To what extent does the probabilistic solution of the Buffon problem and it's variations give insight into approximating π?

Candidate Code: jjm399

Pagecount: 18

1 Introduction

My interest in methods of approximation arose while reading an article by Gregory Galperin, *”Playing Pool with π”*[1], where he found a way of approximating π by looking at the number of bounces made by billiards bouncing off each other and the sides of a pool table. After reading, I was curious to learn how π even popped up in these calculations, and wondered by which other methods π could be approximated. While doing further research, I stumbled upon the ideas of incorporating probability in numerical approximation. More specifically, I read of the work of the 18th century mathematician and naturalist, Georges-Louis Leclerc, Comte de Buffon, which Galperin had actually mentioned within the prologue of his work as well! I read how specifically he investigated the probability of a needle with some length l, landing upon on a dividing line separating two parallel planks separated by an equal spacing d, which is seen visually as such, where the black bars are needles, and the lengths l and d are shown:

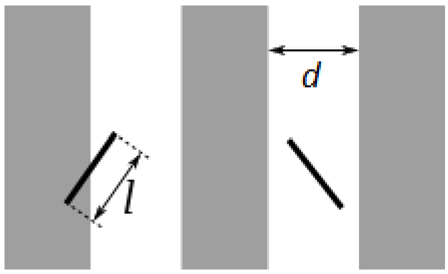


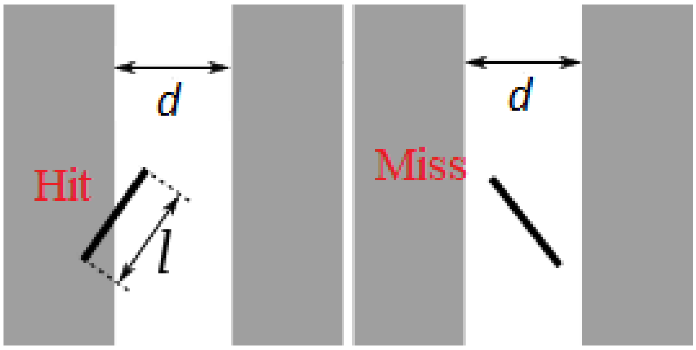
Figure 1: Diagram of Buffon’s Variation

Buffon himself actually calculated this probability analytically, making the first initial connection, although not examined until quite some time after his writings, between probability and numerical approximation. From Buffon’s work thus, variations of the problem were developed, one of the most similar ones being that of Laplace’s variation. Given this Buffon’s problem and it’s connection to numerical approximation, I wondered on the practicality and viability of perhaps using probability to actually measure a value of π, despite Galperin himself saying *”He/she can easily observe that it takes a lot of drops to get a more or less good precision for π. - no one can guarantee any specific precision in calculation of with the use of this method.*” [1], I wanted to see if this was actually true, and to what extent it was true. Being quite intrigued on the topic of statistics and probability from my Mathematics Analysis and Approaches class, I wondered whether by using this acquired knowledge and perhaps my own research, I could make some insights into probabilistic methods of approximating π, by more specifically examining Buffon’s method and mentioning it’s variations. So to focus my research, I narrowed down my research question to be: To what extent does the probabilistic solution of the Buffon problem and it’s variations give insight into approximating π?

2 Buffon’s Problem

In order to determine anything about the viability and practicality of using probabilistic equations to approximate π, examining the derivation and origin of such an equation could perhaps yield insights into this direction.

Thus consider the experiment as follows. We may drop a needle from some height unto the surface in Figure 1. We can observe the fall of the needle, and thus also record whether the needle lands on one of the intersections or not.



(a) Diagram of a Hit (b) Diagram of a Miss

Figure 2: Diagrams of a Hit and a Miss

Since we are interested in observing the probability of the intersection of the needle, we can say that the needles which intersect are denoted as successes and the ones which don't to be failures, also noting that the drop of one needle doesn't effect the drop of another, hence allowing us to denote them as independent. To quantify these results, let H be the number of hits, and N be the total number of needles, and so $N - H$ is the number of failures. Since we are again interested in the probability of intersection, we are namely interested in H, we can call our 'random variable'.

Definition 2.1. A random variable quantifies the concepts of success and failure, such that if we have a random variable X, with some parameter m which can be measured as a success or a failure, we can say that

$$X = \begin{cases} 1 & m = \text{success} \\ 0 & m = \text{failure} \end{cases} \tag{1}$$

Organizing our experiment, we see that we have a fixed number of trials, N, two measures classifying a success and a failure, and a constant probability of success. Thus, we see that we can classify our experiment by a binomial distribution! Using this fact, we can denote H to be a binomial random variable with parameters N and p such that $H \sim B(N, p)$

Thus let us explore the solution of finding this parameter p within the given problem of the dropping of needles. This would allow us to perform manipulations to begin exploring the viability of the method to approximate π. The proof which is given actually to show a solution to the problem avoids the route many solutions take, which embed in them the issue of using the value of π in the solution and specifically in the distribution of variables within the problem. And since we are curious to find the viability of this probabilistic approach to approximating π, we cannot already know the value of π, which, had we considered a different approach would force us into a paradox. Hence, to formally formulate this problem, consider the following:

Theorem 1. *Given that a needle of length l is dropped unto an equally spaced rectangular tiled floor, with the distance between the tiles being some distance d such that $l < d$, then the probability this needle crosses one of dividing lines of the tiles is*

$$p = \frac{2}{\pi}\frac{l}{d}.$$

Proof. The proof of such a value can follow from a generalization made on the expectation of a randomly dropped needle, the definition of which follows as such

Definition 2.2. The expectation of a random variable X is defined to be [2]

$$E(X) = \sum_x xP(X = x)$$

Hence, if a needle is dropped, regardless of it's length with respect to that of the distance d of the rectangular tiled floor, the expectation of the number of total crossings will be of the form

$$E = p_1 + 2p_2 + 3p_3 + \ldots + np_n$$

where p_n denotes the probability of the needle of some length will cross an exactly n number of dividing lines. What can be noted here is that the probability of which Buffon's variation on the problem asks for is atleast a single crossing, which is the sum

$$p = p_1 + p_2 + \ldots + p_n$$

There is omission of cases where the needle randomly falls exactly upon a dividing line, or with one of it's ends on the dividing line, as these events have probability of zero. In addition, if the needle length is 'short', or in other words, it's length is less than that of the distance d between the dividing lines, the probabilities p_n for $n > 1$ all go to 0, $p_2 = p_3 = \ldots = 0$, hence producing an expectation $E = p$, reducing the problem to simply calculating this expectation.

