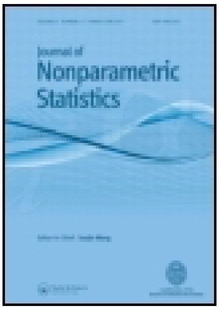

Journal of Nonparametric Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gnst20

Higher-order expansions and efficiencies of tests based on spacings

Sherzod M. Mirakhmedov [a] & S. Rao Jammalamadaka [b]

[a] Institute of Mathematics and Information Technologies, Tashkent, Uzbekistan

[b] Department of Statistics and Applied Probability, University of California Santa Barbara, Santa Barbara, CA, USA

Version of record first published: 13 Feb 2013.

To cite this article: Sherzod M. Mirakhmedov & S. Rao Jammalamadaka (2013): Higher-order expansions and efficiencies of tests based on spacings, Journal of Nonparametric Statistics, DOI:10.1080/10485252.2012.755530

To link to this article: http://dx.doi.org/10.1080/10485252.2012.755530

Journal of Nonparametric Statistics, 2013
http://dx.doi.org/10.1080/10485252.2012.755530

Higher-order expansions and efficiencies of tests based on spacings

Sherzod M. Mirakhmedov[a] and S. Rao Jammalamadaka[b]*

[a]*Institute of Mathematics and Information Technologies, Tashkent, Uzbekistan;* [b]*Department of Statistics and Applied Probability, University of California Santa Barbara, Santa Barbara, CA, USA*

(*Received 6 January 2012; final version received 28 November 2012*)

Statistics based on spacings, or the gaps between points, have been widely used in many contexts, primarily in testing goodness of fit. This paper derives Edgeworth-type asymptotic expansions for the sum of functions of s-step spacings where s, the order of spacings, may increase together with the sample size n. When s is fixed, it is known that only the Greenwood test, based on the sum of squares of these spacings, is first-order asymptotically efficient. In contrast, it is shown here that if s goes to infinity, there exist many other tests which are first-order efficient. We introduce and study the second-order efficiency of such tests and show that if s is sufficiently large relative to n, the Greenwood test is no longer second-order efficient. Interestingly, we see that the common phenomenon of first-order efficiency implying second-order efficiency does not hold true in this situation.

Keywords: spacings; goodness-of-fit tests; higher-order expansions; asymptotic efficiencies; Pitman efficiency; second-order efficiency

2000 Mathematics Subject Classifications: 62G10; 62G20; 60F10

1. Introduction and review

Let $X_1, X_2, \ldots, X_{n-1}$ be independent random variables (r.v.s) with a common absolutely continuous distribution function F with support on the unit interval [0,1]. Let $X_{1,n} \le X_{2,n} \le \cdots \le X_{n-1,n}$ be the corresponding order statistics. For an integer $s \ge 1$, the disjoint spacings of step size or 'order' s are defined as

$$W_{m,s} = X_{ms,n} - X_{(m-1)s,n}, \quad m = 1, 2, \ldots, N,$$

with the notation $X_{0,n} = 0$, $X_{n,n} = 1$, and $N = n/s$, which is assumed to be an integer, without loss of generality, for the asymptotic theory. We allow for s to increase with n, i.e. $s = s(n)$, but with the proviso that $s = o(n)$. When $s = 1$, these are called one-step spacings or simply spacings.

One may, in general, consider what have been called 'asymmetric' statistics of the form $\mathbb{T}_N(W) = \sum_{m=1}^{N} f_m(nW_{m,s})$, where f_m, $m = 1, 2, \ldots, N$, are a given sequence of real functions defined on non-negative reals. But for the reasons given below, the class of 'symmetric' statistics, where $\{f_m,\ m = 1, 2, \ldots, N\}$ are the same for all m, is of considerably more interest. Thus, for a given real function f on the non-negative real axis, most of this paper (except Section 2) is

*Corresponding author. Email: rao@pstat.ucsb.edu

concerned with symmetric class of test statistics

$$\mathbb{S}_N(W) = \sum_{m=1}^{N} f(nW_{m,s}). \tag{1}$$

The spacings clearly depend on n, and the function f may also depend on n in the cases we consider, but for simplicity of notation we suppress the extra subscript n. Most common and well-known examples of spacings tests of the form (1) are the so-called Greenwood statistic

$$G_N^2 = \sum_{m=1}^{N} (nW_{m,s} - s)^2,$$

the Log-spacings statistic which is sometimes called the Moran statistic

$$M_N = \sum_{m=1}^{N} \log(nW_{m,s}),$$

the entropy-type statistic

$$H_N = \sum_{m=1}^{N} nW_{m,s} \log(nW_{m,s}),$$

the Kimball statistic

$$K_{\gamma,N} = \sum_{m=1}^{N} (nW_{m,s})^{\gamma}, \quad \gamma(\gamma - 1) \neq 0, \quad \gamma > -\frac{s}{2},$$

and the Rao statistic

$$R_N = \sum_{m=1}^{N} |nW_{m,s} - s|.$$

We are interested here in using statistics of the form (1) in goodness-of-fit problems, which, through a probability integral transformation, can be reduced to testing the null hypothesis of uniformity, i.e.

$$H_0 \colon F(x) = x, \quad 0 \leq x \leq 1.$$

Statistics of the type (1) are of great interest in several contexts including hypothesis testing and reliability (see, for instance, Pyke 1965, for a somewhat back dated but a very useful review), circular data analysis where they play a pivotal role because they provide a maximal invariant under the rotation group (see, e.g., Jammalamadaka and SenGupta 2001), and spacings-based parameter estimation (see, e.g., Ghosh and Jammalamadaka 2000), just to name a few applications. Correspondingly, there is a rich literature devoted to such statistics and their use. Given that these spacings are highly dependent random quantities with a Dirichlet distribution under H_0 in finite samples, large-sample theory is the main avenue for studying such statistics. For some exact distribution theory in finite samples, see, e.g., Rao (1976) and Rao and Sobel (1980).

Further, in trying to decide which of these tests performs better, one needs to investigate their asymptotic efficiencies (AEs). Among others, comparison of asymptotic local powers of such tests is one important way to of assessing their AEs. In order to study the local powers of spacings statistics, we consider a sequence of alternatives

$$H_{1n} \colon F_n(x) = x + L(x)\delta(n), \quad 0 \leq x \leq 1, \tag{2}$$

where $\delta(n) \to 0$, as $n \to \infty$, and $L(x)$ are 'sufficiently smooth' (to be made precise later in Section 3).

The alternatives (2) need some explanation. When $\delta(n) = n^{-1/2}$, the locally most powerful test among the asymmetric class, based on one-step spacings (or s-step spacings for any fixed s), is linear in the spacings and is thus a weighted linear combination of the order statistics (see, e.g., Holst and Rao 1981). The test statistics depend specifically on the particular alternative, i.e. the function $L(x)$, and perform poorly against any other alternative, and therefore are not very interesting. On the other hand, if one considers symmetric statistics in spacings, no power is obtained for alternatives at a 'distance' of $n^{-1/2}$ and this is connected with a result of Chibisov (1961). However, as shown in Rao and Sethuraman (1975), power for such symmetric tests based on one-step spacings can only be obtained when $\delta(n) = n^{-1/4}$, but however as a bonus, one obtains 'omnibus' tests, i.e. ones whose form does not depend on the alternative $L(x)$. Statistics based on s-step (or higher order) spacings are shown to be asymptotically more efficient (see, e.g., Del Pino 1979; Rao and Kuo 1984) with their Pitman efficiency increasing monotonically with any finite step size s. This led to investigations about letting the step size s also increase as a function of the sample size n. Jammalamadaka, Zhou, and Tiwari (1989) have shown that symmetric tests can discriminate alternatives (2) with $\delta(n) = (ns)^{-1/4}$ which are somewhat closer to the rate $n^{-1/2}$, if $s \to \infty$. Also, it is shown that the optimal choice of the spacings step is $s = O(n^{3/5})$ to maximise the Pitman efficiency.

Note that the spacings can be considered as grouping of the observations with random end points for the class intervals and, as a result, one should expect some loss in information contained in the sample. Indeed, there is a duality between the chi-square tests which compare the observed and expected *frequencies* holding the number of classes fixed, and the s-step Greenwood test which compares the observed and expected *cell-lengths* holding the observed number in each cell fixed at the step size s. Jammalamadaka and Tiwari (1985, 1987) show that the Greenwood test based on spacings is superior to a comparable chi-square test for any fixed s, but when $s \to \infty$, these two tests have asymptotically the same Pitman efficiency (cf. Jammalamadaka et al. 1989).

When $\delta(n) = (ns)^{-1/4}$, the alternatives converge to the null at a rate that keeps the asymptotic power bounded away from the 'level' of the test and 1, and form a family of Pitman alternatives. The Greenwood test is known to be optimal within this family, i.e. in terms of the Pitman AE (see, for instance, Sethuraman and Rao 1970; Del Pino 1979; Rao and Kuo 1984; Jammalamadaka et al. 1989; Mirakhmedov and Naeem 2008). Another alternative family arises when one assumes that $\delta(n)$ remains constant, i.e. alternatives do not approach the hypothesis. Zhou and Jammalamadaka (1989) showed that for this family of alternatives, the Greenwood test is inferior to the Log-spacings test in terms of the Bahadur AE. Furthermore, these alternatives (2) with $\delta(n) \to 0$ and $\delta(n)(ns)^{1/4} \to \infty$ provide yet another family of intermediate alternatives. Mirakhmedov and Naeem (2008) and Mirakhmedov (2010) showed that the Greenwood test is still optimal within this subfamily of intermediate alternatives with $\delta(n) = o(ns^2)^{-1/6}$ but is inferior to tests satisfying the Cramér condition (such as the Log-spacings and the Rao statistics) within intermediate alternatives when $\delta(n)(ns^2)^{1/6} \to \infty$ (see Mirakhmedov 2010, for details).

As stated above, for the alternatives (2) with $\delta(n) = (ns)^{-1/4}$, the Greenwood test is asymptotically most powerful (AMP) within this class (1) for all finite step sizes s and is the unique AMP test for any fixed s. Nevertheless, as we show in Section 3, if $s \to \infty$ then there are also other AMP tests, e.g. the Log-spacings, Kimball's, and Entropy-type tests. One should therefore consider further comparison of these tests based on higher-order expansions for the power functions of these tests. We introduce and study the second-order efficiency for such AMP tests in Section 3. We show that for $s \gg n^{1/3}$, the Greenwood test loses the optimality property, in the sense of not being second-order efficient. Interestingly, we see that the common phenomenon of first-order efficiency implying second-order efficiency (see Bickel, Chibisov, and van Zwet 1981) does not hold true in this situation.

We now mention a few papers that deal with limit theorems for spacings statistics. Results on the central limit theorem and Berry–Esseen bounds for the remainder term, both under the null hypothesis and under the alternatives (2), have been studied by Rao and Sethuraman (1975), Del Pino (1979), Does and Klaassen (1984), Deheuvels (1985), Kuo and Jammalamadaka (1984), Mirakhmedov (2005), and Baryshnikov, Penrose, and Yukich (2009), among others. Edgeworth-type asymptotic expansions under the null hypothesis have been derived by Does, Helmers, and Klaassen (1987) and Kalandarov (2001). Large deviation results were obtained by Zhou and Jammalamadaka (1989), Mirakhmedov (2006), and Mirakhmedov, Tirmizi, and Naeem (2011).

The current work aims to do several things. Our primary goal is to unify as well as generalise many existing results, so that we cover many facets of the spacings theory, including (i) fixed and increasing step sizes, (ii) different types of AEs, e.g. the Pitman, Bahadur, and intermediate efficiencies, and (iii) providing general results for deriving higher-order asymptotic expansions, which underlie the discussion of second-order efficiencies of spacings tests.

