Article

Entrance Fees and a Bayesian Approach to the St. Petersburg Paradox

Diego Marcondes †, Cláudia Peixoto *,†, Kdson Souza † and Sergio Wechsler †

Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, 05508-090 São Paulo, Brazil; diegormmarcondes@gmail.com (D.M.); kdson19@gmail.com (K.S.); sw@ime.usp.br (S.W.)

* Correspondence: claudiap@ime.usp.br; Tel.: +55-11-3091-6142

† These authors contributed equally to this work.

Abstract: In *An Introduction to Probability Theory and its Applications*, W. Feller established a way of ending the St. Petersburg paradox by the introduction of an entrance fee, and provided it for the case in which the game is played with a fair coin. A natural generalization of his method is to establish the entrance fee for the case in which the probability of heads is θ ($0 < \theta < 1/2$). The deduction of those fees is the main result of Section 2. We then propose a Bayesian approach to the problem. When the probability of heads is θ ($1/2 < \theta < 1$) the expected gain of the St. Petersburg game is finite, therefore there is no paradox. However, if one takes θ as a random variable assuming values in $(1/2, 1)$ the paradox may hold, which is counter-intuitive. In Section 3 we determine necessary conditions for the absence of paradox in the Bayesian approach and in Section 4 we establish the entrance fee for the case in which θ is uniformly distributed in $(1/2, 1)$, for in this case there is a paradox.

Keywords: St. Petersburg paradox; entrance fees; Bayesian analysis

MSC: 60F05

1. Introduction

The St. Petersburg paradox was first discussed in letters between Pierre Rémond de Montmort and Nicholas Bernoulli, dated from 1713, and was supposedly invented by the latter [1]. The paradox has been since then one of the most famous examples in probability theory and has been generalized by economists, philosophers and mathematicians, being widely applied in decision theory and game theory. The paradox arises from a very simple coin tossing game. A player tosses a fair coin until it falls heads. If this occurs at the ith toss, the player receives 2^i money unities. However, the paradox appears when one engages in determining the fair quantity that should be paid by a player to play a trial of this game. Indeed, in the eighteenth century, when the paradox was first studied, the main goal of probability theory was to answer questions that involved gambling, especially the problem of defining fair amounts that should be paid to play a certain game or if a game was interrupted before its end. Those amounts were then named *moral value* or *moral price* and are what we call *expected utility* nowadays. It is intuitive to take the expected gain at a trial of the game or the expected gain of a player if the game were not to end now, respectively, as *moral values*, and that is what the first probabilists established as those fair amounts. Nevertheless, when the expected gain was not finite, they would not know what to do, and that is exactly what happens in the St. Petersburg game. In fact, let X denote the gain of a player at a trial of the game. Then, the expected gain at a trial of the game is given by:

$$E(X) = \sum_{i=1}^{\infty} 2^i \frac{1}{2^i} = \infty,$$

since the random variable that represents the toss in which the first heads appears has a geometric distribution with parameter 1/2.

Besides the fact that the gain is indeed a random variable with infinite expected value, the St. Petersburg paradox has drawn the attention of many mathematicians and philosophers throughout the years because it deals with a game that nobody in his right state of mind would want to pay large amounts to play, and that is what has intrigued great mathematicians such as Laplace, Cramer, and the Bernoulli family. Indeed, at that time, random variables with infinite expected value were quite disturbing, for not even the simplest laws of large numbers had been proven and the difference between a random variable and its expected value was not yet clear. In this scenario, the St. Petersburg paradox emerged and took its place in the probability theory.

The magnitude of the paradox may be exemplified by the comparison between the odds of wining in the lottery and recovering the value paid to play the game, as shown in [2], in the following way. Suppose a lottery that pays 50 million money unities to the gambler that hits the 6 numbers drawn out of sixty. For simplification, suppose that there is no possibility of a draw and that the winner always gets the whole 50 million. Now, presume a gambler can bet on 16 numbers, i.e., bet on $\binom{16}{6} = 8008$ different sequences of 6 numbers, paying $10,000$ money unities, and let p be his probability of wining. Then, it is straightforward that:

$$p = \frac{\binom{16}{6}}{\binom{60}{6}} = 1.59 \times 10^{-4}.$$

On the other hand, if one pays $10,000$ money unities to play the St. Petersburg game, he will get his money back (and eventually profit) if the first heads appears on the 14th toss or later. Therefore, letting q be the probability of the gambler recovering its money, it is easy to see that:

$$q = \mathbb{P}\{X \geq 14\} = 1 - \mathbb{P}\{X < 14\} = 1 - \sum_{i=1}^{13} \frac{1}{2^i} = 1.22 \times 10^{-4}$$

and $p > q$. Then, winning the lottery can be more likely than recovering the money in a St. Petersburg game. Of course, it is true only for a $10,000$ money unities (or more) lottery bet. Thus, if one is interested in investing some money, the lottery may be a better investment than the St. Petersburg game. I would rather take my chances on Wall Street, for p and q are far too small.

Nevertheless, the paradox gets even more interesting when we compare the lottery in the scenario above and the St. Petersburg game in terms of their expected gain. On the one hand, as seen before, the expected gain in the St. Petersburg game is infinite. On the other hand, the expected gain at the lottery on the scenario above is:

$$E(\text{gain at the lottery}) = 50p \times 10^6 - 10^4 \approx -2050.$$

Here the St. Petersburg paradox is brought to a twenty-first century context, for which even the layman can understand its most intrinsic problem: how can one expect to win an infinite amount of money, but have at the same time less probability of winning anything at all than someone who expects to lose money? Now imagine the impact of this result on the eighteenth century mathematicians, who did not know the modern probability theory, which greatly facilitates understanding of the St. Petersburg paradox.

