

Experiment 1

Data Analysis

1.1 Introduction

Some degree of uncertainty exists in any measurement. This uncertainty is called the *error*. Here, an error does not mean a mistake, and neither it is the discrepancy between the student-measured value and that given in a textbook. Rather, it is a quantitative indication of the probability that a further measurement under the same conditions would give a result that falls within a specified range of the reported value. The determination of the error associated with a measurement is part of the general problem of data analysis. In this introduction only a few of the important concepts will be covered.

The other, equally important part of this introductory lab, is to learn to make good use of the computer tools available to you. Modern science is impossible without computer-assisted measurement tools, and one needs to become proficient in their use. This lab is performed on the computer workstations in the lab, and no actual data is collected, but the scripts you will develop during this lab will be the starting point for all subsequent experiments. In fact, it should become a habit to prepare the skeleton scripts in advance of the lab, so that the data acquisition and preliminary analysis can be performed quickly and efficiently.

“Scripts” referred to in the previous paragraphs are, in essence, simple computer programs. We will be executing (running) these scripts within the environment of **eXtrema**, a powerful data analysis and plotting package developed at the TRIUMF accelerator in Vancouver, BC, but the principle of saving sequences of commands in scripts and then running those scripts, applying them to a variety of data files, applies equally well to **Matlab/octave**, **gnuplot**, **python** or any number of other environment. We chose to use **eXtrema** because its syntax is particularly simple, and it is particularly well-suited to the analysis of data in Physics.

It is helpful to develop good habits in organizing your data and files in a meaningful structured manner. To begin, make sure you have a convenient working directory and open a terminal window in it. It is assumed you are familiar with the basic operations of a command-line interface in Linux. The following reminder may help to get you going:

```
mkdir -p ~/2P20/Lab1
cd ~/2P20/Lab1
```

Within that terminal window, you can then start our chosen data analysis program by simply typing **eXtrema** at the prompt. All of the script files for this lab should be placed in this directory.

Introduction to eXtrema

❗ Complete an introductory **eXtrema** tutorial, at <http://www.physics.brocku.ca/Labs/eXtrema>.

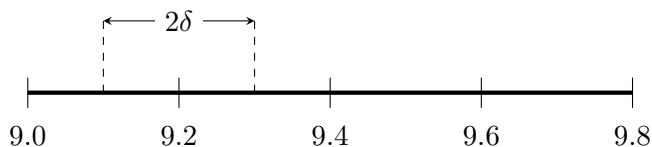


Figure 1.1: When the scale ticks are 0.2 units apart, any measurement within $\delta = 0.1$ units of each tick will be recorded as having the value of the tick itself. The range of true values reported as, say 9.2, will be 9.2 ± 0.1 , meaning anywhere between 9.1 and 9.3, so within a bin of size $2\delta = 0.2$ units which is the precision of the scale.

Table 1.1: Sample experimental measurements of L and T for a pendulum.

Trial #1	Length (cm)	Trial #2	Time (seconds)
1	88.90	1	1.88
2	88.88	2	1.86
3	88.92	3	1.88
4	88.91	4	1.90
5	88.89	5	1.88
6	88.90	6	1.92
7	88.88	7	1.88
8	88.91	8	1.90
9	88.91	9	1.90
10	88.93	10	1.90
11	88.89		
12	88.91		

1.2 Instrumental errors

An analog scale, such as that on a meter stick, consists of a series of equally spaced reference points, or ticks. The spacing between these ticks establishes the precision of the scale. When a single measurement is performed using such an analogue “device”, the best one can do is to report the value of the tick nearest to the measurement. In essence, any measurement within one-half of a step between ticks will be reported as having the value of the tick. This is illustrated in Fig. 1.1. This principle is extended to digital measurements, where the actual readout can be recorded unambiguously, with the understanding that the implied precision is still \pm the smallest digital step of the instrument, typically ± 0.5 in the least significant digit of the display.

