Z. Wahrscheinlichkeitstheorie verw. Gebiete
35, 213–220 (1976)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

The Approximation of Partial Sums of Independent RV's

Péter Major

Mathematical Institute of the Hungarian Academy of Sciences
Reáltanoda u. 13–15, H-1053 Budapest, Hungary

Given the partial sums of i.i.d.r.v.s with a distribution function $F(x)$, $\int x\,dF(x)=0$, $\int x^2\,dF(x)=1$ we want to approximate these r.v.s with the partial sums of independent standard normal variables. This problem was discussed in papers [1] and [2]. The case when the third moment of the distribution does not exist was omitted. The aim of this paper is to fill this gap.

Introduction

In papers [1] and [2] a sequence of partial sums of i.i.d.r.v.s was approximated with normal summands. The case when the third moment of the summands does not exist was omitted because some statements of the paper were based on the Berry-Essen theorem which holds only if the third moment exists. Here we will prove somewhat weaker results in the case when the third moment does not exist and show that these weaker results are sufficient to make an optimal construction possible. We could choose the same construction as in [1] and [2] but we choose another one which is just as good and simpler. Similar constructions are also found in [3, 4] and [5].

We prove the following:

Theorem 1. *Let $F(x)$ be a distribution function such that $\int x\,dF(x)=0$, $\int x^2\,dF(x)=1$ and $\int x^2 g(|x|)\,dF(x)<\infty$ where $g(x)$ satisfies the following conditions*

i) *$g(x)/x$ is monotonically decreasing,*

ii) *$\bar{g}(x)=g(x)/x^{\varepsilon}$ is monotonically increasing with an appropriate $\varepsilon>0$.*

Then one can construct a sequence of i.i.d.r.v.s $X_1, X_2, \ldots$ with distribution $F(x)$ and another sequence $Y_1, Y_2, \ldots$ with standard normal distribution in such a way that the sequences

$$S_n=\sum_{k=1}^{n} X_k, \qquad T_n=\sum_{k=1}^{n} Y_k \qquad (n=1,2,\ldots),$$

satisfy the relation

$$P\left(\limsup \frac{|S_n - T_n|}{h(n)} \leqq L\right) = 1 \tag{1}$$

with an appropriate constant L. Here $h(x)$ denotes the inverse of the function $x^2 g(x)$.

Instead of i) we could write

(i′) $\frac{g(x)}{x^{1+\varepsilon}}$ is monotonically decreasing. (ε is sufficiently small.)

Observe, that if $\int x^2 g(|x|)\, dF(x) < \infty$ then there is a function $u(x) \geqq 0$, $u(x) \to \infty$ such that $\int x^2 g(|x|)\, u(|x|)\, dF(x) < \infty$. Applying Theorem 1 with $g(x)\, u(x)$ instead of $g(x)$ we obtain the following

Corollary. *Under the conditions of Theorem 1 one can construct the sequences $X_1, X_2, \ldots$ $Y_1, Y_2, \ldots$ in such a way that*

$$\frac{S_n - T_n}{h(n)} \to 0, \quad \text{with probability one.}$$

Specifically choosing $g(x) = |x|^{r-2}$, $2 < r \leqq 3$ we obtain that if $\int |x|^r\, dF(x) < \infty$ $2 < r \leqq 3$ then

$$\frac{S_n - T_n}{n^{\frac{1}{r}}} \to 0. \tag{2′}$$

Using the results of [2] one sees that (2′) holds true for every $r > 2$.

Proofs. To prove our Theorem we need the following lemmas.