To proceed further with the proof, it will be necessary to show what is called the linearity of the expectation, in that specifically, $E(X + Y) = E(X) + E(Y)$.

Theorem 2. *Given random variables X and Y, along with some constant $k \in \mathbb{R}$, we have that $E(kX+kY)$=$kE(X)+kE(Y)$.*

Proof. The proof for the above can be constructed as such,

Lemma 3. For some constant k and a given random variable X, it is true that $E(kX) = kE(X)$.

Proof. It can be assumed that $k \neq 0$ as the statement becomes trivial for $k = 0$. From this, it can be said that by definition of expectation,

$$\begin{aligned} E(kX) &= \sum_x P(kX = x) \\ &= k \sum_x \left(\frac{x}{k}\right) P(X = \frac{x}{k}) \end{aligned}$$

letting $\omega = \frac{x}{k}$, and as a result summing over ω, it is obtained that $E(kX) = k\sum_\omega \omega P(X = \omega) = kE(X)$ ■

Further, we can show linearity of this expectation function by considering E(X+Y) for some random variables X

and Y, such that

$$\begin{aligned}
E(X+Y) &= \sum_x \sum_y (x+y) P(X=x \cap Y=y) \\
&= \sum_x \sum_y x P(X=x \cap Y=y) + \sum_x \sum_y y P(X=x \cap Y=y) \\
&= \sum_x x \sum_y P(X=x \cap Y=y) + \sum_y y \sum_x P(X=x \cap Y=y)
\end{aligned}$$

and from this recognizing that $\sum_y P(X=x \cap Y=y) = P(X=x)$ and similarly $\sum_x P(X=x \cap Y=y) = P(Y=y)$, we get

$$\begin{aligned}
E(X+Y) &= \sum_x x \sum_y P(X=x \cap Y=y) + \sum_y y \sum_x P(X=x \cap Y=y) \\
&= \sum_x P(X=x) + \sum_y y P(Y=y) \\
&= E(X) + E(Y)
\end{aligned}$$

by definition. And hence, by Lemma 1, and the above, we have that the expectation is linear. ∎

Now the expectation as a function of the length l of the needle can be considered as $E(l)$, such that, to make use of the linearity of the expectation, l can be considered as a sum of two parts of the needle, $l = x + y$ as can be seen in Figure 1. Namely, consider now $E(x)$, which is strictly monotone for $x \geq 0$, hence producing some $E(x) = kx$, where $k = E(1)$ [3]. To investigate k more appropriately, consider dropping instead of straight needles, needles of any closed-form shape, such as a polygon. Let any polygon, denoted by P, have a total length l, such that the total number of crossings once dropped randomly is the sum of the number of crossings produced by the pieces which compose it, giving an expectation of $E = kl$, by Theorem 2. Given this, now consider a circle, C, with some diameter of d, with a length exactly equal to $l = d\pi$. Assuming that this circle is now dropped unto the same equally spaced rectangular tiles placed parallel to each other as such:

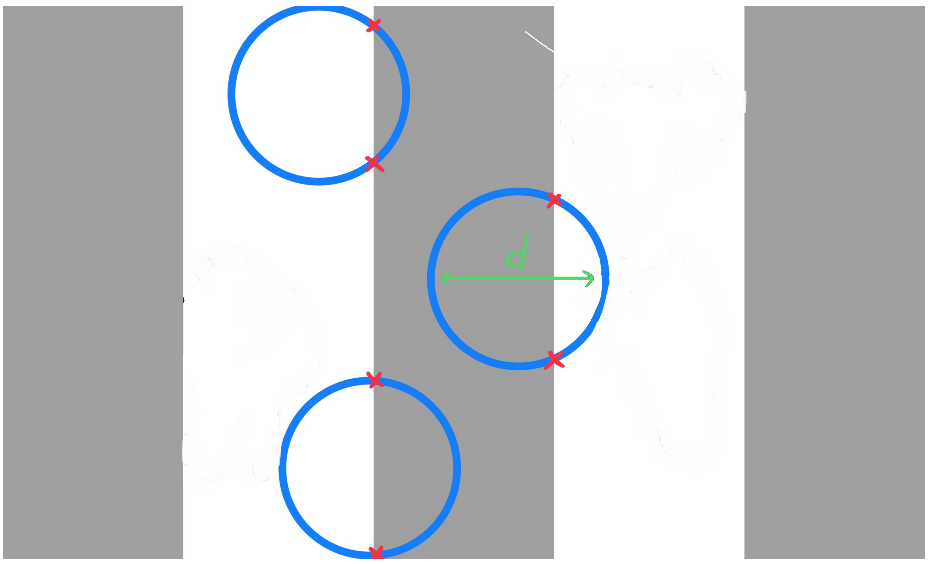


Figure 3: Diagram of Circles Dropped Unto given Surface

, it can be seen that it crosses atleast 2 dividing lines no matter the position dropped, such that $E_C = 2$. Now assume, along with the circle, two polygons, P_i, and P^i, one which inscribes, and the other circumscribes the circle C respectively, with i denoting the number of are also dropped along side C as shown in Figure 4:

