Asymptotic Expansion of Uniform Distribution on a Circle

Muhammad Naeem*
Deanship of Preparatory Year Program Umm Al Qura University, Makkah Mukarramah, KSA; naeemtazkeer@yahoo.com

Keywords: Asymptotic Normality, Edgeworth Expansion, IID Random Vectors, Random Variables, Uniform Spacing

1. Introduction

Let \( X_1 \leq X_2 \ldots \leq X_{n-1} \) be the order statistics of the sample \( X_1, X_2, \ldots, X_n \) define the spacing as \( D_{m,n} = X_{m,n} - X_{(m-1),n} \), \( m = 1, 2, \ldots, k \in \mathbb{N} \), \( D_k = 1 - X_{k,n} \) and \( \mathcal{D} = (D_{1,n}, \ldots, D_{k,n}) \) with notation \( X_{0,n} = 0 \) and \( X_{n,n} = 1 \), we consider the tests based on statistics of the form

\[
R_k(\mathcal{D}) = \sum_{m=1}^{k} g(nD_{m,n}, m/k), \quad (1.1)
\]

Where function \( g(u, y) \) is defined on \([0, \infty] \times [0,1] \). Statistics \( R_k(\mathcal{D}) \) is called symmetric if \( g(u, y) = g(u) \) i.e. does not depend on \( y \); otherwise it is said to be a non–symmetric. Non–Symmetric statistics arises, also, when the spacing \( D_{m,n} \) are reduced to the uniform spacing.

The statistics of type \( R_k(\mathcal{D}) \) that are most popular, are generated by the following functions:

- The linear statistic: \( g(x, y) = l(y)x \), where \( l(y) \) is a real function defined on \([0,1] \).
- Greenwood' statistic: \( g(x, y) = x^2 \); for example.
- The log-spacing statistics : \( g(x, y) = \log x \).
- Entropy-type spacing statistics: \( g(x, y) = x \log x \).
- Generalized Rao’s statistics: \( g(x, y) = |x - k|_r \), \( r = 0 \).

The first test statistic based on simple spacing \( D_{m,n} \) was introduced by Greenwood in connection with testing whether certain events such as the spread of disease occur at random on the time axis. The Greenwood statistics corresponds to the case \( g(u, y) = u^2 \). Since then many tests based on spacing have been proposed in the literature for a good survey. Several famous papers provided a unified treatment of the asymptotic distribution theory for symmetric simple spacing statistics. There are so many random variables for which the exact distributions in manageable form do not exist. Therefore, it is a common practice that the applied statisticians are making their efforts to derive limit theorems in order to approximate
the exact distribution to know asymptotic properties of random variables. In particular the normal distribution is often used for this kind of approximations, due to the fact that many statistics are at first order asymptotically equivalent to a sum of i.i.d.r.v.s. In order to obtain better rate of convergence in the Central Limit Theorem a standard technique is to replace the normal distribution by the so called Edgeworth expansion, which is determined by the higher moments of the distributions. The advantage of the Edgeworth series is that the error is controlled, so it is a true asymptotic expansion. Some authors calculated Edgeworth expansion of spacing statistics for small to moderate sample sizes. The validity of formal Edgeworth expansions has been proved under suitable assumptions in articles. It is not naive to find out Edgeworth approximation for a function that is a special case of (1.1) for example in the authors discussed an Edgeworth approximation for g(u) = u^2. In their paper, Does et al. used the characterization of Le Cam, and established an Edgeworth type expansion for the sum of a function of uniform spacing. In this paper we represent S_n as sum of sine function of uniform spacing, using the same conditions and assumptions and derive Edgeworth type expansion for it. This will obviously provide more general results regarding the random variable discussed above. This paper is organized as follows. In section 2 we formulate our Theorem, Preliminary Lemmas are stated in section 3, Section 4 deals with the distribution of random variable S_n and contains the proof of the Theorem.

