2016

A Computational And Theoretical Exploration of the St. Petersburg Paradox

Alexander Olivero
Butler University, atoliver@butler.edu

BUTLER UNIVERSITY HONORS PROGRAM

Honors Thesis Certification

Please type all information in this section:

Applicant: Alexander Thomas Olivero
(Name as it is to appear on diploma)

Thesis title: A Computational and Theoretical Exploration of the St. Petersburg Paradox

Intended date of commencement: May 2016

Read, approved, and signed by:

Thesis adviser(s): William Johnston — May 3, 2016
Date
Ankur Gupta (by Wm Johnston, on his behalf) — "
Date

Reader(s): Chris Wilson " " — "
Date

Certified by: Rusty Jones — 5/4/16
Director, Honors Program — Date

For Honors Program use:

Level of Honors conferred: University ______

Departmental ______

A Computational and Theoretical Exploration of the St. Petersburg Paradox

A Thesis

Presented to the Department of Mathematics

and

College of Liberal Arts and Sciences

of

Butler University

In Partial Fulfillment

of the Requirement for Departmental Honors

Alexander Thomas Olivero

Abstract

This thesis displays a sample distribution, generated from both a simulation (for large n) by computer program and explicitly calculated (for smaller n), that is not governed by the Central Limit Theorem and, in fact seems to display chaotic behavior. To our knowledge, the explicit calculation of the sample distribution function is new. This project outlines the results that have found a relation to number theory in a probabilistic game that has perplexed mathematicians for hundreds of years.

Acknowledgements

I would like to thank my advisers Dr. Bill Johnston and Dr. Ankur Gupta for all of their assistance on this project. It would not have been such a success without your guidance.

Contents

1 Introduction

In 1713, Nikolaus Bernoulli sent a letter to French mathematician De Mondmort in which he defined a theoretical game of chance [1]. The game works in this manner: The player flips a fair, two-sided coin. If the coin tosses heads on the first flip, the house pays \$2. If the first heads tossed occurs on the second toss, the house pays \$4. If it happens on the third toss, the house pays \$8. To generalize, the house pays out larger powers of two as the first heads occurs on later tosses: $\$2^n$ if it is on the n^{th} toss. The paradox occurs when applying the probabilities to each of these situations to find the expected value for the game. The paradox is explained in formal mathematical symbols as:

$$E[X] = (\tfrac{1}{2})(2) + (\tfrac{1}{4})(4) + (\tfrac{1}{8})(8) + ... + (\tfrac{1}{2^n})(2^n) + ...$$
$$E[X] = 1 + 1 + 1 + 1 + ...$$
$$E[X] = \infty,$$

where $E[X]$ is defined as the expected value for X.

By its mathematical definition, a player of this game should expect to win an infinite amount of money. But, surely no one would dare pay any large sum of money on this game in which half of the time the player will only receive \$2 as the payout — and that is the paradox!

Since this problem was originally presented, countless mathematicians have attempted to understand it in more detail. Daniel Bernoulli became aware of the problem through correspondence with his cousin and gave it a title when he published it in the St. Petersburg Academy Proceedings [1]. He investigated the calculation of the expected utility of the game, rather than the expected payout. This work examines the satisfaction produced by a dollar outcome. In economics, this relates very closely to the idea of diminishing returns. At a certain point, a consumer receives no more gratification from having more of a product [4]. Since this initial investigation, economists such as Adam Smith and John Maynard Keynes have investigated this problem. A multitude of papers have been published examining the implications of this simple problem.

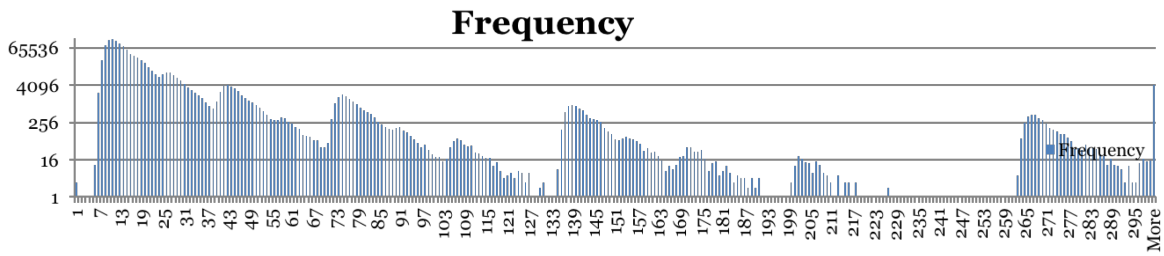


Figure 1: A Sample Probability Distribution Histogram for $n = 2048$ plays of the game

An 18th century French mathematician Compte de Buffon was so concerned that real-life experience would not match this theoretically-calculated expected value that he decided to play the game 2048 times and see what happened by running the experiment by hand with an actual coin [2]. The story goes that he paid a child to flip a coin and record what had happened.

In 2011, mathematicians Dominic Klyve and Anna Lauren repeated this procedure with a computer. In fact, they did more: they played the game 2048 times (calling it "the Buffon experiment"), but then performed what is known as a Monte-Carlo method.

Definition 1.1. *A Monte-Carlo method is a technique that averages a large number of simulated results in order to numerically approximate the solution of a mathematical problem that studies the distribution of some random variable, often generated by a computer.*

Simply put, they did what Buffon did, but iterated it one million times. Using the Monte-Carlo method proves to be an effective means to examine large numbers of plays of the game. Simulating the game allows the observer to notice trends to identify defining characteristics. The result of this simulation — the frequency of each average payout — yields an interesting graph, called a histogram. This thesis project began by generating such a histogram for Buffon's experiment. This result is seen in Figure 1.

Definition 1.2. *We define a sample probability histogram as a graphical representation of an estimate obtained via sampling or simulation of the probability distribution of a continuous variable.*

Klyve and Lauren first generated such a histogram for the St. Petersburg Paradox problem in 2011. They note that "[the sample probability histogram] is surprising. Its comb-like, fractaline quality demands explanation.[2]" This project replicates this work but also takes it farther by finding a mathematical explanation for the properties of this paradox.

The average payouts that are displayed on this histogram are known to mathematicians as random variable values for this experiment. Simply put, a random variable is a numerical value assigned to an outcome of a probabilistic experiment. A more formal definition will be provided later in this paper. The work of this thesis aims to describe the probability distribution associated with this random variable.

There are many interesting aspects left to be explored following this work. Obviously there is something unique happening to cause the spikes in probability. By the Central Limit Theorem, one might expect this distribution to approach a normal bell-shaped curve. This is clearly not the case, and this project will provide a proof for why this paradox violates this fundamental concept from probability theory — one of the most important theorems in this subfield. Additionally, there are many intervals of average payouts for which the probability is zero for this distribution. Such intervals of zero probability do not typically occur for distributions of averages of random variables. These characteristics demand and deserve explanation, as they seem quite bizarre to have come from a simple coin flipping game.