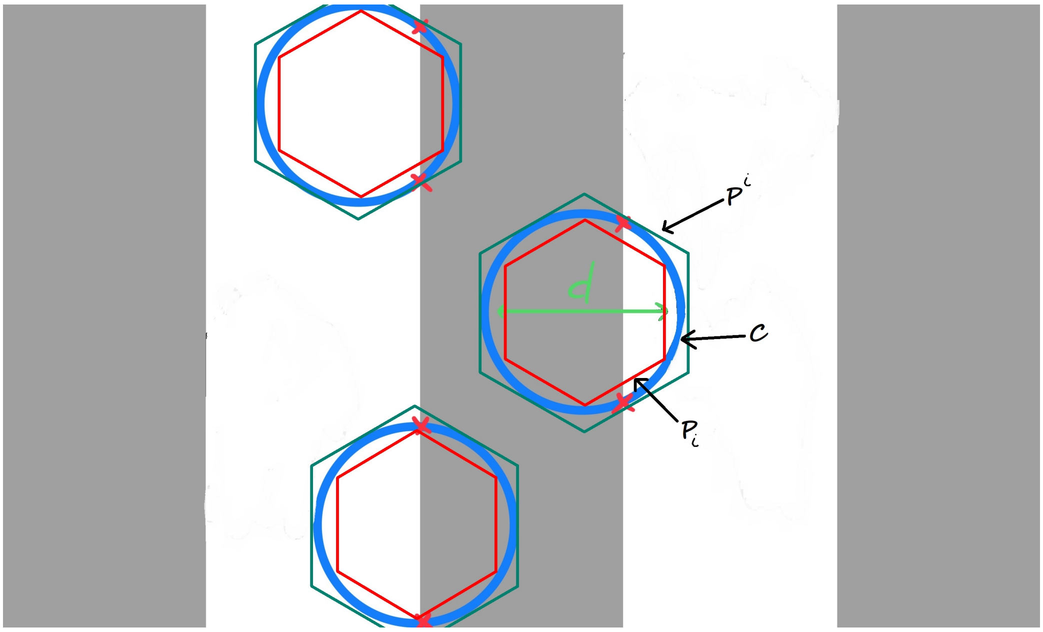


Figure 4: Diagram of Circles with Inscribing and Circumscribing Polygons Dropped Unto given Surface

It can be said that the expectations follow a relationship $E_{P_i} \leq E_{P^i}$. Particularly, the relationship

$$E_{P_i} \leq E_C \leq E_{P^i}$$

is true. However it should be noticed here that polygons P_i, and P^i are merely approximations of C, such that if we were to increase the number of sides of these polygons, they would produce continuously better an better approximations of a perfect circle! So therefore we can say that

$$E_{P_i} \leq E_C \leq E_{P^i}$$
$$\lim_{i \to \infty} l(P_i) \leq d\pi \leq \lim_{i \to \infty} l(P^i)$$

thus yielding, an expression for k,

$$k = \frac{2}{\pi}\frac{1}{d}$$

which is analogous to that of the general case for l, and where k is the probability of crossing, p, such that

$$p = \frac{2}{\pi}\frac{l}{d} \tag{2}$$

■

And thus we see that we have derived the solution for the probability of a needle intersecting the dividing lines of parallel planks. So from here, we can perform numerous re-arrangements to conduct a further statistical and probabilistic analysis to continue to try and answer the research question proposed earlier.

3 Approximating π by Probability

As we have determined the parameter p of our random variable H we can do some basic algebraic re-arrangement to see,

$$\begin{aligned} p &= \frac{2}{\pi}\frac{l}{d} \\ \Leftrightarrow \pi &= \frac{2}{p}\frac{l}{d}. \end{aligned}$$

Thus showing clearly the connection between the probability p and π with a clear expression. Which then prompts the question of in this re-arranged formula, how to make it such that we actually estimate p. Although a smart individual may presume this probability may simply be the proportion of successes to trials, this is not initially mathematically evident. And so we must undertake a quick proof to see how we may denote this parameter p. Hence we consider a specific case of a an experiment with $N = 8$ trials, with some probability p of success and $1 - p$ of failure. Let X denote the number of successes in this experiment such that $X \sim B(8, p)$. The probability distribution of such a random variable can be given in terms of a binomial distribution, such that if $X = 5$

$$P(X = 5) = \binom{8}{5} p^5 (1 - p)^3.$$

meaning that the likelihood of a total of 5 successes occurring out of a total of $N = 8$ trials is proportional to $p^5(1-p)^3$ for some parameter $p \in (0, 1)$. The following is the graph of observed as a result:

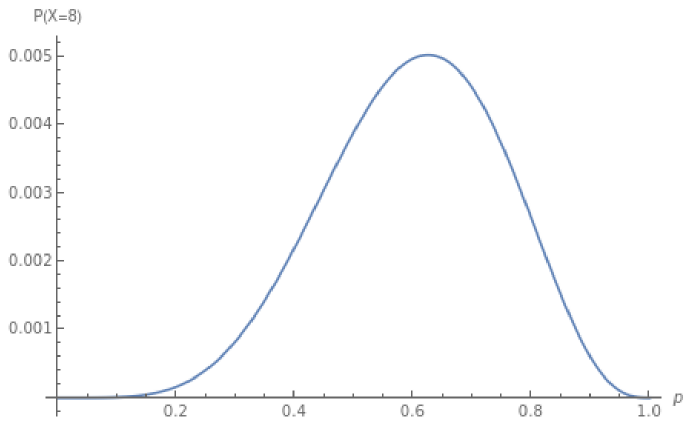


Figure 5: Graph of $P(X = 8)$ with Varied p

It can be seen here clearly that the graph actually has a maximum as p varies! Infact, we see that this maximum can be found by simply taking the derivative of this function and setting it equal to 0, methods familiar from early calculus classes. In an attempt to generalize, let us then consider an m number of experiments, trials, with a fixed N number of attempts, with a probability of p success. Further assume that these trials and attempts are independent such that $X_i \sim B(N, p)$,$i = 1, 2, 3, \ldots$. Thus consider the random variable Y representing the outcomes of all of trials such that $Y = (X_1, X_2, \ldots, X_m)$. From properties of independent events, as also presented in the Mathematics Analysis and Approches HL Booklet [2], it is known that

$$P(X = x \cap Y = y) = P(X = x) \cdot P(Y = y)$$

allowing us to generalize in saying that for m such independent events, we get that namely $P(Y = k)$, where $k \in \mathbb{N}$ and $x \in \mathbb{N}$ denote the number of successes is simply

$$P(Y = k) = P\left(\bigcap_{i=1}^{m} X_i = x_i\right) = \prod_{i=1}^{m} P(X_i = x_i). \tag{3}$$

