Boletín de la Sociedad Matemática Mexicana Vol. 27 No. 1, 1982

THE ORDER OF APPROXIMATION IN THE RANDOM CENTRAL LIMIT THEOREM FOR MAXIMUM SUMS IN NON-IDENTICAL CASE

BY A. K. BASU

1. Introduction

Let $\{X_n\}$ be a sequence of independent but not necessarily identically distributed random variables such that $EX_k = \mu_k$, Var $X_k = \sigma_k^2$, and $\beta_k^{2+\delta} = E\,|\,X_k - \mu_k\,|^{2+\delta} < \infty$ for some $0 < \delta \leq 1$. Let us put

$$S_n = \sum_{k=1}^{n} X_k,\ s_n^2 = \sum_{k=1}^{n} \sigma_k^2,\ A_n = \sum_{k=1}^{n} \mu_k,\ B_n^{2+\delta} = \sum_{k=1}^{n} \beta_k^{2+\delta}.$$

Let $\{N_n\}$ be a sequence of positive integer-valued random variables not necessarily independent of the X_n's such that N_n/n converges in probability to a positive random variables N as $n \to \infty$. Throughout this paper we shall assume that N is independent of the X's and $d_1^2 \leq \sigma_k^2 \leq d_2^2$ for all $k \geq K_0$.

Define

$$S_{N_n} = \sum_{k=1}^{N_n} X_k,\ L_n = \sum_{k=1}^{N_n} \mu_k.$$

Let

$$\Delta_n = \sup_x \left| P\left[\frac{S_n - A_n}{s_n} \leq x\right] - \Phi(x) \right|, \tag{1.1}$$

where Φ denotes the distribution function of the standard normal variate. Let $\bar{S}_n = \max_{1\leq k\leq n} S_k$ and set

$$\bar{\Delta}_n = \sup_x \left| P\left[\frac{\bar{S}_n - A_n}{s_n} \leq x\right] - \Phi(x) \right|$$

if $\mu_k > 0$ for at least one $k = 1, 2, \cdots n$, and

$$\bar{\bar{\Delta}}_n = \sup_x \left| P\left[\frac{\bar{S}_n}{s_n} \leq x\right] - G(x) \right|$$

if $\mu_k = 0$ for all $k = 1, 2, \cdots$, where

$$G(x) = \begin{cases} 2\Phi(x) - 1, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

Define

$$\left.\begin{aligned} \Delta_{1n} &= \sup_x \left| P\left[\frac{S_{N_n} - L_n}{s_{[nN]}} \leq x\right] - \Phi(x) \right| \\ \Delta_{2n} &= \sup_x \left| P\left[\frac{S_{N_n} - L_n}{s_{N_n}} \leq x\right] - \Phi(x) \right| \end{aligned}\right\} \tag{1.2}$$

$$\left.\begin{aligned}\bar{\Delta}_{1n} &= \sup_x \left| P\left[\frac{\bar{S}_{N_n} - L_n}{s_{[nN]}} \le x\right] - \Phi(x) \right| \\ \bar{\Delta}_{2n} &= \sup_x \left| P[(\bar{S}_{N_n} - L_n)/s_{N_n} \le x] - \Phi(x) \right|\end{aligned}\right\} \quad (1.3)$$

$$\left.\begin{aligned}\bar{\bar{\Delta}}_{1n} &= \sup_x \left| P\left[\frac{\bar{S}_{N_n}}{s_{[nN]}} \le x\right] - G(x) \right| \\ \bar{\bar{\Delta}}_{2n} &= \sup_n | P[\bar{S}_{N_n}/s_{N_n} \le x] - G(x) |\end{aligned}\right\} \quad (1.4)$$

Let $\{\epsilon_n\}$ be a sequence of real numbers such that $\epsilon_n \to 0$ as $n \to \infty$ and for all n, $n^{-1} \le \epsilon_n < 1$. The following conditions will be used in the sequel, for some $C_1 > 0$ and $C_2 > 0$, and $0 < \delta \le 1$,

(i)

(a)
$$P\left[N < \frac{C_2}{n\epsilon_n}\right] = O(\epsilon_n^{\delta/2}) \quad (1.5)$$

$$P[| N_n/(nN) - 1] > C_1\epsilon_n | = O(\epsilon_n^{\delta/2}) \quad (1.6)$$

(b) $$P[| s_{N_n}^2/s_{[nN]}^2 - 1 | > C_1\epsilon_n] = O(\epsilon_n^{\delta/2}) \quad (1.7)$$

