

A New Look at the St. Petersburg Paradox

Eden Foley*

Faculty advisor: Dhanuja Kasturiratna[†]

Department of Mathematics and Statistics,

Northern Kentucky University,

Highland Heights, KY 41009, USA

Abstract

The infinite expected value of the St. Petersburg Paradox has been a source of contention within probability theory since its inception in the early 18th century. This work has aimed to avoid supposition and instead has chosen to focus on empirical evidence generated through simulation. Armed with sufficient evidence, this work has modeled the sampling distribution of the St. Petersburg Paradox's mean. This model allows a prospective gambler or casino owner to know whether to partake in the game at a given price. In addition, the resulting model has been discovered to be highly adaptable to other similar distributions. Potential applications of this work to earthquake magnitudes is also discussed.

AMS 2010 subject classification: Primary 62F40; Secondary 62E17

Key words and phrases: Empirical model, Expected value, Law of large numbers, St. Petersburg paradox.

1 Introduction

Since its inception, probability theory has provided key insight for those interested in quantifying risk. It has allowed statisticians, actuaries and others to collect statistics in the real world and then make useful inferences regarding the underlying distribution being sampled. However, there are cases in which a distribution may have such a "heavy tail" that using traditional statistics without caution to quantify the risks is highly inadvisable. Unlike cases such as mortality or human height which have a very limited range of values they can take on, some distributions outcomes can take on a wide range of values in which unlikely events can have a substantial effect on the long-term average. The St. Petersburg Paradox is a famous

*foleye1@nku.edu

[†]kasturiratna1@nku.edu

probability paradox discussed originally in a series of letters in 1713 by Nicholas Bernoulli [1],[2]. In it, the gambler flips a coin until he receives his first head. The distribution of how many flips it will take follows a geometric distribution with probability of one half. The paradox is created due to the fact that the person offering this game to the gambler will have to pay out two dollars for the first flip and then twice as much for each successive flip required. This is a clearly defined distribution, but when attempting to calculate the expected value, the expectation does not converge to a finite value. By multiplying every outcome by its corresponding probability, the expected value appears to be infinite.

$$2 \left(\frac{1}{2}\right) + 4 \left(\frac{1}{4}\right) + 8 \left(\frac{1}{8}\right) + \dots + 2^n \left(\frac{1}{2^n}\right) + \dots = 1 + 1 + 1 + \dots + 1 + \dots = \infty$$

The paradox arises from the discrepancy between the expected value, often thought of as the "fair" value for a game in probability theory and the one experienced in practice. Clearly this game is not worth an infinite amount to the gambler since he is unlikely to receive more than a modest gain and certainly a finite one. Much work has been done to demonstrate the flawed conclusion of the St. Petersburg Paradox being worth an infinite amount. In practice, the expected value can be used in almost all circumstances as a stand in for the break even point of a distribution. However, by definition the expected value is only the value the sample mean converges to as the sample size goes to infinity. Due to the Law of Large Numbers, these two views of expected value rarely conflict as most of the time sample means very quickly converge to their expected value. In the cases where large events play a significant role, the Law of Large Numbers cannot be relied upon to converge the sample mean to its expected value. In these situations, decision making must be done on the basis of an individual's comfort with risk and the sampling distribution of the mean over the number of trials they are interested in attempting.

Since the origins of this paradox with Nicholas Bernoulli, [2], the St. Petersburg Paradox and other probability distributions whose expectation is a diverging series have attracted attention from academia. Work done in the past mainly has focused on justifying the discrepancy between the mathematical expected value and the real life value the game would have to a prospective gambler, [1], [3], [7], [4]. Justification mainly follows two different paths. The first is applying a utility transformation to recognize that there is a diminishing return to money, which has the effect of making the adjusted expected value finite. The other approach is to set some arbitrary high value as the cutoff point in which one disregards all values higher than this as a possible outcome and then adjust the distribution accordingly.