Lemma 1. *Let $X_1, X_2, \ldots$ be i.i.d.r.v.s, $EX_1^2\, g(|X_1|) < \infty$ where $g(x)$ satisfies* (i) *and* (ii). *Define*

$$X_i' = \begin{cases} X_i & \text{if } |X_i| < c\sqrt{n \log n} \\ 0 & \textit{otherwise}, \end{cases}$$

$$\tilde{X}_i = \frac{X_i' - EX_i'}{DX_i'}, \qquad F(x) = P(\tilde{X}_1 < x),$$

and

$$F_n(x) = P\left(\frac{\tilde{X}_1 + \tilde{X}_2 + \cdots + \tilde{X}_n}{\sqrt{n}} < x\right).$$

Then we have

$$1 - F_n(x) = [1 - \Phi(x)]\left[1 + O\left(\frac{1}{\bar{g}(\sqrt{n})}\right)\right],$$

$$F_n(-x) = \Phi(-x)\left[1 + O\left(\frac{1}{\bar{g}(\sqrt{n})}\right)\right]$$

if $0 \leqq x \leqq \frac{\varepsilon}{8c}\sqrt{\log n}$. ($\Phi(x)$ is the standard normal distribution.)

Proof. The proof applies the standard large-deviation technique. We prove only the first relation, the proof of the second one being the same. Define $dV(x)=\frac{e^{tx}\,dF(x)}{R(t)}$, where $R(t)=\int e^{tx}\,dF(x)$ and let $\psi(t)=\log R(t)$. We have

$$1-F_n(x)=1-F^{*(n)}(\sqrt{n}\,x)=\int_{\sqrt{n}\,x}^{\infty} e^{n\psi(t)-ty}\,dV^{*(n)}(y). \tag{2}$$

Provided $t<\frac{\varepsilon}{6c}\frac{\sqrt{\log n}}{\sqrt{n}}$, then some calculations show that $|R(t)|$, $|R'(t)|$, $|R''(t)|$ are bounded by a constant, and

$$|R'''(t)|\leqq\int |x|^3\,e^{tx}\,dF(x)\leqq\int |x|^{3+\frac{\varepsilon}{2}}\,dF(x)$$

$$\leqq\frac{(n\log n)^{\frac{1}{2}+\frac{\varepsilon}{4}}}{g(\sqrt{n\log n})}\int x^2\,g(|x|)\,dF(x)<\frac{\sqrt{n}}{\bar{g}(\sqrt{n})\,n^{\frac{\varepsilon}{5}}}.$$

Thus calculating the derivatives of $\psi(t)$ we obtain that $|\psi(t)|$, $|\psi'(t)|$, $|\psi''(t)|$ are bounded by a constant and

$$\psi'''(t)\leqq K\frac{\sqrt{n}}{\bar{g}(\sqrt{n})\,n^{\frac{\varepsilon}{5}}}.$$

Choose t as the solution of the equation $x=\sqrt{n}\,\psi'(t)$. Then $x<\frac{\varepsilon}{8c}\sqrt{\log n}$ implies that $t\leqq\frac{\varepsilon}{6c}\frac{\sqrt{\log n}}{\sqrt{n}}$. Expressing t as a function of x, the Taylor-series of $t(x)$ gives that

$$t=\frac{x}{\sqrt{n}}+O\left(\frac{x^2}{\bar{g}(\sqrt{n})\,n^{\frac{1}{2}+\frac{\varepsilon}{5}}}\right).$$

We want to approximate $V^{*(n)}(x)$ by a normal distribution with expectation $n\,\psi'(t)$ and variance $n\,\psi''(t)$. Now $V^{*(n)}(x)$ may not have a bounded third moment, but

$$\int x^{2+\frac{\varepsilon}{2}}\bar{g}(|x|)\,dV(x)=\frac{\int x^{2+\frac{\varepsilon}{2}}e^{tx}\bar{g}(|x|)\,dF(x)}{R(t)}\leqq\int x^2 g(x)\,dF(x)\leqq K.$$

Therefore, we can state (see e.g. [6] p. 141) that

$$\sup_x\left|V^{*(n)}(x)-\Phi\left(\frac{x-n\,\psi'(t)}{\sqrt{n\,\psi''(t)}}\right)\right|\leqq\frac{A}{\bar{g}(\sqrt{n})\,n^{\varepsilon/2}}$$

with an appropriate constant A.

