The Annals of Probability
2003, Vol. 31, No. 3, 1228–1240

DONSKER'S THEOREM FOR SELF-NORMALIZED PARTIAL SUMS PROCESSES

BY MIKLÓS CSÖRGŐ,[1] BARBARA SZYSZKOWICZ[1] AND QIYING WANG

Carleton University, Carleton University and Australian National University

Let $X, X_1, X_2, \dots$ be a sequence of nondegenerate i.i.d. random variables with zero means. In this paper we show that a self-normalized version of Donsker's theorem holds only under the assumption that X belongs to the domain of attraction of the normal law. A thus resulting extension of the arc sine law is also discussed. We also establish that a weak invariance principle holds true for self-normalized, self-randomized partial sums processes of independent random variables that are assumed to be symmetric around mean zero, if and only if $\max_{1\le j\le n}|X_j|/V_n \to_P 0$, as $n\to\infty$, where $V_n^2=\sum_{j=1}^n X_j^2$.

1. Introduction and main results. Let $X, X_1, X_2, \dots$ be a sequence of nondegenerate i.i.d. random variables and let

$$S_n=\sum_{j=1}^n X_j, \qquad V_n^2=\sum_{j=1}^n X_j^2, \qquad n=1,2,\dots.$$

The classical weak invariance principle states that, on an appropriate probability space, as $n\to\infty$,

$$\sup_{0\le t\le 1}\left|\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{[nt]}(X_j-EX_j)-\frac{1}{\sqrt{n}}W(nt)\right|=o_P(1) \quad \text{if and only if} \quad \operatorname{Var}(X)=\sigma^2<\infty, \tag{1}$$

where $\{W(t), 0\le t<\infty\}$ is a standard Wiener process. This invariance principle in probability is a stronger version of Donsker's classical functional central limit theorem. The normalizer $(n\sigma^2)^{-1/2}$ in (1) is that in the classical central limit theorem when $\operatorname{Var}(X)<\infty$.

In contrast to the well-known classical central limit theorem, Giné, Götze and Mason (1997) obtained the following self-normalized version of the central limit theorem. As $n\to\infty$,

$$\frac{1}{V_n}\sum_{j=1}^n (X_j-EX_j)\xrightarrow{\mathcal{D}} N(0,1) \quad \text{if and only if} \quad \lim_{x\to\infty}\frac{x^2P(|X|>x)}{EX^2 I_{(|X|\le x)}}=0. \tag{2}$$

Received October 2001; revised May 2002.
[1]Supported by NSERC Canada grants.
AMS 2000 subject classifications. Primary 60F05, 60F17; secondary 62E20.
Key words and phrases. Donsker's theorem, self-normalized sums, arc sine law.

The latter condition is well known to be equivalent to saying that X belongs to the domain of attraction of the normal law. This beautiful theorem was conjectured by Logan, Mallows, Rice and Shepp (1973). For a short summary of developments that have eventually led to Gine, Götze and Mason (1997), we refer to the Introduction of the latter paper.

The result in (2) shows that when the normalizer in the classical central limit theorem is replaced by an appropriate sequence of random variables then the central limit theorem holds under a weaker moment condition than in the classical case. Thus, in the light of (2), it is natural to ask whether a self-normalized version of the weak invariance principle (1) could also hold under the same weaker assumption. As the following theorem shows, the answer to this paramount question is affirmative.

THEOREM 1. *As $n \to \infty$, the following statements are equivalent*:

(a) $EX = 0$ *and X is in the domain of attraction of the normal law*.

(b) $S_{[nt_0]}/V_n \to_{\mathcal{D}} N(0, t_0)$ *for* $t_0 \in (0, 1]$.

(c) $S_{[nt]}/V_n \to_{\mathcal{D}} W(t)$ *on* $(D[0, 1], \rho)$, *where ρ is the sup-norm metric for functions in* $D[0, 1]$, *and* $\{W(t), 0 \le t \le 1\}$ *is a standard Wiener process*.

(d) *On an appropriate probability space for* $X, X_1, X_2, \dots,$ *we can construct a standard Wiener process* $\{W(t), 0 \le t < \infty\}$ *such that*

$$\sup_{0 \le t \le 1} \left| S_{[nt]}/V_n - W(nt)/\sqrt{n} \right| = o_P(1).$$

Assuming appropriate conditions, we mention two immediate analogs of Theorem 1 when $\{X_j, j \ge 1\}$ is a sequence of independent random variables with $EX_j = 0$ and finite variances EX_j^2. Write $s_n^2 = \sum_{j=1}^n EX_j^2$. If the Lindeberg condition holds, namely,

$$\text{for all } \varepsilon > 0, \qquad s_n^{-2} \sum_{j=1}^n EX_j^2 I_{(|X_j| > \varepsilon s_n)} \to 0 \qquad \text{as } n \to \infty,$$

then it is readily seen that $V_n^2/s_n^2 \to_P 1$. Hence it follows easily from classical results [e.g., Prohorov (1956)] that $S_{K_n(t)}/V_n \to_{\mathcal{D}} W(t)$ on $(D[0, 1], \rho)$, where $K_n(t) = \sup\{m : s_m^2 \le t s_n^2\}$.