There have been too many proposed utility models to mention them individually, but the first began as a possible solution from Daniel Bernoulli, [1], who suggested that the "moral value" of $\log x$ to a gain of x .

$$\frac{1}{2} \log(2) + \frac{1}{4} \log(4) + \frac{1}{8} \log(8) + \dots + \frac{1}{2^n} \log(2^n) + \dots$$

$$\sum_{n=1}^{\infty} \frac{1}{2^n} \log(2^n) = \sum_{n=1}^{\infty} \frac{1}{2^n} n \log(2) = \log(2) \sum_{n=1}^{\infty} n \frac{1}{2^n} = 2 \log(2)$$

While this explanation does make the valid point that the marginal value of a dollar diminishes as one accumulates more of them in accordance with the law of diminishing returns, it and other arguments made along similar lines do not answer why the absolute dollar expected value of the original problem is not infinite. Instead they answer the question "Why is the St. Petersburg Paradox not worth an infinite amount to a prospective gambler?"

The other most common argument is one which sets a cutoff value for the payoff for various reasons. One reason might be that there is a maximum that any casino or institution would be willing or able to pay, so all payoffs above this amount must be discarded. The other cause is one of chance. There are events so unlikely that we could live a million lifetimes and still have a very small probability of having them occur. These types of events, as the reasoning goes, should not be included in the decision making process of the prospective gambler. To make this argument clear, let us assume that we do not suspect any institution could afford a payout of 2^{50} or about 1,126 trillion dollars and that we are not likely to get forty-nine tails before our first head in our lifetime. Ignoring the insignificant adjustment for the possibilities we are removing from the distribution on the remaining probabilities, the expectation is remarkably different.

$$\frac{1}{2}(2) + \frac{1}{4}(4) + \frac{1}{8}(8) + \dots + \frac{1}{2^{50}}(2^{50}) = 50$$

Both the arguments above, along with countless other arguments proposed in academia in the past three centuries have "solved" the St. Petersburg Paradox by demonstrating why it is finite. Further, they provide what they believe to be a "fair" value for the game. The problem is that almost every one of the many solutions to the paradoxes proposed in the past result in inconsistent values which they deem as a fair offering price for the game. This work uses simulation to study the actual sampling distribution of gains. While some previous simulation studies have been done, these works have provided scattered conclusions with many questions unanswered. The most sophisticated simulation study of the St. Petersburg Paradox this author has read comes from Ricardo Rodriguez's, [10]. In his abstract, he writes "Even more surprising is the absence of an extensive simulation study of finite St. Petersburg sequences from the literature."

Other than the work described above, simulation studies on the St. Petersburg Paradox have followed two paths. One approach has been to study a single sample size's distribution, [6]. The other main approach has been to simulate time series of the sample average over a large number of trials and observe the behavior of the sample mean, [4], [9]. These two approaches have generated two useful observations regarding the St. Petersburg Paradox. The first is that the mean one arrives at even after a large sample is not enough to predict with any reasonable accuracy the mean one will get if one repeats the experiment again. The second observation is that the sample mean tends to grow logarithmically.

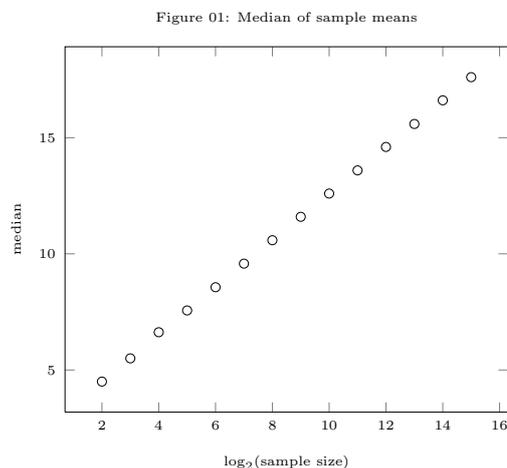
In Rodriguez's simulation study, a more complete work is done. Rodriguez presents in his work the distribution of the sample mean for several sample sizes, which provides a much more complete picture of the St. Petersburg Paradox. The primary result he observes is that increasing the sample size results in a horizontal shift of the distribution of the sample mean and that this distribution seems to retain a similar shape, [10].

