Estimating PI Using Probability

Evan Richard

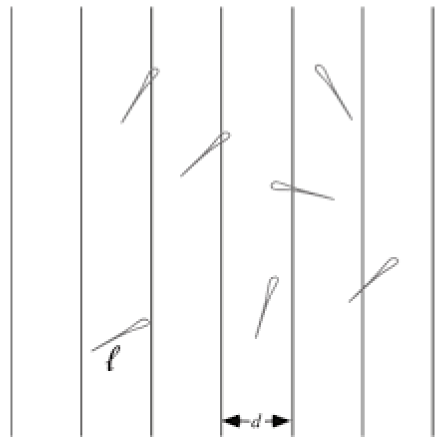
COMPSCI1200

Michael Gosner

April 25, 2024

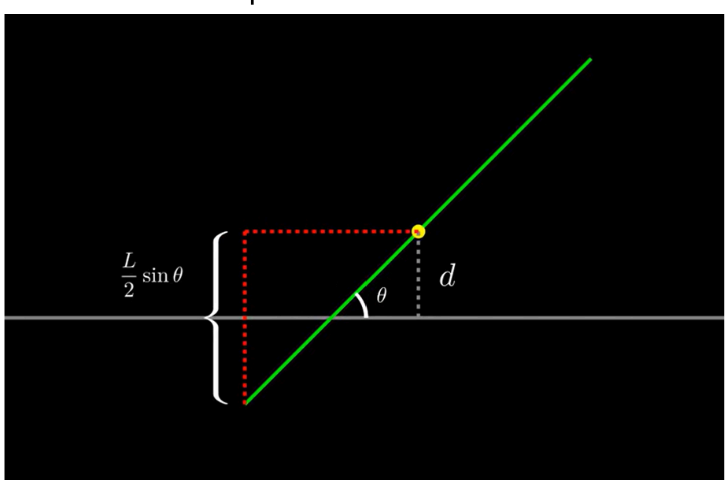
Project Description

In this experiment, I will be using the concepts of Buffon's needle problem in order to estimate the value of pi using the Monte Carlo method. Buffon's needle problem involves the probability that a thrown needle on top of parallel boards will intersect one of the parallel boards. In this case our sample space includes two possible events or outcomes, either the needle is intersecting a line or it is not. By using these simple events, formulas can be derived to estimate pi. I will also be using varying distances of boards, varying lengths of needles, and a varied amount of needles to be dropped to see what gets closest to determining the actual value of pi.



Needle and Board Diagram

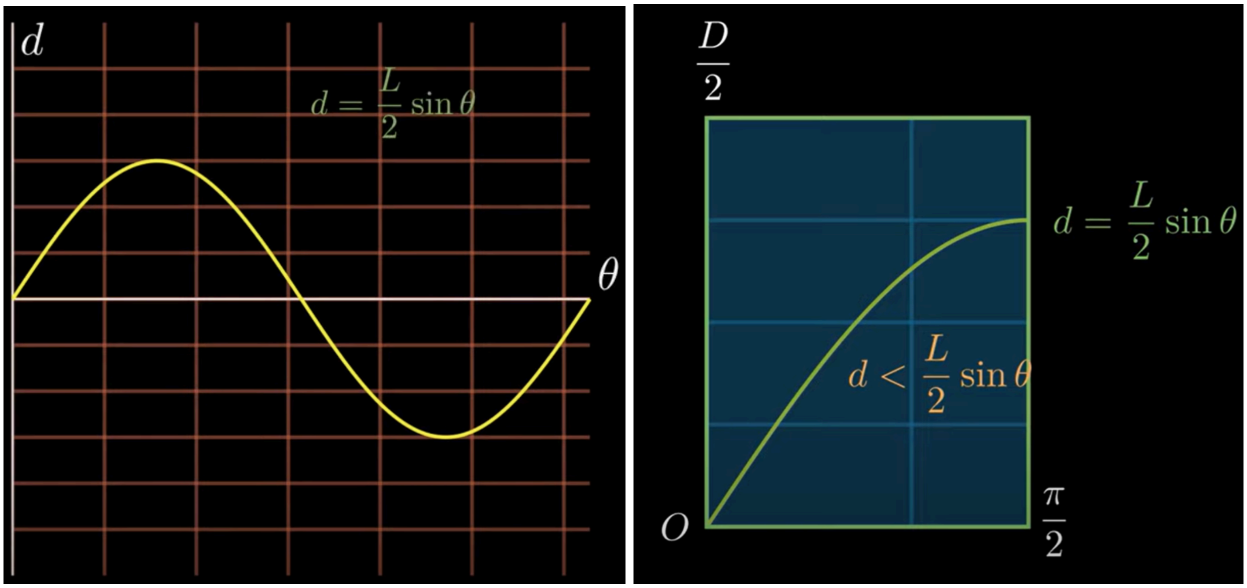
So let's first assume that the needle thrown is of length *l* and that the distance between the parallel boards is of distance *D*. Where both *l* and *D* are of some value greater than 0 and *l* is of some value less than or equal to *D*. So first we would want to figure out what the probability is that a single needle will intersect the parallel boards when thrown.