Determination of a gravitational acceleration g

In a typical experiment, the value of the gravitational acceleration g is measured by determining the period T of a pendulum of length L . A sample data set is reported in Table 1.1. From the data, g can be calculated

from L and T according to:

$$g = \frac{4\pi^2 L}{T^2}.$$

stat

- ❓ What instrumental errors would you infer for the measurements of the length and the time?
- ❗ While a long careful look at the data table might give you some insight, it is helpful to take a quick visual survey at the data. Use your preferred text editor (**gedit** is suggested) to create the following script, or macro file. The manual will refer to it as **ex2.pcm**, but feel free to use any other meaningful name. Use the file extension **.pcm** to save you some typing in the future, as this is the default.¹

```
window 3
L={88.9;88.88;88.92;88.91;88.89;88.9;88.88;88.91;88.91;88.93;88.89;88.91}
graph L

window 4
T={1.88;1.86;1.88;1.9;1.88;1.92;1.88;1.9;1.9;1.9}
graph T
```

Once the file is saved, it can be called and executed in **eXtrema**, like this:

```
@ex2
```

File extension is not necessary, if **.pcm** (or **.stk**) was used as the extension. From now on, every time you make a change in your **ex2.pcm** file, re-run the macro from the **eXtrema** command line, just like above (up-arrow on keyboard will help). Note how the graph command did not have two vectors (x, y) specified but only one. In such cases, the index of the vector is the assumed x -coordinate.

- ❗ Even this very basic first draft is already helpful in visualizing uncertainty of the measurement. We can quickly improve on it by modifying the macro in the following way:
- add restart commands, so that every time we run the script, we start from a clean slate;
 - improve the quality of the graphs by using discrete symbols, unconnected by a line (by convention, lines are reserved for theoretical curves);
 - add axes' labels;
 - calculate some basic statistics on the data, and show it on the graphs;
 - add comments to help understand what and how the macro is doing, and to personalize it.

Below is the result of the outlined modifications. In the future, in your own macros, try to always incorporate similar improvements early on.

```
! ex2.pcm
!   eXtrema macro used in performing Exercise 2 from Lab 1 of PHYS 2P20
!   By: Student B. Diligent, 2022.09
clear
defaults
destroy\all

set xlabel `Trial No.`

window 3
```

¹ **.pcm** stands for **p**hysica **c**ommand **m**acros; **physica** was the **eXtrema** predecessor. **.stk** extension can also be used.

```

L={88.9;88.88;88.92;88.91;88.89;88.9;88.88;88.91;88.91;88.93;88.89;88.91}
set ylabel `Length, cm'
set plotsymbol -15
set %plotsymbolsize 2
set plotsymbolcolor red
scales 0 0 0 88.7 89.1 4
graph L

stat L L_mean\mean L_std\sdev
set plotsymbol 0
set curvelinetype 1
set curvecolor red
graph\overlay {1:len(L)},{L_mean;L_mean}
set curvelinetype 9
graph\overlay {1:len(L)},{L_mean+L_std;L_mean+L_std}
graph\overlay {1:len(L)},{L_mean-L_std;L_mean-L_std}

window 4
T={1.88;1.86;1.88;1.9;1.88;1.92;1.88;1.9;1.9;1.9}
set ylabel `Time, s'
set plotsymbol -17
set %plotsymbolsize 1.5
set plotsymbolcolor blue
graph T

stat T T_mean\mean T_std\sdev
set plotsymbol 0
set curvelinetype 1
set curvecolor blue
graph\overlay {1:len(T)},{T_mean;T_mean}
set curvelinetype 9
graph\overlay {1:len(L)},{T_mean+T_std;T_mean+T_std}
graph\overlay {1:len(L)},{T_mean-T_std;T_mean-T_std}

```

A few explanations:

- Note how we keep changing the setting of `plotsymbol`. A negative value means “do not draw lines between points”, zero means no points drawn, only the line. We normally do not use it, but a positive value would draw both the point symbols and the line between them.
- We employed a slight increase in `plotsymbolsize`, to 1.5% or 2% of the screen width, instead of the default 1%, but in general the default values for most settings in **eXtrema** work really well.
- To draw the horizontal lines we specified the two vectors on-the-fly as $\{x_1;x_2\}, \{y_1;y_2\}$. This corresponds to a pair of endpoints (x_1, y_1) and (x_2, y_2) between which a horizontal line is drawn (note how $y_1 = y_2$ in our script).

Of course, the solid lines on the graphs represent the mean values, and the dashed ones are one standard deviation away.

1.3 Random errors

In those cases in which only a single measurement is possible, the error reported is necessarily the instrumental error described in Section 1.2. If the situation permits, however, it is good practice to repeat a

measurement several times. If the precision of the instrument is low, it may be found that all measurements are found to be the same within the large instrumental error.

