Strong approximation of the St. Petersburg game

István Berkes[1]

Dedicated to Professor Paul Doukhan on the occasion of his 60th birthday

Abstract

Let $X, X_1, X_2, \ldots$ be i.i.d. random variables with $P(X = 2^k) = 2^{-k}$ $(k \in \mathbb{N})$ and let $S_n = \sum_{k=1}^n X_k$. The properties of the sequence S_n have received considerable attention in the literature in connection with the St. Petersburg paradox (Bernoulli 1738). Let $\{Z(t), t \geq 0\}$ be a semistable Lévy process with underlying Lévy measure $\sum_{k \in \mathbb{Z}} 2^{-k} \delta_{2^k}$. For a suitable version of (X_k) and $Z(t)$, we prove the strong approximation $S_n = Z(n) + O(n^{5/6+\varepsilon})$ a.s. This provides the first example for a strong approximation theorem for partial sums of i.i.d. sequences not belonging to the domain of attraction of the normal or stable laws.

Keywords: St. Petersburg game, a.s. invariance principle, semistable laws.

MSC 2000: 60F15, 60F17, 60G50.

1 Introduction

Let $X, X_1, X_2, \ldots$ be i.i.d. random variables with $P(X = 2^k) = 2^{-k}$ $(k = 1, 2, \ldots)$ and put $S_n = \sum_{k=1}^n X_k$. The study of the sequence $\{S_n, n \in \mathbb{N}\}$ has received considerable attention in the literature in connection with the St. Petersburg paradox (Bernoulli 1738) concerning the "fair" entry price for a game where the winnings are distributed according to X. Martin-Löf [14] proved that

$$S_{2^m}/2^m - m \xrightarrow{d} G \tag{1}$$

where G is the semistable distribution with characteristic function $\exp(g(t))$, where

$$g(t) = \sum_{l=-\infty}^{0} (\exp(it2^l) - 1 - it2^l)2^{-l} + \sum_{l=1}^{\infty} (\exp(it2^l) - 1)2^{-l}. \tag{2}$$

[1]) Graz University of Technology, Institute of Statistics, Kopernikusgasse 24, 8010 Graz, Austria. Email: berkes@tugraz.at. Research supported by FWF grant P24302-N18 and OTKA grant K 108615.

He also proved ([14], Theorem 1) that if $n_k \sim \gamma 2^k$, $1 \le \gamma < 2$, then

$$S_{n_k}/n_k - \operatorname{Log} n_k \xrightarrow{d} G_\gamma$$

where G_γ denotes the distribution with characteristic function $\exp(\gamma g(t/\gamma) - it \operatorname{Log} \gamma)$ and $\operatorname{Log} n$ denotes logarithm with base 2. Letting $\gamma_n = n/2^{[\operatorname{Log} n]}$ (where $[\cdot]$ denotes integral part), Csörgő [9] proved that

$$\sup_x \left| P\left(\frac{S_n}{n} - \operatorname{Log} n \le x\right) - G_{\gamma_n}(x)\right| \longrightarrow 0 \quad \text{as } n \to \infty \tag{3}$$

and determined the precise rate of convergence. Relation (3) shows that the class of subsequential limit distributions of $S_n/n - \operatorname{Log} n$ is the class

$$\mathcal{G} = \{G_\gamma,\, 1 < \gamma \le 2\}. \tag{4}$$

Moreover, if n runs through the interval $[2^m, 2^{m+1}]$ then, with error tending to 0 as $m \to \infty$, the distribution of the variable $S_n/n - \operatorname{Log} n$ runs through the elements of the discrete set

$$\{G_\gamma,\, \gamma = 1 + j2^{-m}, j = 0, 1, \dots, 2^m\}.$$

(Note that $G_1 = G_2$, so that the motion is 'circular' in $\mathcal{G}$.) This remarkable behavior was called *merging* in [9]. Csörgő and Dodunekova [11] showed that merging holds for extremal and trimmed sums of the sequence (X_n) as well and Berkes, Horváth and Schauer [5] and del Barrio, Janssen and Pauly [1] proved that the same holds for bootstrapped sums of (X_n).