Needle Intersection ("How to get a Pi Out of needles? | Simulation | Buffon's Needle," 00:01:12 - 00:01:17)

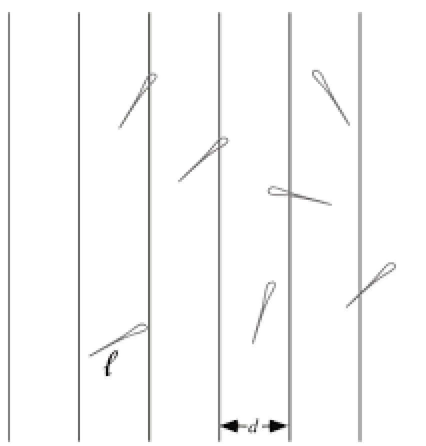
Theta represents the angle at which the needle intersects the parallel line. The yellow dot represents the center of the needle. *d* represents the shortest distance between the yellow dot and the line. The lower half of the needle, where *L* is the entire length of the needle, (below the yellow line) can be represented as $\frac{L}{2}$. So now by using trigonometry, we can find the red vertical distance to be $\frac{L}{2}sin\theta$. With this, we can show that *d* will always be smaller than $\frac{L}{2}sin\theta$

or $d < \frac{L}{2} sin\theta$. Now we know from these equations that θ is constrained from $0 \leq \theta \leq \frac{\pi}{2}$ and that $0 \leq d \leq \frac{D}{2}$.



Needle Probability Curve (“How to get a Pi Out of needles? | Simulation | Buffon’s Needle,” 00:02:18 - 00:03:28)

Above is the graph of the sin curve $d = \frac{L}{2} sin\theta$, where the x-axis is theta and the y-axis is *d* (the distance from the center of the needle to the curve). We can portion this graph (shown on the right) to fit our previously discussed constraints $0 \leq \theta \leq \frac{\pi}{2}$ and $0 \leq d \leq \frac{D}{2}$. Then we can show that the probability that the needle will intersect the parallel lines is equal to the area under the curve divided by the area of the entire rectangle. Doing this we get our equation for probability $P = \frac{2L}{\pi D}$.



Needle and Board Diagram

Finally once again looking at the needle, we can see that approximately the probability of the needle intersecting a line is $P = \frac{x}{n}$ where *x* is the number of needles intersecting and *n* is the total number of needles thrown. We equate our two probabilities $\frac{x}{n} = \frac{2L}{\pi D}$ and solve for pi $\pi = \frac{2nL}{xD}$. Which in theory should be the equation we can use to estimate pi!

My hypothesis for this experiment is that we will see better results when the needle is not only closer to the size of the boards, but when the amount of needles dropped is larger. My reasoning for the second one is that basic logic tells me that as you increase the number of iterations on a probabilistic simulation you will in turn see more accurate results. For the first one, however, I believe this because if our needles are way smaller than the distance between the boards then the probability of the needle intersecting those boards is incredibly low. Which will lead to what I coin right now as “event bias”. The event of a needle not intersecting the board will be much greater than that of the needle intersecting the boards. Which will in turn lead to less accurate results.

