Bernoulli **12**(6), 2006, 971–1002

Pooling strategies for St Petersburg gamblers

SÁNDOR CSÖRGŐ[1] and GORDON SIMONS[2]

[1]*Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, H-6720 Szeged, Hungary. E-mail: csorgo@math.u-szeged.hu*
[2]*Department of Statistics and Operations Research, University of North Carolina, Smith Building, Chapel Hill NC 27599-3260, USA. E-mail: gsimons@email.unc.edu*

Peter offers to play exactly one St Petersburg game with each of $n \geqslant 2$ players, $\mathrm{Paul}_1, \ldots, \mathrm{Paul}_n$, whose conceivable pooling strategies are described by all possible probability distributions $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$. Comparing infinite expectations, we characterize among all $\boldsymbol{p}_n$ those admissible strategies for which the pooled winnings, each distributed as $V_{\boldsymbol{p}_n} = \sum_{k=1}^{n} p_{k,n} X_k$, yield a finite added value for each and every one of $\mathrm{Paul}_1, \ldots, \mathrm{Paul}_n$ in comparison with their individual winnings $X_1, \ldots, X_n$, even though their total winnings $S_n = X_1 + \ldots + X_n$ is the same. We show that the added value of an admissible $\boldsymbol{p}_n$ is just its entropy $H(\boldsymbol{p}_n)$, and we determine the best admissible strategy $\boldsymbol{p}_n^*$. Moreover, for every $n \geqslant 2$ and $\boldsymbol{p}_n$ we construct semistable approximations to $S_{\boldsymbol{p}_n} = V_{\boldsymbol{p}_n} - H(\boldsymbol{p}_n)$. We show in particular that $S_{\boldsymbol{p}_n}$ has a proper semistable asymptotic distribution as $n \to \infty$ along the entire sequence of natural numbers whenever $\max\{p_{1,n}, \ldots, p_{n,n}\} \to 0$ for a sequence $\boldsymbol{p}_n$ of admissible strategies, which is in sharp contrast to S_n/n, and the rate of convergence is very fast for $S_{\boldsymbol{p}_n^*}$.

Keywords: added value; asymptotic distributions; best admissible pooling strategies; comparison of infinite expectations; several players; St Petersburg games

1. Introduction

Peter offers to let Paul toss a fair coin repeatedly until it lands heads and pays him 2^k ducats if this happens on the kth toss, $k \in \mathbb{N} = \{1, 2, \ldots\}$. What is the price for Paul to pay to make the game 'equal and fair'? Since, for Paul's winning X,

$$\boldsymbol{P}\{X = 2^k\} = \frac{1}{2^k} \quad \text{and} \quad \boldsymbol{E}(X) = \sum_{k=1}^{\infty} 2^k \boldsymbol{P}\{X = 2^k\} = \sum_{k=1}^{\infty} \frac{2^k}{2^k} = \infty, \tag{1}$$

it is an infinite number of ducats, but, as Nicolaus Bernoulli wrote, 'there ought not be a sane man who would not happily sell his chance for forty ducats'. This is the St Petersburg paradox. The question was asked by Nicolaus Bernoulli (1713), and the excerpt is from a letter to his cousin Daniel Bernoulli in 1728. Citing his 'most illustrious uncle' and Gabriel Cramer's 1728 contribution through him, Daniel's (1738) St Petersburg paper initiated what, through three centuries of extensive discussions, has become, in Samuelson's (1977) words, 'an honored corner in the memory bank of the cultured analytic mind'. (Since we use the more convenient payoff system 2, 4, 8, ... instead of the original 1, 2, 4, ..., we doubled the 'twenty ducats' in the original text.) Other recent overviews are due to Jorland (1987) and Dutka (1988), and a fuller historical analysis with numerous new findings will appear in Csörgő and Simons (2008).

Since $\boldsymbol{P}\{X \leqslant x\} = \sum_{k=1}^{\lfloor \operatorname{Log} x \rfloor} 2^{-k}$, the distribution function of Paul's gain X is

$$F(x) = \boldsymbol{P}\{X \leqslant x\} = \begin{cases} 0, & \text{if } x < 2, \\ 1 - \dfrac{1}{2^{\lfloor \operatorname{Log} x \rfloor}} = 1 - \dfrac{2^{\langle \operatorname{Log} x \rangle}}{x}, & \text{if } x \geqslant 2, \end{cases} \tag{2}$$

where, with $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ denoting the integers, for any y within the real line $\mathbb{R}$, $\lfloor y \rfloor = \max\{k \in \mathbb{Z} : k \leqslant y\}$ denotes its integer part, $\lceil y \rceil = \min\{k \in \mathbb{Z} : k \geqslant y\} = -\lfloor -y \rfloor$, required later, denotes its 'upper integer part', and $\langle y \rangle = y - \lfloor y \rfloor = y + \lceil -y \rceil$ denotes its fractional part. The symbol Log (with a capital L) denotes the base 2 logarithm, and log will denote the natural logarithm. Beginning with D. Bernoulli (1738), the ideas concerning a single gain X spawned the development of modern economic theory based on the notion of utility. Yet, from a mathematical standpoint, equation (2) summarizes all that can be said about a single X; surely Nicolaus Bernoulli must have calculated that $\boldsymbol{P}\{X > 40\} = \boldsymbol{P}\{X > 32\} = \frac{1}{32} = 0.031\,25$, for example. So the first real mathematical question is Paul's price for the cumulative gain $S_n = X_1 + X_2 + \ldots + X_n$ in n games, where $X_1, X_2, \ldots$ are independent copies of X defined on a probability space $(\Omega, \mathcal{A}, \boldsymbol{P})$.

The subject reached a level of mathematical maturity when Feller (1945) proved a weak law of large numbers for this total gain, stating that

$$\frac{S_n}{n \operatorname{Log} n} \xrightarrow{\boldsymbol{P}} 1 \qquad \text{as } n \to \infty, \tag{3}$$

where $\xrightarrow{\boldsymbol{P}}$ denotes convergence in probability, thereby suggesting (Feller 1968: Section 10.4) that the 'fair price' for n games is $n \operatorname{Log} n$ ducats for large $n \in \mathbb{N}$. Using results from Csörgő and Simons (1996), it is explained in Csörgő (2002) why $n \operatorname{Log} n$ ducats will not satisfy Peter, the banker, and that in general no satisfactory solution can be based on laws of large numbers. But, since the function $2^{\langle \operatorname{Log} x \rangle}$ in (2) is not slowly varying at infinity, the classical Doeblin–Gnedenko criterion (Gnedenko and Kolmogorov 1954: 175) implies that S_n has no asymptotic distribution for any centring and norming sequences as $n \to \infty$ over the entire sequence $\mathbb{N}$ of natural numbers.

However, opening a new phase in the exploration of the problem, Martin-Löf (1985) proved along the subsequence $\{2^k\}_{k=1}^{\infty}$ that $\lim_{k \to \infty} \boldsymbol{P}\{2^{-k} S_{2^k} - k \leqslant x\} = G(x)$, $x \in \mathbb{R}$, for a non-stable semistable distribution function G of exponent 1, specified through its characteristic function given below. As was pointed out by Csörgő and Dodunekova (1991), Martin-Löf's result itself implies that there are as many non-degenerate asymptotic distributions of different types along subsequences of $\mathbb{N}$ as the cardinality of the continuum (and, in a friendly manipulation of his article's title, exactly that many clarifications of the paradox). To find all these different limiting types, they showed that it suffices to restrict attention to subsequences of $n^{-1} S_n - \operatorname{Log} n$, and, with $\gamma_n = n / 2^{\lceil \operatorname{Log} n \rceil} \in (\frac{1}{2}, 1]$, $n \in \mathbb{N}$, for any given subsequence $\{n_k\}_{k=1}^{\infty}$ of $\mathbb{N}$, the sequence $n_k^{-1} S_{n_k} - \operatorname{Log} n_k$ converges in distribution as $k \to \infty$ if and only if $\gamma_{n_k} \overset{\mathbf{cir}}{\to} \gamma \in (\frac{1}{2}, 1]$, meaning that either $\lim_{k\to\infty} \gamma_{n_k} = \gamma$ for some $\gamma \in (\frac{1}{2}, 1]$ or the sequence $\{\gamma_{n_k}\}_{k=1}^{\infty}$ has the limit point $\frac{1}{2}$, and possibly 1 as well, in which case we put $\gamma_{n_k} \overset{\mathbf{cir}}{\to} 1$. For later use we observe that the interval $(\frac{1}{2}, 1]$ becomes a compact space under the *circular* limiting operation '$\overset{\mathbf{cir}}{\to}$'. If this circular convergence $\gamma_{n_k} \overset{\mathbf{cir}}{\to} \gamma$ takes place for some $\gamma \in (\frac{1}{2}, 1]$, as $k \to \infty$, then

$$\lim_{k\to\infty} \boldsymbol{P}\left\{\frac{S_{n_k}}{n_k} - \operatorname{Log} n_k \leqslant x\right\} = G_\gamma(x) = \boldsymbol{P}\{W_\gamma \leqslant x\}, \qquad x \in \mathbb{R},$$

where

$$W_\gamma = \frac{1}{\gamma}\left\{\sum_{m=0}^{-\infty} 2^m \left[Y_m(\gamma) - \frac{\gamma}{2^m}\right] + \sum_{m=1}^{\infty} 2^m\, Y_m(\gamma)\right\} + \operatorname{Log}\frac{1}{\gamma}. \tag{4}$$

Here $\ldots, Y_{-2}(\gamma), Y_{-1}(\gamma), Y_0(\gamma), Y_1(\gamma), Y_2(\gamma), \ldots$ are independent random variables such that

$$\boldsymbol{P}\{Y_m(\gamma) = j\} = \frac{(\gamma/2^m)^j}{j!}\, \mathrm{e}^{-\gamma/2^m}, \qquad j = 0, 1, 2, \ldots,$$

that is, $Y_m(\gamma)$ has the Poisson distribution with mean $\gamma/2^m$, $m \in \mathbb{Z}$. With i standing for the imaginary unit, the limiting characteristic function is

$$\boldsymbol{g}_\gamma(t) = \boldsymbol{E}(\mathrm{e}^{\mathrm{i}tW_\gamma}) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx}\, \mathrm{d}G_\gamma(x) = \mathrm{e}^{y_\gamma(t)}, \qquad t \in \mathbb{R}, \tag{5}$$

where

$$y_\gamma(t) = \mathrm{i}t \operatorname{Log}\frac{1}{\gamma} + \sum_{k=0}^{-\infty}\left(\exp\left\{\frac{\mathrm{i}t2^k}{\gamma}\right\} - 1 - \frac{\mathrm{i}t2^k}{\gamma}\right)\frac{\gamma}{2^k} + \sum_{k=1}^{\infty}\left(\exp\left\{\frac{\mathrm{i}t2^k}{\gamma}\right\} - 1\right)\frac{\gamma}{2^k}.$$

Since $\gamma_{2^k} = 1$ for all $k \in \mathbb{N}$, we see that $G(\cdot) = G_1(\cdot)$ for Martin-Löf's limiting distribution. In general, for every $\gamma \in (\frac{1}{2}, 1]$, $\boldsymbol{E}(|W_\gamma|^\beta) < \infty$ for $\beta \in (0, 1)$ but $\boldsymbol{E}(|W_\gamma|) = \infty$, and it can be shown that $G_\gamma(\cdot)$ is positive and strictly increasing over the whole real line, while it is shown in Lemma 3 of Csörgő (2002) that $G_\gamma(\cdot)$ is infinitely many times differentiable on $\mathbb{R}$, with all derivatives vanishing at $\pm\infty$. In particular, the p-quantile $Q_\gamma(p) = G_\gamma^{-1}(p)$ is unique, $G_\gamma(Q_\gamma(p)) = p$, for all $p \in (0, 1)$. Furthermore, it follows from the standard Lévy form of the infinitely divisible $\boldsymbol{g}_\gamma(\cdot)$, given in Csörgő (2002), that all these subsequential limiting distributions of different types are semistable with exponent 1. There is no limiting distribution for S_n because there are very many.

The trouble with having many asymptotic distributions is resolved by a merging approximation constructed from them: since the class $\mathcal{G} = \{G_\gamma(\cdot) : \frac{1}{2} < \gamma \leqslant 1\}$ of subsequential limiting distributions is indexed by the subsequential circular limits of $\gamma_n = n/2^{\lceil \operatorname{Log} n\rceil}$, with the parameter γ_n describing the location of $n = \gamma_n 2^{\lceil \operatorname{Log} n\rceil}$ between two consecutive powers of 2, it is reasonable to expect that $\boldsymbol{P}\{n^{-1}S_n - \operatorname{Log} n \leqslant x\}$ and $G_{\gamma_n}(x)$ merge together for all $x \in \mathbb{R}$ as $n \to \infty$ along the entire sequence $\mathbb{N}$. In fact, this happens at a fast rate: the special case $p = \frac{1}{2}$ of the case $\alpha = 1$ of Theorem 1 in Csörgő (2002) states that for every $\varepsilon > 0$ there is a threshold $n_\varepsilon \in \mathbb{N}$ such that

$$\sup_{x\in\mathbb{R}}\left|\boldsymbol{P}\left\{\frac{S_n}{n}-\operatorname{Log} n\leqslant x\right\}-G_{\gamma_n}(x)\right|\leqslant(1+\varepsilon)\,\frac{\pi}{8}\,\frac{\operatorname{Log}^2 n}{n} \tag{6}$$

whenever $n \geqslant n_\varepsilon$. It is thought that the rate $O((\operatorname{Log} n)^2/n)$ here is best possible.

Concerning Feller's $n \operatorname{Log} n$ ducats from (3), we have $\boldsymbol{P}\{S_n \leqslant n \operatorname{Log} n\} \approx G_{\gamma_n}(0)$ by (6), and it turns out that $0.2070 \leqslant G_\gamma(0) \leqslant 0.2073$ for all $\gamma \in (\frac{1}{2}, 1]$. With this entrance fee, not only may Paul win hugely, but his approximate probability of winning the series is between 0.7927 and 0.7930. So, clearly, Feller's price is not enough for Peter, who cannot win more than $n(\operatorname{Log} n - 2)$ ducats. Since the left-hand side of (6) may be written as $\sup_{0<p<1}|\boldsymbol{P}\{S_n \leqslant n(Q_{\gamma_n}(p) + \operatorname{Log} n)\} - p|$, if Peter and Paul can agree that Peter should win the series of n games with probability $p = p_n \in (\frac{1}{2}, 1)$, then, attached to this contract, the fair price for Paul to pay for the series is $n(Q_{\gamma_n}(p) + \operatorname{Log} n)$ ducats. There is no equity, to answer Nicolaus Bernoulli's question, but fairness is still possible! The superficially appealing median choice $p = \frac{1}{2}$ ('half the time you win, half the time I win', Paul would argue), for which $2.5844 \leqslant Q_\gamma(\frac{1}{2}) \leqslant 2.6050$ for all $\gamma \in (\frac{1}{2}, 1]$, will still be very far from satisfying for most sane Peters. Noting here only that $Q_\gamma(p)$ is a more violent function of γ for larger values of p, the negotiations in the bargaining process to secure an acceptable value of p will be helped by an extensive mix of graphs and tabular data in Csörgő and Simons (2008).

The whole problem comes from the explosion of expectation in (1) 293 years ago. Our refinement or correction of Feller's variable entrance fee of $\operatorname{Log} n$ ducats to $Q_{\gamma_n}(p) + \operatorname{Log} n$ ducats per game, for some contracted $p \in (\frac{1}{2}, 1)$, is from an asymptotic approximation. One of the aims of the present paper is to show that the necessity of variable entrance fees per game may in fact be established for each fixed $n \geqslant 2$, using nothing more than the notion of expectation itself, a slight extension of the usual form of which may result in a finite mean for the difference of random variables that individually may have infinite expectations in the usual sense.

2. Two Pauls, three Pauls and more: comparison of infinite expectations

The *two-Paul problem*, introduced in Csörgő and Simons (2002), arises when Peter agrees to play exactly one St Petersburg game with each of two players, Paul_1 and Paul_2. Are Paul_1 and Paul_2 better off accepting their individual winnings, X_1 and X_2, say, or agreeing, before they play, to divide their total winnings in half, so that each receives $\frac{1}{2}X_1 + \frac{1}{2}X_2$? Surprisingly, this *averaging strategy* is demonstrably better for both Pauls than the *individualistic strategy* of accepting their individual winnings – the validation of this assertion turning on the observation that the pooled winnings, $\frac{1}{2}X_1 + \frac{1}{2}X_2$, are stochastically larger than the individual winnings, X_1 and X_2. More specifically, we showed in Proposition 1.1 in Csörgő and Simons (2002) the distributional equality

$$X_1 + X_2 = S_2 \overset{\mathcal{D}}{=} T_2 = 2X_1 + X_2 I\{X_2 \leqslant X_1\}, \tag{7}$$

where $I\{A\}$ is the indicator of the event $A \in \mathcal{A}$. (An almost sure version of (7) is presented in Lemma 5 below.) Thus, $T_2 \geqslant 2X_1$, with a strict inequality holding when $X_2 \leqslant X_1$. So, clearly, Paul_1 should prefer $T_2/2$ to X_1, and, consequently, prefer $S_2/2 = \frac{1}{2}X_1 + \frac{1}{2}X_2$ to X_1. Likewise, Paul_2 should prefer $\frac{1}{2}X_1 + \frac{1}{2}X_2$ to X_2.

