International Mathematical Forum, 4, 2009, no. 26, 1281 - 1288

Berry-Esseen Bounds for Random Sums of Non-i.i.d. Random Variables

N. Chaidee[1] and M. Tuntapthai[2]

[1,2]Department of Mathematics, Faculty of Science
Chulalongkorn University, Bangkok 10330, Thailand

[1]National Centres of Excellence in Mathematics, PERDO
Bangkok 10400, Thailand
nattakarn.c@chula.ac.th

Abstract

Let (X_n) be a sequence of independent random variables and N a positive integral-valued random variable such that $N, X_1, X_2, \ldots$ are independent. In 2008, N.Chaidee and K.Neammenee [10] gave Berry-Esseen bounds for random sums in the case of $X_1, X_2, \ldots$ are identically distributed random variables by using Stein's method. In this work, we use the technique from [10] to establish the Berry-Esseen bounds without assuming the identical distributions of $X_1, X_2, \ldots$.

Mathematics Subject Classification: 60F05, 60G05

Keywords: Berry-Esseen bound, Stein's method, concentration inequality, random sums, normal approximation

1 Introduction

The central limit theorem (CLT) is one of the fundamental theorems in probability and applied statistics which gives an approximation for the distribution

[1]The research of the first author was partially supported by National Centres of Excellence in Mathematics and Grants for Development of New Faculty Staff from Chulalongkorn University, Thailand.

of the sum of a large number of independent and identically distributed random variables.

In 1941 and 1945, Berry [1] and Esseen [5] obtained the rate of convergence with assuming the third moment finite. Since then, many authors, e.g. [7], [11], [13], improve the bound in the Berry-Esseen theorem and the method that they used is Fourier method. In 2001, Chen and Shao [9] have used a new method, Stein's method, to establish Berry-Esseen bounds for the sums of independent random variables without assuming the existence of third moment.

In many situations, we would like to apply CLT to randomly indexed sums. The CLT for random sums was obtained by many mathematicians, e.g. [3], [4], [6], [12], and they still used Fourier method. In this paper, we give both uniform and non-uniform Berry-Esseen bounds for random sums in the case of $X_1, X_2, \ldots$ are not necessarily identically distributed random variables by using Stein's method.

2 Uniform Bound

In 1977, Batirov, Manevich and Nagaev [8] obtained a uniform bound for random sums in the case of $X_1, X_2, \ldots$ are independent and not necessarily identically distributed random variables by using Fourier method. The following is their result.

Theorem 1. ([8]) *Let $X_1, X_2, \ldots$ be independent and not necessarily identically distributed random variables with $EX_i = 0$, $EX_i^2 = \sigma_i^2$ and $E|X_i|^3 = \gamma_i < \infty$ and N a positive integral-valued random variable such that $N, X_1, X_2, \ldots$ are independent. Set $s_N^2 = \sum_{i=1}^{N} \sigma_i^2$, $\beta_N = \sum_{i=1}^{N} \gamma_i$ and $W = \frac{W_N}{\sqrt{Es_N^2}}$ where $W_N = X_1 + X_2 + \cdots + X_N$. Then there exists a constant C such that*

$$\sup_{x\in\mathbb{R}} |P(W \le x) - \Phi(x)| \le C\Big(\frac{E\beta_N}{(Es_N^2)^3} + \frac{\sqrt{E|s_N^2 - Es_N^2|}}{(Es_N^2)^2}\Big)$$

where $\Phi(x)$ is the standard normal distribution function.

In 2008, N.Chaidee and K.Neammanee [10] applied Stein's method to obtained the constant C in Theorem 1 in the case of the random variables $X_1, X_2, \ldots$ are identically distributed.