By using a similar method as in the proof of Theorem 1 (cf. Section 2), we can also redefine $\{X_j, j \ge 1\}$ on a richer probability space together with a sequence of independent normal random variables $\{Y_j, j \ge 1\}$ with mean zero and $\mathrm{Var}(Y_j) = \mathrm{Var}(X_j)$ such that

$$\sup_{0 \le t \le 1} \left| S_{[nt]}/V_n - \sum_{j=1}^{[nt]} Y_j/s_n \right| = o_P(1)$$

provided that the Lindeberg condition holds.

Furthermore, we prove also the following result for self-normalized, self-randomized partial sums processes of independent random variables.

THEOREM 2. *Let $X_1, X_2, \dots$ be independent symmetric random variables around mean zero. Then*

$$\max_{1 \le j \le n} |X_j|/V_n \xrightarrow{P} 0 \qquad \text{as } n \to \infty, \tag{3}$$

if and only if

$$S_{\tilde{K}_n(t)}/V_n \xrightarrow{\mathcal{D}} W(t) \qquad \text{on } (D[0,1], \rho), \tag{4}$$

where $\tilde{K}_n(t) = \sup\{m : V_m^2 \le t V_n^2\}$.

We mention that (3) is equivalent to the condition that X is in the domain of attraction of the normal law if $\{X_j, j \ge 1\}$ is a sequence of i.i.d. random variables [cf. O'Brien (1980)]. Also, it is readily seen that the Lindeberg condition implies (3). However, it is not clear at this moment whether or not Theorem 2 still holds for general independent random variables, that is, without assuming $\{X_j, j \ge 1\}$ to be symmetric. In the i.i.d. case, for X being symmetric, Griffin and Mason (1991) attribute to Roy Erikson the proof of (2). That $S_n/V_n \to_D N(0,1)$, as $n \to \infty$, with $X_1, X_2, \dots$ as in Theorem 2, is due to Egorov (1996). This result in turn inspired us to prove Theorem 2.

The proofs of Theorems 1 and 2 will be given in the next section. We conclude this section with some immediate corollaries of Theorem 1, which are also of independent interest. With $x \ge 0$, write

$$G_1(x) = P\Big(\sup_{0 \le t \le 1} W(t) \le x\Big), \qquad G_2(x) = P\Big(\sup_{0 \le t \le 1} |W(t)| \le x\Big),$$

$$G_3(x) = P\Big(\int_0^1 W^2(t)\,dt \le x\Big), \qquad G_4(x) = P\Big(\int_0^1 |W(t)|\,dt \le x\Big).$$

Our first corollary is an extension of the original Erdős and Kac (1946) invariance principle to the corresponding functionals of self-normalized sums.

COROLLARY 1. *Let $EX = 0$ and X be in the domain of attraction of the normal law. Then, as $n \to \infty$, we have*

(i) $P(\max_{1 \le k \le n} S_k/V_n \le x) \to G_1(x)$ *for* $x \ge 0$, *and* $P(\min_{1 \le k \le n} S_k/V_n \le x) \to 1 - G_1(-x)$ *for* $x < 0$;

(ii) $P(\max_{1 \le k \le n} |S_k|/V_n \le x) \to G_2(x)$ *for* $x \ge 0$;

(iii) $P(n^{-1} \sum_{k=1}^n (S_k/V_n)^2 \le x) \to G_3(x)$ *for* $x \ge 0$;

(iv) $P(n^{-1} \sum_{k=1}^n |S_k/V_n| \le x) \to G_4(x)$ *for* $x \ge 0$.

We note in passing that the same results also hold true for the corresponding functionals of $S_{\tilde{K}_n(\cdot)}/V_n$ as in Theorem 2.

Erdős and Kac (1947) gave a further demonstration of their (1946) invariance principle by deducing a general form of Lévy's arc sine law (1939) via assuming a central limit theorem. Namely, let $X_1, X_2, \dots$ be independent random variables with $EX_j = 0$, $EX_j^2 = 1$ and assume that Lindeberg's condition holds true, that is, as $n \to \infty$, we have $n^{-1/2} S_n \to_D N(0, 1)$. Then,

$$\lim_{n\to\infty} P(\Pi_n/n \le x) = (2/\pi) \arcsin \sqrt{x}, \qquad 0 \le x \le 1, \tag{5}$$

where $\Pi_n = \sum_{j=1}^{n} I_{0<S_j<\infty}$, that is, Π_n denotes the number of positive elements in the sequence of $S_1, \dots, S_n$.