We refer to this surprising phenomenon as the *two-Paul paradox*. As with the original one, the heart of this paradox is the *infinite* expectation appearing in (1): were the expectation finite, then X_1 and $\frac{1}{2}X_1 + \frac{1}{2}X_2$ would share the same *finite* expectation – thereby precluding the possibility that one is stochastically larger than the other.

How much better is the averaging strategy? Since, with empty sums understood as 0,

$$\boldsymbol{P}\{X_1 \geqslant 2^k | X_2 = 2^k\} = \boldsymbol{P}\{X_1 \geqslant 2^k\} = 1 - \sum_{j=1}^{k-1} \left(\frac{1}{2}\right)^j = 1 - \left[1 - \frac{1}{2^{k-1}}\right] = \frac{2}{2^k}$$

for all $k \in \mathbb{N}$, and hence $\boldsymbol{P}\{X_1 \geqslant X_2 \,|\, X_2\} = 2/X_2$, we see that

$$\boldsymbol{E}(X_2\, I\{X_2 \leqslant X_1\}) = \boldsymbol{E}(X_2 \boldsymbol{P}\{X_1 \geqslant X_2 | X_2\}) = \boldsymbol{E}\left(X_2 \frac{2}{X_2}\right) = 2. \tag{8}$$

Thus, in this precise sense, the averaging strategy provides an extra ducat $(2/2 = 1)$ of *added value* for each of Paul_1 and Paul_2 in comparison with their individualistic strategy.

More than a charming mathematical curiosity, this example suggests a precise comparison of infinite expectations more generally. Indeed, this is possible for many pairs of random variables ar U and V through the use of the *comparison operator*

$$\boldsymbol{E}[U, V] = \int_{-\infty}^{\infty} [\boldsymbol{P}\{U > x\} - \boldsymbol{P}\{V > x\}]\, \mathrm{d}x. \tag{9}$$

Both versions of the integral in (9) are considered in Csörgő and Simons (2002), Lebesgue and improper Riemann, and the calculus of these operators is described there in Theorem 2.2. One finds, as one would want, that $\boldsymbol{E}[U, V] = \boldsymbol{E}(U) - \boldsymbol{E}(V)$, whenever $\boldsymbol{E}(U)$ and $\boldsymbol{E}(V)$ are defined with at least one of them finite, with the conventions $\pm\infty - c = \pm\infty = c - \mp\infty$ for a finite c. In Csörgő and Simons (2002) the original ideas are attributed to Fréchet and Hoeffding with references, and both referees pointed out that (9) is a special case of a pseudomoment in Zolotarev (1978). Of course, if the Lebesgue version is defined, then so is the improper Riemann, but the examples constructed in Csörgő and Simons (2002) show that the converse is not true in general. The two interpretations are of course the same if the integrand in (9) is non-negative for all $x \in \mathbb{R}$, in particular when U is stochastically larger then V.

Consistent with (8), it is shown in Csörgő and Simons (2002) that $\boldsymbol{E}[\frac{1}{2}X_1 + \frac{1}{2}X_2, X_1] = 1$. Thus the comparison operator defined in (9) properly evaluates the added value, one ducat, secured by each of Paul_1 and Paul_2 when they agree to adopt the averaging strategy described above. In fact, we were able to extend this for the average S_n/n of independent winnings $X_1, X_2, \ldots, X_n$ of $n = 2^k$ Pauls and prove in Proposition 3.1 that $\boldsymbol{E}[S_{2^k}/2^k, X_1] = k$ for every $k \in \mathbb{N}$. Furthermore, it was possible to continue from here and finally arrive in Csörgő and Simons (2002) at Proposition 3.3 and Theorem 5.2, which, accompanying (6), state that

$$\boldsymbol{E}\left[\frac{S_n}{n} - \operatorname{Log} n,\, W_{\gamma_n}\right] = 0, \qquad \text{for all } n \in \mathbb{N}, \tag{10}$$

in the Lebesgue sense, and for specially constructed distributionally equivalent copies $W_{\gamma_n}^{[n]} \stackrel{\mathcal{D}}{=} W_{\gamma_n}$, defined on a rich enough probability space,

$$\boldsymbol{E}\left(\left|\left\{\frac{S_n}{n} - \operatorname{Log} n\right\} - W_{\gamma_n}^{[n]}\right|\right) = O\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right). \tag{11}$$

The statements (6), (10) and (11) all become special cases of results in the present paper.

However, the *three-Paul problem* is more complicated, and more interesting. Suppose Peter agrees to play one St Petersburg game with each of three Pauls: Paul_1, Paul_2 and Paul_3. Denote their respective individual winnings by X_1, X_2 and X_3, and let $S_3 = X_1 + X_2 + X_3$. How does the averaging strategy, which yields $S_3/3$ for each of the Pauls, compare with the individualistic strategy, yielding X_1, X_2 and X_3, respectively? Surprisingly, they are *incomparable* in the sense that the integral

$$\boldsymbol{E}\left[\frac{S_3}{3},\, X_1\right] = \int_0^\infty \left[\boldsymbol{P}\left\{\frac{S_3}{3} > x\right\} - \boldsymbol{P}\{X_1 > x\}\right] \mathrm{d}x \tag{12}$$

is not defined even as an improper Riemann integral. What one finds is that the (Lebesgue) integrals of the positive and negative parts of the integrand in (12) are each equal to $+\infty$. Necessarily, $S_3/3$ cannot be stochastically larger or stochastically smaller than $X_1 : \boldsymbol{P}\{X_1 = 2\} = 0.5$ while $\boldsymbol{P}\{S_3/3 = 2\} = 0.125 < 0.5$, and $\boldsymbol{P}\{X_1 < 8\} = 0.75$ while $\boldsymbol{P}\{S_3/3 < 8\} = 0.761\,718\,75 > 0.75$. Indeed, the plots of the distribution functions of $S_3/3$ and X_1 cross over each other infinitely often, and it is possible to demonstrate that

$$\int_0^b \left[\boldsymbol{P}\left\{\frac{S_3}{3} > x\right\} - \boldsymbol{P}\{X_1 > x\}\right] \mathrm{d}x = \operatorname{Log} 3 + \delta(b) - \delta(3b) + o(1), \qquad \text{as } b \to \infty, \tag{13}$$

where

$$\delta(s) = 1 + \langle \operatorname{Log} s \rangle - 2^{\langle \operatorname{Log} s \rangle}, \qquad s > 0. \tag{14}$$

One finds that $\delta(s)$ is a non-negative periodic function in the transformed variable $u = \operatorname{Log} s$, and that it assumes the value 0 if and only if u is integer valued, that is, if and only if s is an integer power of 2. Since this function plays a role in the main results, we include its graphs in Figure 1. The limit supremum of the integral in (13) as $b \to \infty$ is $1\frac{2}{3}$ and the limit infimum is $1\frac{1}{2}$. So, to conclude: there seems to be no rational justification for the three Pauls to use the averaging strategy. This strategy is incapable of providing added value.

But there are two other pooling strategies for the three Pauls, investigated in Csörgő and Simons (2002), that do yield added value. The simplest calls for each Paul to give all of his winnings to the other two Pauls, half to each. Under this strategy, Paul_1 ends up with $\frac{1}{2}X_2 + \frac{1}{2}X_3$, Paul_2 with $\frac{1}{2}X_1 + \frac{1}{2}X_3$, and Paul_3 with $\frac{1}{2}X_1 + \frac{1}{2}X_2$. Analogous to averaging in the two-Paul problem, this strategy provides one ducat of added value for each of the three Pauls. The second strategy is for each Paul to share one-half of his winnings evenly with the other two Pauls. Under this strategy, Paul_1 ends up with $\frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3$, Paul_2 with

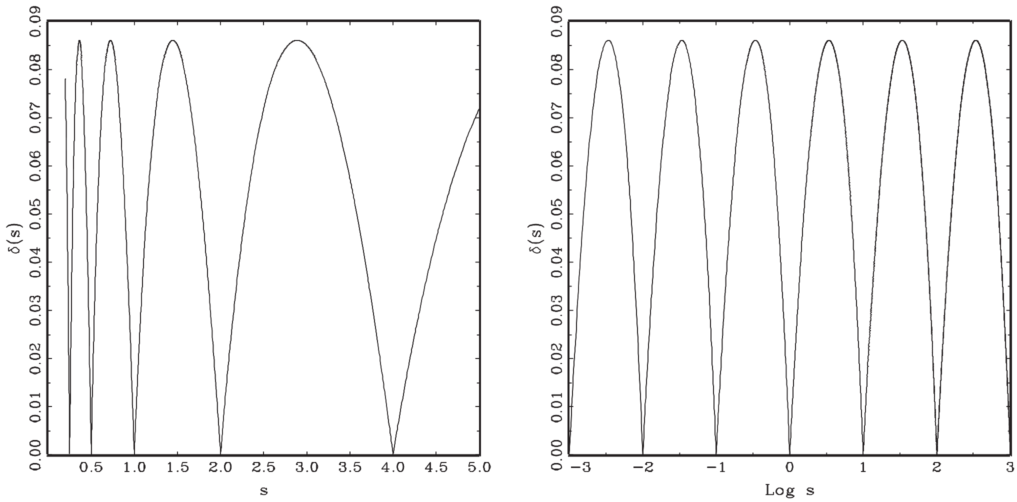

Figure 1. The function $\delta(s)$, $s > 0$.

$\frac{1}{2}X_2 + \frac{1}{4}X_1 + \frac{1}{4}X_3$, and Paul$_3$ with $\frac{1}{2}X_3 + \frac{1}{4}X_1 + \frac{1}{4}X_2$. This strategy provides $1\frac{1}{2}$ ducats of added value for each of the three Pauls.

More generally, let $\boldsymbol{p} = (p_1, p_2, p_3)$ denote an arbitrary vector with non-negative components adding to unity, and consider the pooling strategy which earns Paul$_1$ the amount $p_1X_1 + p_2X_2 + p_3X_3$, Paul$_2$ the amount $p_3X_1 + p_1X_2 + p_2X_3$, and Paul$_3$ the amount $p_2X_1 + p_3X_2 + p_1X_3$. We see that the averaging strategy $\boldsymbol{p} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is not comparable to the individualistic strategy (1, 0, 0), while the pooling strategies $\boldsymbol{p} = (0, \frac{1}{2}, \frac{1}{2})$ and $\boldsymbol{p} = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ are comparable and provide 1 and $1\frac{1}{2}$ extra ducats of added value, respectively. What remains to be determined is the set of comparable vectors $\boldsymbol{p} = (p_1, p_2, p_3)$, and then the $\boldsymbol{p}$ in this set that maximizes the comparison operator $A(\boldsymbol{p}) = \boldsymbol{E}[p_1X_1 + p_2X_2 + p_3X_3, X_1]$. We address these same questions for n Pauls, $n \geqslant 2$.

3. Admissible pooling strategies for n Pauls and semistable approximations

We are assuming that Peter agrees to play exactly one St Petersburg game with each of n players, Paul$_1$, Paul$_2, \ldots,$ Paul$_n$, $n = 2, 3, \ldots$. Their individual winnings are $X_1, X_2, \ldots, X_n$, respectively, independent copies of X as in the Introduction. The focus of attention here is on a *pooling strategy* $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$, consisting of non-negative components that sum to unity, to which all players agree before any of them plays. Under this strategy, Paul$_1$ is to receive the amount $p_{1,n}X_1 + p_{2,n}X_2 + \ldots + p_{n,n}X_n$, Paul$_2$ is to receive the amount $p_{n,n}X_1 + p_{1,n}X_2 + \ldots + p_{n-1,n}X_n$, Paul$_3$ is to receive the amount $p_{n-1,n}X_1 + p_{n,n}X_2 + p_{1,n}X_3 + \ldots + p_{n-2,n}X_n, \ldots,$ and Paul$_n$ is to receive the amount $p_{2,n}X_1 + p_{3,n}X_2 + \ldots + p_{n,n}X_{n-1} + p_{1,n}X_n$. Under these rotating assignments of weights, every bit of all of the individual winnings is distributed. Moreover, the strategy is fair to every Paul in the sense that their winnings have the same distribution and, from the exercise of the strategy, each receives the same *added value* equal to

$$A(\boldsymbol{p}_n) = \boldsymbol{E}[p_{1,n}X_1 + \ldots + p_{n,n}X_n, X_1]$$
$$= \int_0^\infty [\boldsymbol{P}\{p_{1,n}X_1 + \ldots + p_{n,n}X_n > x\} - \boldsymbol{P}\{X_1 > x\}]\,\mathrm{d}x, \tag{15}$$

whenever the integral is defined, so that a comparison is possible. Anticipating the result below, we shall call a strategy $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$ *admissible* if each of its components is either zero or an integer power of 2. Individualistic strategies are thus admissible, otherwise the powers in non-zero components are negative integers. By continuity, $p \operatorname{Log} p$ is interpreted as zero whenever $p = 0$.

Theorem 1. *The n Pauls realize the added value $A(\boldsymbol{p}_n)$ if and only if the pooling strategy $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$ is admissible. Moreover, when $\boldsymbol{p}_n$ is admissible, then the added value $A(\boldsymbol{p}_n)$ is equal to*

$$H(\boldsymbol{p}_n) = -\{p_{1,n} \operatorname{Log} p_{1,n} + \ldots + p_{n,n} \operatorname{Log} p_{n,n}\}, \tag{16}$$

the entropy of $\boldsymbol{p}_n$. Furthermore, the independent St Petersburg variables $X_1, \ldots, X_n$ can be defined on a rich enough probability space that carries, for each admissible strategy $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$, a St Petersburg random variable $X_{\boldsymbol{p}_n}$ and a non-negative random variable $Y_{\boldsymbol{p}_n}$ such that

$$p_{1,n}X_1 + \ldots + p_{n,n}X_n = X_{\boldsymbol{p}_n} + Y_{\boldsymbol{p}_n} \tag{17}$$

almost surely.

Equation (17) in the third statement identifies the source for the added value: it implies that $\sum_{j=1}^n p_{j,n}X_j$ is stochastically larger than X_1 for every admissible strategy $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$, and hence the integral $A(\boldsymbol{p}_n)$ in (15) is in fact finite as a Lebesgue integral, and the added value can be thought of as arising from $Y_{\boldsymbol{p}_n}$, that is, $\boldsymbol{E}(Y_{\boldsymbol{p}_n}) = A(\boldsymbol{p}_n) = H(\boldsymbol{p}_n)$. This is indeed so, as will be pointed out after the proof of Theorem 1. All the proofs are given in the next section.

Significantly, the added value represents, simultaneously for all Pauls, a genuine *anticipated* benefit, arising solely from their agreement to use pooling strategy $\boldsymbol{p}_n$; in no way is this 'sleight of hand'. And yet, paradoxically, they (and Peter) all know that their total winnings are $S_n = X_1 + \ldots + X_n$, the same amount with or without the pooling strategy. Stated in economic terms: through cooperation, the microeconomic perspective is sweetened for all of the Pauls while the macroeconomic perspective is unaltered.

So how well can n Pauls do by pooling? And how do they pool their winnings in order to maximize their added value? These questions are addressed in the next theorem.

Theorem 2. *For every admissible strategy $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$ the entropy $H(\boldsymbol{p}_n)$ in (16) is bounded above by*

$$H_n = \lfloor \mathrm{Log}\, n \rfloor + 2^{\langle \mathrm{Log}\, n \rangle} - 1 = \mathrm{Log}\, n - \delta(n), \tag{18}$$

where $\delta(\cdot)$ *is the function defined in* (14). *Moreover, the bound* H_n *is attainable by means of the admissible strategy*

$$\boldsymbol{p}_n^* = (p_{1,n}^*, \ldots, p_{n,n}^*) = (2p_n^*, \ldots, 2p_n^*, p_n^*, \ldots, p_n^*), \qquad \textit{with } p_n^* = \frac{1}{2^{\lceil \mathrm{Log}\, n \rceil}} = \frac{\gamma_n}{n}, \tag{19}$$

where the numbers of p_n^**s and* $2p_n^*$*s are, respectively,*

$$m_1(n) = 2n - 2^{\lceil \mathrm{Log}\, n \rceil} \quad \textit{and} \quad m_2(n) = 2^{\lceil \mathrm{Log}\, n \rceil} - n\,. \tag{20}$$

Apart from reorderings of the components of $\boldsymbol{p}_n^*$ *in* (19)*, the point of maximum is unique.*

It is well known that the unrestricted maximum of the entropy $H(\boldsymbol{p}_n)$ is attained uniquely when $\boldsymbol{p}_n$ is $\boldsymbol{p}_n^{\diamond} = (1/n, \ldots, 1/n)$, and the maximum is equal to $\mathrm{Log}\, n$. Whenever n is an integer power of 2, this $\boldsymbol{p}_n^{\diamond}$ is admissible, and, correctly, the $\boldsymbol{p}_n^*$ in (19) has only p_n^*s ($m_1(n) = n$ and $m_2(n) = 0$), so that $\boldsymbol{p}_n^* = \boldsymbol{p}_n^{\diamond}$ and (18) reduces to $\mathrm{Log}\, n$, as it should. For other values of n, the unrestricted maximum $\mathrm{Log}\, n$ cannot be obtained and the strictly positive $\delta(n)$, appearing in (18), can be thought of as the 'cost' for n not being an integer power of 2. Since the function $\delta(s)$ is bounded above by $1 - (1 + \log\log 2)/\log 2 \approx 0.0861$, attained whenever $\langle \mathrm{Log}\, s \rangle = -(\log\log 2)/\log 2 \approx 0.5288$ (see Figure 1), this cost is small for all n. And for large n, the relative cost $\delta(n)/\mathrm{Log}\, n$ is negligible.