Let us approximate (2) by

$$u_t=\int_{\sqrt{n}\,x}^{\infty} e^{n\psi(t)-ty}\,d\Phi\left(\frac{y-n\,\psi'(t)}{\sqrt{n\,\psi''(t)}}\right).$$

Substituting $z=\frac{y-n\,\psi'(t)+t\,n\,\psi''(t)}{\sqrt{n\,\psi''(t)}}$ one gets

$$u_t=\exp\left(n\left[\psi(t)-t\,\psi'(t)+\frac{t^2}{2}\,\psi''(t)\right]\right)\cdot\left(1-\Phi(t\sqrt{n\,\psi''(t)}\right)$$

where

$$n\left[\psi(t)-t\,\psi'(t)+\frac{t^2}{2}\psi''(t)\right]=O\left(\frac{x^3}{\bar{g}(\sqrt{n})\,n^{\frac{\varepsilon}{5}}}\right)=O\left(\frac{1}{\bar{g}(\sqrt{n})}\right)$$

and

$$\log\frac{1-\Phi(t\sqrt{n\,\psi''(t)})}{1-\Phi(x)}=O\left(\frac{1}{\bar{g}(\sqrt{n})}\right).$$

Therefore,

$$u_t=[1-\Phi(x)]\left(1+O\left(\frac{1}{\bar{g}(\sqrt{n})}\right)\right).$$

Integration by parts shows that $1-F_n(x)-U_t$ is less than

$$\frac{2A}{\bar{g}(\sqrt{n})\,n^{\frac{\varepsilon}{2}}}\exp\left[n(\psi(t)-t\,\psi'(t))\right]$$

$$\leqq\frac{2A}{\bar{g}(\sqrt{n})\,n^{\frac{\varepsilon}{2}}}\cdot\frac{u_t}{\exp\left(\frac{n t^2\psi''(t)}{2}\right)(1-\Phi(t\sqrt{n\,\psi''(t)})}=O\left(\frac{1-\Phi(x)}{\bar{g}(\sqrt{n})}\right).$$

These estimations prove Lemma 1.

Now define the inverse of a distribution function $F(x)$ as

$$F^{-1}(t)=\sup(y;\,F(y)\leqq t).$$

Then given a uniformly distributed r.v. ξ on [0, 1], the r.v. $F^{-1}(\xi)$ has distribution $F(x)$.

We need the following corollary of Lemma 1.

Corollary. *Let* $\frac{S_n}{\sqrt{n}}=\Phi^{-1}(\xi)$, $\frac{T_n}{\sqrt{n}}=F_n^{-1}(\xi)$ *where* ξ *is a uniformly distributed random variable on* [0, 1]. *The estimation*

$$D^2(S_n-T_n)=O\left(\frac{\sqrt{n}}{\bar{g}(\sqrt{n})}\right)^2$$

holds if we take $c^2\geqq\frac{\varepsilon^3}{128(2+\varepsilon)}$ *in Lemma 1.*

Proof of the Corollary. If $0\leqq S_n\leqq c'\sqrt{n\log n}$, $c'=\frac{\varepsilon}{8c}$, then

$$1-\Phi\left(\frac{S_n}{\sqrt{n}}-\frac{K}{\bar{g}(\sqrt{n})}\right)\geqq 1-F_n\left(\frac{S_n}{\sqrt{n}}\right)$$

$$=\left[1-\Phi\left(\frac{S_n}{\sqrt{n}}\right)\right]\left(1+\Phi\left(\frac{1}{\bar{g}(\sqrt{n})}\right)\right]\geqq 1-\Phi\left(\frac{S_n}{\sqrt{n}}+\frac{K}{\bar{g}(\sqrt{n})}\right)$$

with an appropriate $K>0$; thus $|S_n-T_n|\leqq\frac{K\sqrt{n}}{\bar{g}(\sqrt{n})}$.

By Hölder's inequality,

$$\int_{x>c'\sqrt{\log n}} x^2\, dF_n(x) \leqq P\left(\frac{S_n}{\sqrt{n}} > c'\sqrt{\log n}\right)^{\frac{\varepsilon}{2+\varepsilon}} \left[\int |z|^{2+\varepsilon}\, dF_n(x)\right]^{\frac{2}{2+\varepsilon}}$$

$$\leqq n^{-\frac{c'^2}{2}\cdot\frac{\varepsilon}{2+\varepsilon}} \leqq \frac{1}{n}.$$

The tail of T_n can be similarly estimated. Similar estimations are valid if $S_n \leqq 0$, and these estimations imply the Corollary.