(ii) $$\sum_{n=1}^{\infty} \epsilon_n^{1-\delta/2} P\left[N < \frac{C_2}{n\epsilon_n}\right] < \infty \quad (1.8)$$

(a) $$\sum_{n=1}^{\infty} \epsilon_n^{1-\delta/2} P\left[\left| \frac{N_n}{(nN)} - 1 \right| > C_1\epsilon_n\right] < \infty \quad (1.9)$$

(b) $$\sum_{n=1}^{\infty} \epsilon_n^{1-\delta/2} P[| s_{N_n}^2/s_{[nN]}^2 - 1 | > C_1\epsilon_n] < \infty. \quad (1.10)$$

The following condition will be assumed whenever needed:

$$B_n^{2+\delta}/s_n^{2+\delta} \le n^{-\delta/2},\ 0 < \delta \le 1 \quad (1.11)$$

Note that (1.11) implies the Lindeberg condition.

The purpose of this investigation is to show that given appropriate rates on $\bar{\Delta}_n$ or $\bar{\bar{\Delta}}_n$, we obtain, under conditions (i) or (ii) above, analogous exact rates for $\bar{\Delta}_{in}$ and $\bar{\bar{\Delta}}_{in}$, $i = 1, 2$. The method used in this paper was first introduced by Landers and Rogge (1976, 1977) and later exploited by Ahmad and Basu (1979), and Rychlik (1978). This paper can be regarded as a generalization of Ahmad and Basu (1979) and is an extension of the work by Rychlik (1978) to maximum sums.

2. Lemmas and theorems

The proof of lemma 7 of Landers and Rogge (1976) uses heavily stationarity and this fact has not been noticed by Rychlik (1978), and so his proof of lemma 1 is false. This lemma can be avoided by retaining the maximum and using some martingale inequality for maximum sums.

LEMMA 1. *Let $1 < p < \infty$. If $\{X_n, \mathscr{F}_n, n \geq 1\}$ is a martingale difference sequence, then*

$$E \max_{1 \leq l \leq n} | \textstyle\sum_{i=j+1}^{j+l} X_i |^p \leq C(p)\, E | \sum_{i=j+1}^{j+n} X_i^2 |^{p/2}.$$

In particular for $p = 2$,

$$E \max_{1 \leq l \leq n} | \textstyle\sum_{i=j+1}^{j+l} X_i |^2 \leq C \sum_{i=j+1}^{n} E\, X_i^2.$$

Proof. See Heyde and Hall (1980).

Let $p_x = [x(1 - C_1 \epsilon_n)]$, $q_2 = [x(1 + C_1 \epsilon_n)]$, $x > 0$.

LEMMA 2. *Let $\{X_k\}$ be independent random variables with $EX_k = 0$,* Var $\sum_{k=1}^{n} X_k = s_n^2 < \infty$ *for $n \geq 1$ and assume conditions* (1.5) *and* (1.11) *are satisfied. Then there exists a constant D such that*

$$P[\min_{p_{nN} \leq j \leq q_{nN}} S_j < s_{[nN]} x] - P[\max_{p_{nN} \leq j \leq q_{nN}} S_j < s_{[nN]} x] \leq D \epsilon_n^{\delta/2}$$

for $0 < \delta \leq 1$.

Proof. The difference is bounded above by

$$P[nN < [C_2/\epsilon_n]] + \int_{[C_2/\epsilon_n]} \{P[\min_{p_x \leq j \leq q_x} S_j < s_{[x]} t] - P[\max_{p_x \leq j \leq q_x} S_j < s_{[x]} t]\}\, dP[nN \leq x].$$

For $p < q$ we have

$$P[\min_{p \leq j \leq q} S_j < r] - P[\max_{p \leq j \leq q} S_j < r] = P[S_p < r \leq \max_{p \leq j \leq q} S_j] + P[\min_{p \leq j \leq q} S_j < r \leq S_p]$$

Since we can replace X_i by $-X_i$, it suffices to show for $x \geq [C_2/\epsilon_n]$, $p = p_x$, $q = q_x$ that

$$P[S_p \leq r \leq \max_{p \leq j \leq q} S_j] \leq P[r - H \leq S_p \leq r]$$

where $H = \max_{p \leq j \leq q}(S_j - S_p)$.