A first attempt to solve the paradox was made independently by Daniel Bernoulli in a 1738 paper translated in [3] and by Gabriel Cramer in a 1728 letter to Nicholas Bernoulli available in [1]. In Bernoulli's approach, *the value of an item is not to be based on its price, but rather on the utility it yields*. In fact, the price of an item depends solely on the item itself and is equal for every buyer, although the utility is dependent on the particular circumstances of the purchase and the person making it. In this context, we have his famous assertion that *a gain of one thousand ducats is more significant to a pauper than to a rich man though both gain the same amount*.

In his approach, the *gain* in the St. Petersburg game is replaced by the *utility* it yields, i.e., a *linear* utility function that represents the monetary gain of the game is replaced by a *logarithm* utility function. If a logarithm utility function is used instead, the paradox disappears, for now the expected incremental utility is finite and, therefore, may be used to determine a fair amount to be paid to play a trial of the game. The logarithm utility function takes into account the total wealth of the person willing to play the game, which makes the utility function dependent on who is playing the game, and not only on the game itself, in the following way.

Let w be the total wealth of a player, i.e., all the money he has available to gamble at a given moment. We define the utility of a monetary quantity m as $U(m) = \log m$, in which the logarithm is the natural one. Therefore, letting c be the value that should be paid to play a trial of the game, the expected incremental utility, i.e., how much utility the player expects to gain, is given by:

$$\Delta E(U) = \sum_{i=1}^{\infty} \frac{1}{2^i}\left[U(w + 2^i - c) - U(w)\right] = \left[\sum_{i=1}^{\infty} \frac{1}{2^i}\log(w + 2^i - c)\right] - \log(w) < \infty,$$

and we have an implicit relation between the total wealth of a player and how much he must pay to expect to gain a finite utility $\Delta E(U)$. Table 1 shows the values of c for which $\Delta E(U) = 0$, i.e., the maximum monetary unities someone must be willing to pay to play the game, for different values of w.

Table 1. The maximum value c someone with total wealth w must be willing to pay to play a trial of the St. Petersburg game according to the logarithm utility function, for different values of w.

Total Wealth (w)	Maximum Value (c)	Total Wealth (w)	Maximum Value (c)
2	3.35	500	9.99
4	4	1000	10.96
20	5.77	10,000	14.24
50	6.90	100,000	17.56
100	7.80	500,000	19.88
200	8.73	1,000,000	20.88

Note that, if $w \leq 4$, one must be willing to give all his money and eventually take a loan to play the game. In fact, the game is advantageous for players that have little wealth and, the richer the player is, the lower the percentage of his wealth that he must be willing to pay to play the game.

Although the logarithm utility function solves the St. Petersburg paradox, the paradox immediately reappears if the gain when the first heads occurs at the ith trial is changed from 2^i to $2^{M(i)}$, in which $M(i)$ is another function of i and not the identity one. For example, if $M(i) = 2^i$ the paradox holds even with the logarithm utility function. This fact was outlined by [4] that created the so-called super St. Petersburg paradoxes, in which $M(i)$ may take different forms.

There are many others generalizations and solutions to the St. Petersburg paradox, as shown in [5–7], for example. However, the main goal of this paper is to present and generalize W. Feller's solution by the introduction of an entrance fee, as will be defined on the next section. We generalize his method for the game in which the coin used is not fair, i.e., its probability of heads is $\theta \neq 1/2$, and for the case in which the coin is a random one, i.e., its probability θ of heads is a random variable defined in $(0,1)$.

2. Entrance Fee

A way of ending the paradox, i.e., making the game fair in the sense discussed by [8], is to define an entrance fee e_n that should be paid by a player so he can play n trials of the St. Petersburg game. This solution differs from the solutions given above in the fact that it does not take into account the utility function of a player, although it is rather theoretical, for it determines the cost of the game based on the convergence in probability of a convenient random variable. Furthermore, Feller's solution is given for

the scenario in which the player pays to play n trials of the game, i.e., he pays a quantity to play n trials of the game described above and his gain is the sum of the gains at each trial. In Feller's definition, the entrance fee e_n will be fair if $\frac{S_n}{e_n}$ tends to 1 in probability, in which X_k is the gain at the kth trial of the game and $S_n = \sum_{k=1}^{n} X_k$. If the coin is fair, it was proved by [9] that $e_n = n \log_2 n$.

Note that, of course, $\lim_{n\to\infty} e_n = \infty$. However, e_n gives the *rate* in which the accumulated gain at the trials of the game increases, so that paying the entrance fee e_n may be considered fair, for it will be as close to the accumulated gain in n trials of the game as desired with probability 1 as the number of trials diverges. Although classic and well-known in probability theory, Feller's solution may still be generalized to the case in which the coin used at the game is not fair, in the following way.

As a generalization of the St. Petersburg game, one could toss a coin with probability θ $(0 < \theta < 1)$ of heads. If $0 < \theta \leq 1/2$ the paradox still holds, for:

$$E(X) = \sum_{i=1}^{\infty} 2^i \theta (1 - \theta)^{i-1} \geq \sum_{i=1}^{\infty} 2^i \frac{\theta}{2^{i-1}} = \infty. \tag{1}$$

However, if $1/2 < \theta < 1$,

$$E(X) = \sum_{i=1}^{\infty} 2^i \theta (1 - \theta)^{i-1} = \frac{2\theta}{2\theta - 1} < \infty, \tag{2}$$

so there is no paradox. Applying the same method used by [9], we have, for the game played with a coin with probability θ $(0 < \theta < \frac{1}{2})$ of heads, the entrance fee presented in the theorem below. This has remained an open problem for the last decades and is originally solved in this paper. Proofs for all results are presented in the Appendix A.