On the other hand, if the instrumental error is reduced by going to a more precise measuring device, it is frequently found that the values so determined differ from one another by more than the instrumental error and these differences generally become more noticeable as the accuracy of the instrument increases. As an example, a digital scale readout may fluctuate by more than a single step in the least significant digit of the display. The differences may arise from a number of causes, some of which may be impossible to control (vibrations, temperature fluctuations in the environment, *etc.*), and some that are extremely sensitive to small variations in the experiment line of sight of the experimenter, *etc.*). The measured quantity may also have an intrinsic variability, such as the number of radioactive atoms that decay or the number of electrons emitted per unit time from a filament at some temperature. In all of these cases the resultant uncertainty in the reported value of the quantity measured can best be represented in statistical terms, as a distribution function, *i.e.* the probability of a measurement yielding a certain value q . Note that the uncertainty being described here arises from fluctuations in measurements and not as deviations from a theoretical description of the results. As you may have encountered before, for randomly fluctuating variables, and in the limit of a large number of measurements, the expected probability function approximates the so-called normal (or Gaussian) distribution about the mean

$$P(q) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(q-\langle q \rangle)/2\sigma^2}, \quad (1.1)$$

where $\langle q \rangle$ is the average (mean) value of all measurements of q , and σ is the *standard deviation* which reports the extent of the variations about the mean. The Gaussian distribution is not the only way to describe a distribution of results of imperfect measurements, but it is reasonable, has a simple analytic form, and “is accepted by convention and experimentation to be the most likely distribution for most experiments. In addition, it has the satisfying characteristic that the most probable estimate of the mean [value] from a random sample of observations of q is the average of those observations”² $\langle q \rangle$. There are some subtleties to this result, which we will discuss in the rest of this Section.

Assume we wish to determine the value of some quantity q and we set out to do so by making a set of n measurements, all made under identical conditions so far as they may be controlled, and all having the same associated instrumental error. We call such a set a “sample.” It may be shown (and is intuitively “obvious”) that the single parameter that best represents the sample is the arithmetic mean, defined by

$$\langle q \rangle = \frac{1}{n} \sum_{j=1}^n q_j. \quad (1.2)$$

Stating $\langle q \rangle$ alone, however, is not sufficient to fully characterize the sample of n measurements. To more fully describe the likelihood that a measurement may yield a certain value is to consider the full distribution function of the measured quantity. To determine that, one can construct a frequency histogram, such as the one shown in Figure 1.2, by plotting along the ordinate axis the number of times n_i the measured quantity q falls in the range $q_i \pm \delta$, against the associated value of q_i . The ticks of the horizontal axis mark off segments or “bins” of width 2δ , and i is the index of the bin center points q_i , not the index j of the individual measurements q_j . Of course, to convert a frequency histogram into a probability distribution, one divides by the total number of counts in all bins, $P(q_i) = n_i/n$, since $\sum_{i=1}^n n_i/n = 1$.

A full distribution function is very descriptive, but a simplification can be realized by introducing a simple quantitative measure of “spread” of values about $\langle q \rangle$. Since the individual values of q are distributed about $\langle q \rangle$, we can define the *deviation* of, say, the j^{th} value q_j from the mean as

$$\Delta q_j = (q_j - \langle q \rangle). \quad (1.3)$$

²P.R. Bevington and D.K. Robinson. *Data reduction and error analysis for the physical sciences*. 2nd ed., McGraw Hill, 1992.

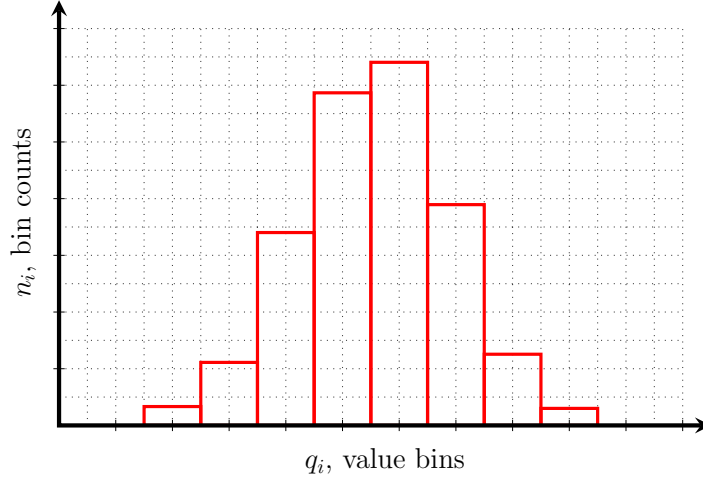


Figure 1.2: Histograms are generated by dividing the range of measurement into “bins” and counting how many individual measurements fall within the range of each bin. In the limit of a large number of measurements of a randomly-varying quantity the histogram shape will approximate a normal distribution.