The paper is organised as follows: we first consider general asymmetric statistics of the form

$$\mathbb{T}_N(D) = \sum_{m=1}^{N} f_m(nD_{m,s}), \tag{3}$$

where $D_{m,s}$ are s-spacings in a sample from $U[0,1]$ distribution, i.e. under the null hypothesis. In Theorem 2.1, we provide Edgeworth expansions of the distribution function, say $P_N(x)$, of $\mathbb{T}_N(D)$ in the form $P_N(x) = \Phi(x) + \varphi(x)(a_{1N}^{(s)}(x)N^{-1/2} + a_{2N}^{(s)}(x)N^{-1} + \cdots + a_{kN}^{(s)}(x)N^{-(k-3)/2}) + r_{kN}$, where $\Phi(x)$ and $\varphi(x)$ are the standard normal distribution function and its density function, respectively, $k \geq 3$, and the remainder term satisfying $r_{kN} = O(N^{-(k-2)/2})$. This expansion is valid for any fixed s, as well as when it may depend on n with $s = o(n)$. The integral equation (11) for the characteristic function (c.f.) of statistics $\mathbb{T}_N(D)$ makes use of known results on the c.f. of a sum of independent random vectors (r.vec.) and provides a general technique for determining the coefficients $a_{jN}^{(s)}(x)$, as given in Equation (22); the explicit form of these coefficients $a_{1N}^{(s)}(x)$ and $a_{2N}^{(s)}(x)$ are obtained for each of the spacings statistics mentioned earlier. Such results not only provide closer approximations of the distribution function relative to the Central Limit Theorem, but are also essential in comparing the higher-order efficiencies of these tests; they are also useful in comparing tests based on moderate samples (see, e.g., Penev and Ruderman 2010). In Section 3, these results are used to get asymptotic expansions for the power functions of these tests and to compare their second-order efficiencies. The proofs of all the main results are presented in the appendix.

Notationally, in what follows c and C, with or without subscripts, are universal positive constants; we use the same symbol θ for a quantity that satisfies $|\theta| \leq 1$ although it varies by context; all asymptotic statements are considered when $n \to \infty$. Throughout the paper, Z and Z_s are independent Gamma r.v.s with the probability density function (pdf)

$$\gamma_s(u) = \frac{u^{s-1}e^{-u}I\{u > 0\}}{\Gamma(s)}, \tag{4}$$

where $\Gamma(s)$ is the gamma function, $\Phi(x)$ and $\varphi(x)$, as mentioned above, are the standard normal distribution function and its pdf, respectively, and $I\{A\}$ is the indicator of the set A.

2. Asymptotic expansion for the sum of functions of uniform spacings

This section deals with the distribution theory under the null hypothesis of uniformity, for the general class of asymmetric spacings statistics (see Theorem 2.1) and then the result specialised to the symmetric class mentioned in Equation (1) (see Theorem 2.3). This latter theorem is used to obtain Edgeworth-type expansions for the standard tests mentioned in Section 1.

It is convenient to use a special notation for the case of observations from the uniform distribution. We will use $U_1, U_2, \dots, U_{n-1}$ for a sample from the uniform [0,1] distribution, $U_{1,n} \le U_{2,n} \le \cdots \le U_{n-1,n}$ for its order statistics, and $D_{m,s} = U_{ms,n} - U_{(m-1)s,n}$ for their s-step spacings. We will now derive Edgeworth-type asymptotic expansions for the general asymmetric statistics of the form

$$\mathbb{T}_N(D) = \sum_{m=1}^{N} f_m(nD_{m,s}),$$

where f_m, $m = 1, 2, \dots, N$, are a given sequence of real functions defined on non-negative real axis.

We refer to $D_i = D_{i,1}$, $i = 1, 2, \dots, n$, as simple uniform spacings, which one may recall are closely tied to exponential r.v.s. If $\mathcal{L}(X)$ denotes the distribution of the r.vec. X, it is well known that $\mathcal{L}(nD_1, \dots, nD_n) = \mathcal{L}((Y_1, \dots, Y_n)| \sum_{i=1}^{n} Y_i = n)$, where $Y_1, Y_2, \dots, Y_n$ are independent r.v.s with common exponential distribution with mean 1. Then the partial sums

$$Z_m =: Z_{m,s} = Y_{(m-1)s+1} + \cdots + Y_{ms}, \quad m = 1, 2, \dots, N$$

have the pdf $\gamma_s(u)$ given in Equation (4). We assume that $Ef_m^2(Z_m) < \infty$. Define

$$\begin{aligned} \varsigma_N &= \sum_{m=1}^{N} Z_m, \quad T_N = \sum_{m=1}^{N} f_m(Z_m), \quad A_N = ET_N, \\ g_m(u) &= f_m(u) - Ef_m(Z_m) - (u-s)\mathrm{cov}(T_N, \varsigma_N)n^{-1}, \\ V_N &= \sum_{m=1}^{N} g_m(Z_m), \quad \sigma_N^2 = \mathrm{Var}\, V_N, \quad \text{and} \quad \mathbb{V}_N(D) = \sum_{m=1}^{N} g_m(nD_{m,s}). \end{aligned} \tag{5}$$

It is easy to check that

$$EV_N = 0, \quad \mathrm{cov}(V_N, \varsigma_N) = 0, \tag{6}$$

and $\sigma_N^2 = (1 - \mathrm{corr}^2(T_N, \varsigma_N))\mathrm{Var}\, T_N$. Since $\mathbb{V}_N(D)$ is just a centred version of $\mathbb{T}_N(D)$, i.e. $\mathbb{V}_N(D) = \mathbb{T}_N(D) - A_N$, we shall work with $\mathbb{V}_N(D)$ instead of $\mathbb{T}_N(D)$; this is more convenient in view of Equation (6). Set

$$\tilde{g}_m = \frac{g_m(Z_m)}{\sigma_N}, \quad \tilde{Z}_m = \frac{Z_m - s}{\sqrt{n}}, \tag{7}$$

$$\Theta_N(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Psi_N(t, \tau)\, \mathrm{d}\tau, \tag{8}$$

with

$$\Psi_N(t, \tau) = \prod_{m=1}^{N} \psi_m(t, \tau), \quad \psi_m(t, \tau) = E \exp\{\mathrm{i}t\tilde{g}_m + \mathrm{i}\tau\tilde{Z}_m\}.$$

As in Mirakhmedov (2005), we have, for any $m \ge 1$, $v \le N$,

$$\int_{-\infty}^{\infty} |\psi_m(t, \tau)\psi_v(t, \tau)|\, \mathrm{d}\tau \le \sqrt{\int_{-\infty}^{\infty} |\psi_m(t, \tau)|^2\, \mathrm{d}\tau \int_{-\infty}^{\infty} |\psi_v(t, \tau)|^2\, \mathrm{d}\tau} \le \sqrt{2\pi N}, \tag{9}$$

which implies

$$\int_{-\infty}^{\infty} |\Psi_N(t,\tau)|\,\mathrm{d}\tau \le \sqrt{2\pi N}. \tag{10}$$

The conditional representation for the spacings in terms of exponential r.v.s mentioned before allows us to write $\mathcal{L}(\mathbb{V}_N(D)) = \mathcal{L}(V_N|\varsigma_N = n)$, and hence $E\exp\{\mathrm{i}t\sigma_N^{-1}\mathbb{V}_N(D)\} = E(\exp\{\mathrm{i}t\sigma_N^{-1}V_N\}|\varsigma_N = n)$. Using Equation (10) and the fact

$$E\exp\{\mathrm{i}t\sigma_N^{-1}V_N + \mathrm{i}\tau(\varsigma_N - n)\} = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\tau(u-n)}\gamma_n(u)E(\exp\{\mathrm{i}t\sigma_N^{-1}V_N\}|\varsigma_N = u)\,\mathrm{d}u,$$

we have, by Fourier inversion, the following Bartlett-type formula:

$$\phi_N(t) \stackrel{\text{def}}{=} E\exp\left\{\mathrm{i}t\frac{\mathbb{V}_N(D)}{\sigma_N}\right\} = \frac{1}{2\pi\sqrt{n}\gamma_n(n)}\int_{-\infty}^{\infty}\Psi_N(t,\tau)\,\mathrm{d}\tau = \frac{\Theta_N(t)}{\Theta_N(0)}, \tag{11}$$

where $\gamma_n(n)$ is the pdf of ς_N as given in Equation (4).

A formal construction of the asymptotic expansion for $\phi_N(t)$ that comes from formula (11) is as follows. The integrand function $\Psi_N(t,\tau)$ is the c.f. of a sum of N independent two-dimensional r.vec.s $(\tilde{g}_m, \tilde{Z}_m)$. Because of Equation (6), this sum has zero expectation, unit covariance matrix, and uncorrelated components. From Bhattacharya and Rao (1976, Chapter 2) (which will be referred to as BR from now on), it is well known that under suitable conditions, the c.f. $\Psi_N(t,\tau)$ can be approximated by a power series in $N^{-1/2}$, coefficients of which are polynomials with respect to (wrt) t and τ containing the factor $\exp\{-(t^2+\tau^2)/2\}$. Hence that series can be integrated wrt the variable τ over the interval $(-\infty,\infty)$. As a result of this integration, we obtain a power series in $N^{-1/2}$, say $Q_N(t)$ (see Equation (19)). Now we replace $\sqrt{2\pi n}\gamma_n(n) = \Theta_N(0)$ by its series approximation, which is actually $Q_N(0)$. Finally, we obtain the asymptotic expansion for $\phi_N(t)$ by dividing $Q_N(t)$ by $Q_N(0)$.

The procedure outlined above needs somewhat long and complex calculations, although manageable. Assume that $E|g_m(Z_m)|^k < \infty$, $k \ge 3$, and set,

$$\rho_{j,N} = \sum_{m=1}^{N} E|\tilde{g}_m|^j, \tag{12}$$

where $\tilde{g}_m$ is as in Equation (7). Let $P_{m,N}(t,\tau)$, $m = 1,2,\ldots$, be the well-known polynomials wrt τ and t from the theory of asymptotic expansion of the c.f. of a sum of independent r.vec.s (see BR, p. 52), in our case concerning the vector sum $(\tilde{g}_1,\tilde{Z}_1) + \cdots + (\tilde{g}_N,\tilde{Z}_N)$; the degree of $P_{m,N}(t,\tau)$ is $3m$ and the minimal degree is $m+2$; the coefficients of $P_{m,N}(t,\tau)$ only involve the cumulants of the r.vec.s $(\tilde{g}_1,\tilde{Z}_1),\ldots,(\tilde{g}_N,\tilde{Z}_N)$ of order $(m+2)$ or less. In particular,

$$\frac{1}{\sqrt{N}}P_{1,N}(t,\tau) = \frac{i^3}{6}\sum_{m=1}^{N}E(t\tilde{g}_m + \tau\tilde{Z}_m)^3,$$

$$\frac{1}{N}P_{2,N}(t,\tau) = \frac{i^4}{24}\sum_{m=1}^{N}(E(t\tilde{g}_m + \tau\tilde{Z}_m)^4 - 3(E(t\tilde{g}_m + \tau\tilde{Z}_m)^2)^2) + \frac{1}{2N}P_{1,N}^2(t,\tau).$$

Define functions $G_{m,N}(t)$ where $G_{0,N}(t) = 1$ and

$$G_{m,N}(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}P_{m,N}(t,\tau)\exp\left\{-\frac{\tau^2}{2}\right\}\mathrm{d}\tau, \quad m = 1,2,\ldots. \tag{13}$$

Using Equation (6) and the notation

$$\alpha_{i,j,N} = \sum_{m=1}^{N} E\tilde{g}_m^i \tilde{Z}_m^j, \tag{14}$$

we find

$$\frac{1}{\sqrt{N}} G_{1,N}(t) = \frac{(\mathrm{i}t)^3}{6}\alpha_{3,0,N} - \frac{\mathrm{i}t}{2}\alpha_{1,2,N}, \tag{15}$$

$$\frac{1}{N} G_{2,N}(t) = \frac{(\mathrm{i}t)^6}{72}\alpha_{3,0,N}^2 + \frac{(\mathrm{i}t)^4}{24}\left[\alpha_{4,0,N} - 3\sum_{m=1}^{N}\alpha_{20m}^2 - 3\alpha_{2,1,N}^2 - \alpha_{3,0,N}\alpha_{1,2,N}\right]$$
$$+ \frac{(\mathrm{i}t)^2}{8}\left[3\alpha_{1,2,N}^2 + \frac{2}{\sqrt{Ns}}\alpha_{2,1,N} - 2\alpha_{2,2,N} + 4\sum_{m=1}^{N}(E\tilde{g}_m\tilde{Z}_m)^2 + \frac{1}{N}\right] - \frac{1}{12Ns}.$$

