Asymptotics for sums of a function of normalized independent sums

Kamil Kosiński

▶ **To cite this version:**

Kamil Kosiński. Asymptotics for sums of a function of normalized independent sums. Statistics and Probability Letters, 2009, 79 (4), pp.415. 10.1016/j.spl.2008.09.013 . hal-00505538

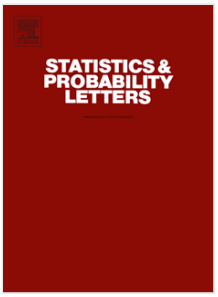
HAL Id: hal-00505538
https://hal.science/hal-00505538

Submitted on 24 Jul 2010

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.

Accepted Manuscript



Asymptotics for sums of a function of normalized independent sums

Kamil Kosiński

PII: S0167-7152(08)00427-6
DOI: 10.1016/j.spl.2008.09.013
Reference: STAPRO 5209

To appear in: *Statistics and Probability Letters*

Received date: 27 August 2008
Revised date: 14 September 2008
Accepted date: 15 September 2008

Please cite this article as: Kosiński, K., Asymptotics for sums of a function of normalized independent sums. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.09.013

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Asymptotics for sums of a function of normalized independent sums

Kamil Kosiński*

Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski, Warszawa, Poland

Abstract

We derive a central limit theorem for sums of a function of independent sums of independent and identically distributed random variables. In particular we show that previously known result from Rempała and Wesołowski (Statist. Probab. Lett. 74, 129), which can be obtained by applying the logarithm as the function, holds true under weaker assumptions.

Key words: Products of sums of iid rv's, Limit distribution, Central limit theorem, Lognormal distribution
2000 MSC: G0F05

1. Introduction

Throughout this paper let $(X_{k,n})_{k=1,\ldots,n}$; $n = 1, 2\ldots$ be a triangular array of independent and identically distributed (iid) random variables (rv's) with the same distribution as X. Let us define the (mutually independent) partial sums $S_{n,n} = \sum_{k=1}^{n} X_{k,n}$. We will work with real functions f defined at least on an interval I such that $\mathbb{P}(X \in I) = 1$. We will also write $\log^+ x$ for $\log(x \vee 1)$.

The asymptotic behavior of the product of partial sums of a sequence of iid rv's has been studied in several papers (see e.g. Lu and Qi (2004) for a brief review). In particular it was shown in Rempała and Wesołowski (2002) that if (X_n) is a sequence of iid positive square integrable rv's with $\mathbb{E}X_1 = \mu$, $\mathrm{Var}(X_1) = \sigma^2 > 0$, then setting $S_n = \sum_{k=1}^{n} X_k$ and $\gamma = \sigma/\mu$ we have as

*Corresponding author.
Email addresses: kamil.kosinski@gmail.com (Kamil Kosiński)

ACCEPTED MANUSCRIPT

$n \to \infty$

$$\left(\prod_{k=1}^{n} \frac{S_k}{k\mu}\right)^{\frac{1}{\gamma\sqrt{n}}} \xrightarrow{d} e^{\sqrt{2}\mathcal{N}},$$

where $\xrightarrow{d}$ stands for convergence in distribution and $\mathcal{N}$ is a standard normal random variable. This result was extended in Qi (2003) and Lu and Qi (2004) to a general limit theorem covering the case when the underlying distribution is integrable and belongs to the domain of attraction of a stable law with index from the interval $[1, 2]$.

This study brought an interest to the array case, where we no longer consider a sequence (X_n) but a triangular array $(X_{k,n})$. In Rempała and Wesołowski (2005) the analogous result was obtained, namely

$$\left(n^{\frac{\gamma^2}{2}} \prod_{k=1}^{n} \frac{S_{k,k}}{k\mu}\right)^{\frac{1}{\gamma\sqrt{\log n}}} \xrightarrow{d} e^{\mathcal{N}}, \tag{1}$$

under the assumption $\mathbb{E}|X|^p < \infty$ for some $p > 2$.

