Measurement and Uncertainty[1]

Introduction

Measurement attempts to quantify some attribute (the **measurand**). It does this by comparing–typically with an instrument–the attribute with a commonly adopted standard reference (such as a clock, a ruler, a mass scale, etc.). The standard reference used defines the **units** of the measurement (s, m, kg, etc.).

- Units used must be specified for a measurement to mean anything.

The outcome may be compared to measurements of similar attributes among objects or events.

A measurement and the measurand are not the same thing: the measurement is an estimate of the measurand. The final result of a measurement is an interval. Thus, measurement (in contrast to "counting") is a statistical procedure that samples (chooses one among many possible values from) one or more **probability distributions**, which are characterized, for example, by a central value and a width.

- Different ways of measuring can–but may very well not–obtain (essentially) the same value of the measurand.
- Repeating measurements in a certain way may result in a distribution of measured values, rather than a single, repeated value.
- Each way of measuring most likely samples a different probability distribution, which is described mathematically with a **probability density function**, or **PDF**.
  - The distribution of measured values may differ, perhaps systematically, when measured in different ways. For example,
    - the central value of each distribution may differ;
    - each distribution may spread around the central value differently.

Numerous factors influence a measurement. The number of factors, and the degree of control of these factors, affects the size of the **uncertainty**. Formally, uncertainty defines an interval within which the measurand is presumed, with a definable probability, to reside. This interval is called a **confidence interval**. It indicates, with a probability, how likely it is that the measurand lies within its boundaries.

- The uncertainty must be stated for a quantitative measurement to mean anything.
- Measurement uncertainty should not be confused with mistakes or errors. "Human error" is not a valid source of uncertainty. If a mistake is made, repeat the measurement without the mistake.

[1] PHYS 263 is a physics course, so this unit is written for physics students. While the basic concepts are the same regardless of discipline, similar presentations are available for other majors:

Beginner's guide to measurement in electronic and electrical engineering

Beginner's guide to measurement in mechanical engineering

This is not all there is to making a measurement. In professional physics experimentation, at least as much time and effort are spent determining the uncertainty and other characterizations of a measurement as are spent estimating the value of the measurand. Ultimately, the components of a measurement include:

- a symbol representing the measurand;
- the best estimate of the value of the measurand;
- the (total) uncertainty of the estimate;
- units (if any);
- the confidence interval; and
- the PDF used to estimate the total uncertainty and confidence interval,

such that the result of a measurement should be reported as:

> $Z =$ (estimated value of the measurand ± uncertainty) units (confidence interval, PDF)

where, again, the confidence interval is a percentage indicating the probability that the measurand lies within the range defined by ± the uncertainty of the estimated value.

> *In terms of the numbers reported, the precision (that is, the number of decimal places) of the uncertainty determines the precision of the measurement. The uncertainty may contain no more than 2 significant figures.*[2] *In any case, the number of decimal places for the estimated value of the measurand must match the number of decimal places of the uncertainty, even if the number of significant figures of the value differs from that of the uncertainty.*

For example, if the uncertainty in a measurement of resistance was determined to be 1.25 Ω, it would be reported to be 1.3 Ω (if the experimenter chooses to report the uncertainty with two significant figures, 1 Ω if the choice is one significant figure). If the measurement of a resister's resistance yielded 17.34 Ω, it would be reported, then, as, say

> $R = (17.3 \pm 1.3)\Omega$ (58%, Rectangular PDF)

If the measurement yielded 1.73 Ω, it would be reported

> $R = (1.7 \pm 1.3)\Omega$ (58%, Rectangular PDF)

If the measurement yielded 0.17 Ω, it would be reported

[2] The point is that the uncertainty indicates, among other things, the limits of the measurement estimation: nothing concrete can be said beyond the first or second decimal place at which uncertainty arises. Some authorities argue that only 1 significant figure is sufficient; the National Institute of Standards and Technology (NIST) and others suggest 2 significant figures. See the discussion below regarding fractional uncertainty for a suggestion as to when 1 significant figure may be justified.

$R = (0.2 \pm 1.3)\Omega$ (58%, Rectangular PDF)

If the measurement yielded 0.02 Ω, it would be reported

$R = (0.0 \pm 1.3)\Omega$ (58%, Rectangular PDF)

Note that the PDF and confidence interval are just examples here. The experimenter chooses and determines these.

The quality of a measurement is related to the uncertainty in two complementary dimensions: **accuracy** and **precision**. An accurate measurement includes the true value of the measurand within the confidence interval around the estimate. This can never be irrefutably known to be the case, but can be confirmed or refuted by additional measurements, particularly if made with different techniques. A precise measurement is one in which the ratio of the uncertainty to the measured value–known as the **relative uncertainty**–is small. By comparing this ratio, measurements can be said to be more or less precise than other measurements. But note that increasing precision may decrease accuracy, while decreasing it may increase accuracy. This is what is meant by calling these complementary.

Types of Uncertainty

In practice, it's useful to distinquish uncertainty by two operationally different methods:

- statistical methods, whose results are called type A uncertainty
- all other methods, whose results are called type B uncertainty

Both methods quantify random and systematic effects. Modern metrology (the science of measurement) doesn't distinguish random from systematic uncertainties, but type A from type B uncertainties.

Statistical, or type A, uncertainties are those assumed to be due to random fluctuations among measurements, the result of, say, changes in the state of the measurand (e.g., the surface on which the measurand rests vibrates while measuring its length), changes in the state of the measuring instrument (e.g., reading the ruler from different angles during a series of length measurements), changes in the state of the environment (e.g., the heating/cooling system goes on and off), changes in the measuring procedure (e.g., different persons take turns measuring the length). Note that though these affects may occur randomly, the result may be systematic (i.e., the length of solids tends to be proportional to the temperature).

Type B uncertainties often dominate type A uncertainties. They are quantified primarily through scientific judgment based on previous measurements, experience or general knowledge, manufacturer's specifications, or calibration or other reported data. Some of these, though quoted, may require interpretation by the experimenter: are the specifications one or several units of standard uncertainty; what PDF was used; and so forth. For many purposes, it's safe to assume, with regard to the measuring instrument, that:

- if all the values within clearly-defined limits are equally likely, the instrument scale can be represented by a rectangular PDF.
- if likely values cluster around a single value within clearly-defined limits, the instrument scale can be represented by a triangular PDF.

