Statistics & Random Error

Experiment Two

Physics 191
Michigan State University

Before Lab

- **Due at the beginning of class (10 pt)**: answer the pre-lab theory questions in Section 1
- Carefully read the entire lab guide

Overview

This is a single-week experiment designed to highlight the inherent randomness in physical measurements. Your lab report should be formatted as explained in the *Lab Report Guidelines* document.

- Observe the random character of human reaction time
- Distinguish between systematic and random errors
- Study the properties of the normal distribution (bell curve)
- Interpret the meaning of standard deviation
- Quantify confidence intervals

Motivation

Quantitative experiments always involve some form of uncertainty estimation. As you learned in the previous lab, there is not an explicit set of rules defining exactly how an uncertainty should be determined. Instead, an experimenter must estimate uncertainties based on the unique set of conditions under which the experiment was performed. Imagine doing some experiment in a modern science lab. Then imagine using the exact same equipment to perform the exact same experiment, but this time outside in a blizzard. Clearly, all sources of error would be exacerbated by the extreme weather. There is no rule about how blizzards effect errors. Instead, the experimenter must quantify the effects of the blizzard.

To further complicate things is the issue of *confidence* in a measured value. Consider the statement, "it takes me 8 ± 3 minutes to ride the bus to school." This seems like a reasonable estimation for most situations. If the bus happens to break down one day and it takes 25 minutes to get to school, should the new estimate be 8 ± 17 minutes? This seems grossly overstated. While it is possible to have a *very* slow trip, the chances of such a delay are quite unlikely. One significant delay does not mean you should start getting out of bed 15 minutes earlier.

This highlights an important issue with uncertainties. There is a level of confidence associated with any uncertainty. Maybe the 8 ± 3 minute estimate for the travel time has a 95 % confident limit to account for the one slow trip (this amounts to 1 problem for every 20 commutes). Such a high confidence level is *why* you feel safe keeping your alarm set for the same time. The confidence associated with a value can be categorically determined using statistics. The principle is to repeat the measurement process many times and examine the *distribution* of measured values. This statistical process is the subject of this experiment.

Statistical Theory and Rocks

Statistical analysis forms the basis of experimental science. Measurements of all types are subject to random error on some level. Recall that the signature of random error is that repeated measurements yield a range of different values. Consider a simple game in which two players wish to determine who is better at throwing rocks. Standing by a small stream, each player tries to throw as many rocks into the stream as possible. If the players stand too far away, then it would be difficult to throw a rock far enough to cross the stream, resulting in more rocks on one side than the other. This is a **systematic** issue, as all positions would be *biased* in one direction. Our players choose short enough distance that they are equally likely to overshoot or undershoot the stream. This symmetry is a hallmark of **random** error.

Figure 1: The distribution of rock positions for each player with the maximum deviations for each player shown.

Figure 1 shows the positions of the rocks thrown by each player. There is no apparent bias. According to the pictures, who is better at throwing? The answer obviously depends on what is meant by "better". Player 2 landed more rocks in the stream, but many rocks are much farther away. While player 1 only has four rocks in the stream, their throws were more consistent. The most natural place to start is with the average rock position for each player. In the most general sense, for a set of measurements $\{x_1, x_2, ..., x_N\}$, the arithmetic **mean**[1] is

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

Calculated in this way, the random fluctuations tend to cancel out. **If this notation is unfamiliar, take a detour to the Appendix to read about summation**. For each rock on one side of the stream, there is a corresponding rock on the opposite side. By the symmetry of the rock positions, the mean distance from the stream for both players would be close to zero. Evidently, the mean rock position alone is not a good indication of the winner.

It may be useful to consider who had the worst throw. The **deviation** is a measure of how far a single value is from the mean,

$$d_i = x_i - \bar{x}. \tag{2}$$

According to Figure 1, Player 2 has the largest deviation shown as the distance d_2. This, too, is insufficient for characterizing the skill of the player. Even the best rock throwers will have some outliers. The next logical step in this progression is to consider the average deviation,

$$\bar{d} = \frac{1}{N-1}\sum_{i=1}^{N} d_i \tag{3}$$

where $(N-1)$ as opposed to N is in the denominator to account for the fact that the data have already been used once to determine the mean, reducing the number of degrees of freedom by 1. This calculation

[1] The symbol $\bar{x}$ is pronounced "x-bar" where the "bar" indicates the mean of x values. The placement of a bar over a variable is the standard notation for a mean.

is also fruitless as the deviations d_i can be positive or negative (depending on the side of the stream in this example). As a result, positions on opposite sides will cancel and both players will have $\bar{d} \approx 0$.

