



King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 022

March 1981

**On the Deviation Between the Distributions of Sums
and Maximum Sums of IDD Random Vectors**

Ibrahim A. Ahmad

On the Deviation Between the Distributions of Sums
and Maximum Sums of IID Random Vectors

By

Ibrahim A. Ahmad*
Department of Mathematical Sciences
University of Petroleum and Minerals
Dhahran, Saudi Arabia

Abstract

For a sequence of independent and identically distributed random vectors, with finite moment of order less than or equal to the second, the rate at which the deviation between the distribution functions of the vectors of partial sums and maximums of partial sums is obtained both when the sample size is fixed and when it is random, satisfying certain regularity conditions. When the second moments exist the rate is of order $n^{-1/4}$ (in the fixed sample size case). Two applications are given, first, we compliment some recent work of Ahmad (1979, J. Mult. Analysis, 9, 214-222) on rates of convergence for the vector of maximum sums and second, we obtain rates of convergence of the concentration functions of maximum sums for both the fixed and random sample size cases.

AMS 1970 Subject Classification: Primary: 60F05

Key Words and Phrases: Vector of maximum sums, Independent random variables, distribution functions, concentration function.

1. Introduction.

Let $\underline{X}_1, \dots, \underline{X}_n, \dots$, be a sequence of independent identically distributed k -dimensional random vectors such that $E\underline{X}_1 = \underline{\mu} > \underline{0}$ ($\underline{\mu} > \underline{0}$ means $\mu_i > 0, i = 1, \dots, k$).

Define

$$\underline{S}_n = \sum_{j=1}^n \underline{X}_j = (S_{1n}, \dots, S_{kn})', \text{ with } S_{in} = \sum_{j=1}^n X_{ij}, i = 1, \dots, k, \quad (1.1)$$

and set

$$\underline{S}_n^* = (\max_{1 \leq j \leq n} S_{1j}, \dots, \max_{1 \leq j \leq n} S_{kj})' = (S_{1n}^*, \dots, S_{kn}^*)', \text{ say} \quad (1.2)$$

It is well-known that if $EX_{i1}^2 < \infty, i = 1, \dots, k$, then

$$\Delta_n = \sup_{\underline{x}} |P[S_{1n} - n\mu_1 \leq \sigma_1 x_1 \sqrt{n}, \dots, S_{kn} - n\mu_k \leq \sigma_k x_k \sqrt{n}] - \phi_R(\underline{x})| \rightarrow 0, \quad (1.3)$$

as $n \rightarrow \infty$,

and

$$\Delta_n^* = \sup_{\underline{x}} |P[S_{1n}^* - n\mu_1 \leq \sigma_1 x_1 \sqrt{n}, \dots, S_{kn}^* - n\mu_k \leq \sigma_k x_k \sqrt{n}] - \phi_R(\underline{x})| \rightarrow 0, \quad (1.4)$$

as $n \rightarrow \infty$,

where $\phi_R(\cdot)$ denote the distribution function of a multivariate normal variate with mean vector $\underline{0}$ and covariance matrix R . Recently, Ahmad [1], developed the rates of convergence of Δ_n^* in terms of those of Δ_n .

Precisely, he showed that if $\Delta_n = O(n^{-\delta/2})$, then $\Delta_n^* = O(n^{-\delta/2}), 0 < \delta \leq 1$ and that if $\sum_{n=1}^{\infty} n^{-1+\delta/2} \Delta_n < \infty$

then $\sum_{n=1}^{\infty} n^{-1+\delta/2} \Delta_n^* < \infty, 0 < \delta < 1$. In this second part

the case $\delta = 0$ was left unanswered. One immediate consequence of the first result reported here is to find the answer.

The center of the proofs developed in Ahmad [1] is to obtain rates of convergence of the deviation between the distribution function of the vector of sums and the vector of maximum sums. Precisely, it is proved that if $E|X_{i1}|^{2+\delta} < \infty$, $0 < \delta \leq 1$, $i = 1, \dots, k$, then

$$\sup_{\underline{x}} |P[\underline{S}_n \leq \underline{x}] - P[\underline{S}_n^* \leq \underline{x}]| = O(n^{-\delta/2}), \quad 0 < \delta \leq 1. \quad (1.5)$$

Thus it seemed natural to ask about the case when moments less than or equal to the second are the only moments we can assume finite? This is one of the questions we attempt to answer in the present investigation.