- if likely values cluster around a single value, but the limits of possible values are unclear or undefined, the data sample a normal (Gaussian) PDF.

Characterizing Probability Density Functions

A number of different statistics are used to characterize a PDF. The two most common ones are mean and standard deviation.

For an arbitrary distribution, the **mean** can be found, for N discrete values, x_i, with the formula

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

while for a continuous distribution, where the PDF is represented by $f(x)$,

$$\bar{x} = \int_{-\infty}^{\infty} x f(x) dx \tag{2}$$

A statistical measure of the spread or width of a distribution around its mean is given by the **variance** or, in the same units as the mean, the square root of the variance, also known as the **standard deviation**. For discrete values, the variance is calculated:

$$s^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1} \tag{3}$$

or for a continuous distribution,

$$s^2 = \int_{-\infty}^{\infty} (x-\bar{x})^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \bar{x}^2 \tag{4}$$

is the variance, and the standard deviation is $s = \sqrt{s^2}$.

The standard deviation is a statistical quantity, and therefore a type A uncertainty, which, when combined properly with type B uncertainties, gives the **standard uncertainty** of the measurement.

Rectangular Probability Density Function

When any value between defined limits is equally possible, the probability distribution is described by a **rectangular PDF**. Because the integral over all possible values of a PDF equals 1, if the half-width of the rectangle is a, then the full width is $2a$, and the height of the rectangle is $1/2a$. A rectangle is obviously symmetric around its center, which must therefore be the mean of the distribution. Thus, the PDF is defined by

$$f(x) = \begin{cases} \frac{1}{2a}, & (\bar{x} - a) < x < (\bar{x} + a) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

1. **(a) Show that the variance of a rectangular PDF of half-width a is**

$$\boldsymbol{s^2_{\textbf{rectangle}} = \frac{a^2}{3}}$$

and that the standard deviation is

$$\boldsymbol{s_{\text{rectangle}} = \frac{a}{\sqrt{3}}}$$

(b) **Show that the proportion of the total rectanglar PDF that is bounded by $\bar{x} \pm s_{\text{rectangle}}$ is approximately 58%.**

A stable digital reading gives a value whose precision is limited by the number of digits displayed. The measurand could have any value between the next higher or lower reading in the last decimal place, known as the **least significant digit**.[3] This suggests that reading the instrument samples a rectangular PDF. Assuming this to be the case, then the uncertainty due to the instrument's finite scale is $s^{\text{scale}}_{\text{rectangle}} = \frac{a}{\sqrt{3}}$, where a is the half-width of the scale. This is the same as the statistical uncertainty of a continuous rectangular distribution, but here the rectangular PDF is assumed; no amount of data will show this. So the uncertainty due to the digital instrument's finite scale is a type B uncertainty. Repeated measurements of the same measurand with the same digital instrument typically exhibit small statistical uncertainty (type A), so the (type B) uncertainty associated with the finite scale of the instrument is likely to be much larger. The instrument may also have limited intrinsic accuracy (check specifications or perform a calibration), another type B uncertainty. All contribute to the total uncertainty.

The Triangular Probability Density Function

Each of the six faces of a fair die is equally likely to face up after the die is rolled. This means that any of the values on the face of a die, 1 to 6, is equally likely to be the outcome of a roll, which is equivalent to saying that the probability of getting a number betweeen 1 and 6 (inclusive) is 1/6, and zero otherwise.

2. **Complete Table 1 for the case when two fair dice are rolled. [“Unique Combinations” means different ways of getting a sum–(2,3) and (3,2) are unique combinations that sum to 5; ”Number” is the number of unique combinations that make the sum; “Probability” is the ratio of number of unique combinations for one sum to the total number of possible unique combinations.]**

[3] It may be that the digital instrument rounds its reading, such that the measurand could have any value between halfway to the next lower reading and halfway to the next higher value.

Table 1: Probabilities for two fair dice.

Sum	Unique Combinations	Number	Probability
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			

Scatter Plots

When two or more quantities are measured in the same experiment, it is usually the case that one of the variables (the **independent variable**) is varied in order to determine how the other variable(s) [the **dependent variable(s)**] consequently vary. Each measurement, regardless of variable, will be subject to uncertainties, and these need to be estimated. The independent variable should never be relatively more uncertain than the dependent variable(s). Otherwise, no relationship suggested by a fit to the data could be taken seriously. (Fitting will be discussed below.) Ideally, relative uncertainties associated with the independent variable should be negligible in comparison to those of the dependent variable(s).

The relationship between an independent variable and some number of dependent variables is shown graphically in a **scatter plot** or 2-dimensional plot, in which the independent variable is given by the abscissa (x-axis) and the dependent variable(s) by the ordinate (y-axis). When making a scatterplot, be sure to title it and label its axes (including units). It is also important to show uncertainties (as so-called **error bars**) associated with the points on the plot.

3. **A frictionless cart with inertia $m = 0.5$ kg is rolled along a smooth horizontal surface. The cart subsequently (at $t = 10$ s) begins rolling up an equally smooth incline of 25 degrees (0.44 radians). Its velocity is measured with a radar gun that logs the velocity every second for the next 15 seconds. With your programming language of choice, make a scatterplot of the data collected:**

t [s]	v [m/s]	s_v [m/s]
10	5	9
11	6	5
12	1	10
13	3	2
14	0	3
15	0.1	0.2
16	0	4
17	-2	4
18	-2	7
19	-2	13
20	-2	17
21	-9	6
22	-15	1
23	-5	21
24	-7	22
25	-7	25

The plot should look something like the one in Figure 1

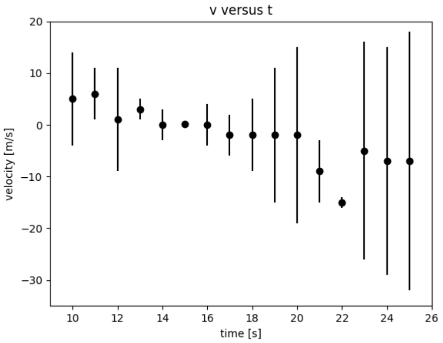

Figure 1: Scatterplot of velocity versus time. Error bars indicate 1 standard uncertainty in the measurement.