Since we know that Y is a binomial random variable, we know that it's probability distribution can be given by

$$P(X_i = x_i) = \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i}$$

thereby giving, by substituting back into Equation (1) that

$$P(Y = k) = \prod_{i=1}^{m} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} \tag{4}$$

For sake of clarity, the new function in Equation (2) will be denoted a function of p and x such that

$$L(p, x) = \prod_{i=1}^{m} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} \tag{5}$$

As we also saw in the specific case, this function must have some local maxima on the interval $p \in (0, 1)$ the proof of which follows from the mean value theorem specifically. So we can now take the derivative of this function with respect to the parameter p. This is important to find the p which allow for maximum probability of a number of successes occurring. In order to make this process simpler, we may take the logarithm of L, $\log(L(p, x))$. But what is $\log(L(p, x))$? Expanding L we get

$$L(p, x) = \binom{n}{x_1} p^{x_1}(1-p)^{n-x_1} \cdot \binom{n}{x_2} p^{x_2}(1-p)^{n-x_2} \cdot \binom{n}{x_3} p^{x_3}(1-p)^{n-x_3} \cdot \ldots \cdot \binom{n}{x_m} p^{x_m}(1-p)^{n-x_m}$$

So the logarithm of the mth term being

$$\log\left(\binom{n}{x_m} p^{x_m}(1-p)^{n-x_m}\right) = \log\left(\binom{n}{x_m}\right) + x_m \log(p) + (n - x_m)\log(1-p)$$

thus it follows that

$$\begin{aligned}
L(p, x) &= \left(\log\left(\binom{n}{x_1}\right) + x_1 \log(p) + (n - x_1)\log(1-p)\right) + \ldots + \left(\log\left(\binom{n}{x_m}\right) + x_m \log(p) + (n - x_m)\log(1-p)\right) \\
L(p, x) &= \sum_{i=1}^{m} \log\left(\binom{n}{x_i}\right) + x_i \log(p) + (n - x_i)\log(1-p) \\
L(p, x) &= \sum_{i=1}^{m} \log\left(\binom{n}{x_i}\right) + \sum_{i=1}^{m} x_i \log(p) + \sum_{i=1}^{m} (n - x_i)\log(1-p).
\end{aligned}$$

Which means that we can take the derivative, $\frac{dL}{dp}$ such that

$$\begin{aligned}
\frac{dL}{dp} &= \frac{d}{dp}\sum_{i=1}^{m} \log\left(\binom{n}{x_i}\right) + \frac{d}{dp}\sum_{i=1}^{m} x_i \log(p) + \frac{d}{dp}\sum_{i=1}^{m}(n - x_i)\log(1-p) \\
\frac{dL}{dp} &= 0 + \sum_{i=1}^{m} \frac{x_i}{p} - \sum_{i=1}^{m} \frac{n - x_i}{1-p}
\end{aligned}$$

thus setting this equal to 0, we get that

$$\sum_{i=1}^{m} \frac{x_i}{p} - \sum_{i=1}^{m} \frac{n - x_i}{1 - p} = 0$$

$$(1-p)mx = p\sum_{i=1}^{m}(n - x_i)$$

$$mx = pn \cdot m$$

$$\hat{p} = \frac{x}{n}$$

And hence we see that the best parameter p given by $\hat{p}$ is the number of total successes over an n number of attempts attempt, as had been hinted at previously. Thus we see that for the specific case initially discussed, with a total of $n = 8$ trials, and 5 successes, the maximum likelihood of 5 successes is this simply $\hat{p} = \frac{5}{8} = 0.625$. More importantly however, we see that infact in the scenario Buffon's problem, where we have H as successes and N as total number of trials, we see that we can write the parameter p as $\hat{p} = \frac{H}{N}$, thus giving the maximum likelihood that a number of hits occurs in successive independent trials.

Namely, we can substitute this value of p into Equation (1) such that we now get the relation

$$\frac{H}{N} = \frac{2}{\pi}\frac{l}{d}. \tag{6}$$

We can again re-arrange this relationship, solving for π to see that

$$\pi \approx \frac{2l}{d}\frac{N}{H}. \tag{7}$$

For some brief reflection on notation, the $\approx$ symbol is used here since π is known to be irrational an so it cannot be represented as a ratio of two integer numbers. We could similarly choose to write $\hat{pi}$, so as to indicate the approximation, which will be used from now on. Similarly, the equality, =, for p is used as writing p doesn't assume a specific ratio, but the limit as more and more needles are thrown.

So given that we have found some $\hat{p}$ expression which maximizes the likelihood of a hit, we could examine in a practical sense, given that we would get different H values for each trial of Buffon's needle, how $\hat{p}$ varies across these trials. Consider performing Buffon's needle across many trials, such that each trial has a certain N number of throws and H number of hits associated with it. By using the formula in Equation (6) we could perhaps gain estimations on π itself as was done previously. We can use a simulation [4], to perform this experiment! Performing these trials with a constant $N = 250$ and $l = d = 1$ we could obtain a table such as this:

Table 1: Table of Hits and Approximated π Values for 6 Trials

Trial (T)	Hits	Approximate Value of π
T_1	155	3.2258
T_2	156	3.2051
T_3	168	2.9761
T_4	170	2.9411
T_5	160	3.1250
T_6	161	3.1055

We can thus notice that each $\hat{p}$ is different from the other, causing some error in the approximation of π. But how big is this error on average? We could try subtracting such that we see the difference $\hat{p} - p$, but then we run into the issue that perhaps negative values might cancel out the positive ones, and thus we could then try and square the difference. Visually, we can think of this as the distance of one value from some known value. So say we wished to find some parameters a, b such that we got a $1:1$ relationship with a target equation, say $ax + b = 2x + 5$, but in this case we have no idea what the target equation is.