These past studies have significant limitations. No study it seems as of yet has tackled the issue of modeling the actual distribution of sample means itself. Providing mainly qualitative observations, previous simulation work has done little to move the St. Petersburg Paradox from being largely a philosophical question into a tangible one. Given the many highly skewed distributions in the real world now studied predominantly through the lens of fractal geometry rather than probability theory, being capable of mathematically modeling the most prominent diverging probability distribution seems worthwhile and is the primary focus of this work.

In section 2 the sampling distribution of the mean is modeled for any sample size. In section 3 this model is tested and found to accurately model the underlying distribution. In section 4, it is discovered that this model can be extended to other distributions with similar characteristics through a simple transformation. Finally, in section 5 the potential applicability to the real world is demonstrated through the successful modeling of earthquake magnitudes. In all, this work manages to go above its original ambitions of modeling the St. Petersburg Paradox and has managed to model an entire class of distributions, including some which occur in practice.

2 Creating an Empirical Model

The dataset used in this work consisted of a million sample means for sample sizes of $2, 4, 8, 16, \dots, 2^{15}$ of the St. Petersburg Paradox. For example, to create the sampling distribution of the mean for 8 trials, you generate 8 trials of the St. Petersburg Paradox and take their average. Then you repeat this process a million times and use the data generated to study the distribution of the sample mean of 8 trials.



As a start to studying the sampling distribution of means, it is first important to find some measure of central tendency. Since the sampling distribution of the mean for the St. Peters-

burg Paradox is extremely skewed, the median is used as the measure of central tendency in this study.

The median of the distribution of sample mean for each sample size is calculated and plotted in Figure 01, which shows the median of the sampling distribution of the mean grows linearly as the sample size grows exponentially. The fact that the median grows as the sample size increases is a result of the inclusion of more unlikely events occurring. This is only the fiftieth percentile of the distribution. To have a useful model of the distribution, a model must be developed for all other percentiles. Because the median growth is linear, the distribution can be modeled in relation to the median, a potentially simpler task. Since standard deviation and other traditional measures of central tendency are not useful in studying this distribution, some other measure is necessary to study the distribution. After experimentation the relationship chosen to be used in this work is a logarithmic one. Transforming percentiles by $\log_{\text{median}}(\text{mean gain})$ provides a way to model the percentiles of the distribution for different sample sizes.

Figure 02: Percentiles of the distribution of mean gain for 2^4 trials

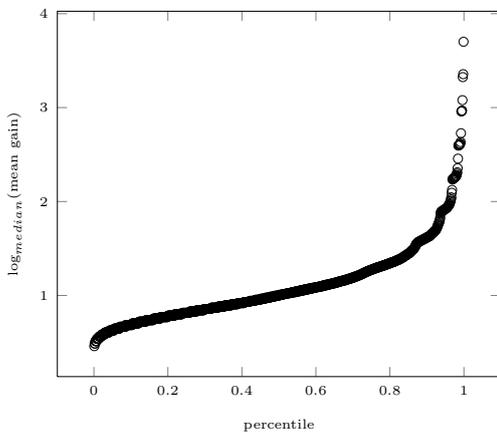
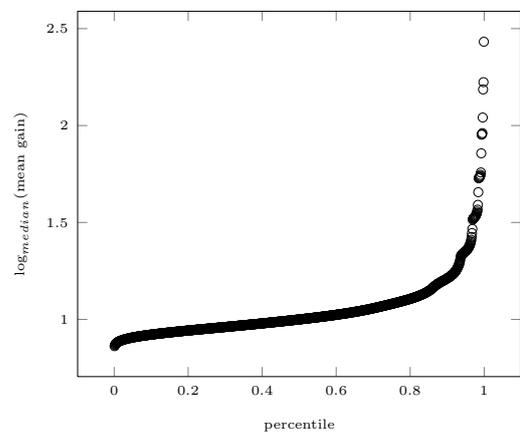