The purpose of this paper is to show that the above result holds true under the assumption $\mathbb{E}X^2(\log^+ |X|)^{1/2} < \infty$. However, it is no longer true in general when only $\mathbb{E}X^2 < \infty$ is assumed. We will show that under this assumption, different normalisation is needed. Furthermore, we will set our discussion in a more general setting. It is straightforward that (1) is a simple corollary from

$$\frac{\sum_{k=1}^{n} f(S_{k,k}/k) - b_n}{a_n} \xrightarrow{d} \mathcal{N},$$

if one sets $f(x) = \log x$ and chooses the sequences a_n, b_n properly.

2. Main result

Theorem 1. *Suppose that $\mathbb{E}|X|^2 < \infty$ and denote $\mu = \mathbb{E}X$, $\sigma^2 = \mathrm{Var}(X)$. Let f be a real function with bounded third derivative on some neighbourhood of μ. Then as $n \to \infty$*

$$\frac{\sum_{k=1}^{n} f(S_{k,k}/k) - b_n}{a_n} \xrightarrow{d} \sigma f'(\mu)\mathcal{N},$$

where

$$a_n = \sqrt{\log n}\,, \quad b_n = nf(\mu) + \frac{f''(\mu)}{2} \sum_{k=1}^{n} \frac{1}{k} \mathbb{E}|X - \mu|^2 1_{\{|X-\mu| \le \sigma k\}}.$$

ACCEPTED MANUSCRIPT

Remark 1. If we strengthen the assumption of the square integrability of random variable X to $\mathbb{E}X^2(\log^+|X|)^{1/2}<\infty$, then we can take the sequence $\tilde{b}_n = nf(\mu)+\frac{f''(\mu)\sigma^2}{2}\log n$ instead of b_n. To see this we should show $\tilde{b}_n - b_n = o(\sqrt{\log n})$, which since $\log n - \sum_{k=1}^n \frac{1}{k} = O(1)$ is equivalent to

$$Q_n := \sum_{k=1}^n \frac{1}{k}\left(\sigma^2 - \mathbb{E}|X-\mu|^2 1_{\{|X-\mu|\le \sigma k\}}\right) = o(\sqrt{\log n}).$$

Observe that Q_n is positive and

$$\begin{aligned} Q_n &= \sum_{k=1}^n \frac{1}{k}\mathbb{E}|X-\mu|^2 1_{\{|X-\mu|>\sigma k\}} = \mathbb{E}|X-\mu|^2 \sum_{\sigma k<|X-\mu|,\, k\le n} \frac{1}{k} \\ &\sim \mathbb{E}|X-\mu|^2 \log^+\left(n \wedge (|X-\mu|/\sigma)\right) =: \tilde{Q}_n. \end{aligned}$$

Therefore, if $\mathbb{E}X^2(\log^+|X|)^{1/2}<\infty$, we can use the Dominated Convergence Theorem and infer that $\tilde{b}_n - b_n = o(\sqrt{\log n})$.

Remark 2. On the other hand, if for some $\varepsilon>0$ we define a random variable X by setting $\mathbb{P}(X=\pm k_n) = C/(2k_n^2 n^2)$ and $\mathbb{P}(X=0) = 1-\sum_n \mathbb{P}(|X|=k_n)$, where $k_n = e^{n^{2+\varepsilon}}$ and $C = 6/\pi^2$. Then we simply have $\mu=0$, $\sigma^2=1$ and $\mathbb{E}X^2(\log^+|X|)^{1/2}=\infty$. Although, one can check that $\limsup_n Q_n/\sqrt{\log n} = \limsup_n \tilde{Q}_n/\sqrt{\log n} = \infty$, which means that we cannot use $\tilde{b}_n$ in general.