Notice also when n is not an integer power of 2 that the two different component values of the strategy appearing in (19), p_n^* and $2p_n^*$, 'straddle' the value $1/n$. So in a certain sense, the maximizing strategy $\boldsymbol{p}_n^*$ is as close to the unrestricted maximal point $(1/n, \ldots, 1/n)$ as it can be while maintaining the requirement that it be admissible. The first ten values of H_n, $n \geqslant 2$, with their maximizing strategies, are as follows:

$H_2 = 1$, $\quad \boldsymbol{p}_2^* = (\frac{1}{2}, \frac{1}{2})$, $\qquad H_7 = 2\frac{3}{4}$, $\quad \boldsymbol{p}_7^* = (\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$,

$H_3 = 1\frac{1}{2}$, $\quad \boldsymbol{p}_3^* = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, $\qquad H_8 = 3$, $\quad \boldsymbol{p}_8^* = (\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$,

$H_4 = 2$, $\quad \boldsymbol{p}_4^* = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, $\qquad H_9 = 3\frac{1}{8}$, $\quad \boldsymbol{p}_9^* = (\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16})$,

$H_5 = 2\frac{1}{4}$, $\quad \boldsymbol{p}_5^* = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$, $\qquad H_{10} = 3\frac{2}{8}$, $\quad \boldsymbol{p}_{10}^* = (\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16})$,

$H_6 = 2\frac{2}{4}$, $\quad \boldsymbol{p}_6^* = (\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$, $\qquad H_{11} = 3\frac{3}{8}$, $\quad \boldsymbol{p}_{11}^* = (\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16})$.

The evolving pattern as n grows is that every time n increases by one, a single component $1/2^j$ with smaller exponent j is replaced by two components, each equal to $1/2^{j+1}$.

The main focus now is the study of the distribution of $\sum_{k=1}^n p_{k,n} X_k$. Allowing any probability distribution $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$ for a strategy, consider

$$S_{\boldsymbol{p}_n} = \sum_{k=1}^n p_{k,n} X_k - H(\boldsymbol{p}_n) = \sum_{k=1}^n p_{k,n}(X_k + \mathrm{Log}\, p_{k,n}). \tag{21}$$

In this compact notation the centred average gain $S_n/n - \mathrm{Log}\, n$ discussed in the first two sections appears as $S_{\boldsymbol{p}_n^{\diamond}}$ for $\boldsymbol{p}_n^{\diamond} = (1/n, \ldots, 1/n)$. Parallel to this, let

$$W_{\boldsymbol{p}_n} = \sum_{k=1}^{n} p_{k,n} W_1^{(k)} - H(\boldsymbol{p}_n) = \sum_{k=1}^{n} p_{k,n}(W_1^{(k)} + \operatorname{Log} p_{k,n}), \tag{22}$$

where $W_1^{(1)}, W_1^{(2)}, \ldots$ are independent copies of Martin-Löf's (1985) generic asymptotic random variable W_1, with distribution function $G_1(x) = \boldsymbol{P}\{W_1 \leqslant x\}$, $x \in \mathbb{R}$, and characteristic function given by $\boldsymbol{g}_1(t) = \boldsymbol{E}(\mathrm{e}^{\mathrm{i}tW_1}) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx}\, \mathrm{d}G_1(x) = \mathrm{e}^{y_1(t)}$, where

$$y_1(t) = \sum_{k=0}^{-\infty} (\mathrm{e}^{\mathrm{i}t2^k} - 1 - \mathrm{i}t2^k)\frac{1}{2^k} + \sum_{k=1}^{\infty} (\mathrm{e}^{\mathrm{i}t2^k} - 1)\frac{1}{2^k}, \qquad t \in \mathbb{R}, \tag{23}$$

the special case $\gamma = 1$ of W_γ and $\boldsymbol{g}_\gamma(\cdot)$ in (4) and (5). The description of the sufficiently rich probability space in the theorem below is given in the proof of Lemma 8 in Section 4.

Theorem 3. *For every $n = 2, 3, \ldots$ and any strategy $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$,*

$$\boldsymbol{E}[S_{\boldsymbol{p}_n}, W_{\boldsymbol{p}_n}] = 0. \tag{24}$$

Furthermore, on a rich enough probability space there exist distributionally equivalent copies $W_{\boldsymbol{p}_n}^{[n]} \stackrel{\mathcal{D}}{=} W_{\boldsymbol{p}_n}$ such that

$$\boldsymbol{E}(|S_{\boldsymbol{p}_n} - W_{\boldsymbol{p}_n}^{[n]}|) \leqslant \frac{C_n}{\sqrt{\bar{\gamma}_n}} \sqrt{\frac{r_n}{2^{r_n}}} = C_n \sqrt{\bar{p}_n r_n}, \tag{25}$$

where the numbers $\bar{p}_n \in (0, 1)$, $r_n \in \mathbb{N}$ and $\bar{\gamma}_n \in (\frac{1}{2}, 1]$ are given as

$$\bar{p}_n = \max\{p_{1,n}, \ldots, p_{n,n}\},\ r_n = \left\lceil \operatorname{Log} \frac{1}{\bar{p}_n} \right\rceil \text{ and } \bar{\gamma} = \frac{1}{\bar{p}_n 2^{r_n}},$$

and, with the constant C greater than 16.587,

$$C_n = \frac{4\sqrt{2}}{\sqrt{2}-1} \frac{\sqrt{\bar{\gamma}_n}}{\sqrt{r_n}} + \frac{1}{2}\sum_{k=0}^{\infty} \sqrt{\frac{r_n + k}{r_n\, 2^k}} \leqslant C := \frac{4\sqrt{2}}{\sqrt{2}-1} + \frac{1}{2}\sum_{k=0}^{\infty} \sqrt{\frac{k+1}{2^k}} < 16.588. \tag{26}$$

Consequently,

$$\sup_{x\in\mathbb{R}} |\boldsymbol{P}\{S_{\boldsymbol{p}_n} \leqslant x\} - \boldsymbol{P}\{W_{\boldsymbol{p}_n} \leqslant x\}| \leqslant \frac{\sqrt{2C_n}}{\bar{\gamma}_n^{1/4}} \left(\frac{r_n}{2^{r_n}}\right)^{1/4} = \sqrt{2C_n}(\bar{p}_n r_n)^{1/4}. \tag{27}$$

The proof of (25) is unusual and reveals the underlying reason for the approximation very clearly: it is the result of an infinite sequence of almost sure coupling statements in Lemma 8, in each step of which we double the number of the n gambling Pauls, and finally apply Martin-Löf's limit theorem. Lemma 9 controls the remainder term in the construction and yields the bound in (25). The bound in (27) is $\sqrt{2}$ times the square root of the bound in (25), where 2 is the reciprocal of the bound on the density in (32) below. We emphasize that the approximation in (27) holds for each fixed n and is applicable even when $\bar{p}_n \not\to 0$, as in Example 2 below; here and henceforth all asymptotic relationships are meant as $n \to \infty$ unless otherwise specified. If, however, $\bar{p}_n \to 0$, then the uniform rate of approximation in (27) is far from what may be achieved by Fourier methods. The latter is contained in the next theorem, the only genuinely asymptotic result in the paper, where the rate obtained, the order of the upper bound, is thought to be best possible.

Theorem 4. *For any sequence of strategies* $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$ *for which* $\bar{p}_n = \max\{p_{1,n}, \ldots, p_{n,n}\} \to 0$*, for every* $\varepsilon > 0$ *there is a threshold* $n_\varepsilon \in \mathbb{N}$ *such that*

$$\sup_{x\in\mathbb{R}} |\boldsymbol{P}\{S_{\boldsymbol{p}_n} \leqslant x\} - \boldsymbol{P}\{W_{\boldsymbol{p}_n} \leqslant x\}| \leqslant (1+\varepsilon)\,\frac{\pi}{8}\,\bar{p}_n \operatorname{Log}^2 \frac{1}{\bar{p}_n} \tag{28}$$

whenever $n \geqslant n_\varepsilon$.

To elucidate the contents of Theorems 3 and 4, we observe that the characteristic function of the approximating distribution function $G_{\boldsymbol{p}_n}(x) := \boldsymbol{P}\{W_{\boldsymbol{p}_n} \leqslant x\}$, $x \in \mathbb{R}$, is

$$\begin{aligned} \boldsymbol{g}_{\boldsymbol{p}_n}(t) := \boldsymbol{E}(\mathrm{e}^{\mathrm{i}tW_{\boldsymbol{p}_n}}) &= \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}G_{\boldsymbol{p}_n}(x) = \prod_{k=1}^{n} \mathrm{e}^{\mathrm{i}tp_{k,n}\operatorname{Log} p_{k,n}}\,\boldsymbol{g}_1(p_{k,n}t) \\ &= \prod_{k=1}^{n} \mathrm{e}^{\mathrm{i}tp_{k,n}\operatorname{Log} p_{k,n} + y_1(p_{k,n}t)}, \qquad t \in \mathbb{R}, \end{aligned} \tag{29}$$

where $\boldsymbol{g}_1(\cdot) = \mathrm{e}^{y_1(\cdot)}$ is Martin-Löf's characteristic function with $y_1(\cdot)$ given in (23). The semistable nature of the distribution of W_1 is expressed by the 'scaling law' found by Martin-Löf (1985) in his Theorem 2:

$$y_1(2^m s) = 2^m y_1(s) - \mathrm{i}\,2^m ms, \qquad s \in \mathbb{R}, \text{ for every } m \in \mathbb{Z},$$

which can be checked directly from (23). Setting

$$r_{k,n} = \left\lceil \operatorname{Log} \frac{1}{p_{k,n}} \right\rceil \quad \text{and} \quad \gamma_{k,n} = \frac{2^{-r_{k,n}}}{p_{k,n}},$$

so that

$$2^{r_{k,n}-1} < \frac{1}{p_{k,n}} \leqslant 2^{r_{k,n}}$$

and hence $p_{k,n} = 2^{-r_{k,n}}/\gamma_{k,n}$ and $\gamma_{k,n} = 2^{-\langle \log p_{k,n} \rangle} \in (\frac{1}{2}, 1]$ for all $k \in \{1, \ldots, n\}$ for which $p_{k,n} > 0$, and using this scaling law, for every $t \in \mathbb{R}$ we obtain

$$\begin{aligned} \boldsymbol{g}_{\boldsymbol{p}_n}(t) &= \prod_{\{1\leqslant k\leqslant n\,:\,p_{k,n}>0\}} \exp\left\{\mathrm{i}t\,\frac{2^{-r_{k,n}}}{\gamma_{k,n}}\left[-r_{k,n} + \operatorname{Log}\frac{1}{\gamma_{k,n}}\right] + y_1\left(2^{-r_{k,n}}\,\frac{t}{\gamma_{k,n}}\right)\right\} \\ &= \exp\left\{\sum_{\{1\leqslant k\leqslant n\,:\,p_{k,n}>0\}} \left[\frac{\mathrm{i}t}{2^{r_{k,n}}}\,\frac{1}{\gamma_{k,n}}\operatorname{Log}\frac{1}{\gamma_{k,n}} + \frac{y_1(t/\gamma_{k,n})}{2^{r_{k,n}}}\right]\right\} \\ &= \exp\left\{\sum_{k=1}^{n} p_{k,n}\left[\mathrm{i}t\operatorname{Log}\frac{1}{\gamma_{k,n}} + \gamma_{k,n}\,y_1\left(\frac{t}{\gamma_{k,n}}\right)\right]\right\} = \exp\left\{\sum_{k=1}^{n} p_{k,n} y_{\gamma_{k,n}}(t)\right\}, \end{aligned} \tag{30}$$

comparing (5) and (23). For the real part $\operatorname{Re} y_1(\cdot)$ of $y_1(\cdot)$ we know from (4) in Csörgő (2002) that

$$\operatorname{Re} y_1(t) \leqslant -\frac{2}{\pi}|t|, \qquad \text{for all } t \in \mathbb{R}, \tag{31}$$

so that the function $t \mapsto |t|^r |\boldsymbol{g}_{p_n}(t)|$ is integrable over $\mathbb{R}$ for any power $r \in \mathbb{N}$, and hence, by a standard result in Fourier analysis, the distribution function $G_{\boldsymbol{p}_n}(\cdot)$ is infinitely many times differentiable on $\mathbb{R}$, with all derivatives vanishing at $\pm\infty$. In particular, by the density inversion theorem,

$$\sup_{x\in\mathbb{R}} |G'_{\boldsymbol{p}_n}(x)| \leqslant \frac{1}{2\pi}\int_{-\infty}^{\infty} |\boldsymbol{g}_{\boldsymbol{p}_n}(t)|\,\mathrm{d}t \leqslant \frac{1}{\pi}\int_0^{\infty} \exp\left\{-\frac{2}{\pi}\sum_{k=1}^{n} p_{k,n}\frac{t\gamma_{k,n}}{\gamma_{k,n}}\right\}\mathrm{d}t = \frac{1}{\pi}\int_0^{\infty} \mathrm{e}^{-2t/\pi}\,\mathrm{d}t = \frac{1}{2} \tag{32}$$

for the corresponding density function $G'_{\boldsymbol{p}_n}(\cdot)$, uniformly for all strategies $\boldsymbol{p}_n$.

Now, for an admissible strategy $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$, $\gamma_{1,n} = \ldots = \gamma_{n,n} = 1 = \bar{\gamma}_n$, so the characteristic function $\boldsymbol{g}_{\boldsymbol{p}_n}(\cdot)$ in (30) becomes $\boldsymbol{g}_{\boldsymbol{p}_n}(t) = \exp\{y_1(t)\sum_{k=1}^n p_{k,n}\} = \mathrm{e}^{y_1(t)} = \boldsymbol{g}_1(\mathrm{t}) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}G_1(x)$, $t \in \mathbb{R}$. This implies the equality $W_{\boldsymbol{p}_n} \stackrel{\mathcal{D}}{=} W_1$ in distribution, meaning that $G_{\boldsymbol{p}_n}(\cdot) = G_1(\cdot)$. In particular, as perhaps the biggest mathematical surprise so far in this disquisition, we see that whenever $\bar{p}_n \to 0$, admissible winnings have a proper asymptotic distribution in the classical sense as $n \to \infty$ *along the entire sequence* $\mathbb{N}$, *always* converging to Martin-Löf's (1985) generic asymptotic random variable W_1.

Corollary 1. *If $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$ is any admissible strategy for $n = 2, 3, \ldots,$ then $\boldsymbol{E}[S_{\boldsymbol{p}_n}, W_1] = 0$, inequality* (25) *holds with $W^{[n]}_{\boldsymbol{p}_n} \stackrel{\mathcal{D}}{=} W_1$ and $\bar{\gamma}_n = 1$, and*

$$\sup_{x\in\mathbb{R}} |\boldsymbol{P}\{S_{\boldsymbol{p}_n} \leqslant x\} - G_1(x)| \leqslant \sqrt{2C_n}\left(\frac{r_n}{2^{r_n}}\right)^{1/4},$$

where $r_n \in \mathbb{N}$ is defined by the equation $\max\{p_{1,n}, \ldots, p_{n,n}\} = 2^{-r_n}$ and, with $\bar{\gamma}_n = 1$ in it, the constant C_n and $C \geqslant C_n$ are as in (26)*. Furthermore, if $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$ is any sequence of admissible strategies such that $\bar{p}_n = \max\{p_{1,n}, \ldots, p_{n,n}\} \to 0$, then for every $\varepsilon > 0$ there is a threshold $n_\varepsilon \in \mathbb{N}$ such that*

$$\sup_{x\in\mathbb{R}} |\boldsymbol{P}\{S_{\boldsymbol{p}_n} \leqslant x\} - G_1(x)| \leqslant (1+\varepsilon)\,\frac{\pi}{8}\;\bar{p}_n \operatorname{Log}^2\frac{1}{\bar{p}_n} = (1+\varepsilon)\,\frac{\pi}{8}\,\frac{r_n^2}{2^{r_n}}, \tag{33}$$

whenever $n \geqslant n_\varepsilon$.

Of course, if $l_n = 2^{r_n}$ foolish executive Pauls force $\boldsymbol{p}^\bullet_n$ with $p^\bullet_{1,n} = \ldots = p^\bullet_{l_n,n} = 2^{-r_n-1}$, $p^\bullet_{l_n+1,n} = \ldots = p^\bullet_{l_n+m_n,n} = 2^{-s_n-1}$, where $m_n = 2^{s_n}$ with $s_n = \lfloor \operatorname{Log}(n - l_n) \rfloor$, and $p^\bullet_{l_n+m_n+1,n} = \ldots = p^\bullet_{n,n} = 0$, then they not only reduce the expectations of all $n \geqslant 65\,536$ Pauls to only $H(\boldsymbol{p}^\bullet_n) = 1 + 2^{-1}(r_n + s_n)$ extra ducats, but also make the rate in (33) very slow when $r_n = \lfloor \operatorname{Log}\operatorname{Log}\operatorname{Log}\operatorname{Log} n \rfloor$.