Figure 03: Percentiles of the distribution of mean gain for 2^{15} trials



We observe in Figure 02 and Figure 03 that there is a strong similarity between the plots of the sampling distribution of the mean gain for different sample sizes. The first observation is that with apparent asymptotes at 0 and 1, the percentiles seem to be a version of the logit function, the inverse of the famous logistic function. With the percentiles transformed by $\log\left(\frac{p}{1-p}\right)$, where p is the percentile, the function becomes much more clear (Figure 04).

Figure 04: Transformed version of Figure 02

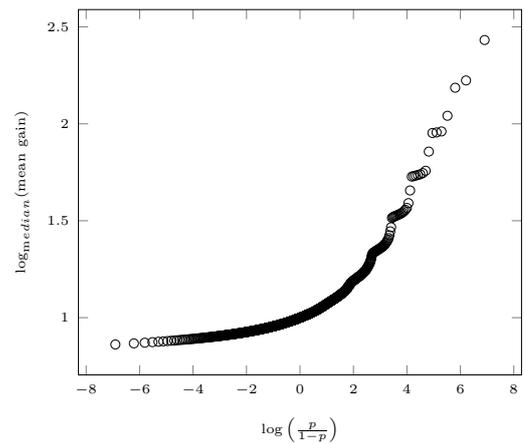


Figure 05: Growth in parameter a

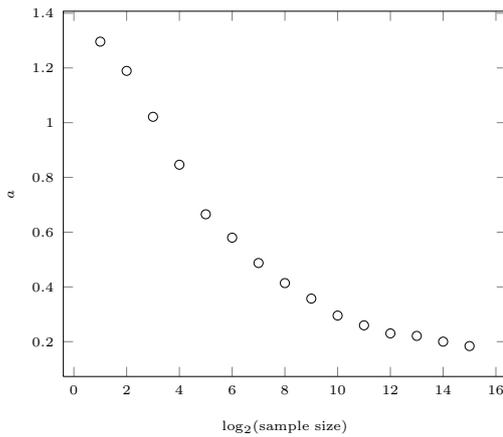
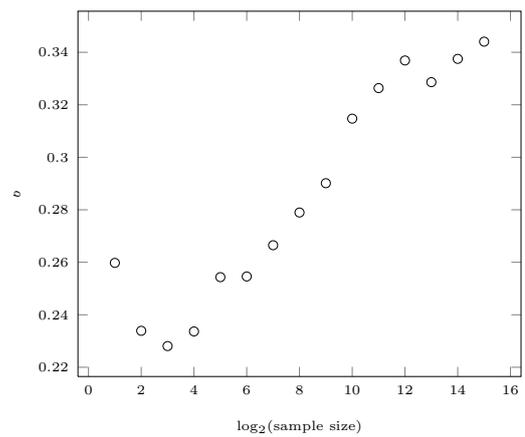


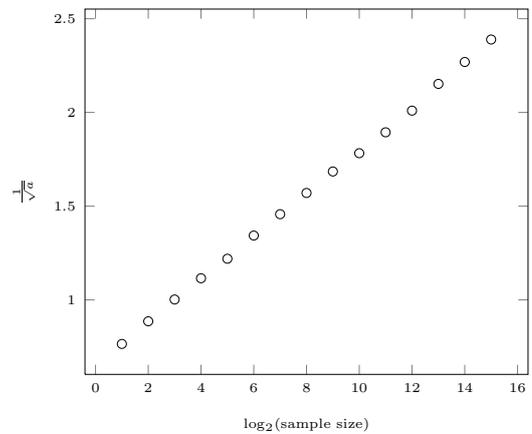
Figure 06: Growth in parameter b



The same observations made for the distribution of sample means for 2^4 trials can be made for all other sample sizes studied. Therefore, every sample size can be modeled using nonlinear regression with an exponential function in the form $ae^{bx} + c$. Further, we observed that the coefficient c was very close to $1 - a$ for every sample size tested. Thus, the model was conveniently simplified to only having two parameters in $a(e^{bx} - 1) + 1$.

