Localized large sums of random variables

Kevin Ford 1,∗

Department of Mathematics, University of Illinois at Urbana-Champaign, 1409 West Green St., Urbana, IL, 61801, USA

Gérald Tenenbaum

Institut Élie Cartan, Université Henri–Poincaré Nancy 1, B.P. 239, 54506 Vandœuvre-lès-Nancy Cedex, France

Abstract

We study large partial sums, localized with respect to the sums of variances, of a sequence of centered random variables. An application is given to the distribution of prime factors of typical integers.

Key words: sums of random variables, prime divisors
1991 MSC: 60G50, 11N25

Dedicated to the memory of Walter Philipp

1 Introduction

Consider random variables $X_1, X_2, \ldots$ with $\mathbb{E} X_j = 0$ and $\mathbb{E} X_j^2 = \sigma_j^2$. Let

$$S_n = X_1 + \cdots + X_n, \quad s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2,$$

and assume that (a) $s_n \to \infty$ as $n \to \infty$.  

∗ Corresponding author.
Email addresses: ford@math.uiuc.edu (Kevin Ford), gerald.tenenbaum@iecn.u-nancy.fr (Gérald Tenenbaum).

1 Research supported by National Science Foundation Grants DMS-0301083 and DMS-0555367

Preprint submitted to Elsevier 13 March 2007
Given a positive function \( f_N \geq 1 + 1/N \), we are interested in the behavior of

\[
I = \liminf_{N \to \infty} \max_{N < s_n \leq N f_N} |S_n|/s_n.
\]

If we replace \( \liminf \) by \( \limsup \), it immediately follows from the law of the iterated logarithm that \( I = \infty \) almost surely when \( f_N \) is bounded. Our results answer a question originally raised, in oral form, by A. Sárközy and for which a partial answer had previously been given by the second author, see Chap. 3 of Oon (2005).

2 Independent random variables

Assume that the \( X_j \) are independent. Then \( \mathbb{E}S_n^2 = s_n^2 \). In addition to condition (a), we will work with two other mild assumptions, (b) \( s_{j+1}/s_j \ll 1 \) when \( s_j > 0 \) and (c) for every \( \lambda > 0 \), there is a constant \( c_\lambda > 0 \) such that if \( n \) is large enough and \( s_m^2 > 2s_n^2 \), then

\[
\mathbb{P}(|S_m - S_n| \geq \lambda s_m) \geq c_\lambda.
\]

Condition (b) says that no term in \( S_n \) dominates the others. Condition (c) follows if the Central Limit Theorem (CLT) holds for the sequence of \( S_n \), since CLT for \( S_n \) implies CLT for \( S_m - S_n \) as \( (m-n) \to \infty \). For example, (c) holds for i.i.d. random variables, under the Lindeberg condition

\[
\forall \varepsilon > 0, \lim_{n \to \infty} \sum_{1 \leq j \leq n} \mathbb{E}\left(\frac{X_j^2}{s_n^2} : |X_j| > \varepsilon s_n\right) = 0
\]

and the stronger Lyapunov condition

\[
\exists \delta > 0 : \sum_{1 \leq j \leq n} \mathbb{E}|X_j|^{2+\delta} = o(s_n^{2+\delta}).
\]

Condition (c) is weaker, however, than CLT.

**Theorem 1**

(i) Suppose (a), (b), and \( f_N = (\log N)^M \) for some constant \( M > 0 \). Then \( I < \infty \) almost surely.

(ii) Suppose (a), (b), (c) and \( f_N = (\log N)^{\xi(N)} \) with \( \xi(N) \) tending monotonically to \( \infty \). Then \( I = \infty \) almost surely.

**Remark.** In the first statement of the theorem we show in fact that almost surely

\[
I \leq 15\sqrt{M + 1}\left(\max_{s_j > 0} s_{j+1}/s_j\right)^2.
\]

**Lemma 2 (Kolmogorov’s inequality, 1929)** We have

\[
\mathbb{P}(\max_{1 \leq j \leq k} |S_j| \geq \lambda s_k) \leq 1/\lambda^2 \quad (k \geq 1).
\]
Proof of Theorem 1. By (a) and (b), there is a constant $D$ so that $s_{j+1}/s_j \leq D$ for all large $j$. Define

$$h(n) := \max\{k : s_k^2 \leq n\} \quad (n \in \mathbb{N}^*),$$

so that the conditions $N < s_n^2 \leq Nf_N$ and $h(N) < n \leq h(Nf_N)$ are equivalent.