However, if n wise Pauls use the best admissible strategy $\boldsymbol{p}^*_n = (p^*_{1,n}, \ldots, p^*_{n,n})$ from Theorem 2, then $\bar{p}^*_n = \max\{p^*_{1,n}, \ldots, p^*_{n,n}\} = 2\gamma_n/n$, and so

$$\sup_{x\in\mathbb{R}}|\boldsymbol{P}\{S_{\boldsymbol{p}_n^*} \leqslant x\} - G_1(x)| \leqslant (1+\varepsilon)\,\frac{\pi\,\gamma_n}{4}\,\frac{\operatorname{Log}^2 n}{n}, \qquad n \geqslant n_\varepsilon, \tag{34}$$

with the best admissible entropy $H(\boldsymbol{p}_n^*) = \operatorname{Log} n - \delta(n)$ in $S_{\boldsymbol{p}_n^*} = \sum_{k=1}^n p_{k,n}^* X_k - H(\boldsymbol{p}_n^*)$.

On the other hand, if $\boldsymbol{p}_n$ is $\boldsymbol{p}_n^\diamond = (1/n, \ldots, 1/n)$, the generally inadmissible uniform averaging strategy for every $n = 2, 3, \ldots,$ then $r_{1,n} = \ldots = r_{n,n} = \lceil \operatorname{Log} n \rceil$ and $\gamma_{1,n} = \ldots = \gamma_{n,n} = \bar{\gamma}_n = n/2^{\lceil \operatorname{Log} n \rceil} = \gamma_n$, so that, again by (30), $\boldsymbol{g}_{\boldsymbol{p}_n^\diamond}(\cdot) = \mathrm{e}^{y_{\gamma_n}(\cdot)} = \boldsymbol{g}_{\gamma_n}(\cdot)$. Thus $W_{\boldsymbol{p}_n^\diamond} \stackrel{\mathcal{D}}{=} W_{\gamma_n}$, and all the results in (10), (11) and (6) follow from Theorems 3 and 4.

The merge theorem in (6) is of mutual interest to Peter and the n Pauls. If they agree on $p = p_n \in (\frac{1}{2}, 1)$, Peter's winning probability for the series of n games, then Peter expects for his net gain, and the n Pauls expect for their joint net loss, that

$$\boldsymbol{P}\{n[Q_{\gamma_n}(p) + \operatorname{Log} n] - S_n \geqslant x\} = \boldsymbol{P}\left\{\frac{S_n}{n} - \operatorname{Log} n \leqslant Q_{\gamma_n}(p) - \frac{x}{n}\right\} \approx G_{\gamma_n}\left(Q_{\gamma_n}(p) - \frac{x}{n}\right)$$

for every $x \in \mathbb{R}$, which is p for $x = 0$, where the order of approximation is $O((\operatorname{Log} n)^2/n)$, since $\{Q_{\gamma_n}(p_n)\}$ is a bounded sequence if $0 < \liminf_{n\to\infty} p_n \leqslant \limsup_{n\to\infty} p_n < 1$. But now each of $\mathrm{Paul}_1, \ldots, \mathrm{Paul}_n$, using the best admissible strategy among themselves, is also interested in the distribution of his personal net winning in his one game, for which

$$\begin{aligned}\boldsymbol{P}\left\{\sum_{k=1}^n p_{k,n}^* X_k - [Q_{\gamma_n}(p) + \operatorname{Log} n] > x\right\} &= \boldsymbol{P}\left\{\sum_{k=1}^n p_{k,n}^* X_k - H(\boldsymbol{p}_n^*) > x + Q_{\gamma_n}(p) + \delta(n)\right\}\\ &\approx 1 - G_1(x + Q_{\gamma_n}(p) + \delta(n))\end{aligned}$$

for every $x \in \mathbb{R}$ by (34), where, again, the order of approximation is $O((\operatorname{Log} n)^2/n)$. Peter, representing the bank or the insurance company, may not be interested in the fine structure of cooperation on the other side. However, in order to facilitate the possibility of an unequal but fair resolution, and a sense of comfort among the customers, he must tell each Paul not only the number of other participants, but also their names and (e-mail) addresses. Numerical and graphical illustrations will appear in Csörgő and Simons (2008).

We consider two more illustrative examples as special cases of Theorems 3 and 4.

Example 1. Suppose that $n - n/2$ bright Pauls are unable to convince otherwise the remaining $n/2$ dumb Pauls, who collectively want to use the halved average strategy $\boldsymbol{p}_{\lfloor n/2 \rfloor}^\diamond/2$, so that the bright Pauls are forced to use the halved best-admissible strategy $\boldsymbol{p}_{n-\lfloor n/2 \rfloor}^*/2$ only among themselves. Then $\boldsymbol{g}_{\boldsymbol{p}_n}(t) = \exp\{y_1(t)/2\}\exp\{y_{\gamma_{\lfloor n/2 \rfloor}}(t)/2\}$, $t \in \mathbb{R}$, for the united strategy $\boldsymbol{p}_n$ of the n Pauls, so that the distribution of $W_{\boldsymbol{p}_n}$ is a convolution of 'spectrally halved' versions of those in Corollary 1 and (6) with the adjusted $\gamma_{\lfloor n/2 \rfloor}$.

Example 2. Consider the strategy $\boldsymbol{p}_n = (p, (1-p)/(n-1), \ldots, (1-p)/(n-1))$ for some $p \in (0, 1)$ and $n > 1/p$. This will be popular for n Pauls who are cautious but firmly trust their own luck, and may be viewed as a generally inadmissible extension of $\boldsymbol{p}_3 = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ in Section 2. Here we obtain $\boldsymbol{g}_{\boldsymbol{p}_n}(t) = \exp\{p y_{\gamma(p)}(t)\}\exp\{(1-p) y_{\gamma(p,n)}(t)\}$, $t \in \mathbb{R}$, for $\gamma(p) = 2^{-\lceil \operatorname{Log}(1/p) \rceil}/p$ and $\gamma(p, n) = (n-1)2^{-\lceil \operatorname{Log}((n-1)/(1-p)) \rceil}/(1-p)$, so that the distribution of $W_{\boldsymbol{p}_n}$ is again a convolution of two spectrally rescaled members of the class $\mathcal{G}$ discussed in Section 1. If $p = 1/6\,500\,000\,000$, for example, then, calculating the constant in (27) exactly, $\sup_{x\in\mathbb{R}}|\boldsymbol{P}\{S_{\boldsymbol{p}_n} \leqslant x\} - \boldsymbol{P}\{W_{\boldsymbol{p}_n} \leqslant x\}| < 0.0246$ for all $n \geqslant 6\,500\,000\,001$, that is, for all St Petersburg gamblers on Earth at this moment.

Finally, with stochastic compactness of sequences $\{S_{\boldsymbol{p}_n}\}$ in mind, notice that

$$\boldsymbol{g}_{\boldsymbol{p}_n}(t) = \exp\left\{\int_{1/2}^{1}\left[\mathrm{i}t\,\mathrm{Log}\,\frac{1}{\gamma} + \gamma y_1\left(\frac{t}{\gamma}\right)\right]\mathrm{d}T_{\boldsymbol{p}_n}(\gamma)\right\} = \exp\left\{\int_{1/2}^{1} y_\gamma(t)\,\mathrm{d}T_{\boldsymbol{p}_n}(\gamma)\right\} \tag{35}$$

for all $t \in \mathbb{R}$ by (30) and (5), for every strategy $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$, where $T_{\boldsymbol{p}_n}(\cdot)$ is a distribution function on $(\frac{1}{2}, 1]$ defined by

$$T_{\boldsymbol{p}_n}(\gamma) = \sum_{\{1 \leqslant k \leqslant n:\ p_{k,n}>0\}} p_{k,n} I\{\gamma_{k,n} \leqslant \gamma\}, \qquad \frac{1}{2} < \gamma \leqslant 1, \tag{36}$$

since $\gamma_{k,n} = 2^{-\langle \mathrm{Log}\, p_{k,n}\rangle} \in (\frac{1}{2}, 1]$ when $p_{k,n} > 0$, hereafter called the *parameter distribution function* associated with $S_{\boldsymbol{p}_n}$. In turn, since the right-hand-side Lévy function, appearing in the canonical Lévy form (Gnedenko and Kolmogorov 1954: 83–84) of $\boldsymbol{g}_\gamma(\cdot)$ in (5) is $R_\gamma(x) = -2^{\langle \mathrm{Log}\,(\gamma x)\rangle}/x$, $x > 0$, for each $\gamma \in (\frac{1}{2}, 1]$, it follows from the details in Csörgő (2002: 823–824) that the Lévy form of $\boldsymbol{g}_{\boldsymbol{p}_n}(\cdot)$ in (35), for the infinitely divisible distribution of $W_{\boldsymbol{p}_n}$, assumes the form

$$\boldsymbol{g}_{\boldsymbol{p}_n}(t) = \exp\left\{\mathrm{i}t\int_{1/2}^{1}\left(u_\gamma + \mathrm{Log}\,\frac{1}{\gamma}\right)\mathrm{d}T_{\boldsymbol{p}_n}(\gamma) + \int_0^\infty\left(\mathrm{e}^{\mathrm{i}tx} - 1 - \frac{\mathrm{i}tx}{1+x^2}\right)\mathrm{d}R_{\boldsymbol{p}_n}(x)\right\}, \qquad t \in \mathbb{R},$$

with $u_\gamma = \sum_{j=1}^\infty \gamma^2/(\gamma^2 + 4^j) - \sum_{j=0}^\infty 1/(1 + \gamma^2 4^j)$ and Lévy function

$$R_{\boldsymbol{p}_n}(x) = \sum_{k=1}^{n} p_{k,n} R_{\gamma_{k,n}}(x) = \int_{1/2}^{1} R_\gamma(x)\,\mathrm{d}T_{\boldsymbol{p}_n}(\gamma) = -\frac{1}{x}\int_{1/2}^{1} 2^{\langle \mathrm{Log}\,(\gamma x)\rangle}\,\mathrm{d}T_{\boldsymbol{p}_n}(\gamma) =: -\frac{L_{\boldsymbol{p}_n}(x)}{x}$$

for all $x > 0$. Since $L_{\boldsymbol{p}_n}(2x) = L_{\boldsymbol{p}_n}(x) \in [1, 2)$, $x > 0$, this distribution is in fact semistable of exponent 1; for discussion and references, see Section 2 of Csörgő (2002).

In what follows, we find it convenient and natural to work with the topology associated with circular convergence $\xrightarrow{\mathbf{cir}}$ (as described just before equation (4)) when working with functions on $(\frac{1}{2}, 1]$, such as $T_{\boldsymbol{p}_n}(\gamma)$, and let $\overset{\mathbf{cir}}{\Longrightarrow}$ denote the associated form of weak convergence for functions defined on the compact space $(\frac{1}{2}, 1]$ (under this topology). In particular, for distribution functions $T_n(\gamma)$ and $T(\gamma)$ defined on $(\frac{1}{2}, 1]$, $T_n(\cdot) \overset{\mathbf{cir}}{\Longrightarrow} T(\cdot)$ requires $T_n(b) - T_n(a) \to T(b) - T(a)$ for every pair of continuity points a and b of $T(\cdot)$ within $(\frac{1}{2}, 1]$, thereby allowing any probability mass that is moving towards $\frac{1}{2}$ as $n \to \infty$ to accumulate, in the limit, at the point 1. Moreover, let $\xrightarrow{\mathcal{D}}$ denote convergence in distribution. Then the following consequence of Theorem 4 is now an easy adaptation of Helly–Bray theory and a classical result (Gnedenko and Kolmogorov 1954: 88–91) for the convergence of infinitely divisible distributions.

Corollary 2. *Let* $\{\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})\}_{n=1}^{\infty}$ *be any sequence of strategies such that* $\bar{p}_n \to 0$. *Then for every subsequence* $\{n_j\}_{j=1}^{\infty} \subset \mathbb{N}$ *there exist a further subsequence* $\{n_{j_m}\}_{m=1}^{\infty}$ *and a distribution function* $T(\cdot)$ *on* $(\frac{1}{2}, 1]$ *such that* $T_{\boldsymbol{p}_{n_{j_m}}}(\cdot) \stackrel{\mathbf{cir}}{\Longrightarrow} T(\cdot)$ *and* $S_{\boldsymbol{p}_{n_{j_m}}} \stackrel{\mathcal{D}}{\to} W_T$ *as* $m \to \infty$, *where* W_T *is a semistable infinitely divisible random variable of exponent* 1, *the characteristic function of which is given by*

$$\boldsymbol{g}_T(t) = \boldsymbol{E}(\mathrm{e}^{\mathrm{i}tW_T}) = \exp\left\{\int_{1/2}^{1} y_\gamma(t)\,\mathrm{d}T(\gamma)\right\}$$

$$= \exp\left\{\mathrm{i}t\int_{1/2}^{1}\left(u_\gamma + \mathrm{Log}\,\frac{1}{\gamma}\right)\mathrm{d}T(\gamma) + \int_0^\infty\left(\mathrm{e}^{\mathrm{i}tx} - 1 - \frac{\mathrm{i}tx}{1+x^2}\right)\mathrm{d}R_T(x)\right\}, \qquad t \in \mathbb{R},$$

where the Lévy function is

$$R_T(x) = \int_{1/2}^{1} R_\gamma(x)\,\mathrm{d}T(\gamma) = -\frac{1}{x}\int_{1/2}^{1} 2^{\langle \mathrm{Log}\,(\gamma x)\rangle}\,\mathrm{d}T(\gamma), \qquad x > 0.$$

Moreover, $\sum_{k=1}^{n} p_{k,n}X_k/H(\boldsymbol{p}_n) \stackrel{\boldsymbol{P}}{\to} 1$ *as* $n \to \infty$ *along the entire sequence* $\mathbb{N}$.

The last statement here generalizes Feller's weak law of large numbers in (3).

Every possible distribution function $T(\cdot)$ on $(\frac{1}{2}, 1]$ occurs as the weak limit of a suitable sequence $\{T_{\boldsymbol{p}_{m_n}}(\cdot)\}_{n=1}^{\infty}$, for which $S_{\boldsymbol{p}_{m_n}} \stackrel{\mathcal{D}}{\to} W_T$. This plausible, but technically non-trivial statement is demonstrated by the following construction of Vilmos Totik: for any given $T(\cdot)$ and $\varepsilon > 0$, first we find $s = s(\varepsilon)$, $n_0 = n_0(\varepsilon) \in \mathbb{N}$, $0 = \sigma_{s+1} < \sigma_s < \ldots < \sigma_1 < \sigma_0 = 1$ and $k_0 = k_{0,n}, \ldots, k_{s+1} = k_{s+1,n} \in \mathbb{N}$, $k_0 + \ldots + k_{s+1} = 2^n$, such that for the intermediate approximation $T^*_{s,n}(\gamma) = \sum_{j=0}^{s+1} k_j 2^{-n} I\{2^{-\sigma_j} \leqslant \gamma\}$ we have $\sup_{1/2 \leqslant \gamma \leqslant 1}|T^*_{s,n}(\gamma) - T(\gamma)| \leqslant \varepsilon$ for all $n \geqslant n_0$, and then for $m = m(s, n) = m_0 + \ldots + m_{s+1} + 1$ we put $\boldsymbol{p}_m = (p_1, \ldots, p_m)$, in which the lth block of $m_l = \lfloor k_l/2^{\sigma_l}\rfloor$ elements are all equal to $2^{\sigma_l}/2^n$, $l = 0, \ldots, s+1$, while $0 \leqslant p_m = 1 - \sum_{j=1}^{m-1} p_j \leqslant 2(s+1)/2^n$. Taking $T_{\boldsymbol{p}_{m(s,n)}}(\cdot)$, pertaining to this $\boldsymbol{p}_{m(s,n)}$ in (36), one can show that $\sup_{1/2 \leqslant \gamma \leqslant 1} |T_{\boldsymbol{p}_{m(s,n)}}(\gamma) - T^*_{s,n}(\gamma)| \leqslant 4(s+1)/2^n$ for all $n \geqslant n_0$.

We conjecture not only that the parameter distribution function T is uniquely determined by the distribution of W_T, but also that the random variables $W_{T^{(1)}}$ and $W_{T^{(2)}}$ are incomparable for different parameter distribution functions $T^{(1)}$ and $T^{(2)}$.

Lastly, we point out that all the results in this paper will likely acquire a more general form that includes those generalized St Petersburg games, considered in Csörgő and Simons (1996) and Csörgő (2002), and in their earlier references, in which all of the Pauls play the game with the same possibly biased coin, for which the probability of heads is some number $p \in (0, 1)$, so that $\boldsymbol{P}\{X = r^k\} = q^{k-1}p$, $k \in \mathbb{N}$, where $q = 1 - p$ and $r = 1/q$. The generalizations are non-trivial, and a student of one of us, Péter Kevei, is working on these problems.

4. Proofs

Lemmas 2–4 are needed for the proof of the first two statements of Theorem 1, while Lemma 1 is used in the proof of Lemma 2.

