Tr. J. of Mathematics
22 (1998) , 53 – 60.

THE NON UNIFORM BOUNDS OF REMAINDER TERM IN CLT FOR THE SUM OF FUNCTIONS OF UNIFORM SPACINGS*

S. Mirakhmedov & U. Kalandarov

Abstract

The non uniform bound of the remainder in the central limit theorem for the sums of functions of uniform spacings is established. The bound depend on the moments of functions of the standard exponential random variables.

1. Introduction

Let $U_1, U_2, \ldots$ be a sequence of independent uniform (0, 1) random variables (r.v.), $0 \equiv U_{0,n} \leq U_{1,n} \leq \cdots \leq U_{n,n-1} \leq U_{n,n} \equiv 1$ denote the ordering of $U_1, \ldots, U_{n-1}$ and $D_{i,n} = U_{i,n} - U_{i-1,n}, \quad i = 1, \ldots, n$, be their spacings, $D = (D_{1n}, \ldots, D_{n,n})$. Let $f_{mn}(y), \quad m = 0, 1, \ldots, n$, be measurable functions of non-negative argument y.

We consider the statistics that are defined as follows

$$R_n(D) = \sum_{m=1}^{n} f_{mn}(nD_{m,n}). \tag{1}$$

The statistics of (1.1) types are called a divisible statistic (DS). A DS is said to be symmetric if the functions f_{mn} are same for all $m = 1, 2, \ldots, n$. Statistics of this form are used in different statistical tasks, for example, for testing uniformity. Random variables of type (1.1) are also used in the problem covers. An excellent survey of first order limit theory for statistics of the form (1.1) were given by Pyke (1965, 1972) according to whom a study of the rate of convergence for sums of functions of uniform spacings is of interest.

The main idea of the proof of our results is following a well-known characterization that has been applied by Le Cam (1958) in order to prove first order limit theorems (see also Pyke (1965), and [3, 4, 5]).

* Research supported by TUBITAK, TURKEY. We are grateful to Hacettepe University, whose hospitality we enjoyed while working on part of this paper.

Let $Y_1, Y_2, \ldots$ be independent exponential r.v.s. with parameter 1 and let

$$Y = (Y_1, \ldots, Y_n), \quad S_n = \sum_{m=1}^{n} (Y_m - 1), \quad R_n(Y) = \sum_{m=1}^{n} f_{mn}(Y_m).$$

Then

$$\mathcal{L}(R_n(D)) = \mathcal{L}(R_n(Y)/S_n = 0), \tag{2}$$

i.e., $R_n(D)$ has the same distribution as a sum of independent a special r.v.s. given another sum of independent r.v.s.

The asymptotically normality and Berry-Esseen bounds for statistics of type (1.1) have been studied by several authors. But the optimal condition $Ef^2(Y_1) < \infty$ for asymptotical normality of symmetric DS was obtained of Beirlant, Janssen and Veraverbeke in 1991 year. The proof in this paper based on another well-known property of uniform spacings: the joint distribution of D is the same as the joint distribution of $(\overline{S}_0 \cdot \overline{S}_n^{-1}, \ldots, \overline{S}_{n-1} \cdot \bar{S}_n^{-1} 1)$, where $\overline{S}_n = S_n + n$ and on a Taylor expansion idea for statistics of type (1.1). Using the representation (1.2) Does and Klaassen (1984) obtained estimates for the rate of convergence in CLT of type $O(n^{-1/2})$. The Lindeberg type condition and uniformly bounds (by n and a argument of distribution function) of remainder in the CLT (Mirakhmedov (1996)), Edgwort type asymptotical expansion (Does, Helmers, Klaassen (1985)) and probability of large deviations (Mirakhmedov (1996)) for statistics of type (1.1) were obtained by using equation (1.2). In this article we will obtain the non uniform bound of the remainder in the CLT. We remark that the functions $f_{mn}(y), \ m = 1, \ldots, n$ may be randomly. In this case we suppose that the $f_{ln}(x_1), \ldots, f_{nn}(x_n)$ are independent r.v.s. not depending on the Y and D for arbitrary set of non negative numbers $x_1, \ldots, x_n$.

In what follows C, C_i are a positive universal constants not depending on the n and distributions of the r.v. $f_{mn}(Y_m), \quad m = 1, \ldots, n, \epsilon$ is a sufficiently small positive value.