Let $Z(t)$ denote the Lévy process defined by

$$E(\exp(iuZ(t)) = \exp(tg(u)). \tag{5}$$

The process $Z(t)$ has been introduced by Martin-Löf [14] who proved the scaling relation

$$g(2^m t) = 2^m (g(t) - imt).$$

From this it follows that the transformation $t \longrightarrow 2t$ does not change the distribution of the process

$$\{Z(t)/t - \operatorname{Log} t,\, t > 0\}. \tag{6}$$

In particular, $Z(2)/2 - 1 \stackrel{d}{=} Z(1)$, and since $Z(2) \stackrel{d}{=} Z(1) \star Z(1)$, the distribution of $Z(1)$ is semistable. In view of the atomic Lévy measure in the characteristic function of $Z(1)$, its distribution is not stable. It also follows that

$$Z(n)/n - \operatorname{Log} n \stackrel{d}{=} Z(\gamma_n)/\gamma_n - \operatorname{Log} \gamma_n \stackrel{d}{=} G_{\gamma_n},$$

showing that the distribution of the sequence $Z(n)/n - \operatorname{Log} n$ exhibits the merging behavior (3) in an ideal way, i.e. the left hand side of (3) is equal to 0 for all n. Hence in analogy with strong approximation theory under finite variances, it is natural to ask if the process $\{S_n, n \ge 1\}$ can be approximated, in the almost sure sense, by the semistable process $\{Z(n),\, n \ge 1\}$ with a good remainder term. Such an approximation would naturally yield much more information on the behavior of the partial sums S_n than their weak limit behavior. The purpose of this paper is to prove such a strong approximation result. More precisely, we will prove the following

Theorem. *Let $X, X_1, X_2, \dots$ be i.i.d. r.v.'s with $P(X = 2^k) = 2^{-k}$ $(k = 1, 2, \dots)$ and let $S_n = \sum_{k \le n} X_k$. Let $Z(t)$ be the Lévy process defined by (5), with g given by (2). Then without changing their distributions, the processes $\{S_n,\ n \ge 1\}$ and $\{Z(n),\ n \ge 1\}$ can be defined on a common probability space such that*

$$|S_n - Z(n)| = O(n^{5/6+\varepsilon}) \qquad a.s. \tag{7}$$

for any $\varepsilon > 0$.

As in the case of i.i.d. sequences with finite variances, our theorem implies the functional (Donsker type) version of (1), as well as the almost sure central limit theorem in [2]. As the deductions are routine, we omit the details.

Our theorem can be extended for the class of i.i.d. sequences $X, X_1, X_2, \dots$ satisfying

$$P(X > x) = c_1 x^{-\alpha} \psi(\log x), \quad P(X \le -x) = c_2 x^{-\alpha} \psi(\log x) \qquad (x \ge x_0)$$

for some $x_0 > 0$, where $c_1 \ge 0$, $c_2 \ge 0$, $0 < \alpha < 2$, are constants and ψ is a bounded periodic function. However, since the proof requires lengthy calculations and no new ideas, we do not give the details here. Note that such i.i.d. sequences belong to the domain of geometric attraction of semistable laws, see Grinevich and Khokhlov [13] for a precise characterization of this class in terms of characteristic functions. Also, as shown by Csörgő and Megyesi [12], for partial sums of i.i.d. sequences belonging to this class, an analogue of the merging relation (3) holds.