Lemma 1. *If U and V are non-negative random variables for which*

$$\boldsymbol{E}(\min(U, V)) < \infty, \tag{37}$$

then, viewed as a Lebesgue integral,

$$A_2 := \int_0^\infty [\boldsymbol{P}\{U+V>x\} - \boldsymbol{P}\{U>x\} - \boldsymbol{P}\{V>x\}]\,\mathrm{d}x = 0.$$

Proof. When the use of Fubini's theorem can be justified in what follows, one has

$$\begin{aligned} A_2 &= \int_0^\infty \boldsymbol{E}(I\{U+V>x\} - I\{U>x\} - I\{V>x\})\,\mathrm{d}x \\ &= \boldsymbol{E}\left(\int_0^\infty [I\{U+V>x\} - I\{U>x\} - I\{V>x\}]\,\mathrm{d}x\right) = \boldsymbol{E}((U+V) - U - V) = 0, \end{aligned}$$

proving the statement. As to this interchange of the integral and the expectation, observe that

$$\begin{aligned} |I\{U+V>x\} - I\{U>x\} - I\{V>x\}| \leqslant{}& I\{\min(U, V)>x\} \\ &+ I\{\max(U, V) \leqslant x < U+V\} \end{aligned}$$

for all $x \geqslant 0$, so that for $Z := \int_0^\infty |I\{U+V>x\} - I\{U>x\} - I\{V>x\}|\,\mathrm{d}x$ we have

$$\begin{aligned} Z &\leqslant \int_0^\infty [I\{\min(U, V)>x\} + I\{\max(U, V) \leqslant x < U+V\}]\,\mathrm{d}x \\ &= \min(U, V) + [U+V-\max(U, V)] = 2\min(U, V). \end{aligned}$$

Thus $\boldsymbol{E}(Z) \leqslant 2\boldsymbol{E}(\min(U, V)) < \infty$ by the assumption in (37). This justifies the use of Fubini's theorem, and completes the proof. □

Lemma 2. *For every $n \in \mathbb{N}$, if $U_1, U_2, \ldots, U_n$ are non-negative random variables for which*

$$\boldsymbol{E}(\min(U_j, U_k)) < \infty, \qquad \textit{for } 1 \leqslant j, k \leqslant n, j \neq k,$$

then, viewed as a Lebesgue integral,

$$A_n := \int_0^\infty \left[\boldsymbol{P}\left\{\sum_{j=1}^n U_j > x\right\} - \sum_{j=1}^n P\{U_j > x\}\right]\mathrm{d}x = 0.$$

Proof. The proof is by induction. For $n = 1$, the assumption is vacuous and the integrand is identically zero. Thus the statement holds for $n = 1$. Next, assume that the statement holds for some $n \in \mathbb{N}$. To establish it for $n + 1$, we have

$$A_{n+1} = \int_0^\infty \left[\boldsymbol{P}\left\{ \sum_{j=1}^{n+1} U_j > x \right\} - \boldsymbol{P}\left\{ \sum_{j=1}^{n} U_j > x \right\} - \boldsymbol{P}\{U_{n+1} > x\} \right] \mathrm{d}x$$
$$+ \int_0^\infty \left[\boldsymbol{P}\left\{ \sum_{j=1}^{n} U_j > x \right\} - \sum_{j=1}^{n} \boldsymbol{P}\{U_j > x\} \right] \mathrm{d}x.$$

The last integral is zero by virtue of the induction hypothesis. So it remains to show that the penultimate integral is also equal to zero. But this follows from Lemma 1 with U and V set equal to $\sum_{j=1}^{n} U_j$ and U_{n+1} respectively, and from the observation that

$$\boldsymbol{E}(\min(U,\, V)) = \boldsymbol{E}\left(\min\left(\sum_{j=1}^{n} U_j,\, U_{n+1} \right) \right) \leqslant \boldsymbol{E}\left(\sum_{j=1}^{n} \min(U_j,\, U_{n+1}) \right)$$
$$= \sum_{j=1}^{n} \boldsymbol{E}(\min(U_j,\, U_{n+1})),$$

which is finite, so that assumption (37) of Lemma 1 holds. This completes the induction step. □

Lemma 3. *If* X_1 *and* X_2 *are independent St Petersburg random variables and* c_1 *and* c_2 *are non-negative constants, then* $\boldsymbol{E}(\min(c_1X_1,\, c_2X_2)) < \infty$.

Proof. Of course, this is obvious if either constant is zero. So assume that $c_1, c_2 > 0$. When $x \geqslant 2\max(c_1,\, c_2)$, the distribution function F appearing in (2) yields

$$\boldsymbol{P}\{\min(c_1X_1,\, c_2X_2) > x\} = \left[1 - F\left(\frac{x}{c_1}\right)\right]\left[1 - F\left(\frac{x}{c_2}\right)\right] < \frac{4c_1c_2}{x^2}.$$

Consequently, $\boldsymbol{E}(\min(c_1X_1,\, c_2X_2)) = \int_0^\infty \boldsymbol{P}\{\min(c_1X_1,\, c_2X_2) > x\}\, \mathrm{d}x$ is finite. □

Lemma 4. *If* X *is a St Petersburg random variable and* $b \geqslant 1$, *then*

$$\int_0^b \boldsymbol{P}\{X > x\}\, \mathrm{d}x = \lfloor \operatorname{Log} b \rfloor + 2^{\langle \operatorname{Log} b \rangle} = 1 + \operatorname{Log} b - \delta(b),$$

with the function $\delta(\cdot)$ *defined in* (14).

Proof. The second equality here follows directly from the definition of $\delta(\cdot)$ in (14). In view of (2), noticing that the formula $\boldsymbol{P}\{X > x\} = 1 - F(x) = 2^{-\lfloor \operatorname{Log} x \rfloor}$ still produces the correct value 1 for $1 \leqslant x < 2$, the first equality requires us to show that $\int_1^b 2^{-\lfloor \operatorname{Log} x \rfloor}\, \mathrm{d}x = \lfloor \operatorname{Log} b \rfloor + 2^{\langle \operatorname{Log} b \rangle} - 1$ for $b \geqslant 1$, that is, $\int_1^{2^c} 2^{-\lfloor \operatorname{Log} x \rfloor}\, \mathrm{d}x = \lfloor c \rfloor + 2^{\langle c \rangle} - 1$ for $c \geqslant 0$. But

$$\int_1^{2^c} 2^{-\lfloor \operatorname{Log} x\rfloor}\,\mathrm{d}x = (\log 2)\int_0^c 2^{-\lfloor y\rfloor}\,2^y\,\mathrm{d}y = (\log 2)\left\{\int_0^{\lfloor c\rfloor} 2^{\langle y\rangle}\,\mathrm{d}y + \int_{\lfloor c\rfloor}^{\lfloor c\rfloor+\langle c\rangle} 2^{\langle y\rangle}\,\mathrm{d}y\right\}$$

$$= (\log 2)\left\{\lfloor c\rfloor\int_0^1 2^y\,\mathrm{d}y + \int_0^{\langle c\rangle} 2^y\,\mathrm{d}y\right\} = \lfloor c\rfloor(2^1 - 2^0) + (2^{\langle c\rangle} - 2^0),$$

which is the desired equation, completing the proof. □

The next lemma is a basic block in the proof of the third statement of Theorem 1. It is the almost sure version of the distributional equation in (7).

Lemma 5. *The independent St Petersburg random variables X and Y can be defined on a rich enough probability space that carries another pair of independent St Petersburg random variables X' and Y' such that $X + Y = 2X' + Y'I\{Y' \leqslant X'\}$ almost surely.*

Proof. Let X and Y be defined on a probability space that carries a third St Petersburg random variable Z such that X, Y, Z are independent. We claim that this space will do if we define

$$X' = X\,I\{X = Y\} + \frac{\max(X,\,Y)}{2}\,I\{X \neq Y\} \quad \text{and}$$

$$Y' = XZI\{X = Y\} + \min(X,\,Y)\,I\{X \neq Y\}.$$

Indeed, checking the algebra first, if $X = Y$, then

$$2X' + Y'I\{Y' \leqslant X'\} = 2X + XZI\{XZ \leqslant X\} = 2X = X + Y,$$

while if $X \neq Y$, then

$$2X' + Y'I\{Y' \leqslant X'\} = 2\,\frac{\max(X,\,Y)}{2} + \min(X,\,Y)\,I\left\{\min(X,\,Y) \leqslant \frac{\max(X,\,Y)}{2}\right\}$$

$$= \max(X,\,Y) + \min(X,\,Y) = X + Y,$$

again, since $\max(X,\,Y)$ is at least twice as large as $\min(X,\,Y)$ if $X \neq Y$.

To check that X' and Y' have the desired joint distribution, note that for all $j,\,k \in \mathbb{N}$,

$$\begin{aligned}
\boldsymbol{P}\{X' = 2^j,\, Y' = 2^k\} &= \boldsymbol{P}\{X' = 2^j,\, Y' = 2^k,\, X = Y\} + \boldsymbol{P}\{X' = 2^j,\, Y' = 2^k,\, X > Y\} \\
&\quad + \boldsymbol{P}\{X' = 2^j,\, Y' = 2^k,\, X < Y\} \\
&= \boldsymbol{P}\{X = 2^j,\, XZ = 2^k,\, X = Y\} + \boldsymbol{P}\{X = 2^{j+1},\, Y = 2^k,\, X > Y\} \\
&\quad + \boldsymbol{P}\{Y = 2^{j+1},\, X = 2^k,\, X < Y\} \\
&= \boldsymbol{P}\{X = 2^j,\, Y = 2^j,\, Z = 2^{k-j}\} + 2\boldsymbol{P}\{X = 2^{j+1},\, Y = 2^k,\, X > Y\} \\
&= \begin{cases} \boldsymbol{P}\{X = 2^j,\, Y = 2^j,\, Z = 2^{k-j}\}, & \text{if } k \geqslant j+1, \\ 2\boldsymbol{P}\{X = 2^{j+1},\, Y = 2^k,\, X > Y\}, & \text{if } k < j+1, \end{cases} \\
&= \begin{cases} \boldsymbol{P}\{X = 2^j,\, Y = 2^j,\, Z = 2^{k-j}\}, & \text{if } k \geqslant j+1, \\ 2\boldsymbol{P}\{X = 2^{j+1},\, Y = 2^k\}, & \text{if } k < j+1, \end{cases}
\end{aligned}$$

so that in either case,

$$\boldsymbol{P}\{X' = 2^j,\, Y' = 2^k\} = \frac{1}{2^j}\frac{1}{2^k} = \boldsymbol{P}\{X = 2^j,\, Y = 2^k\},$$

completely proving the lemma. □

Lemma 6. *The number of the smallest strictly positive components of an admissible strategy $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$, different from individualistic strategies, is even.*

Proof. If not, with the smallest non-zero component of $\boldsymbol{p}_n$ denoted as $1/2^k$ for some $k \in \mathbb{N}$, the integer $2^k = \sum_{j=1}^n 2^k p_{j,n}$ would have to be odd, since the sum would contain an odd number of ones and, possibly, additional even integers. □

Proof of Theorem 1. In the proof of the first two statements we shall restrict our attention to the improper-Riemann-integral interpretation of (15). As already mentioned, it will follow from the third assertion of the theorem, the proof of which is independent of the proof of the first two statements, that whenever (15) is defined as an improper Riemann integral (if and only if $\boldsymbol{p}_n$ is admissible, once the first statement is proved), the integrand in (15) is non-negative, so that (15) is also defined in the more demanding Lebesgue sense, necessarily with the same value. For a given strategy $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$, the integral $A(\boldsymbol{p}_n)$ in (15) in turn is defined in the improper Riemann sense if and only if $A(\boldsymbol{p}_n, b) \to A(\boldsymbol{p}_n)$ as $b \to \infty$, where

$$A(\boldsymbol{p}_n, b) = \int_0^b [\boldsymbol{P}\{p_{1,n}X_1 + \ldots + p_{n,n}X_n > x\} - \boldsymbol{P}\{X_1 > x\}]\,\mathrm{d}x. \tag{38}$$

Surprisingly, it is possible to show that

$$A(\boldsymbol{p}_n, b) = H(\boldsymbol{p}_n) + R(\boldsymbol{p}_n, b) + o(1), \qquad \text{as } b \to \infty, \tag{39}$$

where $H(\boldsymbol{p}_n)$ is the entropy function defined in (16), and

$$R(\boldsymbol{p}_n,\, b) = \delta(b) - [p_{1,n}\delta(b/p_{1,n}) + \ldots + p_{n,n}\delta(b/p_{n,n})], \tag{40}$$

where $\delta(\cdot)$ refers to the function defined in (14) and where $p\delta(b/p)$ is interpreted as zero whenever $p = 0$; it will be clear that the latter convention is in accord with the similar convention for the terms of the entropy function just before the statement of Theorem 1.

The 'if part' in the first statement of the theorem is already obvious from (39) and (40): for whenever a component $p_{j,n}$ is an integer power of 2, $\delta(b/p_{j,n}) = \delta(b)$ for every value of $b > 0$, due to the periodicity property discussed following the definition of the function $\delta(\cdot)$ in (14). Consequently, $R(\boldsymbol{p}_n,\, b) = \delta(b) - \delta(b)\sum_{j=1}^{n} p_{j,n} = 0$ for every $b > 0$, and then $A(\boldsymbol{p}_n,\, b) = H(\boldsymbol{p}_n) + o(1) \to H(\boldsymbol{p}_n)$ as $b \to \infty$. So $A(\boldsymbol{p}_n) = H(\boldsymbol{p}_n)$ for every admissible strategy $\boldsymbol{p}_n = (p_{1,n},\, \ldots,\, p_{n,n})$, as claimed in the second statement.

Conversely, still taking (39) for granted, suppose the improper Riemann integral $A(\boldsymbol{p}_n)$ in (15) is defined, so that $A(\boldsymbol{p}_n,\, b) \to A(\boldsymbol{p}_n)$ as $b \to \infty$. Then it follows from (39) that

$$R(\boldsymbol{p}_n,\, b) = A(\boldsymbol{p}_n) - H(\boldsymbol{p}_n) + o(1), \qquad \text{as } b \to \infty. \tag{41}$$

Now, in view of the previously noted periodicity property of $\delta(s)$ in the variable $u = \operatorname{Log} s$, it follows from the definition in (40) that

$$\begin{aligned} R(\boldsymbol{p}_n,\, 2^k b) &= \delta(2^k b) - [\, p_{1,n}\delta(2^k b/p_{1,n}) + \ldots + p_{n,n}\delta(2^k b/p_{n,n})] \\ &= \delta(b) - [\, p_{1,n}\delta(b/p_{1,n}) + \ldots + p_{n,n}\delta(b/p_{n,n})] = R(\boldsymbol{p}_n,\, b) \end{aligned}$$

for every $k \in \mathbb{N}$. Fixing $b > 0$ and letting $k \to \infty$, so that $2^k b \to \infty$, we see that the $o(1)$ term appearing in (41) must be identically zero. Whence,

$$R(\boldsymbol{p}_n,\, b) = A(\boldsymbol{p}_n) - H(\boldsymbol{p}_n), \qquad b > 0, \tag{42}$$

a constant function in the variable b.

At this point, we must examine more closely the function $\delta(\cdot)$ defined in (14) and graphed in Figure 1. It is apparent that $\delta(s)$ is continuous in s, and even differentiable in s except at integer powers of 2, where $\delta(s) = 0$ and where there are unequal left and right derivatives. Let D_+ denote the right-hand-side differential operator and D_- denote the left-hand-side differential operator, and consider the linear operator $D = D_+ - D_-$ that applies to a subset of real functions on $\mathbb{R}$. As applied to $\delta(\cdot)$, it can be easily checked that

$$D_+\delta(s)|_{s=2^k} = \frac{1}{2^k}\left(\frac{1}{\log 2} - 1\right) > 0 \quad \text{and} \quad D_-\delta(s)|_{s=2^k} = \frac{1}{2^k}\left(\frac{1}{\log 2} - 2\right) < 0$$

for any $k \in \mathbb{Z}$, the multiplying factor of $1/2^k$ in the first of these being approximately 0.4426 while that in the second -0.5574, in accord with the first graph in Figure 1. So,

$$D\delta(s) = \begin{cases} \dfrac{1}{2^k}, & \text{for } s = 2^k \text{ when } k \in \mathbb{Z}, \\ 0, & \text{for all other } s > 0, \end{cases}$$

from which, when $p_{j,n} > 0$, we find that

$$D\, p_{j,n}\delta(b/p_{j,n}) = \begin{cases} \dfrac{1}{2^k}, & \text{for } b = 2^k\, p_{j,n} \text{ when } k \in \mathbb{Z}, \\ 0, & \text{for all other } b > 0, \end{cases}$$

$j = 1, \ldots, n$. Viewing $R(\boldsymbol{p}_n, b)$ as a function of $b > 0$, using (40) and the last two formulae for $s = 1 = b$, we obtain

$$DR(\boldsymbol{p}_n, b)|_{b=1} = 1 - \sum_{j \in A} p_{j,n},$$

where A is the set of indices $j \in \{1, \ldots, n\}$ for which $p_{j,n}$ is an integer power of 2. But $DR(\boldsymbol{p}_n, b) = 0$ by (42) for every $b > 0$, and so $\sum_{j \in A} p_{j,n} = 1$. So every non-zero $p_{j,n}$ is an integer power of 2. This completes the 'only if' part for the first statement.