$$\begin{aligned}
&P[r - H \leq S_p \leq r] \\
&= \int P[r - h \leq S_p \leq r]\, dP[H \leq r] \\
&\leq C\, B_p^{2+\delta}/s_p^{2+\delta} + \int | \Phi(r/s_p) - \Phi((r - h))/s_p) |\, dP[H \leq r] \\
&\leq C p^{-\delta/2} + EH/s_p && \text{(by condition 1.11)} \\
&\leq (D/\epsilon_n)^{-\delta/2} + (EH^2)^{1/2}/s_p && \text{(since } p \geq D[C_2/\epsilon_n] \geq D_3/\epsilon_n\text{)} \\
&\leq D_1 \epsilon_n^{\delta/2} + (\textstyle\sum_{i=p}^{q} \sigma_i^2 / \sum_{i=1}^{p} \sigma_i^2)^{1/2} && \text{(from lemma 1)} \\
&\leq D_1 \epsilon_n^{\delta/2} + D_2 \epsilon_n^{1/2} \leq D \epsilon_n^{\delta/2}.
\end{aligned}$$

THEOREM 2.1.

(a) *If condition (ia) is satisfied and if* $\overline{\Delta}_n = O(n^{-\delta/2})$, *then* $\overline{\Delta}_{1n} = O(\epsilon_n{}^{\delta/2})$ *for* $0 < \delta \le 1$.

(b) *If condition (ia) is satisfied and if* $\overline{\overline{\Delta}}_n = O(n^{-\delta/2})$, *then* $\overline{\overline{\Delta}}_{1n} = O(\epsilon_n{}^{\delta/2})$, *for* $0 < \delta \le 1$.

THEOREM 2.2

(a) *If condition (iia) is satisfied, if* $\epsilon_n = O(n^{-1})$ *and if* $\sum_{n=1}^{\infty} n^{-1+\delta/2}\overline{\Delta}_{1n} < \infty$, *then* $\sum_{n=1}^{\infty} \epsilon_n{}^{1-\delta/2}\overline{\Delta}_{1n} < \infty$, *for* $0 < \delta \le 1$.

(b) *If condition (iia) is satisfied, if* $\epsilon_n = O(n^{-1})$ *and if* $\sum_{n=1}^{\infty} n^{-1+\delta/2}\overline{\overline{\Delta}}_n < \infty$, *then* $\sum_{n=1}^{\infty} \epsilon_n{}^{1-\delta/2}\overline{\overline{\Delta}}_{1n} < \infty$, *for* $0 < \delta \le 1$.

Remark. Theorem 2.1 is true for $\overline{\Delta}_{2n}$ and $\overline{\overline{\Delta}}_{2n}$ provided (ib) also holds. Similarly, Theorem 2.2 is valid for $\overline{\Delta}_{2n}$ and $\overline{\overline{\Delta}}_{2n}$ provided (iib) also holds. This follows from the fact that $[\,|\,s_{N_n}/s_{[nN]} - 1\,| > (C_1\epsilon_n)^{1/2}] \subseteq [\,|\,s_{N_n}{}^2/s_{[nN]}{}^2 - 1\,| \ge C_1\epsilon_n]$ and $P[\,|\,s_{N_n}{}^2/s_{[nN]}{}^2 - 1\,| > C_1\epsilon_n] = O(\epsilon_n{}^{\delta/2})$, and lemma 1 of Michel and Pfanzagl (1971).

Proof of Theorem 2.1. Define $I_n = \{k:[nN](1 - C_1\epsilon_n) \le k \le [nN](1 + C_1\epsilon_n)\}$.

Then, $\overline{\Delta}_{1n} \le \sup_x |\,P[(\overline{S}_{N_n} - L_n)/s_{[nN]} \le x,\ N_n \in I_n] - \Phi(x)\,| + P[N_n \notin I_n] \le \sup_x |\,P[(\overline{S}_{N_n} - L_n)/s_{[nN]} \le x,\ N_n \in I_n] - \Phi(x)\,| + O(\epsilon_n{}^{\delta/2})$.