Next, we created a model for a and b with respect to the sample size. Above in Figure 05 and Figure 06 are the parameters a and b in $a(e^{bx} - 1) + 1$ for each sampling distribution. As can be seen, the plot of the b parameter is much more volatile than a . However, it does have a strong linear relationship with the sample size. After performing linear regression on the b parameter, it allows for a more precise relationship for a by repeating the nonlinear regression with b fixed in the form $a(e^{\hat{b}x} - 1) + 1$, where \hat{b} represents the predicted value of b for each respective sample size. Through observation, it is discovered that a can be modeled linearly after replacing a by the reciprocal of the square root of a , (Figure 07).

Figure 07: Linearize parameter a



At this point we have an acceptable model for the sampling distribution of the means for any sample size. The function for the sampling distribution of the means is below, where n is the $\log_2(\text{sample size})$ and p is the percentile between 0 and 1.

$$f(n, p) = (2.53 + n) \left[\frac{1}{(.65 + 0.115n)^2} \left(\frac{p}{1-p} \right)^{0.2 + 0.01n} + 1 - \frac{1}{(.65 + 0.115n)^2} \right]$$

For example, we can take a look at the famous case of 2048 trials by Compte de Buffon in which he arrived at \$9.82 as the mean gain from his experiment,[3]. According to the model that we have found, that value would be only the 14th percentile of the distribution. If a casino had done a similar experiment to Compte de Buffon and set the entry fee at \$9.82 then there would be an 86% chance of loss. The 86th percentile, an event of similar likelihood on the higher end of the distribution, would be \$28.83.

For a more meaningful case, let us think of a gambler who plans on playing the St. Petersburg Paradox 1024 times and wants to be 95% sure of not losing money. The highest price this gambler would accept as entry fee would be $f(10, 0.05) = \$7.93$.

3 Determining Goodness-of-fit

Comparing the actual percentile values to the ones predicted is not particularly useful for determining goodness of fit because it would result in inevitable illusory increases in the error for higher percentiles, which grow in magnitude very quickly, and for larger sample sizes which tend to also have larger values being studied. Since the model is built to fit the power one must raise the median to get the percentile, comparing the power one must exponentiate median to in order to get the percentile to the one predicted by the model would be a better test of how well the model performs. The following table shows the difference between the predicted log transformed mean gain and those observed in simulation of 10, 000 trials of each sample size independently of the data used to create the model itself. For testing purposes, generating sample sizes of 10, 000 have seemed to consistently to represent a pretty precise picture of the underlying distribution itself, and is assumed to be large enough to test the quality of fit of the model to the data.

Table 01: Difference between observed and predicted percentiles of mean gain

	Percentile										
	.20	.40	.50	.60	.65	.70	.75	.80	.85	.90	.95
2^3	0.0228	0.0087	0.0016	0.0198	0.0041	0.0153	0.0530	0.0435	0.0399	0.0684	0.0949
2^5	0.0171	0.0012	0.0031	0.0024	0.0017	0.0045	0.0159	0.0096	0.0265	0.0354	0.0699
2^7	0.0135	0.0044	0.0001	0.0051	0.0065	0.0013	0.0035	0.0008	0.0111	0.0202	0.0594
2^9	0.0141	0.0045	0.0003	0.0037	0.0043	0.0027	0.0007	0.0005	0.0049	0.0169	0.0473
2^{11}	0.0141	0.0037	0.0008	0.0039	0.0052	0.0001	0.0016	0.0002	0.0110	0.0096	0.0368
2^{13}	0.0135	0.0031	0.0003	0.0024	0.0040	0.0018	0.0005	0.0015	0.0036	0.0092	0.0416
2^{15}	0.0123	0.0034	0.0002	0.0032	0.0037	0.0018	0.0034	0.0048	0.0007	0.0014	0.0298

Based on the results in Table 01, the model seems to be a good match of the actual percentiles observed in simulation. This model seems to be by far the most accurate predictor of results of the St. Petersburg Paradox created to date.

