Experiment: Dartboard Statistics

Goals

- Observe how natural, random processes translate into statistical data.
- Apply statistical analysis to a set of data.
- Practice using a scripting/programming language to perform data analysis.
- Practice carefully recording the setup, procedures, results, analyses, and conclusions in a lab book for a simple experiment.

Personal Protective Equipment & Safety

In addition to the standard safety rules for Second-year Physics Lab Courses, this lab involves darts. The Physics Undergraduate Labs (UGL) will supply personal protective equipment. The following safety rules will be in effect for this lab:

- No food or drink will be allowed in the labs.
- Wear long pants, a dress, or the equivalent.
- Wear close-toed shoes.

Required, Suggested, & Optional Equipment for Students

- Lab Book #1 & Pens (Required)
- Safety Glass (Required); Either UGL-Supplied or Student-Owned
- Ruler (Suggested)
- USB key (Suggested)
- Laptop (Suggested)

Background

Random processes

See: Ch 2, Hughes & Hase (H&H)

Throughout the physical world, some degree of randomness is inherent to most processes. Whether it is due to thermal variations, quantum fluctuations, or human imprecision, the outcome of a measured value is almost always affected by these random processes. As a result, there is a distribution of results in any measurement, and we can quantify the degree of randomness by evaluating the spread of these measurement results.

For processes that are truly random, that are independent from each other, and whose fluctuations add simply, a binomial distribution describes the distribution of possible outcomes. If you imagine beginning with a "true" value, and then subject that to a random influence that either (with equal probability) adds or subtracts a small error value, the new value is either bigger or smaller by that amount. If this process is then repeated, one can add the possible pathways to each outcome (see Fig. 1) and determine with what probability a particular measurement will be made. This leads to the binomial distribution. (For an excellent demonstration of this, look up a Galton board: http://galtonboard.com/)

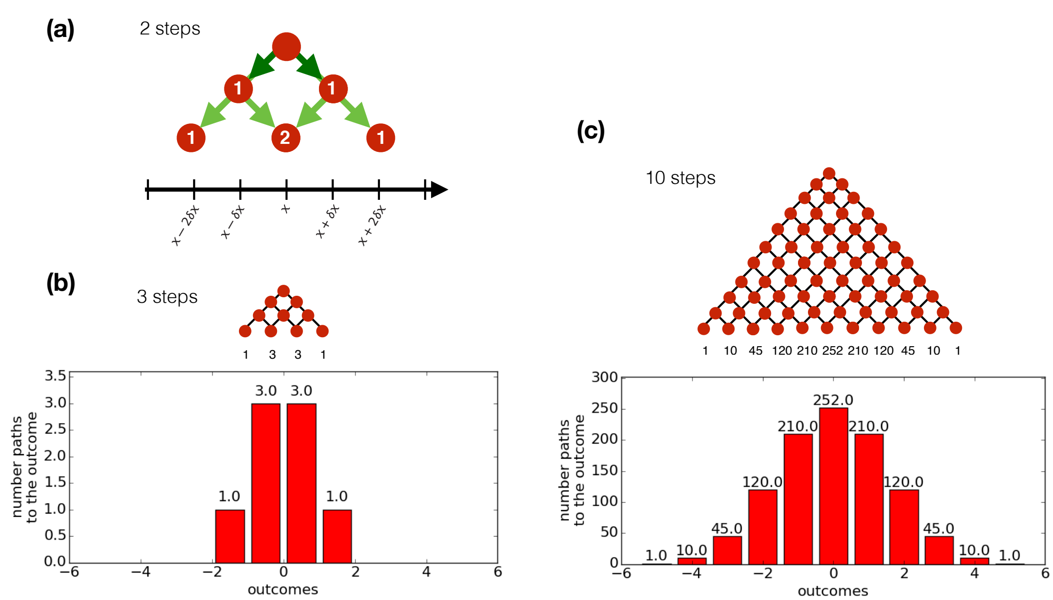

Figure 1: Binomial distribution. (a) Model for statistical distribution.

In the limit of many small random fluctuations that add to each other, the binomial distribution is well approximated as a Gaussian,

$$G(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \tag{1}$$

where x represents the possible measurements outcome, μ is the (theoretical) true value the measurement is trying to determine, and σ is the parameter that characterizes the width of distribution.