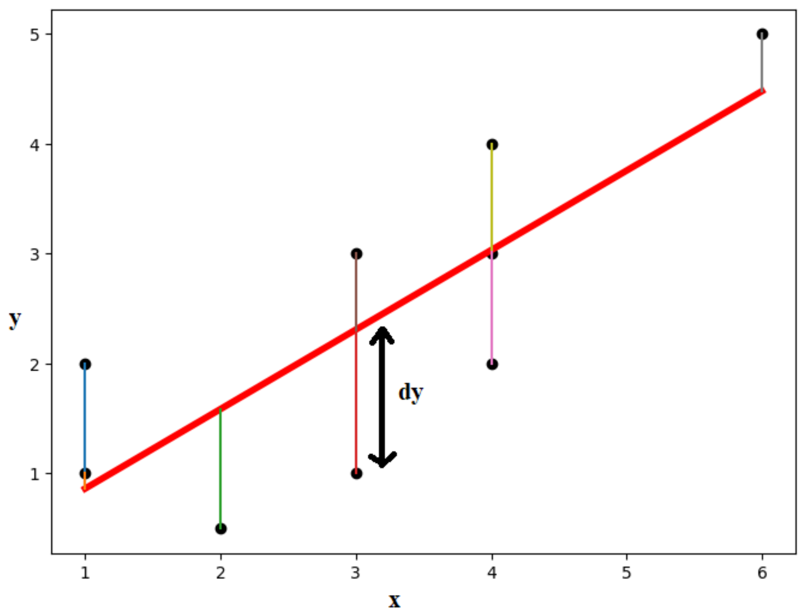


Figure 6: Graph of Collected Sample Data (Black Dots) and Target Equation (Red)

Then, we could simply utilize the Pythagorean distance formula which says that the square of the distance, d between two points is the sum in the change in the y, Δy and x, Δx [2]. So that we get

$$d^2 = \Delta y^2 + \Delta x^2$$

such that we get the distance between a collected sample point and the true value, in other words, the error. And so this actually works quite well, and thus we find that the squared error in each of our trials as such:

Table 2: Table of Hits, Approximated π Values, and Squared Errors for 6 Trials

Trial (T)	Hits	Approximate Value of π (5 s.f)	Squared Error (3 s.f)
T_1	155	3.2258	$7.09 \cdot 10^{-3}$
T_2	156	3.2051	$4.03 \cdot 10^{-3}$
T_3	168	2.9761	$2.73 \cdot 10^{-2}$
T_4	170	2.9411	$4.01 \cdot 10^{-2}$
T_5	160	3.1250	$2.75 \cdot 10^{-4}$
T_6	161	3.1055	$1.29 \cdot 10^{-3}$

But then again, we have conducted a total of 6 trials, and hence we could think of taking the average squared error as a way of seeing how good the parameter $\hat{p}$ is as a means of approximating π! Doing this process, we get namely that the average squared error, e, is:

$$e = \frac{7.09 \cdot 10^{-3} + 4.03 \cdot 10^{-3} + 2.73 \cdot 10^{-2} + 4.01 \cdot 10^{-2} + 2.75 \cdot 10^{-4} + 1.29 \cdot 10^{-3}}{6}$$
$$e = 1.33 \cdot 10^{-2}$$

Attempting to generalize, we could survey an m number of trials each giving some parameter $\hat{p}_m$ from a true p. Thus the mth squared error being $(\hat{p}_m - p)^2$, and hence the mth average square error as such:

$$e(p, m) = \frac{(\hat{p}_1 - p)^2 + (\hat{p}_2 - p)^2 + \ldots + (\hat{p}_m - p)^2}{m}$$
$$e(p, m) = \frac{1}{m}\sum_{i=1}^{m} (\hat{p}_i - p)^2$$

What is also curious in this process is also the fact that actually the bigger the error, un-squared, the bigger the squared error. Which means that the average squared error, e gets much bigger. So perhaps outliers in a collected sample would disproportionately effect the mean squared error, meaning that attempting to minimize it would ultimately fail in a finite collected sample.

We actually further notice that here, we can utilize the definition of expectation, Definition 2.1, from earlier! In fact, we get, by the definition of expectation, that the average squared error of the parameter $\hat{p}$ is

$$e(p, m) = E\left[(\hat{p}_m - p)^2\right]. \tag{8}$$

Given that we have such a relationship, we can reflect back on the meaning of the statement itself. Well, we were wishing to find out how exactly the value of this error varies across many trials, and for this we chose to examine the average squared error. But perhaps we could make use of another statistical tool to our advantage. Therefore, let us define the concept of variance,

Definition 3.1. The definition of the variance of a random variable is given by

$$Var(X) = E[(X - \mu)^2] \tag{9}$$

where $E[X]$ is the expectation and μ is the mean. This definition of variance applies for random variables which are continuous and discrete.

From this definition we see that infact the definition of the error function e looks quite similar to this. So maybe we could try and expand the contents of e and see what we get.