We first consider the case when $f_N := (\log N)^M$. Let

$$N_j := j^{(M+3)j}, \quad t(j) := \lfloor (M+1)(\log j)/\log 2 \rfloor, \quad H_j := 2^{t(j)},$$

and

$$U_j := h(N_j), \quad U_{j,t} := h(2^t N_j) \quad (0 \leq t \leq t(j)), \quad V_j := h(H_j N_j) = U_{j,t(j)}.$$

It is possible that $U_{j,t+1} = U_{j,t}$ for some $t$. Note that for large $j$, $H_j N_j \geq N_j f_N$.

Let $k$ be a constant depending only on $M$ and $D$. For $j \geq 1$ define the events

$$A_j := \{|S_{V_j}| \leq s_{U_{j+1}}\},$$

$$B_j := \bigcap_{0 \leq t \leq t(j)-1} B_{j,t} \quad \text{where} \quad B_{j,t} := \left\{ \max_{U_{j,t+1} \leq n \leq U_{j+1,t+1}} |S_{U_{j+1,t+1}} - S_n| \leq k s_{U_{j+1,t}} \right\},$$

$$C_j := \{|S_{U_{j+1}} - S_{V_j}| \leq 2 s_{U_{j+1}}\}.$$

By (b) and the definition of $h(N)$, we have

$$D^{-1}\sqrt{2^t N_j} \leq s_{U_{j,t}} \leq \sqrt{2^t N_j}$$

for all $j, t$. It follows from Lemma 2 that

$$\mathbb{P}(\overline{A_j}) \leq D^2 H_j N_j \leq \frac{D^2}{j^2}.$$

Thus, $\sum_{j \geq 1} \mathbb{P}(\overline{A_j}) < \infty$ and hence almost surely there is a $j_0$ so that $A_j$ occurs for $j \geq j_0$. Applying Lemma 2 again yields

$$\mathbb{P}(\overline{B_{j,t}}) \leq \frac{s_{U_{j+1,t+1}}^2 - s_{U_{j+1,t}}^2}{k^2 s_{U_{j+1,t}}^2} \leq \frac{D^2 2^t N_j}{k^2 2^t N_j} = \frac{2D^2}{k^2}.$$

If $k = 3D\sqrt{M} + 1$, then

$$\mathbb{P}(B_j) \geq \left(1 - \frac{2D^2}{k^2}\right)^{t(j)} \geq \frac{1}{j^{1/2}}$$

for large $j$. Also by Lemma 2, $\mathbb{P}(C_j) \geq \frac{3}{4}$, and since $B_j$ and $C_j$ are independent,

$$\sum_{j \geq 1} \mathbb{P}(B_j C_j) = \infty.$$
Since the events $B_jC_j$ are independent, the Borel–Cantelli lemma implies that almost surely the events $B_jC_j$ occur infinitely often. Thus, the event $A_jB_jC_j$ occurs for an infinite sequence of integers $j$. Take such a index $j$, let $n \in [U_{j+1}, V_{j+1}]$ and $U_{j+1, g-1} < n \leq U_{j+1, g}$, where $1 \leq g \leq t(j + 1)$. We have by several applications of (1)

$$|S_n| \leq |S_{V_j}| + |S_{U_{j+1} - S_{V_j}}| + \sum_{0 \leq t \leq g-2} |S_{U_{j+1,t}} - S_{U_{j+1,t-1}}| + |S_n - S_{U_{j+1,g-1}}|$$

$$\leq 3s_{U_{j+1}} + k \sum_{0 \leq t \leq g-1} s_{U_{j+1,t}}$$

$$\leq \left\{3 + k(1 + 2^{1/2} + \cdots + 2^{(g-1)/2})\right\} \sqrt{N_{j+1}}$$

$$\leq 5k \sqrt{2^{g-1}N_{j+1}}$$

$$\leq 5kDs_n = 15D^2(M + 1)^{1/2}s_n.$$  

This completes the proof of part (i) of the theorem, since

$$V_{j+1} \geq h\left(\frac{1}{2}j^{M+1}N_j\right) \geq h(N_j \log^M N_j)$$

for large $j$.