It remains to validate (39). To this end, set $U_j = p_{j,n} X_j$ in Lemma 2 for $j = 1, \ldots, n$. By Lemma 3 the condition $\boldsymbol{E}(\min(U_j, U_k)) < \infty$ for $1 \leqslant j, k \leqslant n, j \neq k$, is met, so that

$$\int_0^\infty \left[\boldsymbol{P}\left\{ \sum_{j=1}^n p_{j,n} X_j > x \right\} - \sum_{j=1}^n \boldsymbol{P}\{p_{j,n} X_j > x\} \right] \mathrm{d}x = 0.$$

Consequently,

$$\int_0^b \boldsymbol{P}\left\{ \sum_{j=1}^n p_{j,n} X_j > x \right\} \mathrm{d}x = \sum_{j=1}^n \int_0^b \boldsymbol{P}\{p_{j,n} X_j > x\}\, \mathrm{d}x + o(1)$$

as $b \to \infty$, so that, from (38),

$$\begin{aligned} A(\boldsymbol{p}_n, b) &= \sum_{j=1}^n \int_0^b \boldsymbol{P}\{p_{j,n} X_j > x\}\, \mathrm{d}x - \int_0^b \boldsymbol{P}\{X_1 > x\}\, \mathrm{d}x + o(1) \\ &= \sum_{\{1 \leqslant j \leqslant n:\ p_{j,n} > 0\}} p_{j,n} \int_0^{b/p_{j,n}} \boldsymbol{P}\{X_1 > y\}\, \mathrm{d}y - \int_0^b \boldsymbol{P}\{X_1 > x\}\, \mathrm{d}x + o(1) \end{aligned}$$

as $b \to \infty$. All of the latter integrals can be evaluated for sufficiently large values of b by applying Lemma 4. With these applications, and with the definitions of $H(\boldsymbol{p}_n)$ and $R(\boldsymbol{p}_n, b)$ in (16) and (40), we obtain

$$\begin{aligned} A(\boldsymbol{p}_n, b) &= \sum_{\{1 \leqslant j \leqslant n:\ p_{j,n} > 0\}} p_{j,n} \left[1 + \operatorname{Log} \frac{b}{p_{j,n}} - \delta\left(\frac{b}{p_{j,n}} \right) \right] - [1 + \operatorname{Log} b - \delta(b)] + o(1) \\ &= \sum_{\{1 \leqslant j \leqslant n:\ p_{j,n} > 0\}} p_{j,n} \operatorname{Log} \frac{1}{p_{j,n}} + \left\{ \delta(b) - \sum_{\{1 \leqslant j \leqslant n:\ p_{j,n} > 0\}} p_{j,n} \delta\left(\frac{b}{p_{j,n}} \right) \right\} + o(1) \\ &= H(\boldsymbol{p}_n) + R(\boldsymbol{p}_n, b) + o(1) \qquad \text{as } b \to \infty. \end{aligned}$$

This completes the validation of (39), and hence the proof of the first two statements.

We turn to the proof of the third assertion. Given an arbitrary admissible strategy

$\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$, let $m = m_n \geqslant 1$ denote the number of its non-zero components. Starting with this strategy $\boldsymbol{p}_n^{(1)} = \boldsymbol{p}_n$ and the independent St Petersburg random variables $X_1^{(1)} = X_1, \ldots, X_n^{(1)} = X_n$, we shall define recursively a sequence of admissible strategies $\boldsymbol{p}_n^{(i)} = (p_{1,n}^{(i)}, \ldots, p_{n,n}^{(i)})$, $i = 2, \ldots, m$, terminating with $\boldsymbol{p}_n^{(m)} = (1, 0, \ldots, 0)$, and corresponding sequences of independent St Petersburg random variables $X_1^{(i)}, \ldots, X_n^{(i)}$, $i = 2, \ldots, m$, which yield a monotone non-increasing sequence of pooled winnings:

$$\begin{aligned} p_{1,n}X_1 + \ldots + p_{n,n}X_n &= p_{1,n}^{(1)}X_1^{(1)} + \ldots + p_{n,n}^{(1)}X_n^{(1)} \\ &\geqslant p_{1,n}^{(2)}X_1^{(2)} + \ldots + p_{n,n}^{(2)}X_n^{(2)} \geqslant \ldots \\ &\geqslant p_{1,n}^{(m)}X_1^{(m)} + \ldots + p_{n,n}^{(m)}X_n^{(m)} \\ &= X_1^{(m)} \end{aligned}$$

almost surely. To complete the proof, following this construction, one only needs to set $X_{\boldsymbol{p}_n} = X_1^{(m)}$ and $Y_{\boldsymbol{p}_n} = p_{1,n}X_1 + \ldots + p_{n,n}X_n - X_{\boldsymbol{p}_n}$.

Now to the details of the construction: beginning with $i = 2$, but otherwise proceeding from the already constructed $(i-1)$th level in exactly the same way at each step, we seek out within the strategy $\boldsymbol{p}_n^{(i-1)} = (p_{1,n}^{(i-1)}, \ldots, p_{n,n}^{(i-1)})$ a pair of the smallest non-zero components, $p_{j,n}^{(i-1)}$ and $p_{k,n}^{(i-1)}$, say, with $1 \leqslant j < k \leqslant n$. Since all of our constructed strategies are admissible, beginning with the given admissible strategy $\boldsymbol{p}_n^{(1)} = \boldsymbol{p}_n$, it follows by Lemma 6 that $p_{j,n}^{(i-1)}$ and $p_{k,n}^{(i-1)}$ are equal. Call the common value $c = c_{j,k;n}^{(i-1)}$. Then let

$$p_{l,n}^{(i)} = \begin{cases} 2c, & \text{for } l = j, \\ 0, & \text{for } l = k, \\ p_{l,n}^{(i-1)}, & \text{for } l = 1, \ldots, n, \ l \neq j \text{ and } l \neq k. \end{cases}$$

Finally, identifying the St Petersburg random variables X and Y appearing in the statement of Lemma 5 with the variables $X_j^{(i-1)}$ and $X_k^{(i-1)}$ appearing in the $(i-1)$th stage of the construction, corresponding to the two smallest components of the strategy singled out, let $X'_{j,k;i}$ and $Y'_{j,k;i}$ be the new independent St Petersburg random variables, also independent of $X_l^{(i-1)}$, $l = 1, \ldots, n$, $l \neq j$, $l \neq k$, provided by Lemma 5, and set

$$X_l^{(i)} = \begin{cases} X'_{j,k;i}, & \text{for } l = j, \\ Y'_{j,k;i}, & \text{for } l = k, \\ X_l^{(i-1)}, & \text{for } l = 1, \ldots, n, \ l \neq j \text{ and } l \neq k. \end{cases}$$

Clearly, the new strategy $\boldsymbol{p}_n^{(i)} = (p_{1,n}^{(i)}, \ldots, p_{n,n}^{(i)})$ is admissible and

$$
\begin{aligned}
p_{1,n}^{(i)}X_1^{(i)} + \ldots + p_{n,n}^{(i)}X_n^{(i)} &= p_{1,n}^{(i-1)}X_1^{(i-1)} + \ldots + p_{n,n}^{(i-1)}X_n^{(i-1)} \\
&\quad - c[X_j^{(i-1)} + X_k^{(i-1)}] + 2c\,X'_{j,k;i} + 0 \cdot Y'_{j,k;i} \\
&= p_{1,n}^{(i-1)}X_1^{(i-1)} + \ldots + p_{n,n}^{(i-1)}X_n^{(i-1)} \\
&\quad - c[2X'_{j,k;i} + Y'_{j,k;i}I\{Y'_{j,k;i} \leqslant X'_{j,k;i}\}] + 2cX'_{j,k;i} \\
&= p_{1,n}^{(i-1)}X_1^{(i-1)} + \ldots + p_{n,n}^{(i-1)}X_n^{(i-1)} - cY'_{j,k;i}\,I\{Y'_{j,k;i} \leqslant X'_{j,k;i}\} \\
&\geqslant p_{1,n}^{(i-1)}X_1^{(i-1)} + \ldots + p_{n,n}^{(i-1)}X_n^{(i-1)}
\end{aligned}
$$

almost surely by Lemma 5. This completes the recursive construction for indices $i = 2, \ldots, m$. It should be noted that the number of non-zero components in $\boldsymbol{p}_n^{(i)}$ is one fewer than in $\boldsymbol{p}_n^{(i-1)}$. By the time the index i reaches m, the strategy $\boldsymbol{p}_n^{(m)}$ has only one non-zero component, and it takes the stated form $(1, 0, \ldots, 0)$. The proof of the third statement of the theorem is now complete, and so the whole theorem is proved. □

In line with what it should be, one can show – after carefully tracing through the construction above in the proof of the third assertion of Theorem 1, and using (8) – that $\boldsymbol{E}(Y_{\boldsymbol{p}_n}) = H(\boldsymbol{p}_n)$. Thus the third statement of Theorem 1 provides a stochastic route for evaluating the added value of an admissible strategy $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})$.

Proof of Theorem 2. Simplifying notation, we shall let Π_n denote the collection of all n-dimensional admissible strategies $\boldsymbol{p} = (p_1, \ldots, p_n) = (p_{1,n}, \ldots, p_{n,n}) = \boldsymbol{p}_n$ with non-increasing components $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_n$. Since the value of $H(\boldsymbol{p}) = H(\boldsymbol{p}_n)$ is independent of the order of the components of $\boldsymbol{p}$, the ordering of the component sizes within Π_n is imposed as a matter of convenience. By Lemma 6, for $\boldsymbol{p} \in \Pi_n$, the last two terms, $u_{\boldsymbol{p}} = p_{n-1}$ and $v_{\boldsymbol{p}} = p_n$, can assume one of the three patterns: $(u_{\boldsymbol{p}}, v_{\boldsymbol{p}}) = (0, 0)$, $(1/2^k, 0)$ or $(1/2^k, 1/2^k)$ for some $k \in \mathbb{N}$. In all three cases, the sum $u_{\boldsymbol{p}} + v_{\boldsymbol{p}}$ is either zero or an integer power of 2. Thus, when $n \geqslant 3$, we can form a new admissible strategy $\hat{\boldsymbol{p}} = (\hat{p}_1, \ldots, \hat{p}_{n-1}) \in \Pi_{n-1}$ consisting of the $n-1$ components $p_1, \ldots, p_{n-2}, u_{\boldsymbol{p}} + v_{\boldsymbol{p}}$, ordered from largest to smallest. Further, we observe (with $0\,\mathrm{Log}\,0 = 0$) that

$$
\begin{aligned}
H(\boldsymbol{p}) - H(\hat{\boldsymbol{p}}) &= -\sum_{j=1}^{n} p_j \,\mathrm{Log}\, p_j + \sum_{j=1}^{n-1} \hat{p}_j \,\mathrm{Log}\, \hat{p}_j \\
&= (u_{\boldsymbol{p}} + v_{\boldsymbol{p}})\mathrm{Log}\,(u_{\boldsymbol{p}} + v_{\boldsymbol{p}}) - u_{\boldsymbol{p}}\,\mathrm{Log}\,u_{\boldsymbol{p}} - v_{\boldsymbol{p}}\,\mathrm{Log}\,v_{\boldsymbol{p}} \\
&= \begin{cases} 0, & \text{when } v_{\boldsymbol{p}} = 0, \\ 2u_{\boldsymbol{p}}\,\mathrm{Log}\,2u_{\boldsymbol{p}} - 2u_{\boldsymbol{p}}\,\mathrm{Log}\,u_{\boldsymbol{p}} = 2u_{\boldsymbol{p}}, & \text{when } v_{\boldsymbol{p}} > 0. \end{cases}
\end{aligned}
$$

Thus

$$H(\boldsymbol{p}) = \begin{cases} H(\hat{\boldsymbol{p}}), & \text{when } v_{\boldsymbol{p}} = 0, \\ H(\hat{\boldsymbol{p}}) + 2u_{\boldsymbol{p}}, & \text{when } v_{\boldsymbol{p}} > 0. \end{cases} \tag{43}$$

Equation (43) will play an essential role in an induction argument below.

After these preliminaries, we also note that the second equality in (18) follows directly from the definition of the function $\delta(\cdot)$ in (14). Moreover, writing any $n = 2, 3, \ldots$ as $n = 2^{\lfloor \mathrm{Log}\, n \rfloor} + r_n$ with the remainder term $r_n \in \{0, 1, \ldots, 2^{\lfloor \mathrm{Log}\, n \rfloor} - 1\}$, we see that

$$H_n = \lfloor \mathrm{Log}\, n \rfloor + 2^{\langle \mathrm{Log}\, n \rangle} - 1 = \lfloor \mathrm{Log}\, n \rfloor + \frac{n}{2^{\lfloor \mathrm{Log}\, n \rfloor}} - 1 = \lfloor \mathrm{Log}\, n \rfloor + \frac{2^{\lfloor \mathrm{Log}\, n \rfloor} + r_n}{2^{\lfloor \mathrm{Log}\, n \rfloor}} - 1$$
$$= \lfloor \mathrm{Log}\, n \rfloor + \frac{r_n}{2^{\lfloor \mathrm{Log}\, n \rfloor}},$$

so what we have to prove is that

$$H(\boldsymbol{p}) \leqslant k + \frac{r}{2^k}, \qquad \boldsymbol{p} \in \Pi_{2^k+r},\ r = 0, 1, \ldots, 2^k - 1;\ k \in \mathbb{N}, \tag{44}$$

and that equality is attained in (44) for a given k and r, uniquely within Π_{2^k+r}, when the first $2^k - r$ components of $\boldsymbol{p} = \boldsymbol{p}^* = \boldsymbol{p}_n^* = \boldsymbol{p}_{2^k+r}^* = (p_1^*, \ldots, p_{2^k+r}^*)$ are equal to $1/2^k$ and the remaining $2r$ components are equal to $1/2^{k+1}$. To see that this is consistent with the description of $\boldsymbol{p}_n^*$ in (19) and (20), we note that when $n = 2^k + r$ with $r > 0$, then $\lceil \mathrm{Log}\, n \rceil = \lceil \mathrm{Log}\,(2^k + r) \rceil = k + 1$, so that, appropriately, $m_2(n) = 2^{k+1} - (2^k + r) = 2^k - r$ and $2p_n^* = 2/2^{k+1} = 1/2^k$. Likewise, appropriately, $m_1(n) = 2(2^k + r) - 2^{k+1} = 2r$ and $p_n^* = 1/2^{k+1}$. Also, when $n = 2^k$, then $\lceil \mathrm{Log}\, n \rceil = \lceil \mathrm{Log}\, 2^k \rceil = k$, so that $m_2(n) = 2^k - (2^k + 0) = 0$, $m_1(n) = 2(2^k + 0) - 2^k = 2^k$, and, appropriately, $\boldsymbol{p}_n^* = \boldsymbol{p}_{2^k}^*$ is a vector with all of its components equal to $p_{2^k}^* = 1/2^k$.

Addressing the claim of equality first, for the strategy $\boldsymbol{p}_{2^k+r}^* = (p_1^*, \ldots, p_{2^k+r}^*)$ described above we have

$$\sum_{j=1}^{2^k+r} p_j^* = \frac{2^k - r}{2^k} + \frac{2r}{2^{k+1}} = 1$$

and

$$H(\boldsymbol{p}_{2^k+r}^*) = \sum_{j=1}^{2^k+r} p_j^* \,\mathrm{Log}\, \frac{1}{p_j^*} = (2^k - r)\, \frac{k}{2^k} + 2r\, \frac{k+1}{2^{k+1}} = \frac{k2^k + r}{2^k} = k + \frac{r}{2^k}\,.$$

It remains to prove (44) and the attendant claim that $\boldsymbol{p}_{2^k+r}^*$ is the only strategy in Π_{2^k+r} for which equality obtains in (44). By the unconstrained result mentioned after Theorem 2 in the previous section, this statement is valid for every $k \in \mathbb{N}$ when $r = 0$. Suppose that the statement is true for some $k \in \mathbb{N}$ and $r \in \{0, 1, \ldots, 2^k - 2\}$, and consider $n = 2^k + r + 1$ and any $\boldsymbol{p} = (p_1, \ldots, p_n) \in \Pi_{2^k+r+1}$. Let $u_{\boldsymbol{p}}$ and $v_{\boldsymbol{p}}$ be the penultimate and ultimate components of $\boldsymbol{p}$. Then with $\hat{\boldsymbol{p}} \in \Pi_{2^k+r}$ defined from $\boldsymbol{p}$ as described above, we can conclude from (43) and from part of the induction hypothesis that

$$\max_{\boldsymbol{p}\in\Pi_n} H(\boldsymbol{p}) \leqslant \max_{\hat{\boldsymbol{p}}\in\Pi_{n-1}} H(\hat{\boldsymbol{p}}) + 2\max_{\{\boldsymbol{p}\in\Pi_n: v_{\boldsymbol{p}}>0\}} u_{\boldsymbol{p}} \leqslant k + \frac{r}{2^k} + 2\max_{\{\boldsymbol{p}\in\Pi_n: v_{\boldsymbol{p}}>0\}} u_{\boldsymbol{p}}. \tag{45}$$

Thus, to show the inequality

$$H(\boldsymbol{p}) \leqslant k + \frac{r+1}{2^k}, \qquad \boldsymbol{p} \in \Pi_{2^k+r+1}, \tag{46}$$

the case $v_{\boldsymbol{p}} = 0$ being of no relevance here, it is only necessary to show that $u_{\boldsymbol{p}}$ can never exceed $1/2^{k+1}$ when $u_{\boldsymbol{p}} = v_{\boldsymbol{p}} > 0$. If this were not so, we would need to have a $\boldsymbol{p} = (p_1, \ldots, p_n) \in \Pi_{2^k+r+1}$ for which $u_{\boldsymbol{p}} = v_{\boldsymbol{p}} \geqslant 1/2^k$ and, consequently,

$$p_1 + \ldots + p_n \geqslant \frac{n}{2^k} = \frac{2^k + r + 1}{2^k} > 1.$$

So, necessarily, $u_{\boldsymbol{p}} = v_{\boldsymbol{p}} \leqslant 1/2^{k+1}$, and (46) follows.