Lévy (1939) found this arc sine law for Brownian motion (Wiener process) and also referred to connection with the coin tossing game. For an insightful treatise on Lévy's method, we refer to Takács (1981). In addition to Lévy's method for Brownin motion and the Erdős and Kac (1947) invariance principle for obtaining (5) as stated here for partial sums of independent random variables having second moments, we mention also that Sparre and Andersen (1949) discovered a combinatorial proof that revealed the surprising fact that the arc sine law also held true for partial sums of i.i.d. ramdom variables with a continuous and symmetric distribution whose second moment is not neceessarily finite. In this regard then it is interesting to note that another direct application of Theorem 1 yields the following result in the i.i.d case.

COROLLARY 2. *Assume that* $EX = 0$ *and* X *is in the domain of attraction of the normal law*. *Then* (5) *holds true in this case as well.*

Further to Theorem 1, we note also that in Csörgő, Szyszkowicz and Wang (CsSzW) (2001) we prove optimal weighted approximations for the sequence of self-normalized partial sum processes $\{S_{[nt]}/V_n, 0 \le t \le 1\}$, while in CsSzW (2003) we investigate the asymptotic behaviour in distribution of $\max_{1\le k\le n} S_k/V_k$ as well as the LIL for S_n/V_n.

2. Proofs.

PROOF OF THEOREM 1. The statement (b) implies (a) by an immediate restatement of Theorem 3.3 of Giné, Götze and Mason (1997). It is obvious that (d) implies (c) and hence also (b). So, it only needs to be shown that (a) implies (d).

For the sake of proving the latter, we first provide and list some lemmas that are also of independent interest. For convenience, throughout the paper,

$$l(x) := EX^2 I_{(|X|\le x)},$$

and we shall denote an absolute constant by A, which may differ from one place to another in the text.

LEMMA 1. *The following statements are equivalent*:

(a) $l(x)$ *is a slowly varying function at* ∞;
(b) $x^2 P(|X| > x) = o(l(x))$;
(c) $x E|X| I_{(|X|>x)} = o(l(x))$;
(d) $E|X|^\alpha I_{(|X|\le x)} = o(x^{\alpha-2} l(x))$ *for* $\alpha > 2$.

PROOF. It follows from Theorem 2 of Feller [(1966), page 275] that (a) holds if and only if (b) does. If (b) holds, then (c) follows from Lemma 6.2 of Griffin and Kuelbs (1989) with $\theta \downarrow 0$, and by noting that

$$\begin{aligned} E|X|^\alpha I_{(|X|\le x)} &= \int_0^x y^\alpha \, dP(|X| \le x) \\ &= x^\alpha P(|X| > x) + \alpha \int_0^x y^{\alpha-1} P(|X| > y)\, dy, \end{aligned} \tag{6}$$

we get (d). On the other hand, it can be easily shown that (c) implies (b) and (d) implies (b) via using (6) again. Therefore, the proof of Lemma 1 is now complete. □

The next result is due to Sakhanenko (1980, 1984, 1985).

LEMMA 2. *Let* $X_1, X_2, \ldots$ *be independent random variables with* $EX_j = 0$ *and* $\sigma_j^2 = EX_j^2 < \infty$ *for each* $j \ge 1$. *Then we can redefine* $\{X_j, j \ge 1\}$ *on a richer probability space together with a sequence of independent* $N(0, 1)$ *random variables*, Y_j, $j \ge 1$, *such that for every* $p > 2$ *and* $x > 0$,

$$P\left\{\max_{i\le n} \left| \sum_{j=1}^i X_j - \sum_{j=1}^i \sigma_j Y_j \right| \ge x \right\} \le (Ap)^p x^{-p} \sum_{j=1}^n E|X_j|^p,$$

where A *is an absolute positive constant.*

LEMMA 3. *Let* a_j, $j \ge 1$, *be a sequence of nonnegative constants and put* $A(n) = \sum_{j=1}^n a_j$. *If* $a_{n+1}/A(n) \to 0$ *as* $n \to \infty$, *then we have*

$$A^{-1/2}(n) \sum_{j=1}^{n-1} a_{j+1} A^{-1/2}(j) = O(1). \tag{7}$$

If in addition $A([tn])/A(n) \to 1$ *for any* $t > 0$ *as* $n \to \infty$, *then*

$$[nA(n)]^{-1/2} \sum_{j=1}^{n-1} j^{1/2} a_{j+1} A^{-1/2}(j) = o(1) \tag{8}$$

and

$$\frac{1}{nA(n)} \sum_{j=1}^{n-1} j\, a_{j+1} = o(1). \tag{9}$$

PROOF. To prove (7), we assume without loss of generality that $a_{j+1}/A(j) \le 1/2$ for all $j \ge 1$. Noting that $1 \le A(j+1)/A(j) \le 3/2$ for $j \ge 1$ and $\sqrt{1+y} \le 1 + y/2$ for $y \ge 0$, we get