We have that e is given as:

$$e(p,m) = E\left[(\hat{p}_m - p)^2\right].$$

we can add and subsequently subtract a $E[\hat{p}]$, such that

$$e(p,m) = E\left[(\hat{p} - E[\hat{p}] + E[\hat{p}] - p)^2\right]$$

and then expanding as such

$$\begin{aligned} e(p,m) &= E\left[(\hat{p} - E[\hat{p}] + E[\hat{p}] - p)^2\right] \\ &= E\left[(\hat{p} - E[\hat{p}])^2 + 2(\hat{p} - E[\hat{p}]) \cdot (E[\hat{p}] - p) + (E[\hat{p}] - p)^2\right]. \end{aligned}$$

And thus by the linearity of the expectation, or namely, Theorem 2 and Lemma 1, we have that

$$\begin{aligned} e(p,m) &= E\left[(\hat{p} - E[\hat{p}])^2\right] + E\left[2(\hat{p} - E[\hat{p}])(E[\hat{p}] - p)\right] + E\left[(E[\hat{p}] - p)^2\right] \\ &= E\left[(\hat{p} - E[\hat{p}])^2\right] + 2(E[\hat{p}] - p)\, E\left[\hat{p} - E[\hat{p}]\right] + (E[\hat{p}] - p)^2 \\ &= E\left[(\hat{p} - E[\hat{p}])^2\right] + (E[\hat{p}] - p)^2 \end{aligned}$$

and by definition of variance, we get that

$$e(p,m) = Var(\hat{p}) + (E[\hat{p}] - p)^2. \tag{10}$$

Thus more specifically we notice that e can also be written as a function of the variance of the parameter p and another term. This relationship works perfectly for our purposes, as we could theoretically, and as shown experimentally previously, calculate the average squared error, e and thus attempt to minimize this error as much as possible. In other words, by which methods can we realistically minimize e?

In a finite context, to provide an approximation means to have some, be it small or large, error, and within that context, from Equation (9) we have seen that this error is equivalent to the variance of that same parameter. So to minimize we can take the approach of actually just simply ridding ourselves of the extra term $(E[\hat{p}] - p)^2$, such that $(E[\hat{p}] - p)^2 = 0$. We reserve the right to do this due to the fact that the parameter $\hat{p}$ in this case actually is completely unbiased, noticing that here $(E[\hat{p}] - p)^2$ is the bias term, in that it adds some bias to an otherwise un-biased parameter.

Going back to the context of the Buffon needle problem, and specifically Equation (6), we can write the difference of the approximated and true value as

$$\hat{\pi} - \pi = \frac{2l}{d}\frac{N}{H} - \frac{2l}{d}\frac{1}{p} = \frac{2l}{d}\left(\frac{N}{H} - \frac{1}{p}\right) = \frac{2l}{d}\left(\frac{1}{\hat{p}} - \frac{1}{p}\right). \tag{11}$$

So we see that to find the average squared error in the estimation, $e(p, m)$, as also outlined previously in Equation (9), of π, $\hat{\pi}$, we must take the variance of the term $\frac{2l}{d}\frac{N}{H}$. Thus, with the substitution of $R = \frac{l}{d}$ that the variance is

$$\begin{aligned} Var(\hat{\pi}) = Var\left(\frac{d}{2l}\frac{H}{N}\right) = Var\left(\frac{1}{2R}\frac{H}{N}\right) &= \frac{1}{(2NR)^2}Var(H) \\ &= \frac{1}{2N^2R^2}N\pi(1-\pi) \\ &= \frac{\pi(1-\pi)}{2NR} \end{aligned}$$

The $Var(H)$ term is calculated to be $N\pi(1-\pi)$, since we know H is a binomial random variable with parameters N and p, the variance would be equivalent to $Np(1-p)$ where in this case p is π. Given that this estimation, $\hat{\pi}$, is unbiased, as also explained previously, then the variance of this estimation is equivalent to $e(\pi, m)$ of $\hat{\pi}$,

$$Var(\hat{\pi}) = \frac{\pi(1-\pi)}{2NR} = e(\pi, m). \tag{12}$$

As also explained previously, choosing values of d and l such that their ratio $\frac{l}{d} = 1$ would give the most optimal approximation the fastest. Thus we may choose l and d such that we have $l = d = R = 1$. And further yet, we may choose to apply what is called the delta-method[5], [6], which gives us in the end

$$e(\pi) = \frac{\pi^2}{N}\left(\frac{\pi}{2} - 1\right).$$

Thus by using the methods outlined previously, we had managed to obtain the average squared error of any experiment one could choose to conduct using the Buffon needle experiment! Meaning that if we wanted to obtain an approximation of π to a total of 3 decimal digits, or 4 significant figures, $\pi = 3.141$, we would obtain that the average squared error, and as a consequence of Equation (10), the asymptotic variance in $\hat{\pi}$ in an ever increasing asymptotic number of experiments, m, would be roughly:

$$Var(\hat{\pi}) \approx \frac{5.628}{N} \tag{13}$$

However before proceeding to explore the extensions to the original Buffon problem, we can examine the nature of this relationship. Say, what would happen if $N \to \infty$? Well, using some basic laws of limits, we see that infact

$$\lim_{N\to\infty} Var(\hat{\pi}) = 0.$$

Which makes quite a lot of sense if one actually thinks about throwing an infinite number of needles. If we throw an infinite number of needles, we would be able to perfectly get the value of π. But since we cannot actually thrown an infinite number of needles, the question of, at which extent of needles thrown, do we have a variance which can give a reasonable approximation of π? Throwing $N = 100$ needles gives us

$$Var(\hat{\pi}) \approx \frac{5.628}{100} \approx 0.0562$$

and throwing $N = 1000$ needles

$$Var(\hat{\pi}) \approx \frac{5.628}{1000} \approx 0.00562. \tag{14}$$

And so we see that throwing $N = 1000$ needles gives a reasonable probability that we will be able to approximate π to at least 2 decimal places. And the reason for this being a probability and not a certainty is that the average squared error, is, as stated in it's name, an average, so in finitely many throws, there will always be some outlier or skew which means that this variance is in fact not a certainty. However, taking for example $N = 10000$ needles, is not only computationally quite easy to compute for a computer, it gives a variance of roughly $5.628 \cdot 10^{-4}$, quite low. And again here it is not necessary to show this in powers of 10, but for the sake of understanding the true magnitude of this many numbers, it is quite useful in order to quantify. So in sum by applying this method, we have found the efficiency, or how fast we can approximate π to a reasonably accurate degree.