We might be tempted to use the average of the deviations, *i.e.*,

$$\langle \Delta q \rangle = \frac{1}{n} \sum_{j=1}^n \Delta q_j,$$

as a measure of the spread of q values in our sample. However, this is not a fruitful approach (we will test and confirm that in Section 1.3). Instead, we calculate the mean *squared* deviation, called the *variance*, and use it as a measure of the spread; the square root of the variance is the *standard deviation* σ :

$$\sigma^2 = \langle (\Delta q)^2 \rangle = \lim_{n \rightarrow \infty} \left[\frac{1}{n} \sum_{j=1}^n (q_j - \langle q \rangle)^2 \right]. \quad (1.4)$$

A true measure of σ can only be obtained in the limit of an infinite number of measurements; for a finite n , careful analysis shows that a more appropriate measure is the *standard deviation of the sample* S , defined through

$$S^2 = \frac{1}{n-1} \sum_{j=1}^n (q_j - \langle q \rangle)^2 \quad (1.5)$$

which is what we will use from now on. This is also what **eXtrema** reports as the “standard deviation” in all of its statistical functions.

S is a measure of the width of the histogram; a large S value corresponds to a wide spread of values around the average value. If one further measurement of q was done, there is a 68% probability that it will fall in the range $\langle q \rangle \pm S$, a 95% chance that it will fall in the range $\langle q \rangle \pm 2S$. S is therefore an indication of the error of a particular (single) measurement. The term *error* here does not refer here to a deviation from some theoretical or “known” value. Since we do not know what the true value is, this is at best an *estimate* of the error inherent in our measurement.

In those experimentally rare situations when the “true mean” of the distribution is known, we could take several samples, calculate the mean of each sample and see how these means are distributed. The standard deviation of this “distribution of means” is called the *standard deviation of the mean*, and it determines the probability that our sample mean $\langle q \rangle$ lies within a certain range of the true mean. Statistical theory shows that we can calculate it as $\sigma_m = S/\sqrt{n-1}$. As we increase the number of samples, the mean will

get closer to the true mean value, S will approach the true width of the underlying true distribution, and $\sigma_m \rightarrow 0$. Note this difference: S (calculated over n measurements) predicts the precision of a *single* additional measurement, whereas σ_m predicts the precision of the average of *many* sets of measurements. As n increases, S may not change very much, while σ_m will decrease.

In practice, for less than an infinite number of measurements, our best efforts can only yield *estimates* of the underlying “true” distributions, and we take our *sample variance* S^2 as the best estimate of the variance σ^2 , in order to estimate the uncertainty σ_q of our measurement $\langle q \rangle$. The result of a set of n repeated measurements quoted as

$$q = \langle q \rangle \pm S_q$$

represents our best estimate of the true uncertainty σ_q in the measurement of q . The two quantities, S_q and σ_q , are different, but in practice S_q is the best measurement-based estimate of σ_q .

Basics of statistical analysis in eXtrema

- ❗ Show (analytically) that the average deviation $\langle \Delta q_j \rangle$ is always zero. Confirm (numerically) for our L and T .
- ❗ In Section 1.2 we used **eXtrema**’s **statistics** command to calculate averages and standard deviations. This time, calculate the standard deviation *of the sample* for L and T , but from first principles, as required by Equation 1.5, and verify that what **eXtrema** reported are the same numbers that are obtained “by hand”. No need to break out a handheld calculator, use **eXtrema** itself to do these step-by-step calculations: **eXtrema** is an excellent general-purpose calculator.