It seems likely that the exponent 5/6 in (7) is far from optimal, but since for applications all exponents < 1 suffice and we do not know the optimal exponent, we will not investigate this problem here. Finding the optimal remainder term is unsolved even in the case of stable limit distributions. In the case of symmetric X, upper bounds for the remainder term in the stable case are given in [3], [6], [15], while lower bounds are given in [6]. For example, in [6] it is shown that if X is symmetric with

$$P(X > x) = (c + \beta(x)) x^{-\alpha}, \quad 0 < \alpha < 2,\ x \ge x_0$$

where $\beta(x) = (\log x)^{-\gamma}$, $\gamma > 0$, then the partial sums $\sum_{k=1}^n X_k$ can be approximated with a stable Lévy process $Z(n)$ with a.s. remainder term $O(n^{1/\alpha}(\log n)^\tau)$ for $\tau = 1/\alpha - \gamma/\alpha + \varepsilon$, but not for $\tau = 1/\alpha - 2\gamma/\alpha - \gamma - \varepsilon$. Similar results hold for slower decreasing functions $\beta(x)$. On the other hand, the proof of lower bounds in [6] breaks down if β decreases at least polynomially, thus even in the simplest symmetric case when $P(X > x) = cx^{-\alpha}$ $(0 < \alpha < 2, x \ge x_0)$ no lower bounds are known. In case of the St. Petersburg variable X it follows from the results of [4] that the difference $|P(X > x) - P(Y > x)|$ of the tails of X and the limiting semistable variable Y is $O(x^{-(1+\gamma)})$ for some $\gamma > 0$ and again the method of [6] yields no lower bounds in the invariance principle. As Csörgő [9] showed, the precise convergence speed in (3) is $O((\log n)^2/n)$ (cf. also Lemma 2 below), which is better than the classical Berry-Esseen bound $O(n^{-1/2})$ for i.i.d. sequences with finite third moments. Thus it is conceivable that the remainder term $O(n^{5/6+\varepsilon})$ in our strong approximation theorem can be improved beyond $O(n^{1/2})$, but this remains open.

2 Proof

Let $Y_1, Y_2, \dots$ be i.i.d. random variables with distribution G having characteristic function $\exp(g(t))$ with g defined by (2). Then letting $Z^*(n) = \sum_{k=1}^n Y_k$, the processes $\{Z(n),\ n \ge 1\}$ and $\{Z^*(n),\ n \ge 1\}$ have the same distribution and thus our theorem states equivalently that the sequences $(X_k), (Y_k)$ can be defined jointly on a suitable probability space such that

$$\sum_{k=1}^{n} (X_k - Y_k) = O(n^{5/6+\varepsilon}) \qquad \text{a.s.} \tag{8}$$

Our proof will use a modification of the standard blocking technique. Using a remainder term in the merging theorem in [9], the blocking method yields the approximation (8) along a polynomially growing sequence (t_k) of n's. Unfortunately, the fluctuation of the partial sums of X_n and Y_n in the intervals $[t_k, t_{k+1}]$ are too large for extending the approximation (8) for all n. However, as we are going to see, the difficulty is caused by a single large term X_i and Y_j within $[t_k, t_{k+1}]$, and using a special coupling ensuring that the indices of the maximal terms of the sequences (X_n) and (Y_n) in the blocks $[t_k, t_{k+1}]$ coincide, then removing these terms and using a minimax inequality of Billingsley [8] instead of a standard maximal inequality resolves the difficulty. This idea was used by Berkes, Dabrowski, Dehling and Philipp [3] in the context of stable $\mathbb{R}^d$-valued sequences and appears to have many further applications for heavy tailed sequences.

Lemma 1. *We have*

$$P\left\{\left|\frac{S_n}{n} - \operatorname{Log} n\right| > x\right\} \le \frac{9}{x} \qquad (x \ge 9,\ n = 1, 2, \dots). \tag{9}$$

Relation (9) remains valid if we replace S_n by $\tilde{S}_n = \sum_{i \le n} \tilde{X}_i$, where $\tilde{X}_1, \tilde{X}_2, \dots$ are i.i.d. random variables with characteristic function $\exp(g(u))$.