It is this cancellation feature that motivates the definition of standard deviation[2],

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} d_i^2}. \tag{4}$$

Formulated in this way, it is clear how the standard deviation relates to the simple measure of the deviation of a given point. Squaring the deviations eliminates the possibility of a negative number in the sum. The result is a quantity that characterizes the spread of a set of values. The most general expression for the **standard deviation** is

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{5}$$

which results from combining equations 2 and 4. Now we are in a good position to say who is a better thrower. While the mean rock position for both players is about the same, player 1 has a much smaller standard deviation–their rocks were thrown with more precision.

Histograms

A histogram is a type of chart that displays the number of times an event occurred. The vertical axis counts the number of times a value (or range of values) was obtained. The horizontal axis represents the variable in question. The number of counts associated with a specific bin depends on the width of the bin. This is illustrated in Figure 2. All four histograms were generated with the same data set, but different numbers of bins were used in each instance. On the left, only five bins were used so each bin contains a large range of values. As a result, the number of counts is very large. As the number of bins increases, the counts per bin decreases and the shape of the histogram approaches the theoretical bell-shaped curve (shown in red). Given a sufficient number of measurements subject only to random fluctuations, the associated histogram will always approach a bell shaped distribution.

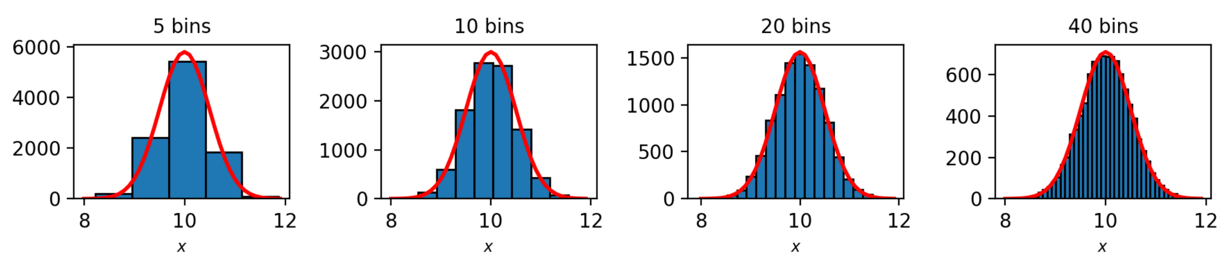

Figure 2: Histograms of 10000 samples with a mean of 10 and standard deviation of 0.5.

The Normal Distribution

The concepts outlined above are universal for any set of randomly distributed values, which are said to be *normally* distributed. The normal distribution is often referred to as a bell curve because of its symmetrical bell shape. The **normal distribution** is a type of function that describes probabilities associated with a set of values which is entirely characterized by their mean and standard deviation. This statistical framework defines the confidence level associated with a measurement uncertainty.

In the previous experiment, you measured distances using a ruler. You likely assigned the *minimum* uncertainty to your measurements because repetition would yield the same length each time. Now imagine

[2]The symbol σ is a lowercase greek "sigma."

measuring the period of a pendulum with a stopwatch. Because the measurement depends on reaction time, repeated measurements will result in slightly different values. The variation between measurements increases the uncertainty; the smallest digit on the stopwatch would not account for the fluctuations.

When successive measurements of the same quantity yield results with random variations, the mean and standard deviation are used to determine the best estimate of the measurement and its uncertainty.