Following Landers and Rogge (1977) and Theorem 3 of Ahmad and Basu (1979), we need to show that $|\,J(x)\,| = O(\epsilon_n{}^{\delta/2})$ where

$$J(x) = P[(\overline{S}_{N_n} - L_n)/s_{[nN]} \le x] - P\left[\frac{\overline{S}_{[nN]} - \sum_{k=1}^{[nN]} \mu_k}{s_{[nN]}} \le x\right]$$

Let $b_n(x) = L_n + x\, s_{[nN]}$. Then

$$\begin{aligned} J(x) &\le P[(\overline{S}_{N_n} - L_n)/s_{[nN]} \le x,\ N_n \in I_n] + P[N_n \notin I_n] \\ &\quad - P[(\overline{S}_{[nN]} - \textstyle\sum_1^{[nN]} \mu_k)/s_{[nN]} \le x] \\ &\le O(\epsilon_n{}^{\delta/2}) + P[\overline{S}_{N_n} \le b_n(x),\ N_n \in I_n] - P[\overline{S}_{[nN]} \le \textstyle\sum_1^{[nN]} \mu_k + xs_{[nN]}] \\ &= O(\epsilon_n{}^{\delta/2}) + P[\overline{S}_{[(nN)(1-C_1\epsilon_n)]} \le b_n(x)] - P[\overline{S}_{[nN]} < \textstyle\sum_1^{[nN]} \mu_k + x\, s_{[nN]}] \end{aligned}$$

Since N is independent of the X_i's, we have

$$\begin{aligned} J(x) &\le O(\epsilon_n{}^{\delta/2}) + \textstyle\sum_{k=1}^{[C_2/\epsilon_n]} P[[nN] = k] + \sum_{k=[C_2/\epsilon_n]+1}^{\infty} P[[nN] = k] \\ &\quad \cdot \{P[\overline{S}_{[k(1-C_1\epsilon_n)]} \le A_k(1 + C_1\epsilon_n) + xs_k] - P[\overline{S}_k \le A_k + xs_k]\} \\ &\le O(\epsilon_n{}^{\delta/2}) + P\left[N < \frac{C_2}{n\epsilon_n}\right] + \textstyle\sum_{k=[C_2/\epsilon_n]+1}^{\infty} P[[nN] = k]\, \dfrac{C}{k^{\delta/2}} = O(\epsilon_n{}^{\delta/2}), \end{aligned}$$

where C is a constant independent of k.

In a similar fashion we can also show that $-J(x) = O(\epsilon_n^{\delta/2})$.

$$I = \sup_x | P[(\bar{S}_{[nN]} - \sum_1^{[nN]} \mu_k)/s_{[nN]} \leq x] - \Phi(x) |$$
$$\leq \sum_{k=1}^{[C_2/\epsilon_n]-1} P[[nN] = k] + \sum_{k=[C_2/\epsilon_n]}^{\infty} P[[nN] = k] \bar{\Delta}_k$$
$$= P[N < C_2/n\epsilon_n] + \sum_{k=[C_2/\epsilon_n]}^{\infty} P[[nN] = k] \frac{C}{k^{\delta/2}} = O(\epsilon_n^{\delta/2})$$

for some constant C.

The proof that $\sum_{n=1}^{\infty} \epsilon_n^{1-\delta/2} \bar{\Delta}_{1n} < \infty$ follows the same way as before and as Ahmad and Basu (1979). The proof that $\sum_{n=1}^{\infty} \epsilon_n^{1-\delta/2}\bar{\Delta}_{2n} < \infty$ follows from the next well known lemma.

LEMMA A. *Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables and let $\{a_n\}$ be a sequence of nonnegative real numbers such that $a_n = O(n^{-1})$. Assume that for some $\alpha > 0$,*

$$\sum_{n=1}^{\infty} a_n^{\alpha} | P[X_n \leq x] - \Phi(x) | < \infty \quad \textit{and} \quad \sum_{n=1}^{\infty} a_n^{\alpha} P[| Y_n - 1 | \geq a_n] < \infty,$$

then

$$\sum_{n=1}^{\infty} a_n^{\alpha} | P[X_n \leq xY_n] - \Phi(x) | < \infty$$

(this follows from lemma 1 of Michel and Pfanzagl (1971)).

First we show that $\sum_{n=1}^{\infty} \epsilon_n^{1-\delta/2}\bar{\Delta}_{[nN]} < \infty$, but we can easily see that

$$\bar{\Delta}_{[nN]} \leq P[N < C_2 n\epsilon_n] + \sum_{k=[C_2/\epsilon_n]+1}^{\infty} \bar{\Delta}_k P[[nN] = k].$$

Hence

$$\sum_{n=1}^{\infty} \epsilon_n^{1-\delta/2}\bar{\Delta}_{[nN]}$$
$$\leq \sum_{n=1}^{\infty} \epsilon_n^{1-\delta/2} P\left[N < \frac{C_2}{n\epsilon_n}\right] + \sum_{n=1}^{\infty} \epsilon_n^{1-\delta/2} \sum_{k=[C_2/\epsilon_n]+1}^{\infty} P([nN] = k),$$

where first term is finite by condition (1.8) and the rest follows as in the previous case.