We note that $E(Z_m - s)^l$ is a polynomial of degree $\lfloor l/2 \rfloor$ wrt s, and also

$$\rho_{j,N} \geq N^{-(j-2)/2}, \quad j \geq 2. \tag{16}$$

Because of this and Lemma 9.5 of BR (p. 71), we have

$$|P_{m,N}(t,\tau)| \leq C(m)(1 + (t^2 + \tau^2)^{3m/2})\rho_{k,N}^{m/(k-2)}, \tag{17}$$

so that

$$|G_{m,N}(t,\tau)| \leq C_1(m)(1 + t^{3m})\rho_{k,N}^{m/(k-2)}. \tag{18}$$

Set

$$Q_{r,N}(t) = \mathrm{e}^{-t^2/2}\sum_{m=0}^{r-3} N^{-m/2} G_{m,N}(t), \quad r \leq k. \tag{19}$$

Using Stirling's formula

$$v! = \sqrt{2\pi v}\, v^v \exp\left\{-v + \sum_{j=1}^{l}\frac{B_{2j}}{2j(2j-1)v^{2j-1}} + O\left(\frac{1}{v^{2l}}\right)\right\},$$

where B_{2j} are Bernoulli numbers (see Abramowitz and Stegun 1972, p. 257), we have

$$\Theta_N^{-1}(0) = (\sqrt{2\pi n}\gamma_n(n))^{-1} = \frac{(n-1)!}{\sqrt{2\pi n} n^{n-1}\mathrm{e}^{-n}} = \sum_{m=0}^{\lfloor (k-2)/2 \rfloor} b(m) n^{-m} + o(n^{-\lfloor (k-2)/2 \rfloor}), \tag{20}$$

where $b(m)$ are constants depending only on m. In particular, $b(0) = 1$, $b(1) = -\frac{7}{12}$. In view of Equations (16) and (18), we use Equation (20) to define the functions $r_{m,N}(t)$ from the equality

$$(\sqrt{2\pi n}\gamma_n(n))^{-1} Q_{k,N}(t) = \mathrm{e}^{-t^2/2}(\Upsilon_{k,N}(t) + O(N^{-(k-2)/2})),$$

where

$$\Upsilon_{j,N}(t) = \mathrm{e}^{-t^2/2}\sum_{m=0}^{j-3} N^{-m/2} r_{m,N}(t).$$

In particular, we have

$$\Upsilon_{3,N}(t) = \exp\left\{-\frac{t^2}{2}\right\},$$

$$\Upsilon_{4,N}(t) = \exp\left\{-\frac{t^2}{2}\right\}(1 + N^{-1/2}G_{1,N}(t)), \tag{21}$$

$$\Upsilon_{5,N}(t) = \exp\left\{-\frac{t^2}{2}\right\}(1 + N^{-1/2}G_{1,N}(t)) + N^{-1}(G_{2,N}(t) - G_{2,N}(0)).$$

Let $\Re_{j,N}(x)$ be defined by the equation $\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx}\, \mathrm{d}\Re_{j,N}(x) = \Upsilon_{j,N}(t)$. Such $\Re_{j,N}(x)$ can be obtained by formally substituting $((-1)^{\nu}/\sqrt{2\pi})(\mathrm{d}^{\nu-1}/\mathrm{d}x^{\nu-1})\,\mathrm{e}^{-x^2/2}$ in place of $(\mathrm{i}t)^{\nu}\,\mathrm{e}^{-t^2/2}$ for each ν in the expression for $\Upsilon_{j,N}(t)$ (see Lemma 7.2 of BR). Using Equation (15), we have, in particular,

$$\Re_{3,N}(x) = \Phi(x),$$

$$\Re_{4,N}(x) = \Phi(x) - \varphi(x)\left[\frac{H_2(x)}{6}\alpha_{3,0,N} - \frac{1}{2}\alpha_{1,2,N}\right],$$

$$\begin{aligned}\Re_{5,N}(x) = \Phi(x) - \varphi(x)\Bigg[&\frac{H_2(x)}{6}\alpha_{3,0,N} - \frac{1}{2}\alpha_{1,2,N} + \frac{H_5(x)}{72}\alpha_{3,0,N}^2 \\ &+ \frac{H_3(x)}{24}\left(\alpha_{4,0,N} - 3\sum_{m=1}^{N}\alpha_{20m}^2 - 3\alpha_{2,1,N}^2 - \alpha_{3,0,N}\alpha_{1,2,N}\right) \\ &+ \frac{H_1(x)}{8}\left(3\alpha_{1,2,N}^2 + \frac{2}{\sqrt{Ns}}\alpha_{2,1,N} - 2\alpha_{2,2,N} + 4\sum_{m=1}^{N}(E\tilde{g}_m\tilde{Z}_m)^2 + \frac{1}{N}\right)\Bigg],\end{aligned} \tag{22}$$

where $H_{\nu}(x)$ is the νth-order Hermite–Chebishev polynomial, namely $H_1(x) = x, H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$, and $H_5(x) = x^5 - 10x^3 + 15x$.

We thus arrive at the main result of this section,

THEOREM 2.1 *Consider the general asymmetric spacings statistic $\mathbb{T}_N(D)$ given in Equation* (3). *Then for any integer $k \geq 3$, there exists a constant $C > 0$ such that*

$$\sup_x |P(\mathbb{T}_N(D) < x\sigma_N + A_N) - \Re_{k,N}(x)| \leq C(\rho_{k,N} + I\{k > 3\}\exp\{-N(1-\chi_N)\}\sqrt{nN}\log\rho_{k,N}^{-1}),$$

where A_N, σ_N^2, $\rho_{k,N}$, and $\Re_{k,N}(x)$ are defined in Equations (5), (12), *and* (22), *and*

$$\chi_N = \frac{1}{N}\sum_{m=1}^{N}\sup_{\substack{c(\rho_{3,N}\sigma_N)^{-1}\leq|t|\leq c(\rho_{k,N}\sigma_N)^{-1}\\ \tau\in(-\infty,\infty)}} |E\exp\{\mathrm{i}tf_m(Z_{m,s}) + \mathrm{i}\tau Z_{m,s}\}| \tag{23}$$

for a constant $c > 0$.

When $k = 3$, we obtain the following special case (which corresponds to Corollary 2 of Mirakhmedov 2005).

COROLLARY 2.1 *There exists a positive constant C such that*

$$\sup_x |P(\mathbb{T}_N(D) < x\sigma_N + A_N) - \Phi(x)| \le C\rho_{3,N}.$$

Results on asymptotic expansions for the moments of this statistic are of interest on their own. From Lemma A.3, we obtain the following asymptotic expansion results, the second part of which improves Lemma 3.4 of Does and Klaassen (1984).

THEOREM 2.2 (1) *If* $E|g_m(Z_m)|^3 < \infty$ *then*

$$E\mathbb{T}_N(D) = \sum_{m=1}^{N} Ef_m(Z_m) + c_1\theta_1\sigma_N\rho_{3,N} \quad \textit{and} \quad \mathrm{Var}\,\mathbb{T}_N(D) = \sigma_N^2(1 + c_2\theta_2\rho_{3,N}).$$

(2) *If* $E|g_m(Z_m)|^4 < \infty$ *then*

$$E\mathbb{T}_N(D) = \sum_{m=1}^{N} Ef_m(Z_m) - \frac{1}{2n}\sum_{m=1}^{N} Eg_m(Z_m)(Z_m - s)^2 + c_3\theta_1\sigma_N\rho_{4,N},$$

and
(3) *if* $E|g_m(Z_m)|^5 < \infty$ *then*

$$\begin{aligned}&\mathrm{Var}\,\mathbb{T}_N(D)\\ &= \sigma_N^2\left(1 + \frac{1}{4}\left(3\alpha_{1,2,N}^2 + \frac{2}{\sqrt{n}}\alpha_{2,1,N} - 2\alpha_{2,2,N} + 4\sum_{m=1}^{N}(E\tilde{g}_m\tilde{Z}_m)^2 + \frac{1}{N}\right) + c_4\theta_2\rho_{5,N}\right),\end{aligned}$$

with the notation used in Equations (5)*,* (12)*, and* (14).

Now we return to the symmetric statistic (1) of interest, for which we consider a two-term expansion. In this case, $g_m(u)$ reduces to

$$g(u) = f(u) - Ef(Z) - (u - s)s^{-1}\,\mathrm{cov}(f(Z), Z),$$

with

$$\mathrm{Var}\,g(Z) = (1 - \mathrm{corr}^2(f(Z), Z))\mathrm{Var}(f(Z)) = \sigma^2 \quad \text{say.} \tag{24}$$

Define

$$\chi = \sup_{(t,\tau)\in\Omega} |E\exp\{\mathrm{i}tf(Z) + \mathrm{i}\tau Z\}|, \tag{25}$$

where $\Omega = \{(t,\tau) : c\sigma^2/E|g(Z)|^3 \le |t| \le cN^{1/2}\sigma^3/E|g(Z)|^4, \tau \in (-\infty, \infty)\}$. Let P_0 stand for the probability under the null hypothesis H_0.

THEOREM 2.3 *Assume that* $E|g(Z)|^4 < \infty$*. Let* $\mathbb{S}_N(W)$ *be the general symmetric statistic defined in Equation* (1)*. Then there exists a constant C such that*

$$\begin{aligned}&\sup_x \left| P_0\left\{\frac{\mathbb{S}_N(W) - NEf(Z)}{\sigma\sqrt{N}} < x\right\} - \Phi(x) - \frac{\varphi(x)}{\sqrt{N}}\left[\frac{(1-x^2)Eg^3(Z)}{6\sigma^3} + \frac{Eg(Z)\,(Z-s)^2}{2s\sigma}\right]\right| \\ &\le C\left(\frac{Eg^4(Z)}{N\sigma^4} + \sqrt{Nn}\log N\, e^{-N(1-\chi)}\right).\end{aligned} \tag{26}$$

Remark 2.1 From Lemma 3.1 of Does et al. (1987), we have the following.

If $(c,d) \subset (0,\infty)$ is an open interval on which f is almost everywhere differentiable with derivative f' which is not essentially constant on (c,d), then the following condition, known as Cramer's condition

$$\limsup_{|(t,\tau)|\to\infty} |E\exp\{\mathrm{i}tf(Z)+\mathrm{i}\tau Z\}| < 1, \tag{27}$$

is fulfilled. Hence, the second term inside the brackets on the rhs of Equation (20) is exponentially small whenever

$$\frac{\sigma^2}{E|g(Z)|^3} \geq c > 0. \tag{28}$$

Remark 2.2 Recall the representation $\mathcal{L}(\mathbb{V}_N(D)) = \mathcal{L}(V_N|\varsigma_N = n)$ stated at the beginning of this section. In view of this, the first term in the square bracket in Equation (26) corresponds to the standardised third-order cumulant of V_N, which is a sum of independent r.v.s; the second term is the result of the fact that the statistic $\mathbb{V}_N(D)$ is defined by centralising $\mathbb{T}_N(D)$ by A_N, which is asymptotic rather than being the exact value of the expectation, namely $E\mathbb{T}_N(D)$ (see Part 2 of Theorem 2.2).

We now discuss how this Theorem 2.3 applies to each of the special cases of symmetric statistics that were mentioned in Section 1 (computational details are omitted).

(A) *Greenwood statistic*. For the Greenwood statistic, we have $f(x) = (x-s)^2$.