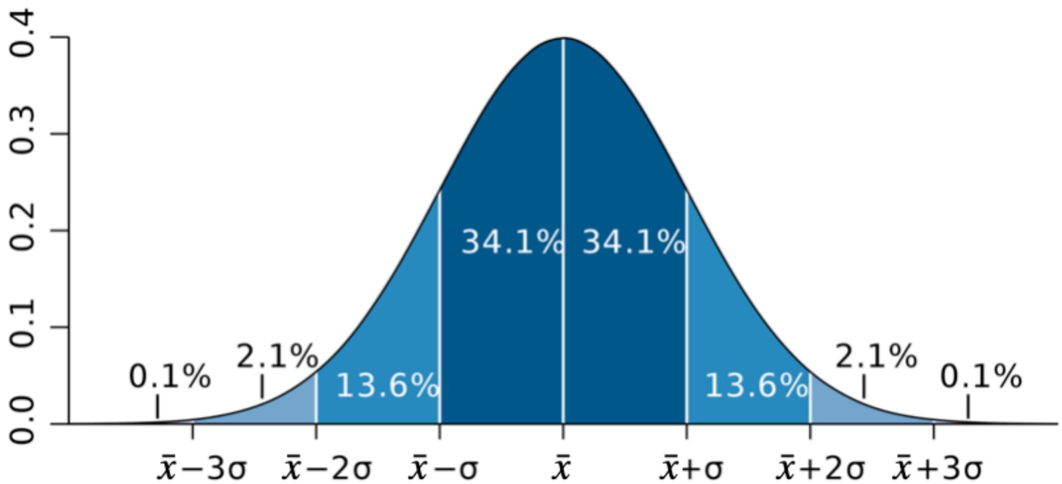

Figure 3

Figure 3 is a plot of the normal distribution with a mean value $\bar{x}$ and a standard deviation σ. Remember, σ is a just number determined from the full set of measurements of the quantity x. The percentages indicated under the curve correspond to the fraction of values which occur over the specified interval. For example, 34.1% of values are between $\bar{x}$ and $\bar{x}+\sigma$. Put another way, if a number is sampled randomly from the set, there is a 34.1% chance that it will fall between $\bar{x}$ and $\bar{x}+\sigma$.

The probability of ascertaining a number between $\bar{x}-\sigma$ and $\bar{x}+\sigma$ is 68.2% (shown in dark blue). The most probable value is at the peak of the curve, corresponding to the mean. Values far to the left or right of the mean are exceedingly less likely to be measured. Making a single measurement of x amounts to sampling a value from all possibilities along the horizontal axis.

Range	$\bar{x}\pm\sigma$	$\bar{x}\pm2\sigma$	$\bar{x}\pm3\sigma$
Measurements within range	68%	95%	99.7%
Measurements outside range	32%	5%	0.3%
	1 in 3	1 in 20	1 in 400

We can now return to the problem of assigning uncertainties. For a single measurement of x subject to random variations, the **uncertainty in a single measurement** is conventionally 1 standard deviation, $\delta x = \sigma$. This means that there is a 33% chance that an additional measurement would result in a number outside of the dark blue range in Figure 3.

Now consider the *uncertainty in the mean of a set of measurements*. This concept is best understood by looking at the histograms in Figure 4. Each histogram was generated with a true mean value of $\bar{x} = 10$. The leftmost histogram has the fewest measurements, and it is difficult to determine the mean by visual inspection (the highest count is closer to $x = 11$). The number of measurements increases in each plot to the right. In the rightmost plot, the histogram is well formed making it obvious that the mean (the peak) is located at $x = 10$. This is a clear indication that increasing the number of measurements must increase the confidence in the mean value.

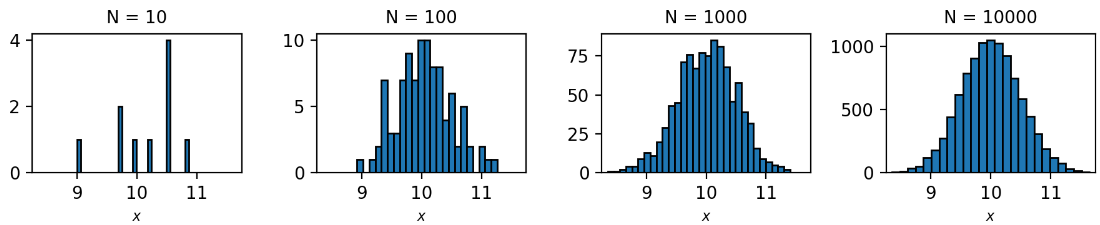

Figure 4: Histograms made by sampling a normal distribution with increasing numbers samples N

The uncertainty in the mean $\bar{x}$ of a set of measurements can be determined applying the uncertainty propagation formula to the definition of the mean in Equation 1 (this is one of the pre-lab theory questions). The **uncertainty in the mean** is called the **standard error**,

$$\delta\bar{x} = \frac{\sigma}{\sqrt{N}} \tag{6}$$

where N is the total number of measurements. Take a minute to digest this equation. For a set of measurements $\{x_1, x_2, ..., x_N\}$:

- $\bar{x}$ is the average of the set
- $\delta\bar{x} =$ "uncertainty in the mean"
- σ is the standard deviation of the set
- N is the total number of measurements

The histograms in Figure 4 show qualitatively that the uncertainty in the mean decreases as the number of measurements increases. Equation 6 quantifies this observation.

