The non uniform bounds of remainder term in CLT for the sum of functions of k-spacings

Lazareva Viktoriya Andreyevna
V.I. Romanovski Institute of Mathematics, Academy of Sciences of Uzbekistan
1748mailbox@mail.ru

Abstract. The paper establishes a non-uniform bound on the remainder term in the central limit theorem for sum of functions of disjoint uniform k-spacings, where the step size k may increase together with the sample size.

MSC (2020): 60F05; 60E05; 62G30.

Key words: Uniform spacings, central limit theorem, non-uniform bound, exponential distribution.

Introduction

Let $U_1, U_2, \dots, U_{n-1}$ be an ordered sample of size $n-1$ from a uniform $[0,1]$ distribution, $U_0 = 0$, $U_n = 1$, $G_i^{(k)} = U_{ik} - U_{(i-1)k}$, $i = 1, 2, \dots, N'$, $G_{N+1}^{(k)} = 1 - U_{N'k}$ their k-spacings, where $N' = \lfloor n/k \rfloor$ is the integer part of n/k. Let $N = N'$ if n/k is an integer and $N = N' + 1$ otherwise, and let $f_m(u)$, $m = 1, 2, \dots, N$ be a set of measurable functions. We consider the statistics of the type

$$R_n(G) = \sum_{m=1}^{N} f_m\left(nG_m^{(k)}\right), \tag{1}$$

where $k = k(n)$ may increase to infinity jointly with n.

Obviously, all these random variables (r.v.s) depend on the sample size n. However, we omit here and below the corresponding suffix for notational simplicity.

Statistics of the type (1) are of great interest in several contexts, including hypothesis testing and reliability, circular data analysis where they play a pivotal role because they provide a maximal invariant under the rotation group, and spacings-based parameter estimation, just to name a few applications. There is a huge literature devoted to such statistics and their use; see [5], [3] and references therein. Given that these spacings are highly dependent random quantities with a Dirichlet distribution in finite samples, large-sample theory is the main avenue for studying such statistics.

Statistics of the form $R_n(G)$, which represent sums of functions of disjoint uniform k-spacings, are also employed in applied tasks such as testing for dispersive ordering and addressing problems related to random coverage of the circle. This is particularly significant when k increases with the sample size n, as it enables the construction of goodness-of-fit tests based on statistics of this type to verify hypotheses about the form of the distribution, discriminating alternatives at distances $(nk)^{-1/4}$ that can be arbitrarily close to $n^{-1/2}$ (see [3]).

In [2] a bound of the Berry-Esseen type was obtained, whereas in [6] a non-uniform bound on the remainder term is established in the case $k = 1$.

In this work we obtain a non-uniform estimate that holds for arbitrary integer $k = k(n)$, including the case when k increases together with the sample size n. Thus our result extends results of [2] and [6].

Most common and well-known examples of spacings tests of the form (1) are the so-called Greenwood statistic $G_N^2 = \sum_{m=1}^{N}(nG_m^{(k)} - k)^2$ and the Log-spacings statistic $M_N = \sum_{m=1}^{N} \log(nG_m^{(k)})$.

The main results of the present paper are given in Results. The proof of the assertions of Results is presented in Proofs.

In what follows, C, C_i are positive constants and D_t^k is the k-th order derivative.

Results

We suppose that the moments used below exist. Set $G = (G_1, \dots, G_N)$, where $G_i = G_i^{(1)}$, $i = 1, 2, \dots, n$, and let $Y_1, Y_2, \dots$ be independent standard exponentially distributed r.v.s, $Y = (Y_1, \dots, Y_n)$, $S_n = Y_1 + \dots + Y_n$. Then $\mathcal{L}(nG) = \mathcal{L}(Y/S_n = n)$, where $\mathcal{L}(X)$ denotes the distribution of the random vector X. Put