Now let us take any positive (i.e. $I\subset(0,\infty)$), nondegenerate random variable X with $\mathbb{E}X^2(\log^+|X|)^{1/2}<\infty$ and $f(x)=\mu\log(x/\mu)$. Theorem 1 yields (1), that is the result from Rempała and Wesołowski (2005) however under weaker assumptions. Our argument shows that their result holds true even under the assumption of square integrability, although the normalizing sequences should be different. Namely, instead of the term $n^{\frac{\gamma^2}{2}}$ in (1) we should have $\exp(\frac{1}{2\mu^2}\sum_{k=1}^n \frac{1}{k}\mathbb{E}|X-\mu|^2 1_{\{|X-\mu|\le\sigma k\}})$.

The proof of Theorem 1 relies on the Taylor's expansion of a function f in the neighbourhood of μ. Linear term in this expansion obeys a version of the classical Central Limit Theorem. Other terms are negligible mainly due to the Strong Law of Large Numbers (SLLN). This assertion will be made valid by a series of lemmas.

Lemma 2. *Under the assumptions of Theorem 1 with $\sigma>0$*

$$\frac{1}{\sigma\sqrt{\log n}}\sum_{k=1}^n\left(\frac{S_{k,k}}{k}-\mu\right)\xrightarrow{d}\mathcal{N} \quad \text{as } n\to\infty.$$

ACCEPTED MANUSCRIPT

PROOF. We may assume $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$ by a simple normalization argument. Since

$$\operatorname{Var}\left(\sum_{k=1}^{n} \frac{S_{k,k}}{k}\right) = \sum_{k=1}^{n} \frac{\operatorname{Var}(S_{k,k})}{k^2} = \sum_{k=1}^{n} \frac{1}{k} \sim \log n,$$

then to complete the proof it is sufficient to show the Lindeberg condition for the array $(\frac{S_{k,k}}{k\sqrt{\log n}})_{k\le n}$, that is

$$\bigwedge_{r>0} \quad \frac{1}{\log n} \sum_{k=1}^{n} \mathbb{E}\left(\frac{S_{k,k}}{k}\right)^2 1_{\{|S_{k,k}/k| > r\sqrt{\log n}\}} = o(1). \tag{2}$$

Since $\{(S_{k,k}/\sqrt{k})^2\}$ is uniformly integrable,

$$\sup_{k\in\mathbb{N}} \mathbb{E}\left(\frac{S_{k,k}}{\sqrt{k}}\right)^2 1_{\{|S_{k,k}/\sqrt{k}| > r\sqrt{\log n}\}} \to 0 \quad \text{as } n \to \infty,$$

proving (2). □

To establish the SLLN we will refer to Hsu-Robbins (1947) Law of Large Numbers (cf. Li et al. (1995) for partial bibliographies and brief discussions).

Lemma 3 (Hsu-Robbins LLN)**.** *For a sequence (X_n) of iid rv's with $\mathbb{E}X_1 = 0$ and $\mathbb{E}X_1^2 < \infty$ the series*

$$\sum_{n=1}^{\infty} \mathbb{P}(|S_n/n| > t) \tag{3}$$

converges for every $t > 0$.

The condition (3) implies $S_n/n \to 0$ a.s. under the Borel-Cantelli lemma. Moreover if X_1 in Lemma 3 has the same distribution as X in Theorem 1, then $\mathbb{P}(|S_n/n| > t) = \mathbb{P}(|S_{n,n}/n| > t)$, i.e $S_{n,n}/n \to 0$ a.s. as well.

Before we proceed, we need some technical results derived from the elementary fact about the moments of sums of iid rv's (e.g. Hall and Heyde, 1980, p. 23).

Lemma 4 (Rosenthal's inequality)**.** *If (X_n) is a sequence of iid rv's with the zero mean, then for any $p \ge 2$*

$$\mathbb{E}|S_n|^p \le C_p \left(n\mathbb{E}|X_1|^p + n^{p/2}(\mathbb{E}X_1^2)^{p/2}\right),$$

where C_p is a constant depending only on p.