We need the next lemma in order to work with truncated random variables instead of the original ones.

Lemma 2. *Let $X_1, X_2, \ldots$ be i.i.d.r.v.s with d.f. $F(x)$, $EX_1=0$, $EX_1^2=1$, $EX_1^2\, g(|X_1|) < \infty$ where $g(x)$ is as in Theorem 1. Define $\alpha_n = \dfrac{\left[g^{-1}\left(\frac{n}{\log n}\right)\right]^2}{\log n}$ where $g^{-1}(x)$ is the inverse of $g(x)$. Let v_n, $n=1,2,\ldots$ be a monotone increasing sequence of positive integers such that $v_n < B\,\alpha_n$ with appropriate $B>0$.*

Define

$$\bar{X}_k = \begin{cases} X_k & \text{if } |X_k| < g^{-1}\left(\dfrac{n}{\log n}\right);\ \sum_{j=1}^{n-1} v_j \leqq k < \sum_{j=1}^{n} v_j \\ 0 & \text{otherwise} \end{cases}$$

and

$$\tilde{X}_k = \frac{\bar{X}_k - E\bar{X}_k}{D\bar{X}_k}.$$

Then the sequences $S_n = \sum_{j=1}^{n} X_j$, $\bar{S}_n = \sum_{j=1}^{n} \bar{X}_j$ and $\tilde{S}_n = \sum_{j=1}^{n} \tilde{X}_j$ $n=1,2,\ldots$ satisfy the relations

a) $|S_n - \bar{S}_n| \leqq K(\omega)$ *with probability one,*

b) $\dfrac{\tilde{S}_n - \bar{S}_n}{h(n)} \to 0$ *with probability one*

($h(x)$ is again as in Theorem 1, that is the inverse of $x^2 g(x)$).

Let us remark that $h(n\,\alpha_n) = g^{-1}\left(\frac{n}{\log n}\right)$, since $n\,\alpha_n = \left[g^{-1}\left(\frac{n}{\log n}\right)\right]^2 \cdot \frac{n}{\log n}$.

Proof of Lemma 2. Denote $g^{-1}\left(\frac{n}{\log n}\right)$ by u_n.

a) We have

$$\sum_{k=1}^{\infty} P(X_k \neq \bar{X}_k) = \sum_{n=1}^{\infty} v_n\, P(|X_1| \geqq u_n)$$

$$\leqq B \sum_n n\,\alpha_n\, P(u_n < |X_1| < u_{n+1})$$

$$= B \sum \int_{u_n}^{u_{n+1}} u_n^2\, g(u_n)\, dF(x) \leqq B \int x^2 g(|x|)\, dF(x) < \infty.$$

Therefore $X_k(\omega) \neq \bar{X}_k(\omega)$ only finitely many times.

b) It is enough to prove the estimations

$$\text{b 1)}\quad \sum_k \frac{E(\tilde{X}_k - \bar{X}_k)}{h(k)} < \infty \quad \text{and} \quad \text{b 2)}\quad \sum \frac{D^2(\tilde{X}_k - \bar{X}_k)}{h^2(k)} < \infty.$$

$$\text{b 1)}\quad \sum_k \frac{E(\tilde{X}_k - \bar{X}_k)}{h(k)} \leqq c \sum \frac{\alpha_n}{h(n\,\alpha_n)} \int_{|x|>u_n} |x|\, dF(x)$$

$$\leqq \sum \frac{c n \alpha_n}{u_n} \int_{u_{n+1}>|x|>u_n} |x|\, dF(x) \leqq c' \int_{-\infty}^{\infty} x^2 g(|x|)\, dF(x) < \infty.$$

$$\text{b 2)}\quad D^2(\tilde{X}_m - \bar{X}_m) = (1 - D\bar{X}_m)^2 \leqq 1 - D^2\,\bar{X}_m = 1 - E\bar{X}_m^2 + (E\bar{X}_m)^2.$$