Theorem 1. *Let $\alpha = \frac{1}{1-\theta}$. If $0 < \theta < 1/2$ and $e_n = \frac{2\theta}{1-2\theta} n \left(n^{-1+\log_\alpha 2} \log_\alpha n - 1 \right)$ then*

$$\lim_{n\to\infty} \mathbb{P}\left\{ \left| \frac{S_n}{e_n} - 1 \right| > \delta \right\} = 0.$$

Figure 1 shows the entrance fees for selected n and θ. It is important to note that the limit of e_n, as defined in Theorem 1, as θ increases to 1/2, does not equal the entrance fee $n \log_2 n$ for the fair coin game. In fact,

$$\lim_{\theta \to 1/2} e_n = \lim_{\theta \to 1/2} \frac{2\theta}{1 - 2\theta} n \left(\log_2 n - 1 \right) = \infty$$

for all values of n. Therefore, the entrance fee has an interesting discontinuity property at $\theta = 1/2$. Furthermore, as the value of θ decreases, the entrance fee increases for each n. This behaviour is expected for as the value of θ decreases, it will take longer for the first heads to show up and the player can expect to gain more money.
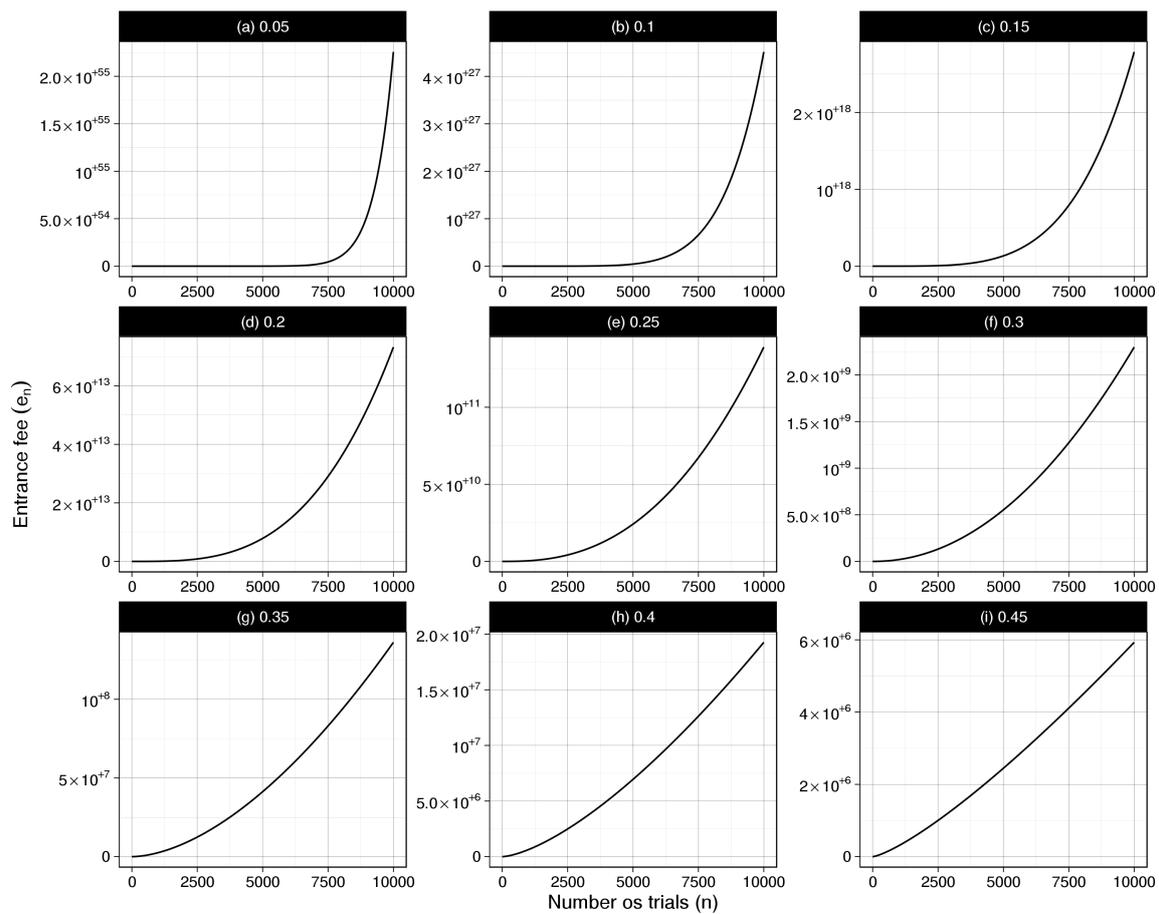
Figure 1. Entrance fees for different values of *n* and for $\theta \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$. Each figure presents the graph of e_n as a function of *n*, for selected values of θ, that are the values on top of each graph. The x-axis presents the values of *n* and y-axis the values of e_n, for each considered θ.

3. Bayesian Approach

Persi Diaconis and Susan Holmes wrote that there is much further work to be done by incorporating uncertainty and prior knowledge into basic probability calculations and developed a Bayesian version of three classical problems in [10]. Following this advice, we develop a Bayesian approach to the St. Petersburg paradox.

The Bayesian paradigm started with Reverend Thomas Bayes, whose work led to what is known today as Bayes theorem, and was first published posthumously in 1763 by Richard Price [11]. Although his work inspired great probabilists and statisticians of the following centuries, it was not until the second half of the twentieth century that the Bayesian paradigm was popularized within the statistical society, for the development of computer technology made it possible to apply Bayesian analysis in a variety of theoretical and practical problems.