2. Results

We suppose that the moments which we use below exists. Let

$$\begin{aligned}
\rho_n &= corr(R_n(Y), S_n); \\
X_m(u) &= f_{mn}(u) - Ef_{mn}(Y_m) - \rho_n \sqrt{DR_n(Y)/n}(u-1); \\
T_n(D) &= \sum_{m=1}^{n} X_m(nD_{mn}), \ldots, T_n(Y) = \sum_{m=1}^{n} X_m(Y_m).
\end{aligned}$$

Note that $\sigma_n^2 \equiv DT_n(Y) = (1 - \rho_n^2) DR_n(Y)$ and

$$ET_n(Y) = 0, \ldots, cov(T_n(Y), S_n) = 0. \tag{3}$$

Obviously that $T_n(D) = R_n(D) - ER_n(Y)$. Therefore without loss generality we can consider the statistic $T_n(D)$ instead of $R_n(D)$. From definition of σ_n^2 it follows that $\sigma_n^2 = 0$ if and only if $f_{mn}(y) = Cy + a_m$, where constants a_m are arbitrary and C does not depend on m for all $m = 1, \ldots, n$. In what follows we suppose $\sigma_n^2 > 0$ for all $n = 1, 2, \ldots$.

Let

$$\overline{X}_n = X_m(Y_m)/\sigma_n \cdots \overline{Y}_n = (Y_m - 1)/\sqrt{n}, \quad \beta_{kn} = \sum_{m=1}^{n} E|\overline{X}_m|^k,$$

$\Phi(x)$ be the standard normal distribution function and $P_n(x) = P\{T_n(D) < x\sigma_n\}$.

Theorem. *There is a constant $C > 0$ such that for arbitrary integer $s > 2$ and $n > s+1$*

$$\Delta_n(x) = |P_n(x) - \Phi(x)| \le \frac{C(s)}{1 + |x|^{s-2}} \left(\beta_{3n} + \beta_{sn} + \frac{1}{\sqrt{n}} \right).$$

In Section 3 of Pyke (1965) some examples of DS are given. In particular, several functions for symmetrical DS are related to $g(x) = x^r$, $r > 0$, $r \neq 1, \ldots g_2(x) = (x-1)^2, \ldots g_3(x) = |x-1|$, $g_4(x) = \log x$ and are all included in theorem. For these functions from the theorem it follows the estimate: for arbitrary fixed s there exist constant $C(s)$ such that

$$\Delta_n(x) \le \frac{C(s)}{\sqrt{n}(1 + |x|^s)}.$$

But for another examples $g_5(x) = x^{-1}$ (Pyke (1965)) our theorem is useless.

3. Proof

Let $\varphi_n(t)$ be the characteristic function of the random variables (r.v.) $T_n(D)/\sigma_n$. By Corollary 11.5 and Lemma 11.6 of Bhattacharya, Rao (1976) one has for arbitrary $T > 0$

$$(1 + |x|^{s-2})\Delta_n(x) \le C_0 \max_{0 \le k \le s} \int_{|t| \le T} \left| D_t^k (\varphi_n(t) - \exp\left\{-\frac{t^2}{2}\right\}) \right| dt + \frac{C_1}{T} \tag{4}$$

where D_t^k denote k-th derivation. It is well known (see, for example, Pyke (1965)) that we can choose a regular version of the conditional distribution of $T_n(Y)$ given $\overline{S}_n \equiv n^{-1/2} S_n = x$ such that

$$\varphi_n(t) = E(\exp\{it\sigma_n^{-1} T_n(Y)\}/\overline{S}_n = 0). \tag{5}$$

Let $p_n(x)$ be a density of the r.v. $\overline{S}_n$, and

$$\Psi_{mn}(t, \tau) = E \exp\{it\overline{X}_m + i\tau\overline{Y}_m\}, \quad \Psi_n(t, \tau) = \prod_{m=1}^{n} \Psi_{mn}(t, \tau).$$

Lemma 1. *We have*

$$\varphi_n(t) = \frac{1}{2\pi p_n(0)_\infty} \int_{-\infty}^{\infty} \Psi_n(t,\tau) d\tau.$$

Proof. With the aim of Plancherel's identity (cf. Theorem 4.1 of Bhattacharya and R. Ranga Rao (1976)) we check that for all t (see also, Does and Klaassen (1984))

$$\int_{-\infty}^{\infty} |\Psi_{in}(t,\tau)|^2 \, d\tau = 2\pi\sqrt{n} \int_0^{\infty} \exp(-2u) du = \pi\sqrt{n}.$$