On the other hand, if we let $\{N_n\}$ be a sequence of integer valued random variables not necessarily independent of $\{X_n\}$ and such that (N_n/n) converges in probability to a positive random variable N , independent of $\{X_n\}$ and such that $E(N) < \infty$, then it is possible to prove that if $\{\epsilon_n\}$ is a sequence of real numbers such that $\epsilon_n \geq n^{-1}$ for all $n \geq 1$, and that for some constant C_1 and C_2 the following condition holds:

$$(i) \quad P[|N_n/N - 1| > C_1 \epsilon_n] = O(\epsilon_n^{\delta/2}), \quad \text{and}$$

$$(ii) \quad P[N < C_2/n\epsilon_n] = O(\epsilon_n^{\delta/2}),$$

and if

$$E|X_{i1}|^{2+\delta} < \infty, \quad i = 1, \dots, k, \quad 0 < \delta \leq 1, \quad \text{then}$$

$$\sup_{\underline{x}} |P[\underline{S}_{N_n} \leq \underline{x}] - P[\underline{S}_{N_n}^* \leq \underline{x}]| = O(\epsilon_n^{\delta/2}), \quad 0 < \delta \leq 1. \quad (1.5)$$

A careful inspection of the proof of Theorem 2.2, to follow, reveals that the same argument can be used to show (1.5) by an application of Theorem 1 of Ahmad [1]. Thus our second goal is to deal with the random size case when only

$$E|X_{i1}|^{1+\alpha} < \infty, \quad i = 1, \dots, k \quad \text{and} \quad 0 < \alpha \leq 1.$$

In Section 3 two applications are presented, first we use Theorem 2.1 to show that if $\sum_{n=1}^{\infty} n^{-1} \Delta_n < \infty$, then $\sum_{n=1}^{\infty} n^{-1} \Delta_n^* < \infty$, and in the second application we obtain rates of convergence of the concentration function of the maximum of partial sums (we discuss the case $k = 1$), both for fixed and random sample sizes.

2. Main Results.

THEOREM 2.1. Let $\{X_n\}$ be a sequence of iid random vectors such that $EX_1 = \underline{\mu} > \underline{0}$. If $E|X_{i1} - \mu_i|^{1+\alpha} < \infty$ for some $0 < \alpha < 1, i=1, 2, \dots, k$, then

$$\sup_{\underline{x}} |P[S_n \leq \underline{x}] - P[S_n^* \leq \underline{x}]| = O(n^{-\alpha/2(1+\alpha)}). \quad (2.1)$$

PROOF. Clearly, for any \underline{x} ,

$$\begin{aligned} P[S_n \leq \underline{x}] - P[S_n^* \leq \underline{x}] &= P[S_n \leq \underline{x}, \bigcup_{i=1}^k \{S_{in}^* > x_i\}] \leq \sum_{i=1}^k P[S_n \leq \underline{x}, S_{in}^* > x_i] \\ &\leq \sum_{i=1}^k P[S_{in} \leq x_i, S_{in}^* > x_i] = \sum_{i=1}^k \{P[S_{in} \leq x_i] - P[S_{in}^* \leq x_i]\}. \end{aligned} \quad (2.2)$$

Thus it suffices to show that if $E|X_{i1} - \mu_i|^{1+\alpha} < \infty$, then

$$\sup_{x_i} |P[S_{in} \leq x_i] - P[S_{in}^* \leq x_i]| = O(n^{-\alpha/2(1+\alpha)}), \quad i = 1, 2, \dots, k.$$