In the centuries following Bayes work, another branch of statistics, the frequentist one, was developed and exhaustively studied by probabilists, mathematicians and, of course, statisticians. The main characteristic of the frequentist statistics is that it treats the probability of an event as the limit of the ratio between the number of times the event occurs and the number of times the experiment, in which the event is a possible outcome, is repeated, the limit being taken as the latter diverges. The sharper reader may immediately conclude that this interpretation of probability has an intrinsic problem: not every experiment can be *replicated*. Consider, for example, the probability of rain tomorrow. By a frequentist perspective, this probability is the limit of the ratio between the number of times that *rain* occurs tomorrow and the number of times that tomorrow *happens*. However, it is naive

to think of probability in this way, for tomorrow will happen only once. Therefore, there is a need for a more general interpretation of probability, and that is where the Bayesian approach comes into play.

In the Bayesian approach, the concept of probability may be interpreted as a measure of knowledge: how much is known about the odds of an event to occur. Of course, the concept of probability in Bayesian statistics may be considered by a frequentist statistician as subjective, for each person may have different knowledge about the likelihood of an event and, therefore, different people may have distinct conclusions analysing the same data. The debate between the Bayesian and the frequentist approach has been going on for the last decades, and we are not going to get into its merits. However, for a succinct and introductory presentation of the Bayesian approach see [12].

In the context of the St. Petersburg paradox, the main difference between the frequentist and Bayesian approach manifests in the way in which the parameter θ, i.e., the probability of heads of the game's coin, is treated. In the frequentist approach, θ is a *number* in the interval $(0, 1)$, either known or unknown. If it is known, as it was in Sections 1 and 2, there is nothing to be done from a statistical point of view. However, if it is unknown, it may be estimated and hypotheses about it may be tested. On the other hand, in the Bayesian approach, the parameter θ is a random variable, i.e, has a cumulative probability function $F(a) = \mathbb{P}\{\theta \leq a\}$, that is called the priori distribution of θ. The St. Petersburg game played with a coin with probability θ_0 may also be interpreted from a Bayesian point of view, as it is enough to take:

$$F(a) = \begin{cases} 1, & \text{if } a \geq \theta_0 \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

However, in this section, we will determine conditions on F for which the paradox does not hold, i.e., $E(X) < \infty$. For this purpose, we have first to define the probability distribution of X (the gain at a trial of the St. Petersburg game) from a Bayesian point of view. Applying basic properties of the probability measure we have:

$$\mathbb{P}\{X = x\} = \int_0^1 \mathbb{P}\{X = x|\theta\}dF(\theta) = \int_0^1 \theta(1-\theta)^{x-1}dF(\theta), \tag{4}$$

for $X|\theta$ has a geometric distribution with parameter θ. Therefore, the expected gain $E(X)$ at a trial of the game is given, applying (4), by:

$$E(X) = \sum_{x=1}^{\infty} x \int_0^1 \mathbb{P}\{X = x|\theta\}dF(\theta) = \int_0^1 \sum_{x=1}^{\infty} x\mathbb{P}\{X = x|\theta\}dF(\theta) = \int_0^1 E(X|\theta)dF(\theta)$$

If F is given by (3), it was proved in (1) and (2) that if $\theta_0 \leq 1/2$ the paradox holds, whilst if $\theta_0 > 1/2$ there is no paradox. This conclusion is the same from a frequentist point of view, and therefore is not of interest. The Bayesian approach gets more appealing when F takes other forms, distinct from (3). As a motivation for the Bayesian approach, we present the case in which θ is uniformly distributed in $(1/2, 1)$, which raises some interesting questions about the Bayesian approach.

From (2), we know that if $1/2 < \theta < 1$, the expected gain of the game is finite, therefore there is no paradox. However, suppose one incorporates prior knowledge to θ and takes it as a random variable assuming values in $(1/2, 1)$ with probability 1 and distribution F. It would be expected that the paradox would not hold for any distribution F with such properties. However, this does not happen.

Suppose θ is uniformly distributed in $(1/2, 1)$. Then,

$$E(X) = \int_{1/2}^{1} \sum_{x=1}^{\infty} 2^x \theta(1-\theta)^{x-1} 2d\theta = \sum_{x=1}^{\infty} 2^{x+1} \int_0^{1/2} (1-p)p^{x-1}dp = \infty. \tag{5}$$

The result (5) raises a couple of interesting questions about the Bayesian approach to the St. Petersburg paradox. Firstly, one may ask for which prior distributions F the paradox

holds. Theorem 2 answers this question, giving a sufficient condition for the absence of paradox in the Bayesian approach. Then, one could inquire what is the entrance fee for the case in which θ is a random variable uniformly distributed in $(1/2, 1)$. In Section 4, we establish the entrance fee for this case.

We now treat the case in which the probability θ of heads in each trial of the game is a random variable, defined as $(\Omega, \mathbb{F}, \mathbb{P})$, taking values in the interval $(1/2, 1)$ with probability 1 and distribution F. The case in which $F(1/2) > 0$ is not of interest, for the paradox holds for any F with this property. Therefore, in this scenario, the expected gain at a trial of the game is given by:

$$E(X) = \int_{1/2}^{1} E(X|\theta)dF(\theta) = \int_{1/2}^{1/2+\epsilon} E(X|\theta)dF(\theta) + \int_{1/2+\epsilon}^{1} E(X|\theta)dF(\theta)$$

for any fixed $0 < \epsilon < 1/2$. From now on, consider ϵ fixed in $(0, 1/2)$, unless said otherwise. Now,

$$\int_{1/2+\epsilon}^{1} E(X|\theta)dF(\theta) \leq \int_{1/2+\epsilon}^{1} E(X|\theta = 1/2 + \epsilon)dF(\theta) \leq \frac{1+2\epsilon}{2\epsilon} < \infty$$

so that the paradox will exist if, and only if, $\int_{1/2}^{1/2+\epsilon} E(X|\theta)dF(\theta) = \infty$. The lemma below gives a sufficient condition for the absence of paradox and will be used to prove the main result of this section.