This thesis only begins to scratch the surface of the depth of this amazing paradox. There is a powerful connection to powers of two that arises from the construction of the game. There is an application of a computer science program with algorithmic properties that is used as a tool for further understanding of the game. Additionally, this project stumbled upon a link to the mathematical subfield of Number Theory, specifically regarding partitions of numbers. Even after all of this investigation, there remain aspects to be explored by future mathematicians. What might one find in exploring the number of coin tosses as the examined random variable? What might occur if the coin is not equally weighted between heads and tails – further, what might occur if the coins weight changed after each toss? The game itself has a certain breadth of sophistication allowing mathematicians and economists alike to dive into problem that will have ever more results, observations, and applications.

2 Computer Simulation

To gain a fundamental understanding of the algorithm used to simulate the St. Petersburg Paradox, this project included the writing of a C++ program. This program allows the programmer to select a number of games for which the program will play, then averages the payout, and stores the data into a text file. The program then exports the data to Microsoft Excel and converts it into a histogram. The output is then displayed for the selected parameter.

The program works by utilizing a function called *cointoss* that plays the game once and returns the value of the payout for that play of the game. This function sets an integer counter variable x to 0 and sets another integer variable *coin* to 0 as well. *coin* equalling zero is equivalent to being in the state of tails. A while loop then "flips" the coin by setting the variable number either to 0 or 1, and then increments the counter variable x by one. This continues while *coin* is equal to 0 ("tails"). If the random value selected is one, the "player has flipped heads" and the while loop breaks, returning 2^x, where x is the counter variable.

The function *buffonexp* creates an array of size n which is determined by the user. A "for loop" then fills each bin of the array with a run of the function *cointoss*. A new for loop then creates a variable *sum* and adds the value of each bin. This is equivalent to adding the payout of each of the n games that have been played. Finally, this function creates a double variable *avg* and divides sum by n, and returns this value. This gives the average payout for the n plays of the game.

The main portion of the program essentially repeats the process of *buffonexp*, except it creates an array of size 1,000,000 and utilizes a while loop to fill each bin with a call of *buffonexp*. Additionally, this function opens a text file and uses a file stream object to export each of the values to the text file. At the end of the program, the average of the average payouts is displayed for the user.

After the computer program is completed, the user then copies the data from the text file and pastes it into a column of Microsoft Excel. After selecting all of the column, the user selects histogram for the data toolbar to create the desired sample probability histogram.

This project is in no way, shape, or form intended to prove impressive from a computer science standpoint. If fact, this code is a "brute force" method to accomplish the goal of completing these calculations and is not optimized for running time. This work might make some computer scientists cringe, but its simplicity is all that is required to accomplish this mathematical investigation.

This C++ code is found as Appendix A. While simple to understand, the insight it provides into the St. Petersburg Parodox proved invaluable—setting the stage for an impressive investigation of the probabilistic and number theoretic characteristics.

The following represent the simulated outputs for two and three plays of the game (Figures 2 and 3, respectively), which will be investigated in a formal, mathematical manner.

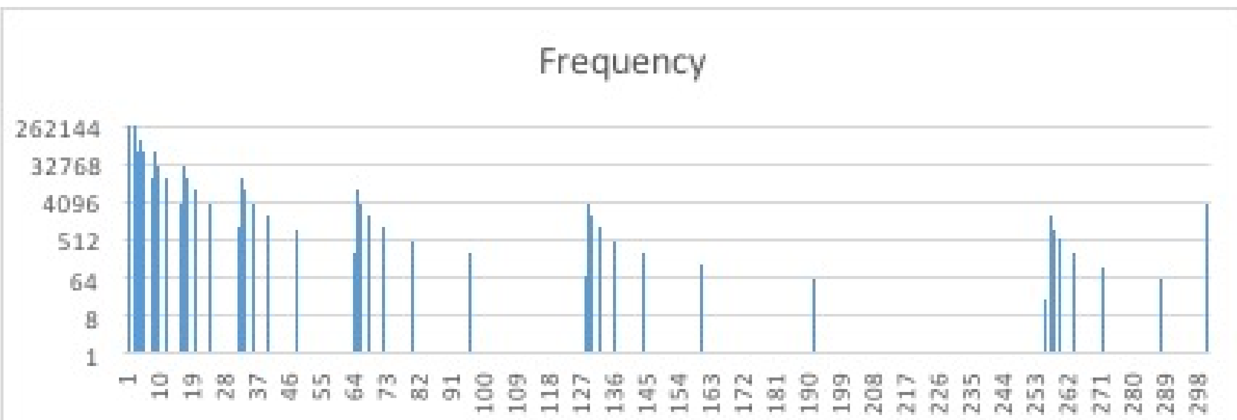


Figure 2: A Sample Probability Histogram of average payouts for $n = 2$ plays of the game iterated using the Monte Carlo Method

The sample histogram for $n = 3$ plays of the game appears in Figure 3 on the next page. See Appendix C additional sample histograms created by this simulation, for $n = 4$, $n = 10$, $n = 4096$, 8192, 16384, and 32768.

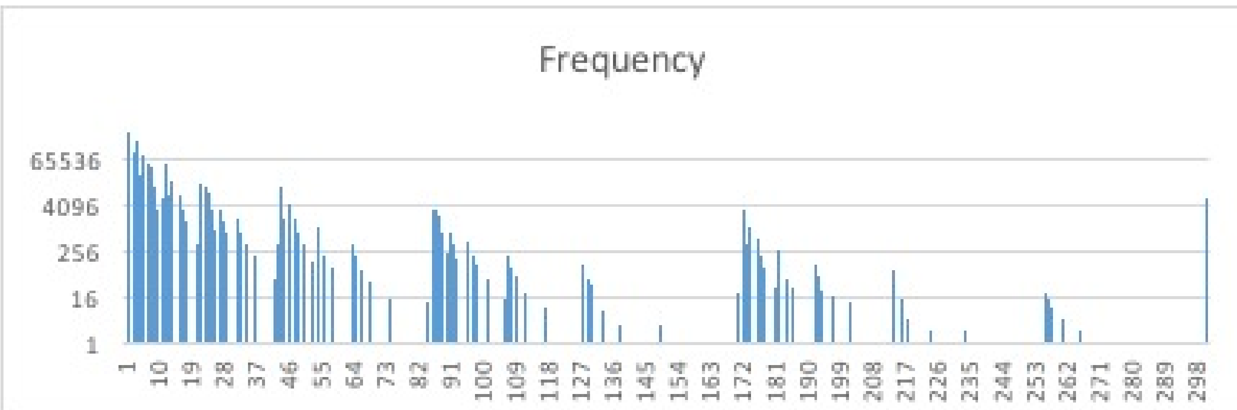


Figure 3: A Sample Probability Histogram of average payouts for $n = 3$ plays of the game iterated using the Monte Carlo Method

3 Mathematical Analysis for Two and Three Plays of the Game

Upon completing an examination of the computer simulation, the attention is now turned to the explicit mathematical investigation of the true population probability distribution for the St. Petersburg Paradox. This portion of the project began with tedious calculations of specific cases, which in turn led to insights as to why the probability histograms have such bizarre shapes..