2. Result

Let X_j, j = 1, 2, …, n be the independent and identically distributed random vectors which take their places on the unit circumference. Denote by S_n the area of the convex polygon having X_j, j = 1, 2, …, n as vertices, then obviously S_n →π a.s. as n → ∞. Let the successive arc-length or spacing be denoted D_j, D_{j+1}…D_{n} such that D_1 + D_2 + … + D_n = 1. Such spacing has been widely studied; see the review paper by Pyke. Let Φ(x) be the standard normal distribution and

\[
\tilde{F}_n(x) = \Phi(x) - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[ 1 + \frac{1}{6} \frac{x^3}{3} - \frac{1}{3} \frac{x^4}{3} \right] \left( 1 + \frac{1}{24} \frac{934}{425} x^5 - \frac{1}{72} \frac{20}{3} x^6 \right) + \frac{1}{8} \left[ \frac{5}{12} \frac{20}{3} + \frac{69}{25} + 4\frac{\sqrt{15}}{3} x \right] + \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} n \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[ 1 + \frac{1}{6} \frac{x^3}{3} - \frac{1}{3} \frac{x^4}{3} \right] \left( 1 + \frac{1}{24} \frac{934}{425} x^5 - \frac{1}{72} \frac{20}{3} x^6 \right) + \frac{1}{8} \left[ \frac{5}{12} \frac{20}{3} + \frac{69}{25} + 4\frac{\sqrt{15}}{3} x \right] + \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} n \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[ 1 + \frac{1}{6} \frac{x^3}{3} - \frac{1}{3} \frac{x^4}{3} \right] \left( 1 + \frac{1}{24} \frac{934}{425} x^5 - \frac{1}{72} \frac{20}{3} x^6 \right) + \frac{1}{8} \left[ \frac{5}{12} \frac{20}{3} + \frac{69}{25} + 4\frac{\sqrt{15}}{3} x \right]
\]

Theorem: Let \( \Delta_n = \pi - S_n \) with \( \tilde{F}_n(x) \) as in (2.1) while \( a_n = 8\pi^2 / n^2 \) and \( \beta_n^2 = (19(46\pi^3))/n^5 \). Then

\[
P \{ (\Delta_n - a_n) / \beta_n \leq x \} = \tilde{F}_n(x) + o(1/n), x \in R \text{ as } n \to \infty.
\]

3. Preliminary Lemmas

The aim of this section is to settle some technical points. Only statements of the required lemmas are presented here. For proof one should see 31.

Lemma 1: Let \( g : [0, \infty) \to R \) be a nonlinear measurable function whose derivative exists and is not necessarily constant on \( (c, d) < (0, \infty) \) such that \( E g'(y) < \infty \).

Lemma 2: Let \( x \) be a random variable taking values in \( R^n \), the distribution of which is absolutely continuous on some Borel set \( B \) with \( P(X \in B) > 0 \). Let \( f : R^n \to R^k \) be a measurable function which is Lebesgue almost everywhere differentiable on \( B \) with the \( k \times m \) matrix \( f \) as differential. If all \( \gamma \in R^k \setminus \{0\} \) satisfy \( P(f(X) \gamma = 0) / X \in B) < 1 \). Then \( \lim_{|x| \to \infty} E e^{i\gamma f(x)} < 1 \) holds.

4. Proof

Let \( U_1, U_2, ..., U_n \) be an ordered sequence of independent uniform \( (0,1) \) random variables. The uniform spacing are defined as \( G_j = U_{j+1} - U_j, j = 1, 2, ..., n \). Let \( Y_j, j = 1, 2, ..., n \) be the polar angles corresponding to \( X_j, j = 1, 2, ..., n \) and \( \psi_1, \psi_2, ..., \psi_n \) be the ordered statistics corresponding to \( \psi_j, j = 1, 2, ..., n-1 \) with notations \( \phi_0 = 0 \) and \( \phi_n = 2\pi \) then
\[ S_n = \frac{1}{2} \sum_{j=1}^{n} \sin(\phi_j - \phi_{j-1}). \]

Let \( \xi_1, \xi_2, \ldots, \xi_n \) be i.i.d. exponential random variables then \( Y_i = \frac{n}{2\pi} \xi_i \) is exponentially distributed with parameter 1. By putting \( \zeta = (\xi_1, \xi_2, \ldots, \xi_n) \).

Then \( \mathcal{G}(\phi_1, \phi_2, \ldots, \phi_{n-1}) = \mathcal{G}(\xi_1, \xi_2, \ldots, \xi_{n-1})/\xi_n = 2\pi) \).

Now \( 2\Delta_n = 2\pi - 2S_n - \sum_{j=1}^{n} \left[ (\phi_j - \phi_{j-1}) - \sin(\phi_j - \phi_{j-1}) \right] \).