4 Generalizing Model to Other Similar Distributions

For applications beyond the St. Petersburg Paradox, one would ideally be able to handle other distributions in which the median of their sampling distribution of the mean grows linearly as their sample size grows exponentially. Through experimentation it can be shown that any geometric or exponential random variable where every outcome is raised to power so that the expectation can be written as $\sum_{i=1}^{\infty} c$ or $\int_0^{\infty} cdx$, where c is any constant, will have its median grow linearly as its sample size increases exponentially. For example, if you take an exponential variable with mean $E(x) = 1$ and then transform that variable by e^x , then the expectation is $\int_0^{\infty} e^x e^{-x} dx = \int_0^{\infty} dx = \infty$. A case with a different geometric variable would be one with mean $E(y) = 3$ that is transformed by 3^y which creates the expectation $\sum_{i=1}^{\infty} \frac{2}{3} \times 3^i \times 3^{-i} = \sum_{i=1}^{\infty} \frac{2}{3} = \infty$.

Using various continuous and discrete cases of distributions similar to the St. Petersburg Paradox with different median lines, it was observed that any two distributions with the same median growth will have very similar sampling distributions of their means. This allows for the modeling of any distribution with a linear median growth as the sample size increases exponentially to be modeled by the St. Petersburg Paradox to change its median line of $y = x + 2.53$ (the base) to match the distribution you are interested in. Below are the errors similar to the previous error table the discrete and continuous versions of the St. Petersburg Paradox described above. The error tables are generated by comparing the actual percentiles of 10,000 trials of each sample size and the St. Petersburg Paradox adjusted to match the other's median line.

Table 02: Difference between observed and predicted percentiles for geometric distribution

	Percentile										
	.20	.40	.50	.60	.65	.70	.75	.80	.85	.90	.95
2^3	0.0157	0.0008	0.0199	0.0044	0.0092	0.0142	0.0783	0.0270	0.0008	0.0568	0.0099
2^5	0.0094	0.0028	0.0086	0.0005	0.0136	0.0169	0.0090	0.0019	0.0327	0.0672	0.0133
2^7	0.0117	0.0020	0.0029	0.0022	0.0000	0.0054	0.0082	0.0064	0.0426	0.0113	0.1114
2^9	0.0141	0.0054	0.0023	0.0037	0.0068	0.0061	0.0046	0.0162	0.0044	0.0002	0.0452
2^{11}	0.0141	0.0044	0.0015	0.0028	0.0007	0.0035	0.0011	0.0037	0.0086	0.0575	0.0012
2^{13}	0.0110	0.0011	0.0026	0.0026	0.0019	0.0055	0.0088	0.0120	0.0046	0.0163	0.0234
2^{15}	0.0105	0.0046	0.0021	0.0005	0.0031	0.0051	0.0060	0.0020	0.0106	0.0053	0.0621

