HAL open science

Localized large sums of random variables

Kevin Ford, Gérald Tenenbaum

To cite this version:

Kevin Ford, Gérald Tenenbaum. Localized large sums of random variables. 2007. hal-00091337v2

HAL Id: hal-00091337

https://hal.science/hal-00091337v2

Preprint submitted on 13 Mar 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Given a positive function $f_N \geqslant 1 + 1/N$, we are interested in the behavior of

$$I = \liminf_{N\to\infty} \max_{N<s_n^2\leqslant N f_N} |S_n|/s_n.$$

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

$$\mathbb{P}\left(|S_m - S_n| \geqslant \lambda s_m\right) \geqslant c_\lambda.$$

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

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \sum_{1\leqslant j\leqslant n} \mathbb{E}\left(X_j^2/s_n^2 : |X_j| > \varepsilon s_n\right) = 0$$

and the stronger Lyapunov condition

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

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 ∞. Then $I = \infty$ almost surely.

Remark. In the first statement of the theorem we show in fact that almost surely $I \leqslant 15\sqrt{M+1}(\max_{s_j>0} s_{j+1}/s_j)^2$.

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

$$\mathbb{P}(\max_{1\leqslant j\leqslant k} |S_j| \geqslant \lambda s_k) \leqslant 1/\lambda^2 \qquad (k \geqslant 1).$$

Proof of Theorem 1. By (a) and (b), there is a constant D so that $s_{j+1}/s_j \leqslant D$ for all large j. Define

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

so that the conditions $N < s_n^2 \leqslant N f_N$ and $h(N) < n \leqslant h(N f_N)$ 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

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

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

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

$$A_j := \{|S_{V_j}| \leqslant s_{U_{j+1}}\},$$
$$B_j := \bigcap_{0 \leqslant t \leqslant t(j)-1} B_{j,t} \;\text{ where }\; B_{j,t} := \left\{ \max_{U_{j+1,t} \leqslant n \leqslant U_{j+1,t+1}} |S_{U_{j+1,t+1}} - S_n| \leqslant k s_{U_{j+1,t}} \right\},$$
$$C_j := \{|S_{U_{j+1}} - S_{V_j}| \leqslant 2 s_{U_{j+1}}\}.$$

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

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

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

$$\mathbb{P}(\overline{A_j}) \leqslant D^2 \frac{H_j N_j}{N_{j+1}} \leqslant \frac{D^2}{j^2}.$$

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

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

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

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

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

$$\sum_{j \geqslant 1} \mathbb{P}(B_j C_j) = \infty.$$

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 $U_{j+1,g-1} < n \leqslant U_{j+1,g}$, where $1 \leqslant g \leqslant t(j+1)$. We have by several applications of (1)

$$\begin{aligned}
|S_n| &\leqslant |S_{V_j}| + |S_{U_{j+1}} - S_{V_j}| + \sum_{0 \leqslant t \leqslant g-2} |S_{U_{j+1,t}} - S_{U_{j+1,t+1}}| + |S_n - S_{U_{j+1,g-1}}| \\
&\leqslant 3s_{U_{j+1}} + k \sum_{0 \leqslant t \leqslant g-1} s_{U_{j+1,t}} \\
&\leqslant \left\{3 + k(1 + 2^{1/2} + \cdots + 2^{(g-1)/2})\right\}\sqrt{N_{j+1}} \\
&\leqslant 5k\sqrt{2^{g-1}N_{j+1}} \\
&\leqslant 5kDs_n = 15D^2(M+1)^{1/2}s_n.
\end{aligned}$$

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

$$V_{j+1} \geqslant h(\tfrac{1}{2}j^{M+1}N_j) \geqslant h(N_j \log^M N_j)$$

for large j.

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

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

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

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

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

We have

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

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

$$\begin{aligned}
\mathbb{P}(Y_j \leqslant \lambda/2) &\leqslant \prod_{1 \leqslant t \leqslant u(j)} \mathbb{P}\left(|S_{U_{j,t}^*} - S_{U_{j,t-1}^*}| \leqslant \tfrac{1}{2}\lambda(s_{U_{j,t}^*} + s_{U_{j,t-1}^*})\right) \\
&\leqslant \prod_{1 \leqslant t \leqslant u(j)} \mathbb{P}\left(|S_{U_{j,t}^*} - S_{U_{j,t-1}^*}| \leqslant \lambda\sqrt{K^t N_j^*}\right) \\
&\leqslant (1 - c_\lambda)^{u(j)} \leqslant \frac{1}{j^2}.
\end{aligned}$$

Thus

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

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

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*

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

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 principles. We assume that (d) there exists a sequence of i.i.d. normal random variables Y_j with $\mathbb{E}Y_j^2 = \sigma_j^2$, defined on the same probability space as the sequence of X_j, and such that if $Z_n = Y_1 + \cdots + Y_n$, then

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

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 certain mixing conditions, and lacunary sequences $X_j = \{n_j\omega\}$ with $\inf n_{j+1}/n_j > 1$, ω 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 ∞ and set $f_N := (\log N)^{\xi(N)}$. Then $I = \infty$ almost surely.