From this and Holder's inequality it follows that

$$\int_{-\infty}^{\infty} |\Psi_{im}(t,\tau)\Psi_{km}(t,\tau)| d\tau \le \pi\sqrt{n}, \quad 1 \le i, \;\; k \le n \tag{6}$$

and hence

$$\int_{-\infty}^{\infty} |\Psi_n(t,\tau)| d\tau \le \pi\sqrt{n}. \tag{7}$$

In view of (3.3) and (3.5), Fourier inversion of

$$\Psi_n(t,\tau) = \int_{-\infty}^{\infty} \exp\{i\tau x\} p_n(x) E(\exp\{it\sigma_n^{-1} T_n(Y)\}/\overline{S}_n = x) dx$$

yields Lemma 1. Put $K_n = C_2 \min(\sqrt{n}, \beta_{3n}^{-1}), \ldots C_2 = 1 - 6/\sqrt{37}$,

$$\begin{aligned}
A(t,\tau) &= \{(t,\tau) : |t| \le K_n, \;\; \tau \in (-\infty, \infty)\} \\
A_1(t,\tau) &= \{(t,\tau) : |t| \le C_4 \beta_{sn}^{-1/s}, \;\; |\tau \;\; \le C_4 n^{(s-2)/2}\} \\
A_2(t,\tau) &= \left\{(t,\tau) : |t| \le K_n, \;\; |\tau| \le \frac{1}{6}\sqrt{n}\right\} \\
A_3(t,\tau) &= \left\{(t,\tau) : |t| \le K_n, \;\; |\tau| > \frac{1}{6}\sqrt{n}\right\},
\end{aligned}$$

where $0 < C_4 < 1/6$. From Lemma 1 we have

$$\begin{aligned}
J_k &\equiv \int_{|t| \le K_n(\epsilon)} |D_t^k(\varphi_n(t) - \exp\{-t^2/2\})| dt \\
&\le \frac{1}{2\pi p_n(0)} \left[\int\int_{A_1(t,\tau)} |D_t^k(\Psi_n(t,\tau) - \exp\{-(t^2+\tau^2)/2\})| dt d\tau \right. \\
&+ \int\int_{A_2(t,\tau) - A_1(t,\tau)} |D_t^k \Psi_n(t,\tau)| dt d\tau + \int\int_{A_3(t,\tau)} |D_t^k \Psi_n(t,\tau)| dt d\tau \\
&+ \left. \int\int_{A(t,\tau) - A_1(t,\tau)} |t|^k \exp\{-(t^2+\tau^2)/2\} dt d\tau \right] \\
&+ \frac{1}{\sqrt{2\pi}} \left| \frac{1}{\sqrt{2\pi} p_n(0)} - 1 \right| \int\int_{A(t,\tau)} |t|^k \exp\{-(t^2+\tau^2)/2\} dt d\tau. \qquad (8)
\end{aligned}$$

Let the symbols $J_{1k}, J_{2k}, J_{3k}, J_{4k}$ be summands in the brackets, correspondingly, and J_{5k} be the outside of the summand bracket on the right hand side of (3.5). □

Lemma 2. *1. If $(t,\tau) \in A_1(t,\tau)$ then for each $k:\ 0 \le k \le s$ there is a constant $C_5(s,k)$ such that*

$$|D_t^k(\Psi_n(t,\tau) - \exp\{-(t^2+\tau^2)/2\})| \le C_5(s,k)\left(\beta_{3n} + \beta_{sn} + \frac{1}{\sqrt{n}}\right)$$
$$\left(1 + |t|^{3(s-2)+k} + |t|^{3(s-2)4k}\right)\exp\{-(t^2+\tau^2)/4\}.$$

2. If $(t,\tau) \in A_2(t,\tau)$, then for $k = 0,1$

$$|D_t^k\Psi_n(t,\tau)| \le (t^2+\tau^2)^k \exp\{-(t^2+\tau^2)/4\}.$$

Proof. Let $P_m(t,\tau),\quad r = 1,2,\ldots$ be a well-known polinomials on the theory of asymptotical expansion of a characteristic functions of a sum of independent r.v.s. (see, Bhattachareya, Rao (1976), the functions $\tilde{P}_r(iBt, \{\chi_\nu\})$, p.52, 82).