With hopefully no confusion we shall drop the suffix i from now

on. Let $a_n = n^{1/2(1+\alpha)}$,

$$\begin{aligned}
& P[S_n \leq x] - P[S_n^* \leq x] \\
&= P[S_n \leq x, S_n^* > x] \leq P[S_n \leq x, S_n^* > x, S_n^* - S_n \leq a_n] + P[S_n^* - S_n \geq a_n] \\
&\leq P[x - a_n < S_n \leq x] + P[S_n^* - S_n \geq a_n] \leq Q(S_n, a_n) + P[S_n^* - S_n \geq a_n],
\end{aligned} \tag{2.3}$$

where $Q(\xi, x) = \sup_z P[z < \xi \leq z + x]$ is the concentration function of the random variable ξ . Petrov [3] proved that if $\{X_n\}$ are nondegenerate, then $Q(S_n, x) \leq K \frac{x+1}{\sqrt{n}}$ for all $x \geq 0$, and all $n \geq 1$, where K is a positive constant independent of n and x . Thus

$$Q(S_n, a_n) \leq K \frac{1+a_n}{\sqrt{n}} \leq Kn^{-1/2+1/2(1+\alpha)} = Kn^{-\alpha/2(1+\alpha)}. \tag{2.4}$$

Next, let us evaluate an upper bound for $P[S_n^* - S_n \geq a_n]$.

$$\begin{aligned}
& P[S_n^* - S_n \geq a_n] \\
&= P[\max\{(-X_2 - X_3 - \dots - X_n), (-X_3 - \dots - X_n), \dots, (-X_n), 0\} \geq a_n] \\
&= P[\max\{(Y_1 + Y_2 + \dots + Y_{n-1}), (Y_1 + Y_2 + \dots + Y_{n-2}), \dots, (Y_1), 0\} \geq a_n] \\
&\leq P[\max\{0, Y_1, (Y_1 + Y_2), \dots, (Y_1 + Y_2 + \dots + Y_n)\} \geq a_n] = P[\max_{1 \leq j \leq n} T_j \geq a_n], \text{ say} \\
&\leq \sum_{k=0}^{m-1} P[\max\{T_{p_k+1}, \dots, T_{p_{k+1}}\} \geq a_n],
\end{aligned} \tag{2.5}$$

where $Y_j = -X_{n-j+1}$, $T_j = \sum_{i=1}^j Y_i$, $j = 1, \dots, n$,

$l=p_0 < p_1 < p_2 < \dots < p_{m-1} < p_m = n$ and $m = m(n)$ are integers such that $p_1 = [a_n]$ and $p_k = 2^{k-1}[a_n]$ with $[x]$ denoting the largest integer less than or equal to x , and m is such that $2^{m-2}a_n \leq n \leq 2^{m-1}a_n$. Next, note that

$$\begin{aligned}
& P[\max\{T_{p_k+1}, \dots, T_{p_{k+1}}\} \geq a_n] \\
&\leq P[\max\{|T_{p_k+1} - ET_{p_k+1}|, \dots, |T_{p_{k+1}} - ET_{p_{k+1}}|\} \geq a_n - ET_{p_k}]
\end{aligned}$$

$$\begin{aligned}
&\leq P[\max\{|T_1 - ET_1|, \dots, |T_{P_{k+1}} - ET_{P_{k+1}}|\} \geq a_n - ET_{P_k}] \\
&\leq (a_n + p_k \mu)^{-(1+\alpha)} E[\max\{|T_1 - ET_1|^{1+\alpha}, \dots, |T_{P_{k+1}} - ET_{P_{k+1}}|^{1+\alpha}\}] \\
&\leq (a_n + p_k \mu)^{-(1+\alpha)} C_\alpha \sum_{\ell=1}^{P_{k+1}} E|Y_\ell - EY_\ell|^{1+\alpha} \\
&\leq C_\alpha^* p_{k+1} / (a_n + p_k \mu)^{(1+\alpha)}, \tag{2.6}
\end{aligned}$$

where the first inequality is obtained since $ET_n = -n\mu_1 < 0$ and with C_α and C_α^* positive constants independent of n . From (2.5) and (2.6) it follows that,

$$\begin{aligned}
P[S_n^* - S_n \geq a_n] &\leq \tilde{C}_\alpha \sum_{k=0}^{m-1} \frac{p_{k+1}}{(a_n + p_k)^{1+\alpha}} \leq \tilde{C}_\alpha \left[\frac{p_1}{a_n^{1+\alpha}} + \sum_{k=1}^{m-1} \frac{p_{k+1}}{p_k^{1+\alpha}} \right] \\
&\leq \tilde{C}_\alpha \left[\frac{p_1}{a_n^{1+\alpha}} + \sum_{k=1}^{\infty} \frac{2^k [a_n]}{(2^{k-1} [a_n])^{1+\alpha}} \right] \leq \tilde{C}_\alpha [a_n]^{-\alpha} (1 + 2^{1+\alpha} \sum_{k=1}^{\infty} 2^{-k\alpha}) \\
&= \tilde{C}_\alpha^* [a_n]^{-\alpha}, \text{ say.} \tag{2.7}
\end{aligned}$$

Thus the theorem is proved. \square

A random sum version of Theorem 2.1 is given next.