Single Measurement	$x \pm \sigma$
Mean Value	$\bar{x} \pm \frac{\sigma}{\sqrt{N}}$

Discrepancies and the Statistical t-test

Many experiments are designed to test an existing model by measuring some quantity and comparing to a known reference value. This reference value could be determined theoretically or by previous experiments. It is useful to compare the measured value with the reference by computing the **discrepancy**,

$$\text{discrepancy} = |x_{ref} - x_m|$$

where x_{ref} and x_m are the reference and measured values, respectively. If the measured value is exactly equal to the reference, the discrepancy is zero (the two numbers are in perfect agreement). The t-test is a measure of the *significance* of the discrepancy based on the confidence intervals associated with the measurement distribution. It is defined as

$$t = \frac{|x_{ref} - x_m|}{\delta\bar{x}} \tag{7}$$

where $\delta\bar{x}$ is the standard error (i.e. the uncertainty in x_m). Equation 7 can be manipulated to get $|x_{ref} - x_m| = t \cdot \delta\bar{x}$. Expressed in this way, it is clear that t counts the number of standard errors by which the two values differ.

Rule: A discrepancy is considered significant if $t > 2$.

As an example, imagine buying a spring with a manufacturer stated spring constant of $k_{ref} = 7.10$ N/kg. In the lab, you perform a series of measurements and determine $k_m = 7.21 \pm 0.03$ N/kg. Plugging these values into Equation 7 gives

$$t = \frac{|7.10 - 7.21|}{0.03} = 3.7.$$

Because $t > 2$, the discrepancy is considered significant. The probability of obtaining such a large discrepancy by random chance alone is effectively zero. Interpretation of this result depends on the context of the experiment. It can indicate a problem with a measurement or, far less often, a problem with a theory.

1 Theory Questions

1. (1 pt) Is high precision indicated by a small or a large standard deviation?

2. The goal of this problem is to prove the validity of the standard error as the uncertainty in the mean. Consider a set of N measurements of the time period of a clock $\{T_1 \pm \delta T_1,\ T_2 \pm \delta T_2, ...,\ T_N \pm \delta T_N\}$ where each measurement is subject to significant random error from human reaction time. The mean of the set of measurements is

$$\bar{T} = \frac{T_1 + T_2 + ... + T_N}{N} \tag{8}$$

and the standard deviation is σ. This is a function of N variables. The uncertainty propagation formula[3] can be applied to the mean to get an expression for the uncertainty,

$$\delta\bar{T} = \sqrt{\left(\frac{\partial\bar{T}}{\partial T_1} \cdot \delta T_1\right)^2 + \left(\frac{\partial\bar{T}}{\partial T_2} \cdot \delta T_2\right)^2 + ... + \left(\frac{\partial\bar{T}}{\partial T_N} \cdot \delta T_N\right)^2}. \tag{9}$$

where there are N total terms under the square root (one for each variable). In the following steps, your goal is to show that Equation 9 reduces to Equation 6. *Hint*: you do not need to explicitly evaluate any sums for this problem.

   a. (1 pt) From the reading, the uncertainty in a single measurement of an arbitrary quantity x is $\delta x = \sigma$. What are the individual uncertainties $\delta T_1, \delta T_2, ..., \delta T_N$?

   b. (1 pt) Determining the uncertainty involves evaluating N derivatives. What is the value of the first derivative, $\partial\bar{T}/\partial T_1$? *Hint:* take a derivative of the equation for $\bar{T}$ with respect to T_1, treating all other variables as constants (N is also a constant).

   c. (1 pt) Show that

$$\frac{\partial\bar{T}}{\partial T_1} = \frac{\partial\bar{T}}{\partial T_2} = ... = \frac{\partial\bar{T}}{\partial T_N}.$$

   Hint: if you get stuck, pretend there are only three variables ($N = 3$) and write out Equation 8 with just T_1, T_2, T_3. Then, take a derivate with respect to each variable.

   d. (1 pt) Combine your results to show that $\delta\bar{T} = \sigma/\sqrt{N}$

3. Because of time constraints, we are often limited to very few measurements in this class. The goal of this problem is to check the validity of the standard error as the uncertainty in in the mean in a case of only two measurements ($N = 2$). Imagine making a measurement of a quantity x two times and obtaining values x_1 and x_2. We would like to determine an expression for $\delta\bar{x}$ in terms of x_1 and x_2.