$$Z_{m,k} = Y_{(m-1)k+1} + \dots + Y_{mk}$$

where $m = 1, 2, \dots, N'$,

$$Z_{N,k} = Z_{N'+1,k} = Y_{N'k+1} + \dots + Y_n$$

if n/k is not an integer, and $Z_{N,k} = 0$ otherwise; $S_{N,k} = Z_{1,k} + \dots + Z_{N,k}$;

$$R_n(Z) = \sum_{m=1}^{N} f_m(Z_{m,k}); \quad \rho = \operatorname{corr}(R_N(Z), S_{N,k});$$

$$g_m(u) = f_m(u) - Ef_m(Z_{m,k}) - (u-k)\rho\sqrt{\operatorname{Var} R_N(Z)/Nk}$$

$$T_N(G) = \sum_{m=1}^{N} g_m(nG_m^{(k)}), \quad T_N(Z) = \sum_{m=1}^{N} g_m(Z_{m,k}).$$

Note that $\sigma_N^2 = \operatorname{Var} T_N(Z) = (1-\rho^2)\operatorname{Var} R_N(Z)$ and

$$ET_N(Z) = 0, \quad \operatorname{cov}(T_N(Z), S_{N,k}) = 0. \tag{2}$$

From the definition of σ_N^2, it follows that $\sigma_N^2 = 0$ if and only if $f_m(u) = Cu + b_m$, $m = 1, 2, \dots, N$, where constants b_m are arbitrary and C does not depend on m for all $m = 1, \dots, N$. We suppose that $\sigma_N^2 > 0$ for all $N = 1, 2, \dots$.

Put $\bar{g}_m = g_m(Z_{m,k})/\sigma_N$, $\bar{Z}_{m,k} = (Z_{m,k} - k)/\sqrt{Nk}$

$$\beta_{jN} = \sum_{m=1}^{N} E|\bar{g}_m|^j, \quad P_N(x) = P[T_N(G) < x\sigma_N], \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{t^2}{2}\right\} dt. \tag{3}$$

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

$$\Delta_N(x) = |P_N(x) - \Phi(x)| \le \frac{C(s)}{1+|x|^{s-2}} (\beta_{3N} + \beta_{sN}).$$

For the Greenwood statistic $f_m(u) = (u-k)^2$ we have $\sigma_N^2 = 2Nk(k+1)$.

Corollary 1. Applying the theorem for $s = 4$, a positive constant $C > 0$ exists such that, whenever $n > 5$,

$$\left| P\left(\frac{G_N^2 - EG_N^2}{\sigma_N} < x\right) - \Phi(x) \right| \le \frac{C}{1+|x|^2} \left(\frac{15k^3 + 222k^2 + 579k + 372}{Nk(k+1)^2}\right)^{1/2}$$

For the Log-spacings statistic $f_m(u) = \log u$ we have $\sigma_N^2 = N(\zeta(2,k) - k^{-1})$.

Corollary 2. From the theorem with $s = 4$, there is some $C > 0$ ensuring that for $n > 5$,

$$\left| P\left(\frac{M_N - EM_N}{\sigma_N} < x\right) - \Phi(x) \right| \le \frac{C}{1+|x|^2} \left(\frac{3k^{-2} - 2k^{-3} - 6k^{-1}\zeta(2,k) + 3\zeta^2(2,k) + 6\zeta(4,k)}{N(\zeta(2,k) - k^{-1})^2}\right)^{1/2}$$

The assertions in the corollaries can be obtained using the well-known inequality $\beta_{3,N} \le \beta_{4,N}^{1/2}$ and by direct calculation of $\beta_{4,N}$ for the corresponding statistics, which were computed in [3].

Proofs

For simplicity of computations, we suppose that n/k is an integer; the case when n/k is not an integer needs only quite clear additional computations.