For the proof of (9) see Berkes, Csáki and Csörgő [2]; the proof of the second relation is similar.

Lemma 2. *For any $n \ge 1$ we have*

$$\pi\left(dist\left(\frac{S_n}{n} - \operatorname{Log} n\right),\ G_{\gamma_n}\right) \le C\frac{(\log n)^2}{n} \tag{10}$$

for some absolute constant $C > 0$, where π denotes the Prohorov distance.

This follows from Theorem 1 of Csörgő [9].

Lemma 3. *Let $\{\tau_j : 1 \le j \le n\}$ be a finite sequence of i.i.d. random variables with sum $S = \sum_{j\le n} \tau_j$. Assume that the distribution function of $|\tau_1|$ is continuous. Then L, defined by $|\tau_L| = \max_{1\le j\le n} |\tau_j|$, is with probability one, a well defined random variable that is independent of S and has uniform distribution on $\{1, 2, \dots, n\}$.*

Clearly, the distribution of L is uniform on $\{1, 2, \dots, n\}$; in fact, it is uniform conditionally on any symmetric function of $\tau_1, \dots, \tau_n$, i.e. it is independent of S.

Proof of the Theorem. We first enlarge the probability space to carry an i.i.d. sequence (ζ_n) of standard normal r.v.'s, which is also independent of (X_n). By the LIL for (ζ_n), it suffices to prove the theorem for the sequence (X_n^*) where $X_n^* = X_n + \zeta_n$. Also,

$$P\left(\left|n^{-1}\sum_{k=1}^{n}\zeta_k\right| \ge x\right) \le \exp(-nx^2/2) \le C/x, \qquad (x \ge 1) \tag{11}$$

with some constant C and thus Lemma 1 remains valid, with possibly different constants, for the sequence (X_k^*). Further, (11) implies that

$$P\left(\left|n^{-1}\sum_{k=1}^{n}\zeta_k\right| \ge (\log n)/\sqrt{n}\right) \le C''(\log n)^2/\sqrt{n},$$

and thus Lemma 2 also remains valid for (X_n^*) with $(\log n)^2/n$ on the right hand side of (10) replaced by $(\log n)^2/\sqrt{n}$. Since in the rest of the proof of the theorem we use the properties of the St. Petersburg sequence (X_n) only through Lemmas 1 and 2, in the sequel we can drop the stars and let X_n denote the perturbed version of X_n. As a consequence, the X_n have continuous distribution.

Let

$$t_k = [k^\rho], \quad n_k = t_{k+1} - t_k, \quad H_k = (t_k, t_{k+1}] \tag{12}$$

for some $\rho > 3$ chosen suitably later and

$$\xi_k = n_k^{-1}\left(\sum_{j\in H_k} X_j - n_k \operatorname{Log} n_k\right). \tag{13}$$

The modified version of Lemma 2 implies that the Prohorov distance of the distribution of ξ_k and of G_{γ_k} is $\ll (\log n_k)^2/\sqrt{n_k}$ and since the underlying probability space is atomless (because of the continuity of the distribution of the X_k's), the proof of Theorem 2 of Berkes and Philipp [7] shows that on the same probability space there exists a sequence $\{\eta_k, k \ge 1\}$ of independent random variables such that η_k is measurable with respect to $\sigma\{\xi_1, \dots, \xi_k\}$, it has distribution G_{γ_k} and

$$P\{|\xi_k - \eta_k| \ge \alpha_k\} \le \alpha_k \tag{14}$$

where

$$\alpha_k \ll \frac{(\log n_k)^2}{\sqrt{n_k}} \ll \frac{(\log k)^2}{k^{(\rho-1)/2}}. \qquad (15)$$

Here, and in the sequel, $\ll$ means the same as the O notation.