By (d),

$$I = O(1) + \liminf_{N\to\infty} \max_{N<s_n^2\leqslant Nf_N} |Z_n|/s_n,$$

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

$$T_n = \sum_{p \leqslant n} Y_p, \qquad S_n = \sum_{p \leqslant n} X_p.$$

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

$$s_n^2 = \sum_{p \leqslant n} \frac{1}{p} - \frac{1}{p^2} = \log_2 n + O(1).$$

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

Let $\omega(m, t)$ denote the number of distinct prime factors of m which are $\leqslant t$. The sequence $\{T_n : n \geqslant 1\}$ 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 $2 \leqslant r \leqslant x$ and $r = x^{1/u}$, $\boldsymbol{\omega}_r(m) = (\omega(m, 1), \ldots, \omega(m, r))$ and suppose Q is any subset of $\mathbb{Z}^r$. Then, given arbitrary $c < 1$, and uniformly in x, r and Q, we have

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

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

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

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 $g^2 f_{g^2} \leqslant \log_2 m$ for large m, then for a set of integers m of natural density 1,[2] we have*

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

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

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

$$I_m := \min_{\substack{g(m)\leqslant N \\ Nf_N \leqslant \log_2 m}} \max_{N\leqslant \log_2 t \leqslant N f_N} \varrho(m,t).$$

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,

$$\mathbb{P}\left(\bigcap_{J\leqslant j\leqslant 3J/2} \overline{A_j B_j C_j}\right) \leqslant \sum_{J\leqslant j\leqslant 3J/2} \frac{D^2}{j^2} + \prod_{J\leqslant j\leqslant 3J/2}\left(1-\frac{3}{4\sqrt{j}}\right) \ll \frac{1}{J}.$$

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

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

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

$$\widetilde{\varrho}(m,t) := \frac{\left|\omega(m,t) - \sum_{p\leqslant t} 1/p\right|}{\sqrt{\sum_{p\leqslant t}(1-1/p)/p}}.$$

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

$$\frac{1}{x}\left|\left\{m \leqslant x : \min_{G\leqslant N\leqslant G^{5/3}} \max_{h(N)<n\leqslant h(Nf_N)} \widetilde{\varrho}(m,t) \leqslant K\right\}\right| \geqslant 1 - O\left(\frac{\log_2 G}{\log G}\right).$$

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

The second part is similar. Note that $\omega(n,x) - \omega\left(n, x^{1/\sqrt{\log_2 x}}\right) \leqslant \sqrt{\log_2 x}$ for $n \leqslant x$, 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,

$$\begin{aligned}
&\frac{1}{x}\left|\left\{m \leqslant x : \min_{\substack{N\geqslant g \\ Nf_N\leqslant \log_2 m}} \max_{N<\log_2 t\leqslant Nf_N} \widetilde{\varrho}(m,t) \leqslant K\right\}\right| \\
&\qquad \leqslant \frac{1}{x}\left|\left\{\sqrt{x} \leqslant m \leqslant x : \min_{\substack{N\geqslant g \\ Nf_N\leqslant \mathscr{L}(x)}} \max_{N\leqslant \log_2 t\leqslant Nf_N} \widetilde{\varrho}(t) \leqslant K+2\right\}\right| + \frac{1}{\sqrt{x}} \\
&\qquad \leqslant \mathbb{P}\left(\inf_{\substack{N\geqslant g \\ Nf_N\leqslant \mathscr{L}(x)}} \max_{h(N)<n\leqslant h(Nf_N)} \frac{|S_n|}{s_n} \leqslant K+2\right) + O\left(\frac{1}{\log_2 x}\right),
\end{aligned}$$

where $\mathscr{L}(x) := \log_2 x - \frac{1}{2}\log_3 x$. Since $f_N \leqslant N$, we have $N^*_{j+1} \leqslant (N^*_j)^2$ in the notation of the proof of Theorem 1. The interval

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

therefore contains at least one interval J_j. By the proof of Theorem 1, for large x, the probability above does not exceed $\sum_{j \geqslant j_0} 1/j^2 \leqslant 1/(j_0 - 1)$, 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 $r(m,t) = \sum_{p^a \| m} r(p^a)$ 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

P.D.T.A. Elliott, 1979. *Probabilistic Number Theory, I*, Springer-Verlag, New York.

S.M. Oon, 2005. *Construction des suites binaires pseudo-aléatoires*, Thèse d'université, Nancy, Université Henri Poincaré–Nancy 1, UFR STMIA , 110 pp.

W. Philipp, 1986. *Invariance principles for independent and weakly dependent random variables*, in Dependence in Probability and Statistics (Oberwolfach, 1985), 225–268, Progr. Probab. Statist. **11**, Birkhäuser Boston, Boston, MA.

G. Tenenbaum, 1999. Crible d'Ératosthène et modèle de Kubilius, in: K. Győry, H. Iwaniec, J. Urbanowicz (eds.), *Number Theory in Progress*, Proceedings of the conference in honor of Andrzej Schinzel, Zakopane, Poland 1997, 1099–1129, Walter de Gruyter, Berlin, New York.