4 Variations of Buffon's Needle

After publishing his original work on the now dubbed Buffon needle problem, other individuals wondered on how to alter the problem to perhaps reach a more general answer. One of the most well known individuals was Pierre-Simon de Laplace. He considered extending the grid, and considering a grid of lines spaced out by some distances a and b which were perpendicular to each other. Figure 7 shows this:

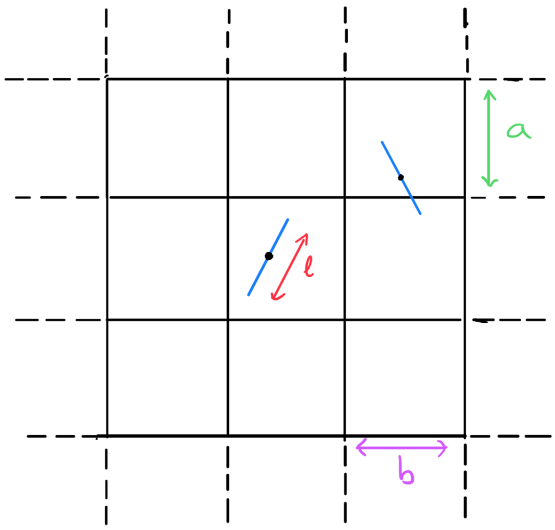


Figure 7: Diagram of Laplace's Extension

In this case, a solution to the problem was found by [7], using a paradoxical method, to be exactly

$$p = \frac{2l(a+b) - l^2}{\pi ab} \tag{15}$$

As in, this is the probability that given that a needle of length l is dropped unto two sets of equally spaced parallel lines, of spacing length a and b, with $l < a, b$ where one set is perpendicular to the other, the needle crosses one dividing line across one set of parallel lines. We can actually notice some similarities in this solution to that of Buffon's. Infact, we could wonder, what would happen if say the distances for one set of dividing lines increased continuously? In other words, what would happen if a or b increased to infinity? Let's try and compute the limit as $a \to \infty$.

$$\begin{aligned}
\lim_{a\to\infty}\left(\frac{2l(a+b)-l^2}{\pi ab}\right) &= \lim_{a\to\infty}\left(\frac{2l(a+b)}{\pi ab}-\frac{l^2}{\pi ab}\right) \\
&= \lim_{a\to\infty}\left(\frac{2la+2lb}{\pi ab}\right) - \lim_{a\to\infty}\left(\frac{l^2}{\pi ab}\right) \\
&= \lim_{a\to\infty}\left(\frac{2la}{\pi ab}\right) + \lim_{a\to\infty}\left(\frac{2lb}{\pi ab}\right) - \lim_{a\to\infty}\left(\frac{l^2}{\pi ab}\right) \\
&= \lim_{a\to\infty}\left(\frac{2l}{\pi b}\right) + \lim_{a\to\infty}\left(\frac{2l}{\pi a}\right) - \lim_{a\to\infty}\left(\frac{l^2}{\pi ab}\right) \\
&= \frac{2l}{\pi b}.
\end{aligned}$$

So infact, Buffon's problem is actually a limiting case of Laplace's variation! Although this is very intriguing, what insight does it give into approximations of π? Well, as we have seen with Buffon's problem, we can conduct a similar statistical analysis on this variation of Buffon's problem too. We follow the same procedure of re-arrangement, calculating the variance to find the average squared error, and applying the delta method to find the asymptotic variance which in turn, given that we plug in a value of π say $\pi = 3.141$ to which we want to approximate to, we can find the general efficiency, as we did with Buffon's solution.

We can think to even further generalize this approach. Perhaps we could consider including 3 sets of lines in our grid. This would result in some sort of polygonal pattern, and a visualization can be drawn as such:

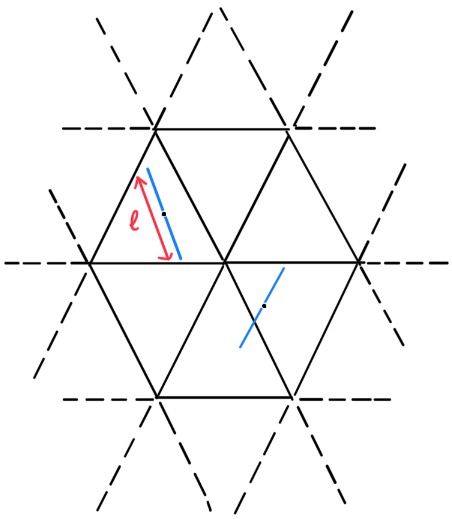


Figure 8: Diagram of the Triple Grid Variation

And so as we can see in Figure 8 that we have the 'triple grid' variation of the problem, with needles in blue, and the length l of the needles indicated in red. The dashed lines indicate that the pattern in the solid black lines continue indefinitely.

And so we may choose to repeat the process again for this variation as well, giving a relation for the average squared error of the approximation. So we may actually spot a certain pattern forming here, where we continuously add more and more sets of grid lines. So infact here the question of what actually would happen to the efficiency of the approximation as we add more and more grid lines could be posed. Or rather more formally, would adding more and more grid lines, yield better approximations of π than by keeping the set of grid lines at a minimum? So if we were to denote the number grid lines by M, would it be true that for some constants $c_1 < c_2$ where $c_1, c_2 \in \mathbb{N}$

$$\lim_{M \to c_1} Var(\hat{\pi}) < \lim_{M \to c_2} Var(\hat{\pi}) \tag{16}$$

at some finite N? Which could also be generalized to the question of whether or not for some constant $c_i \in \mathbb{N}$, such that $c_i > c_{i-1}, c_{i-2}, \ldots, c_2, c_1$

$$\lim_{M \to c_1} Var(\hat{\pi}) < \lim_{M \to c_2} Var(\hat{\pi}) < \ldots < \lim_{M \to c_i} Var(\hat{\pi}) \tag{17}$$

And this is mainly for some finite, and not infinite c_i, since we are interested in the computational feasibility of such a calculation, and thus we cannot truly work with infinities. So we see that by performing a set procedure of statistical analysis and attempting to generalize the original Buffon's variation, we get the extent to which these probabilistic methods and problems can approximate π to a 'reasonable' degree. But, then again, what is a 'reasonable' degree? I have been using this term and it's synonyms to quantify the accuracy of an approximation, yet these terms are very loosely defined. So perhaps this could mean, even in a more general context outside this specific investigation, a 'fast' and 'significant' approximation. So that it's fast in that with a few trials, a 'significant, or, to a couple decimal places can be achieved.