ACCEPTED MANUSCRIPT

We will use this version of the Rosenthal's inequality to prove the following lemma, which on the other hand will simplify a number of steps in the next lemma. The later will play a crucial role in the proof of the main theorem.

Lemma 5. *Under the assumptions of Theorem 1, for every $p > 2$*

$$\mathbb{E}\sum_{k=1}^{\infty}\left(\frac{|\tilde{T}_k|}{k}\right)^p < \infty,$$

where $\tilde{T}_k = \sum_{i=1}^{k}\left(X_{i,k}1_{\{|X_{i,k}|\le k\}} - \mathbb{E}X_{i,k}1_{\{|X_{i,k}|\le k\}}\right)$.

Proof. Let $Y_k \stackrel{d}{=} X_{i,k}1_{\{|X_{i,k}|\le k\}} - \mathbb{E}X_{i,k}1_{\{|X_{i,k}|\le k\}}$, then by Lemma 4

$$\begin{aligned}
\mathbb{E}\sum_{k=1}^{\infty}\left(\frac{|\tilde{T}_k|}{k}\right)^p &\le C_p\sum_{k=1}^{\infty}\frac{1}{k^p}\left(k\mathbb{E}|Y_k|^p + k^{p/2}(\mathbb{E}Y_k^2)^{p/2}\right)\\
&\le C_p 2^p\left(\sum_{k=1}^{\infty}\frac{1}{k^{p-1}}\mathbb{E}|X|^p 1_{\{|X|\le k\}} + \sum_{k=1}^{\infty}\frac{1}{k^{p/2}}(\mathbb{E}X^2)^{p/2}\right)\\
&= C_p 2^p\left(\mathbb{E}|X|^p\sum_{k\ge|X|}^{\infty}k^{1-p} + (\mathbb{E}X^2)^{p/2}\sum_{k=1}^{\infty}\frac{1}{k^{p/2}}\right)\\
&\le C_p 2^p\left(C\mathbb{E}|X|^2 + (\mathbb{E}X^2)^{p/2}\sum_{k=1}^{\infty}\frac{1}{k^{p/2}}\right) < \infty,
\end{aligned}$$

for some positive constant C. □

Lemma 6. *Under the assumptions of Theorem 1 we have*

$$\sum_{k=1}^{n}\left[\left(\frac{S_{k,k}-k\mu}{k}\right)^2 - \frac{\mathbb{E}|X-\mu|^2 1_{\{|X-\mu|\le\sigma k\}}}{k}\right] = O_{\mathbb{P}}(1), \tag{4}$$

$$\sum_{k=1}^{n}\left|\frac{S_{k,k}-k\mu}{k}\right|^3 = O_{\mathbb{P}}(1). \tag{5}$$

Proof. To simplify the notation we will write S_k for $S_{k,k}$. First note that to show (4) it is sufficient to prove that

$$\sum_{k=1}^{n}\left(\frac{S_k^2}{k^2} - \frac{\mathbb{E}|X|^2 1_{\{|X|\le k\}}}{k}\right) = O_{\mathbb{P}}(1),$$

ACCEPTED MANUSCRIPT

for a normalized random variable X. Take any $\varepsilon > 0$ and let $T_k := \sum_{i=1}^{k} X_{i,k} 1_{\{|X_{i,k}| \le k\}}$. Then $\sum_{k=1}^{\infty} \mathbb{P}(S_k \neq T_k) \le \sum_{k=1}^{\infty} k\mathbb{P}(|X| > k) < \infty$, because $\mathbb{E}X^2 < \infty$. Hence we can take R big enough that $\sum_{k=R}^{\infty} \mathbb{P}(S_k \neq T_k) < \varepsilon/3$ and M big enough that

$$\mathbb{P}\left(\left|\sum_{k=1}^{R-1}\left(\frac{S_k^2}{k^2} - \frac{\mathbb{E}|X|^2 1_{\{|X|\le k\}}}{k}\right)\right| > M/2\right) < \varepsilon/3.$$