THEOREM 2.2. Let $\{X_n\}$ be a sequence of iid random vectors such that $EX_1 = \underline{\mu} > \underline{0}$ and assume that $E|X_{i1} - \mu_i|^{1+\alpha} < \infty$, $i=1,2,\dots,k$, for some $0 < \alpha \leq 1$. Further, let $\{N_n\}$ be a sequence of integer-valued random variables (not necessarily independent of the $\{X_i\}$) such that $\frac{N_n}{n}$ converges in probability to a positive random variable N independent of the X_i 's. Assume that $E(N) < \infty$ and that there exists positive constants C_1 and C_2 and a sequence $\{\epsilon_n\}$ of reals such that,

(i) $\epsilon_n \geq n^{-1}$ and $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

(ii) $P\left\{ \left| \frac{N_n}{[nN]} - 1 \right| > C_1 \epsilon_n \right\} = o(\epsilon_n^{\alpha/2(1+\alpha)})$, where $[x]$ denote the

largest integer less than or equal to x .

$$(iii) P[N < C_2/n\epsilon_n] = o(\epsilon_n^{\alpha/2(1+\alpha)}).$$

Then

$$\sup_{\underline{x}} |P[S_{N_n} \leq \underline{x}] - P[S^*_{N_n} \leq \underline{x}]| = o(\epsilon_n^{\alpha/2(1+\alpha)}). \quad (2.8)$$

PROOF. Again as in Theorem 2.1 it suffices to prove the theorem for the univariate case. First we show that for any real x ,

$$P[S_{[nN](1+C_1\epsilon_n)} \leq x] - P[S^*_{[nN](1+C_1\epsilon_n)} \leq x] = o(\epsilon_n^{\alpha/2(1+\alpha)}). \quad (2.9)$$

Since N is independent of the X_i 's we obtain from Condition (iii) that

$$\begin{aligned} & P[S_{[nN](1+C_1\epsilon_n)} \leq x] - P[S^*_{[nN](1+C_1\epsilon_n)} \leq x] \\ &= P[S_{[nN](1+C_1\epsilon_n)} \leq x, S^*_{[nN](1+C_1\epsilon_n)} > x] \\ &= \sum_{\ell=1}^{\infty} P[S_{\ell} \leq x, S^*_{\ell} > x] P([nN](1+C_1\epsilon_n) = \ell) \\ &\leq P[nN < C_2/\epsilon_n(1+C_1\epsilon_n)] + \sum_{\ell=[C_2/\epsilon_n]}^{\infty} P[S_{\ell} \leq x, S^*_{\ell} > x] P([nN](1+C_1\epsilon_n) = \ell) \\ &\leq P[N < C_2/n\epsilon_n] + \sum_{\ell=[C_2/\epsilon_n]}^{\infty} O(\ell^{-\alpha/2(1+\alpha)}) P([nN](1+C_1\epsilon_n) = \ell) \\ &= o(\epsilon_n^{\alpha/2(1+\alpha)}) + o(\epsilon_n^{\alpha/2(1+\alpha)}). \end{aligned} \quad (2.10)$$

Let $I_n = \{k \mid [nN](1-C_1\epsilon_n) \leq k \leq [nN](1+C_1\epsilon_n)\}$. Note that $P[N_n \notin I_n] = o(\epsilon_n^{\alpha/2(1+\alpha)})$. Thus for any real number x ,

$$\begin{aligned} & P[S_{N_n} \leq x] - P[S^*_{N_n} \leq x] \\ &= P[S_{N_n} \leq x, S^*_{N_n} > x] \leq P[S_{N_n} \leq x, S^*_{N_n} > x, N_n \in I_n] + P[N_n \notin I_n] \\ &\leq P[S_{N_n} \leq x, S^*_{N_n} \geq x, N_n \in I_n] + o(\epsilon_n^{\alpha/2(1+\alpha)}). \end{aligned} \quad (2.11)$$