So therefore, there is left only another avenue of investigation within which we could into, and more specifically manipulations of these probabilistic experiment to give possible well-known generators of π to an accurate degree, which would also provide a very thorough contribution to the investigation at hand.

5 Conclusion

We started with a simple yet quite elegant problem in which we considered dropping a needle on a surface with a set of parallel lines equally spaced from one another. We managed to provide a solution to the problem of finding the probability of a random needle crossing at least a single dividing line on the surface, noticing that in this result was π! We used this result to begin to ponder the research question regarding the viability of actually using some sort of numerical approximation by simply re-arranging the equation to solve for π. Using some calculus techniques and observations, we managed to find an expression for the probability p such that the probability of a needle hitting a dividing line was maximised. From here, a statistical analysis involving the average squared error and the variance of the approximations from data collected of π was conducted to arrive in Equation (13)

$$Var(\hat{\pi}) = \frac{5.628\ldots}{N}.$$

After this observation, an extension in the form of an increasing number of sets of dividing lines was proposed, ending with a conjecture that increasing the number of dividing lines, M, to some $c_i > c_{i-1}, c_{i-2}, \ldots, c_2, c_1$ would actually give a faster and better approximation for π.

Although the analysis conducted did serve to answer the research question quite thoroughly, it still leaves room for gaps. Namely, a notable reference and subsequent analysis to past literature, through the work of Italian mathematician Lazzarini [8] can be commented upon. This is such that perhaps by manipulating the values of d and l, we may obtain possible generators of well-known approximations to π, such as $\pi \approx \frac{355}{113}$, as was the case for Lazzarini's experiments. Though given the fact of the sheer number of possible needles that can be thrown and also hits recorded, we may think that this result is quite extraordinary. We can quickly gain an insight as to the true probability of such a result being obtained by utilizing Equation (13), since we can simply substitute $N = 3408$ to obtain

$$Var(\hat{\pi}) = \frac{5.628}{N} \Rightarrow \frac{5.628}{3408} \approx 0.00165 \tag{18}$$

Thus meaning that at best, π can be approximated to roughly 3-4 decimal places, which the approximation of Lazzarini surpasses by a whole 2-3 decimal places, putting into great question his result.

However, the question of whether or not Lazzarini actually performed this experiment is not really important, but rather the method of choosing l and d to give these well-known approximations. It is not impossible that these ratios arise themselves out of natural experimental methods, since it is the viability of obtaining a good approximation of π we are investigating, these ratios are simply a few out of the many many possibilities which could occur. Thus stumbling on these few ratios would actually be quite comparable to attempting to find a needle in a haystack.

Further yet, I also thought of extending this problem of Buffon's needle even further. Namely, what would happen if we were to consider 'dropping' a needle in a 3 dimensional space? So basically by considering how would these spaces would look like, what would the resulting probability be? I think this could be a very interesting extension to make to the problem, and to conduct a subsequent statistical analysis upon. Also, perhaps analysing the cases where the length of the needle is such that $l > d$ could also be examined for the original problem and it's extensions.

In general however, we see that these probabilistic methods of approximating π can be used to approximate π to a very decent error difference. Especially with the computational power the average layman possess, this process can be automated such that the experiment is conducted indefinitely to observe approximated values of π. Methods similar to that of Lazzarini, where the length of the needle and the spacing between parallel planks are manipulated in order to have the possibility of producing known rational fraction approximations of π, are valid to the extent that the values for N and H are recorded for every experiment conducted, and are indeed random. Thus we can conclusively conclude that although yes, these probabilistic methods can be used to approximate π, they would require a great number of needles thrown, to even achieve a degree of approximation to say 4 decimal places. Although, by using different variations of the problem, it could be feasible that an decent approximation can be obtained with only a few hundred needles.

References

[1] G. Galperin, “Playing pool with (the number from a billiard point of view),” *Regular and Chaotic Dynamics*, vol. 8, Jan. 2003, Date Accessed: 12/28/2020. DOI: 10.1070/RD2003v008n04ABEH000252.

[2] *Mathamatics: Analysis and approaches formula booklet*, Date Accessed: 2/6/2021. [Online]. Available: https://www.ibdocuments.com/IB20DOCUMENTS/Data20and20Formula20Booklets/Mathematics20Analysis20and20Approaches/Formula20Booklet20-20English.pdf.

[3] M. Aigner, Z. G. M., and K. H. Hofmann, “Analysis, buffon’s needle problem,” in *Proofs from the Book*. Springer, 2018, Date Accessed: 1/10/2021.

[4] E. Siniksaran, *Throwing buffon’s needle with mathematica*, Date Accessed: 12/28/2020, 2011. [Online]. Available: https://www.mathematica-journal.com/volume/v11i1/.

[5] A. Papanicolaou, *Taylor approximation and the delta method*, Date Accessed: 1/18/2021, Apr. 2009.

[6] J. Ver Hoef, “Who invented the delta method?” *American Statistician - AMER STATIST*, vol. 66, May 2012, Date Accessed: 1/18/2021. DOI: 10.1080/00031305.2012.687494.

[7] J. V. Uspenskij, “Probabilities in continuum, laplace’s problem,” in *Introduction to mathematical probability*. McGraw-Hill, 1937, Date Accessed: 1/11/2021.

[8] M. Lazzarini, “Un’applicazione del calcolo della probabilita alla ricerca sperimentale di un valore approssimato di π,” *Periodico di Matematica per l’Insegnamento Secondario*, 2nd ser., vol. 3, pp. 140–143, 1901, Date Accessed: 1/16/2021.