$$\begin{aligned}
I(n) &:= A^{-1/2}(n) \sum_{j=1}^{n-1} a_{j+1} A^{-1/2}(j) \\
&= A^{-1/2}(n) \sum_{j=1}^{n-1} A^{1/2}(j+1)\big[\big(1 + a_{j+1}A^{-1}(j)\big)^{1/2} - 1\big] \\
&\quad + A^{-1/2}(n) \sum_{j=1}^{n-1} \big(A^{1/2}(j+1) - A^{1/2}(j)\big) \\
&\le \tfrac{1}{2} A^{-1/2}(n) \sum_{j=1}^{n-1} a_{j+1} A^{1/2}(j+1) A^{-1}(j) + 1 - a_1 A^{-1/2}(n) \\
&\le \tfrac{1}{2}\Big(\tfrac{3}{2}\Big)^{1/2} I(n) + 1.
\end{aligned}$$

This implies (7), since $\frac{1}{2}(\frac{3}{2})^{1/2} < 1$.

If $A([tn])/A(n) \to 1$, then for any $t > 0$,

$$\frac{A(n) - A([tn])}{A^{1/2}(n)A^{1/2}([tn])} \to 0 \qquad \text{as } n \to \infty.$$

Hence, by using (7), letting $n \to \infty$ and then $t \to 0$, we have

$$\begin{aligned}
&[nA(n)]^{-1/2} \sum_{j=1}^{n-1} j^{1/2} a_{j+1} A^{-1/2}(j) \\
&\quad \le t^{1/2} A^{-1/2}(n) \sum_{j=1}^{[tn]} a_{j+1} A^{-1/2}(j) + A^{-1/2}(n) \sum_{j=[tn]}^{n-1} a_{j+1} A^{-1/2}(j) \\
&\quad \le O(1)t^{1/2} + \frac{A(n) - A([tn])}{A^{1/2}(n)A^{1/2}([tn])} = o(1).
\end{aligned}$$

This gives (8). The proof of (9) is similar, and hence omitted. This also completes the proof of Lemma 3. □

We now are ready to prove that (a) implies (d). Put $b = \inf\{x \ge 1 : l(x) > 0\}$ and

$$\eta_j = \inf\left\{s : s \ge b+1, \frac{l(s)}{s^2} \le \frac{1}{j}\right\}, \qquad j = 1, 2, \ldots.$$

Furthermore, let $B_n^2 = nl(\eta_n)$,

$$X_j^* = X_j I_{(|X_j| \le \eta_j)} \quad \text{and} \quad S_n^* = \sum_{j=1}^{n} X_j^*.$$

By Lemma 2, we can redefine $\{X_j, j \ge 1\}$ on a richer probability space together with a sequence of independent $N(0, 1)$ random variables, $Y_j,\ j \ge 1$, such that for any $x > 0$,

$$\begin{aligned} &P\left\{\max_{i\le n}\left|\sum_{j=1}^{i}(X_j^* - EX_j^*) - \sum_{j=1}^{i}\sigma_j^* Y_j\right| \ge x\right\} \\ &\qquad \le Ax^{-3}\sum_{j=1}^{n} E|X|^3 I_{(|X|\le\eta_j)}, \end{aligned} \tag{10}$$

where $\sigma_j^{*2} = \mathrm{Var}(X_j^*)$. Let $\{W(t), 0 \le t < \infty\}$ be a standard Wiener process such that

$$W(n) = \sum_{j=1}^{n} Y_j, \qquad n = 1, 2, 3, \ldots.$$

We have

$$\begin{aligned} &\sup_{1/n\le t\le 1}\left|S_{[nt]}/V_n - W(nt)/\sqrt{n}\right| \\ &\quad \le \sup_{1/n\le t\le 1}\left|B_n^{-1}\sum_{j=1}^{[nt]}\sigma_j^* Y_j - n^{-1/2}W(nt)\right| \\ &\qquad + \sup_{1/n\le t\le 1}\left|B_n^{-1}\big(S_{[nt]}^* - ES_{[nt]}^*\big) - B_n^{-1}\sum_{j=1}^{[nt]}\sigma_j^* Y_j\right| \\ &\qquad + \sup_{1/n\le t\le 1}\left|\frac{1}{V_n}S_{[nt]} - \frac{1}{B_n}\big(S_{[nt]}^* - ES_{[nt]}^*\big)\right| \\ &\quad := I_1(n) + I_2(n) + I_3(n). \end{aligned} \tag{11}$$

Therefore, (d) will follow from (a) if we can prove that, as $n \to \infty$,

$$I_j(n) = o_P(1), \qquad j = 1, 2, 3, \tag{12}$$

on assuming that $EX = 0$ and X is in the domain of attaction of the normal law.