Theorem 2. ([10]) *Let $X_1, X_2, \ldots$ be independent identically distributed random variables with $EX_i = 0$, $EX_i^2 = \sigma^2$ and $E|X_i|^3 = \gamma < \infty$ and N a positive*

integral-valued random variable with $EN^{3/2} < \infty$ *such that* $N, X_1, X_2, \dots$ *are independent. Set* $W = \dfrac{W_N}{\sqrt{EN}\sigma}$ *where* $W_N = X_1 + X_2 + \cdots + X_N$. *Then we have*

$$\sup_{x\in\mathbb{R}}|P(W \le x) - \Phi(x)| \le 10.32\delta + \frac{0.125}{\sqrt{EN}} + \frac{1.5\delta EN^{3/2}}{(EN)^{3/2}} + \frac{E|N - EN|}{EN}$$

where $\delta = \dfrac{\gamma}{\sqrt{EN}\sigma^3}$.

In this section, we generalize Theorem 2 in the case of $X_1, X_2, \dots$ are not necessarily identically distributed random variables. The following is our main result.

Theorem 3. *Under the condition in Theorem 1, we have*

$$\sup_{x\in\mathbb{R}}|P(W \le x) - \Phi(x)| \le 6.875\frac{E[\delta_N\sqrt{Es_N^2}]}{\sqrt{Es_N^2}} + 1.5\frac{E[\delta_N(s_N^2)^{3/2}]}{(Es_N^2)^{3/2}} + 0.945\frac{E[\delta_N s_N^2]}{Es_N^2} + 100E\delta_N^2 + \frac{E|s_N^2 - Es_N^2|}{Es_N^2}$$

where $\delta_N = \dfrac{\beta_N}{s_N^2\sqrt{Es_N^2}}$.

Proof. For each $n \in \mathbb{N}$ and $i = 1, 2, \dots, n$, let

$$s_n^2 = \sum_{i=1}^n \sigma_i^2, \quad \beta_n = \sum_{i=1}^n \gamma_i, \quad \delta_n = \frac{\beta_n}{s_n^2\sqrt{Es_N^2}},$$
$$Y_i = \frac{X_i}{\sqrt{Es_N^2}}, \quad W_n = Y_1 + Y_2 + \cdots + Y_n, \quad W_n^{(i)} = W_n - Y_i,$$
$$M_i(t) = E[Y_i(I(0 \le t \le Y_i) - I(Y_i \le t < 0))] \text{ and } p_n = P(N = n).$$

Follows the argument in [10], p.726 and p.728, we obtain

$$\begin{aligned} &P(W \le x) - \Phi(x) \\ &= \sum_{n=1}^\infty p_n\Big(\sum_{i=1}^n E\int_{-\infty}^\infty f'(W_n)M_i(t)dt - \sum_{i=1}^n E\int_{-\infty}^\infty f'(W_n^{(i)} + t)M_i(t)dt\Big) \\ &\quad + \sum_{n=1}^\infty p_n Ef'(W_n)\Big(1 - \sum_{i=1}^n \int_{-\infty}^\infty M_i(t)dt\Big) \\ &:= A_1 + A_2 \end{aligned} \tag{1}$$

where

$$A_1 \leq \sum_{n=1}^{\infty} p_n \sum_{i=1}^{n} E \int_{-\infty}^{\infty} P(x-t < W_n^{(i)} \leq x - Y_i \mid Y_i) M_i(t) dt$$
$$+ \frac{3}{2Es_N^2} \sum_{n=1}^{\infty} p_n \delta_n s_n^2 \Big(\sqrt{\frac{s_n^2}{Es_N^2}} + 0.63\Big)$$

and

$$|A_2| = \sum_{n=1}^{\infty} p_n \Big| Ef'(W_n)\Big(1 - \sum_{i=1}^{n} EY_i^2\Big)\Big| \leq \sum_{n=1}^{\infty} p_n \Big|1 - \frac{s_n^2}{Es_N^2}\Big| = \frac{E|s_N^2 - Es_N^2|}{Es_N^2}. \quad (2)$$