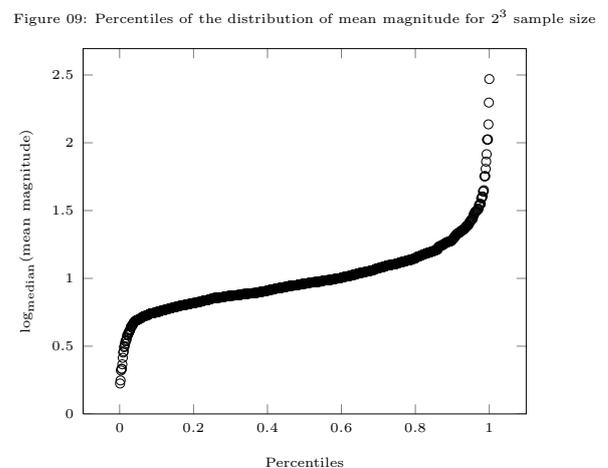
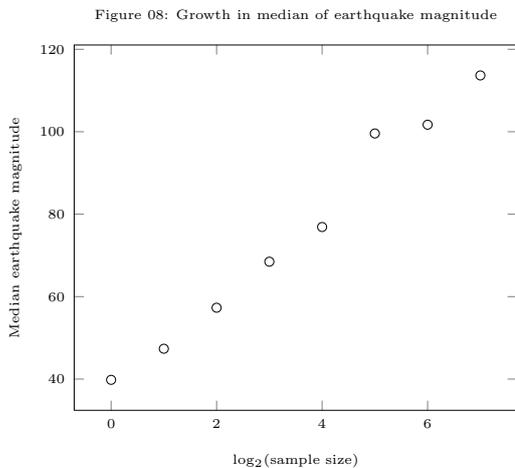
The errors are calculated identically to how they were before with the St. Petersburg error table. The values are the difference between the observed exponents from the median and those predicted by the model created for the St. Petersburg Paradox after adjusting it to have an identical median line. As can be seen, the errors generated are competitive with the errors seen before with the St. Petersburg Paradox itself, which indicates that the St. Petersburg Paradox can be used to model other distributions with a similar growth in the median of their sampling distribution of the mean.

Table 03: Difference between observed and predicted percentiles for exponential distribution

	Percentile										
	.20	.40	.50	.60	.65	.70	.75	.80	.85	.90	.95
2^3	0.0207	0.0028	0.0030	0.0182	0.0215	0.0354	0.0459	0.0532	0.0839	0.1117	0.1826
2^5	0.0179	0.0072	0.0032	0.0007	0.0020	0.0061	0.0083	0.0134	0.0225	0.0423	0.0735
2^7	0.0155	0.0043	0.0010	0.0014	0.0001	0.0036	0.0101	0.0111	0.0181	0.0324	0.0596
2^9	0.0145	0.0030	0.0000	0.0011	0.0021	0.0020	0.0032	0.0004	0.0114	0.0164	0.0354
2^{11}	0.0134	0.0039	0.0004	0.0033	0.0052	0.0040	0.0014	0.0004	0.0076	0.0202	0.0598
2^{13}	0.0145	0.0046	0.0005	0.0051	0.0067	0.0067	0.0067	0.0035	0.0009	0.0061	0.0461
2^{15}	0.0131	0.0035	0.0018	0.0024	0.0039	0.0051	0.0043	0.0028	0.0016	0.0034	0.0249

5 Applying Model to Earthquake Data

Earthquake magnitudes seemed an ideal candidate for applying the new St. Petersburg Paradox model to the real world. Extensive and accurate seismological records allow for a meaningful look at the distribution of larger sampling distributions. As can be seen below in Figure 08, the median of average magnitude grows in a linear way as the sample size increases exponentially. The data used in this analysis come from the University of St. Louis seismological data which is available on their website. Also, it is important to note that the data below is not on the Richter scale, but rather on an absolute scale of magnitude where an earthquake of magnitude 1 is the base unit.



As can be seen above in Figure 09, the percentiles of the earthquake magnitudes follow a similar shape to the percentiles of the St. Petersburg Paradox when transformed by taking the log base median in the same way as the St. Petersburg Paradox, which supports there being a connection to the St. Petersburg Paradox. Below are the actual percentiles observed in the earthquake data for the average magnitudes and those predicted by the St. Petersburg Paradox model by adjusting the model to have the median line match those observed in the earthquake data.