Statistical analysis

See: Ch 2, Hughes & Hase (H&H)

In order to analyze any set of data, some assumptions must be made regarding the random processes involved. In this experiment, if we assume that each throw is independent from the other and is a random process (without bias), we expect a "normal" or "Gaussian" distribution. From the data collected, we can extract several pieces of information, and we are interested in the centre of our measurements, the width (or spread) of our measurements, and the uncertainty in determining the centre.

To find the expected centre from a repeated set of N measurements of some quantity x, we can take the mean of the outcomes $x_1, x_2, x_3, \ldots$,

$$\bar{x} = \frac{1}{N}\left(x_1 + x_2 + x_3 \cdots\right) = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{2}$$

which is our best guess as to the value of μ. (H&H §2.2)

We can also compute a measure of the width, or spread, of the measurements using a quantity called the "standard deviation" of the sample. Using all of the data, we can compute

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

where σ_{N-1} (also called σ_{sample}) is our best guess as to the value of σ. This value gives us a region of confidence: we believe that, with $\sim$2/3 probability, the next measurement we take will be found between $(\bar{x} - \sigma_{N-1})$ and $(\bar{x} + \sigma_{N-1})$. (H&H §2.3)

For random processes, in the limit of an infinite number of unbiased measurements, normal distribution defined above (Eq. 1) should arise. In this case, we can see why we defined it in terms of σ: the standard deviation of this distribution (also known as the standard deviation of the population) is σ. The probability of finding results between $\bar{x} + \sigma$ and $\bar{x} - \sigma$ is

$$P(x; \mu - \sigma, \mu + \sigma) = \int_{\mu-\sigma}^{\mu+\sigma} G(x; \mu, \sigma)\ dx = \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu-\sigma}^{\mu+\sigma} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] dx = 0.683 \tag{4}$$

So, our interpretation of the standard deviation is justified: we expect results to fall within one standard deviation about 68% of the time. (H&H §2.5)

The last piece of information we can get about a set of measurements of x is the uncertainty in the mean value – letting us ask the question: how well do we know the expected value of x? The more measurements we make, the better we expect to know the mean (H&H §2.7, 4.5). To quantify this, we may use the "standard error" (or "standard deviation of the mean"):

$$\alpha = \frac{\sigma_{N-1}}{\sqrt{N}}. \tag{5}$$

By plotting a histogram of the set of data against a normal distribution curve, a simple check of the goodness-of-fit can be made to this model of unbiased, random measurements.

Methods

Apparatus

In this quick experiment, the apparatus consists of a "dartboard" that, unlike ones with which you may be familiar, is sectioned into 41 vertical, numbered bins (from -20 to $+20$) that are separated along the horizontal direction (see Fig. 2). This configuration tests your ability to throw darts accurately along just one dimension. There will be five darts with each dartboard setup.