The significant portion to note for this analysis of the game is that, instead of analyzing the game itself, we analyze a given number of plays of the game in which the random variable is the average payout.

Definition 3.1. *A random variable assigns a number to each possible outcome in a random experiment.*

A simple example of a random variable would be the value of a rolled die. In this experiment, we could assign a random variable X to be the number of marks on the side that is face up. If one were to roll the die and have it land on the side with four marks then, we would define the random variable's value as $X = 4$.

Definition 3.2. *Events A and B are independent when $P[A \text{ and } B] = P[A] \cdot P[B]$.*

In probability theory, it is important to note if events are independent, as it affects the calculation of the probability for that scenario. For example, if one wished to know the probability of rolling a 2 on a fair die and then rolling a 6 on that same die the probability would be $\frac{1}{36}$. In particular, since each event is independent and they have a $\frac{1}{6}$ probability, the probability of rolling a 2 and then a 6 is $\frac{1}{6} \cdot \frac{1}{6}$. This concept will be utilized frequently throughout this project, with clarity of its application becoming apparent through examples.

Instead of examining plays of the game, we fix the average payout and find the values of individual payouts that combine to form it. For example, an average of \$8 on three plays of the game could come form payouts of 8,8, and 8 or from payouts of 4,4, and 16. This approach then utilizes a new random variable, which is the payout on an individual play of the game. Then, by associating the probabilities of each of these individual payouts, the probability of the given average payout can be found. We will also use the following concept.

Definition 3.3. *A discrete random variable is one which has a value that is typically obtained by counting something. The number of values the random variable assigns is either finite or countably infinite. Any discrete random variable X is described by its probability mass function $P[X = x]$, where x is a value that X assigns.*

Because tossing a coin produces what are typically discrete random variables and because $\overline{X_n}$ = the average payout on n tosses, is described through a probability mass function. The first step toward understanding and trying to describe $\overline{X_n}'s$ probability mass function is to examine small values of n. Trying to determine this function for $n = 2048$ would be extremely difficult.

Therefore, this investigation starts first with the simple case for $n = 2$ plays of the game. It is first important to observe a notational subtlety. Instead of examining probabilities for $\overline{X_n}$, look at $n{\cdot}\overline{X_n}$. This minor change will the work save from constantly dealing with fractions. The mathematical analysis of this random variable $n \cdot \overline{X_n}$ is identical to the analysis of $\overline{X_n}$. Because the random variable is constructed to be the average payout, the value shall be restricted to the even numbers as there is no odd parity possible. This can be observed simply by the fact that the smallest such number of coin tosses in a given play of the game is one. Raising this as a power of two will give 2. It follows that all sums of powers of two are even, by definition. The set of possible payouts are all even integers, and the sum of any number of even integers will always itself be an even integer.

Definition 3.4. *The notation (x_1, x_2) represents the outcome on two plays of the game, where the payout of the first play is x_1 and the payout of the second play is x_2.*

With this notation, examine the smallest such payout for two plays of the game: \$4—tossing heads on the first toss for both plays of the game. This is represented as $P[2\overline{X_2} = 4] = P(2, 2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2^2}$. The preceding notation $P[2\overline{X_2} = 4]$ is defined as the probability that the payout for two plays of the game will average to be \$ 4.

Because there are only two plays of the game, it is quite easy to develop the following formula for the expected payout of two plays of the game.

Theorem 3.5.

$$P[\overline{X}_2 = (2^k + 2^j)/2] = \begin{cases} (\frac{1}{2})^{j+k} & \textit{if } j = k \\ (\frac{1}{2})^{j+k-1} & \textit{if } j \neq k \end{cases}$$

Proof. For the first case, for which $j = k$, using the definition of the probabililty and the independence of each play of the game, $P[2\overline{X}_2 = (2^k + 2^j)] = (\frac{1}{2})^j \cdot (\frac{1}{2})^k = (\frac{1}{2})^{j+k}$. For the second case, $P[2\overline{X}_2 = (2^k + 2^j)] = 2(\frac{1}{2})^j \cdot (\frac{1}{2})^k = 2 \cdot (\frac{1}{2})^{j+k} = (\frac{1}{2})^{j+k-1}$.

□

Moving beyond the simplistic case of $n = 2$, a similar notation for n plays of the game is used.

Definition 3.6. *The notation $(x_1, x_2, ..., x_n)$ represents the outcome on n plays of the game, where the payout of the i^{th} play is x_i, for $i = 1, 2, ..., n$.*

This convention for the probabilities is adopted for the remainder of this paper. So for $n = 3$ plays of the game, a similar example would be $P[3\overline{X_3} = 6] = P(2, 2, 2) = \frac{1}{2^3}$, which represents the probability of earning a payout of \$2 on each play of three distinct games. The following represents the explicitly calculated values for a few different probabilities for $n = 3$ plays of the game:

$P[3\overline{X_3} = 8] = 3 \cdot P(2, 2, 4) = \frac{3}{2^4}$

$P[3\overline{X_3} = 10] = 3 \cdot P(2, 2, 6) = \frac{3}{2^5}$

$P[3\overline{X_3} = 12] = 3 \cdot P(2, 2, 8) + 3! \cdot P(2, 4, 6) = \frac{7}{2^6}$

$P[3\overline{X_3} = 20] = 3 \cdot P(8, 8, 4) + 3 \cdot P(16, 2, 2) = \frac{3}{2^{10}} + \frac{3 \cdot 16}{2^6} = \frac{7}{2^6}$

$P[3\overline{X_3} = 50] = 3! \cdot P(32, 16, 2) = \frac{3!}{2^{10}}$

By explicitly exploring these probabilities, the structure of the values gave insight to find trends and create conjectures. The first identifiable feature of these values is a formula that explains the values for powers of two for the average payout that is being calculated. One of the first observations one would make when viewing these values is the special case that occurs around powers of two. In the form of an equation:

Equation 3.7. $P[\overline{X}_3 = 2^k] = 7/2^{3k}$

Proof. $P[3\overline{X_3} = 3 \cdot 2^k] = P[(2^k, 2^k, 2^k)] + 3 \cdot P[(2^{k+1}, 2^{k-1}, 2^{k-1})] = \frac{7}{2^{3k}}.$

□

Below are examples of the equation confirming the values of each probability.

Example 1: $P[3\overline{X}_3 = 24] = 3 \cdot P[(16, 4, 4)] + P[(8, 8, 8)] = 3/2^8 + 1/2^9 = 7/2^9$

Example 2: $P[3\overline{X}_3 = 48] = 3 \cdot P[(32, 8, 8)] + P[(16, 16, 16)] = 3/2^11 + 1/2^12 = 7/2^{12}$

It is no surprise that there exists a formula that is unique to values of the form 2^k. The intuition here is in the number of ways a number can be represented as the sum of powers of two. In the examples above, there are multiple ways to represent each number as a sum of powers of two. This significantly increases the probability of this expected payout occurring as there are more contributing factors. This begins to provide insight into the "jumps" that occur on the sample probability distribution histogram. These "jumps" are the values for which there is an increase in probability. The general trend for the probability distribution is decreasing, but occasionally, there are areas where the probability increases quickly and then begins to decrease again. Clearly, when there is more than one way for a given number to be represented, there will be a higher probability for that number.