Lemma 1. *Suppose that F is absolutely continuous in $(1/2, 1/2 + \epsilon)$ and $F'(\theta) = f(\theta), \theta \in (1/2, 1/2 + \epsilon)$. If $f(\theta) \leq c(2\theta - 1), \forall \theta \in (1/2, 1/2 + \epsilon)$ and for some $c \in \mathbb{R}_{+}$, there is no paradox.*

Furthermore, the lemma bellow gives an obvious condition for the absence of paradox.

Lemma 2. *If there exists an $\epsilon > 0$ such that $F(1/2 + \epsilon) = 0$, then $\int_{1/2}^{1/2+\epsilon} E(X|\theta)dF(\theta) < \infty$.*

If F is discrete or singular on an ϵ'-neighbourhood of $1/2$, and there is no $\epsilon < \epsilon'$ such that $F(1/2 + \epsilon) = 0$, then there may or may not be paradox. For example, if F is such that $\mathbb{P}\{\theta = 1/2 + 1/m\} = 1/2^m, m = 1, 2, \ldots$, the paradox does not hold. On the other hand, if θ is so that $\mathbb{P}\{\theta = 1/2 + 1/2^{m+2}\} = 1/2^m, m = 1, 2, \ldots$, the paradox holds. Now, if F is absolutely continuous in $(1/2, 1/2 + \epsilon)$, the paradox holds if, and only if, the limit of the probability density of θ at $1/2$ is greater than zero.

Theorem 2. *Suppose that F is absolutely continuous in $(1/2, 1/2 + \epsilon)$ and $F'(\theta) = f(\theta), \theta \in (1/2, 1/2 + \epsilon)$. There is no paradox if, and only if, $\lim_{\theta \to 1/2} f(\theta) = 0$.*

In summary, there will be no paradox if there is no probability mass on $1/2$ nor around it. Then, if F is discrete or singular, there may or may not be paradox. However, if F is absolutely continuous in a neighbourhood of $1/2$, there will be no paradox if, and only if, the probability density tends to zero as θ tends to $1/2$. Otherwise, there would be a probability mass in *every* neighbourhood ϵ' of $1/2$.

4. Uniformly Distributed θ

The method of [9] will now be applied to find the entrance fee for the game in which the probability θ of heads in each trial is a random variable with uniform distribution in $(1/2, 1)$, i.e., $f(\theta) = 2\mathbb{1}_{\{\theta \in (1/2,1)\}}$. As $\lim_{\theta \to 1/2} f(\theta) = 2$, by Theorem 2, in this scenario there is a paradox. It will be shown that the entrance fee in this case is $e_n = n \log_e(\log_2 n)$.

Theorem 3. *If the probability of heads θ in each trial of the game is a random variable uniformly distributed in* $(1/2, 1)$ *and* $e_n = n \log_e(\log_2 n)$ *then for all* $\delta > 0$

$$\lim_{n \to \infty} \mathbb{P}\left\{ \left| \frac{S_n}{e_n} - 1 \right| > \delta \right\} = 0.$$

Figure 2 shows the entrance fee for different values of n.
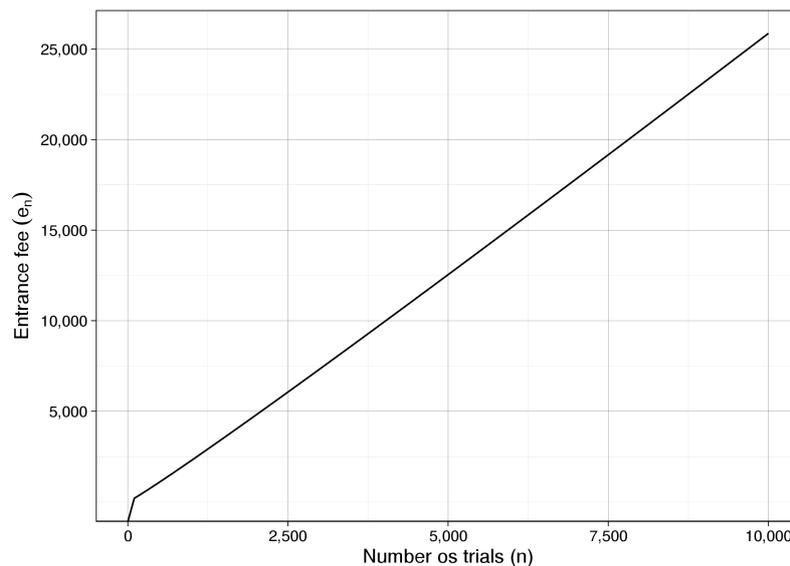
Figure 2. Entrance fees for different values of n for θ uniformly distributed in $(1/2, 1)$.

Note that the entrance fee in this case is smaller than the entrance fee for the case in which $\theta = 1/2$ with probability 1. It is expected, for the probability mass that was first concentrated on $1/2$ is now spread equally over the interval $(1/2, 1)$, an interval in which there is no paradox. In fact, what is really interesting in this Bayesian approach is the fact that if θ is uniformly distributed in $(1/2, 1)$, the probability of being chosen a θ, for which the paradox holds, is zero, although the expected gain is infinite and, overall, there is a paradox. A suggestion for further studies would be to establish the entrance fee for random θ with other distributions, e.g., truncated Pareto-like distributions with scale parameter equal to $1/2$ and beta distributions.