THEOREM 2.4 *The asymptotic expansion for the Greenwood statistic is given by*

$$P_0\left\{\frac{G_N^2 - Ns}{\sqrt{2Ns(s+1)}} < x\right\}$$
$$= \Phi(x) + \frac{\varphi(x)}{\sqrt{N}}\left(\frac{\sqrt{2}(s^2+5s+4)(1-x^2)}{3(s+1)^{3/2}\sqrt{s}} + \sqrt{\frac{1}{2}\left(1+\frac{1}{s}\right)}\right) + O\left(\frac{1}{N}\right).$$

(B) *Log-spacings statistic*. For the Log-spacings statistic, we have $f(x) = \log x$. For integers $m \geq 1$ and $k \geq 1$, let

$$\psi(m) = \sum_{j=1}^{m-1}\frac{1}{j} \quad \text{and} \quad \zeta(m,k) = \sum_{j=k}^{\infty}\frac{1}{j^m}$$

be the digamma function and Hurwitz zeta function, respectively.

THEOREM 2.5 *The asymptotic expansion for the Log-spacings statistic is given by*

$$P_0\left\{\frac{M_N - N\psi(s)}{\sqrt{N(\zeta(2,s)-s^{-1})}} < x\right\}$$
$$= \Phi(x) + \frac{\varphi(x)}{\sqrt{N}}\left(\frac{(1-2s^2\zeta(3,s))(1-x^2)}{6(s\zeta(2,s)-1)^{3/2}\sqrt{s}} + \frac{1}{2\sqrt{s}(s\zeta(2,s)-1)^{1/2}}\right) + O\left(\frac{1}{N}\right).$$

(C) *Entropy-type statistic*. In this case, $f(x) = x\ln x$. Applying these in Theorem 2.3, we obtain the following.

THEOREM 2.6 *If $s \to \infty$, then the asymptotic expansion for the Entropy statistic is given by*

$$P_0\left\{\frac{H_N - N\psi(s+1)}{\sqrt{Ns(s+1)\zeta(2,s) - s}} < x\right\} = \Phi(x) + \frac{\varphi(x)}{\sqrt{N}}\left(5\sqrt{2}(1-x^2) + \frac{3\sqrt{2}}{2}\right) + O\left(\frac{1}{\sqrt{ns}} + \frac{1}{N}\right).$$

(*D*) *Kimball statistics*. Here, we have $f(x) = x^\gamma$, with $\gamma(\gamma - 1) \neq 0$. Put $q(u) = \Gamma(s+u)/\Gamma(s)$.

THEOREM 2.7 *If $s \to \infty$ and $\gamma < 1$, then the asymptotic expansion for the Kimball statistic is given by*

$$P_0\left\{\frac{K_{\gamma,N} - Nq(\gamma)}{\sqrt{q(2\gamma) - (1+\gamma^2 s^{-1})q^2(\gamma)}} < x\right\}$$
$$= \Phi(x) + \frac{\varphi(x)}{\sqrt{N}}\left[\frac{\sqrt{2}(\gamma^2 + 3\gamma - 2)(1-x^2)}{12\gamma(\gamma-1)} + \frac{\sqrt{2}}{2}\right] + O\left(\frac{1}{\sqrt{ns}} + \frac{1}{N}\right).$$

(*E*) *Rao's statistic*. In this case, $f(x) = |x - s|$. Let $\Gamma(m, y) = \int_y^\infty u^{m-1}\,\mathrm{e}^{-u}\,\mathrm{d}u$, $y > 0$, denote the upper incomplete gamma function and with $\gamma_m(u)$ as in Equation (4), define

$$a(s) = 2s\gamma_s(s), \quad b(s) = 2\,\mathrm{e}^{-s}\sum_{m=0}^{s}\frac{s^m}{m!} - 1, \quad d(s) = \frac{4}{\Gamma(s)}(s\Gamma(s,s) + s^s\,\mathrm{e}^{-s}(1+s)) - 2s,$$
$$\kappa(s) = Eg^3(z) = (3b^2(s)+1)\,\mathrm{d}(s) + s(3a(s)(b^2(s)-1) - 2b(s)(b^2(s)+3)) + 2a^3(s),$$
$$\sigma^2(s) = s\left[1 - 4s\gamma_s^2(s) - \left(1 - 2\,\mathrm{e}^{-s}\sum_{m=0}^{s}\frac{s^m}{m!}\right)^2\right].$$

With these notation and using Theorem 2.3, we obtain the following theorem.

THEOREM 2.8 *Let $s \geq 1$ be fixed. Then the asymptotic expansion for the Rao statistic is given by*

$$P_0\left\{\frac{R_N - Na(s)}{\sigma(s)\sqrt{N}} < x\right\} = \Phi(x) + \frac{\varphi(x)}{\sqrt{N}}\left[\frac{\kappa(s)(1-x^2)}{6\sigma^3(s)} + \frac{d(s)}{2s\sigma(s)}\right] + O\left(\frac{1}{N}\right).$$

If we let $s \to \infty$, then the second term on the rhs in the asymptotic expansion for the Rao statistic takes the following simple form:

$$\frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi N}}\left[\frac{\sqrt{2}\pi(1-x^2)}{3(\pi-2)^{3/2}} + \sqrt{\frac{2}{\pi-2}}\right].$$

Remark 2.3

(i) Theorems 2.6 and 2.7 are presented for the case $s \to \infty$ in order to have reasonably nice expressions; an asymptotic expansion, which is valid for any fixed integer $s \geq 1$, can be written using exact formulas (A20)–(A24); however, they are very messy and are avoided. Indeed Theorem 2.7 is still valid for any γ which is not zero or one, for any *fixed* $s \geq 1$, because in this case, condition (2.24) is satisfied.

(ii) The conditions '$\gamma < 1$' in Theorem 2.7 and '$s \geq 1$ fixed' in Theorem 2.8 are technical and are assumed in order to satisfy the simple sufficient condition (2.24). However, we conjecture but cannot yet prove that a result similar to Lemma A.5 should be true for Kimball's and Rao's statistics, which then allows us to remove these restrictive conditions.

(iii) Note that in the case when $s \to \infty$, more terms in the asymptotic expansion might be needed (as in the case of triangular arrays) and cannot be written simply as series in powers of $N^{-1/2}$ since the asymptotic behaviour of each term depends on how fast s increases. For instance, this phenomenon is reflected in terms involving $1/\sqrt{ns}$ in the remainder term of Theorems 2.6 and 2.7; note that this term will be dominating if the number of spacings $N \gg n^{2/3}$.

3. Asymptotic efficiencies

In this section, we consider testing the null hypothesis of uniformity against the sequence of alternatives

$$H_1\colon F_n(x) = x + L(x)(ns)^{-1/4}, \quad 0 \le x \le 1, \tag{29}$$

where $L(0) = L(1) = 0$, $L'(x) = l(x)$ is continuous on [0,1] and $l'(x)$ exists.

We consider tests based on symmetric statistics $\mathbb{S}_N(W)$ and assume that we reject H_0 for large values of the statistic $\mathbb{S}_N(W)$. We will refer to such a test as an f-test, for short.

We continue using the notation of previous sections, in particular, $g(Z)$ and σ^2 as defined in Equation (24). Also let $\beta_{\mathbb{S}_N(W)}(l)$ stand for the power function of the test based on the statistic $\mathbb{S}_N(W)$, and with ZZ a r.v. with the pdf $\gamma s(u)$ given in Equation (4), define

$$\mathbb{Z}_2 = Z^2 - 2(s+1)Z + s(s+1),$$
$$\mathbb{Z}_3 = Z^3 - 3(s+2)Z^2 + 3(s+1)(s+2)Z - s(s+1)(s+2),$$
$$\mu(s,f) = \operatorname{corr}(g(Z), \mathbb{Z}_2), \quad \nu(s,f) = \operatorname{corr}(g(Z), \mathbb{Z}_3), \quad \|l\|_k^k = \int_0^1 l^k(x)\,\mathrm{d}x, \tag{30}$$
$$\Delta_\omega(l,s,f) = \|l\|_2^2\,\mu(s,f)\sqrt{\frac{(s+1)}{2s}} - u_\omega, \quad u_\omega = \Phi^{-1}(1-\omega).$$

Let $C^{(k)}$ stand for the class of functions with continuous derivatives of order $j = 0, 1, \ldots, k$.

Theorem 3.1

(i) *Let* $f \in C^{(2)}$, $E|g(Z)|^3/\sigma^3\sqrt{N} \to 0$, *as* $n \to \infty$, *then*

$$\beta_{\mathbb{S}_N(W)}(l) = \Phi(\Delta_\omega(l,s,f)) + o(1). \tag{31}$$

(ii) *Let* $f \in C^{(3)}$, $E|g(Z)|^4 < \infty$, *and the condition* (22) *be fulfilled. Then*

$$\begin{aligned}\beta_{\mathbb{S}_N(W)}(l(x)) = {} & \Phi(\Delta_\omega(l,s,f)) \\ & - \varphi(\Delta_\omega(l,s,f))\left(\frac{(1-\Delta_\omega^2(l,s,f))Eg^3(Z)}{6\sigma^3\sqrt{N}} + \sqrt{\frac{s+1}{2Ns}}\mu(s,f)\right. \\ & \left. + \frac{\|l\|_3^3}{\sqrt{6}}\frac{\nu(s,f)}{N^{1/4}}(1+o(1))\right) + O\left(\frac{1}{N}\max\left(1, \frac{Eg^4(Z)}{\sigma^4}\right)\right).\end{aligned} \tag{32}$$

From Equations (31) and (32), we see that the function $\mu(s,f)$ plays a key role in determining the asymptotic nature of the f-tests. To see the significance and meaning of $\mu(s,f)$, we note (cf. Del Pino 1979)

$$\mu(s,f) = \operatorname{corr}(g(Z), (Z-s)^2) = \operatorname{corr}_0(\mathbb{S}_N(W), G_N^2(W))(1+o(1)), \tag{33}$$

so that $|\mu(s,f)| \le 1$, and $|\mu(s,f)| = 1$ for any s, only for the Greenwood statistic G_N^2, i.e. the Greenwood test is AMP within the class of f-tests; it is the unique AMP test for any fixed s. But if $s \to \infty$ we can and do have additional AMP tests, because there exist other f-tests for which $|\mu(s,f)| \to 1$ as $s \to \infty$. For example, by direct calculations, we find:

for the Log-statistic

$$\mu^2(s, \log x) = \frac{1}{2(s+1)(s\zeta(2,s)-1)} = 1 - \frac{1}{s} + O\left(\frac{1}{s^2}\right), \tag{34}$$

for entropy-type statistic

$$\mu^2(s, x\log x) = \frac{1}{2(s+1)[(s+1)\zeta(2,s+1)-1]} = 1 - \frac{1}{3s} + O\left(\frac{1}{s^2}\right), \tag{35}$$

and for Kimball statistics

$$\mu^2(s, x^\gamma) = \frac{\gamma^2(\gamma-1)^2 q^2(\gamma)}{2s(s+1)\left(q(2\gamma) - (1+\gamma^2 s^{-1})q^2(\gamma)\right)} = 1 - \frac{(\gamma-2)^2}{3s}\left(1 + O\left(\frac{1}{s}\right)\right). \tag{36}$$

Hence, all of these three statistics will also generate AMP tests. However, for Rao's statistic

$$\mu^2(s, |x-s|) = \frac{2(s+1)\gamma_{s+2}^2(s)}{1 - 4s\gamma_s^2(s) - (1 - 2\,\mathrm{e}^{-s}\sum_{m=0}^{s} s^m/m!)^2} = \frac{1}{\pi - 2}\left(1 + O\left(\frac{1}{s}\right)\right), \tag{37}$$

so that Rao's statistic does not generate an AMP test even as $s \to \infty$. Numerical values of $\mu(s,f)$ for these statistics are presented in Tables 1 and 2, for some reasonable step sizes s.

The values on the first row correspond to the Pitman efficiencies obtained in Sethuraman and Rao (1970) for one-step spacings. As $s \to \infty$, the efficiencies increase monotonically to 1 as expected, except for Rao's test.