Therefore,

\[ \mathcal{G}(2\Delta_n) = \mathcal{G} \left[ \sum_{j=1}^{n} \left( \zeta_j - \zeta_{j-1} \right) - \sin(\zeta_j - \zeta_{j-1}) \right] = 2\pi \]

\[ \mathcal{G}(4\pi/n \Delta_n) = \mathcal{G} \left[ \sum_{j=1}^{n} (Y_j - \sin Y_j)/Y = n \right]. \]

Define a statistics as \( Q_n = \sum_{m=1}^{n} g(nG_{m,n}) \), \( n = 1, 2, \ldots \)

where \( Q_n \) is a special case of (1.1) based on uniform spacing formed on the circumference and \( g \) satisfies the conditions settled in lemma–1 and lemma–2. Consider a random variable with \( g(Y) = Y - \sin Y \) where \( Y \sim \text{Exp}\{1\} \) becomes a special case of \( Q_n \). If \( F_n \) is the distribution of \( \left( T_n - ET_n \right)/\sqrt{\text{Var}T_n} \) and \( \tilde{F}_n \) is as in (2.1) then

\[ \lim n \sup_{x \in \mathbb{R}} \left| F_n(x) - \tilde{F}_n(x) \right| \leq o(1), x \in \mathbb{R}. \]

Now, although long but simple calculations lead one to get \( \mu \equiv E(Y - \sin Y) = 1/2 \) and \( \sigma^2 \equiv \text{Var}(Y - \sin Y) = \frac{\sqrt{23}}{2\sqrt{5}} \) and \( \rho^2 \equiv \text{Cov}(Y - \sin Y) = 1. \)

Introduce \( \tilde{g}(Y) = \frac{g(Y) - \mu - (Y - 1)}{(\sigma^2 - \rho^2)^{1/2}} = \frac{\sqrt{5}(1 - 2\sin Y)}{\sqrt{3}} \).

Then \( k_3 = 34, k_4 = \frac{934}{425}, a = -\frac{1}{2} E\tilde{g}(Y)(Y - 1)^2 = \frac{\sqrt{5}}{2\sqrt{3}} \) and \( b = 3(Eg^2(Y)(Y - 1)^2) = 69/25 + 4\sqrt{15} \).

The Edgeworth type expansion \( \tilde{F}_n \) of \( \tilde{G} \) is as given in (2.1).

Note that \( E(Y - \sin Y)^4 = \frac{499271}{1700} \). So the first condition of Lemma–1 is satisfied. By taking \( m = 1, k = 1 \), \( f(x) = \left( x, \frac{\sqrt{5}(1 - 2\sin x)}{\sqrt{3}} \right) \) and \( B = (0, \infty) \) in Lemma–2 and let \( y = (c, d) \) then \( f(x)^y \gamma = \left[ x, \frac{\sqrt{5}(1 - 2\sin x)}{\sqrt{3}} \right] [c \ d]^T = cx + \frac{\sqrt{5}(1 - 2\sin x)}{\sqrt{3}} d \). For, \( f(x)^T \cdot \gamma = 0 \) three cases arises

(i) \( c = 0, d \neq 0 \), (ii) \( c \neq 0, d = 0 \), (iii) \( c \neq 0, d \neq 0 \). For all the three possible cases \( \sup_{(s,t) \in B} \left| Q(s, t) \right| \leq 1 \) also \( \frac{d}{dy}(y - \sin y) = 1 - \cos y \) is not constant on \( (0, \infty) \). Hence by lemma–1

\[ \lim n \sup_{x \in \mathbb{R}} \left| F_n(x) - \tilde{F}_n(x) \right| = o(1), x \in \mathbb{R} \]

This completes the proof.

5. Conclusion

The Edgeworth series approximate the density function in terms of its first four moments thus providing an improvement to the central limit theorem. The theorem proved in this paper proposes an improvement in the central limit theorems proved in 15,16.

6. Acknowledgment

The author would like to pay extreme gratitude to the editor and unknown referee whose valuable suggestions provided a proper make up to the paper.

7. References

1. Mirakhmedov SM, Naeeem M. Asymptotic Properties of the Goodness-of-Fit Tests based on Spacing. Pak J Statistics. 2008; 24(4):253–68.
2. Holst L, Rao JS. Asymptotic spacing theory with applications to the two sample problem. Canadian J Statist. 1981; 9:603–10.

3. Kuo M, Jammalamadaka SR. Limit theory and efficiencies for tests based on higher ordered spacing. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions; Calcutta. 1981. p. 333–52.