Define L_k by $|X_{t_k+L_k}| = \max_{j\in H_k}|X_j|$. Since the X_k have continuous distribution, Lemma 3 shows that L_k is defined uniquely with probability 1, has uniform distribution on $(0, n_k] \cap \mathbf{Z}$ and is independent of ξ_k, and consequently of the whole sequence $\{\xi_k, k \geq 1\}$. Since the L_k are independent, it follows that $\{L_k,\, k \geq 1\}$ is independent of $\{\xi_k,\, k \geq 1\}$ and since η_k is measurable with respect to $\sigma\{\xi_1, \ldots, \xi_k\}$, it follows that $\{L_k,\, k \geq 1\}$ is independent of $\{\eta_k,\, k \geq 1\}$. In other words, the joint distribution of L_k and η_k is the same as that of L_k and ξ_k, a fact that will enable us to guarantee that the L_k will also be the location of the maximum the block $\{Y_j,\, j \in H_k\}$ of the approximating sequence (Y_k) (still to be constructed).

Let $\{Y_i, i \geq 1\}$ be a sequence of independent random variables, defined on some probability space and with common characteristic function $\exp(g(u))$. Denote by L_k^* the random variable defined by $|Y_{t_k+L_k^*}| = \max_{j\in H_k}|Y_j|$. Since the distribution of Y_i is continuous (in fact, Y_i has an infinitely many times differentiable density, see Csörgő [10]), by Lemma 3 L_k^* is well-defined, has uniform distribution on $(0, n_k] \cap \mathbf{Z}$ and is independent of

$$\eta_k^* = n_k^{-1}\left(\sum_{j\in H_k} Y_j - n_k \operatorname{Log} n_k\right).$$

As we noted above, η_k^* has distribution G_{γ_k} and thus the sequence $\{(\eta_k^*, L_k^*), k \geq 1\}$ has the same distribution as $\{(\eta_k, L_k),\ k \geq 1\}$. We apply Lemma A1 of Berkes and Philipp [7] to the joint law F of the sequences $\{\xi_i, i \geq 1, \eta_k, k \geq 1\}$ and $\{(\eta_k, L_k), k \geq 1\}$ and the joint law G of the sequences $\{(\eta_k^*, L_k^*), k \geq 1\}$ and $\{Y_i, i \geq 1\}$ and the spaces $S_1 = \mathbf{R}^\infty \times \mathbf{R}^\infty$, $S_2 = (\mathbf{R} \times \mathbf{N})^\infty$, $S_3 = \mathbf{R}^\infty$. We obtain a joint law Q with marginals F and G, which we realize on some probability space Ω'. Hence, keeping the same notation we can set $\eta_k = \eta_k^*$ and $L_k = L_k^*$.

In summary, we have redefined the sequences $\{X_i, i \geq 1\}$, $\{\xi_k, k \geq 1\}$ and $\{L_k, k \geq 1\}$ without changing their joint law on a (possibly) new probability space, together with a sequence $\{Y_i, i \geq 1\}$ of i.i.d. random variables with common characteristic function $\exp(g(u))$ with the following properties:

$$\eta_k = n_k^{-1}\left(\sum_{i\in H_k} Y_i - n_k \operatorname{Log} n_k\right), \quad |Y_{t_k+L_k}| = \max_{i\in H_k}|Y_i|, \qquad (16)$$

i.e. the location $t_k + L_k$ of $\max_{i\in H_k}|X_i|$ and $\max_{i\in H_k}|Y_i|$ is the same.