You may want to cut and paste the commands one-by-one into the command prompt, to see the effect of each one. Afterwards, you can execute the entire macro at once.

```
! ex3.pcm
!   eXtrema macro used in performing Exercise 3 from Lab 1 of PHYS 2P20
!   It assumes @ex2 has been called previously, so that L and T are known vectors.
!   If not, uncomment the first line of code here
!   By: Student B. Diligent, 2022.09

!@ex2

= ` ***** Calculation of basic statistical parameters for L '

N=len(L)
= sum(L)
L_m=sum(L)/N

= sum(L-L_m) ! pause and think here...
= sum((L-L_m)^2)
L_s2=sum((L-L_m)^2)/(N-1)
L_s=sqrt(L_s2)

= ` Previously calculated mean(L) = '//rchar(L_mean)//` +/- '//rchar(L_std)
= ` Manually calculated   mean(L) = '//rchar(L_m)//` +/- '//rchar(L_s)

= ` '

= ` ***** Calculation of basic statistical parameters for T '

N=len(T)
= sum(T)
T_m=sum(T)/N
```

```
=sum(T-T_m) ! pause and think here...
=sum((T-T_m)^2)
T_s2=sum((T-T_m)^2)/(N-1)
T_s=sqrt(T_s2)

=Previously calculated mean(T) = '//rchar(T_mean)//` +/- '//rchar(T_std)
=Manually calculated mean(T) = '//rchar(T_m)//` +/- '//rchar(T_s)
```

Constructing a distribution function

The mean and the standard deviation are the simplest forms of statistical analysis of the data. A more complete picture is provided by the full probability distribution for the scatter of the data. For this, we need to construct the histograms of our distributions. With only a relatively few points in our datasets, our histograms can only have a small number of bins.

❗ Construct histograms for L and T .

The following **eXtrema** macro is offered without comment, but it does perform the required operations. Study it, use the Help facility as needed, and add comments in appropriate places, to ensure you fully understand all of the commands used here. For example, the size of **plotsymbol** for histograms means something different, use the Help facility to figure out what the commands do, and enter appropriate comments in the right places in the macro file. You are expected to include this modified, fully commented macro file, in your first lab report.

```
! ex4.pcm
! eXtrema macro used in Section 1.3.2, Lab 1 of PHYS 2P20
! By: Student B. Diligent, 2022.09
clear
defaults
destroy\all

L={88.9;88.88;88.92;88.91;88.89;88.9;88.88;88.91;88.91;88.93;88.89;88.91}
N_L=len(L)

window 5
set
xlabel `Trial No.'
ylabel `Length, cm'
%xlaxis 15
plotsymbol -15
plotsymbolcolor red
curvelinetype 1
curvecolour red

scales 0 0 0 88.84 88.98 7
graph L
stat L L_mean\mean L_std\sdev
set plotsymbol 0
graph\overlay {1:len(L)},{L_mean;L_mean}
set curvelinetype 9
graph\overlay {1:len(L)},{L_mean+L_std;L_mean+L_std}
graph\overlay {1:len(L)},{L_mean-L_std;L_mean-L_std}

get ymin L_1 ! match the range of the histogram to what's displayed
get ymax L_2
get ynlincs steps
L_step=(L_2-L_1)/steps
```



```

bins=[L_1-3*L_step:L_2+3*L_step:L_step]
bin L,bins,counts

window 6
scales L_1 L_2 0 0 max(counts)+2 0
set
  xlabel `Length, cm'
  ylabel `Count'
  %xlaxis 15
  %ylaxis 15
  curvelinetype 1
  curvecolour red

graph\histogram bins,counts
set curvelinetype 9
set curvecolour blue
generate ll bins[1],,bins[#] 200
graph\overlay ll,normal(ll,L_mean,L_std)*(max(bins)-min(bins))/len(bins)*N_L

T={1.88;1.86;1.88;1.9;1.88;1.92;1.88;1.9;1.9;1.9}
N_T=len(T)

window 7
set
  xlabel `Trial No.'
  ylabel `Time, s'
  %xlaxis 15
  plotsymbol -17
  %plotsymbolsize 1.5
  plotsymbolcolor red
  curvelinetype 1
  curvecolor red

scales 0 0 0 1.82 1.96 7
graph T

stat T T_mean\mean T_std\sdev
set plotsymbol 0
graph\overlay {1:len(T)},{T_mean;T_mean}
set curvelinetype 9
graph\overlay {1:len(T)},{T_mean+T_std;T_mean+T_std}
graph\overlay {1:len(T)},{T_mean-T_std;T_mean-T_std}

get ymin T_1          ! match the range of the histogram to what is displayed
get ymax T_2
get nlyinc steps
T_step=(T_2-T_1)/steps

window 8
bins=[T_1-3*T_step:T_2+3*T_step:T_step]
bin T bins,counts
scales T_1 T_2 0 0 max(counts)+2 0
set
  ylabel `Count'
  xlabel `Time, s'
  %xlaxis 15
  %ylaxis 15