An understanding of the "jumps" was a significant first step towards clearly defining how this paradox works, but another portion of the graph demanded explanation—specifically the large "gaps" where the probability is zero. The following theorems and equations provide an explanation for the when the probability for a given value will be zero and when a gap exists, as well as how large that gap will be, where the probability is zero.

Theorem 3.8. *For $x \in \{6, 7, 8, ...\}$ and if x cannot be written as $2^j + 2^k + 2^l$ for $j < k < l$ or if x has an even number of ones in its base two expansion, then $P[3\overline{X}_3 = x] = 0$.*

Proof. If x cannot be written as $2^j + 2^k + 2^l$ for $j < k < l$ and if the probability is not zero, then there must be at least two of the three plays where heads was first tossed on the same numbered coin toss. Say $j = k$, or (if heads is tossed on the same numbered toss in all three plays) $j = k = l$. Then either $x = 2^{j+1} + 2^l$ or $x = 3 \cdot 2^j = 2^{j+1} + 2^j$, which in either case would mean x has an even number of ones in its base two expansion. □

Theorem 3.9. *Suppose positive integer powers satisfy $n > m > k$. For any x with $2^n + 2^m + 2^k + 2 \leq x \leq 2^n + 2^{m+1} - 2$, $P[3\overline{X}_3 = x] = 0$.*

As an example for the above theorem, choose n=9, m=5, and k=4, so that the corresponding powers of 2 are 512, 32, and 16. By the above theorem, $P[3\overline{X}_3 = x] = 0$ for every value x between 562 and 574, inclusive.

View below the mathematically calculated probability distribution for $n = 3$ up to the average payout of \$62. Already becoming more dense than $n = 2$, this histogram gives a visual for the theorems described above as well as verifies the computer simulation's representation. Note the "jumps" as well as the "gaps" for which the probability is zero.

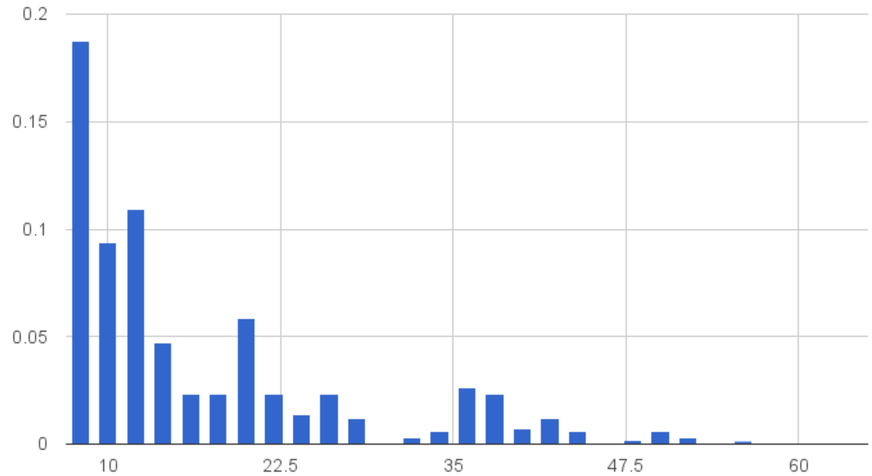


Figure 4: The Explicitly Calculated Probability Distribution Histogram for $n = 3$ plays of the game

4 The Random variable Y = the number of coin tosses

Examining the expected payout as the random variable is quite different from examining the number of coins tossed for each play of the game. This variable presents itself in an interesting manner because of its direct connection to the construction of the game. Instead of focusing on the payout, simply examining the variable of the number of coin tosses leads to a closer insight into the infinite expected value. The difference is subtle, but simple. Not raising this value as a power of two significantly changes the analysis of probabilities. This aspect of the St. Petersburg Paradox is closely related to an important type of random variable—one that arises often in probability theory—and is known as the Negative Binomial Random Variable.

The construction of the game leads to a connection with what is known as the Binomial Distribution. There are three defining characteristics for the Binomial Distribution [3], which are all met by the characteristics of tossing a coin.

- Each trial must have just two outcomes, a success and a failure – we assign heads as a success and tails as a failure
- The probability of a success must always be the same – the probability of a heads is .5
- Each trail is independent of any other trial – the outcome of one coin flip does not effect the outcome of another

From this explanation, it is clear that each coin toss in the construction of the St. Petersburg Paradox meets these characteristics.

Furthermore, we understand a more well-defined subcategory of the Binomial Distribution, specifically the Negative Binomial Distribution.

Definition 4.1. *A statistical experiment is a Negative Binomial Experiment if the following properties hold [5]:*

- *The experiment consists of n repeated trials.*

- *Each trial must have just two outcomes: a success or a failure.*
- *The probability of a success is the same for every trial.*
- *Each trail is independent of any other trial.*
- *The experiment continues until a predetermined number of successes has occurred*

From examining the explicit calculations of for probabilities of each scenario, the following equation was found to represent each probability.

Label X_{NB} the random variable counting the total number of trials seen to obtain heads in n plays of the game.

$$P[\overline{Y_n} = k] = P[X_{NB} = n(k-1)] = \tbinom{n-1}{k-1}(\tfrac{1}{2})^{n+k}$$

While this equation seemed to be quite the insightful result for the problem, it was quickly found that this indeed is the Negative Binomial Mass Function. While the result was not new, it certainly proves to be an interesting example of when the Negative Binomial Distribution applies [5]. The insight to the big problem of explaining the paradox is minimal from this result, but this is a unique way to determine the equation for the probability of a negative binomial distribution.

5 The Central Limit Theorem

One of the most fundamental concepts from probability theory is the Central Limit Theorem. This theorem governs most random experiments that arise. The most common feature that is seen when studying distributions of sample averages is that they follow an approximately normal "bell-shaped curve" that is quite familiar. This theorem gives a powerful insight into any random variable that it governs, but the investigation of this thesis problem leads one to believe that it does not apply to the random variable of the average payout of a given number of plays of the St. Petersburg Paradox game. The Central Limit Theore will now be formally stated.

Theorem 5.1. *Suppose that $X_1, X_2, ...$ is a sequence of independent and identically distributed random variables, each with finite expected value μ and finite nonzero standard deviation σ. Let Z_n be the standardized version of $\overline{X_n}$, i.e.*

$$Z_n = \frac{\overline{X_n} - \mu}{(\sigma/\sqrt{n})}.$$

Then as $n \to \infty$, $Z_n \to \mathcal{N}(0, 1)$, the normal random variable with mean 0 and standard deviation 1.

Stated in plain words: Given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent and identically distributed random variables, each with a well-defined finite expected value and well-defined finite variance, will be approximately normally distributed, regardless of the underlying distribution.