Furthermore

$$\sum \frac{(E\bar{X}_k)^2}{h^2(k)} \leqq c \sum \frac{E\bar{X}_k}{h(k)} < \infty,$$

and

$$\sum \frac{1 - E\bar{X}_k^2}{h^2(k)} \leqq \sum \frac{c n \alpha_n}{h^2(n\,\alpha_n)} \int_{u_n<|x|<u_{n+1}} |x|^2\, dF(x)$$

$$\leqq c' \int x^2\, g(|x|)\, dF(x) < \infty.$$

Proof of Theorem 1. Define $S_n, \bar{S}_n$ and $\tilde{S}_n$ as in Lemma 2. (We define v_n later.) Because of Lemma 2 it is enough to prove formula (1) writing $\tilde{S}_n$ instead of S_n.

Denote $\gamma_n = \sum_{j=1}^{n} \alpha_j$. Using the notation of our previous lemmas, our construction will be the following.

Let $\xi_1, \xi_2, \ldots$ be i.i.d.r.v.s and let ξ_1 be uniformly distributed on $[0, 1]$. Let $t > 0$ be a sufficiently small fixed number, denote $[t(\gamma_n - \gamma_{n-1})]$ by v_n and put $w_n = \sum_{k=1}^{n} v_k$. We construct the random variables

$$\left.\begin{aligned} \frac{\tilde{S}_{w_{n+1}} - \tilde{S}_{w_n}}{\sqrt{v_n}} &= F_{u_n}^{-1}(\xi_n) \\ \frac{T_{w_{n+1}} - T_{w_n}}{\sqrt{v_n}} &= \Phi^{-1}(\xi_n). \end{aligned}\right. \qquad n = 1, 2, \ldots.$$

($F_n(\sqrt{n}x)$ is the distribution of $\tilde{S}_n$.)

Complete the sequences $\tilde{S}_{w_n}, T_{w_n}$ into two sequences $\tilde{S}_n, T_n$ so that the joint distribution of the sequences $\tilde{S}_n$ and T_n $n = 1, 2, \ldots$ be the prescribed one, while $\tilde{S}_{w_n}, T_{w_n}$ should agree with the previously constructed $\tilde{S}_{w_n}$ and T_{w_n}. We prove that these $\tilde{S}_n$ and T_n satisfy (1).

First we state that

$$\frac{\tilde{S}_{w_n} - T_{w_n}}{h(n\,\alpha_n)} \to 0.$$

Since $E(\tilde{S}_{w_n}-T_{w_n})=0$ this relation follows from the estimation

$$\sum_n \frac{D^2((\tilde{S}_{w_n}-\tilde{S}_{w_{n-1}})-(T_{w_n}-T_{w_{n-1}}))}{h^2(n\,\alpha_n)}$$

$$\leqq K\cdot\sum \frac{\alpha_n}{h^2(n\,\alpha_n)\,\bar{g}(\sqrt{\alpha_n})^2}$$

$$=K\sum \frac{\alpha_n}{\bar{g}(\sqrt{\alpha_n})^2\, g^{-1}\left(\frac{n}{\log n}\right)^2}=\sum \frac{K}{\bar{g}(\sqrt{\alpha_n})^2 \log n}$$

$$\leqq K\sum \frac{\alpha^{2\varepsilon}(\log n)^2}{n^2}<\infty$$

if ε is chosen sufficiently small.

Here we used the estimation

$$\frac{g(\sqrt{\alpha_n})}{\sqrt{\alpha_n}}\geqq\frac{g(\sqrt{\alpha_n \log n})}{\sqrt{\alpha_n \log n}},$$

which is true by (i), and thus

$$g(\sqrt{\alpha_n})\geqq\frac{g(\sqrt{\alpha_n \log n})}{\sqrt{\log n}}=\frac{n}{(\log n)^{\frac{3}{2}}}.$$

It remains to prove that

$$\limsup_n \sup_{w_n<k<w_{n+1}} \frac{|\tilde{S}_k-\tilde{S}_{w_n}|}{h(n\,\alpha_n)}<L, \tag{3}$$

and the same relation holds writing T instead of S.