From Theorem 9.11 (Bhattachareya, Rao (1976)) and properties (2.1) it follows that there is constant $C_6(s,k)$ such that for each $k: 0 \le k \le s, \ldots$ and $(t,\tau) \in A_1(t,\tau)$ the inequality

$$\left| D_t^k(\Psi_n(t,\tau) - \exp\left\{-t^2+\tau^2/2\right\}\left(1 + \sum_{r=1}^{s-3}\frac{P_r(t,\tau)}{n^{r/2}}\right)\right|$$

$$\le C_6(s,k)\left(\beta_{sn} + n^{-(s-2)/2}\right)\left(1 + (t^2+\tau^2)^{3(s-2)+k}\right)\exp\{-(t^2+\tau^2)/4\} \qquad (9)$$

is hold true. A same reasoning as in proof Lemma 9.5 (Bhattachareya, Rao (1976, p.71) give us that

$$|n^{-r/2}P_m(t,\tau)| \le C_7(r)(\beta_{r+2,n} + n^{-r/2})(1 + (t^2+\tau^2)^{3r-k}) \qquad (10)$$

The inequalities (3.5) and (3.6) and $\beta_{kn} \le \beta_{3n} + \beta_{sn},\quad 3 \le k \le s$, implies part 1 of Lemma 2.

Put $J_r = (j_1, \ldots, j_r)$ is an r subset of the set $N = (1, \ldots, n),\quad r \ge 0,\quad J_0 = \emptyset$, and (J_r) is collection of all J_r. It is easy to see that

$$|D_t^k\Psi_n(t,\tau)| \le \sum_{r=0}^{k} C(r,k) \sum_{(J_r)} \prod_{i \in N - J_f} |\Psi_{in}(t,\tau)| \prod_{j \in J_f} |D_t^{\gamma_j}\Psi_{jn}(t,\tau)|, \qquad (11)$$

where $\gamma_{j_1}, \dots, \gamma_{j_r}$ are non negative integers such that $\gamma_{j_1} + \cdots + \gamma_{j_f} = k$. We have

$$\begin{aligned} |\Psi_{mn}(t,\tau)|^2 &\le 1 - E(t\overline{X}_m + \tau\overline{Y}_m)^2 + \frac{2}{3}E|t\overline{X}_m + \tau\overline{Y}_m|^3 \\ &\le \exp\{-E(t\overline{X}_m + \tau\overline{Y}_m)^2 + \frac{2}{3}E\ |t\overline{X}_m + \tau\overline{Y}_m|^3\}. \end{aligned}$$

Thereto, using inequality between moments of r.v.s, we get

$$\exp\left\{E(t\overline{X}_m + \tau\overline{Y}_m)^2 - \frac{2}{3}E\ |t\overline{X}_m + \tau\overline{Y}_m|^3\right\} \le e^{1/3}$$

since $\max_{y\ge 0}(y^2 - \frac{2}{3}y^3) \le \frac{1}{3}$. Hence, recollecting (2.2) and that $|a+b|^3 \le 4(|a|^3 + |b|^3)$, we obtain

$$\prod_{i\in N-J_f} |\Psi_{in}(t,\tau)| \le e^{r/3}\exp\{-(t^2+\tau^2)/4\}. \qquad (12)$$

Obviously

$$|D_t\Psi_{mn}(t,\tau)| \le (|t| + |\tau|)E\overline{X}_m^2 + |\tau|E\overline{Y}_m^2, \qquad (13)$$

and

$$|D_t^k\Psi_{mn}(t,\tau)| \le E|\overline{X}_m|^k \le E\overline{X}_m^2 + E|\overline{X}_m|^s.$$

Putting $d_m = \max\{(|t| + |\tau|)E\overline{X}_m^2 + |\tau|E\overline{Y}_m^2, \dots E\overline{X}_m^2 + E|\overline{X}_m|^s\}$ we get

$$\sum_{(J_r)}\prod_{j\in J_f}|D_t^{\gamma_j}\Psi_{jn}(t,\tau)| \le \left(\sum_{m=1}^n d_m\right)^r \le C(s)\max(1, |t| + |\tau|)^r. \qquad (14)$$

The second part of Lemma 2 follows from (3.8), (3.9), (3.11).

Using Lemma 2 we find that

$$|J_{1k} + J_{2k}| \le C(s)(\beta_{sn} + n^{-(s-2)/2}). \qquad (15)$$

With the aid of the inequality $x < \exp(x-1)$ we have for any $m = 1, \dots, n$ and $(t,\tau) \in A_3(t,\tau)$

$$\begin{aligned} |\Psi_{mn}(t,\tau)| &= |E\exp\{i\tau\overline{Y}_m\}(\exp\{it\overline{X}_m\} - 1)| \le |E\exp\{it\overline{Y}_m\}| \\ &\quad + |t|E|\overline{X}_m| \le \exp\{-(1 - |E\exp\{i\tau\overline{Y}_m\}|) + |t|E|\overline{X}_m|\} \\ &\le \exp\{-(1 - (1+\tau^2/n)^{-1/2}) + |t|E|\overline{X}_m|\} \qquad (16) \\ &\le \exp\{-2C_0 + |t|E|\overline{X}_m|\} \qquad (17) \end{aligned}$$

because $|E\exp\{i\tau\overline{Y}_m\}| = (1 + \tau^2/\sqrt{n})^{-1/2}$ and $|\tau| > \sqrt{n}/6$.