Now suppose $f_N = (\log N)^{\xi(N)}$ with $\xi(N)$ tending monotonically to $\infty$.

Let $\lambda > 0$ be arbitrary and define $K := 2D^2$. Let $N_1^*$ be so large that $f_{N_1^*} \geq K$. For $j \geq 1$ let $N_{j+1}^* = N_j^*K^{u(j)}$, where $u(j) := [\log f_{N_j^*} / \log K]$. Put

$$U_j^* := h(N_j^*), \quad U_{j,t}^* := h(K^tN_j^*) (0 \leq t \leq u(j)).$$

Let $J := [U_j^*, U_{j+1}^*]$ and

$$Y_j := \max_{n \in J_j} |S_n| / s_n.$$  

We have

$$u(j) \geq 1 \Rightarrow N_{j+1}^* \geq KN_j^* \Rightarrow u(j) / \log j \to \infty.$$  

Therefore, by (c), if $j$ is sufficiently large then

$$\mathbb{P}(Y_j \leq \lambda / 2) \leq \prod_{1 \leq t \leq u(j)} \mathbb{P}\left(|S_{U_j^*, t} - S_{U_{j,t-1}^*, t}| \leq \frac{1}{2} \lambda(s_{U_j^*, t} + s_{U_{j,t-1}^*, t+1})\right)$$

$$\leq \prod_{1 \leq t \leq u(j)} \mathbb{P}\left(|S_{U_j^*, t} - S_{U_{j,t-1}^*, t}| \leq \lambda \sqrt{K^tN_j^*}\right)$$

$$\leq (1 - c\lambda)^{u(j)} \leq \frac{1}{j^2}.$$  

Thus

$$\sum_{j \geq 1} \mathbb{P}(Y_j \leq \lambda / 2) < \infty.$$  

Almost surely, $Y_k \leq \lambda / 2$ for only finitely many $k$.  

4
Theorem 1 has an analog for Brownian motion, which follows from Theorem 1 and
the invariance principle.

**Theorem 3** Let $W(t)$ be Brownian motion on $[0, \infty)$. If $f_N = (\log N)^M$ with fixed
$M > 0$, then almost surely

$$I = \liminf_{N \to \infty} \max_{N < t \leq N f_N} \frac{|W(t)|}{\sqrt{t}} < \infty.$$  

If $f_N = (\log N)^{\xi(N)}$ with $\xi(N) \to \infty$, then $I = \infty$ almost surely.

Theorem 3 can be proved directly and more swiftly using the methods used to
establish Theorem 1. By invariance principles (e.g. Philipp, 1986), one may deduce
from Theorem 3 a version of Theorem 1 where stronger hypotheses on the $X_j$ are
assumed. As it stands, now, however, Theorem 1 does not follow from Theorem 3.

3 Dependent random variables

The conclusions of Theorem 1 can also be shown to hold for certain sequences of
weakly dependent random variables by making use of almost sure invariance prin-
ciples. We assume that (d) there exists a sequence of i.i.d. normal random variables
$Y_j$ with $E Y_j^2 = \sigma_j^2$, defined on the same probability space as the sequence of $X_j$,
and such that if $Z_n = Y_1 + \cdots + Y_n$, then

$$|S_n - Z_n| = O(s_n) \quad \text{a.s.}$$

Of course the variables $Y_j$ are dependent on the $X_j$, but not on each other. Property
(d) has been proved for martingale difference sequences, sequences satisfying cer-
tain mixing conditions, and lacunary sequences $X_j = \{n_j \omega\}$ with $\inf n_{j+1}/n_j > 1$,
$\omega$ uniformly distributed in $[0, 1]$ and $\{x\}$ is the fractional part of $x$. See e.g. Philipp
(1986) for a survey of such results.

**Theorem 4** (i) Suppose (a), (b), and (d). If $f_N := (\log N)^M$ for some constant
$M > 0$, then $I < \infty$ almost surely.