So all we need to show is

$$\mathbb{P}\left(\left|\sum_{k=R}^{n}\left(\frac{T_k^2}{k^2} - \frac{\mathbb{E}|X|^2 1_{\{|X|\le k\}}}{k}\right)\right| > M/2\right) < \varepsilon/3,$$

which is implied by

$$\sum_{k=1}^{n} \frac{T_k^2 - b_k}{k^2} = O_{\mathbb{P}}(1) \tag{6}$$

with $b_k := k\mathbb{E}|X|^2 1_{\{|X|\le k\}}$. Observe that

$$T_k = \sum_{i=1}^{k} \left(X_{i,k} 1_{\{|X_{i,k}|\le k\}} - \mathbb{E}X_{i,k} 1_{\{|X_{i,k}|\le k\}}\right) + c_k := \tilde{T}_k + c_k,$$

where $c_k = k\mathbb{E}X 1_{\{|X|\le k\}}$ and

$$\sum_{k=1}^{n} \frac{T_k^2 - b_k}{k^2} = \sum_{k=1}^{n} \frac{\tilde{T}_k^2 - b_k}{k^2} + \sum_{k=1}^{n} \frac{c_k^2}{k^2} + 2\sum_{k=1}^{n} \frac{c_k \tilde{T}_k}{k^2} := I_1 + I_2 + I_3.$$

Recall that $\mathbb{E}X = 0$ so that

$$|c_k| = |k\mathbb{E}X 1_{\{|X|\le k\}}| = |k\mathbb{E}X 1_{\{|X|> k\}}| \le \mathbb{E}|X|^2 = 1,$$

thus $I_2 = O(1)$. We also have $I_3 = O_{\mathbb{P}}(1)$ because I_3 is bounded in L^2

$$\begin{aligned}\mathbb{E}\left(\sum_{k=1}^{n} \frac{c_k \tilde{T}_k}{k^2}\right)^2 &= \sum_{k=1}^{n} \mathbb{E}\left(\frac{c_k \tilde{T}_k}{k^2}\right)^2 \le \sum_{k=1}^{n} \frac{1}{k^4} k \operatorname{Var}(X 1_{\{|X|\le k\}}) \\ &\le \mathbb{E}X^2 \sum_{k=1}^{n} \frac{1}{k^3} = O(1).\end{aligned}$$

ACCEPTED MANUSCRIPT

I_1 can be rewritten as

$$I_1 = \sum_{k=1}^{n} \frac{\tilde{T}_k^2 - \mathbb{E}\tilde{T}_k^2}{k^2} - \sum_{k=1}^{n} \frac{c_k^2}{k^3} := I_{11} - I_{12}.$$

But $I_{12} = O(1)$ since $|c_k| \leq 1$. So in order to show (6) it is enough to show that

$$K_n := \sum_{k=1}^{n} \frac{\tilde{T}_k^2 - \mathbb{E}\tilde{T}_k^2}{k^2} = O_{\mathbb{P}}(1),$$

where $\tilde{T}_k$ is a sum of independent, mean zero rv's with the same distribution as $X1_{\{|X|\leq k\}} - \mathbb{E}X1_{\{|X|\leq k\}}$. This however follows from Lemma 5 with $p = 4$. Indeed

$$\mathbb{E}K_n^2 = \operatorname{Var}(K_n) = \sum_{k=1}^{n} \frac{1}{k^4} \operatorname{Var}(\tilde{T}_k^2) \leq \mathbb{E}\sum_{k=1}^{\infty} \left(\frac{\tilde{T}_k}{k}\right)^4 < \infty,$$

so the proof of (4) is complete.