Using Lemma 1 with an approximately chosen L

$$P\left(\sup_{w_n<k\leqq w_{n+1}} \frac{S_k-S_{w_n}}{h(n\,\alpha_n)}>L\right)\leqq 2P\left(\frac{\tilde{S}_{w_{n+1}}-\tilde{S}_{w_n}}{h(n\,\alpha_n)}>\frac{L}{2}\right)$$

$$=2P\left(\frac{\tilde{S}_{w_{n+1}}-\tilde{S}_{w_n}}{\sqrt{v_{n+1}}}>\frac{L}{2}\,\frac{h(n\alpha_n)}{\sqrt{v_{n+1}}}\right)$$

$$\leqq 2P\left(\frac{\tilde{S}_{w_{n+1}}-\tilde{S}_{w_n}}{\sqrt{v_{n+1}}}>L'\sqrt{\log v_n}\right)\leqq\frac{1}{n^2}.$$

Thus the Borel-Cantelli lemma implies (3). This relation can be similarly proved with T_n-s instead of S_n-s.

Finally we state the following

Theorem 2. *Let $F(x)$ and $h(x)$ be as in Theorem 1. Then for any x, $x^2 g(x)\geqq n$, $x<\sqrt{n\log n}$ there exists two finite sequences $X_1, X_2, \ldots X_n$ resp. $Y_1, Y_2, \ldots Y_n$ of i.i.d.r.v.s with d.f. $F(x)$ resp. $\Phi(x)$ such that the sequences $S_k=\sum_{i=1}^{k} X_i$, $T_k=\sum_{i=1}^{k} Y_i$*

$k \leqq 1, \ldots n$ satisfy the relation

$$P(\sup_{k \leqq n} |S_k - T_k| > x) = \frac{o(n)}{x^2\, g(x)}. \tag{4}$$

The proof is similar to that of Theorem 1, so we give only a rough sketch of it. It is enough to prove this formula with $O(\cdot)$ instead of $o(\cdot)$ and then the same remark can be applied as after Theorem 1.

Let us truncate the r.v.-s X_i at $\frac{x}{3}$, and denote the partial sums made from the standardized form of these truncated variables by $\tilde{S}_k$, $k=1,2,\ldots n$. Approximate the variables $\tilde{S}_{jm} - \tilde{S}_{(j-1)m}$ by $T_{jm} - T_{(j-1)m}$ where $m = t\frac{x^2}{\log x}$, $j=1,2,\ldots\frac{n}{m}$ and $t>0$ is an appropriate fixed constant, in the same way as it was done in the corollary of Lemma 1, and complete the sequences $\tilde{S}$ and T as in the construction of Theorem 1. Then similar arguments like in the proof of Theorem 1 show that (4) holds with $\tilde{S}_k$ instead of S_k and $\frac{x}{2}$ instead of x.

On the other hand

$$P\left(\sup |S_k - \tilde{S}_k| > \frac{x}{2}\right) = \frac{O(n)}{x^2\, g(x)},$$

and these estimations prove Theorem 2.

References

1. Komlós, J., Major, P., Tusnády, G.: An approximation of partial sums of independent RV's and the sample DF, I. Z. Wahrscheinlichkeitstheorie verw. Gebiete **32**, 111–131 (1975)
2. Komlós, J., Major, P., Tusnády, G.: An approximation of partial sums of independent RV's and the sample DF, II. Z. Wahrscheinlichkeitstheorie verw. Gebiete **34**, 33–58 (1976)
3. Bártfai, P.: Über die Entfernung der Irrfahrtswege. Studia Sci. Math. Hungar **5**, 41–49 (1970)
4. Csörgö, M., Révész, P.: A new method to prove Strassen type laws of invariance principle I. Z. Wahrscheinlichkeitstheorie verw. Gebiete **31**, 255–259 (1975)
5. Borovkov, A. A.: On the rate of convergence for the invariance principle (in Russian). Teor. Verojatnost. i Primenen. XVIII. **2**, 217–234 (1973)
6. Petrov, V.V.: On sums of independent random variables. (In Russian) Moscow: Nauka 1972

Received September 2, 1975