This together with (13) yields:

$$\sum_{j\in H_k}(X_j - Y_j) = n_k(\xi_k - \eta_k). \qquad (17)$$

Using (14) and (15) and since $\rho > 3$ implies $\sum_{k=1}^{\infty} \alpha_k < \infty$, we get, using the Borel–Cantelli lemma,

$$|\xi_k - \eta_k| \ll \alpha_k \ll \frac{(\log k)^2}{k^{(\rho-1)/2}} \quad \text{a.s. as} \quad k \to \infty$$

and hence using (12) we find

$$\left| \sum_{i \le t_k} (X_i - Y_i) \right| \ll \sum_{j \le k-1} n_j \alpha_j \ll k^{(\rho+1)/2} (\log k)^2 \ll t_k^{(\rho+1)/2\rho} (\log t_k)^2 \quad \text{a.s.} \tag{18}$$

This estimates the difference $|\sum_{i \le n} (X_i - Y_i)|$ for all n of the form $n = t_k$. For general n we need the following lemmas.

Lemma 4. *With probability 1 we have for any $\varepsilon > 0$ and sufficiently large k that*

$$\max_{n \in H_k} \min \left\{ \left| \sum_{t_k < j \le n} X_j \right|, \left| \sum_{n < j \le t_{k+1}} X_j \right| \right\} \le 2 t_k^{1-1/(2\rho)+\varepsilon} \tag{19}$$

and a similar statement holds for the Y_j's.

Proof. Let $a_0 = 0$ and $a_j = j \operatorname{Log} j$ for $j \ge 1$. We claim that

$$P(|(S_j - S_i) - (a_j - a_i)| \ge \lambda) \le \frac{18(j-i)}{\lambda} \operatorname{Log} N \qquad \text{for } 1 \le i < j \le N. \tag{20}$$

Clearly, (20) holds for $\lambda < 18(j-i) \operatorname{Log} N$, since then the right hand side exceeds 1. Assume now $\lambda \ge 18(j-i) \operatorname{Log} N$. Then we have, observing that $|a_j - a_i| \le 2(j-i) \operatorname{Log} N$ by the mean value theorem and trivially $a_{j-i} \le (j-i) \operatorname{Log} N$, we get

$$\begin{aligned} &P(|(S_j - S_i) - (a_j - a_i)| \ge \lambda) = P(|S_{j-i} - (a_j - a_i)| \ge \lambda) \\ &\le P(|S_{j-i}| \ge 8\lambda/9) \le P(|S_{j-i} - a_{j-i}| \ge \lambda/2) \le 18(j-i)/\lambda, \end{aligned}$$

where in the last step we used Lemma 1. Thus we proved (20) and letting $\bar{X}_k = X_k - (a_k - a_{k-1})$, $\bar{S}_n = \sum_{k \le n} \bar{X}_k = S_n - a_n$, we get by the independence of the $\bar{X}_j$ for any $1 \le i \le j \le k \le N$ and $\lambda > 0$,

$$P\left\{|\bar{S}_j - \bar{S}_i| \ge \lambda, |\bar{S}_k - \bar{S}_j| \ge \lambda\right\} \le \frac{324}{\lambda^2}(j-i)(k-j) \operatorname{Log}^2 N \le \frac{324}{\lambda^2}(k-i)^2 \operatorname{Log}^2 N.$$

Hence using Theorem 12.1 of Billingsley [8] with $\gamma = 1$, $\alpha = 1$ and $u_j = 18 \operatorname{Log} N$, we get for any $N \ge 1$ and $\lambda > 0$,

$$P\left\{ \max_{1 \le k \le N} \min\{|\bar{S}_k|, |\bar{S}_N - \bar{S}_k|\} \ge \lambda \right\} \le C \frac{1}{\lambda^2} N^2 \operatorname{Log}^2 N \tag{21}$$

for some absolute constant $C > 0$. Clearly, replacing $\bar{S}_k$ and $\bar{S}_N$ in (21) with S_k and S_N, the random variable in the brackets on the left hand side of (21) changes at most by $N \operatorname{Log} N$ and thus

$$P\left\{ \max_{1 \le k \le N} \min\{|S_k|, |S_N - S_k|\} \ge \lambda + N \operatorname{Log} N \right\} \le C \frac{1}{\lambda^2} N^2 \operatorname{Log}^2 N. \tag{22}$$

Hence choosing $N = n_k$, $\lambda = t_k^{1-1/(2\rho)+\varepsilon}$ and using stationarity and the Borel-Cantelli lemma, we get the statement of Lemma 4 for the X_j's. The proof for the Y_j's is the same.