Next, we will give a bound of

$$P(a \leq W_n^{(i)} \leq b) \text{ for all } a < b.$$

For a fix number n and for $i = 1, 2, \ldots, n$ and for all $a < b$, define $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(t) = \begin{cases} -\frac{1}{2}(b-a) - \delta_n & \text{for } t < a - \delta_n, \\ t - \frac{1}{2}(b+a) & \text{for } a - \delta_n \leq t \leq b + \delta_n, \\ \frac{1}{2}(b-a) + \delta_n & \text{for } t \geq b + \delta_n \end{cases}$$

and using the argument of uniform concentration inequality in [10], p.725, we obtain

$$P(a \leq W_n^{(i)} \leq b) \leq \frac{2.5Es_N^2}{s_n^2}\Big\{\Big(\frac{1}{2}(b-a) + 2\delta_n\Big)\sqrt{\frac{s_n^2}{Es_N^2}} + 40\delta_n^2\Big\}$$

which implies that

$$|A_1| \leq 6.875\frac{E[\delta_N\sqrt{s_N^2}]}{\sqrt{Es_N^2}} + 100E\delta_N^2 + 1.5\frac{E[\delta_N(s_N^2)^{3/2}]}{(Es_N^2)^{3/2}} + 0.945\frac{E[\delta_N s_N^2]}{Es_N^2}. \quad (3)$$

Combining (1), (2) and (3), we have the theorem as desired. □

Corollary 4. *If $X_1, X_2, \ldots$ are independent identically distributed random variables, then we have*

$$\sup_{x\in\mathbb{R}} |P(W \leq x) - \Phi(x)| \leq 7.82\delta + \Big(\frac{100\gamma^2}{\sqrt{EN}\sigma^6}\Big)\frac{1}{\sqrt{EN}} + \frac{1.5\delta EN^{3/2}}{(EN)^{3/2}} + \frac{E|N - EN|}{EN}$$

where $\delta = \dfrac{\gamma}{\sqrt{EN}\sigma^3}$.

Observe that in the case of $EN \to \infty$, the bound in Corollary 4 is better than the bound in Theorem 2.

3 Non-Uniform Bound

In this section, we give a non-uniform bound for random sums W. We also use the notation as in the section 2.

In 2008, N. Chaidee and K. Neammanee [10] give a non-uniform bound in normal approximation of W. The following is their result.

Theorem 5. ([10]) *There exists a constant C such that for every real number x,*

$$|P(W \leq x) - \Phi(x)| \leq \frac{C\delta}{(1+|x|)^3}\Big(1 + \frac{EN^5}{(EN)^5}\Big) + \frac{CE|N^2 - (EN)^2|}{(1+|x|)^2(EN)^2}$$

where $\delta = \dfrac{\gamma}{\sqrt{EN}\sigma^3}$.

In this section, we establish a non-uniform bound for W in the case of $X_1, X_2, \ldots$ are not necessarily identically distributed random variables.

Theorem 6. *If $s_n^2 > 4 \max\limits_{1\leq i\leq n} \sigma_i^2$ for all positive integer $n \geq 5$, then there exists a constant C such that for every real number x,*

$$\begin{aligned}
&|P(W \leq x) - \Phi(x)| \\
&\leq \frac{C}{(1+|x|)^3}\Big(E\delta_N + \frac{E[\delta_N (s_N^2)^5]}{(Es_N^2)^5} + \frac{E[\delta_N (s_N^2)^2]}{(Es_N^2)^2} + \frac{E[\delta_N \sqrt{s_N^2}]}{\sqrt{Es_N^2}} + \frac{E[\delta_N (s_N^2)^{3/2}]}{(Es_N^2)^{3/2}}\Big) \\
&\quad + \frac{C}{(1+|x|)^2}\Big(E\delta_N^2 + \frac{E[\delta_N^2 (s_N^2)^5]}{(Es_N^2)^5} + \frac{E|(s_N^2)^2 - (Es_N^2)^2|}{(Es_N^2)^2}\Big)
\end{aligned}$$

where $\delta_N = \dfrac{\beta_N}{s_N^2\sqrt{Es_N^2}}$.