Table 04: Observed and predicted percentiles for earthquake data

		Percentile								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
2^3	observed	1.49	1.58	1.68	1.73	1.83	1.91	2.02	2.13	2.37
	predicted	1.67	1.73	1.78	1.83	1.88	1.94	2.02	2.12	2.31
2^4	observed	1.59	1.68	1.76	1.81	1.87	1.95	2.06	2.18	2.34
	predicted	1.72	1.78	1.83	1.88	1.93	1.98	2.04	2.14	2.28
2^5	observed	1.64	1.77	1.83	1.88	1.99	2.06	2.09	2.18	2.34
	predicted	1.78	1.84	1.88	1.93	1.97	2.02	2.08	2.17	2.30
2^6	observed	1.68	1.80	1.93	1.98	2.00	2.02	2.11	2.16	2.52
	predicted	1.83	1.89	1.93	1.97	2.01	2.06	2.12	2.20	2.32

6 Conclusion

The St. Petersburg Paradox has been studied for hundreds of years under a qualitative lens. The aim of this research has been to put discussion of this topic on a more firm footing based on the actual distribution that occurs in simulation. This work has produced a mathematical model for the distribution of the sample mean for any sample size which is a unique achievement among all previous work involving the St. Petersburg Paradox. Also, the applicability of this model has been demonstrated far beyond the St. Petersburg Paradox itself, which of course one is very unlikely to observe in the real world. Through only a simple transformation, this distribution can be applied to any distribution of sample means from a power series distribution similar to the St. Petersburg Paradox. The extension of this model to other distributions with a similar median growth in simulation and with earthquakes in real life provides a means of finally providing probability theory a way to handle distributions with extremely heavy tails in which the past average may be unrepresentative of the true expected value. While this model only applies to a class of extremely heavy tailed distributions whose median lines grow in a linear fashion as the sample size increases exponentially, this research could be continued to expand other types of scaling distributions whose median growth is not the same. With this approach fully explored, decision makers faced with distributions with extremely heavy tails may for the first time be able quantify the uncertainty they face.

References

- [1] Bernoulli, D.(1738) Specimen theoriae novae de mensura sortis. Commentarii Academiae Scientiarum Imperialis Petropolitanae. *Papers Imp. Acad. Sci. St. Petersburg*, 5, 23-36.

- [2] Bernoulli, N. (1713) correspondence. In "Correspondence of Nicholas Bernoulli concerning the St. Petersburg Game", available in English translation by Richard Pulskamp at <http://www.cs.xu.edu/math/Source/Montmort/stpetersburg.pdf>
- [3] Buffon, G.(1777) Essai darithmetique morale. *Supplement a' l'Histoire naturelle*, 4, 46-148.
- [4] Ceasar, A. (1984) A Monte Carlo simulation related to the St. Petersburg Paradox. *The College Math. J.*, 15, 339-342.
- [5] Dehling, H. G., (1997) Daniel Bernoulli and the St. Petersburg Paradox. *Vierde serie Deel*, 15, 3, 223-227.
- [6] Dutka, J., (1988) On the St. Petersburg Paradox. *Archive for History of Exact Sciences*, 39, 1, 13-39.
- [7] Euler, L. (1862) Vera aestimatio sortis in ludis. *Opera posthuma*, 1, 315-318.
- [8] Hayden, B., Platt, M. (2009) The mean, the median, and the St. Petersburg paradox. *Judgment and Decision Making*, 4, 256-272.
- [9] Liebovitch, L. S., Scheurle, D., (2000) Two lessons from fractals and chaos. *Complexity*, 5, 4, 34-43.
- [10] Rodriguez, R. J., (2006) Finite sequences of St. Petersburg games: Inferences from a simulation study. *Journal of Statistical Computation and Simulation*, 76, 10, 925-933.
- [11] Szekely, G.J., Richards, D. P. (2004) The St. Petersburg paradox and the crash of high-tech stocks in 2000. *The American Statistician* , 58, 3, 225-231.