Define the integer-valued random variables $L_n = [nN](1-C_1\epsilon_n)$

and $M_n = \{[nN](1 + C_1 \epsilon_n)\}$. Then

$$\begin{aligned}
 & P[S_{N_n} \leq x, S_{N_n}^* > x, N_n \in I_n] \\
 &= P[\{S_{L_n} \leq x, S_{L_n}^* > x\} \cup \dots \cup \{S_{M_n} \leq x, S_{M_n}^* > x\}] \\
 &= \sum_{\ell=1}^{\infty} P[[nN] = \ell] P[\{S_{[\ell(1-C_1 \epsilon_n)]} \leq x, S_{[\ell(1-C_1 \epsilon_n)]}^* > x\} \\
 &\quad \cup \dots \cup \{S_{[\ell(1+C_1 \epsilon_n)]} \leq x, S_{[\ell(1+C_1 \epsilon_n)]}^* > x\}] \\
 &\leq P[N < (C_2/n\epsilon_n)] + \sum_{\ell=[C_2/\epsilon_n]}^{\infty} P[[nN] = \ell] \\
 &\quad \sum_{k=[\ell(1-C_1 \epsilon_n)]}^{[\ell(1+C_1 \epsilon_n)]} P[S_k \leq x, S_k^* > x] \\
 &\leq O(\epsilon_n^{\alpha/2(1+\alpha)}) + \sum_{\ell=[C_2/\epsilon_n]}^{\infty} (2\ell C_1 \epsilon_n)^{\alpha} P[[nN] = \ell] O(\epsilon_n^{\alpha/2(1+\alpha)}),
 \end{aligned}
 \tag{2.12}$$

where the second term of the last upper bound follows from Theorem 1 above. Now if $\ell > [C_2/\epsilon_n]$, then $\ell(1 - C_1 \epsilon_n) > Cn$ for some positive constant C . Hence $\epsilon_n \ell(1 - C_1 \epsilon_n) \leq (Cn)^{-1}$ and the second term in the last upper bound of (2.12) is less than or equal to

$$Cn^{-\alpha(2(1+\alpha))} \epsilon_n E([nN]) = O(\epsilon_n^{\alpha/2(1+\alpha)}). \tag{2.13}$$

The proof of the theorem is now complete.

3. Two Applications.

(i) Rates of convergence for the vector of maximum sum:

In the following theorem we employ Theorem 2.1 to complete some recent results reported in Ahmad [1] concerning the rates of

convergence of the vector of maximum sums. Assume that $E|X_{i1}|^2 < \infty$, $i = 1, \dots, k$.

THEOREM 3.1. (i) If $\Delta_n = O(n^{-\delta/2})$, then $\Delta_n^* = O(n^{-\delta/2})$ for all $0 < \delta \leq 1$.

(ii) If $\sum_{n=1}^{\infty} n^{-1+\delta/2} \Delta_n < \infty$, then $\sum_{n=1}^{\infty} n^{-1+\delta/2} \Delta_n^* < \infty$, for all $0 \leq \delta < 1$.

PROOF. Part (i) and Part (ii) for $0 < \delta < 1$ are given in Theorems 1 and 2 of Ahmad [1]. Thus we only need to prove Part (ii) for $\delta = 0$. As in Ahmad [1] we need only to show that for all \underline{x}

$$\sum_{n=1}^{\infty} n^{-1} \{P[\underline{S}_n \leq \underline{x}] - P[\underline{S}_n^* \leq \underline{x}]\} < \infty. \quad (3.1)$$

But since $E|X_{i1}|^2 < \infty$, then by Theorem 2.1, for all \underline{x}

$$P[\underline{S}_n \leq \underline{x}] - P[\underline{S}_n^* \leq \underline{x}] = O(n^{-1/4}). \quad (3.2)$$

Hence

$$\sum_{n=1}^{\infty} n^{-1} \{P[\underline{S}_n < \underline{x}] - P[\underline{S}_n^* < \underline{x}]\} \leq \sum_{n=1}^{N_0} n^{-1} \{P[\underline{S}_n < \underline{x}] - P[\underline{S}_n^* < \underline{x}]\} + C \sum_{n=N_0+1}^{\infty} n^{-1-1/4} < \infty. \quad (3.3)$$

Thus the theorem is proved.