It is well known that one of the certain conditions required is a finite expected value. This characteristic, by definition, excludes a play of the game from being governed by this theorem. But what about the case in which the average payout of multiple plays of the game is examined? One might expect this this to closely resemble a normal distribution, especially with the large number of plays—2048—that the computer program considers. As we see with the graph of the distribution, we find that this is not the case. There is no bell-shape to it at all. Why is this true? The paradox's violation of the Central Limit Theorem is explained in the following theorem.

Theorem 5.2. *Let X Be the payoff random variable that assigns $X = 2^k$ for tossing heads on the k^{th} toss. Since $E[X] = \infty$, $E[\overline{X}_n] = \infty$ for any sample size n.*

Proof. For any $m = 1, 2, 3, ...$ we define a random variable $g_m(X) = \begin{cases} X & \text{if } X \leq 2^m \\ 0 & \text{if } X > 2^m \end{cases}$.

$P[g_m(X) = 0] = \sum_{k=1}^{\infty} \frac{1}{2^{m+k}} = \frac{\frac{1}{2^{m+1}}}{1-\frac{1}{2}} = \frac{1}{2^m}$. Otherwise, $P[g_m(X) = 2^k] = \frac{1}{2^k}$, $k = 1, 2, 3, ..., m$ for $2^k \leq 2^m$. $E[g_m(X)] = \sum x \cdot P[g_m(X) = x] = 0 + \sum_{k=1}^{m} 2^k \cdot \frac{1}{2^k} = \sum_{k=1}^{m} 1 = m$. Therefore, $m = E[g_m(X)] = E[\overline{g_m(X)}_n] \leq E[\overline{X}_n]$ holds for each given n and for all m = 1,2,3, ... We take the limit as $m \to \infty$ to get $E[\overline{X_n}] = \infty$. Since the expected value of X is infinite, the Central Limit Theorem does not apply to ensure normality of $\overline{X}$. A key condition for the Central Limit Theorem has not been met, and we see this take effect in the probability distribution.

□

This proof is crucial to one's understanding of the paradox. It is clear and intuitive that the way in which the game is constructed causes a violation of the Central Limit Theorem. As previously noted, the assignment of random variables is a defining characteristic for a probabilistic experiment. Indeed, the Central Limit Theorem will not kick in to provide a clear and concise explanation of the St. Petersburg Paradox. The true understanding comes from a fundamental understanding of partitioning the integers into powers of two.

6 Binary and Special Binary Representation

In this project, powers of two are important not only to playing the game, but also to the resulting probabilities and equations for this problem. When examining expected payouts, it becomes evident that the payout must be expressed as the sum of powers of two.

Definition 6.1. *Binary representation is the base two representation of some $a \in \mathbb{Z}$. The result is a string of $0's$ and $1's$.*

Suppose one wished to represent the integer 19 as a binary number. The binary representation can be found by finding the largest power of two that can go into the number, marking that as a 1, then subtracting it from the initial integer. Repeat this process on the remainder of the subtraction until the subtraction yields 0. If there is a power of two that is too large to go into the number, we represent that as a 0 and try the next smaller power of two. By following this process, we find that 10011 is the binary representation for the integer 19 since $19 = 16+2+1 = 1\cdot 2^4+0\cdot 2^3+0\cdot 2^2+1\cdot 2^1+1\cdot 2^0$.

Definition 6.2. *For a positive integer $x = x_1x_2...x_n$, expressed in its binary representation, the weight of x, denoted as $wt(x)$ is the number of digits x_i of x, for which $x_i = 1$, $i = 1, \ldots, n$.*

To illustrate from the previous example, $wt(19) = 3$, since there are three ones in the binary expansion 10011. Simply put, the weight of a number is the number of ones in its binary expansion. This fundamental understanding of binary representation and the weight of a number provide a clear understanding of some of the characteristics of the probability distribution histogram for the St. Petersburg Paradox. With a little bit of thought, one would believe the statement that there are some integers for which there is no binary expansion for some given weight.

One restriction that binary notation has for this project is the fact that the powers of two that compose a given integer are required to be distinct. Simply put, a binary representation of a number is unique. This proves to be a disadvantage for analysis of the St. Petersburg Paradox, as it is allowable for a expected payout to be composed in different ways.

For example, one might wish to obtain $P[3\overline{X_3} = 544]$. The standard binary notation for this number is 1000101010, which has weight 4. To stay true to the desired three plays of the game for this case, we would need the weight to be exactly 3. Indeed, one can find that $544 = 16 + 16 + 512$, thus 544 has been represented by the sum of powers of two and should have an assigned probability, but how can this be represented in binary when the 16's position can only be 0 or 1? A simple answer is to allow the binary representation to account for repeated powers of two, not just 0 or 1 powers of two.

Definition 6.3. *Special binary representation is the base two representation of some $a \in \mathbb{Z}$ where repetition of powers of two is allowed. The result is a string of integers that represent the number of powers of two for that value for each place.*

Definition 6.4. *For an element $x = x_1 x_2 ... x_n[2]$, the special weight of x, denoted as $swt(x)$ is defined to be $\sum_{i=1}^{n} x_i$.*

It was shown above that 544 can be represented with three non-distinct powers of two. A simpler example for the special binary representation is as follows:

The binary expansion for 10 is 1010[2]. It is clear that $wt(10) = 2$, but can 10 be represented for $swt(10) = 3$? This is simply done by"demoting" one of the given ones in the binary expansion. Demoting is defined as removing an individual "1" from one decimal place and replacing it by adding two "1's" to the decimal place to the immediate right. For this example, 10 could be represented as 0210 or 1002. To clarify, $10 = 0 \cdot 8 + 2 \cdot 4 + 1 \cdot 2 + 0 \cdot 1 = 1 \cdot 8 + 0 \cdot 4 + 1 \cdot 2 + 0 \cdot 1$. Both special binary representations have special weight three, thus meeting the requirement previously defined. Because 10 has representations for special weight 2 and special weight 3, there would be nonzero probabilities assigned to the expected payout of 10 for both two plays of the game and three plays of the game.

By continuing to demote ones, a number can be broken down into all of its possible special binary representation. The technique used to find these values is to use a tree structure, called a special binary tree, to keep track of each special representation. Each horizontal row on the tree will give all possible representations for a given weight. Knowing this fact, the understanding of the gaps and spikes that occur on the sample probability distribution histogram becomes much more simple. These tree structures' lack of representation for a certain weight explain when gaps will occur, and the width of the tree provides insight to the jumps in probabilities for certain areas of the distribution.

For example, Figure 5, below, gives the complete special binary tree for the number 10.