(ii) Let $\xi(N)$ tend monotonically to $\infty$ and set $f_N := (\log N)^{\xi(N)}$. Then $I = \infty$
almost surely.

By (d),

$$I = O(1) + \liminf_{N \to \infty} \max_{N < s_n \leq N f_N} |Z_n| / s_n,$$

and we apply Theorem 1 to the sequence of $Y_j$. The variable $Z_n$ is normal with
variance $s_n^2$, hence (c) holds.
4 Prime factors of typical integers

Consider a sequence of independent random variables $Y_p$, indexed by prime numbers $p$, such that $\mathbb{P}(Y_p = 1) = 1/p$ and $\mathbb{P}(Y_p = 0) = 1 - 1/p$. We can think of $Y_p$ as modelling whether or not a “random” integer is divisible by $p$. As $\mathbb{E}Y_p = 1/p$, we form the centered r.v.’s $X_p = Y_p - 1/p$ (we may also define $X_j$ for non-prime $j$ to be zero with probability 1). Let

$$T_n = \sum_{p \leq n} Y_p, \quad S_n = \sum_{p \leq n} X_p.$$  

We have $\mathbb{E}X_p^2 = (1 - 1/p)/p$, hence by Mertens’ estimate

$$s_n^2 = \sum_{p \leq n} \frac{1}{p} - \frac{1}{p^2} = \log n + O(1).$$

Here and in the sequel, $\log_k$ denotes, for integer $k \geq 2$, the $k$-fold iterated logarithm. Since $\mathbb{E}|X_p|^3 \leq 1/p$, the Lyapunov condition holds with $\delta = 1$. Then (a), (b) and (c) hold, and therefore the conclusion of Theorem 1 holds. Here take $D = \max_{n \geq 2} s_{n+1}/s_n$ since $s_1 = 0$.

Let $\omega(m, t)$ denote the number of distinct prime factors of $m$ which are $\leq t$. The sequence $\{T_n : n \geq 1\}$ mimics well the behavior of the function $\omega(m, n)$ for a “random” $m$, at least when $n$ is not too close to $m$. This is known as the Kubilius model. It can be made very precise, see (Elliott , 1979, Ch. 3, especially pp. 119–122) and Tenenbaum (1999) for the sharpest estimate known to date. Suppose $r$ is an integer with $2 \leq r \leq x$ and $r = x^{1/u}$, $\omega_r(m) = (\omega(m,1), \ldots, \omega(m,r))$ and suppose $Q$ is any subset of $\mathbb{Z}^r$. Then, given arbitrary $c < 1$, and uniformly in $x, r$ and $Q$, we have

$$\frac{1}{x} |\{m \leq x : \omega_r(m) \in Q\}| = \mathbb{P}((T_1, \ldots, T_r) \in Q) + O \left( x^{-c} + e^{-u \log u} \right). \quad (2)$$

An analog of Theorem 1, established by parallel estimates, provides via (2) information about localized large values of

$$g(m, t) := |\omega(m, t) - \log_2 t|/\sqrt{\log_2 t}.$$  

**Theorem 5** (i) Let $M > 0$ be fixed, $f_N := (\log N)^M$ and put $K := 30D^2 \sqrt{M + 1}$. If $g = g(m) \to \infty$ monotonically as $m \to \infty$ in such a way that $g^2 f_{g^2} \leq \log_2 m$ for large $m$, then for a set of integers $m$ of natural density 1, we have

$$\max_{g(m) \leq N \leq g(m)^2} \min_{N < \log_2 t \in N f_N} g(m, t) \leq K.$$  

2 A subset $\mathcal{A}$ of $\mathbb{N}^+$ is said to have natural density 1 if $|\mathcal{A} \cap [1, x]| = x + o(x)$ as $x \to \infty$. 

6
(ii) Let $\xi(N) \to \infty$ in such a way that $f_N := (\log N)^{\xi(N)} \leq N$. Suppose that $g(m) \to \infty$ monotonically as $m \to \infty$, that $g(m) \leq (\log_2 m)^{1/10}$, and let

$$I_m := \min_{g(m) \leq N} \max_{N \leq \log_2 m} g(m, t).$$

Then, $I_m \to \infty$ on a set of integers $m$ of natural density 1.