4. **With some programming language (not Excel), plot the probabilities in Table 1 against the associated sums to reproduce Figure 2.**

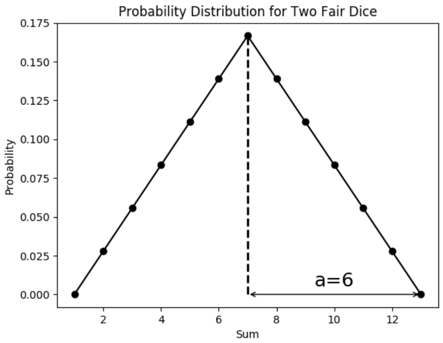

Figure 2: Probability distribution for two fair dice. Note the symmetry of the distribution (an isosceles triangle). The connected dots define a triangular PDF for the roll of two fair dice. The symmetry implies that the center of the distribution is also the mean. Also shown is the half-width, a.

The symmetry of this distribution implies that the mean, median, and mode are all the same: the center of the distribution. In general, a symmetric triangular probability distribution has a mean at the middle of the distribution, a half-width a, and a height $1/a$.

Treating the sums of two rolled dice as a continuous distribution, with x the sum of two dice faces, the PDF is

$$f(x) = \begin{cases} \frac{1}{36}(x-1), & 1 \leq x \leq 7 \\ \frac{1}{36}(13-x), & 7 < x \leq 13 \\ 0 & \text{Otherwise} \end{cases} \tag{6}$$

so that the integral for the mean becomes

$$\bar{x} = \frac{1}{36}\left[\int_1^7 x(x-1)dx + \int_7^{13} x(13-x)dx\right] \tag{7}$$

5. **Check that the three ways [central value, discrete distribution, continuous distribution] of determining the mean of the "perfect" triangular distribution all give the same answer.**

6. **Show that the standard deviation of the "perfect" two dice triangular PDF equals $\sqrt{6}$, and that, in general, the standard deviation of a triangular PDF is:**

$$s_{\text{triangle}} = \frac{a}{\sqrt{6}}$$

   where a is the PDF half-width.

7. **By integrating the triangular PDF between $\pm s_{\text{triangle}}$ of the mean, or with methods of analytic geometry (areas of triangles), show that the area bounded by ± 1 triangular PDF standard deviation is approximately 65%**

of the total area of the triangle. [Note that the total area under a PDF is 1, by definition of probability.]

Recall that the area bounded by ±1 rectangular PDF standard deviation is approximately 58%. It is important to recognize that the one standard-uncertainty interval (once type B uncertainties are combined with the type A uncertainties) around the central value has a different interpretation for different PDFs.

- The PDF sampled must be specified for a measurement to mean anything.

Histograms

A **histogram** is a kind of (vertical) bar graph displaying how numerical data are distributed. There is no separation between histogram bars. The bars correspond to intervals of values of the quantity represented along the horizontal axis. The intervals are referred to as **bins**. Dividing the horizontal axis into bins is a preliminary step to filling in the histogram. The height of each bar is related to the number or fraction of data falling within the bin interval.

8. **Simulate 36 rolls of a pair of dice by writing a program (not in Excel) that adds a pair of random integers between 1 and 6 (inclusive), each with equal likelihood, 36 times. Have the program plot a histogram of the results. Be sure to title, and label the axes of, the histogram. Set the abscissa (x-axis) limits to 0 - 14. The result should look something like Figure 3. How does the simulation compare to the "perfect" distribution?**

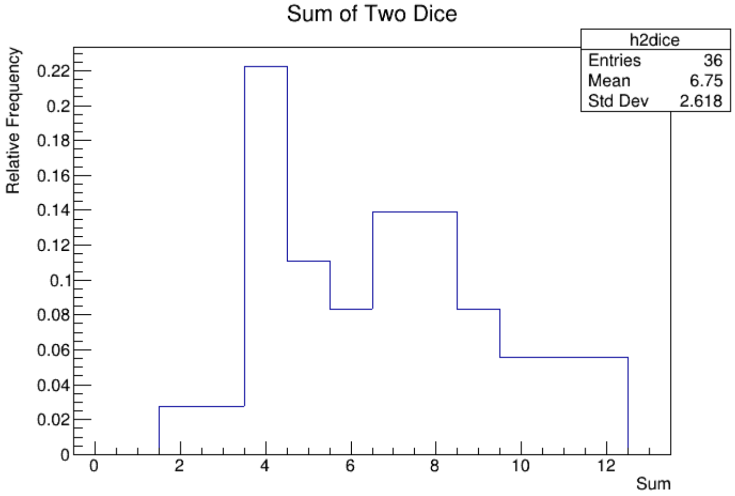

Figure 3: Histogram of a simulation of two rolled dice. The ordinate (y-axis) is "relative frequency" (also known as the "empirical probability", which here is the bin frequency divided by 36). This makes the comparison with the "perfect" (or "theoretical") distribution easier, but is not necessary.

Summing the face values of two fair dice is a counting exercise, not an experiment, but the triangular PDF is applicable to many experimental situations, particularly if the uncertainty is clearly bounded both above and below the measured value. An example of such a situation is a measurement with a ruler in which the edge being measured is unquestionably between two ruled hash marks. Recall, though, that assigning a half-width is a judgment call, a compromise between accuracy and precision, so, perhaps, the hash marks are clearly farther apart than necessary, or the edge being measured is not halfway between the marks.

In general, analog measuring instruments display (continuously) varying scale *positions*, in contrast to digital measuring instruments, which display *numbers* with finite significant digits. Reading an analog scale is a measurement process in and of itself, with its own PDF resulting from such effects as shifting parallax, indicator jitter, and clarity of scale, in addition to scale resolution. Some of these may affect statistical uncertainty (type A), but others of these have to be estimated by judgment (type B uncertainty).

When analog jitter is small, as for steady digital measurements, limits beyond which the measurand cannot fall can be determined, but, unlike in the digital case, selecting these limits–and thereby the half-width–is a judgment call, one that requires a compromise between accuracy (the true value of the measurand is included in the range of the full-width) and precision (minimizing the uncertainty). Each limit should be sufficiently far from the central value that its (the limit's) probability is zero, but the interval between the limits should be as narrow as possible to minimize the estimated uncertainty. Also different from digital measurements, values of analog measurements are usually less likely to be nearer the limits than to be centered between them. The uncertainty introduced by such an analog scale is well-approximated by a triangular PDF, and is a type B uncertainty. Its size is the same as that of a triangular data set, like the two-dice distribution: $s_{\text{triangular}}^{\text{scale}} = a/\sqrt{6}$, where a is half the separation of the limits. The standard deviation of a series of measurements of the same measurand with the same analog instrument is unlikely to be zero, and may be roughly the same magnitude as the scale uncertainty. Other uncertainties, including those mentioned above, also contribute to the total uncertainty. These must be included to better estimate the total measurement uncertainty.