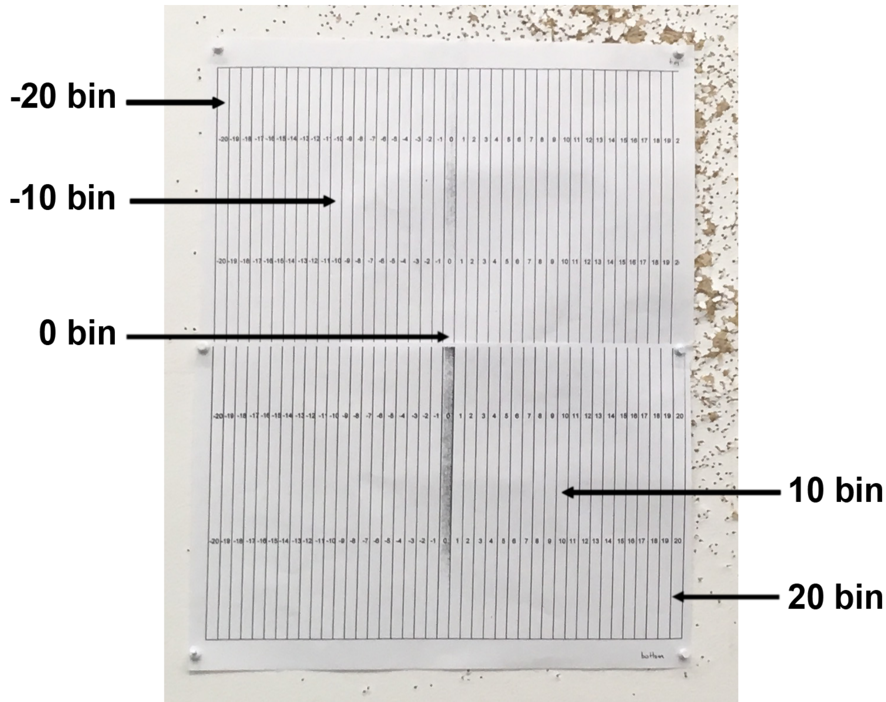

Figure 2: Picture of the dartboard setup.

Procedures

Record all data in your lab book, such as the set-up (labelled sketches, i.e., schematics will demonstrate your methodology), well-labelled tables of your data, and any "diary-like" observations as you proceed through the experiment. These notes, as well as any tables of numbers containing your data, should be neat and well labelled, allowing future readers to understand what was done during the experiment (including yourself for your later analysis).

1. Find a lab partner (or two if directed to do so in the lab) to work with, collect a set of darts, and choose one of the dartboard setups.
2. The board will already be placed against an open section of wall with at least 2 m of space before it. Choose a distance, about 2 m away from the dartboard to use as the line from which you will throw your darts. Make sure to record the uncertainty associated with this measurement. Mark the distance with a piece of tape.
3. Standing at the measured distance, each partner will throw darts at the board while a lab partner records which bin the darts land in. It may be more efficient to throw the full set of five at a time,

then record the five results in a single row of your lab book. You can decide how to alternate between partners throwing the darts and recording the results, but eventually, each person should perform a minimum of 160 throws, and record their partner's data for a similar number of throws. Treat the first 10 throws of each partner as warmup; the remaining 150 throws should be recorded for each person individually in the lab book of the person throwing the darts.

4. If you miss the dartboard with a dart or the dart falls out, rethrow that dart.

5. After finishing 150 valid throws for each partner and recording the data, along with your observations, have an instructor sign your lab notebook to verify your data.

Data processing and analysis

The objectives for the analyses in this lab activity are to have you understand how to process data using excel or a scripting language like Python, how to extract data from a file, and how to do statistical analysis on these data. While you will record your data immediately and do some initial analysis, you will do a more full analysis later in the semester. As such, you will submit two parts to this lab activity:

Part I) Raw Data & Initial Analysis: You will create a "comma-separated values" (csv) file of your data, submit this on e-class, do some simple analysis, and turn in your lab book.
(This part will be graded under the "Lab Book" rubric for 3% of your total course grade.)

Part II) Class-wide Data & Final Analysis: You will analyse the statistics of both your dart throws and those of the entire class together, and comment on the statistical outcomes of the process.
(This part will be graded under the "Lab Book" rubric for 3% of your total course grade.)

<u>Part I: Raw Data</u>

You will submit your 150 data points as a comma-separated values file. Organize the data in two columns, "Valid Dart Thrown" and "Dart Strike," where there will be one row for each valid dart thrown (in addition to a row containing the text that labels the data). Use the following file naming convention:
"Lastname_firstname_dartboard.csv" (eg, LeBlanc_Lindsay_dartboard.csv).

- ◇ Enter your data into a computer in a csv format. There are various ways to do this:
  - Using a spreadsheet program (Excel, Numbers, Google Sheets) and saving the data as .csv.
  - Using a text editor, and saving the data as .csv.
  - Any other way you prefer.

  The first few lines of the CSV that you create will look like:

  ```
  Valid Dart Thrown, Dart Strike
  1, 2
  2, -6
  3, 9
  ...
  150, 3
  ```

- ◇ Upload your (properly named) .csv file to e-class in the specified format. This will allow us to automatically compile a complete class list that we will distribute for Part II of this lab.