Set $\varphi_N(t) = E\exp\{itT_N(G)/\sigma_N\}$. By Corollary 11.5 and Lemma 11.6 of [1] 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. Denote

$$\psi_m(t,\tau) = E\exp\left\{it\bar{g}_m + i\tau\bar{Z}_{m,k}\right\}, \quad \Psi_N(t,\tau) = \prod_{m=1}^{N} \psi_m(t,\tau)$$

By equality (4) of [2] we have

$$\varphi_N(t) = \frac{1}{2\pi\sqrt{n}p_n(n)} \int_{-\infty}^{\infty} \Psi_N(t,\tau)d\tau. \tag{5}$$

Set

$$A(t,\tau) = \{(t,\tau) : |t| \le C_4\beta_{3N}^{-1}, \tau \in (-\infty,\infty)\}$$
$$A_1(t,\tau) = \left\{(t,\tau) : |t| \le \beta_{sN}^{-1/s},\ |\tau| \le N^{(s-2)/2s}\right\},$$
$$A_2(t,\tau) = \left\{(t,\tau) : |t| \le C_4\beta_{3N}^{-1},\ |\tau| \le \frac{1}{6}\sqrt{N}\right\},$$
$$A_3(t,\tau) = \left\{(t,\tau) : |t| \le C_4\beta_{3N}^{-1}, |\tau| > \frac{1}{6}\sqrt{N}\right\},$$

We need to choose C_4 sufficiently small for (16) to hold.

Using representation (5) for $\varphi_N(t)$ and taking into account inequality (4), we have

$$\begin{aligned} J_k &:= \int_{|t|\le\beta_{3N}^{-1}} |D_t^k(\varphi_N(t) - \exp\{-t^2/2\})|\,dt \\ &\le \frac{1}{2\pi p_n(n)}\Bigg[\iint_{A_1(t,\tau)} |D_t^k(\Psi_N(t,\tau) - \exp\{-(t^2+\tau^2)/2\})|\,dt\,d\tau \\ &\quad + \iint_{A_2(t,\tau)-A_1(t,\tau)} |D_t^k\Psi_N(t,\tau)|\,dt\,d\tau + \iint_{A_3(t,\tau)} |D_t^k\Psi_N(t,\tau)|\,dt\,d\tau \\ &\quad + \iint_{A(t,\tau)-A_1(t,\tau)} |t|^k \exp\{-(t^2+\tau^2)/2\}\,dt\,d\tau\Bigg] \\ &+ \frac{1}{\sqrt{2\pi}}\left|\frac{1}{\sqrt{2\pi p_n(n)}} - 1\right| \iint_{A(t,\tau)} |t|^k \exp\{-(t^2+\tau^2)/2\}\,dt\,d\tau. \end{aligned} \tag{6}$$

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 (6).

Lemma 1. 1. If $(t,\tau) \in A_1(t,\tau)$ then for each $k : 0 \le k \le s$ there is a constant $C_6(s,k)$ such that

$$|D_t^k(\Psi_N(t,\tau) - \exp\{-(t^2+\tau^2)/2\})| \le C_6(s,k)\,(\beta_{3N} + \beta_{sN})$$
$$\left(1 + |t|^{3(s-2)+k} + |\tau|^{3(s-2)+k}\right)\exp\{-(t^2+\tau^2)/4\}.$$

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

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

Proof. Let $P_r(t,\tau), r=1,2,\dots$ be a well-known polynomials on the theory of asymptotical expansion of a characteristic functions of a sum of independent r.v.s. (see, [1], the functions $\tilde{P}_r(iBt,\{\chi_\nu\})$, p.52, 82). From Theorem 9.11 ([1]) and properties (2) it follows that there is constant $C_7(s,k)$ such that for each $k: 0\le k\le s,\dots$ and $(t,\tau)\in A_1(t,\tau)$ the inequality