The Normal (Gaussian) PDF

The normal or Gaussian PDF, recognizable as the familiar bell-shaped curve, has the mathematical form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{8}$$

where μ is the value of the mean, median, and mode of the distribution (assuming an infinite number of measurements), and σ is the standard deviation [standard (statistical) uncertainty] of the distribution, which is the uncertainty of an *individual* measurement taken with a given instrument and technique. For a finite number of measurements, N, these statistics are estimated by the mean, Equation 1, and the standard uncertainty, the square root of the variance, Equation 3.

Approximately 68% of the normal distribution lies within ±1 standard deviation of the mean, compared to 58% of the rectangular distribution and 65% of the triangular distribution. The small difference between the spread of triangular and normal distributions means that they are often interchangeable: whichever is easier to use under the circumstances may be assumed.

9. **Measure your reaction time by starting a digital stop watch and trying to stop it exactly at 2.00 s. After practicing a few times, record 15 or 20 attempts. Decide which uncertainty–statistical, scale precision (number of displayed decimal places), or accuracy (from manufacturer's specification)–dominates the spread in your results. Justify your answer quantitatively. [The best estimate of your reaction time is the total uncertainty in the data around 2.00 s.]**

10. **Measure 20 times with a digital stop watch the time it takes an 8.5-in by 11-in piece of paper (letter) to fall from shoulder height to the floor (only trials in which the paper hits nothing on the way down should be included).**

    (a) **Histogram the results. Which PDF (rectangle, triangle, normal) best describes the shape of the histogram?**

    (b) **Determine the (statistical) standard uncertainty of the data. How does this number compare with the (statistical) standard uncertainty around 2.00 s of your reaction time?**

    (c) **List all non-statistical (type B) uncertainties. Estimate their magnitude and identify their associated PDF.**

Confidence Levels

The reason for identifying the probability distribution function (PDF) being sampled by a measurement is to be able to interpret quantitatively the standard uncertainty of the measurement. Recall that for a rectangular distribution, $s_{\text{rectangle}} = a/\sqrt{3}$ (where a is the half-width of the rectangle, including both type A and type B uncertainties), and the fractional area of the distribution bounded by $\pm s_{\text{rectangle}}$ is 58%. For a triangular distribution, $s_{\text{triangle}} = a/\sqrt{6}$ (where a is the half-width of the triangle, including both type A and type B uncertainties), and the fractional area of the distribution bounded by $\pm s_{\text{triangle}}$ is 65%. And for a normal (Gaussian) distribution, for which the statistical standard uncertainty is the standard deviation, the fractional area of the distribution bounded by $\pm s_{\text{normal}}$ (which includes both type A and type B uncertainties) is 68%. Then, for example, if the central value of a measurement (or, more likely, a series of measurements) is x, then $x \pm s_{\text{rectangle}}$ indicates that, at the 58% confidence level, the measurand is between $\mu - s_{\text{rectangle}}$ and $\mu + s_{\text{rectangle}}$. If the chosen PDF were triangular or normal (Gaussian), the confidence level is 65% or 68%.

11. (a) **What confidence levels are being claimed by $x \pm 2s_{\text{normal}}$?**

    (b) **What confidence levels are being claimed by $x \pm 3s_{\text{normal}}$?**

    (c) **What coverage factor n [the number (not necessarily an integer) of normal (Gaussian) standard uncertainties, $s_{\text{normal}} \to ns_{\text{normal}}$] gives a 90% confidence level?**

Combining Uncertainties

Repeated measurements of the same measurand with the same digital instrument typically exhibit small statistical uncertainty (type A), so the (type B) uncertainty associated with the finite scale of the instrument is likely to be much larger. Both uncertainties may in turn be smaller than the intrinsic accuracy (also type B) of the instrument (check specifications or perform a calibration). Each uncertainty contributes to the total uncertainty and should be identified explicitly in any report of the measurement and combined into a total uncertainty.

On the other hand, the statistical uncertainty (type A) of repeated measurements of the same measurand with the same analog instrument may be roughly the same magnitude as the (type B) uncertainties of the scale and other contributions (see discussion above). As in the digital case, each uncertainty contributes to the total uncertainty and should be identified explicitly in any report of the measurement and combined into a total uncertainty.

Unless there is reason to think otherwise, all contributions should be independent of one another, and so (as will be soon shown) added in quadrature. So, for example, in the digital case:

$$s_{\text{total}} = \sqrt{(s_{\text{statistical}})^2 + (s_{\text{scale}})^2 + (s_{\text{accuracy}})^2} \tag{9}$$

Because estimated standard uncertainties of different PDFs bound different confidence intervals, Equation 9 is not valid generally. When combining estimated standard uncertainties of different PDFs, the experimenter must decide which PDF, and thus which confidence interval, will ultimately be reported and scale the uncertainties representing other PDFs accordingly. For example, an experimenter intending to report a rectangular PDF from a combination of an uncertainty representing a rectangular PDF with an uncertainty representing a triangular PDF, must scale the triangle uncertainty to the size of the rectangle uncertainty:

$$s_{\text{triangle}} \Rightarrow \frac{0.58}{0.65} s_{\text{triangle}} = s_{\text{rectangle}} \tag{10}$$

for this uncertainty.

Equation 9 should be written, then, as

$$s_{\text{total}} = \sqrt{(k_{\text{statistical}} s_{\text{statistical}})^2 + (k_{\text{scale}} s_{\text{scale}})^2 + (k_{\text{accuracy}} s_{\text{accuracy}})^2} \tag{11}$$

where each of the k_i is the appropriate scale factor. Of course, if one contribution is much smaller than the others, it may be ignored in this calculation, though it still must be noted in the report (so its absence from the s_{total} calculation is justified). A measurement m should then be reported as:

> m = (estimated value of the measurand ± total uncertainty) units (confidence interval, PDF)

where the PDF is the one chosen by the experimenter to represent the measurement, the total uncertainty is as given by Equation 11 (with the relevant uncertainties scaled as necessary), and the confidence interval corresponds to the chosen PDF. The total uncertainty may be scaled by a coverage factor, as well.