Appendix A

Proof 1 (Theorem 1). Let $\{(U_k, V_k) : k = 1, 2, \dots\}$ be a sequence of random variables such that:

$$\begin{cases} (U_k, V_k) = (X_k, 0), & \text{if } X_k \leq (n \log_\alpha(\log_\alpha n))^{\log_\alpha 2} \\ (U_k, V_k) = (0, X_k), & \text{if } X_k > (n \log_\alpha(\log_\alpha n))^{\log_\alpha 2}. \end{cases}$$

Therefore,

$$\mathbb{P}\left\{ \left| \frac{S_n}{e_n} - 1 \right| > \delta \right\} \leq \mathbb{P}\left\{ \left| \sum_{k=1}^{n} U_k - e_n \right| > \delta e_n \right\} + \mathbb{P}\left\{ \sum_{k=1}^{n} V_k \neq 0 \right\}. \tag{A1}$$

Let r be the greatest integer such that $\alpha^r \leq n \log_\alpha(\log_\alpha n) < \alpha^{r+1}$. Then,

$$r = \log_\alpha \alpha^r \leq \log_\alpha(n \log_\alpha(\log_\alpha n)) < \log_\alpha \alpha^{r+1} = r+1$$
$$\implies 2^r \leq 2^{\log_\alpha(n \log_\alpha(\log_\alpha n))} \leq 2^{r+1}.$$

Now, defining $\beta = n \log_\alpha(\log_\alpha n)$,

$$2^{\log_\alpha(n \log_\alpha(\log_\alpha n))} = 2^{\log_\alpha \beta} = \beta^{\log_\beta 2^{\log_\alpha \beta}} = \beta^{\log_\alpha 2} = (n \log_\alpha(\log_\alpha n))^{\log_\alpha 2}$$
$$\implies 2^r \leq (n \log_\alpha(\log_\alpha n))^{\log_\alpha 2} \leq 2^{r+1}.$$

Applying the inequalities above to the second term of the right side of (A1) we get:

$$\mathbb{P}\left\{ \sum_{k=1}^n V_k \neq 0 \right\} \leq n\mathbb{P}\{X_1 > (n \log_\alpha(\log_\alpha n))^{\log_\alpha 2}\} \leq n\mathbb{P}\{X_1 > 2^r\} \leq$$

$$\leq n(1-\theta)^r \leq \frac{n(1-\theta)^{-1}}{\alpha^{r+1}} \leq \frac{n(1-\theta)^{-1}}{n \log_\alpha(\log_\alpha n)} \xrightarrow{n \to \infty} 0.$$

Note that,

$$E(U_k) = \sum_{i=1}^r 2^i \theta(1-\theta)^{i-1} = \frac{2\theta}{1-2\theta}\left\{[2(1-\theta)]^r - 1\right\}$$

$$E(U_k^2) = \sum_{i=1}^r 2^{2i}\theta(1-\theta)^{i-1} = \frac{4\theta}{3-4\theta}\left\{[4(1-\theta)]^r - 1\right\}$$

$$Var\left(\sum_{k=1}^n U_k - nE(U_k) \right) \leq nE(U_k^2) \leq 2n[4(1-\theta)]^r = 2n\alpha^{\log_\alpha[4(1-\theta)]^r} \leq$$

$$\leq 2n[\alpha^r]^{-1+2\log_\alpha 2} \leq 2n(n \log_\alpha(\log_\alpha n))^{-1+2\log_\alpha 2}.$$

Taking,

$$e_n = \frac{2\theta}{1-2\theta} n(n^{-1+\log_\alpha 2} \log_\alpha n - 1) \geq \frac{\theta}{1-2\theta} n\left(n^{-1+\log_\alpha 2} \log_\alpha n \right)$$

and applying the Chebyshev's inequality to the first term of the right side of (A1) we get:

$$\mathbb{P}\left\{ \left| \sum_{k=1}^n U_k - nE(U_k) \right| > \delta e_n \right\} \leq \frac{2n^{2\log_\alpha 2}(\log_\alpha(\log_\alpha n))^{-1+2\log_\alpha 2}}{\delta^2 \left(\frac{\theta}{1-2\theta}\right)^2 n^{2\log_\alpha 2}(\log_\alpha n)^2} \leq$$

$$\leq \frac{2(\log_\alpha(\log_\alpha n))^{-1+2\log_\alpha 2}}{\delta^2 \left(\frac{\theta}{1-2\theta}\right)^2 (\log_\alpha n)^2} \xrightarrow{n \to \infty} 0.$$

Now, it is enough to prove that $nE(U_k)$ may be approximated by e_n for large values of n. By definition,

$$\alpha^r \le n \log_\alpha(\log_\alpha n) < \alpha^{r+1} \implies \log_\alpha n \le r \le \log_\alpha n + \log_\alpha(\log_\alpha(\log_\alpha n))$$

$$\implies \lim_{n\to\infty} \frac{\log_\alpha n}{\log_\alpha n + \log_{2(1-\theta)}(\log_\alpha n)} \le \lim_{n\to\infty} \frac{r}{\log_\alpha n + \log_{2(1-\theta)}(\log_\alpha n)}$$

$$\le \lim_{n\to\infty} \frac{\log_\alpha n + \log_\alpha(\log_\alpha(\log_\alpha n))}{\log_\alpha n + \log_{2(1-\theta)}(\log_\alpha n)}$$

$$\implies 1 \le \lim_{n\to\infty} \frac{r}{\log_\alpha n + \log_{2(1-\theta)}(\log_\alpha n)} \le 1 \implies r \approx \log_\alpha n + \log_{2(1-\theta)}(\log_\alpha n)$$

$$\implies r \approx \log_\alpha n + (\log_{2(1-\theta)} \alpha)(\log_\alpha(\log_\alpha n)) \implies r \approx \log_\alpha[n(\log_\alpha n)^{\log_{2(1-\theta)} \alpha}]$$

$$\implies \alpha^r \approx n(\log_\alpha n)^{\log_{2(1-\theta)} \alpha}.$$

Hence,

$$[2(1-\theta)]^r = \alpha^{\log_\alpha[2(1-\theta)]^r} = (\alpha^r)^{\log_\alpha[2(1-\theta)]} \approx [n(\log_\alpha n)^{\log_{2(1-\theta)} \alpha}]^{\log_\alpha[2(1-\theta)]}$$

