurements to analyze. Instead, we must somehow estimate the uncertainty of a single measurement. This is necessarily somewhat subjective. If only a few measurements are available, it is more reasonable to use the entire range covered by the measurements to define the uncertainty instead of calculating the standard deviation of the mean. If only one measurement is available the resolution of the device and the variation in the quantity measured are important guides. For example, the resolution of a meter stick is 1 mm. If it is used to measure the length of a rectangular steel plate, an uncertainty of 1 mm, or perhaps even 0.5 mm, is reasonable. If I use the same meter stick to measure the height of a small child, issues of variable posture and how I line up the stick lead to an uncertainty of as much as 1 cm. Be conservative with your estimates. That is, when in doubt, it is a good policy to report a larger uncertainty.

Propagation of Uncertainties in Calculations

Frequently, calculations involve one or more measured quantities, and we need to determine how the uncertainties in input quantities translate into the uncertainty in the result. The guidelines below cover all of the possibilities. Always check that the result and its corresponding uncertainty have the same units. If they do not, something went wrong.

Addition/Subtraction

When adding or subtracting, add absolute uncertainties in quadrature.

For example, if $d = a + b - c$, then

$$\sigma_d = \sqrt{\sigma_a^2 + \sigma_b^2 + \sigma_c^2} \tag{4}$$

Multiplication/Division

When multiplying or dividing, add relative (fractional) uncertainties in quadrature.

For example, if $d = \frac{ab}{c}$, then

$$\frac{\sigma_d}{d} = \sqrt{\left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2 + \left(\frac{\sigma_c}{c}\right)^2} \tag{5}$$

When raising a value to a power, multiply its relative error by the power. For example, if $d = \frac{a^l b^m}{c^n}$

$$\frac{\sigma_d}{d} = \sqrt{\left(l\frac{\sigma_a}{a}\right)^2 + \left(m\frac{\sigma_b}{b}\right)^2 + \left(n\frac{\sigma_c}{c}\right)^2} \tag{6}$$

In General (Approximately)

Use first derivatives to determine the approximate variation of the result due to the uncertainty in each measured quantity. Then, add them in quadrature

If a quantity f is a function of the measured quantities $a, b, c, ...$, then

$$\sigma_f = \sqrt{\left(\frac{\partial f}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial f}{\partial b}\right)^2 \sigma_b^2 + \left(\frac{\partial f}{\partial c}\right)^2 \sigma_c^2 + ...} \tag{7}$$

In General (Better Approximation)

When calculating a result which depends on measured input quantities, determine the variations in the result due to each input quantity, and add the variations in quadrature. In some cases, upper and lower uncertainties differ.

For example, if $d = a \log b$, the individual variances are

$$\sigma_{da+} = |(a + \sigma_a)\log b - d| \tag{8}$$

$$\sigma_{da-} = |(a - \sigma_a)\log b - d| \tag{9}$$

$$\sigma_{db+} = |a\log(b + \sigma_b) - d| \tag{10}$$

$$\sigma_{db-} = |a\log(b - \sigma_b) - d| \tag{11}$$

and the upper and lower uncertainties are

$$\sigma_{d+} = \sqrt{\sigma_{da+}^2 + \sigma_{db+}^2} \tag{12}$$

$$\sigma_{d-} = \sqrt{\sigma_{da-}^2 + \sigma_{db-}^2} \tag{13}$$

This kind of analysis is a good job for a spreadsheet or a Python program. This method is approximate, because interactions between variables are neglected.

Reporting Results with Uncertainties

Explicit Uncertainties

Results with uncertainties are typically reported in the form

$$x \pm \sigma_x \tag{14}$$

Units are *always* included, and are usually given after the result and its uncertainty. It is common practice to round uncertainties to one significant figure. Results should be rounded off to the decimal place of the corresponding uncertainties. For example, if an analysis of several measurements of my height reveals an average of $\bar{h} = 1.8037$ m with a standard deviation of the mean of $\sigma_{\bar{h}} = 0.00566$ m, I report my height as $h = 1.804 \pm 0.006$ m. The form

$$x(\sigma_x) \tag{15}$$

is also sometimes used, where the uncertainty is given as a single digit. In this form, my height is $h = 1.804(6)$ m. The uncertainty is assumed to be in the last reported digit of the result. With asymmetric uncertainties, one uses the form

$$x^{+\sigma_{x+}}_{-\sigma_{x-}} \tag{16}$$

Always include explicit uncertainties when reporting results.

Significant Figures

Often, a numerical value is given without an explicit uncertainty. You should never do this when reporting results of measurements in lab, but we do this all the time when we are solving problems in class, for homework, and on exams. When you write a numerical value without an uncertainty, you imply an uncertainty with the number of **significant figures** you include.

- The implied uncertainty is in the last digit on the right – the **least significant digit**.

  When I say that I am 182 *cm tall, the implied uncertainty is in the 2.*

- The number of significant figures = the number of digits ignoring leading zeros.

  0.000368 *has 3 significant figures. None of the zeros are significant, because they are leading zeros.*

- Notation: Where trailing zeros make the number of significant figures ambiguous, put a bar over the least significant figure, or use scientific notation. (This is also a way to indicate the least significant digit when keeping extra digits in intermediate values in multi-step calculations to avoid round-off error.)

  $52,710$ *might have has 4 or 5 significant figures. The zero is needed in order to express the number, either way. Let's say it has 4 significant figures. It can be written unambiguously as* $52,7\bar{1}0$ *or* 5.271×10^4*.*

- Addition/subtraction: Round the result to the last decimal place of the least precise input.

$$5.827 + 0.61 = 6.4\bar{3}7 = 6.43$$

- Multiplication/division: The result has the same number of significant figures as the input with the smaller number of significant figures.

$$5.827 \times 0.61 = 3.\bar{5}5447 = 3.6$$

L.A. Riley (`lriley@ursinus.edu`), updated June 2021