Proof. In the view of Theorem 3, we may without loss of generality, assume $x \geq 4$. Follow the argument in [10], p.735-736, we have

$$\begin{aligned}
&|P(W \leq x) - \Phi(x)| \\
&\leq \Big|\sum_{n=1}^{\infty} p_n \sum_{i=1}^{n} E\Big(I(|Y_i| \leq 1 + \frac{x}{4}) \int_{-\infty}^{\infty} (f'(W_n^{(i)} + Y_i) - f'(W_n^{(i)} + t))M_i(t)dt\Big)\Big| \\
&\quad + \Big|\sum_{n=1}^{\infty} p_n \sum_{i=1}^{n} E\Big(I(|Y_i| > 1 + \frac{x}{4}) \int_{-\infty}^{\infty} (f'(W_n^{(i)} + Y_i) - f'(W_n^{(i)} + t))M_i(t)dt\Big)\Big| \\
&\quad + \Big|\sum_{n=1}^{\infty} p_n \Big(Ef'(W_n) - \sum_{i=1}^{n} E\int_{-\infty}^{\infty} f'(W_n)M_i(t)dt\Big)\Big| \\
&:= |A_1| + |A_2| + |A_3|,
\end{aligned}$$

$$|A_1| \le \Big| \sum_{n=1}^{\infty} p_n \sum_{i=1}^{n} E\Big(I(|Y_i| \le 1 + \frac{x}{4}) \int_{-\infty}^{\infty} M_i(t) \int_{t}^{Y_i} Eg(W_n^{(i)} + u) du dt \Big) \Big|$$
$$+ \sum_{n=1}^{\infty} p_n \sum_{i=1}^{n} E\Big(I(|Y_i| \le 1 + \frac{x}{4})$$
$$\times \int_{-\infty}^{\infty} P(x - \max(Y_i, t) < W_n^{(t)} \le x - \min(Y_i, t)|Y_i) M_i(t) dt \Big)$$
$$:= |A_{11}| + |A_{12}|,$$

and

$$|A_{11}| \le \frac{C}{(1+x)^3} \Big(E\delta_N + \frac{E[\delta_N (s_N^2)^5]}{(Es_N^2)^5} \Big) + \frac{C}{(1+x)^2} \Big(E\delta_N^2 + \frac{E[\delta_N^2 (s_N^2)^5]}{(Es_N^2)^5} \Big),$$
$$|A_2| \le \frac{CE[\delta_N (s_N^2)^2]}{(1+x^3)(Es_N^2)^2},$$
$$|A_3| \le \frac{CE|(s_N^2)^2 - (Es_N^2)^2|}{(1+x)^2 (Es_N^2)^2}$$

where $g(w) = (wf(w))'$.

To bound A_{12}, we use the non-uniform concentration inequality : for a positive integer $n \ge 5$ and $0 \le a < b$. If $s_n^2 > 4 \max\limits_{1 \le i \le n} \sigma_i^2$, then there exists a constant C such that

$$P(a \le W_n^{(i)} \le b) \le \frac{C(b - a + \delta_n)}{(1+a)^3} \Big(\frac{Es_N^2}{s_n^2} + \frac{(s_n^2)^4}{(Es_N^2)^4} \Big). \tag{4}$$

To prove the above inequality, for each $n \in \mathbb{N}$, we define

$$f_n(x) = \begin{cases} 0 & \text{for } \ x < a + r - \kappa_n, \\ (1 + x - r + \kappa_n)^3 (x - a - r + \kappa_n) & \text{for } \ a + r - \kappa_n \le x \le b + r + \kappa_n, \\ (1 + x - r + \kappa_n)^3 (b - a + 2\kappa_n) & \text{for } \ x > b + r + \kappa_n \end{cases}$$

where $\kappa_n = 4\delta_n$.