We now proceed to prove (12). Since X belongs to the domain of attraction of the normal law, that is, $x^2 P(|X| > x) = o(l(x))$, by Lemma 1 $l(x)$ is a slowing varying function at ∞. Hence it can be easily shown that

$$jl(\eta_j) \leq \eta_j^2 \leq (j+1)l(\eta_j) \qquad \text{for } j \geq 1, \tag{13}$$

$$l(\eta_{j+1})/l(\eta_j) \to 1 \qquad \text{as } j \to \infty, \tag{14}$$

$$|EX_j^*| \leq E|X|I_{(|X|>\eta_j)} = o\big(\eta_j^{-1}l(\eta_j)\big) = o(B_j/j) \qquad \text{as } j \to \infty, \tag{15}$$

$$\mathrm{Var}(X_j^*) = EX_j^{*2} - (EX_j^*)^2 = \big(1+o(1)\big)l(\eta_j) \qquad \text{as } j \to \infty, \tag{16}$$

$$E|X_j^*|^3 \leq E|X|^3 I_{(|X|\leq\eta_n)} = o\big(\eta_n l(\eta_n)\big) = o(B_n^3/n) \qquad \text{as } j \to \infty, \tag{17}$$

and, as $n \to \infty$,

$$\frac{1}{B_n^2}\sum_{j=1}^n X_j^2 \to 1 \qquad \text{in probability.} \tag{18}$$

Let $\eta_0 = 0$. Noting that $l(\eta_n) = \sum_{k=1}^n EX^2 I_{(\eta_{k-1}<|X|\leq\eta_k)}$ and that (14) implies $l(\eta_{[tn]})/l(\eta_n) \to 1$ for any fixed $t > 0$, by (13) and Lemma 3, we get, as $n \to \infty$,

$$\begin{aligned} &\frac{1}{nl(\eta_n)}\sum_{j=1}^n EX^2 I_{(\eta_j<|X|\leq\eta_n)} \\ &\qquad \leq \frac{1}{nl(\eta_n)}\sum_{k=1}^{n-1}(k+1)EX^2 I_{(\eta_k<|X|\leq\eta_{k+1})} \to 0 \end{aligned} \tag{19}$$

and

$$\begin{aligned} &\frac{1}{n^{1/2}l^{1/2}(\eta_n)}\sum_{j=1}^n E|X|I_{(\eta_j<|X|\leq\eta_n)} \\ &\qquad \leq \frac{1}{n^{1/2}l^{1/2}(\eta_n)}\sum_{j=1}^{n-1} jE|X|I_{(\eta_j<|X|\leq\eta_{j+1})} \\ &\qquad \leq \frac{1}{n^{1/2}l^{1/2}(\eta_n)}\sum_{j=1}^{n-1}\frac{j^{1/2}}{l^{1/2}(\eta_j)}EX^2 I_{(\eta_j<|X|\leq\eta_{j+1})} \to 0. \end{aligned} \tag{20}$$

It follows from (16) and (19) that

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^n\left(\frac{\sigma_j^*}{l^{1/2}(\eta_n)} - 1\right)^2 &\leq \frac{2}{n}\sum_{j=1}^n\frac{o(1)l(\eta_j)}{l(\eta_n)} + \frac{2}{nl(n)}\sum_{j=1}^n\big(l^{1/2}(\eta_j) - l^{1/2}(\eta_n)\big)^2 \\ &\leq o(1) + \frac{2}{nl(n)}\sum_{j=1}^n EX^2 I_{(\eta_j<|X|\leq\eta_n)} \\ &\to 0 \qquad \text{as } n \to \infty. \end{aligned}$$

This, together with Kolmogrov's inequality, implies that for any $\varepsilon > 0$,

$$P\left(\max_{1\le k\le n}\left|\sum_{j=1}^{k}\left(\frac{\sigma_j^*}{l^{1/2}(\eta_n)}-1\right)Y_j\right|\ge \varepsilon n^{1/2}\right)\le \frac{1}{\varepsilon^2 n}\sum_{j=1}^{n}\left(\frac{\sigma_j^*}{l^{1/2}(\eta_n)}-1\right)^2\to 0,$$

and hence, as $n\to\infty$, $I_1(n)=o_P(1)$.