```

```

plotsymbol 0
curvelinetype 9
curvecolor red

graph\histogram bins,counts
generate tt bins[1],,bins[#] 200
set curvecolor blue
graph\overlay tt,normal(tt,T_mean,T_std)*(max(bins)-min(bins))/len(bins)*N_T

window 0
set
color black
textalign 1
textinteractive 0
%xtextlocation 12
%ytextlocation 87
%textheight 3.5

text '<b1>a)'
set %ytextlocation 37
text '<b1>b)'
set %xtextlocation 62
set %ytextlocation 87
text '<b1>c)'
set %ytextlocation 37
text '<b1>d)'

```

1.4 Propagation of errors

Very frequently a parameter of interest is not measured directly, but is deduced from one or more parameters that are measured, through some functional relationship. For example, the gravitational constant g is approximately given in terms of length L of a pendulum and the period T of its swing by

$$g = \frac{4\pi^2 L}{T^2} . \quad (1.6)$$

If measurements of both T and L have errors associated with them, we need to know how these errors combine to give rise to an uncertainty in g .

In the following we shall consider a general case, in which the quantity z is a function of other directly measurable parameters x, y, \dots , expressed as

$$z = z(x, y, \dots) . \quad (1.7)$$

The most probable value of z is given by

$$\langle z \rangle = z(\langle x \rangle, \langle y \rangle, \dots) , \quad (1.8)$$

and the uncertainty in $\langle z \rangle$ can be obtained by examining the scatter of values resulting from the calculations using the individual measurements x_i, y_i, \dots , namely

$$z_i = z(x_i, y_i, \dots) . \quad (1.9)$$

The variance σ_z^2 (and therefore the standard deviation $\sigma_z = \sqrt{\sigma_z^2}$) are calculated using

$$\sigma_z^2 = \lim_{n \rightarrow \infty} \left[\frac{1}{n} \sum (z_i - \langle z \rangle)^2 \right] . \quad (1.10)$$

Using a Taylor series' expansion, Equation 1.10 can be written as

$$\begin{aligned}\sigma_z^2 &\approx \lim_{n \rightarrow \infty} \frac{1}{n} \sum \left[(x_i - \langle x \rangle) \left(\frac{\partial z}{\partial x} \right) + (y_i - \langle y \rangle) \left(\frac{\partial z}{\partial y} \right) + \dots \right]^2 \\ &\approx \lim_{n \rightarrow \infty} \frac{1}{n} \sum \left[(x_i - \langle x \rangle)^2 \left(\frac{\partial z}{\partial x} \right)^2 + (y_i - \langle y \rangle)^2 \left(\frac{\partial z}{\partial y} \right)^2 + 2(x_i - \langle x \rangle)(y_i - \langle y \rangle) \left(\frac{\partial z}{\partial x} \right) \left(\frac{\partial z}{\partial y} \right) + \dots \right]\end{aligned}\quad (1.11)$$

where the partial derivatives are evaluated with all other variables held constant at their mean values, *i.e.* $(\partial z / \partial x)_{\langle y \rangle, \dots}$ is implied. The first two terms recognizably contain σ_x^2 and σ_y^2 ; the third term arises from products of deviations in x, y, \dots which are expected to be uncorrelated for independently measured quantities, and thus can be expected to average out to zero. This third term involves *covariance* of x and y , defined as

$$\sigma_{xy}^2 \equiv \lim_{n \rightarrow \infty} \left[\frac{1}{n} \sum (x_i - \langle x \rangle)(y_i - \langle y \rangle) \right]$$