4. Jammalamadaka SR, Goria MN. A goodness of fit test based on Gini index spacing. Statist and Prob Letters. 2004; 68:177–87.

5. Naeem M. Asymptotic Efficiencies of the Goodness–of–Fit Test based on Linear statistics. Accepted in NTMSCI. 2015.

6. Greenwood M. The statistical study of infectious diseases. J Roy Statist Soc A. 1946; 109:85–110.

7. del Pino GE. On the asymptotic distribution of k–spacing with applications to goodness–of–fit tests. Ann Statist. 1979; 7:1058–65.

8. Xian Z, Jammalamadaka SR. Bahadur efficiencies of spacing tests for goodness of fit. Ann Inst Statist Math. 1989; 28:783–6.

9. Darling DA. On a class of problems related to the random division of an interval. Ann Math Statist. 1953; 24:239–53.

10. Deheuvels P, Derzko G. Exact laws for sums of logarithms of uniform spacing. Austrian J Statist. 2003; 32:29–47.

11. Jammalamadaka, Tiwari. Efficiencies of some disjoint spacing tests relative to a chi-square test. In: Madan P, Valaplanja JP, Wertz W. Perspectives and New Directions in Theoretica and Applied Statistics I. JohnWiley; 1987. p. 311–7.

12. Jammalamadaka SR, Zhou X, Tiwari RC. Asymptotic efficiencies of spacing tests for goodness of fit. Metrika. 1989; 36:355–77.

13. Bartoszewicz J. Bahadur and Hodges–Lehmann approximate efficiencies of tests based on spacings. Statist Probab Lett. 1995; 23:211–20.

14. Rao JS. Bahadur efficiencies of some tests for uniformity on the circle. Ann Math Statist. 1972; 43:468–79.

15. Nagajov AV, Goldfield SM. The limit theorem for the uniform distribution on the circumference. Dresden: Wiss zeit der Tech University; 1989.

16. Naeem M. On Random Covering of a Circle. Journal of Prime Research in Mathematics. 2008; 4:127–31.

17. Mirakhmedov SM, Naeem M. Asymptotical Efficiency of the Goodness of Fit Test based on Extreme k–spacing Statistic. J Appl Probabl Statist. 2008; 3(1):65–75.

18. Jammalamadaka SR, Wells MT. A test of goodness–of–fit based on extreme spacing with some efficiency comparisons. Metrika. 1988; 35:223–32.

19. Pyke R. Spacings. J Roy Stat Soc B. 1965; 27:395–449.

20. Deheuvels P. Spacings and applications. In: Mogyorodi J, editor, Probability and Statistics Decisions Theory A. 1985. p. 1–30.

21. Chibisov DM. On the test of fit based on sample spacing. Theor Probabl Appl. 1961; 6:325–9.

22. Rao JS, Sethuraman J. Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors. Ann Statist. 1975; 3:299–313.

23. Ghosh K. Some contributions to inference using spacing [PhD thesis]. Santa Barbara, USA: Department of Statistics and Applied Probability. University of California; 1997.

24. Ghosh K, Jammalamadaka SR. Some recent results on inference based on spacings. In: Puri ML, editor. Asymptotic in Statistics and Probability. VSP; 2000. p. 185–96.

25. Le Cam L. Un theoreme sur le division de unintervalle per despoints pris au hazard. Publ Inst Statist Univ Paris. 1958; 7:7–16.

26. Kale BK. Unified derivation of tests of goodness of fit based on spacings. Sankhya Ser A. 1969; 31:43–8.

27. Ghosh K, Jammalamadaka SR. Small sample approximation for spacing statistics. Journal of Statistical Planning and Inference. 1998; 69:245–61.

28. Kallenberg WCM. Interpretation and Manipulation of Edgeworth Expansion. Ann Inst Statist Math. 1993; 45(2):341–51.

29. Bhattacharya RN, Ghosh JK. On the validity of the formal Edgeworth expansion. Ann Statist. 1978; 6:434–51.

30. Does RJMM, Helmers R, Klaassen CAJ. Approximating the distribution of Greenwood's Statistics. Statistica Neerlandica. 1988; 42(3).

31. Does RJMM, Helmers R, Klaassen CAJ. On Edgeworth expansion for the sum of a function of uniform spacings. J Stat Plann Infer. 1987; 17:149–57.