To estimate $I_2(n)$, let $Z_j = X_j^* - EX_j^* - \sigma_j^* Y_j$. It follows from (10) and (17) that for any $\varepsilon > 0$,

$$\begin{aligned}
&P\left(\sup_{1/n\le t\le 1}\left|B_n^{-1}\sum_{j=1}^{[nt]}Z_j\right|\ge\varepsilon\right)\\
&\quad= P\left(\max_{1\le k\le n}\left|\sum_{j=1}^{k}Z_j\right|\ge \varepsilon B_n\right)\\
&\quad\le \frac{A}{(\varepsilon B_n)^3}\sum_{k=1}^{n}E|X|^3 I_{(|X|\le\eta_j)}\le \frac{A}{(\varepsilon B_n)^3}nE|X|^3 I_{(|X|\le\eta_n)}\\
&\quad= o(1)\qquad \text{as } n\to\infty.
\end{aligned}$$

This gives that $I_2(n)=o_P(1)$ as $n\to\infty$.

As to $I_3(n)$, we have

$$\begin{aligned}
I_3(n)&\le \frac{1}{V_n}\sup_{1/n\le t\le 1}\left|S_{[nt]}-S_{[nt]}^*+ES_{[nt]}^*\right|\\
&\qquad+\left|\frac{1}{V_n}-\frac{1}{B_n}\right|\sup_{1/n\le t\le 1}\left|S_{[nt]}^*-ES_{[nt]}^*\right|\\
&\le \frac{1}{V_n}\sum_{j=1}^{n}\left(|X_j|I_{(|X_j|>\eta_j)}+E|X_j|I_{(|X_j|>\eta_j)}\right)\\
&\qquad+\left|\frac{B_n}{V_n}-1\right|\sup_{1/n\le t\le 1}B_n^{-1}\left|S_{[nt]}^*-ES_{[nt]}^*\right|\\
&:= I_3^{(1)}(n)+I_3^{(2)}(n).
\end{aligned}\tag{21}$$

By Markov's inequality, (13), (15) and (20), we obtain, for any $\varepsilon > 0$,

$$\begin{aligned}
&P\left(B_n^{-1}\sum_{j=1}^{n}\left(|X_j|I_{(|X_j|>\eta_j)}+E|X_j|I_{(|X_j|>\eta_j)}\right)\ge\varepsilon\right)\\
&\quad\le \frac{2n}{\varepsilon B_n}E|X|I_{(|X|>\eta_n)}+\frac{2\varepsilon^{-1}}{n^{1/2}l^{1/2}(\eta_n)}\sum_{k=1}^{n}E|X|I_{(\eta_k<|X|\le\eta_n)}\\
&\quad\to 0\qquad \text{as } n\to\infty,
\end{aligned}$$

and hence, as $n \to \infty$, $B_n^{-1} \sum_{j=1}^n (|X_j| I_{(|X_j|>\eta_j)} + E|X_j| I_{(|X_j|>\eta_j)}) = o_P(1)$. This, together with (18), implies that, as $n \to \infty$,

$$I_3^{(1)}(n) = \frac{B_n}{V_n} B_n^{-1} \sum_{j=1}^n \big(|X_j| I_{(|X_j|>\eta_j)} + E|X_j| I_{(|X_j|>\eta_j)}\big) = o_P(1). \tag{22}$$

We continue to use the notations $I_1(n)$ and $I_2(n)$ introduced in (11). Noting that $\sup_{0<t\le 1} |n^{-1/2} W(nt)| = O_P(1)$ and using the estimators above for $I_1(n)$ and $I_2(n)$, it can be easily shown that

$$\begin{aligned}
&\sup_{1/n\le t\le 1} B_n^{-1} \big|S_{[nt]}^* - E S_{[nt]}^*\big| \\
&\quad \le I_1(n) + I_2(n) + \sup_{0<t\le 1} \big|n^{-1/2} W(nt)\big| = O_P(1)
\end{aligned}$$

as $n \to \infty$. Hence, by using (18) again, as $n \to \infty$, we obtain

$$I_3^{(2)}(n) = \left|\frac{B_n}{V_n} - 1\right| \sup_{1/n\le t\le 1} B_n^{-1} \big|S_{[nt]}^* - E S_{[nt]}^*\big| = o_P(1). \tag{23}$$

It now follows from (21)–(23) that, as $n \to \infty$,

$$I_3(n) \le I_3^{(1)}(n) + I_3^{(2)}(n) = o_P(1).$$

On collecting the estimators above for $I_j(n)$, $j = 1, 2, 3$, we obtain the desired (12). This also completes the proof of Theorem 1. $\square$

PROOF OF THEOREM 2. If the statement (4) holds, then $S_n/V_n \to_D N(0, 1)$ as $n \to \infty$, and hence (3) follows from Theorem 2 of Egorov (1996).