Similar arguments apply to all other pairs of variables, and thus Equation 1.11 gets simplified:

$$\begin{aligned}\sigma_z^2 &\approx \sigma_x^2 \left(\frac{\partial z}{\partial x} \right)^2 + \sigma_y^2 \left(\frac{\partial z}{\partial y} \right)^2 + \dots + \sigma_{xy}^2 \left(\frac{\partial z}{\partial x} \right) \left(\frac{\partial z}{\partial y} \right) + \dots \\ &\approx \sigma_x^2 \left(\frac{\partial z}{\partial x} \right)^2 + \sigma_y^2 \left(\frac{\partial z}{\partial y} \right)^2 + \dots\end{aligned}\quad (1.12)$$

with similar terms for all other variables. Equation 1.12 is called *error propagation equation*. It is worth repeating here that the latter, abbreviated form of Equation 1.12 is only valid in the case of completely uncorrelated fluctuations in the contributing variables. If an erroneous measurement x_i is always accompanied by a compensating erroneous measurement y_i , the covariance term may yield a significant erroneous increase (or decrease, since a positive contribution is not assured) in the calculated value of σ_z .

A weighted sum

If $z = ax \pm by$, where a and b are constants, the partial derivatives are simply $(\partial z / \partial x) = a$ and $(\partial z / \partial y) = \pm b$. Equation 1.12 immediately yields

$$\sigma_z^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2 \pm ab \sigma_{xy}^2 \approx a^2 \sigma_x^2 + b^2 \sigma_y^2.$$

One can say that errors are additive here.

Products and ratios

If $z = \pm axy$ (a is again a constant), then $(\partial z / \partial x) = \pm ay$ and $(\partial z / \partial y) = \pm ax$ and the expression for σ_z becomes

$$\frac{\sigma_z^2}{z^2} = \frac{\sigma_x^2}{x^2} + \frac{\sigma_y^2}{y^2} + 2 \frac{\sigma_{xy}^2}{xy},$$

and similarly for $z = \pm a(x/y)$,

$$\frac{\sigma_z^2}{z^2} = \frac{\sigma_x^2}{x^2} + \frac{\sigma_y^2}{y^2} - 2 \frac{\sigma_{xy}^2}{xy}.$$

If covariance σ_{xy}^2 is small, *i.e.* fluctuations in x and y are uncorrelated, both expressions reduce to

$$\frac{\sigma_z^2}{z^2} \approx \frac{\sigma_x^2}{x^2} + \frac{\sigma_y^2}{y^2}.$$

Here, the relative errors are additive.

Powers

If $z = ax^{\pm b}$, a and b constants, then

$$\left(\frac{\partial z}{\partial x}\right) = \pm abx^{\pm b-1} = \pm b \frac{z}{x}$$

and the relative errors are related by

$$\frac{\sigma_z}{z} = \pm b \frac{\sigma_x}{x}.$$

The above rules can be combined, for example:

$$z(x, y, r, \dots) = \frac{x^p y^q \dots}{r^t \dots} \longrightarrow \frac{\sigma_z^2}{z^2} \approx p^2 \frac{\sigma_x^2}{x^2} + q^2 \frac{\sigma_y^2}{y^2} + t^2 \frac{\sigma_r^2}{r^2} + \dots \quad (1.13)$$

Computational exercises

- ① Calculate the average value of g using Equation 1.6. One of the following two methods requires that there is a T_j value for every L_j value, so first trim both vectors to the shorter of the two vector lengths.

```
! Define a number of useful physical constants, we need Pi
define\constants
```

```
vector L T min(len(L),len(T))
! First method, use mean L and T values
g1=4*Pi^2*mean(L)*1e-2/mean(T)^2
! Second method, calculate N values of g, and mean of the results
g2=mean(4*Pi^2*L*1e-2/T^2)
=Compare: '//rchar(g1)//` vs. '//rchar(g2)
```

- ① Calculate (on paper) $(\partial g / \partial L)_T$ and $(\partial g / \partial T)_L$ from Equation 1.6. Evaluate the numerical values of these partial derivatives using mean values $\langle L \rangle$ and $\langle T \rangle$.
- ① Using **eXtrema**'s commands, estimate σ_L and σ_T , and combine with the values for the above partial derivatives to calculate σ_g directly, as in Equation 1.12.
- ① Yet another approach is to combine relative errors, as in Equation 1.13.

```
! First method, combine relative errors
dg1=g1*sqrt((stdev(L)/mean(L))^2+(2*stdev(T)/mean(T))^2)
! Second method, calculate N values of g, and stdev of the results
dg2=stdev(4*Pi^2*L*1e-2/T^2)
=Compare: '//rchar(g1)//`+/-'//rchar(dg1)//` vs. '//rchar(g2)//`+/-'//rchar(dg2)
```

Are all of your results in agreement?