$$\implies [2(1-\theta)]^r \approx n^{-1+\log_\alpha 2} \log_\alpha n$$

$$\implies nE(U_k) = \frac{2\theta}{1-2\theta}n\Big\{[2(1-\theta)]^r - 1\Big\} \approx \frac{2\theta}{1-2\theta}n\Big\{n^{-1+\log_\alpha 2} \log_\alpha n - 1\Big\} = e_n$$

for large values of n. Therefore, for $e_n = \frac{2\theta}{1-2\theta}n\Big\{n^{-1+\log_\alpha 2} \log_\alpha n - 1\Big\}$,

$$\mathbb{P}\Big\{\Big|\sum_{k=1}^{n} U_k - n\frac{2\theta}{1-2\theta}\Big[n^{-1+\log_\alpha 2} \log_\alpha n - 1\Big]\Big| > \delta e_n\Big\} \xrightarrow{n\to\infty} 0$$

$$\implies \mathbb{P}\Big\{\Big|\sum_{k=1}^{n} U_k - e_n\Big| > \delta e_n\Big\} \xrightarrow{n\to\infty} 0 \implies \lim_{n\to\infty} \mathbb{P}\Big\{\Big|\frac{S_n}{e_n} - 1\Big| > \delta\Big\} = 0.$$

$\square$

Proof 2 (Lemma 1).

$$\lim_{\theta\to1/2} E(X|\theta)f(\theta) = \lim_{\theta\to1/2} \frac{2\theta}{2\theta-1}f(\theta) \le \lim_{\theta\to1/2} \frac{2\theta}{2\theta-1}c(2\theta-1) = c < \infty$$

$$\implies \int_{1/2}^{1/2+\epsilon} E(X|\theta)dF(\theta) < \infty.$$

$\square$

Proof 3 (Lemma 2). Clearly $\int_{1/2}^{1/2+\epsilon} E(X|\theta)dF(\theta) = 0 < \infty.$ $\square$

Proof 4 (Theorem 2). ($\implies$) Consider the extension g of f on $(1/2 - \epsilon, 1/2 + \epsilon)$ given by:

$$g(\theta) = f(\theta)\mathbb{1}_{\{\theta\in[1/2,1/2+\epsilon)\}} - f(1-\theta)\mathbb{1}_{\{\theta\in(1/2-\epsilon,1/2)\}}$$

in which $\mathbb{1}$ is the indicator function. Note that,

$$\int_{1/2}^{1/2+\epsilon} E(X|\theta)dF(\theta) = \int_{1/2}^{1/2+\epsilon} E(X|\theta)g(\theta)d\theta.$$

As g is odd around $1/2$ and continuous for an ϵ small enough, $g'(1/2)$ exists and $g''(1/2) = 0$. Therefore, for an ϵ small enough, g' and g'' exist for $\theta \in I = (1/2, 1/2 + \epsilon)$. Consider ϵ such that g' and g'' exist on I and note that $0 \le g'(\theta) < \infty, \theta \in I$, for f is a probability density.

On the one hand, if $g''(\theta) > 0$ for $\theta \in I$, g is concave upward and,

$$g(\theta) \leq \frac{g(1/2+\epsilon)}{2\epsilon}(2\theta - 1), \forall \theta \in I.$$

Taking $c = \frac{g(1/2+\epsilon)}{2\epsilon}$, the result follows from Lemma 1.

On the other hand, if $g''(\theta) \leq 0$ for $\theta \in I$, by the Taylor expansion of g around $1/2$, for $1/2 \leq \bar{\theta} \leq \theta$ and $\forall \theta \in I$,

$$g(\theta) = g'(1/2)(\theta - 1/2) + \frac{g''(\bar{\theta})}{2}(\theta - 1/2)^2 \leq \frac{g'(1/2)}{2}(2\theta - 1).$$

Taking $c = \frac{g'(1/2)}{2}$ the result follows from Lemma 1.

($\Longleftarrow$) Suppose that there is no paradox, but $\lim_{\theta \to 1/2} f(\theta) = \xi > 0$. Then,

$$\infty > E(X) \geq \int_{1/2}^{1/+\epsilon} E(X|\theta)dF(\theta) \geq \lim_{\theta \to 1/2} \frac{2\theta\xi\epsilon'}{2\theta - 1} = \infty, 0 < \epsilon' < \epsilon.$$

$\square$

Proof 5 (Theorem 3). Let $\{(U_k, V_k) : k = 1, 2, \dots\}$ be a sequence of random variables such that,

$$\begin{cases} (U_k, V_k) = (X_k, 0), & \text{if } X_k \leq n\log_2(\log_2 n) \\ (U_k, V_k) = (0, X_k), & \text{if } X_k > n\log_2(\log_2 n). \end{cases}$$

Therefore,

$$\mathbb{P}\left\{\left|\frac{S_n}{e_n} - 1\right| > \delta\right\} \leq \mathbb{P}\left\{\left|\sum_{k=1}^{n} U_k - e_n\right| > \delta e_n\right\} + \mathbb{P}\left\{\sum_{k=1}^{n} V_k \neq 0\right\}. \tag{A2}$$

Defining r as the greatest integer such that $2^r \leq n\log_2(\log_2 n) < 2^{r+1}$ implies that:

$$\log_2 n < r \leq \log_2 n + \log_2(\log_2(\log_2 n)).$$

Applying the inequalities above to the second term of the right side of (A2) we get:

$$\mathbb{P}\left\{\sum_{k=1}^{n} V_k \neq 0\right\} \leq n\mathbb{P}\{X_1 > n\log_2(\log_2 n)\} \leq n\mathbb{P}\{X_1 > 2^r\} \leq n(1-\theta)^r \leq$$

$$\leq n(1/2)^r = \frac{2n}{2^{r+1}} \leq \frac{2n}{n\log_2(\log_2 n)} \xrightarrow{n \to \infty} 0.$$