We next prove that (3) implies (4). We assume without loss of generality that $X_1, \ldots, X_n$ are defined on a probability space which also supports a sequence of independent Rademacher random variables $\varepsilon_1, \ldots, \varepsilon_n$ that are independent of $X_1, \ldots, X_n$. In view of symmetry of X_j and independence of X_j and ε_j, it is readily seen that

$$\sum_{j=1}^{\tilde{K}_n(t)} X_j / V_n \stackrel{D}{=} \sum_{j=1}^{\tilde{K}_n(t)} X_j \varepsilon_j / V_n, \qquad 0 \le t \le 1, \tag{24}$$

where $\stackrel{D}{=}$ denotes equality in distribution. Write $X_n(t) = \sum_{j=1}^{\tilde{K}_n(t)} X_j \varepsilon_j / V_n$. By using (24) and classical methods of weak convergence [cf. Billingsley (1968), Chapters 2 and 3], it suffices to show that

(a) for all $0 \le t_1 \le t_2 \le \cdots \le t_k \le 1$ and $k \ge 1$,

$$\big(X_n(t_1), \ldots, X_n(t_k)\big) \xrightarrow{D} \big(W(t_1), \ldots, W(t_k)\big), \tag{25}$$

and that

(b) for each $\varepsilon > 0$,

$$\lim_{h\to 0}\limsup_{n\to\infty} P\biggl(\sup_{|t-s|\le h}\bigl|X_n(t)-X_n(s)\bigr| > \varepsilon\biggr) = 0. \tag{26}$$

We first verify tightness, that is, (b). Let P' and E' denote conditional probability and conditional expectation respectively, given $X_1, X_2, \dots$. Recalling the definition of $\tilde{K}_n(t)$, it is readily seen that for any $\varepsilon > 0$,

$$\begin{aligned}
&P'\biggl(\sup_{kh<t\le\min\{(k+1)h,1\}}|X_n(t)-X_n(kh)| > \varepsilon\biggr)\\
&\quad\le P'\Biggl(\max_{\tilde{K}_n(kh)\le r-1\le\tau_n(h)}\Biggl|\sum_{j=\tilde{K}_n(kh)+1}^{r} X_j\varepsilon_j\Biggr| > \varepsilon V_n\Biggr)\\
&\quad\le 4\varepsilon^{-4}V_n^{-4}E'\Biggl(\sum_{j=\tilde{K}_n(kh)+1}^{\tau_n(h)+1} X_j\varepsilon_j\Biggr)^4\\
&\quad\le A\varepsilon^{-4}V_n^{-4}\Biggl(\sum_{j=\tilde{K}_n(kh)+1}^{\tau_n(h)+1} X_j^2\Biggr)^2\\
&\quad\le A\varepsilon^{-4}V_n^{-2}\sum_{j=\tilde{K}_n(kh)+1}^{\tau_n(h)+1} X_j^2\Bigl(h+\max_{1\le j\le n} X_j^2/V_n^2\Bigr),
\end{aligned}$$

where $\tau_n(h) = \min\bigl[\tilde{K}_n\{(k+1)h\}, n-1\bigr]$. Therefore, for any $\varepsilon > 0$,

$$\begin{aligned}
&P\biggl(\sup_{|t-s|\le h}\bigl|X_n(t)-X_n(s)\bigr| > \varepsilon\biggr)\\
&\quad\le \sum_{k:kh\le 1} P\biggl(\sup_{kh<t\le\min\{(k+1)h,1\}}|X_n(t)-X_n(kh)| > \varepsilon - \max_{1\le j\le n}|X_j|/V_n\biggr)\\
&\quad\le \sum_{k:kh\le 1} E\biggl[P'\biggl(\sup_{kh<t\le\min\{(k+1)h,1\}}|X_n(t)-X_n(kh)| > \varepsilon/2\biggr)\biggr] \qquad (27)\\
&\qquad + \frac{1}{h}P\Bigl(\max_{1\le j\le n}|X_j| \ge \varepsilon V_n/2\Bigr)\\
&\quad\le A\varepsilon^{-4}\Bigl(h + h^{-1}E\max_{1\le j\le n} X_j^2/V_n^2\Bigr).
\end{aligned}$$

Noting that (3) implies that $\lim_{n\to\infty} E\max_{1\le j\le n} X_j^2/V_n^2 = 0$, (26) follows from (27) by letting $n\to\infty$ first and then $h\to 0$.

The proof of convergence of finite-dimensional distributions as in (25) is similar to that of Theorem 2 of Egorov (1996) with minor modifications, and hence details are omitted. The proof of Theorem 2 is now complete. □

PROOF OF COROLLARY 2. The proof of Corollary 2 coincides with that of Billingsley [(1968), page 138], establishing (5) in the i.i.d. case. We rewrite it here for convenience, for it is a short one.