3. Rates of convergence for partial sums

In this section we exploit the methods of Landers and Rogge (1977), and their extensions by Rychlik (1978) and obtain various generalizations of their results.

THEOREM 3.1. (a) *If conditions* (1.5), (1.6) *and* (1.11) *are satisfied then* $\Delta_{1n} = O(\epsilon_n^{\delta/2})$. *If* (1.7) *is also satisfied, then* $\Delta_{2n} = O(\epsilon_n^{\delta/2})$.

(b) *If conditions* (1.8) *and* (1.9) *are satisfied and if* $\sum_{n=1}^{\infty} n^{-1+\delta/2}\Delta_n < \infty$, $0 \leq \delta < 1$, *then*

$$\sum_{n=1}^{\infty} \epsilon_n^{1-\delta/2}\Delta_{in} < \infty,\ i = 1, 2,\ 0 \leq \delta < 1.$$

Remark: (1) As we shall see in the proof of Theorem 3.1 (a), it suffices to assume that $\epsilon_n \geq n^{-1}$ for all ≥ 1. The condition $\epsilon_n = O(n^{-1})$ is necessary, however, to obtain the second part of the theorem. (2) $\Delta_n = O(n^{-\delta/2})$ holds by Berry-Esseen theorem under the condition (1.11).

Proof. Without loss of generality we may assume that $\mu_k = 0$ for $k \geq 1$. Let $b_n(x) = xS_{[nN]}$.

Following the arguments of Landers and Rogge (1977) first we show that

$$\Delta_{[nN]} = \sup_x P[S_{[nN]} \leq b_n(x)] - \Phi(x) \mid = O(\epsilon_n{}^{\delta/2}).$$

Since N is independent of the X_i's and since $\Delta_n = O(n^{-\delta/2})$,

$$\begin{aligned} \Delta_{[nN]} &= \sup_x \mid \textstyle\sum_{k=1}^{\infty} P[S_k \leq xs_k]P[nN = k] - \Phi(x) \mid \\ &\leq \textstyle\sum_{k=1}^{[C_2/\epsilon_n]-1} P[nN = k] + \sum_{k=[C_2/\epsilon_n]}^{\infty} \Delta_k P[nN = k] \\ &= P[N < C_2/n\epsilon_n] + \textstyle\sum_{k=[C_2/\epsilon_n]}^{\infty} P[nN = k]C/k^{\delta/2} \\ &= O(\epsilon_n{}^{\delta/2}) + O(\epsilon_n{}^{\delta/2}), \end{aligned}$$

where C is an absolute constant independent of n, and since $\epsilon_n \geq 1/n$.

Now combining with lemma 2 and the arguments of Landers and Rogge (1977) $\Delta_{1n} = O(\epsilon_n{}^{\delta/2})$.

To prove $\Delta_{2n} = O(\epsilon_n{}^{\delta/2})$ it suffices to observe that

$$[\mid s_{N_n}/s_{[nN]} - 1 \mid \geq \epsilon_n{}^{1/2}] \subseteq [\mid s_{N_n}{}^2/s_{[nN]}{}^2 - 1 \mid \geq \epsilon_n],$$

condition (1.6) and lemma of Michel and Pfanzagl (1971). The rest of the proof follows same as Landers and Rogge (1977).

CENTRO DE INVESTIGACIÓN DEL IPN, MÉXICO, D. F.

REFERENCES

[1] I. A. AHMAD, AND A. K. BASU, *On rates of convergence in the random CLT for partial sums and maximum partial sums.* To appear in Sankhya.

[2] D. LANDERS, AND L. ROGGE, *The exact approximation order in the CLT for random summation.* Z. Wahrscheinlichkeitstheorie und verw. Gebiete, **36** (1976), 269–283.

[3] D. LANDERS, AND L. ROGGE, *A counter example in approximation theory of random summation.* Ann. Prob., **5** (1977), 1018–1023.

[4] R. MICHEL, AND J. PFANZAGL, *The accuracy of the normal approximation for minimum contrast estimator.* Z. Wahrscheinlichkeitstheorie und verw. Gebiete, **18** (1971), 73–94.

[5] Z. RYCHLIK, The Order of Approximation in the Random Central Limit Theorem. Lecture Notes in Mathematics, No. 656, Springer-Verlag, (1978), 225–236.

[6] C. C. HEYDE, AND P. HALL, Martingale Limit Theory and its Application, Academic Press, New York, 1980.