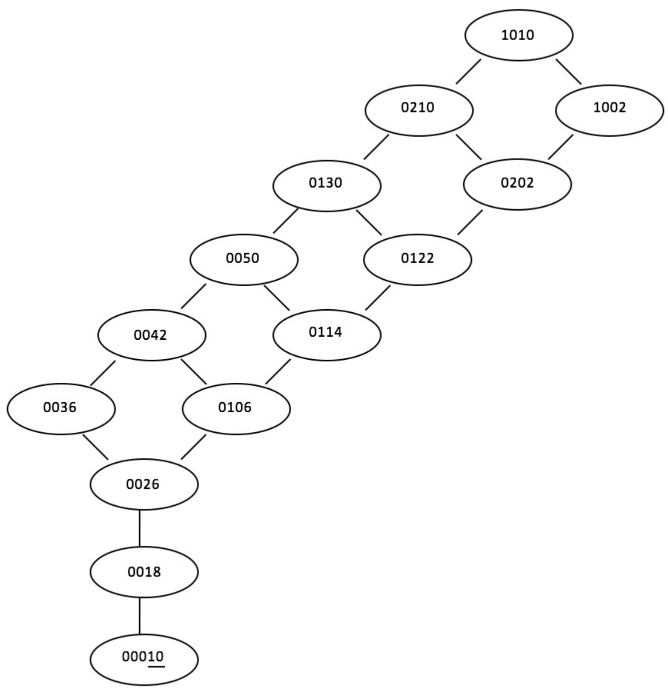


Figure 5: Special Binary Tree for 10

Additionally, see Appendix B for the complete special binary tree for the number 20. Note how quickly this tree begins to widen and how deep it is. Clearly, the larger a number is, the more possible representations in special binary there are.

Theorem 6.5. *Given k coin tosses and a (sufficiently large) n, n can be written as k powers of 2 exactly when there exists a binary expansion for which the $swt(n) = k$.*

Proof. *Because each coin toss represents a distinct power of two, when a value is obtained, it marks that value "on" for the binary expansion. When each power of two* *has been represented, the binary expansion for an integer has been obtained, specifically with weight k.*

This theorem provides a substantial insight into the probability distribution for the St. Petersburg Paradox. For all $x > k$, there is no special binary expansion for a number with $swt(x)$. This follow from the fact that there are no more special binary representations beyond that for which $swt(x) = x$. This explains zeros for probabilities in the distrubuiton. There are integers that cannot be written as the sum of non-distinct powers of two and this tree shows which ones those are. Generalizing this method for creating trees—creating a computer program to generate all of them for some upper bound—would provide the ability to know where gaps and where spikes will occur for a given number of plays of the game.

Theorem 6.6. *The width of a special binary tree will show the given number for that tree will cause a "jump" on the probability distribution. Additionally, if a number has no special binary representation for a given weight (i.e. that number does not appear on its special binary tree), then it will have a zero probability on the probability distribution for that number of plays of the game.*

Proof. *The number of nodes on each horizontal level of the special binary tree represent distinct possibilities of possible payouts for a given special weight. The more nodes on a given horizontal level will result in a higher probability of that value arising. If a number has no special binary representation, it is impossible for that value to be obtained for that given weight, which implies the probability of that value occurring is zero.*

It is important to note that this notation gives an extra term that should not be considered for examining the St. Petersburg Paradox. The rightmost digit of the special binary representation, by this construction, can be thought of as the coefficient on a payout of $\$2^0 = \$1$. It has been discussed that this is not a possible payout, so any special binary notation that has a value greater than zero in the rightmost position should be discarded as uninteresting to the application of this representation to the St. Petersburg Paradox game.

For the special binary tree for the number 10 (refer again to Figure 5), one can see that the last five rows (corresponding to $n = 6, 7, 8, 9$, and 10 plays of the game) do not have any element in the row that have a nonzero rightmost digit, and hence they do not apply. Rows $n = 2, 3, 4$, and 5 each have at least one element with a nonzero entry in the last row, and so 10 will be an average payout value for $n\overline{X_n}$ with positive probability for each of these four n values, and only these four. The example of the special binary tree for the number 20 found in Appendix B gives a more powerful insight as to how significantly different a representation can be for a weight.

This project concludes with the above theorem regarding special binary trees, but leaves room for much more investigation. One could easily generate a computer program to develop these trees. Being able to quickly evaluate these trees would provide a truly powerful understanding of how the probabilities for a given number occur.

7 Conclusion

This project has approached the St. Petersburg Paradox in many different ways: through computer simulation, using probability theory, and using number theory. Clearly this is a problem that has many connections to various sub-fields of mathematics.

While the algorithm itself that was used to simulate the game is not interesting from a computer science standpoint, it proved to be a powerful tool in understanding the problem. The sample probability histograms created by this simulation were indeed what sparked this project. As the original mathematicians who created them noted, it "demanded explanation" [2].

By explicitly calculating probabilities, this project began to pick up steam, finding simple trends that seemed to appear in the investigation of $n = 2$ and 3 plays of the game. These equations developed nicely into theorems with succinct proofs that the probability distributions displayed. Additionally, a strong relation to the Negative Binomial Random Variable was observed, which is one of the most important distributions from Probability Theory. This result was not unexpected, as flipping a coin is the most simple example that is almost always used to illustrate this concept when it is taught in a course.

It is no surprise that there is an inherent connection to powers of two in how the probability distribution is developed. One might wish to relate the expected payouts to a binary representation, but the key insight for this thesis was the fundamental understanding that repetitions of each power of two must be allowed. This led to developing "special binary notation," which does just that. This representation allows for a quick calculation of how to many possible ways an integer can be represented as the sum of non-distinct powers of two.

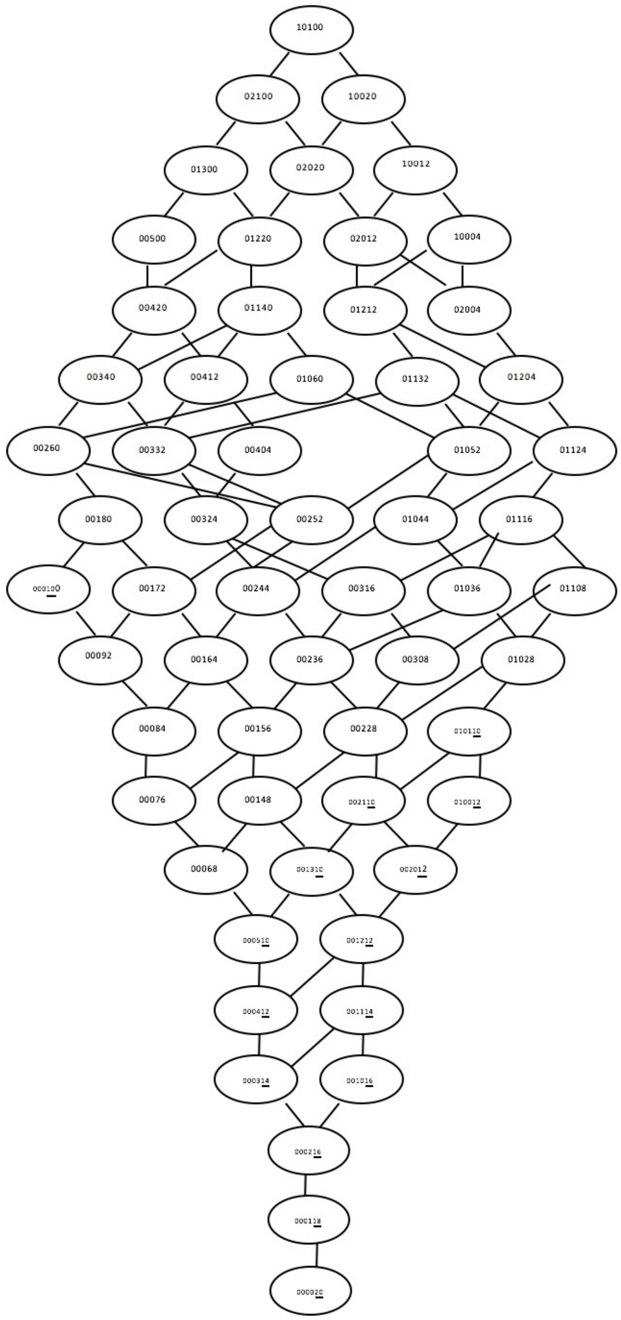
There is a huge amount of potential for continued work in this area. This thesis has only begun to scratch the surface of the probabilistic and number theoretic connections to the St. Petersburg Paradox. One might wish to continue investigating special binary trees, which might lead to insights not only to this subject, but also to the partition function. Additionally, more theorems could be developed with the intention of explicitly describing the probability distribution for this probability experiment with $n = 4, 5, \ldots$. Clearly, the St. Petersburg Paradox could continue to provide fruitful mathematical study, providing many more results for probability, computer science, and number theory.