The case of equality in (46) arises from two equalities in (45). Necessarily, a $\boldsymbol{p} \in \Pi_{2^k+r+1}$ for which equality obtains in (46) must be such that $\hat{\boldsymbol{p}} = \boldsymbol{p}^*_{n-1} = \boldsymbol{p}^*_{2^k+r}$, which by the other part of the induction hypothesis is the unique maximizing member of $\Pi_{n-1} = \Pi_{2^k+r}$, and $u_{\boldsymbol{p}} = v_{\boldsymbol{p}} = 1/2^{k+1}$. The first condition says that

$$\hat{\boldsymbol{p}} = \Bigg(\underbrace{\frac{1}{2^k}, \ldots, \frac{1}{2^k}}_{2^k - r \text{ times}}, \underbrace{\frac{1}{2^{k+1}}, \ldots, \frac{1}{2^{k+1}}}_{2r \text{ times}}\Bigg).$$

Quite obviously, this and the second condition can hold together if and only if

$$\boldsymbol{p} = \Bigg(\underbrace{\frac{1}{2^k}, \ldots, \frac{1}{2^k}}_{2^k - (r+1) \text{ times}}, \underbrace{\frac{1}{2^{k+1}}, \ldots, \frac{1}{2^{k+1}}}_{2(r+1) \text{ times}}\Bigg),$$

which is $\boldsymbol{p}^*_n = \boldsymbol{p}^*_{2^k+r+1}$, completing the proof of the theorem. □

The next lemma is for the proof of (24) in Theorem 3. As discussed in Csörgő and Simons (2002), this lemma is not true in general: it is valid here because of the independence assumptions.

Lemma 7. *If $U_1, \ldots, U_n$ are independent random variables and $V_1, \ldots, V_n$ are also independent random variables, $n \geqslant 2$, such that $\boldsymbol{E}[U_1, V_1], \ldots, \boldsymbol{E}[U_n, V_n]$ are all finite either in the Lebesgue or in the improper Riemann sense, then*

$$\boldsymbol{E}_n := \boldsymbol{E}[U_1 + \ldots + U_n, V_1 + \ldots + V_n] = \boldsymbol{E}[U_1, V_1] + \ldots + \boldsymbol{E}[U_n, V_n]$$

in the corresponding sense.

Proof. It suffices to prove the statement for $n = 2$, which we formally do in the Lebesgue case; the other case is analogous. Using first the law of total probability twice, combined with independence, adding and subtracting a term, integrating by parts and changing variables, by Fubini's theorem we obtain

$$
\begin{aligned}
\boldsymbol{E}_2 &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} \boldsymbol{P}\{U_1 + v > x\}\,\mathrm{d}\boldsymbol{P}\{U_2 \leqslant v\} - \int_{-\infty}^{\infty} \boldsymbol{P}\{V_1 + v > x\}\,\mathrm{d}\boldsymbol{P}\{V_2 \leqslant v\}\right]\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} \boldsymbol{P}\{U_1 + v > x\}\,\mathrm{d}_v(\boldsymbol{P}\{U_2 \leqslant v\} - \boldsymbol{P}\{V_2 \leqslant v\})\right]\mathrm{d}x \\
&\quad + \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} (\boldsymbol{P}\{U_1 + v > x\} - \boldsymbol{P}\{V_1 + v > x\})\,\mathrm{d}_v\boldsymbol{P}\{V_2 \leqslant v\}\right]\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} (\boldsymbol{P}\{V_2 + y > x\} - \boldsymbol{P}\{U_2 + y > x\})\,\mathrm{d}_y\boldsymbol{P}\{U_1 > y\}\right]\mathrm{d}x \\
&\quad + \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} (\boldsymbol{P}\{U_1 + v > x\} - \boldsymbol{P}\{V_1 + v > x\})\,\mathrm{d}_v\boldsymbol{P}\{V_2 \leqslant v\}\right]\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \boldsymbol{E}[V_2,\, U_2]\,\mathrm{d}_v\boldsymbol{P}\{U_1 > y\} + \int_{-\infty}^{\infty} \boldsymbol{E}[U_1,\, V_1]\,\mathrm{d}_v\boldsymbol{P}\{V_2 \leqslant v\} \\
&= \boldsymbol{E}[U_1,\, V_1] - \boldsymbol{E}[V_2,\, U_2] = \boldsymbol{E}[U_1,\, V_1] + \boldsymbol{E}[U_2,\, V_2],
\end{aligned}
$$

where, in the fourth and sixth equations, we also used properties (10c) and (10a) of Theorem 2.2 in Csörgő and Simons (2002). □

The last two lemmas were roughly described following Theorem 3. We use the notation $S_{\boldsymbol{p}_n}$ from (21) with the emphasized convention that such centred sums are always built on n independent St Petersburg random variables for any strategy $\boldsymbol{p}_n$. For any given $n = 2, 3, \ldots$ and strategy $\boldsymbol{p}_n = (p_{1,n},\, p_{2,n},\, \ldots,\, p_{n,n})$ and every $k \in \{0, 1, 2, \ldots\}$ we consider the associated strategy

$$
\boldsymbol{p}_n^k = (p_{1,n2^k},\, p_{2,n2^k},\, \ldots,\, p_{n2^k,n2^k}) = \left(\underbrace{\frac{p_{1,n}}{2^k}, \ldots, \frac{p_{1,n}}{2^k}}_{2^k \text{ times}}, \underbrace{\frac{p_{2,n}}{2^k}, \ldots, \frac{p_{2,n}}{2^k}}_{2^k \text{ times}}, \ldots, \underbrace{\frac{p_{n,n}}{2^k}, \ldots, \frac{p_{n,n}}{2^k}}_{2^k \text{ times}}\right)
$$

of $n2^k$ components, so that $\boldsymbol{p}_n^0 = \boldsymbol{p}_n$. Thus, according to our convention, $S_{\boldsymbol{p}_n^k}$ is based on $n2^k$ independent St Petersburg random variables. Whenever $S_{\boldsymbol{p}_n^j}$ is defined on the probability space of Lemma 8 below, for some $j \in \{0, 1, 2, \ldots\}$, based on the independent St Petersburg random variables $X_{1,n2^j}, \ldots, X_{n2^j,n2^j}$, the space will be rich enough to carry an independent sequence $X^*_{1,n2^j}, \ldots, X^*_{n2^j,n2^j}$ of independent St Petersburg random variables, from which we form the independent random variables

$$
Z_{1,n2^j} = \frac{X^*_{1,n2^j} I\{X^*_{1,n2^j} \leqslant X_{1,n2^j}\}}{2} - 1,\ \ldots,\ Z_{n2^j,n2^j} = \frac{X^*_{n2^j,n2^j} I\{X^*_{n2^j,n2^j} \leqslant X_{n2^j,n2^j}\}}{2} - 1
$$

and define

$$T_{\boldsymbol{p}_n^j} = \sum_{m=1}^{n2^j} p_{m,n2^j}\, Z_{m,n2^j} = \frac{p_{1,n}}{2^j}\sum_{m=1}^{2^j} Z_{m,n2^j} + \ldots + \frac{p_{n,n}}{2^j}\sum_{m=1}^{2^j} Z_{(n-1)2^j+m,n2^j}\,.$$

We note right away that $Z_{m,n2^j}$ has mean zero by (8) for all $m = 1, \ldots, n2^j$, and hence $\boldsymbol{E}(T_{\boldsymbol{p}_n^j}) = 0$ whenever $T_{\boldsymbol{p}_n^j}$ is defined.

Lemma 8. *For any fixed* $n = 2, 3, \ldots$ *and strategy* $\boldsymbol{p}_n = (p_{1,n}, p_{2,n}, \ldots, p_{n,n})$ *the independent St Petersburg random variables* $X_1, X_2, \ldots, X_n$*, on which* $S_{\boldsymbol{p}_n}$ *is based, can be given on a rich enough probability space* $(\Omega, \mathcal{A}, \boldsymbol{P})$ *such that*

$$S_{\boldsymbol{p}_n^k} = S_{\boldsymbol{p}_n} + \sum_{j=0}^{k-1} T_{\boldsymbol{p}_n^j} \qquad \textit{for every } k \in \mathbb{N}, \tag{47}$$

almost surely.

Proof. Let the probability space $(\Omega, \mathcal{A}, \boldsymbol{P})$ be such that, besides the given initial sequence $\{X_{1,n2^0}, X_{2,n2^0}, \ldots, X_{n,n2^0}\} = \{X_1, X_2, \ldots, X_n\}$, it carries all the independent sequences

$$\{X^*_{1,n2^j}, \ldots, X^*_{n2^j,n2^j}\} \quad \text{and} \quad \{B^*_{1,n2^j}, \ldots, B^*_{n2^j,n2^j}\}, \qquad j = 0, 1, 2, \ldots,$$

that are independent of the initial sequence $\{X_1, X_2, \ldots, X_n\}$ such that for each $j \in \{0, 1, 2, \ldots\}$ the sequence $\{X^*_{1,n2^j}, \ldots, X^*_{n2^j,n2^j}\}$ consists of independent St Petersburg random variables, while the sequence $\{B^*_{1,n2^j}, \ldots, B^*_{n2^j,n2^j}\}$ consists of independent mean-one-half Bernoulli random variables. We claim that this space does the job.

Starting from the initial sequence $\{X_{1,n2^0}, X_{2,n2^0}, \ldots, X_{n,n2^0}\} = \{X_1, X_2, \ldots, X_n\}$, we inductively construct the desired sequences for (47). Consider a sequence $\boldsymbol{X}_{n2^k} = \{X_{1,n2^k}, \ldots, X_{n2^k,n2^k}\}$ of independent St Petersburg random variables for some $k \in \{0, 1, 2, \ldots\}$. Match this with $\boldsymbol{X}^*_{n2^k} = \{X^*_{1,n2^k}, \ldots, X^*_{n2^k,n2^k}\}$, and construct $Z_{1,n2^k}, \ldots, Z_{n2^k,n2^k}$ and $T_{\boldsymbol{p}_n^k}$ from the two sequences as described above. Finally, define the independent random variables $\{Y_{1,n2^{k+1}}, \ldots, Y_{n2^{k+1},n2^{k+1}}\} =: \boldsymbol{Y}_{n2^{k+1}}$ in the following way:

$$\begin{aligned} Y_{2i-1,n2^{k+1}} = {} & X_{i,n2^k} I\{X_{i,n2^k} < X^*_{i,n2^k}\} + 2X_{i,n2^k} I\{X^*_{i,n2^k} \leqslant X_{i,n2^k}, B_{i,n2^k} = 0\} \\ & + X^*_{i,n2^k} I\{X^*_{i,n2^k} \leqslant X_{i,n2^k}, B_{i,n2^k} = 1\} \end{aligned}$$

and

$$\begin{aligned} Y_{2i,n2^{k+1}} = {} & X_{i,n2^k} I\{X_{i,n2^k} < X^*_{i,n2^k}\} + 2X_{i,n2^k} I\{X^*_{i,n2^k} \leqslant X_{i,n2^k}, B_{i,n2^k} = 1\} \\ & + X^*_{i,n2^k} I\{X^*_{i,n2^k} \leqslant X_{i,n2^k}, B_{i,n2^k} = 0\} \end{aligned}$$

for $i = 1, \ldots, n2^k$. The proof of Theorem 5.1 in Csörgő and Simons (2002) establishes that all the components in $\boldsymbol{Y}_{n2^{k+1}}$ are St Petersburg random variables and

$$p_{i,n2^k}\left[\frac{Y_{2i-1,n2^{k+1}} + Y_{2i,n2^{k+1}}}{2} - 1\right] = p_{i,n2^k}[X_{i,n2^k} + Z_{i,n2^k}], \qquad i = 1, \ldots, n2^k,$$

from which, summing over all indices $i = 1, \ldots, n2^k$,

$$\sum_{i=1}^{n2^{k+1}} p_{i,n2^{k+1}} Y_{i,n2^{k+1}} - 1 = \sum_{i=1}^{n2^k} p_{i,n2^k} X_{i,n2^k} + T_{\boldsymbol{p}_n^k}$$

almost surely. Building $S_{\boldsymbol{p}_n^{k+1}}$ on the sequence $\boldsymbol{X}_{n2^{k+1}} = \boldsymbol{Y}_{n2^{k+1}}$, this may be rewritten as

$$S_{\boldsymbol{p}_n^{k+1}} + H(\boldsymbol{p}_n^{k+1}) - 1 = S_{\boldsymbol{p}_n^k} + H(\boldsymbol{p}_n^k) + T_{\boldsymbol{p}_n^k},$$

which, since, clearly,

$$H(\boldsymbol{p}_n^j) = H(\boldsymbol{p}_n) + j \qquad \text{for all} \quad n \in \{2, 3, \ldots\} \text{ and } j \in \{0, 1, 2, \ldots\}, \tag{48}$$

yields $S_{\boldsymbol{p}_n^{k+1}} = S_{\boldsymbol{p}_n^k} + T_{\boldsymbol{p}_n^k}$. On the one hand, this gives $S_{\boldsymbol{p}_n^1} = S_{\boldsymbol{p}_n^0} + T_{\boldsymbol{p}_n^0}$ to initiate the induction and, on the other, using the induction hypothesis that the equality in (47) holds for some $k \in \{0, 1, 2, \ldots\}$, also that $S_{\boldsymbol{p}_n^{k+1}} = S_{\boldsymbol{p}_n} + \sum_{j=0}^{k} T_{\boldsymbol{p}_n^j}$ holds almost surely. □

Lemma 9. *For every* $m \in \{2, 3, \ldots\}$ *and strategy* $\boldsymbol{p}_m = (p_{1,m}, p_{2,m}, \ldots, p_{m,m})$ *with* $\bar{p}_m = \max\{p_{1,m}, \ldots, p_{m,m}\}$, *let* $X_1, \ldots, X_m, X_1^*, \ldots, X_m{}^*$ *be independent St Petersburg random variables and* $Z_j = \frac{1}{2} X_j^* I\{X_j^* \leqslant X_j\} - 1$, $j = 1, \ldots, m$. *Then*

$$\boldsymbol{E}(|T_{\boldsymbol{p}_m}|) = \boldsymbol{E}(|p_{1,m} Z_1 + \ldots + p_{1,m} Z_m|) < \frac{1}{2\sqrt{\bar{\gamma}_m}} \sqrt{\frac{r_m}{2^{r_m}}} + \frac{4}{\sqrt{2^{r_m}}},$$

where $r_m = \lceil \operatorname{Log}(1/\bar{p}_m) \rceil \in \mathbb{N}$ *and* $\bar{\gamma}_m = 2^{-r_m}/\bar{p}_m \in (\frac{1}{2}, 1]$ *for which* $\bar{p}_m = 2^{-r_m}/\bar{\gamma}_m$.

Proof. Adjusting the end of the proof of Lemma 5.1 in Csörgő and Simons (2002), from the average of $Z_1, \ldots, Z_m$ to their present linear combination, for all $k \in \{0, 1, 2, \ldots\}$ we obtain

$$\boldsymbol{E}(|T_{\boldsymbol{p}_m}|) < \sqrt{\sum_{j=1}^{m} p_{j,m}^2} \sqrt{\frac{k}{2} + \frac{2}{2^k}} \leqslant \sqrt{\bar{p}_m} \sqrt{\frac{k}{2} + \frac{2}{2^k}}.$$

Choosing $k = \lfloor r_m/2 \rfloor \geqslant 0$, the lemma follows by elementary manipulation. □

Applying Lemma 9 to any of the $T_{\boldsymbol{p}_n^j}$ in (47), since $r_{n2^j} = r_n + j$ and $\bar{\gamma}_{n2^j} = \bar{\gamma}_n$ for every $n \in \{2, 3, \ldots\}$ and $j \in \{0, 1, 2, \ldots\}$, as is easy to see, we obtain

$$\begin{aligned} \sum_{j=0}^{\infty} \boldsymbol{E}(|T_{\boldsymbol{p}_n^j}|) &< \sum_{j=0}^{\infty} \left[\frac{1}{2\sqrt{\bar{\gamma}_n}} \sqrt{\frac{r_n + j}{2^{r_n + j}}} + \frac{4}{\sqrt{2^{r_n + j}}} \right] \\ &= \frac{1}{\sqrt{\bar{\gamma}_n}} \sqrt{\frac{r_n}{2^{r_n}}} \sum_{j=0}^{\infty} \left[\frac{1}{2} \sqrt{\frac{1 + j/r_n}{2^j}} + \frac{4\sqrt{\bar{\gamma}_n}}{\sqrt{r_n}\sqrt{2^j}} \right] = \frac{C_n}{\sqrt{\bar{\gamma}_n}} \sqrt{\frac{r_n}{2^{r_n}}}, \end{aligned} \tag{49}$$

identifying the constant C_n in (26) of Theorem 3 by elementary calculation. This implies that the right-hand side of (47) converges almost surely on the probability space of Lemma 8, as $k \to \infty$, to the limiting random variable $S_{\boldsymbol{p}_n} + \sum_{j=0}^{\infty} T_{\boldsymbol{p}_n^j}$, which we denote by $W_{\boldsymbol{p}_n}^{[n]}$. On the other hand, using (48), for the left-hand side random variable in (47) we notice that

$$S_{\boldsymbol{p}_n^k} \stackrel{\mathcal{D}}{=} p_{1,n}\left[\frac{\sum_{j=1}^{2^k} X_j}{2^k} - k\right] + \ldots + p_{n,n}\left[\frac{\sum_{j=(n-1)2^k+1}^{n2^k} X_j}{2^k} - k\right] - H(\boldsymbol{p}_n)$$

for a sequence $X_1, \ldots, X_{n2^k}$ of independent St Petersburg random variables, and hence by Martin-Löf's (1985) Theorem 1, or the corresponding special case of a result in Csörgő and Dodunekova (1991), both described in Section 1, $S_{\boldsymbol{p}_n^k}$ converges in distribution as $k \to \infty$ to the random variable $W_{\boldsymbol{p}_n} = p_{1,n} W_1^{(1)} + \ldots + p_{n,n} W_1^{(n)} - H(\boldsymbol{p}_n)$, as defined in (22). In conclusion,

$$W_{\boldsymbol{p}_n} \stackrel{\mathcal{D}}{=} W_{\boldsymbol{p}_n}^{[n]} = S_{\boldsymbol{p}_n} + \sum_{j=0}^{\infty} T_{\boldsymbol{p}_n^j}, \tag{50}$$

where the right-hand side random variable is well defined on the probability space of Lemma 8.