$$\left|D_t^k\left(\Psi_N(t,\tau)-\exp\left\{-\frac{t^2+\tau^2}{2}\right\}\left(1+\sum_{r=1}^{s-3}\frac{P_r(t,\tau)}{N^{r/2}}\right)\right)\right|$$
$$\le C_7(s,k)\left(\beta_{sN}+N^{-(s-2)/2}\right)\left(1+(t^2+\tau^2)^{3(s-2)+k}\right)\exp\{-(t^2+\tau^2)/4\} \qquad (7)$$

holds true. A same reasoning as in proof Lemma 9.5 ([1], p.71) give us that

$$|N^{-r/2}P_r(t,\tau)|\le C_8(r)(\beta_{r+2,N}+N^{-r/2})(1+(t^2+\tau^2)^{3r-k}) \qquad (8)$$

The inequalities (7) and (8) and $\beta_{kN}\le\beta_{3N}+\beta_{sN}, 3\le k\le s$, implies part 1 of Lemma 1.

Put $Q_r=(q_1,\dots,q_r)$ is an r subset of the set $L=(1,\dots,N), r\ge 0, Q_0=\emptyset$, and (Q_r) is collection of all Q_r. It is easy to see that

$$|D_t^k\Psi_N(t,\tau)|\le\sum_{r=0}^{k}C(r,k)\sum_{(Q_r)}\prod_{i\in L-Q_r}|\Psi_{iN}(t,\tau)|\prod_{q\in Q_r}|D_t^{\gamma_q}\Psi_{qN}(t,\tau)|, \qquad (9)$$

where $\gamma_{q_1},\dots,\gamma_{q_r}$ are non negative integers such that $\gamma_{q_1}+\dots+\gamma_{q_r}=k$. We have

$$|\psi_m(t,\tau)|^2\le 1-E(t\bar{g}_m+\tau\bar{Z}_{m,k})^2+\frac{2}{3}E|t\bar{g}_m+\tau\bar{Z}_{m,k}|^3$$
$$\le\exp\{-E(t\bar{g}_m+\tau\bar{Z}_{m,k})^2+\frac{2}{3}E|t\bar{g}_m+\tau\bar{Z}_{m,k}|^3\}. \qquad (10)$$

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

$$\exp\left\{E(t\bar{g}_m+\tau\bar{Z}_{m,k})^2-\frac{2}{3}E|t\bar{g}_m+\tau\bar{Z}_{m,k}|^3\right\}\le e^{1/3}$$

since $\max_{y\ge0}(y^2-\frac{2}{3}y^3)\le\frac{1}{3}$. Hence, recollecting (3) and that $|a+b|^3\le 4(|a|^3+|b|^3)$, we obtain

$$\prod_{i\in L-Q_r}|\psi_i(t,\tau)|\le e^{r/3}\exp\{-(t^2+\tau^2)/4\}. \qquad (11)$$

Obviously

$$|D_t\psi_m(t,\tau)|\le E|\bar{g}_m|, \qquad (12)$$

and

$$|D_t^k\psi_m(t,\tau)|\le E|\bar{g}_m|^k\le E\bar{g}_m^2+E|\bar{g}_m|^s \qquad (13)$$

Putting $d_m=\max\{E|\bar{g}_m|, E\bar{g}_m^2+E|\bar{g}_m|^s\}$ we get

$$\sum_{(Q_r)}\prod_{q\in Q_r}|D_t^{\gamma_q}\psi_q(t,\tau)|\le\left(\sum_{m=1}^{N}d_m\right)^r\le C(s)\max(1,|t|+|\tau|)^r, \qquad (14)$$

The second part of Lemma 1 follows from (9), (11), and (14).

Using Lemma 1 we find that

$$|J_{1k}+J_{2k}|\le C(s)(\beta_{sN}+N^{-(s-2)/2})\le C(s)(\beta_{3N}+\beta_{sN}), \qquad (15)$$

since, from the definition in (3), one has the lower bound $N^{-(s-2)/2}\le\beta_{sN}$.