Experimental Data

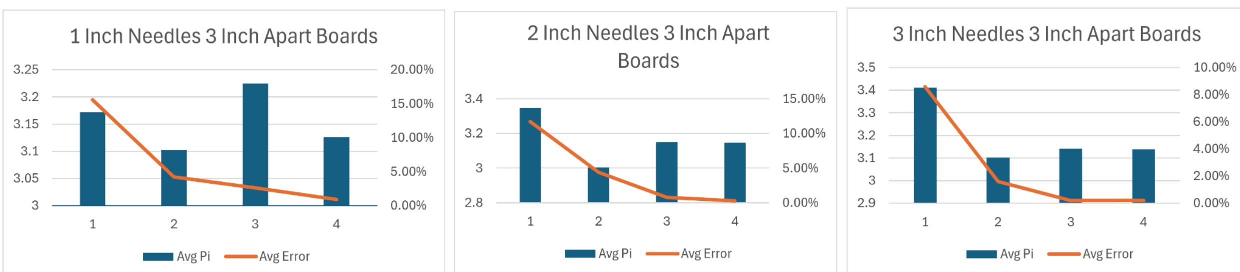


Figure 1

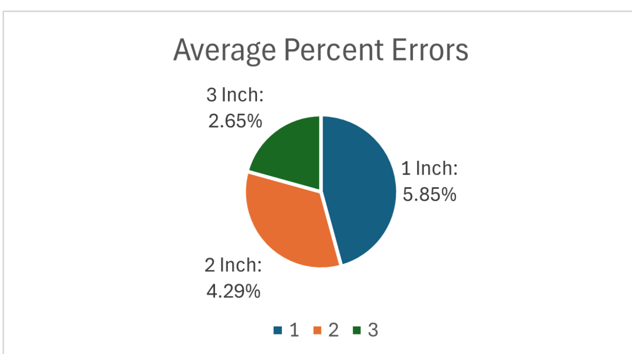


Figure 2

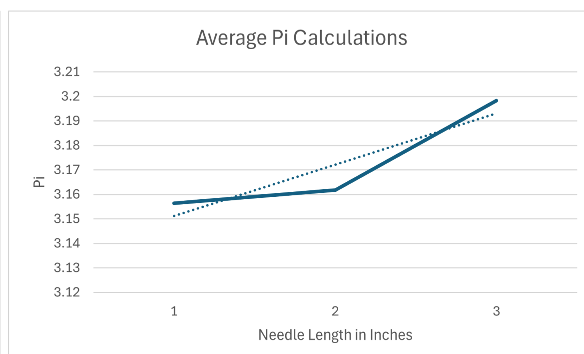


Figure 3

Data Analysis

First of all, all of these charts along with collected data can be seen on the Excel sheet in the zip file. All data was collected by running the mock simulation three times for each set of data. For example, the first run was 100 iterations of 1-inch needles thrown into boards 3 inches apart done 3 times to obtain an average. This was to try and decrease the effect that probability had on the results.

Starting with Figure 1 we can see many things. First of all for all of them as the number of iterations increases the average error decreases dramatically. Not only does it do this for every individual chart, but, it does this when comparing all three charts to each other. This means we might be able to conclude that as iterations increase, the accuracy the program had at estimating pi also increases. Also as the length of the needle approaches the length that the boards are apart the average error decreases. This would prove both my hypotheses correct that results would be more accurate with more iterations and with needles closer to the length between the boards.

Figure 2 also shows an interesting trend in the data. When taking the average percent errors between all three of the different needle lengths (1, 2, and 3 inches) the data shows that the percent error goes down. So this coincides with Figure 1 in further proving that as the needle approaches the length of the boards, the program is more accurate at estimating pi.

Finally, figure 3 shows an interesting trend that I can't quite explain. Even though the data becomes more accurate at estimating Pi, the average of the Pi that is estimated increases. Now this is very convoluted and the graph doesn't really do it justice but here is kind of my explanation. In my data, it seems that at the very beginning when testing sample sizes of 100, the average estimate for Pi is very high. However, what's more interesting is that this very high value for Pi only increases as the length of the needle approaches the board distance. At the same time as the needle becomes longer, the estimated Pi will approach the actual Pi at a much faster rate when increasing the sample sizes. So really the reason the averages are going up is because of that first 100 sample size test.

These final results were confusing but my hypothesis for this is because of what I talked about earlier with "event bias". Because we are dealing with a probabilistic program a smaller sample size will just tend to give more skewed results with smaller sample sizes. The only way I would know to fix this is to first of all do more test with larger sample sizes and see if the trend continues. If it doesn't it might just prove "event bias".

Citations Page

"How to get a Pi Out of needles? | Simulation | Buffon's Needle." *Youtube*, uploaded by piJack Mathman, 11 September 2021, https://www.youtube.com/watch?v=OlSOhe7a00g&ab_channel=piJackMathman