Let

$$Z_j = Y_j I(|Y_j| \le 1 + a) - EY_j I(|Y_j| \le 1 + a), \eta_j = |Z_j| \min(\kappa_n, |Z_j|) \ \text{ and } \ U = \sum_{\substack{j=1 \\ j \ne i}}^{n} \eta_j.$$

By the same the argument in [10], p.729-733, we have

$$EU \ge \frac{s_n^2}{4Es_N^2} + \frac{(s_n^2 - 4\sigma_i)}{4Es_N^2} \ge \frac{s_n^2}{4Es_N^2}$$

where we have use the fact that $s_n^2 > 4 \max_{1\le i\le n} \sigma_i^2$ in the last inequality which yields the concentration inequality (4).

Applying the concentration inequality, we obtain

$$|A_{12}| \le \frac{C}{(1+x)^3}\Big(E\delta_N + \frac{E[\delta_N (s_N^2)^5]}{(Es_N^2)^5}\Big).$$

This is complete the proof. □

Corollary 7. *Assume that* $s_n^2 > 4 \max_{1\le i\le n} \sigma_i^2$ *for all positive integer* $n \ge 5$. *If* $X_1, X_2, \ldots$ *are independent identically distributed random variables, then there exists a constant* C *such that for every real number* x,

$$\begin{aligned}|P(W \le x) - \Phi(x)| \le & \frac{C\delta}{(1+|x|)^3}\Big(1 + \frac{EN^{3/2}}{(EN)^{3/2}} + \frac{EN^2}{(EN)^2} + \frac{EN^5}{(EN)^5}\Big) \\ & + \frac{C\delta^2}{(1+|x|)^2}\Big(1 + \frac{EN^5}{(EN)^5}\Big) + \frac{C\delta}{(1+|x|)^2}\frac{E|N^2 - (EN)^2|}{(EN)^2}\end{aligned}$$

where $\delta = \dfrac{\gamma}{\sqrt{EN}\sigma^3}$.

Acknowledgements. We are greatly indepted to K.Neammanee for his generous comments and suggestions which have helped improving our work.

References

[1] A.C.Berry, The accuracy of Gaussian apporximation to the sum of independent variables. Trans. Amer. Math Soc., **49**(1941), 122-136.

[2] A.D.Barbour and L.H.Y.Chen, An introduction to Stein's method, Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore, **4**(2005), Singapore university press and world scientific publishing, Singapore.

[3] A.Krajka and Z.Rychilk, Necessary and sufficint conditions for weak convergence of random sums of independent random variables. Comment. Math. Univ. Carolin., **34**(1993), 465-482.

[4] A.Nakas, Estimate of the remainder term in the central limit theorem with a random number of random summands. Liet. Mat. Rink., **12**(1972), 157-163.

[5] C.G.Esseen, Fourier analysis of distributionfunctions: a mathematical study of the Laplace-Gaussian law. Acta math, **77**(1945), 11-25.

[6] D.Landers and L.Rogge, The exact approximation order in the central limit theorem for random summation. Z. Wahrsch. Vers. Gebiete, **36**(1976), 269-283.

[7] I.S.Shiganov, A refinement of the upper bound of the constant in the remainder term of the centrtal limit theorem. J. Soviet Math., **3**(1986), 171-182.

[8] Kh.Batirov, D.V.Manevich and S. V. Nagaev, The Esseen inequality for sum of a random number of differently distributed random variables. Matematicheskie Zametki, **22**(1980), 143-146.

[9] L.H.Y.Chen and Q.M.Shao, A non-uniform Berry-Esseen bound via Stein's method. Probab. Theory Related Fields, **120**(2001), 236-254.

[10] N.Chaidee and K.Neammanee, Berry-Esseen bound for independent random sum via Stein's method. International Mathematical Forum, **15**(2008), 721-738.

[11] P.Van Beeck, An application of Fourier methods to the the problem of sharpening the Berry-Esseen inequality. Z. Wahrsdh. Verw. Gebiete, 23(1972), 187-196.

[12] P.Vellaisamy and B.Chaudhuri, Poisson and compound Poisson approximations for random sums of random variables. J. Applied Prob., **33**(1996), 127-137.

[13] W.Feller, On the Berry-Esseen theorem. Z. Wahrsch. Verw. Gebite, **10**(1986), 261-268.

Received: September, 2008