Table 1. Numerical values of $\mu^2(s,f)$ for the Log-spacings, Rao, and entropy-type statistics.

s	$\mu^2(s, \log x)$	$\mu^2(s, \lvert x-s \rvert)$	$\mu^2(s, x\log x)$
1	0.3876	0.5727	0.8625
2	0.5750	0.6759	0.9019
3	0.6764	0.7264	0.9239
4	0.7391	0.7565	0.9380
5	0.7816	0.7765	0.9476
10	0.8798	0.8218	0.9706
20	0.9368	0.8476	0.9844
30	0.9571	0.8567	0.9894
50	0.9740	0.8643	0.9935
100	0.9872	0.8701	0.9967

Table 2. Numerical values of $\mu^2(s,f)$ for the Kimball statistics.

	γ								
s	$-1/2$	$-1/4$	1/4	1/2	3/2	2	5/2	3	4
$\mu^2(S, X^\gamma)$									
1	Does not exist	0.2072	0.5424	0.6723	0.9678	1.00	0.9725	0.9000	0.6792
2	0.3162	0.4536	0.6802	0.7695	0.9766	1.00	0.9793	0.9230	0.7407
3	0.4839	0.5858	0.7554	0.8229	0.9817	1.00	0.9834	0.9375	0.7826
4	0.5860	0.6670	0.8023	0.8565	0.9850	1.00	0.9861	0.9473	0.8129
5	0.6545	0.7216	0.8342	0.8794	0.9873	1.00	0.9881	0.9545	0.8358
10	0.8110	0.8473	0.9084	0.9330	0.9928	1.00	0.9930	0.9729	0.8983
20	0.909	0.9198	0.9517	0.9646	0.9961	1.00	0.9962	0.9850	0.9423
30	0.9328	0.9456	0.9672	0.9759	0.9973	1.00	0.9973	0.9896	0.9597
50	0.9592	0.9670	0.9800	0.9853	0.9984	1.00	0.9983	0.9936	0.9749
100	0.9794	0.9833	0.9899	0.9928	0.9992	1.00	0.9992	0.9967	0.9870

The AMP tests can now be compared with each other by using the higher-order terms of the asymptotic expansion of their powers, using Theorem 3.1.

Let $\mathbb{S}_{1,N}(W)$ and $\mathbb{S}_{2,N}(W)$ be any two statistics of the form (1) with kernel functions f_1 and f_2, and let $\beta_{\mathbb{S}_{1,N}(W)}(l)$ and $\beta_{\mathbb{S}_{2,N}(W)}(l)$ be their respective power functions.

Definition 3.1 *The AMP f_1-test is called 'second-order asymptotically efficient' wrt f_2-test if*

$$\lim_{N\to\infty} \sqrt{N}(\beta_{\mathbb{S}_{1,N}(W)}(l) - \beta_{\mathbb{S}_{2,N}(W)}(l)) \geq 0.$$

In particular, applying Equation (32) to the specific examples of symmetric statistics and putting $\Delta_\omega(l) = 2^{-1/2}\,\|l\|_2^2 - u_\omega$, we can show (see the appendix)

$$\beta_{G_N^2}(l) = \Phi(\Delta_\omega(l)) - \varphi(\Delta_\omega(l))\sqrt{\frac{2}{N}}\left(\frac{1-\Delta_\omega^2(l)}{3} + \frac{1}{2}\right)\left(1 + O\left(\frac{1}{s}\right)\right) + O\left(\frac{1}{N}\right), \tag{38}$$

$$\begin{aligned}\beta_{M_N}(l) = \Phi(\Delta_\omega(l)) - \varphi(\Delta_\omega(l))\Bigg(&\frac{\|l\|_2^2}{2\sqrt{2}\,s}(1+O(s^{-1})) \\ &+ \sqrt{\frac{2}{N}}\left(\frac{1-\Delta_\omega^2(l)}{3} + \frac{1}{2}\right) + \frac{\sqrt{2}\,\|l\|_3^3}{3(ns)^{1/4}}\Bigg) + O\left(\frac{1}{N}\right),\end{aligned} \tag{39}$$

$$\begin{aligned}\beta_{H_N}(l) = \Phi(\Delta_\omega(l)) - \varphi(\Delta_\omega(l))\Bigg(&\frac{\|l\|_2^2}{6\sqrt{2}\,s}(1+O(s^{-1})) \\ &+ \sqrt{\frac{2}{N}}\left(5(1-\Delta_\omega^2(l)) + \frac{3}{2}\right) + \frac{\|l\|_3^3}{6(ns)^{1/4}}\Bigg) + O\left(\frac{1}{\sqrt{ns}} + \frac{1}{N}\right),\end{aligned} \tag{40}$$

$$\begin{aligned}\beta_{K_N}(l) = \Phi(\Delta_\omega(l)) - \varphi(\Delta_\omega(l))\Bigg(&\frac{(\gamma-2)^2\,\|l\|_2^2}{6\sqrt{2}\,s}(1+O(s^{-1})) \\ &+ \sqrt{\frac{2}{N}}\left(\frac{(\gamma^2+3\gamma-2)(1-\Delta_\omega^2(l))}{12\gamma(\gamma-1)} + \frac{1}{2}\right) + \frac{\sqrt{2}(\gamma-2)^2\,\|l\|_3^3}{6(ns)^{1/4}}\Bigg) + O\left(\frac{1}{\sqrt{ns}} + \frac{1}{N}\right).\end{aligned} \tag{41}$$

Now we consider the problem of second-order asymptotic efficiency (SOAE) within this class of AMP tests, especially when compared with the Greenwood test. Any of these tests can be SOAE iff $s^{-1} = o(N^{-1/2})$, corresponding to the case when the step size s is higher order than $n^{1/3}$, i.e. $s \gg n^{1/3}$; note that in this situation we have $(ns)^{-1/4} = o(N^{-1/2})$. For such an s, comparison of Equations (38)–(40) shows that the Log-test is SOAE but entropy-type is not. Note that $\beta_{K_{2,N}}(\Delta_\omega) = \beta_{G_N^2}(\Delta_\omega)$ as it should be. On the other hand, a comparison of Equation (38) with Equation (41) shows that the Kimball test $K_{\gamma,N}$ is SOAE wrt to the Greenwood test for $\gamma \in (-s/2, 0) \cup [1/3, 1) \cup [2, \infty)$, while failing to be so for values of $\gamma \in (0, 1/3) \cup (1, 2)$.

These results show that the well-known phenomenon of first-order efficiency generally implying second-order efficiency (see, e.g., Bickel et al. 1981) does not hold true here. This effect arises mainly because we are dealing with the case when the step of the spacings s depends on the sample size n. This is similar to the phenomenon that has been observed for tests based on grouped data when the number of groups increases along with the sample size (see, e.g., Kallenberg 1985; Ivchenko and Mirakhmedov 1992).

Acknowledgements

We wish to thank Professor Konstantin Borovkov whose advice on the proof of Lemma A.5 was very helpful. We would also like to express our gratitude to the referee and the Associate Editor for a careful reading of the manuscript and for very constructive comments that considerably improved the presentation of the paper.

References

Abramowitz, M., and Stegun, I.A. (1972), *Handbook of Mathematical Functions with Formulas*, Graphs and Mathematical Tables, New York: Dover.

Baryshnikov, Y., Penrose, M.D., and Yukich, J.E. (2009), 'Gaussian Limits for Generalized Spacings', *Annals of Applied Probability*, 19, 158–185.

Bhattacharya, R.N., and Rao, R.R. (1976), *Normal Approximation and Asymptotic Expansions*, New York: Wiley.

Bickel, P.J., Chibisov, D.M., and van Zwet, W.R. (1981), 'On Efficiency of First and Second Order', *International Statistical Review*, 49, 169–175.

Chibisov, D.M. (1961), 'On Tests of Fit Based on Sample Spacings', *Theory of Probability & Its Application*, 6, 325–329.

Deheuvels, P. (1985), 'Spacings and Applications', in *Probability and Statistics Decisions Theory*, eds. F. Konecny, J. Mogyoródi, and W. Wertz, Dordrecht: Reidel, pp. 1–30.

Del Pino, G.E. (1979), 'On Asymptotic Distribution of k-Spacings with Applications to Goodness of Fit Tests', *Annals of Statistics*, 7, 1058–1065.

Does, R.J.M.M., and Klaassen, C.A.J. (1984), 'The Berry–Esseen Theorem for Function of Uniform Spacings', *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65, 461–471.

Does, R.J.M.M., Helmers, R., and Klaassen, C.A.J. (1987), 'On the Edgeworth Expansion for the Sum of a Function of Uniform Spacings', *Journal of Statistical Planning and Inference*, 17, 149–157.

Feller, W. (1971), *An Introduction to Probability Theory and its Applications* (Vol. 2), New York, London.

Ghosh, K., and Jammalamadaka, S.R. (2000), 'Some Recent Results on Inference Based on Spacings', in *Asymptotic in Statistics and Probability*, ed. M.L. Puri, Utrecht: VSP, pp. 185–196.

Gradshteyn, L.S., and Ryzhik, I.M. (2000), *Tables of Integrals, Series, and Products* (6th ed.), ed. Alan Jeffrey, New York: AP, p. 572.

Holst, L., and Rao, J.S. (1981), 'Asymptotic Spacings Theory with Applications to the Two-Sample Problem', *Canadian Journal of Statistics*, 9, 79–89.

Ivchenko, G.I., and Mirakhmedov, S.A. (1992), 'On Limit Theorems for Decomposable Statistics and Efficiency of the Corresponding Statistical Tests', *Discrete Mathematics and Applications*, 2, 547–562.

Jammalamadaka, S.R., and SenGupta, A. (2001), *Topics in Circular Statistics*, Singapore: World Scientific Press.

Jammalamadaka, S.R., and Tiwari, R.C. (1985), 'Asymptotic Comparison of Three Tests for Goodness of Fit', *Journal of Statistical Planning and Inference*, 12, 295–304.

Jammalamadaka, S.R. and Tiwari, R.C. (1987), 'Efficiencies of Some Disjoint Spacing Tests Relative to a Chi-Square Test', in *Perspectives and New Directions in Theoretical and Applied Statistics*, eds. J. Vilaplana, and W. Wertz, New York: John Wiley, pp. 311–317.

Jammalamadaka, S.R., Zhou, X., and Tiwari, R.C. (1989), 'Asymptotic Efficiencies of Spacings Tests for Goodness of Fit', *Metrika*, 36, 355–377.

Kalandarov, U.N. (2001), 'About Asymptotic Expansion for Sum of Functions of Uniform Spacings', *Uzbek Mathematical Journal*, 10, 42–50.

Kallenberg, W.C.M. (1985), 'On Moderate and Large Deviations in Multinomial Distributions', *Annals of Statistics*, 13, 1554–1580.

Kuo, M., and Jammalamadaka, S.R. (1984), 'Limit Theory and Efficiencies for tests based on higher order spacings', in *Statistics: Applications and New Directions*, eds. J.K. Ghosh and J. Roy, Calcutta: Indian Statistical Institute, pp. 333–352.

Mirakhmedov, S.A. (2005), 'Lower Estimation of the Remainder Term in the CLT for a Sum of the Functions of k-Spacings', *Statistics & Probability Letters,* 73, 411–424.

Mirakhmedov, S.M. (2006), 'Probability of Large Deviations for the Sum of Functions of Spacings', *International Journal of Mathematics and Mathematical Sciences*, ID 58738, 1–22.

Mirakhmedov, S.M. (2010), 'On Greenwood Goodness-of-Fit Test', *Journal of Statistical Planning and Inference*, 140, 3017–3025, doi:10.1016/j.jspi.2010.04.002.

Mirakhmedov, S.M., and Naeem, M. (2008), 'Asymptotic Properties of the Goodness-of-Fit Tests Based on Spacings', *Pakistan Journal of Statistics*, 24, 253–268.

Mirakhmedov, S.M., Tirmizi, S.I., and Naeem, M. (2011), 'The Cramér-type Large Deviation Theorem for Sums of Functions of Higher Order Non-Overlapping Spacings', *Metrika*, 74, 33–54, doi:10.1007/s00184-009-0288-6.