12. Report in full the measurements made in Problems 9 and 10

Uncertainty Propagation[4]

Whenever more than one uncertainty is involved, whether because a single measurement is subject to multiple sources of uncertainty, or because a result combines multiple measurements each with its own uncertainty (often both are involved), the total uncertainty is obtained by combining the uncertainties in a procedure referred to as **propagation of uncertainties**.

If z results from combining (through some mathematical function) multiple measurements, x_i, each with its own standard uncertainty, s_i, (i enumerates each measurement), then $z = z(x_i)$. In the general case, the combined uncertainty in z is calculated:

$$s_z^2 = s_c^2 = \sum_{i=1}^{N} \left(\frac{\partial z}{\partial x_i} \right)^2 s_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\partial z}{\partial x_i} \frac{\partial z}{\partial x_j} r_{ij} \tag{12}$$

where N is the total number of measurements, and r_{ij} is the **correlation coefficient**:

$$r_{ij} = \frac{1}{N(N-1) s_i s_j} \sum_{k=1}^{N} (\delta x_k)_i (\delta x_k)_j \tag{13}$$

where $(\delta x_k)_i = x_{k,i} - \bar{x}_i$. The correlation coefficient is a numerical quantity between -1 and $+1$ which indicates the strength of relationship between two quantities (here, measurement i and measurement j). The partial derivatives are known as **sensitivity coefficients**, because they express how much the result changes for a unit change in each contribution (measurement).

When $r_{ij} = 0$, the uncertainties from the two contributions (measurements) are uncorrelated (or independent), and Equation 12 reduces to

$$s_z^2 = s_c^2 = \sum_{n=1}^{N} \left(\frac{\partial z}{\partial x_i} \right)^2 s_i^2 \tag{14}$$

that is, the square of the rate at which the combined measurement changes with a unit change in each measurement times the square of each measurement's standard uncertainty all summed over all measurements that contribute to the combined measurement. The square root of this is the combined standard uncertainty.

If the quantity of interest is the product of two measurements (for example, velocity is the product of the displacement and the inverse of the time interval over which the displacement occurred),

$$z = y_1 y_2$$

which have (independent) uncertainties, s_1 and s_2, then the combined standard uncertainty is

$$s_z = s_c = y_1 y_2 \sqrt{\left(\frac{s_1}{y_1} \right)^2 + \left(\frac{s_2}{y_2} \right)^2}$$

[4] Usually, and problematically, called **error propagation**.

since $\frac{dz}{dy_1} = y_2$ and $\frac{dz}{dy_2} = y_1$.

13. Show that if

$$\boldsymbol{z = \frac{1}{y_1 y_2}}$$

where $\boldsymbol{z}$ is the inverse of the product of two measurements $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ which have independent standard uncertainties $\boldsymbol{s_1}$ and $\boldsymbol{s_2}$, then

$$\boldsymbol{s_z = s_c = \frac{1}{y_1 y_2}\sqrt{\left(\frac{s_1}{y_1}\right)^2 + \left(\frac{s_2}{y_2}\right)^2}}$$

Uncertainty propagation implies that the **uncertainty of a mean** is *not the standard deviation*, which gives the (statistical) uncertainty associated with a single measurement. The mean is the sum of multiple measurements, so uncertainty propagation is required. Assuming the same PDF (same standard deviation for all measurements, $s_{x_i} = s_x$ for all i):

$$s_{\bar{x}} = \frac{1}{N}\sqrt{\sum_{i=1}^{N} s_x^2} = \frac{1}{N}\sqrt{N}\ s_x = \frac{s_x}{\sqrt{N}} \tag{15}$$

In words, the uncertainty of the mean decreases from the standard deviation as the square root of the sample size. Although each measurement (done with the same instrument and technique) is equally uncertain no matter how many measurements are taken, the mean, calculated with these measurements, becomes less uncertain as the number of measurements increases, as the square root of the number of measurements.

Confidence in the estimate of the standard deviation also increases with sample size. According to statisticians, the uncertainty in a standard deviation estimate is

$$s_{s_x} = \frac{s_x}{\sqrt{2(N-2)}} \tag{16}$$

This quantity is known as the **fractional uncertainty** (not to be confused with relative uncertainty, which is the ratio of the total uncertainty to the measurand). It can justify giving the total uncertainty with one significant digit: although convention favors two significant digits, if the fractional uncertainty is no better than 10%, one significant figure is sufficient.

Comparing Measurements

A measurement can be intrinsically interesting, but experiments are often done to test a prediction or to check or verify another measurement. Such tests and checks amount to making a comparison.

A frequently proposed, and *always wrong*, comparison method is the so-called percent-difference. Such a method is worse than useless, as it is entirely misleading. The significance of a difference depends on the magnitude of the difference *relative to the uncertainties involved*, not on the (relative) magnitude of that difference. A big difference is not at all surprising if the uncertainties involved are large; a small difference may be significant if the uncertainties are even smaller. Never claim a percent difference to be indicative of anything.

Problems 8 and 10b ask that two quantities be compared, that is, whether the quantities are (significantly) different. Quantitatively, this is asking if the difference between quantities is (significantly) different from zero. A significant difference is one that is large relative to the uncertainties associated with the values being compared. In Problem 8, the question is whether two distributions' $\bar{x}$ and $s_{\text{statistical}}$ differ significantly from one another, given the uncertainties of each statistic. In Problem 10b, the question is whether the two distributions' $s_{\text{statistical}}$ differs significantly from one another, given the uncertainties of each $s_{\text{statistical}}$.

Comparisons like these, between two quantities, are in fact simple examples of combined measurements: given two quantities y_1 and y_2 with uncertainties s_1 and s_2, then, if $z = y_2 - y_1$ is significantly different from zero with respect to s_1 and s_2, y_1 and y_2 are significantly different from one another (given their respective uncertainties).

From Equation 14 the combined variance of a difference is

$$s_z^2 = s_c^2 = s_1^2 + s_2^2$$

since $\left(\frac{dz}{dy_1}\right)^2 = 1$ and $\left(\frac{dz}{dy_2}\right)^2 = 1$, so the combined standard uncertainty associated with difference is

$$z_c = s_c = \sqrt{s_1^2 + s_2^2}$$

(Recall Equation 9 for the combined uncertainty of three independent uncertainties.)