8 Appendix A

```
#include <iostream>
#include <cstdlib>
#include <math.h>
#include <fstream>

using namespace std;

int cointoss();
int buffonexp();

int main () {
srand(time(0));
int n = 1000000;
int a[n];
ofstream data("data.txt");
for (int x = 0; x < n; x++) {
a[x] = buffonexp();
data << a[x] << '\n';
}
double sum = 0;
for (int x = 0; x < n; x++) {
sum += a[x];
}
data.close();
double avg = sum/(double)n;
cout << avg << endl;



return 0;
```

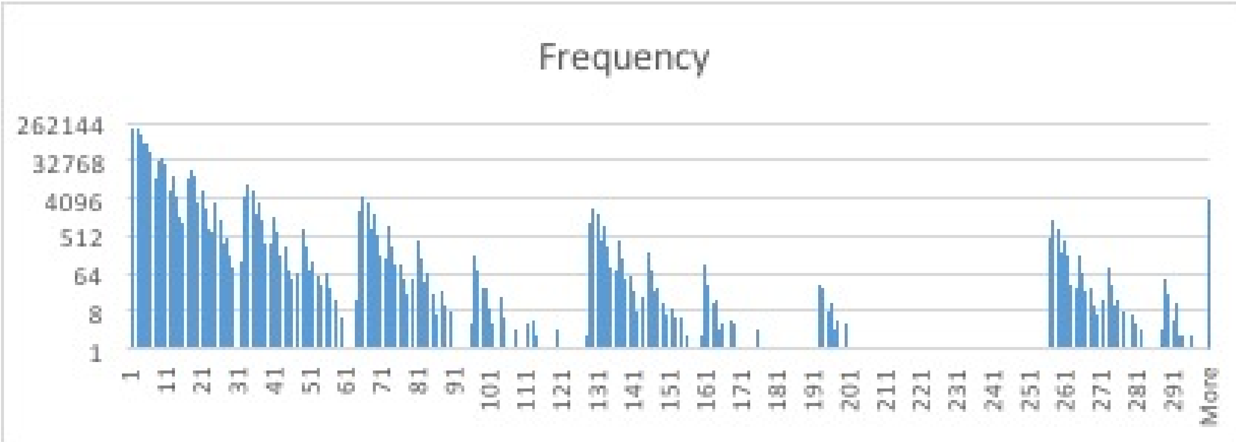
```
}

int buffonexp() {

int n = 2;
int a[n];
for (int x = 0; x < n; x++) {
a[x] = cointoss();
}
double sum = 0;
for (int x = 0; x < n; x++) {
sum += a[x];
}
double avg = sum/n;
return avg;
}

int cointoss() {
int coin = 0;
int x = 0;
while (coin == 0) {
coin = rand() % 2;
x++;
if (coin == 1)
break;
}
int winnings = pow(2,x);
return winnings;
}
```