Penev, S., and Ruderman, A. (2010), 'On the Behavior of Tests Based on Sample Spacings for Moderate Samples', *Journal of Statistical Planning and Inference*, 141, 1240–1249.

Pyke, R. (1965), 'Spacings', *Journal of the Royal Statistical Society, Series B*, 27, 395–449.

Rao, J.S. (1976), 'Some Tests Based on Arc-Lengths for the Circle', *Sankhya: The Indian Journal of Statistics, Series B*, 38, 329–338.

Rao, J.S., and Kuo, M. (1984), 'Asymptotic Results on the Greenwood Statistic and Some of its Generalizations', *Journal of the Royal Statistical Society, Series B*, 36, 226–237.
Rao, J.S., and Sethuraman, J. (1975), 'Weak Convergence of Empirical Distribution Functions of Random Variables Subject to Perturbations and Scale Factors', *Annals of Statistics*, 3, 299–313.
Rao, J.S., and Sobel, M. (1980), 'Incomplete Dirichlet Integrals with Applications to Ordered Uniform Spacings', *Journal of Multivariate Analysis,* 10, 603–610.
Sethuraman, J., and Rao, J.S. (1970), 'Pitman Efficiencies of Tests Based on Spacings', in *Nonparametric Techniques in Statistical Inference*, ed. M.L. Puri, Cambridge, UK: Cambridge University Press, pp. 405–415.
Zhou, X., and Jammalamadaka, S.R. (1989), 'Bahadur Efficiencies of Spacings Tests for Goodness of Fit', *Annals of the Institute of Statistical Mathematics*, 41, 541–553.

Appendix

Recall the notation used in Section 2, and the assumption made, namely that $E|g_m(Z_m)|^k < \infty$ for some integer $k \geq 3$ and $\rho_{k,N} = \sum_{m=1}^{N} E|\tilde{g}_m|^k$. Without loss of generality, we assume that $\rho_{k,N} \leq 1$. In the course of the proofs, we shall omit standard algebra as well as computational details of absolute constants, as we are interested only in the asymptotic results. Set

$$\mathbb{P}_{k,N}(t,\tau) = \mathrm{e}^{-(t^2+\tau^2)/2} \sum_{m=0}^{k-3} N^{-m/2} P_{m,N}(t,\tau),$$

where the polynomials $P_{m,N}(t,\tau)$ are defined just before Equation (13).

LEMMA A.1 *There exist constants* $c_1 > 0, c_2 > 0,$ *and* $C > 0$ *such that if* $|t| \leq c_1 \rho_{k,N}^{-1/k}$, $|\tau| \leq c_2 N^{(k-2)/2k}$, *then for* $l = 0, 1, \dots, k$

$$\left| \frac{\partial^l}{\partial t^l} (\Psi_N(t,\tau) - \mathbb{P}_{k,N}(t,\tau)) \right| \leq C\rho_{k,N}(1 + |t|^{k-l} + |\tau|^{k-l})\, \mathrm{e}^{-(t^2+\tau^2)/4}.$$

Lemma A.1 follows from Theorem 9.11 of BR because of Equation (6), and the facts that $\sum_{m=1}^{N} (\tilde{g}_m, \tilde{Z}_m)$ has unit correlation matrix and $\sum_{m=1}^{N} E|\tilde{Z}_m|^k < C(k)N^{-(k-2)/2}$.

LEMMA A.2 *There exist constants* $c_1 > 0, c_2 > 0,$ *and* $C > 0$ *such that if* $|t| \leq c_1 \rho_{3,N}^{-1}$ *and* $|\tau| \leq c_2 \sqrt{N}$ *then for* $l = 0, 1, \dots, k$

$$\left| \frac{\partial^l}{\partial t^l} \Psi_N(t,\tau) \right| \leq C(1 + |t| + |\tau|)^l \exp\left\{ -\frac{t^2+\tau^2}{6} \right\}.$$

Lemma A.2 can be proved by standard algebra outlined, e.g. in BR (pp. 67, 68, and 125–128).

LEMMA A.3 *There exist constants* $c_1 > 0$ *and* $C > 0$ *such that if* $|t| \leq c_1 \rho_{3,N}^{-1}$ *then for* $l = 0, 1, \dots, k$

$$\left| \frac{\partial^l}{\partial t^l} (\phi_N(t) - \Upsilon_{k,N}(t)) \right| \leq C\rho_{k,N}(1 + |t|^{k-l})\, e^{-t^2/2}.$$

Proof Let $|t| \leq c_1 \rho_{k,N}^{-1/k}$. Equations (8), (13), and (19) imply

$$\begin{aligned}
\left| \frac{\partial^l}{\partial t^l} (\Theta_N(t) - Q_{k,N}(t)) \right| &\leq \frac{1}{\sqrt{2\pi}} \int\limits_{|\tau| \leq c_2 N^{(k-2)/2k}} \left| \frac{\partial^l}{\partial t^l} (\Psi_N(t,\tau) - \mathbb{P}_{k,N}(t,\tau)) \right| \mathrm{d}\tau \\
&\quad + \frac{1}{\sqrt{2\pi}} \int\limits_{c_2 N^{(k-2)/2k} \leq |\tau| \leq c_2\sqrt{N}} \left| \frac{\partial^l}{\partial t^l} \Psi_N(t,\tau) \right| \mathrm{d}\tau + \frac{1}{\sqrt{2\pi}} \int\limits_{|\tau| \geq c_2 N^{(k-2)/2k}} \left| \frac{\partial^l}{\partial t^l} \Psi_N(t,\tau) \right| \mathrm{d}\tau \\
&\quad + \frac{1}{\sqrt{2\pi}} \int\limits_{|\tau| \geq c_2\sqrt{N}} \left| \frac{\partial^l}{\partial t^l} \Psi_N(t,\tau) \right| \mathrm{d}\tau \\
&\overset{\text{def}}{=} (\Im_1 + \Im_2 + \Im_3 + \Im_4).
\end{aligned} \tag{A1}$$

Note that $\rho_{k,N}^{-1/k} \leq \rho_{3,N}^{-1}$, since $\rho_{3,N} \leq \rho_{k,N}^{1/(k-2)} \leq \rho_{k,N}^{1/k}$. Applying Lemmas A.1 and A.2 and the definition of $\mathbb{P}_{k,N}(t,\tau)$ along with the inequalities (16) and (17), we show, after some simplification, that

$$\Im_1 + \Im_2 + \Im_3 \leq C\rho_{k,N}(1 + |t|^{k-l})\, \mathrm{e}^{-(t^2/6)}. \tag{A2}$$

Let $|\tau| \geq c_2\sqrt{N}$. We have

$$|\psi_m(t,\tau)| \leq |E\exp\{\mathrm{i}t\tilde{Z}_m\}(\exp\{\mathrm{i}t\tilde{g}_m\}-1) + E\exp\{\mathrm{i}t\tilde{Z}_m\}| \leq |t|E|\tilde{g}_m| + |E\exp\{\mathrm{i}t\tilde{Z}_m\}|$$
$$\leq \exp\{-(1-|E\exp\{\mathrm{i}\tau\tilde{Z}_m\}|) + |t|E|\tilde{g}_m|\} \leq \exp\{-c_3 + |t|E|\tilde{g}_m|\}, \tag{A3}$$

where we use the inequality $x \leq \mathrm{e}^{x-1}$ and the fact that $1-|E\exp\{\mathrm{i}\tau\tilde{Z}_m\}| = 1-(1+(\tau/\sqrt{Ns})^2)^{-s/2} \geq c_3$. Using the notation $\prod_{m=1}^{N\,(l)}$ for the product over $m = 1,2,\ldots,N$ with some l of the factors replaced by 1, we have

$$\prod_{m=1}^{N} |\psi_m(t,\tau)| \leq \mathrm{e}^{-c_4N}, \quad \prod_{m=1}^{N\,(l)} |\psi_m(t,\tau)| \leq \mathrm{e}^{l}\,\mathrm{e}^{-c_4N}, \tag{A4}$$

using $\sum_{m=1}^{N} E|\tilde{g}_m| \leq \sqrt{N}$, $|t| \leq c_1\rho_{k,N}^{-1/k} \leq c_1N^{(k-2)/2k}$, see Equation (16). We have

$$\left|\frac{\partial}{\partial t}\psi_m(t,\tau)\right| \leq t^2E\tilde{g}_m^2 + |t\tau|E|\tilde{g}_m\tilde{Z}_m| \leq 1.5t^2E\tilde{g}_m^2 + 0.5|\tau|E\tilde{Z}_m^2, \quad \left|\frac{\partial^l}{\partial t^l}\psi_m(t,\tau)\right| \leq E|\tilde{g}_m|^l, \quad l \geq 2. \tag{A5}$$

Using Leibniz's rule for differentiation of the product of N functions, and the relations (A4), (A5), and (9), we obtain after some algebra, similar to that in BR (pp. 127–128),

$$\Im_4 \leq c\exp\{-c_5N\}. \tag{A6}$$

Now, we use Equations (A1), (A2), and (A6) to obtain

$$\left|\frac{\partial^l}{\partial t^l}(\Theta_N(t) - Q_{k,N}(t))\right| \leq C\rho_{k,N}(1+|t|^{k-l})\,\mathrm{e}^{-t^2/6}. \tag{A7}$$

In particular, from this and Equation (20), we have

$$\Theta_N(0) = Q_{k,N}(0) + C\theta n^{-\lfloor (k-2)/2\rfloor} \geq c \quad \text{for some } c > 0. \tag{A8}$$

Equations (11), (A7), and (A8) imply Lemma A.3 after some simple algebra, for $|t| \leq c_1\rho_{k,N}^{-1/k}$.

Now let $c_1\rho_{k,N}^{-1/k} \leq |t| \leq c_1\rho_{3,N}^{-1}$. By virtue of Equations (A8) and (11), we have

$$\left|\frac{\partial^j}{\partial t^j}(\phi_N(t) - \Upsilon_{k,N}(t))\right| \leq C\left[\int_{|\tau|\leq c_2\sqrt{N}} \frac{\partial^j}{\partial t^j}\Psi_N(t,\tau)\,\mathrm{d}\tau + \Im_4\right] + \left|\frac{\partial^j}{\partial t^j}\Upsilon_{k,N}(t)\right|. \tag{A9}$$

Since $\sum_{m=1}^{N} E|\tilde{g}_m| \leq \sqrt{N}$, $|t| \leq c_1\sqrt{N}$, and $\rho_{3,N} \geq N^{-1/2}$, the inequalities in Equation (A4) still hold by choosing $c_1 \leq c_3/2$. Now we use Lemma A.2 and Equation (A6), respectively, for the two terms inside the square bracket in Equation (A9); the proof of Lemma A.3 is complete from Equation (19) and the facts that $|t| \geq c_1\rho_{k,N}^{-1/k}$ and the function $\Upsilon_{k,N}(t)$has the factor $\exp\{-t^2/2\}$. ■

Proof of Theorem 2.1 We have

$$\left|\phi_N(t) - \Upsilon_{k,N}(t)\right| \leq \left|\int_0^t \frac{\mathrm{d}}{\mathrm{d}u}(\phi_N(u) - \Upsilon_{k,N}(u))\,\mathrm{d}u\right| \leq |t|\sup_{|u|\leq|t|}\left|\frac{\mathrm{d}}{\mathrm{d}u}(\phi_N(u) - \Upsilon_{k,N}(u))\right|.$$

Using this and the well-known Esseen's smooth inequality (see, Feller 1971), we have

$$|P(\mathbb{T}_N(D) < x\sigma_N + A_N) - \Re_{k,N}(x)| \leq \frac{1}{\pi}\left[\int_{1\leq|t|\leq c_1\rho_{3,N}^{-1}} |\phi_N(t) - \Upsilon_{k,N}(t)|\,\mathrm{d}t + \int_{0\leq|t|\leq 1}\left|\frac{\mathrm{d}}{\mathrm{d}t}(\phi_N(t) - \Upsilon_{k,N}(t))\right|\mathrm{d}t\right]$$
$$+ \frac{1}{\pi}\int_{c_1\rho_{3,N}^{-1}\leq|t|\leq c_1\rho_{k,N}^{-1}} |t||\phi_N(t) - \Upsilon_{k,N}(t)|\,\mathrm{d}t + \frac{24}{c_1\sqrt{2\pi}}\rho_{k,N}$$
$$= \frac{1}{\pi}[J_0 + J_1] + \frac{1}{\pi}J_2 + \frac{24}{c_1\sqrt{2\pi}}\rho_{k,N}. \tag{A10}$$