We follow the proof of Theorem 1. Keeping the notation introduced there, we see that for large $J$,

$$P \left( \bigcap_{J \leq j \leq 3J/2} A_j \right) \leq \sum_{J \leq j \leq 3J/2} \frac{D^2}{j^2} + \prod_{J \leq j \leq 3J/2} \left( 1 - \frac{3}{4\sqrt{j}} \right) \leq \frac{1}{J}.$$  

For large $G$, define $J$ by $N_{J+1} < G \leq N_{J+2}$. Then $G^{5/3} > N_{\lfloor 3J/2 \rfloor + 2}$ and $J \gg M (\log G)/\log_2 G$. Thus, for large $G$,

$$P \left( \min_{G \leq N \leq G^{5/3}} \max_{h(N) < n \leq h(Nf_N)} \frac{|S_n|}{s_n} \leq K \right) \geq 1 - O \left( \frac{1}{J} \right) \geq 1 - O \left( \frac{\log_2 G}{\log G} \right).$$

The direct number theoretic analog of $|S_n|/s_n$ is

$$\tilde{g}(m, t) := \frac{|\omega(m, t) - \sum_{p \leq t} 1/p|}{\sqrt{\sum_{p \leq t} (1 - 1/p)/p}}.$$  

By (2), if $G$ is large and $G \leq \sqrt{\log_2 x}$ (so that $G^{5/3} f_{G^{5/3}} \leq (\log_2 x)^{7/8}$), then

$$\frac{1}{\sqrt{x}} \left\{ m \leq x : \min_{G \leq N \leq G^{5/3}} \max_{h(N) < n \leq h(Nf_N)} \tilde{g}(m, t) \leq K \right\} \geq 1 - O \left( \frac{\log_2 G}{\log G} \right).$$

Since $\tilde{g}(m, t) = g(m, t) + O \left( 1/\sqrt{\log_2 t} \right)$, the first part of the theorem follows.

The second part is similar. Note that $\omega(n, x) - \omega(n, x^{1/\sqrt{\log_2 x}}) \leq \sqrt{\log_2 x}$ for $n \leq x$, and, for brevity, write $g = g(\sqrt{x})$. By (2) with $u := \sqrt{\log_2 x}$, we have, for any fixed $K$ and large $x$,

$$\frac{1}{\sqrt{x}} \left\{ m \leq \sqrt{x} \leq x : \min_{N \leq g} \max_{N \leq \log_2 x} \tilde{g}(m, t) \leq K \right\} \leq \frac{1}{\sqrt{x}} \left\{ \sqrt{x} \leq m \leq x : \min_{N \leq g} \max_{N \leq \log_2 x} \tilde{g}(t) \leq K + 2 \right\} + \frac{1}{\sqrt{x}}$$

$$\leq P \left( \inf_{N \leq g} \max_{h(N) < n \leq h(Nf_N)} \frac{|S_n|}{s_n} \leq K + 2 \right) + O \left( \frac{1}{\log_2 x} \right),$$
where \( \mathcal{L}(x) := \log_2 x - \frac{1}{2} \log_3 x \). Since \( f_N \leq N \), we have \( N_{j+1}^* \leq (N_j^*)^2 \) in the notation of the proof of Theorem 1. The interval

\[
\left[ (\log_2 x)^{1/10}, \mathcal{L}(x)^{1/2} \right]
\]

therefore contains at least one interval \( J_j \). By the proof of Theorem 1, for large \( x \), the probability above does not exceed \( \sum_{j \geq j_0} 1/j^2 \leq 1/(j_0 - 1) \), where \( j_0 \to \infty \) as \( x \to \infty \).

**Remarks.** The upper bound \( g^2 \) of \( N \) in the first part can be sharpened. By the same methods, similar results can be proved for a wide class of additive arithmetic functions \( r(m,t) = \sum_{p^a \parallel m} r(p^a) \) in place of \( \omega(m,t) \).

**Acknowledgment.** The authors are indebted to Walter Philipp for helpful discussions on the use of almost sure invariance principles.

**References**