   a. (1 pt) Write down the mean $\bar{x}$ in terms of x_1 and x_2.

   b. (1 pt) Use Equation 5 to express the standard deviation in terms of x_1, x_2, and $\bar{x}$

   c. (1 pt) Now, substitute $\bar{x}$ out of your expression for σ by plugging your result from 3a into the standard deviation obtained in 3b. Try to simplify your result. *Hint:* $(x_1 - x_2)^2 = (x_2 - x_1)^2$

   d. (1 pt) The final step is to plug your result from 3c into the equation $\delta\bar{x}$ in terms of x_1 and x_2.

   e. (1 pt) Sketch a number line and label x_1, x_2, and $\bar{x}$. Draw an error bar of length $\delta\bar{x}$ extending in both directions from $\bar{x}$ indicating the uncertainty in the mean. Does the error bar extend beyond the two points x_1 and x_2?

[3] See Fundamentals lab guide for a refresher.

2 The Clock Experiment

At the front of the room is a large digital clock. The clock counts at a constant rate, but **do not assume the rate is one count per second**. The clock will count from 1 to 20, go blank, and then restart. At your bench is a high-precision hand timer. The timer can measure out to 0.001 seconds with an error stated by the manufacturer of 0.0005 seconds for intervals of 10 seconds or less. This intrinsic limitation of the hand timer is *systematic*; it is inherent in the measurement tool and it does not change with repetition. What does change while using the timer is your hand-eye response time. Hand-eye coordination is characteristically random. If you were able to operate the timer indefinitely without any fluctuations, the uncertainty in a single measurement would be that stated by the manufacturer. Unless you're a robot, you will find this an impossible task.

The objective of this experiment is to determine the count rate (seconds per count) of the large digital clock. By using the high-precision hand timer, you will determine if the digital clock has a significant systematic error (counting too fast or slow).

1. (4 pt) Write an introduction paragraph explaining the purpose of the experiment and the important concepts from the reading.

2. (1 pt) Observe the clock. Write down whether you believe the clock is counting reasonably close to one count per second. Before doing any measurements, guess how much your reaction time would vary from one measurement to the next (this is your initial guess for your random error).

3 Measurements

Below is an outline of how you should perform your experiment. *Warning*: the repetitive measurements make this a boring experiment. The purpose is to show the emergence of the normal distribution in your measured data. Future experiments will be more interesting! A general outline of your process is as follows:

- Make tables in Google Sheets or Excel to organize your data. Label the column header for time intervals as T with the appropriate unit.

- One person should time while the other records the data. To avoid an unconscious skewing of data, the person timing should not look at the data until all measurements have been recorded. This is essential; otherwise, you will introduce a bias into your measuring procedure! Make a few practice runs before taking data.

- Choose an interval at least 10 counts long (e.g. from 5 to 15). Avoid starting a timing interval on the first count. Record the initial and final clock count you use to define the interval. Measure the interval 50 times using the hand timer. Use the first three or more counts to develop a tempo for synchronizing your start.

4 Data Analysis

Once your data are organized into tables, create a histogram of your measurements of the time interval, T (all spreadsheet programs have this capability). If your histogram does not look like a normal distribution, take some additional measurements. Your analysis will be better with a well defined distribution.

You should use Excel to perform any statistical calculations. **Ask your TA if you need help doing spreadsheet calculations or making a histogram**. There are two math operations in Excel that will be useful:

- The mean value: "=AVERAGE(...)"

- The sampling standard deviation is "=STDEV.S(...)"

where "..." should be replaced by the cells containing the measurements.

1. (5 pt) Include a labeled histogram showing the distribution of your full set of measurements. Choose a appropriate bin width that shows the character of your distribution.

2. (3 pt) Compute the mean $\bar{T}$, standard deviation σ, and the uncertainty in the mean $\delta\bar{T}$. Be sure to limit the number of decimal places to match the precision of your measurement instrument. **Important**: both σ and $\delta\bar{T}$ have units that must be included.

3. (2 pt) What is your best estimate at the time period of 10 counts? Include the uncertainty in this value.

In the next three questions, you will examine the behavior of the mean, standard deviation, and the uncertainty in the mean as more measurements are made. It may help to consider the behavior as $N \to \infty$.

4. (2 pt) What do you expect to happen to $\bar{T}$ as N increases? Sketch a graph of $\bar{T}$ vs. N.