By Lemma A.3, we obtain

$$J_0 + J_1 \leq C\rho_{k,N}. \tag{A11}$$

Using the inequality $x < \mathrm{e}^{x-1}$ for $c_1\rho_{3,N}^{-1}\sigma_N \leq |t| \leq c_1\rho_{k,N}^{-1}\sigma_N$, we have

$$\prod_{m=3}^{N} |E\exp\{\mathrm{i}tf_m(Z_m) + \mathrm{i}\tau Z_m\}| \leq \mathrm{e}^2\,\mathrm{e}^{-N(1-\chi_N)}, \tag{A12}$$

where χ_N is as defined in Equation (23). On the other hand, as in Equation (11), one can check that

$$\phi_N(t) = \frac{1}{2\pi\gamma_n(n)}\int_{-\infty}^{\infty} E\exp\{\mathrm{i}tf_m(Z_m) + \mathrm{i}\tau Z_m\}\,\mathrm{d}\tau. \tag{A13}$$

Using Equations (9), (A8), and (A12) and the definition of $\Upsilon_{k,N}(t)$, we show that

$$J_2 \leq C\sqrt{nN}\log\rho_{k,N}^{-1}\exp\{-N(1-\chi_N)\}.$$

Theorem 2.1 now follows from Equations (A10) and (A11), last inequality and fact that if $k = 3$ then $J_2 = 0$. ■

Proof of Theorem 2.2 Follows from Lemma A.3 by putting $l = 1$ and $l = 2$. ■

Proof of Theorem 2.3 Follows from Theorem 2.1 by putting $k = 4$ in view of Equations (9) and (16). ■

Proof of Remark 2.1 The proof follows from Lemma 3.1 of Does et al. (1987). ■

LEMMA A.4 *Let X be an r.v. taking values in R^m, the distribution of which is absolutely continuous in some Borel set B with $P\{X \in B\} > 0$. Let $\phi : R^m \to R^k$ be a measurable function, which is Lebesgue almost everywhere differentiable on B with the $k \times m$ matrix ϕ' as the differential. If all $\gamma \in R^k \backslash \{0\}$ satisfy $P\{(\phi')^{\mathrm{T}}\gamma = 0 | X \in B\} < 1$, then* $\limsup_{|v|\to\infty} |E \exp\{i(v^{\mathrm{T}} \cdot \phi(X))\}| < 1$.

Remark 2.1 follows by taking $m = 1$, $k = 2$, $\phi(x) = (x, g(x))$, and $B = (c, d)$ in Lemma A.4.

Proof of Theorem 2.4 Recall that Z is an r.v. with pdf $\gamma_s(u)$. For the Greenwood statistic, $f(x) = (x - s)^2$. By direct calculations, we find

$$Ef(Z) = s, \quad \sigma^2 = 2s(s+1), \quad g(Z) = (Z - s)^2 - 2(Z - s) - s,$$

$$\frac{Eg^3(Z)}{\sigma^3} = \frac{2\sqrt{2}(s^2 + 5s + 4)}{(s+1)^{3/2}\sqrt{s}} = 2\sqrt{2}\left(1 + \frac{7}{2s} + O\left(\frac{1}{s^2}\right)\right), \tag{A14}$$

$$\frac{Eg(Z)Z^2}{s\sigma} = \sqrt{2\left(1 + \frac{1}{s}\right)} = \sqrt{2}\left(1 + \frac{1}{2s} + O\left(\frac{1}{s^2}\right)\right), \tag{A15}$$

$$Eg^4(Z) = 4s(15s^3 + 222s^2 + 579s + 372).$$

Note that the condition (28) is not satisfied in this case; hence Remark 2.1 is not applicable. Nevertheless, one can prove the following. ■

LEMMA A.5 *If $|t|s \geq c > 0$, then $|E\, \mathrm{e}^{\mathrm{i}tZ^2 + \mathrm{i}\tau Z}| \leq c_1 < 1$.*

Proof Let Q_s denote the distribution with pdf $\gamma_s(u)$ and $\Phi_{s,s}$ the normal distribution with both expectation and variance equal to s. Let $T(s)$ be the total variance distance between Q_s and $\Phi_{s,s}$. One can then construct r.v.s Z and X on the same probability space, having distributions Q_s and $\Phi_{s,s}$, respectively, in such a way that $P\{Z \neq X\} = T(s)$. Now

$$\begin{aligned} E\, \mathrm{e}^{\mathrm{i}tZ^2 + \mathrm{i}\tau Z} &= E(\mathrm{e}^{\mathrm{i}tX^2 + \mathrm{i}\tau X}; Z = X) + E(\mathrm{e}^{\mathrm{i}tZ^2 + \mathrm{i}\tau Z}; Z \neq X) \\ &= E(\mathrm{e}^{\mathrm{i}tX^2 + \mathrm{i}\tau X}) - E(\mathrm{e}^{\mathrm{i}tX^2 + \mathrm{i}\tau X}; Z \neq X) + E(\mathrm{e}^{\mathrm{i}tZ^2 + \mathrm{i}\tau Z}; Z \neq X). \end{aligned}$$

Hence $|E\, \mathrm{e}^{\mathrm{i}tZ^2 + \mathrm{i}\tau Z}| \leq |E\, \mathrm{e}^{\mathrm{i}tX^2 + \mathrm{i}\tau X}| + 2T(s)$. It is easy to check that $T(s) = O(s^{-1/2})$ since Q_s is the distribution of a sum of s independent standard exponential r.v.s. On the other hand,

$$E\, \mathrm{e}^{\mathrm{i}tX^2 + \mathrm{i}\tau X} = \frac{1}{\sqrt{1 - \mathrm{i}2ts}} \exp\left\{-\frac{s(\tau + 2st)^2}{2(1 - \mathrm{i}2ts)}\right\}, \quad \text{i.e. } |E\, \mathrm{e}^{\mathrm{i}tX^2 + \mathrm{i}\tau X}| \leq c_1 < 1 \quad \text{as } |t|s \geq c > 0,$$

from which Lemma A.5 follows.

Note that $|E \exp\{\mathrm{i}tG_N^2\}| = |E \exp\{\mathrm{i}t((nD_{1,s})^2 + \cdots + (nD_{N,s})^2)\}|$, so that Theorem 2.4 follows now from Theorem 2.3 with $f(x) = x^2$, and Lemma A.5.

To begin the proof of Theorems 2.5–2.8, we first note that under conditions of these theorems, the corresponding statistics satisfy condition (27), so that χ defined in Equation (25) is <1. ■

Proof of Theorem 2.5 For the Log-spacings statistic, we have $f(x) = \log x$. For integers $m \geq 1$ and $k \geq 1$, note that for the digamma function $\psi(m)$ and the Hurwitz zeta function $\zeta(m, k)$, we have (see Abramowitz and Stegun 1972, p. 261)

$$\psi^{(n)}(s) = (-1)^{n-1}\left[\frac{(n-1)!}{s^n} + \frac{n!}{2s^{n+1}} + \sum_{k=1}^{\infty} B_{2k}\frac{(2k+n-1)!}{(2k)!s^{2k+n}}\right] \tag{A16}$$

and

$$\zeta(n+1, s) = \frac{1}{n!}\left[\frac{(n-1)!}{s^n} + \frac{n!}{2s^{n+1}} + \sum_{k=1}^{\infty} B_{2k}\frac{(2k+n-1)!}{(2k)!s^{2k+n}}\right].$$

In particular,

$$\zeta(2, s) = \frac{1}{s} + \frac{1}{2s^2} + \frac{1}{6s^3} + O\left(\frac{1}{s^4}\right), \quad \zeta(3, s) = \frac{1}{2s^2} + \frac{1}{2s^3} + \frac{1}{4s^4} + O\left(\frac{1}{s^5}\right) \quad \text{as } s \to \infty. \tag{A17}$$

Using formula (4.358) of Gradshteyn and Ryzhik (2000, p. 572) and Equations (A16) and (A17), we find

$$Ef(Z) = E \ln Z = \psi(s), \quad g(Z) = \ln Z - \psi(s) - (Z - s)s^{-1},$$

$$\sigma^2 = \zeta(2, s) - s^{-1} = \frac{1}{2s^2}(1 + O(s^{-1})), \tag{A18}$$

$$Eg^3(Z) = s^{-2} - 2\zeta(3, s) = -\frac{1}{s^3}(1 + O(s^{-1})), \quad Eg(Z)(Z - s)^2 = -1, \tag{A19}$$

$$Eg^4(Z) = 3s^{-2} - 2s^{-3} - 6s^{-1}\zeta(2, s) + 3\zeta^2(2, s) + 6\zeta(4, s) = \frac{15}{4s^4}(1 + O(s^{-1})).$$

These equalities and Theorem 2.4 imply Theorem 2.8. ■

Proof of Theorem 2.6 In this case, $f(x) = x \ln x$. Using formulas (A13) and (A14), one can check that $Ef(Z) = EZ \ln Z = s\psi(s + 1)$, $g(Z) = Z \ln Z - Z\psi(s + 1) - (Z - s)$,

$$\sigma^2 = s(s + 1)\zeta(2, s + 1) - s = \tfrac{1}{2}(1 + O(s^{-1})), \quad Eg(Z)(Z - s)^2 = 3s, \tag{A20}$$

$$\begin{aligned} Eg^3(Z) &= 7s - 2s(s + 1)(s + 2)\zeta(3, s + 2) - \frac{6s(s + 2)}{s + 1} \\ &\quad + 3s\zeta(2, s + 2) + \frac{s(2s + 3)^3}{(s + 1)^2(s + 2)^2} - \frac{s(2s + 3)}{(s + 2)^2}, \end{aligned} \tag{A21}$$

$$Eg^4(Z) = O(1).$$

Applying these in Theorem 2.3 completes the proof of Theorem 2.6. ■

Proof of Theorem 2.7 Now let $f(x) = x^\gamma$, with $\gamma(\gamma - 1) \neq 0$. Put $q(u) = \Gamma(s + u)/\Gamma(s)$. Using the well-known expansion (see Abramowitz and Stegun 1972, p. 258)

$$\ln \Gamma(\kappa) = \left(\kappa - \frac{1}{2}\right) \ln \kappa - \kappa + \frac{\ln(2\pi)}{2} + \frac{1}{12\kappa} + O\left(\frac{1}{\kappa^3}\right) \quad \text{as } \kappa \to \infty,$$

we have $Ef(Z) = q(\gamma)$, $g(Z) = Z^\gamma - q(\gamma) - (Z - s)(\gamma/s)q(\gamma)$,

$$\sigma^2 = q(2\gamma) - \left(1 + \frac{\gamma^2}{s}\right) q^2(\gamma) = \frac{\gamma^2(\gamma - 1)^2}{2s^{2(1-\gamma)}} \left(1 + \frac{4\gamma^2 - 7\gamma + 1}{3s} + O\left(\frac{1}{s^2}\right)\right), \tag{A22}$$

$$\frac{Eg(Z)Z^2}{s\sigma} = \frac{\gamma(\gamma - 1)}{s\sigma} q(\gamma) = \sqrt{2}\left(1 + \frac{\gamma(\gamma - 1)}{2s} + O\left(\frac{1}{s^2}\right)\right), \tag{A23}$$

$$\begin{aligned} Eg^3(Z) &= q(3\gamma) - 3q(\gamma)q(2\gamma)\left(1 + \frac{2\gamma^2}{s}\right) + 2q^3(\gamma)\left(1 + \frac{3\gamma^2}{s} + \frac{3\gamma^3(\gamma + 1)}{2s^2}\right), \\ &= \frac{\gamma^2(\gamma - 1)^2(\gamma^2 + 3\gamma - 2)}{4s^{3(1-\gamma)}}\left(1 + O\left(\frac{1}{s}\right)\right). \end{aligned}$$