Proof of Theorem 3. Since $\boldsymbol{E}[X_1, W_1] = 0$ by the case $n = 1$ in equation (10), the first statement in (24) follows directly from Lemma 7 by properties (10b) and (10c) of Theorem 2.2 in Csörgő and Simons (2002). Next, we have $\boldsymbol{E}(|W_{\boldsymbol{p}_n}^{[n]} - S_{\boldsymbol{p}_n}|) = \boldsymbol{E}(|\sum_{j=0}^{\infty} T_{\boldsymbol{p}_n^j}|) \leqslant \sum_{j=0}^{\infty} \boldsymbol{E}(|T_{\boldsymbol{p}_n^j}|)$, so the second statement in (25) follows from (49).

Finally, we derive (27) from (25). Using the notation in (21), (22) and (50), and denoting the bound in (25) by $\eta_n = C_n \sqrt{r_n/2^{r_n}}/\sqrt{\overline{\gamma}_n} = C_n \sqrt{\overline{p}_n r_n}$, on the special probability space $(\Omega, \mathcal{A}, \boldsymbol{P})$ of Lemma 8 and (25), we have

$$\begin{aligned}\boldsymbol{P}\{S_{\boldsymbol{p}_n} \leqslant x\} = \boldsymbol{P}\{W_{\boldsymbol{p}_n}^{[n]} \leqslant x + (W_{\boldsymbol{p}_n}^{[n]} - S_{\boldsymbol{p}_n}), -\varepsilon_n \leqslant W_{\boldsymbol{p}_n}^{[n]} - S_{\boldsymbol{p}_n} \leqslant \varepsilon_n\}\\ + \boldsymbol{P}\{W_{\boldsymbol{p}_n}^{[n]} \leqslant x + (W_{\boldsymbol{p}_n}^{[n]} - S_{\boldsymbol{p}_n}), |W_{\boldsymbol{p}_n}^{[n]} - S_{\boldsymbol{p}_n}| > \varepsilon_n\}\end{aligned}$$

for every $x \in \mathbb{R}$, where $\varepsilon_n = \sqrt{2\eta_n}$. Since $\boldsymbol{P}\{A \cap B\} \geqslant \boldsymbol{P}\{A\} - \boldsymbol{P}\{B^c\}$ for any two events A and B and the complement B^c of B, using Lemma 9 and the Markov inequality, this implies

$$G_{\boldsymbol{p}_n}(x - \varepsilon_n) - \frac{\eta_n}{\varepsilon_n} \leqslant \boldsymbol{P}\{S_{\boldsymbol{p}_n} \leqslant x\} \leqslant G_{\boldsymbol{p}_n}(x + \varepsilon_n) + \frac{\eta_n}{\varepsilon_n},$$

on any probability space $(\Omega, \mathcal{A}, \boldsymbol{P})$ where the $X_1, \ldots, X_n$ in $S_{\boldsymbol{p}_n}$ are defined. Since for the density function $G'_{\boldsymbol{p}_n}(x)$ of $G_{\boldsymbol{p}_n}(x) = \boldsymbol{P}\{W_{\boldsymbol{p}_n} \leqslant x\}$ we have $\sup_{x\in\mathbb{R}} G'_{\boldsymbol{p}_n}(x) \leqslant \frac{1}{2}$ by (32), from these inequalities, using the Lagrange mean value theorem, we obtain

$$\sup_{x\in\mathbb{R}} |\boldsymbol{P}\{S_{\boldsymbol{p}_n} \leqslant x\} - G_{\boldsymbol{p}_n}(x)| \leqslant \frac{\varepsilon_n}{2} + \frac{\eta_n}{\varepsilon_n} = \sqrt{2\eta_n} = \frac{\sqrt{2C_n}}{\overline{\gamma}_n^{1/4}}\left(\frac{r_n}{2^{r_n}}\right)^{1/4} = \sqrt{2C_n}(\overline{p}_n r_n)^{1/4},$$

proving (27). (The choice of ε_n was to make the two terms in the last bound equal.) □

Proof of Theorem 4. Since the proof is a direct extension of that of (6) in Csörgő (2002), we only sketch it, using all applicable ingredients from Csörgő (2002), including (a close match to) the notation, and concentrating mainly on the differences.

By (29),

$$\boldsymbol{g}_{\boldsymbol{p}_n}(t) = \exp\{-\mathrm{i}tH(\boldsymbol{p}_n)\}\exp\left\{\sum_{k=1}^{n} y_1(p_{k,n}t)\right\},$$

and for $\boldsymbol{f}_{\boldsymbol{p}_n}(t) = \boldsymbol{E}(\mathrm{e}^{\mathrm{i}tS_{\boldsymbol{p}_n}})$ we obtain

$$\boldsymbol{f}_{\boldsymbol{p}_n}(t) = \exp\{-\mathrm{i}tH(\boldsymbol{p}_n)\}\exp\left\{\sum_{k=1}^{n} \log(1 + y_{k,n}(t))\right\},$$

where

$$y_{k,n}(t) = \boldsymbol{E}(\mathrm{e}^{\mathrm{i}tp_{k,n}X} - 1) = \int_0^1 (\exp\{\mathrm{i}tp_{k,n}2^{\lceil \mathrm{Log}\,(1/s)\rceil}\} - 1)\,\mathrm{d}s, \qquad |t| \leqslant T_n.$$

Here $T_n = 2K_n/\bar{p}_n$ for $K_n = 1/\mathrm{Log}\,(1/\bar{p}_n)$ and we take n large enough to make $\bar{p}_{\mathrm{n}}^{1+\kappa} \leqslant K_{\mathrm{n}} \leqslant 4/\mathrm{e}^2$ and $L_n = 2K_n(2 - \mathrm{Log}\,K_n) < 1$ for an arbitrarily given $\kappa > 0$. Of course, $y_{k,n}(t) = 0$ if $p_{k,n} = 0$. If $p_{k,n} > 0$, then for $x_{k,n}(t) = y_{k,n}(t)/p_{k,n}$ we obtain as extensions of the corresponding special cases of (10) and (11) in Csörgő (2002) that

$$\mathrm{Re}\,x_{k,n}(t) \leqslant -\left(\frac{2}{\pi} - \frac{8p_{k,n}}{\pi^2}\,|t|\right)|t|, \qquad t \in \mathbb{R},$$

$$|x_{k,n}(t)| \leqslant \left(2 + \mathrm{Log}\,\frac{2}{p_{k,n}|t|}\right)|t|, \qquad |t| \leqslant T_n.$$

The latter implies $|y_{k,n}(t)| \leqslant L_n$ and, setting $w_{k,n}(t) = \log(1 + y_{k,n}(t)) - y_{k,n}(t)$, also that $|w_{k,n}(t)| \leqslant M_n p_{k,n}^2 |x_{k,n}(t)|^2$ for $|t| \leqslant T_n$, where

$$M_n = \frac{1}{6} + \frac{1}{3}\,\frac{1}{1 - L_n} \to \frac{1}{2}.$$

So we obtain

$$\begin{aligned}
|\boldsymbol{f}_{\boldsymbol{p}_n}(t) - \boldsymbol{g}_{\boldsymbol{p}_n}(t)| &\leqslant \left|\mathrm{e}^{\sum_{k=1}^n \log(1+y_{k,n}(t))} - \mathrm{e}^{\sum_{k=1}^n y_{k,n}(t)}\right| + \left|\mathrm{e}^{\sum_{k=1}^n y_{k,n}(t)} - \mathrm{e}^{\sum_{k=1}^n y_1(p_{k,n}t)}\right| \\
&\leqslant \mathrm{e}^{\sum_{k=1}^n \mathrm{Re}\,y_{k,n}(t)}\mathrm{e}^{\sum_{k=1}^n |w_{k,n}(t)|} \sum_{k=1}^{n} |w_{k,n}(t)| \\
&\quad + \frac{1}{2}\left\{\mathrm{e}^{\sum_{k=1}^n \mathrm{Re}\,y_{k,n}(t)} + \mathrm{e}^{\sum_{k=1}^n \mathrm{Re}\,y_1(p_{k,n}t)}\right\} \sum_{k=1}^{n} |y_{k,n}(t) - y_1(p_{k,n}t)| \\
&=: \boldsymbol{\delta}_n^{(1)}(t) + \boldsymbol{\delta}_n^{(2)}(t)
\end{aligned}$$

as an analogue of (7) in Csörgő (2002). Using the bounds above, the fact that the function $x \mapsto x[3 - \mathrm{Log}\,(p_{k,n}x)]^2$ is increasing on $[0, T_n]$ if $p_{k,n} > 0$, and the fact that for each $|t| \in (0, T_n]$ the function $x \mapsto x[3 - \mathrm{Log}\,(x|t|)]^2$ is also increasing on $[0, \bar{p}_n]$, for every $t \in [-T_n, T_n]$ we obtain

$$\boldsymbol{\delta}_n^{(1)}(t) \leqslant M_n \exp\{-\overline{C}_n t\} \bar{p}_n \{3 - \operatorname{Log}(\bar{p}_n|t|)\}^2 |t|^2,$$

where

$$\overline{C}_n = \frac{2}{\pi} - \frac{16}{\pi^2} K_n - 2K_n M_n (2 - \operatorname{Log} K_n)^2 \to \frac{2}{\pi}.$$

Since by Lemma 4 in Csörgő (2002) we have

$$\sum_{k=1}^{n} |y_{k,n}(t) - y_1(p_{k,n}t)| = \sum_{k=1}^{n} \left| \sum_{j=0}^{\infty} \left[\mathrm{e}^{\mathrm{i}tp_{k,n}/2^j} - 1 - \frac{\mathrm{i}tp_{k,n}}{2^j} \right] 2^j \right| \leqslant \frac{t^2}{2} \sum_{k=1}^{n} p_{k,n}^2 \sum_{j=0}^{\infty} \frac{2^j}{2^{2j}} \leqslant \bar{p}_n t^2,$$

we see that

$$\boldsymbol{\delta}_n^{(2)}(t) \leqslant \frac{1}{2} \left[\exp\left\{ -\left(\frac{2}{\pi} - \frac{16}{\pi^2} K_n \right) |t| \right\} + \exp\left\{ -\frac{2}{\pi} |t| \right\} \right] \bar{p}_n t^2 \qquad \text{for all } t \in \mathbb{R},$$

where we have again used the bound for $\operatorname{Re} x_{k,n}(\cdot)$ and also that for $\operatorname{Re} y_1(\cdot)$ in (31).

By Esseen's classical smoothing lemma (Lemma 1 in Csörgő 2002), using the bound in (32), for the deviation $\Delta_n = \sup_{x\in\mathbb{R}} |\boldsymbol{P}\{S_{\boldsymbol{p}_n} \leqslant x\} - \boldsymbol{P}\{W_{\boldsymbol{p}_n} \leqslant x\}|$ in (28) we obtain

$$\begin{aligned}
\Delta_n &\leqslant \frac{b}{2\pi} \int_{-T_n}^{T_n} \frac{\boldsymbol{\delta}_n^{(1)}(t)}{|t|}\, \mathrm{d}t + \frac{b}{2\pi} \int_{-T_n}^{T_n} \frac{\boldsymbol{\delta}_n^{(2)}(t)}{|t|}\, \mathrm{d}t + c_b \frac{\sup_{x\in\mathbb{R}} G'_{\boldsymbol{p}_n}(x)}{T_n} \\
&\leqslant \left\{ \frac{bM_n}{\pi} \int_0^{2K_n/\bar{p}_n} \mathrm{e}^{-\overline{C}_n t}\, t \left[\frac{2 + \operatorname{Log}(2/\bar{p}_n t)}{\operatorname{Log}(1/\bar{p}_n)} \right]^2 \mathrm{d}t \right\} \bar{p}_n \operatorname{Log}^2 \frac{1}{\bar{p}_n} \\
&\quad + \frac{b}{2\pi} \left\{ \int_0^{\infty} t \exp\left(-\left(\frac{2}{\pi} - \frac{16}{\pi^2} K_n \right) t \right) \mathrm{d}t + \int_0^{\infty} t \exp\left(-\frac{2}{\pi} t \right) \mathrm{d}t \right\} \bar{p}_n + \frac{c_b}{4}\, \bar{p}_n \operatorname{Log} \frac{1}{\bar{p}_n}
\end{aligned}$$

for all n large enough and all $b > 1$, where the constant c_b depends only on b. Since $K_n \to 0$, the coefficient of $\bar{p}_n$ in the second term has a finite limit. But since $K_n \geqslant \bar{p}_\mathrm{n}^{1+\kappa}$, a version of the argument in Csörgő (2002: 842–843) also shows that the limsup of the coefficient of $\bar{p}_n \operatorname{Log}^2(1/\bar{p}_n)$ is not greater than $b(1+\kappa)^2\pi/8$, which proves (28). □

Acknowledgements

This paper is dedicated to Yi-Ching Yao (Institute of Statistical Science, Academia Sinica, Taipei), who first raised the question of working with multiple Pauls. We also thank Erik and Willem van Zwet (Mathematical Institute, University of Leiden) for their remarks on Lemma 5, Péter Kevei (Bolyai Institute, University of Szeged) for spotting some misprints, two referees for their perceptive remarks and we are grateful to Vilmos Totik (Bolyai Institute, University of Szeged) for his construction after Corollary 2. The first author was supported in part by the National Foundation for Scientific Research, Hungary, Grants T-032025, T-034121 and T-048360.

References

Bernoulli, D. (1738) Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, **5**, 175–192. (Reprinted in *Die Werke von Daniel Bernoulli*, Vol. 2, pp. 223–234, Basel: Birkhäuser, 1982. English translation: Exposition of a new theory of the measurement of risk. *Econometrica*, **22**, 23–36, 1954, reprinted in W.J. Baumol and S.M. Goldfeld (eds.), *Precursors in Mathematical Economics: An Anthology*, pp. 15–26, London: London School of Economics and Political Science, 1968; in G.A. Miller (ed.), *Mathematics and Psychology*, pp. 36–52, New York: Wiley, 1964; and in A.N. Page (ed.), *Utility Theory: A Book of Readings*, pp. 199–214, New York: Wiley, 1968.)

Bernoulli, N. (1713) Letter in French to P.R. de Montmort, 9 September. In P.R. de Montmort *Essay d'analyse sur les jeux de hazard*, (2nd edn), pp. 401–402. Paris: Jacque Quillau. (Reprinted in New York: Chelsea, 1980.)

Csörgő, S. (2002) Rates of merge in generalized St. Petersburg games. *Acta Sci. Math.* (*Szeged*), **68**, 815–847.

Csörgő, S. and Dodunekova, R. (1991) Limit theorems for the Petersburg game. In M.G. Hahn, D.M. Mason and D.C. Weiner (eds.), *Sums, Trimmed Sums and Extremes*, Progr. Probab. 23, pp. 285–315. Boston: Birkhäuser.

Csörgő, S. and Simons, G. (1996) A strong law of large numbers for trimmed sums, with applications to generalized St Petersburg games. *Statist. Probab. Lett.*, **26**, 65–73.

Csörgő, S. and Simons, G. (2002) The two-Paul paradox and the comparison of infinite expectations. In I. Berkes, E. Csáki and M. Csörgő (eds), *Limit Theorems in Probability and Statistics*, Vol. I, pp. 427–455. Budapest: János Bolyai Mathematical Society.

Csörgő, S. and Simons, G. (2008) *Resolution of the St. Petersburg Paradox*. Book manuscript in preparation.

Dutka, J. (1988) On the St. Petersburg paradox. *Arch. Hist. Exact Sci.*, **39**, 13–39.

Feller, W. (1945) Note on the law of large numbers and 'fair' games. *Ann. Math. Statist.*, **16**, 301–304.

Feller, W. (1968) *An Introduction to Probability Theory and its Applications*, Vol. I (3rd edn). New York: Wiley.

Gnedenko, B.V. and Kolmogorov, A.N. (1954) *Limit Distributions for Sums of Independent Random Variables*. Cambridge, MA: Addison-Wesley.

Jorland, G. (1987) The Saint Petersburg paradox 1713–1937. In L. Krüger, L.D. Daston and M. Heidelberger (eds), *The Probabilistic Revolution, Vol. I: Ideas in History*, pp. 157–190. Cambridge, MA: MIT Press.

Martin-Löf, A. (1985) A limit theorem which clarifies the 'Petersburg paradox'. *J. Appl. Probab.*, **22**, 634–643.

Samuelson, P.A. (1977) St. Petersburg paradoxes: defanged, dissected, and historically described. *J. Economic Lit.*, **15**, 24–55. (Republished as Chapter 298 in K. Crowley (ed), *The Collected Scientific Papers of Paul A. Samuelson*, Vol. 5, pp. 133–164. Cambridge, MA: MIT Press, 1986.)

Zolotarev, V.M. (1978) Pseudomoments. *Teor. Verojatnost. i Primenen.*, **23**, 284–294 (in Russian).

Received September 2005 and revised February 2006