The result for the comparision, then, is (assuming s_1 and s_2 are independent and sample the same PDF),

$$z = \left[(y_2 - y_1) \pm \sqrt{s_1^2 + s_2^2}\right] \text{ units (CL\%, PDF)}$$

The bigger the difference between measurements y_1 and y_2, the farther $z = y_2 - y_1$ is from zero, but if $\sqrt{s_1^2 + s_2^2}$ is bigger than $|z| = |y_2 - y_1|$, then the interval between $(y_2 - y_1) - \sqrt{s_1^2 + s_2^2}$ and $(y_2 - y_1) + \sqrt{s_1^2 + s_2^2}$ contains zero, and so the difference is not likely significant. But this can be quantified, by assigning a probability that the two quantities are different (or, equivalently, 1 − the probabiliity that they are the same). This probability is found by interpreting the ratio between the measured difference and the combined uncertainty in terms of area under the sampled PDF.

The ratio is commonly referred to as a **t-test** (or **student's t-test**)[5]:

$$t_{\text{difference}} = \frac{|y_2 - y_1|}{\sqrt{s_1^2 + s_2^2}} \tag{17}$$

where y_1 and y_2 are the values being compared, and s_1 and s_2 are their respective standard uncertainties. These should be scaled appropriately, so that the probabilities represented by each are equivalent. Then, the denomenator is itself an uncertainty–that of the difference–with a definite probability associated with the PDF of the difference. The ratio will thus give the number of standard uncertainties of the difference, and can be translated into a confidence level.

[5] And the inverse of the relative uncertainty of the difference.

Under any PDF, if the measured difference is zero, then the probability that the quantities being measured are significantly different is indistinguishable from zero (to the precision of the measurements, the two quantities are the same).

If $|z| = |y_2 - y_1| = \sqrt{s_1^2 + s_2^2}$, then the ratio between the measured difference and the combined uncertainties equals ±1. The probability associated with such a difference depends on the PDF: 58% for the rectangular distribution (as was proved in Problem 1b), 65% for the trianglar PDF (as was proved in Problem 7), and 68% for the normal (Gaussian) PDF. Equivalently, the measurements have a 42% probability of being the same if sampled from a rectangular PDF, a 35% probability of being the same if sampled from a triangular PDF, and a 32% probability of being the same if sampled from a normal (Gaussian) PDF.

Tables 2, 3, and 4 list probabilities associated with some simple ratios of measured differences to combined uncertainty for rectangular, triangular, and normal (Gaussian) PDFs. Probabilites associated with other ratios can be found in tables or calculated on-line, although extrapolating between values should be close enough–and exact for the rectangular PDF.

Table 2: Significance probabilities; rectangular PDF.

$\frac{\lvert y_2-y_1\rvert}{\sqrt{s_1^2+s_2^2}}$	Difference Probability [%]	No Difference Probability [%]
0	0	100
0.5	29	71
1	58	42
1.5	87	13
1.732	100	0

Table 3: Significance probabilities; triangular PDF.

$\frac{\lvert y_2-y_1\rvert}{\sqrt{s_1^2+s_2^2}}$	Difference Probability [%]	No Difference Probability [%]
0	0	100
0.5	37	63
1	65	35
1.5	85	15
2	97	3
2.45	100	0

In practice, the experimenter chooses the PDF for the probability test, scaling all uncertainties as necessary, and sets the confidence interval (probability) for interpreting the result. In particular physics, for example, the convention is to test results against the normal (Gaussian) PDF. A 5 standard uncertainty difference $\left(\frac{|y_2-y_1|}{\sqrt{s_1^2+s_2^2}} > 5\right)$ is required to claim a

Table 4: Significance probabilities; normal (Gaussian) PDF.

$\frac{\lvert y_2-y_1\rvert}{\sqrt{s_1^2+s_2^2}}$	Difference Probability [%]	No Difference Probability [%]
0	0	100
0.5	38	62
1	68	32
1.5	87	13
1.64	90	10
2	95	5
2.33	98	2
2.5	99	1
3	99.7	0.3
5	99.99994	0.00006

discovery, while a 3 standard uncertainty difference is required to confirm the claim. These are very strict conditions; not all scientists outside of particle physics follow this convention.

14. **Interpret the results of Problems 8 and 10b: are the differences significant? Justify the interpretation quantitatively.**

15. **Measure with a ruler the circumference and diameter of a ring. Does the ratio of these measurements agree with expectation? [Consider all uncertainties of each measurement. If the same ruler is used for both measurements, then most type B uncertainties are perfectly correlated, so, in Equation 12, r_{ij} is $+1$ when i and j both refer to the perfectly correlated type B uncertainties, such as ruler resolution, and 0 otherwise (this is a rough approximation, which ordinarily would have to be tested).]**

Fitting Data

Practically everything taught in physics courses (and in most other courses, as well) is divorced from reality: frictionless surfaces; perfectly spherical, homogeneous planets; no air resistance; and so on. There is nothing wrong with this. In fact, fruitful idealizations like these provide clues about the nature of complicating factors (like friction) when experiment is compared to the predictions of such idealizations.

The formal name for scientific idealizing is **modeling**. A model aids understanding and guides imagination by attempting to abstract from the complications of the real world fundamental quantities and patterns relating them.

Among the goals of experiment are testing and suggesting models. In physics, a model is operationalized as an equation or set of equations that is supposed to describe and explain some phenomenon, while predicting what will happen during an investigation of some aspect of that phenomenon. The experimentalist collects data and with them checks the match between equations and behavior. Alternatively, when an experiment encounters a phenomenon about which no model exists, an empirical or functional description of the data in terms of a curve can suggest an appropriate model. Either way, it should be clear that fitting curves to data is an important scientific practice.

Mathematically, of course, curves are equations or functions relating an independent variable to one or more dependent variables. Such functions are themselves characterized by their parameters. For example, a linear model, $y = mx + b$, is characterized by two parameters, the slope, m, and the intercept, b. An exponential model, $y = Ae^{bx}$, is also characterized by two parameters, the coefficient, A, and the rate constant, b. Determining the parameters of an equation/function/model is referred to as **fitting**.

The most basic test of a model is determining how well its functional form follows the data. The model may also predict the magnitude and/or sign of the parameters. Once the parameters of a fit to the data are found (along with the uncertainties associated with each parameter), the quality of the model can be assessed.

While it's not too hard to write a program that fits a function to data, fitting packages are available for most scripting languages.[6] Most packages require as input:

- the function (model);
- the values of the independent variable;
- the associated values of the dependent variable(s);
- initial guesses for the parameters.