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

$$\begin{aligned}
|\psi_m(t,\tau)| &= \left|E\exp\{i\tau\bar{Z}_{m,k}\}(\exp\{it\bar{g}_m\}-1)\right| \le \left|E\exp\{i\tau\bar{Z}_{m,k}\}\right| \\
&+ |t|E|\bar{g}_m| \le \exp\left\{-\left(1-\left|E\exp\{i\tau\bar{Z}_{m,k}\}\right|\right)+|t|E|\bar{g}_m|\right\} \\
&\le \exp\left\{-2C_0+|t|E|\bar{g}_m|\right\}
\end{aligned} \tag{16}$$

because $\left|E\exp\{i\tau\bar{Z}_{m,k}\}\right| = \left(1+\tau^2/(N)\right)^{-k/2}$ and $|\tau| > \frac{1}{6}\sqrt{N}$. Using (12), and that $|D_t\psi_m(t,\tau)| \le E|\bar{g}_m|$ we get

$$\sum_{(Q_r)}\left(\prod_{q\in Q_r}|D_t^{\gamma_q}\psi_q(t,\tau)|\right) \le \left(\sum_{m=1}^{N}\left(E|\bar{g}_m|+E\bar{g}_m^2+E|\bar{g}_m|^s\right)\right)^r \le C(r)N^{r/2}. \tag{17}$$

From (9), (16), (17) and (6) we have

$$J_{3k} \le C(k)N^{(k+1)/2}\exp\{-C_0N\} \tag{18}$$

because $N > s+1$, and $|t| \le C_4\beta_{3N}^{-1} \le C_0\sqrt{N}$. For J_{4k} the obvious estimate

$$|J_{4k}| \le C(\beta_{sN}+N^{-(s-2)/2}) \le C(\beta_{3N}+\beta_{sN}) \tag{19}$$

is true.

Since $p_n(n) = n^{n-1}(n-1)!^{-1}\exp(-n)$ then with the aid of Stirling's formula we obtain

$$\left|\frac{1}{\sqrt{2\pi\sqrt{n}p_n(n)}}-1\right| \le \frac{C}{n}. \tag{20}$$

Hence the bound

$$|J_{5k}| \le \frac{C}{n} \le C(\beta_{3N}+\beta_{sN}) \tag{21}$$

is true (noting that $\beta_{3N} \ge 1/\sqrt{N}$ and $N \approx n/k$).

In (6) we replace summands by their bounds from (8), (15), (18), (19) and (21). Then we find

$$|J_k| \le C\left(\beta_{3N}+\beta_{sN}\right). \tag{22}$$

Putting in (4) $T = C_4\beta_{3N}^{-1}$ and using (22) we complete the proof of the theorem.

REFERENCES

1. Bhattacharya R.N., and Rao R. R. Normal Approximation and Asymptotic Expansions. New York, Wiley. (1976).
2. Mirakhmedov, S. A. Lower estimation of the remainder term in the CLT for a sum of the functions of k-spacings. *Statist. Probab. Lett.* **73** (2005), no. 4, 411–424.
3. Mirakhmedov, S. M. Jammalamadaka, S. R. Higher-order expansions and efficiencies of tests based on spacings. *J. Nonparametr. Stat.* **25** (2013), no. 2, 339–359.
4. Holst, L. Asymptotic normality of sum-functions of spacings. *Ann. Probab.* **7** (1979), 1066–1072.
5. Pyke, R. Spacings (with discussion). *J. Roy. Statist. Soc. Ser. B* **27** (1965), 395–449.
6. Mirakhmedov, S.; Kalandarov, U. The non-uniform bounds of remainder term in CLT for the sum of functions of uniform spacings. *Turkish J. Math.* **22** (1998), no. 1, 53–60.

7. Petrov, V. V. Sums of independent random variables *Science*. Moscow. 1972.

Cite this article

Lazareva V.A. *The non uniform bounds of remainder term in CLT for the sum of functions of k-spacings*, **Acta NUUz**, 2026, No 2/1, pp. 55-60.