Using (3.10), and that $|D_t\Psi_{mn}(t,\tau)| \le E|\overline{X}_m|$ we get

$$\sum_{(J_r)}\left(\prod_{j\in J_f}|D_t^{\beta_j}\Psi_{jn}(t,\tau)|\right) \le \left(\sum_{m=1}^n\left(E|\overline{X}_m| + E\overline{X}_m^2 + E|\overline{X}_m|^s\right)\right)^r \le C(r)n^{r/2}. \qquad (18)$$

From (3.8), (3.13), (3.14) and (3.3) we have

$$J_{3k} \leq C(k)n^{(r+1)/2}\exp\{-C_0 n\} \tag{19}$$

because $n > s+1$, and $|t| \leq K_n \leq C_0\sqrt{n}$.

For J_{4k} the obviously estimate

$$|J_{4k}| \leq C(\beta_{sn} + n^{(s-2)/2}) \tag{20}$$

is true.

Since $p_n(0) = n^{n-\frac{1}{2}}(n!)^{-1}\exp(-n)$ then with the aid Stirlings formula we obtain

$$\left|\frac{1}{\sqrt{2\pi}p_n(0)} - 1\right| \leq \frac{C}{n}. \tag{21}$$

Hence the bound

$$|J_{5k}| \leq \frac{C}{n} \tag{22}$$

is true.

In the (3.4) we replace summands by its bounds from (3.7), (3.12), (3.15), (3.16), (3.18) and for factor $p_n(0)$ using the (3.17). Then we find

$$|J_k| \leq C\left(\beta_{3n} + \beta_{sn} + \frac{1}{\sqrt{n}}\right). \tag{23}$$

Putting in (3.1) $T = K_n$ and using (3.19) we complete the proof of the theorem. □

References

[1] Pyke, R., Spacings, *J. Roy. Statist. Soc. ser.* B **27**, 395-449 (1965).

[2] Pyke, R., Spacings revisited, Proc. Sixth Berkeley Sympos. Math. Statist. Probability, **1**, 417-427 (1972).

[3] Does, R.J.M.M., Klaassen, C.A.J., The Berry-Esseen theorem for the functions of uniform spacings, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **65**, 461-471 (1984).

[4] Does, R.J.M.M., Klaassen, C.A.L., Second order asymptotics for statistics based on uniform spacings, *Asymptotic Statysics.* **2**, Madl. P. and Huskova M. eds., 231-239, Amsterdam, North Holland (1984).

[5] Beirlant, J., Janssen, P., Vereverbeke, M., On the asymptotic normality of functions of uniform spacings, *Canadian J. Statist.*, **19**, 93-101 (1991).

[6] Mirakhmedov, Sh.A., The limit theorems for the conditional distributions of sums of independent random variables, *Discrete Math. Appl.*, **1**, No.4, 519-542 (1995).

[7] Bhattacharya, P.H., and R. Ranga Rao, Normal Approximation and Asymptotic Expansions, New York, Wiley (1976).

[8] Le Cam, L. Un theoreme sur la division d'un interivable par des points pris an hasard. *Publ. Inst. Statist. Univ.*, Paris **7**, 7-16 (1958).

[9] Feller, W., An introduction to probability theory and its applications, **11**, New York, London, Sydney (1966).

[10] Does, R.J.M.M., Helmers, R., Klaassen, C.A.J., On Edgeworth expansion for the sum of a function of uniform spacings, *Journal of Statistics Planning and Inference*, **7**, 149-157 (1987).

[11] Mirakhmedov, Sh.A., Probability of large deviations for the sum a functions of spacings, *Uzb. Fanlar Acad.*, **4**, 8-11 (1997).

[12] Mirakhmedov, Sh.A., The bounds of the remainder term in CLT for the sum of functions of uniform spacings, *Theory Probabl. Appl.*, (to appear).

Sherzod MIRAKHMEDOV & Utkur KALANDAROV
Department of Mathematics
Tashkent State University
700095 Tashkent - UZBEKISTAN

Received 13.05.1996