(ii) Rates of convergence of the concentration function of maximum sum: In this application we assume that X_1, \dots, X_n are univariate. The results apply to the multivariate case without difficulty. Recall the definition of the concentration function of a random variable X ; $Q(X, x) = \sup_y P[y < X \leq x + y]$, $x \geq 0$. Let $S_n^* = \max_{1 \leq j \leq n} S_j$ with $S_j = \sum_{\ell=1}^j X_{\ell}$, $j=1, \dots, n$. We seek rates at which $Q(S_n^*, x)$ and $Q(S_{N_n}^*, x)$ diminish to 0 as $n \rightarrow \infty$. Although the rates we obtain below are not as good as those of S_n (of order

$n^{-1/2}$) itself we believe that they are new and hope that they will stimulate interest for further improvements.

THEOREM 3.2. If $Q(S_n, x) = o(n^{-1/2})$ and if $E|X_1|^{1+\alpha} < \infty$, then for any $0 < \alpha \leq 1$

$$Q(S_n^*, x) = o(n^{-\alpha/2(1+\alpha)}). \quad (3.4)$$

PROOF. Note that

$$\begin{aligned} Q(S_n^*, x) &= \sup_y |P[S_n^* \leq x+y] - P[S_n^* \leq x]| \\ &\leq \sup_y \{P[S_n \leq x+y] - P[S_n^* \leq x+y]\} + \{P[S_n \leq x] - P[S_n^* \leq x]\} + Q(S_n, x) \\ &\leq o(n^{-\alpha/2(1+\alpha)}) + o(n^{-\alpha/2(1+\alpha)}) + o(n^{-1/2}) = o(n^{-\alpha/2(1+\alpha)}), \end{aligned} \quad (3.5)$$

by an application of Theorem 2.1 and using the assumption that $Q(S_n, x) = o(n^{-1/2})$.

Note that a sufficient condition for $Q(S_n, x) = o(n^{-1/2})$, is that, see Petrov [3], X be nondegenerate. Next, we give a random sample size version of the above result using Theorem 2.2.

THEOREM 3.3. If $Q(S_{N_n}, x) = o(\epsilon_n^{1/2})$, if $E|X_1|^{1+\alpha} < \infty$, and if $\{N_n\}$ satisfy the conditions of Theorem 2.2, then for any $0 < \alpha \leq 1$,

$$Q(S_{N_n}^*, x) = o(\epsilon_n^{\alpha/2(1+\alpha)}). \quad (3.6)$$

PROOF. Again we easily see that

$$\begin{aligned}
Q(S_{N_n}^*, x) &\leq \sup_y \{P[S_{N_n}^* \leq x+y] - P[S_{N_n}^* \leq x]\} + \{P[S_{N_n} \leq x] - P[S_{N_n}^* \leq x]\} + Q(S_{N_n}, x) \\
&= O(\epsilon_n^{\alpha/2(1+\alpha)}) + O(\epsilon_n^{\alpha/2(1+\alpha)}) + O(\epsilon_n^{1/2}) = O(\epsilon_n^{\alpha/2(1+\alpha)}),
\end{aligned}$$

by an application of Theorem 2.2 and the assumption that

$$Q(S_{N_n}, x) = O(\epsilon_n^{1/2}). \quad \square$$

We remark here that Ahmad [2] has recently shown that under the conditions of Theorem 2.2 concerning N_n , if $Q(S_n, x) = O(n^{-1/2})$, then $Q(S_{N_n}, x) = O(\epsilon_n^{1/2})$, then, e.g., if X is nondegenerate we have $Q(S_{N_n}, x) = O(\epsilon_n^{1/2})$ and thus Theorem 3.2 applies.

REFERENCES

1. I.A. AHMAD, "A note on rates of convergence in the multidimensional CLT for maximum partial sums", J. Multivariate Analysis, 9, (1979), 214-222.
2. I.A. AHMAD, "On concentration functions of random sums of independent random variables", J. London Math. Soc., 21, (1980), 189-192.
3. V.V. PETROV, "On an estimation of the concentration function of a sum of independent random variables", Theor. Probability Appl., 15, (1970), 701-703.