Lemma 5. *With probability* 1 *there exists a* k_0 *such that for all* $k \ge k_0$ *there is at most one index* $j \in H_k$ *with* $|X_j| > t_k^{1-1/(2\rho)+\varepsilon}$.

Proof. Since $P(|X_1| > t) = O(1/t)$, we have

$$P\left\{\min(|X_i|, |X_j|) > t_k^{1-1/(2\rho)+\varepsilon} \text{ for some } i \ne j \in H_k\right\} \le n_k^2 P^2\left\{|X_1| > t_k^{1-1/(2\rho)+\varepsilon}\right\} \le$$
$$\ll k^{2\rho-2} t_k^{-(2-1/\rho+2\varepsilon)} \ll k^{-(1+2\rho\varepsilon)}$$

by (12). The result follows now from the Borel–Cantelli lemma.

We now show that with probability 1 for sufficiently large k we have

$$\max_{n\in H_k} \min\left\{\left|\sum_{t_k<j\le n}(X_j - Y_j)\right|, \left|\sum_{n<j\le t_{k+1}}(X_j - Y_j)\right|\right\} \le 16 t_k^{1-1/(2\rho)+\varepsilon}. \tag{23}$$

In other words, with probability 1 for any $n \in H_k$, $k \ge k_0$, $\sum_{j\le n}(X_j - Y_j)$ differs from one of the sums $\sum_{j\le t_k}(X_j - Y_j)$ and $\sum_{j\le t_{k+1}}(X_j - Y_j)$ by at most $16 t_k^{1-1/(2\rho)+\varepsilon}$. Then choosing ρ close to 3, (18) and (23) imply conclusion (8) of our theorem.

To prove (23), assume first that

$$\omega \in E_k := \left\{\left|\sum_{j\in H_k} X_j\right| \ge 6 t_k^{1-1/(2\rho)+\varepsilon}\right\}.$$

Then by Lemma 4, for every $n \in H_k$ one of the sums in the brackets in (19) is $\le 2t_k^{1-1/(2\rho)+\varepsilon}$ in absolute value and the other is $\ge 4t_k^{1-1/(2\rho)+\varepsilon}$. Let n run through the interval H_k and for each n we consider which of the two sums in (19) is smaller in absolute value. For the smallest value $n = t_k + 1$, we have $|X_n| \le 2t_k^{1-1/(2\rho)+\varepsilon}$ a.s. by $P(|X_n| > t) = O(1/t)$, $t_k = [k^\rho]$, $\rho > 3$ and the Borel-Cantelli lemma. Thus for this n, the first sum in (19) is smaller. For the same reason, for $n = t_{k+1} - 1$ the second sum is smaller. Hence if n runs through H_k, at least at one location n the minimum in (19) must switch from the first sum to the second sum. Clearly, at this location we have $|X_j| \ge 2t_k^{1-1/(2\rho)+\varepsilon}$ and thus by Lemma 5 there is with probability 1 at most one switch and this occurs at the index n where $|X_n|$ takes its maximum over H_k, i.e. at $n = t_k + L_k$. (Since the location of the maximum is uniformly distributed over H_k, the Borel-Cantelli lemma shows that with probability 1 for $k \ge k_0$ the maximal term cannot occur for $n = t_k + 1$ or $n = t_{k+1}$, and thus the switch occurs inside H_k.) Thus we proved that with probability 1 for $k \ge k_0$ we have

$$\left|\sum_{t_k<j\le n} X_j\right| \le 2t_k^{1-1/(2\rho)+\varepsilon}, \quad t_k < n < t_k + L_k \tag{24}$$

and

$$\Big|\sum_{n<j\le t_{k+1}} X_j\Big| \le 2t_k^{1-1/(2\rho)+\varepsilon}, \quad t_k+L_k\le n\le t_{k+1}. \tag{25}$$