Note that,

$$E(U_k) = \int_{1/2}^{1} \sum_{i=1}^{r} 2^i \theta(1-\theta)^{i-1} 2d\theta = \sum_{i=1}^{r} 2^{i+1} \int_0^{1/2}(1-p)p^{i-1}dp = \sum_{i=1}^{r}\left[\frac{2}{i} - \frac{1}{i+1}\right]$$

$$= \frac{r}{r+1} + \sum_{i=1}^{r}\frac{1}{i}.$$

Now,

$$\sum_{i=1}^{r}\frac{1}{i} = 1 + \sum_{i=2}^{r}\frac{1}{i} \leq 1 + \log_e r \text{ and } \sum_{i=1}^{r}\frac{1}{i} \geq \int_1^{r+1}\frac{dx}{x} = \log_e(r+1)$$

so that,

$$\log_e r \le \log_e(r+1) \le \sum_{i=1}^{r} \frac{1}{i} \le E(U_k) \le \frac{r}{r+1} + 1 + \log_e r \le 2 + \log_e r.$$

The moment of order two of U_k is given by:

$$E(U_k^2) = \int_{1/2}^{1} \sum_{i=1}^{r} 2^{2i} \theta(1-\theta)^{i-1} 2 d\theta = \sum_{i=1}^{r} 2^{2i+1} \int_{0}^{1/2} (1-p)p^{i-1}dp =$$

$$= \sum_{i=1}^{r} \left[\frac{2^{i+1}}{i} - \frac{2^i}{i+1} \right] \le \sum_{i=1}^{r} \frac{2^{i+1}}{i} \le \sum_{i=1}^{r} 2^{i+1} \le 2^{r+2} = (4)2^r \le 4n \log_2(\log_2 n).$$

Finally,

$$Var\left(\sum_{k=1}^{n} U_k - nE(U_k) \right) \le nVar(U_k) \le nE(U_k^2) \le 4n^2 \log_2(\log_2 n).$$

Taking $e_n = n \log_e(\log_2 n)$ and applying the Chebyshev's inequality to the first term of the right side of (A2) we get:

$$\mathbb{P}\left\{ \left| \sum_{k=1}^{n} U_k - n\left[\frac{r}{r+1} + \sum_{i=1}^{r} \frac{1}{i} \right] \right| > \delta e_n \right\} \le \frac{4n^2 \log_2(\log_2 n)}{\delta^2 n^2 (\log_e(\log_2 n))^2} \le$$

$$\le \frac{4}{\delta^2 (\log_e 2) \log_e(\log_2 n)} \xrightarrow{n \to \infty} 0.$$

Again, it is enough to show that $nE(U_k)$ may be approximated by e_n for large values of n. From the inequalities established above:

$$\log_e(\log_2 n) \le \log_e r \le E(U_k) \le 2 + \log_e r \le 2 + \log_e(\log_2 n + \log_2(\log_2(\log_2 n))) \le$$
$$\le 2 + \log_e(\log_2 n + \log_2 n) \le 2 + \log_e 2 + \log_e(\log_2 n) \le 3 + \log_e(\log_2 n)$$

and,

$$1 = \lim_{n \to \infty} \frac{\log_e(\log_2 n)}{\log_e(\log_2 n)} \le \lim_{n \to \infty} \frac{E(U_k)}{\log_e(\log_2 n)} \le \lim_{n \to \infty} \frac{3 + \log_e(\log_2 n)}{\log_e(\log_2 n)} = 1.$$

$\square$

References

1. Van der Waerden, B.L. *Die Werke von Jakob Bernoulli*; Springer: Basel, Switzerland, 1975.
2. Peixoto, C.M.; Wechsler, S.; de São Petersburgo, O.P. *Boletim da Seção Brasileira da ISBA* **2013**, *6*, 6–7.
3. Bernoulli, D. Exposition of a new theory on the measurement of risk. *Econometrica* **1954**, *22*, 23–36.
4. Menger, K. Das Unsicherheitsmoment in der Wertlehre Betrachtungen im Anschluß an das sogenannte Petersburger Spiel. *Zeitschrift für Nationalökonomie* **1934**, *5*, 459–485. doi:10.1007/BF01311578.
5. Peters, O. The time resolution of the St Petersburg paradox. *Philos. Trans. R. Soc. A* **2011**, *369*, 4913–4931.
6. Samuelson, P. A. St. Petersburg paradoxes: Defanged, dissected, and historically described. *J. Econ. Lit.* **1977**, *15*, 24–55.
7. Seidl, C. The St. Petersburg paradox at 300. *J. Risk Uncertain.* **2013**, *46*, 247–264.
8. Feller, W. Note on the law of large numbers and fair games. *Ann. Math. Stat.* **1945**, *16*, 3, 301–304.
9. Feller, W. Chapter X: Law of Large Numbers. In *An Introduction to Probability Theory and Its Applications*; John Wiley & Sons: New York, NY, USA, 1968; Volume I, pp. 243–263.

10. Diaconis, P.; Holmes, S. A Bayesian peek into Feller volume I. *Sankhyā Ser. A* **2002**, *64*, 820–841.

11. Bayes, T.; Price, R.; Canton, J. *An Essay Towards Solving a Problem in the Doctrine of Chances*; Royal Society: London, UK, 1763.

12. Lee, P. M. Chapter 1 Preliminaries. In *Bayesian Statistics: An Introduction*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1997; pp. 1–27