For $x(t) \in D[0, 1]$, let $\lambda(x(\cdot))$ be the Lebesgue measure of the set of t for which $x(t) > 0$. Then λ is measurable with respect to $(D, \mathcal{D})$, where $\mathcal{D}$ denotes the σ-field of subsets of D generated by the finite-dimensional subsets of D, and is continuous except on a set of Wiener measure 0. Now, if $S_n(t) := S_{[nt]}/V_n$, then $\lambda(S_n(\dots))$ is exactly $1/n$ times the number of positive sums among $S_1, \dots, S_{n-1}$. Hence, Theorem 1 and the continuous mapping theorem imply that (5) holds. This completes the proof of Corollary 2. □

Acknowledgments. The authors thank a referee and Editors for their valuable comments and suggestions.

NOTE ADDED IN PROOF. While attending the 8th International Vilnius Conference on Probability Theory and Mathematical Statistics in June 2002, we learned that, using a different method than that of CsSzW (2001) and the present paper, the equivalence of (a), (b) and (c) as in Theorem 1 was also proved in Račkauskas, A. and Suquet, Ch. (2000, 2001).

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Mesures*. Wiley, New York.

CSÖRGŐ, M., SZYSZKOWICZ, B. and WANG, Q. (2001). Donsker's theorem and weighted approximations for self-normalized partial sums processes. Technical Report 360, Laboratory for Research in Statistics and Probability, Carleton Univ. and Univ. of Ottawa.

CSÖRGŐ, M., SZYSZKOWICZ, B. and WANG, Q. (2003). Darling–Erdős theorems for self-normalized sums. *Ann. Probab.* **31** 676–692.

EGOROV, V. A. (1996). On the asymptotic behavior of self-normalized sums of random variables. *Teor. Veroyatnost. i Primenen.* **41** 643–650.

ERDŐS, P. and KAC, M. (1946). On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.* **52** 292–302.

ERDŐS, P. and KAC, M. (1947). On the number of positive sums of independent random variables. *Bull. Amer. Math. Soc.* **53** 1011–1020.

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

GINE, E., GÖTZE, F. and MASON, D. M. (1997). When is the Student t-statistic asymptotically standard normal? *Ann. Probab.* **25** 1514–1531.

GRIFFIN, P. S. and KUELBS, J. D. (1989). Self-normalized laws of the iterated logatithm. *Ann. Probab.* **17** 1571–1601.

GRIFFIN, P. S. and MASON, D. M. (1991). On the asymptotic normality of self-normalized sums. *Proc. Cambridge Phil. Soc.* **109** 597–610.

LÉVY, P. (1939). Sur certains processus stochastiques homogènes. *Compositio Math.* **7** 283–339.

LOGAN, B. F., MALLOWS, C. L., RICE, S. O. and SHEPP, L. A. (1973). Limit distributions of self-normalized sums. *Ann. Probab.* **2** 642–651.

O'BRIEN, G. L. (1980). A limit theorem for sample maxima and heavy branches in Galton–Watson trees. *J. Appl. Probab.* **17** 539–545.

PROHOROV, YU. V. (1956). Covergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.* **1** 157–214.

RAČKAUSKAS, A. and SUQUET, CH. (2000). Convergence of self-normalized partial sums processes in $C[0,1]$ and $D[0,1]$. *Pub. IRMA Lille 53—VI*. Preprint.

RAČKAUSKAS, A. and SUQUET, CH. (2001). Invariance principles for adaptive self-normalized partial sums processes. *Stochastic Process. Appl.* **95** 63–81.

SAKHANENKO, A. I. (1980). On unimprovable estimates of the rate of convergence in invariance principle. *Colloquia Math. Soc. János Bolyai* **32** 779–783.

SAKHANENKO, A. I. (1984). On estimates of the rate of convergence in the invariance principle. In *Advances in Probability Theory*: *Limit Theorems and Related Problems* (A. A. Borovkov, ed.) 124–135. Springer, New York.

SAKHANENKO, A. I. (1985). Convergence rate in the invariance principle for non-identically distributed variables with exponential moments. In *Advances in Probability Theory*: *Limit Theorems for Sums of Random Variables* (A. A. Borovkov, ed.) 2–73. Springer, New York.

SPARRE-ANDERSEN, E. (1949). On the number of positive sums of random variables. *Skand. Aktuarietidskrift* 27–36.

TAKÁCS, L. (1981). The arc sine law of Paul Lévy. In *Contributions to Probability. A Collection of Papers Dedicated to Eugene Lukács* (J. Gani and V. K. Rohatgi, eds.). Academic Press, New York.

M. CSÖRGŐ
B. SZYSZKOWICZ
SCHOOL OF MATHEMATICS
AND STATISTICS
CARLETON UNIVERSITY
1125 COLONEL BY DRIVE
OTTAWA
CANADA K1S 5B6
E-MAIL: mcsorgo@math.carleton.ca
bszyszko@math.carleton.ca

Q. WANG
CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
MATHEMATICAL SCIENCES INSTITUTE
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA, ACT 0200
AUSTRALIA
E-MAIL: qiying@wintermute.anu.edu.au