The same conclusion holds if $\omega\in E_k^c$, with the constant 2 in (24), (25) replaced by 8, since in this case Lemma 4 implies that both sums in (19) are at most $8t_k^{1-1/(2\rho)+\varepsilon}$. These inequalities remain valid if X_j is replaced by Y_j since the locations of the maxima in the blocks are the same for the X and the Y process. Thus we proved

$$\Big|\sum_{t_k<j\le n} (X_j-Y_j)\Big| \le 16t_k^{1-1/(2\rho)+\varepsilon}, \quad t_k<n<t_k+L_k$$

and

$$\Big|\sum_{n<j\le t_{k+1}} (X_j-Y_j)\Big| \le 16t_k^{1-1/(2\rho)+\varepsilon}, \quad t_k+L_k\le n\le t_{k+1},$$

completing the proof of (23).

Acknowledgement. The author is indebted to an anonymous referee for his/her valuable comments and suggestions leading to a substantial improvement of the paper.

References

[1] del Barrio, E., Janssen, A. and Pauly, M.: The $m(n)$ out of $k(n)$ bootstrap for partial sums of St. Petersburg type games. Electron. Commun. Probab. 18 (2013), 1–10.

[2] Berkes, I., Csáki, E. and Csörgő, S.: Almost sure limit theorems for the St. Petersburg game. Stat. Prob. Letters 45 (1999), 23–30.

[3] Berkes, I., Dabrowski, A., Dehling, H., and Philipp, W.: A strong approximation theorem for sums of random vectors in the domain of attraction to a stable law. Acta Math. Hung. 48 (1986), 161–172.

[4] Berkes, I, Győrfi, L. and Kevei, P.: Tail probabilities of St. Petersburg sums, trimmed sums, and their limit. J. Theor. Probability, to appear.

[5] Berkes, I., Horváth, L. and Schauer, J.: Noncentral limit theorems for random selections. Prob. Theory Rel. Fields 147 (2010), 449-479.

[6] Berkes, I. and Dehling, H.: Almost sure and weak invariance principles for random variables attracted by a stable law. Probab. Theory Rel. Fields 83 (1989), 331-353.

[7] Berkes, I. and Philipp, W.: Approximation theorems for independent and weakly dependent random vectors, Annals of Probability, 7 (1979), 29–54.

[8] Billingsley, P. Convergence of probability measures. Wiley, 1968.

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

[10] Csörgő, S.: Fourier analysis of semistable distributions. Acta Appl. Math. 96 (2007), 159–174.

[11] Csörgő, S. and Dodunekova, R.: Limit theorems for the Petersburg game, in: Trimmed sums and extremes (M. G. Hahn, D. M. Mason, D. C. Weiner, eds.) pp. 285–315. Progress in Probability, Vol. 23, Birkhäuser, Boston 1991.

[12] Csörgő, S., Megyesi, Z.: Merging to semistable laws. Teor. veroyatnost. i primenen. 47 (2002), 90-109 in Russian). [English version: Theory Probab. Appl. 47 (2002), 17–33.

[13] Grinevich, I.V. and Khokhlov, Y.S.: The domains of attraction of semistable laws. Teor. veroyatn. primen. 40 (1995), 417–422 (in Russian). English version: Probab. Theory Appl. 40 (1995), 361–366.

[14] Martin-Löf, A.: A limit theorem which clarifies the 'Petersburg paradox'. J. Appl. Probability 22 (1985), 634–643.

[15] Stout, W.: Almost sure invariance principles when $EX_1^2 = \infty$. Z. Wahrscheinlichkeitstheorie verw. Gebiete 49 (1979), 23–32.