Part I: Initial Analysis

- Manually make a histogram of your dart strikes in your lab book. Organize your bins so that there are a times you achieved a score in each bin. This will be useful for checking your programming work.
- In a Python (or Matlab) script, import your dartboard data from a .csv file. You should end up with an array-like variable that contains 150 entries.
- Calculate the sample mean, standard deviation, and standard uncertainty of the mean of the data using built-in functions in Python (or Matlab). Record these in your lab book and clearly label/box these values.
- Create a set of bins that span your data and use the histogram functions in Python (or Matlab) to bin the data. You may choose the width of the bins as you see fit.
- Using Python of Matlab make a plot of the histogram. Print out the plot and paste it in to your lab book; make sure to include a brief plot title and caption.
- Download the Python (.py) or Matlab (.m) code that you used to calculate statistics, make the histogram, and plots. Paste a printout of this code in your lab book.
- Turn in your lab book.

Part I will be evaluated under the standard "Lab Book" rubric, with the minor change that submission of a properly-formatted and named .csv file online will be evaluated in the "Data collection, interim tables, graphs, schematics, consideration of uncertainties" section.

Important Notes: Read carefully the Lab Book Guidelines to include In-Lab and Post-Lab Notes. Make sure you add your observations about the experiment and discussion/conclusion requested for each tasks. In addition, don't forget to include setup detail including diagram (with appropriate and clear labels/captions), uncertainties including distance uncertainty and justifications. The Post-Lab Notes is mandatory, make sure to include print-outs of your plot, any data analysis, and your final results in the lab book.

Part II: Class-wide Data & Final Analysis

Here you will be analysing both your personal data and the class data as a whole. The latter will be distributed to you after Part I is completed. The major goals here are to calculate characterising statistics, to compare the distribution to a normal distribution, and to conclude if there are any systematic errors (bias) in the dart throws. Some key tasks that you need to complete for the submission of your lab book include:

- Calculate the mean, standard deviation, and standard uncertainty of the mean of both sets of the dart data using standard built-in or imported functions/methods in Python (or Matlab). Record these in your lab book and clearly label/box these values.
- Calculate the mean, standard deviation, and standard uncertainty of the mean of both sets of the dart data by implementing the formulas for each (above) in a Python (or Matlab) script, where the only programming functions/methods you may use are "for loops", standard mathematical operations (addition, subtraction, multiplication, division), and functions/methods that return the length of an array or list. Record these in your lab book and clearly label/box these values. Briefly compare the results of your hand-built functions to the built-in functions.

- Create a set of bins that span each set of dart data and use the histogram functions in Python to bin the data. You may choose the width of the bins as you see fit; however, we recommend that you use one constant bin width for each data set.
- Evaluate whether the collected data fit the normal distribution by displaying a Gaussian curve using Eq. 1 on the same graph as the histogram. Please note that the normalization condition for the Gaussian may depend on your choice of bin size. Print out the plots (one for your data and one for the class-wide data) and paste it in to your lab book; make sure to include a brief plot title and caption. In your lab book, briefly indicate your conclusion about if the data fit the normal distribution.
- Evaluate the fits of both sets of the dart data to the normal distribution using the χ^2-test. Record the results of these tests in your lab book and clearly label/box these values.
- In your lab book, briefly discuss your results and reflect on the differences between your data and the full class set of data.
- In your lab book, conclude whether the data indicates a systematic error (bias) in the dart throws using the statistical parameters describing the data.
- Paste a copy of the Python or Matlab code that you used to perform Part II analyses in your lab book. You do not need to include the raw data tables in your lab book (we will already have these).
- Turn in your lab book.

Part II will be evaluated under an adjusted "Lab Book" rubric. There will be no evaluation for "Preparedness and research planning". We will award 15/30 points for "Data collection, interim tables, graphs, schematics, consideration of uncertainties," removing the need for a schematic, but adding the requirement that important equations are indicated in the lab book. We will award 15/30 points for "Completeness, neatness, logical flow of thought, recording of interim observations". The key tasks identified above will be split among the two sections, with aspects more related to textual summary in the latter.

Important Notes: Read carefully the Lab Book Guidelines to include In-Lab and Post-Lab Notes. Make sure you add your observations about the experiment and discussion/conclusion requested for each tasks. The Post-Lab Notes is mandatory, make sure to include print-outs of your plot, any data analysis, and your final results in the lab book.