Hence,

$$\frac{Eg^3(Z)}{\sigma^3} = \frac{\sqrt{2}(\gamma^2 + 3\gamma - 2)}{2\gamma(\gamma - 1)}\left(1 + O\left(\frac{1}{s}\right)\right). \tag{A24}$$

Also it can be checked that $Eg^4(Z) = O(s^{-4(1-\gamma)})$ and that the condition (28) is fulfilled if $\gamma < 1$ or s, the step of spacings, is fixed. Theorem 2.7 follows. ■

Proof of Theorem 2.8 In this case, $f(x) = |x - s|$. By direct calculation, one can check the following:

$$E|Z - s| = a(s) = \sqrt{\frac{2}{\pi s}}\left(1 + O\left(\frac{1}{s}\right)\right), \quad \text{cov}(f(z), Z) = sb(s) = O(1),$$

$$\text{var}|Z - s| = s(1 - 4s\gamma_s^2(s)) = s\frac{\pi - 2}{\pi}\left(1 + O\left(\frac{1}{s}\right)\right), \quad \sigma^2 = \sigma^2(s) = s\frac{\pi - 2}{\pi}\left(1 + O\left(\frac{1}{s}\right)\right),$$

$$Eg(Z)(Z - s)^2 = E|Z - s|(Z - s)^2 = d(s) = \frac{2\sqrt{2}}{\sqrt{\pi}}s^{3/2}\left(1 + O\left(\frac{1}{s}\right)\right),$$

and $Eg^3(z) = \kappa(s) = 2\sqrt{2/\pi}\, s^{3/2}(1 + O(s^{-1}))$. Hence

$$\frac{Eg^3(Z)}{\sigma^3} = \frac{\sqrt{2}\pi}{(\pi - 2)^{3/2}}\left(1 + O\left(\frac{1}{s}\right)\right) \quad \text{and} \quad \frac{Eg(Z)Z^2}{s\sigma} = \frac{2\sqrt{2}}{\sqrt{\pi - 2}}\left(1 + O\left(\frac{1}{s}\right)\right).$$

Also one can check that $Eg^4(Z)/\sigma^4 = O(1)$, so that Theorem 2.8 follows. ■

Proof of Theorem 3.1 We note that, without loss of generality, one can assume that $\mathrm{cov}(f(Z), Z) = 0$. Otherwise, we can take the function $\tilde{f}(x) = f(x) - s^{-1}\mathrm{cov}(f(Z), Z)(x - s)$ instead of $f(x)$, since $Ef(Z) = E\tilde{f}(Z)$ and $\mathbb{S}_N(W) = \tilde{f}(nW_{1,s}) + \cdots + \tilde{f}(nW_{N,s})$. Similar to the relation (3.8) of Rao and Sethuraman (1975) (see also Equation (4.4) of Kuo and Jammalamadaka 1984), it can be checked that under alternatives (29)

$$W_{m,s} = D_{m,s}(1 - \varepsilon_{m,n}), \tag{A25}$$

where

$$\varepsilon_{m,n} = l\left(\frac{m}{N}\right)\frac{1}{(ns)^{1/4}} - \left[L\left(\frac{m}{N}\right)l'\left(\frac{m}{N}\right) + l^2\left(\frac{m}{N}\right)\right]\frac{1}{(ns)^{1/2}} + r_n,$$

and $(ns)^{3/4}r_n = O(1)$ almost everywhere uniformly in m. We have

$$\int_0^1 l(x)\,\mathrm{d}x = 0, \quad \int_0^1 (L(x)l'(x) + l^2(x))\,\mathrm{d}x = 0, \quad \int_0^1 L(x)l(x)l'(x)\,\mathrm{d}x = -\frac{1}{2}\int_0^1 l^3(x)\,\mathrm{d}x.$$

Using these facts, for any function $\vartheta(x) \in C^{(2)}$, we obtain

$$\frac{1}{N}\sum_{m=1}^{N} E\vartheta(Z(1-\varepsilon_{m,n})) = E\vartheta(Z) + \frac{\|l\|_2^2}{2\sqrt{ns}}EZ^2\vartheta''(Z)(1+o(1)) \tag{A26}$$

and for any function $\vartheta(x) \in C^{(3)}$, we have

$$\frac{1}{N}\sum_{m=1}^{N} E\vartheta(Z(1-\varepsilon_{m,n})) = E\vartheta(Z) + \frac{\|l\|_2^2}{2\sqrt{ns}}EZ^2\vartheta''(Z) - \frac{\|l\|_3^3}{6(ns)^{3/4}}EZ^3\vartheta'''(Z)(1+o(1)). \tag{A27}$$

Integrating by parts, we can see that $EZ\vartheta'(Z) = \mathrm{cov}(\vartheta(Z), Z)$, $EZ^2\vartheta''(Z) = \mathrm{cov}(\vartheta(Z), \mathbb{Z}_2)$, and $EZ^3\vartheta'''(Z) = \mathrm{cov}(\vartheta(Z), \mathbb{Z}_3)$. Using these relations and Equations (A26) and (A27) with $g(x)$, $g^2(x)$, $g^3(x)$, and $g(x)(x-s)^2$, respectively, in place of $\vartheta(x)$, we find: if $f(x) \in C^{(2)}$ and $E|g(Z)|^3 < \infty$, then

$$\frac{1}{N}\sum_{m=1}^{N} Eg(Z(1-\varepsilon_{m,n})) = \frac{\|l\|_2^2}{2\sqrt{ns}}\mathrm{cov}(g(Z), \mathbb{Z}_2)(1+o(1)),$$

$$\begin{aligned}\sigma_N^2 &= \sum_{m=1}^{N} Eg^2(Z(1-\varepsilon_{m,n})) \\ &= N\sigma^2\left(1 + \frac{\|l\|_2^2}{2\sigma^2\sqrt{ns}}Eg^2(Z)\mathbb{Z}_2(1+o(1))\right) \\ &= N\sigma^2(1 + C\theta\rho_{3,N}^{2/3}N^{-1/6}),\end{aligned}$$

and if $f(x) \in C^{(3)}$ and $E|g(Z)|^4 < \infty$

$$\frac{1}{\sigma_N^3}\sum_{m=1}^{N} Eg^3(Z(1-\varepsilon_{m,n})) = Eg^3(Z)(1 + \|l\|_2^2\,(ns)^{-1/2}\mathrm{Z}_2) = Eg^3(Z) + C\theta\rho_{4,N}^{3/4}N^{-1/4}, \tag{A28}$$

$$\frac{1}{\sigma_N ns}\sum_{m=1}^{N} Eg(Z(1-\varepsilon_{m,n}))(Z-s)^2 = \frac{Eg(Z)(Z-s)^2}{\sigma\sqrt{N}s} + C\theta\rho_{4,N}^{1/6}N^{-5/6}, \tag{A29}$$

where $\mathbb{Z}_2$ and $\mathbb{Z}_3$ are as in Equation (30) and $\rho_{j,N} = E|g(Z)|^j/N^{(j-2)/2}\sigma^j$; also a bound for the remainder term is obtained using Holder's inequalities. Noting that $\mathrm{Var}\,\mathbb{Z}_2 = 2s(s+1)$ and $\mathrm{Var}\,\mathbb{Z}_3 = 6s(s+1)(s+2)$, we write

$$\frac{EZ^2f''(Z)}{2\sqrt{ns}} = \frac{E(\mathbb{Z}_2 f(Z))}{2\sqrt{ns}} = (2N)^{-1/2}\sigma\,\mu(s,f)\sqrt{1+s^{-1}},$$

$$\frac{EZ^3f'''(Z)}{6(ns)^{3/4}} = \frac{E(\mathbb{Z}_3 f(Z))}{6(ns)^{3/4}} = \frac{\sigma\,\nu(s,f)}{N^{3/4}\sqrt{6}}\sqrt{\left(1+\frac{1}{s}\right)\left(1+\frac{2}{s}\right)}.$$

On using Equation (A26) with $\vartheta(x) = f(x)$, we find

$$\begin{aligned}\sum_{m=1}^{N} Ef(Z(1-\varepsilon_{m,n})) &= NEf(Z) + \frac{\|l\|_2^2\,\sigma\sqrt{N(s+1)}\mu(s,f)}{\sqrt{2s}} \\ &\quad - \|l\|_3^3\,N^{1/4}\sigma\,\nu(s,f)\sqrt{\frac{1}{6}\left(1+\frac{1}{s}\right)\left(1+\frac{2}{s}\right)}(1+o(1)).\end{aligned}$$

Let P_i and E_i, $i = 0, 1$, denote the probability and expectation under the hypothesis and under the alternative (29), respectively. From Corollary 2.1 and Equation (A25), one can check that

$$E_0 S_N(W) = \sum_{m=1}^{N} Ef(Z) + c\theta\sigma\sqrt{N}\rho_{3,N}, \quad E_1 S_N(W) = \sum_{m=1}^{N} Ef(Z(1-\varepsilon_{m,n})) + c\theta\sigma\sqrt{N}\rho_{3,N},$$

$$\mathrm{Var}_0\, S_N(W) = N\sigma^2(1 + c\theta\rho_{3,N}), \quad \mathrm{Var}_1\, S_N(W) = N\sigma^2(1 + c\theta\rho_{3,N}).$$

Hence,

$$\Delta_N = \frac{E_1 S_N(W) - E_0 S_N(W)}{\mathrm{Var}_1 S_N(W)} = \|l\|_2^2\, \mu(s,f)\sqrt{\frac{(s+1)}{2s}} - \|l\|_3^3\, \frac{\nu(s,f)}{N^{1/4}}\sqrt{\frac{1}{6}\left(1+\frac{1}{s}\right)\left(1+\frac{2}{s}\right)}(1+o(1)). \tag{A30}$$

Finally, the sequence of c.f.s $E\exp\{\mathrm{i}tg(Z(1-\varepsilon_{m,n})) + \mathrm{i}\tau Z\}$ converges to the c.f. $E\ \exp\{\mathrm{i}tg(Z) + \mathrm{i}\tau Z\}$ uniformly in t and τ, so that as $n \to \infty$

$$\sup_{(t,\tau)\in\tilde{\Omega}} |E\exp\{\mathrm{i}tg(Z(1-\varepsilon_{m,n})) + \mathrm{i}\tau Z\}| = \chi + o(1), \tag{A31}$$

where $\tilde{\Omega} = \{(t,\tau) : c\sigma^2/E|g(Z(1-\varepsilon_{m,n})|^3 \le |t| \le c\sigma^3\sqrt{N}/E|g(Z(1-\varepsilon_{m,n})|^4, \tau \in (-\infty,\infty)\}$.

Note that in view of Corollary 2.1, the level-ω critical value of the f-test is equal to $c_\omega = u_\omega\sigma\sqrt{N} + NEf(Z)$, $u_\omega = \Phi^{-1}(1-\omega)$. On the other hand, from Equation (A25),

$$P_1\{S_N(W) > c_\omega\} = P\left\{\sum\nolimits_{m=1}^{N} f_m(D_{m,s}(1-\varepsilon_{m,n})) > c_\omega\right\}.$$

Theorem 3.1 now follows from Corollary 2.1, Theorem 2.3, relations (A28)–(A31), and the fact that

$$\Phi(\Delta_N - u_\omega\sqrt{(1+c\theta\rho_{3,N})^{-1}}) = \Phi(\Delta_\omega(l,s,f)) - \varphi(\Delta_\omega(l,s,f))\, \|l\|_3^3\, \frac{\nu(s,f)}{N^{1/4}}\sqrt{\frac{1}{6}\left(1+\frac{1}{s}\right)\left(1+\frac{2}{s}\right)}(1+o(1)).$$

Equations (38)–(41) follow from Equations (34)–(37), respectively, using Equations (A14), (A15), (A18)–(A24), and Theorem 3.1. ■