5. (2 pt) What do you expect to happen to σ as N increases? Sketch a graph of σ vs. N.

6. (2 pt) What do you expect to happen to $\delta\bar{T}$ as N increases? Sketch a graph of $\delta\bar{T}$ vs. N.

7. (4 pt) In the table below, N corresponds to the total number of measurements. The first row corresponds to $\bar{T}, \sigma$, and $\delta\bar{T}$ for the first three measurements. The second row corresponds to $\bar{T}, \sigma$, and $\delta\bar{T}$ for the first 5 measurements, and so on. Construct this table and compute the quantities for each value of N.

N	$\bar{T}$ (s)	σ (s)	$\delta\bar{T}$ (s)
3			
5			
10			
20			
30			
40			
50			

8. Using Sheets/Excel, make a connected scatter plot of the following relationships. **Important**: each plot must have a caption with a brief interpretation of the result.

   a. (3 pt) $\bar{T}$ vs. N

   b. (3 pt) σ vs. N

   c. (3 pt) $\delta\bar{T}$ vs. N

9. (1 pt) Do your plots accord with your predictions from questions 4, 5, and 6?

5 Drawing Conclusions

Up to this point, you have examined the time period for 10 counts. Next, you will determine the period of a single count. Define the count rate (seconds/count) as R. Then the mean with its uncertainty may be denoted as $\bar{R} \pm \delta R$

1. (2 pt) From your measurements, determine your best estimate at the count rate and its uncertainty. Explain your work.

2. (2 pt) Compare your measured count rate to the reference value $R_{ref} = 1.0$ second/count using the statistical t-test:

$$t = \frac{|R_{ref} - \bar{R}|}{\delta\bar{R}}$$

3. (1 pt) Do your results indicate that the clock has a systematic error?

Appendix: Evaluating Sums

This may be your first exposure to summation notation. This section covers the key aspects relevant to this class. Consider a sequence of 6 numbers, $\{1, 3, 3, 2, 2, 1\}$. The mean (average) value of the sequence is

$$\frac{1+3+3+2+2+1}{6} = 2.$$

Let's take some steps to generalize this calculation. The sequence can be expressed in a more general form as $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ with a mean value $\bar{x} = 2$. Using this notation, the previous equation can be written as

$$\frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6} = \bar{x}.$$

Now, consider a sequence that is N numbers long, $\{x_1, x_2, x_3, ..., x_N\}$. The mean value is found in the same way, but now there may be additional numbers:

$$\frac{x_1 + x_2 + x_3 + ... + x_N}{N} = \frac{1}{N}(x_1 + x_2 + x_3 + ... + x_N) = \bar{x} \tag{10}$$

Summation notation is a compact way of expressing the sum of an arbitrary sequence of numbers. Rather than writing out each term in the sum, Equation 10 can be written as

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

where

- $\sum$ is an uppercase Greek sigma. This symbol indicates the summation operation
- N is the total number of elements in the sequence
- i is called the *index* and is a natural number used to identify each element in the sequence. For example, the $i = 2$ element of the list $[3, 5, 7]$ is 5. The $i = 4$ element of the list $[x_!, x_2, x_3, x_4, x_5]$ is x_4.

At this point, the combination of these symbols can be interpreted. Consider a simple example of a sum from $i = 1$ to $N = 4$

$$\sum_{i=1}^{4} i = 1 + 2 + 3 + 4 = 10.$$

For a more abstract example, consider a sum of elements in the list $[x_1, x_2, x_3, x_4]$,

$$\sum_{i=1}^{4} = x_1 + x_2 + x_3 + x_4$$

Let's look at one example of a more complicated sum. The Shannon diversity index is a way of quantifying the biodiversity in an ecosystem[4]. The equation is a sum,

$$H = -\sum_{i=1}^{N} p_i \ln p_i, \tag{11}$$

where p_i is the fraction of a species within a given population. Consider an ecosystem that contains only three species with fractions $p_1 = 1/10$, $p_2 = 2/5$, and $p_3 = 1/2$.

$$H = -\sum_{i=1}^{3} p_i \ln p_i = -\left(p_1 \ln p_1 + p_2 \ln p_2 + p_3 \ln p_3\right) \tag{12}$$

To test your understanding, verify that $H = 0.943$.

[4] This number is actually a form of entropy, though not in the context of thermodynamics. A larger value of H is an indication of a diverse, healthy ecosystem.