If, as should be the case, the standard uncertainties associated with the dependent value(s) have been determined, these uncertainties should also be provided, as well. Some packages require uncertainties for both the independent and dependent variables, but others assume the uncertainty of the independent variable to be much smaller than the uncertainties of the dependent variable(s), and so ignorable. A fit curve should pass nearer points with smaller error bars, perhaps even missing points with the largest error bars. This procedure is referred to as **weighted fitting** [analogous to averaging data (**weighted mean**) with varying uncertainties].

Such packages return the parameters of the model function along with a **covariance matrix**, the diagonal elements of which are the variances [normal (Gaussian)] of the fit parameters. The square root of these variances is the standard (statistical) uncertainty associated with the respective parameter.

Two measures for quantifying (to some extent) how well a model function fits the data will soon be presented. A more fundamental, graphical procedure should always be carried out first when evaluating fit quality. The procedure involves plotting **residuals**:

$$r_i = y_i - \hat{y}_i \tag{18}$$

where $\hat{y}_i$ is the "predicted" value of a dependent variable for each value of the independent variable. That is, the output of the resulting fit function for each value of the independent variable, x_i, input.

Two plots of residuals should be made and evaluated:

[6] If familiar with python, but not with fitting data with python, take a look at the example code:
`http://physics.gmu.edu/~rubinp/courses/161/curvefit.py`

1. A histogram of r_i, which, if the fit were good, would be normal (Gaussian) with a mean of zero. For unweighted fitting, the mean of the residuals is zero by construction, but the histogram may still be skewed, which suggests a bad fit. A good *weighted* fit yields a normally distributed residual histogram centered on zero.

2. A scatterplot of r_i vs x_i [note that this expression is always in the order ordinate (y-axis) vs abscissa (x-axis)], in which, for a good fit, the points plotted are distributed horizontally (a line through them would have zero slope). A sloped, curved, or periodically varying distribution indicates a poor fit, suggesting a wrong or incomplete model (fit function).

The plots help interpret quantitative measures of fit quality, which may, by accident of the data, be better than they should be.

One measure of fit quality is given by the **coefficient of determination**, R^2, the fraction of the variation in the data accounted for by the fit. A value near 1 suggests the model explains most of the variability of the dependent variable around its mean; a value near zero suggests that the model explains little of the variability. It is calculated from the data and fit results as follows:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{19}$$

where SS_{res} is the sum of squared vertical distances of each point from the fit curve (sum of squared residuals),

$$SS_{res} = \sum_i r_i^2 \tag{20}$$

where r_i is the residual of each y_i relative to $\hat{y}_i$ for each x_i, and SS_{tot} is the sum of squared differences of each y_i from the mean of the y-values (total sum of squared differences),

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \tag{21}$$

R^2 is difficult to interpret quantitatively: how close to 1 is a good fit and how close to 0 is a poor fit? Because SS_{tot} grows faster than SS_{res} (which on average shouldn't change at all), R^2 systematically increases with the number of data points. It also increases with the number of parameters, which makes the model less predictive. Furthermore, since it does not take into account the estimated standard uncertainties of the dependent variable, it tends to be systematically smaller for weighted fits than for unweighted fits. Finally, its value can exceed 1 and be less than 0 for non-linear fits, which can't be interpreted as the fraction of explained variance in the data and suggests that the mean of the dependent variable is actually a better predictor of outcomes than is the model. In short, use it with caution, particularly with non-linear models.

A somewhat better measure of fit quality is the **reduced chi-squared**, or chi-squared per degree of freedom, χ^2_ν, where χ^2 is the sum of squared vertical distances of each point from the fit curve reduced by the estimated variance (estimated standard uncertainty squared) associated with each point,

$$\chi^2 = \sum_i \frac{r_i^2}{s_i^2} \tag{22}$$

where r_i is the residual, and s_i is the estimated standard uncertainty, associated with each value of the dependent variable, while ν is the number of degrees of freedom, which, in this case, is the difference between the number of data points and the number of parameters in the model,

$$\nu = \text{Number of data points} - \text{Number of parameters} \tag{23}$$

Thus,

$$\chi_\nu^2 = \frac{\chi^2}{\nu} \tag{24}$$

The chi-squared is a probability distribution which depends on the number of degrees of freedom. In general, Equation 22 implies that the more the data deviate from the curve (the worse the model fits the data), the larger the value of the chi-squared, but, because it's a sum of positive-definite values, the chi-squared grows as the size of the data set grows. The reduced chi-squared, Equation 24 naturally removes this dependence on the size of the data set.

However, there's another reason for a chi-squared being large, or small: the standard uncertainties are poorly estimated. While $\chi_\nu^2 \gg 1$ does indicate a poor model fit, $\chi_\nu^2 > 1$ may suggest that either the model does not fit the data, or the estimated standard uncertainties of the dependent variable are too small. On the other hand, $\chi_\nu^2 < 1$ indicates that the model fits the data "too well," which could indicate that either the model is too sensitive to statistical fluctuations in the data (which can be the case when, for example, the model has too many parameters), or the estimated standard uncertainties of the dependent variable are too large. In principle, then, only when $\chi_\nu^2 = 1$ (or close to it) can it be comfortably concluded that the model matches the data and the uncertainties in the data have been reasonably estimated. This, too, seems (and is) qualitative, but more advanced statistics are required to do better.

16. **In Table 5 [due to Anscombe, *The American Statistician*, Vol. 27, No. 1 (Feb. 1973), pp. 17-21] are four data sets (the first three use the same x values) that all have the same number of observations, means in x and y, linear fit function ($\hat{y} = 0.5x + 3$), R^2 and residual sum of squares (no uncertainties accompany the values, so a χ^2 cannot be computed).**

    (a) **Plot each set of data and draw the linear fit function.**

    (b) **Calculate the residuals in each case and create the two residuals plots described above.**

    (c) **Interpret the plots and state whether and how the residual plots "explain" the data and the problem (if any) with the fits.**

Table 5: Four sets of orderd-pair data.