To prove (5) it suffices to show that $\sum_{k=1}^{n} \left|\frac{S_k}{k}\right|^3 = O_{\mathbb{P}}(1)$ for normalized X, which by the same arguments as above is implied by

$$\sum_{k=1}^{n} \left|\frac{T_k}{k}\right|^3 = O_{\mathbb{P}}(1).$$

Using the same notation we have $|T_k|^3 = |\tilde{T}_k + c_k|^3 \leq 4|\tilde{T}_k|^3 + 4$. Thus

$$\sum_{k=1}^{n} \left|\frac{T_k}{k}\right|^3 \leq 4\left(\sum_{k=1}^{n} \left|\frac{\tilde{T}_k}{k}\right|^3 + \sum_{k=1}^{n} \frac{1}{k^3}\right) := I_4 + I_5.$$

We obviously have $I_5 = O(1)$ and by Lemma 5 with $p = 3$ we get boundedness of I_4 in L^1, which completes the proof. □

Now we are in the position to prove the main theorem.

Proof of Theorem 1. Take a_n and b_n as in the claim and denote $c_k = \mathbb{E}|X - \mu|^2 1_{\{|X-\mu|\leq \sigma k\}}$ so $b_n = \sum_{k=1}^{n}(f(\mu) + f''(\mu)\frac{c_k}{2k})$. By Taylor's expansion,

$$f\left(\frac{S_{k,k}}{k}\right) = f(\mu) + f'(\mu)\left(\frac{S_{k,k}}{k} - \mu\right) + \frac{f''(\mu)}{2}\left(\frac{S_{k,k}}{k} - \mu\right)^2 + O\left(\left|\frac{S_{k,k}}{k} - \mu\right|^3\right) \text{ a.s.},$$

ACCEPTED MANUSCRIPT

as a consequence of the SLLN and the assumption of boundedness of $f^{(3)}$ around μ. Using Lemma 6 we have

$$\begin{aligned}\frac{\sum_{k=1}^n f(S_{k,k}/k) - b_n}{a_n} &= \frac{f'(\mu)}{a_n}\sum_{k=1}^n\left(\frac{S_{k,k}}{k}-\mu\right) + \frac{f''(\mu)}{2a_n}\sum_{k=1}^n\left[\left(\frac{S_{k,k}}{k}-\mu\right)^2 - \frac{c_k}{k}\right] \\ &\quad + O\left(\frac{1}{a_n}\sum_{k=1}^n\left|\frac{S_{k,k}}{k}-\mu\right|^3\right) \text{ a.s.} \\ &= \frac{f'(\mu)}{a_n}\sum_{k=1}^n\left(\frac{S_{k,k}}{k}-\mu\right) + o_{\mathbb{P}}(1).\end{aligned}$$

By Lemma 2

$$\frac{f'(\mu)}{a_n}\sum_{k=1}^n\left(\frac{S_{k,k}}{k}-\mu\right) \xrightarrow{d} \sigma f'(\mu)\mathcal{N},$$

completing the proof. □

3. Acknowledgements

The author is indebted to Dr. R. Latała for his insightful comments and supervision which led to an improved presentation of the paper. Thanks are also due to the anonymous referee for a very careful review of the original manuscript leading to substantial simplifications.

References

Hall, P., Heyde, C.C., 1980. Martingale limit theory and its application. Academic Press

Hsu, P.L., Robbins, H., 1947. Complete convergence and the law of large numbers. Proc. Nat. Acad. Sci. 33, 25–31.

Li, D., Rao, B., Jiang, T., Wang, X., 1995. Complete convergence and almost sure convergence of weighted sums of random variables. J. Theoret. Probab. 8, 49–76.

Lu, X., Qi, Y., 2004. A note on asymptotic distribution of products of sums. Statist. Probab. Lett. 68, 407–413.

ACCEPTED MANUSCRIPT

Rempała, G., Wesołowski, J., 2002. Asymptotics for products of sums and U-statictics. Electron. Comm. Probab. 7, 45–54.

Rempała, G., Wesołowski, J., 2005. Asymptotics for products of independent sums with an application to Wishart determinants. Statist. Probab. Lett. 74, 129–138.

Qi, Y., 2003. Limit distributions for products of sums. Statist. Probab. Lett. 62, 93–100.

ACCEPTED MANUSCRIPT