9 Appendix B

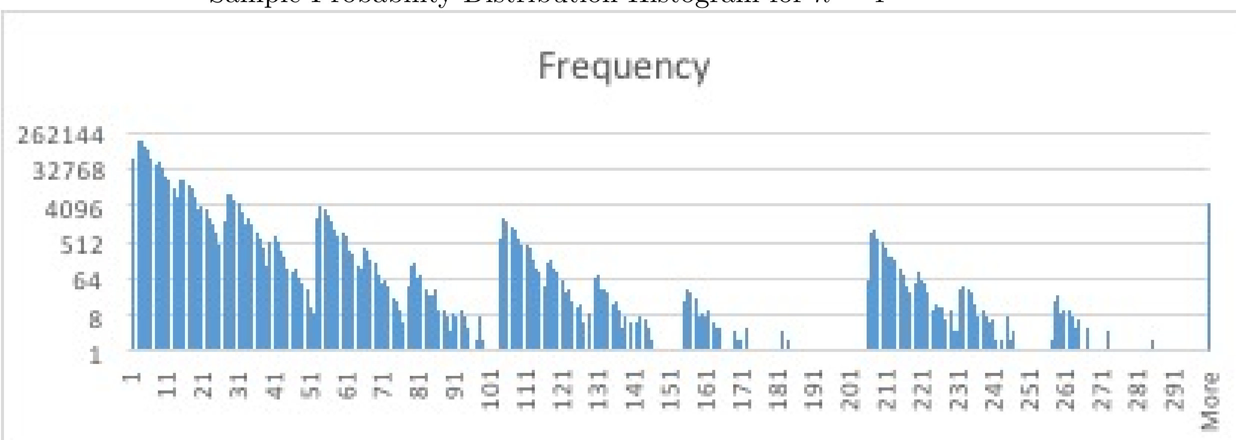


Special Binary Tree for 20

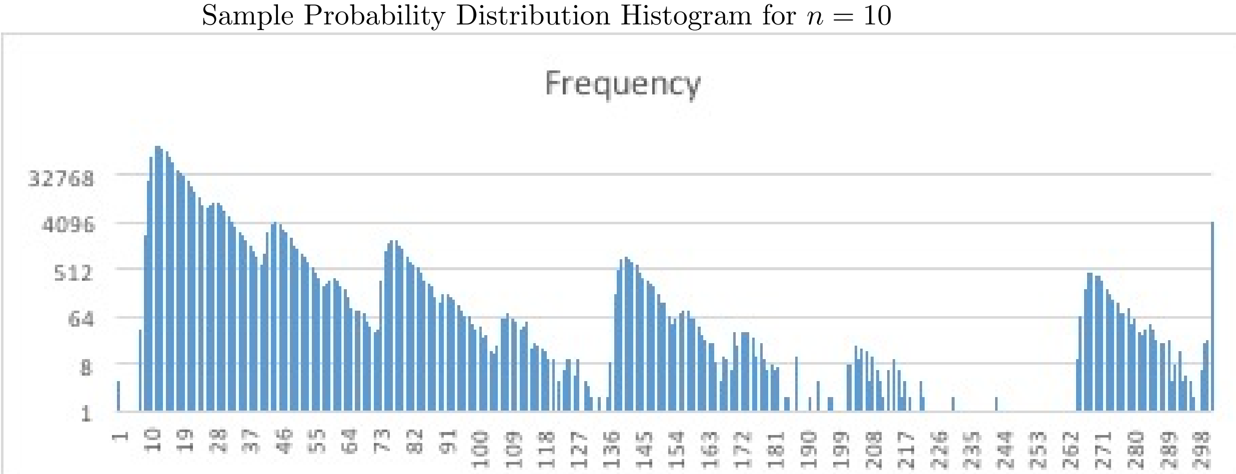
10 Appendix C



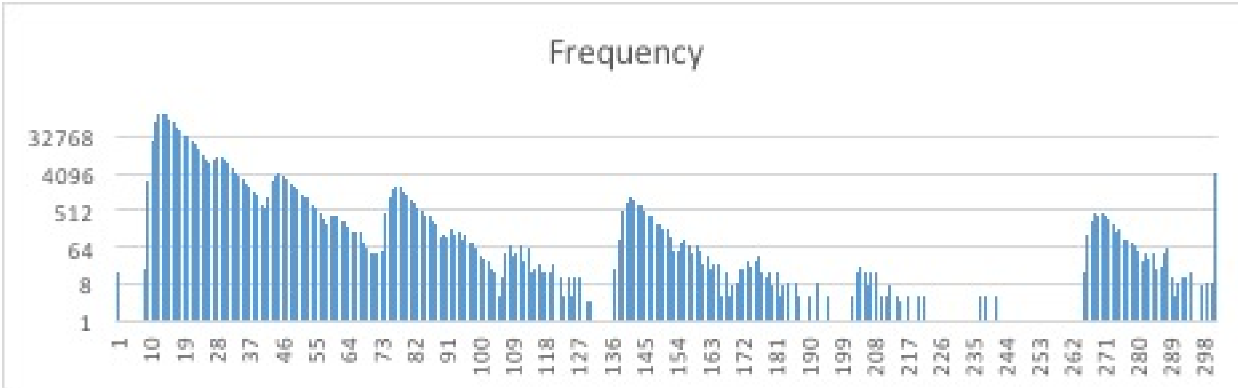
Sample Probability Distribution Histogram for $n = 4$



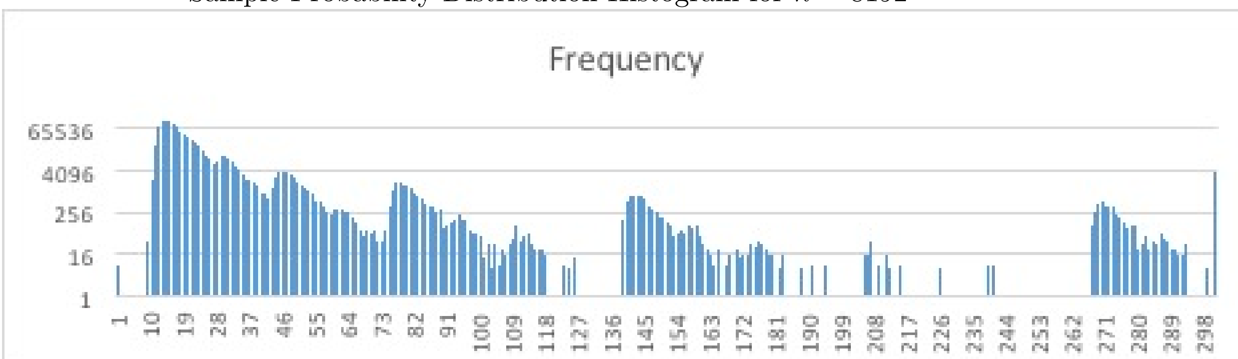
Sample Probability Distribution Histogram for $n = 10$



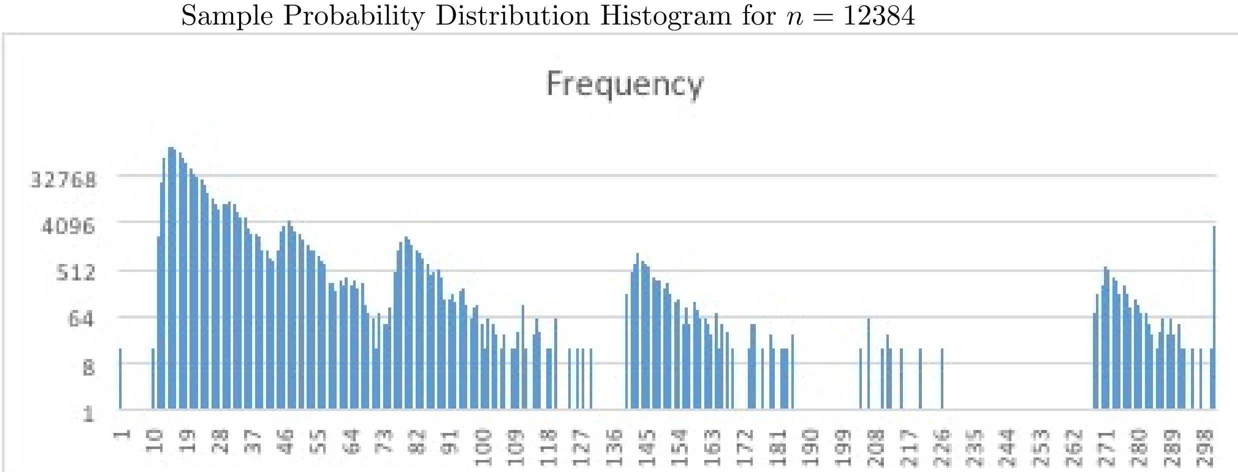
Sample Probability Distribution Histogram for $n = 4092$



Sample Probability Distribution Histogram for $n = 8192$



Sample Probability Distribution Histogram for $n = 12384$



Sample Probability Distribution Histogram for $n = 32769$

11 References

References

[1] Harold Dehling, Daniel Bernoulli and the St. Petersburg Paradox, *Nieuw Archief Voor Wiskunde,*. **15** (1997) 223-227.

[2] Dominic Klyve and Anna Lauren, An Empirical Approach to the St. Petersburg Paradox. *The College Math Journal*. **42** (2011) 260-264.

[3] David Lane, Introduction to Statistics – Chapter 5. (2013) http://onlinestatbook.com/2/probability/binomial.html

[4] Martin, Robert, "The St. Petersburg Paradox", The Stanford Encyclopedia of Philosophy (Summer 2014 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/sum2014/entries/paradox-stpetersburg/.

[5] Negative Binomail Distribution, *Exploring Data* (2016) http://stattrek.com/probability-distributions/negative-binomial.aspx