Data set Variable	1-3 X	1 Y	2 Y	3 Y	4 X	4 Y
Observation no.						
1	10	8.04	9.14	7.46	8	6.58
2	8	6.95	8.14	6.77	8	5.76
3	13	7.58	8.74	12.74	8	7.71
4	9	8.81	8.77	7.11	8	8.84
5	11	8.33	9.26	7.81	8	8.47
6	14	9.96	8.1	8.84	8	7.04
7	6	7.24	6.13	6.08	8	5.25
8	4	4.26	3.1	5.39	19	12.5
9	12	10.84	9.13	8.15	8	5.56
10	7	4.82	7.26	6.42	8	7.91
11	5	5.68	4.74	5.73	8	6.89

[Note that these data exaggerate systematic effects for the sake of demonstration. In real data, such effects can be much more subtle, requiring careful study of the residual plots.]

17. **Assume the acceleration of the cart in Problem 3 was constant. Under this assumption, the measurements of velocity and time should be sufficient to estimate this acceleration and the velocity of the cart before it began rolling up the incline with a linear function.**

    (a) **Fit the data with the velocity a linear function of the time.**

    (b) **Find the coefficient of determination.**

    (c) **Find the reduced chi-squared.**

    (d) **Is the assumption of constant acceleration reasonable?**

    (e) **What was the acceleration?**

    (f) **What was the cart's velocity before it started up the incline?**

As noted previously, if a variable in a scatterplot has significant uncertainties, it must be the dependent variable, or else there is no point in fitting. Weighted fitting assumes uncertainties in only the dependent variable. Should the independent variable be appreciably uncertain (though, again, no more significantly uncertain than the dependent variable), one method for taking this uncertainty into account is to estimate its possible effect on the dependent variable, an approach called **effective variance**, illustrated in Figure 4.

The data are first fit ignoring the independent variable uncertainties. The magnitude of shifts in the values of the dependent variable predicted by the fit parameters is evaluated for each value of the independent variable shifted by its uncertainty to get a set of effective standard uncertainties which should then be added (in quadrature) to the original dependent variable uncertainies.

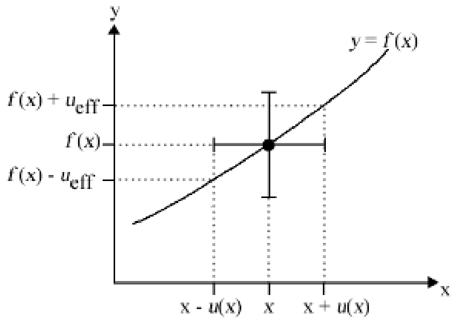

Figure 4: A sketch illustrating how an effective standard uncertainty in the dependent variable due to uncertainty in the independent variable is determined from a preliminary fit ignoring the independent variable uncertainty. The symbol u has the same meaning as s. [From Module 5, *Uncertainty in Physical Measurements*, David N. Harrison.]

Thus, if the curve found by a fit without considering independent variable uncertainties is $\hat{y} = f(x)$. Then,[7] $d\hat{y} = f'(x)dx$ implies $s(\hat{y}_i)_{\text{eff}} = f'(x)s(x_i)$,[8] through which the uncertainty of each value of the dependent variable is adjusted to compensate for the uncertainty in its respective independent variable partner:

$$s(y_i)_{\text{total}} = \sqrt{s(y_i)^2 + s(\hat{y}_i)^2_{\text{eff}}} \tag{25}$$

The data should then be refit using the compensated (total) dependent variable uncertainty.

18. **A frictionless cart of mass M is accelerated by the tension in a massless, inelastic string attached to the side of the cart. The string is hung over a massless, frictionless pulley. A series of weights are attached to the loose end of the string. The force on the cart is then computed with $F = mg$. The cart's acceleration is measured. By appropriately fitting the data in Table 6, determine M, the cart's mass. Provide all details (plots, quantities, etc.) to justify your determination.**

Table 6: Force on, and acceleration of, a frictionless cart under constant tension.

F $[N]$	**a $[m/s^2]$**
$\mathbf{0.25 \pm 0.03}$	$\mathbf{0.6 \pm 0.1}$
$\mathbf{0.74 \pm 0.03}$	$\mathbf{1.4 \pm 0.1}$
$\mathbf{1.23 \pm 0.03}$	$\mathbf{2.4 \pm 0.2}$
$\mathbf{1.72 \pm 0.03}$	$\mathbf{3.4 \pm 0.3}$

[7] Note that u is an alternative symbol for s, the standard uncertainty.

[8] The prime symbol, $'$, means derivative with respect to the independent variable. Thus, $f'(x) \equiv \frac{df(x)}{dx}$.

Counting Experiments

Counting experiments measure (temporal) rates at which something occurs. If the counting is subject to statistical fluctuations (for example, if the time between occurences varies randomly, as is the case for radioactive decays), then the measurement samples a Poisson probability density function:

$$f(N) = \frac{N^N e^{-N}}{N!} \tag{26}$$

where N is the number of counts recorded. The mean of a Poisson distribution is

$$\overline{N} = N, \tag{27}$$

that is, the count itself, and the standard uncertainty is

$$s_N = \sqrt{N}, \tag{28}$$

the square root of the count.

The measured counting rate is

$$n = \frac{N}{t} \tag{29}$$

where t is the time it took to count N events, and the standard uncertainty of the rate is

$$s_n = \frac{\sqrt{N}}{t} = \sqrt{\frac{n}{t}} \tag{30}$$

When N and n are both large ($\gtrsim 100$), the Poisson PDF approximates the shape of a normal (Gaussian) PDF.

Most counting experiments count background events along with signal events, resulting in a rate that is the sum of two rates: $n = n_{\text{signal}} + n_{\text{background}}$. The background rate is typically estimated by counting events when no source is present during an interval $t_{\text{background}}$. The background rate measurement also samples a Poisson PDF.

19. Show that, in the presence of background, the standard uncertainty of the signal rate becomes

$$\boldsymbol{s_{n_{\text{signal}}} = \sqrt{\frac{n}{t} + \frac{n_{\text{background}}}{t_{\text{background}}}}} \tag{31}$$

Given that both signal + background events and background-only events have to be counted in a limited amount of (total) time, the optimal time interval for background counting is related to the interval for signal + background counting by the square root of the ratios of the count rates:

$$t_{\text{background}}^{\text{optimal}} = t\sqrt{\frac{n_{\text{background}}}{n}} \tag{32}$$

20. A Geiger counter records 900 counts in 10 minutes far from a known source of radioactivity, and 1500 counts in 10 minutes near the source.

(a) **What is the counting rate for the source, $n_{\mathbf{source}}$?**

(b) **What would have been the ideal counting interval $t_{\mathbf{background}}$ for the background rate determination, given that the experiment had to be completed in 20 minutes?**