Journal of Statistical Planning and Inference 142 (2012) 673–684

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

U-Statistics based on spacings

David D. Tung [a,*], S. Rao Jammalamadaka [b]

[a] *ALEKS Corporation, Irvine, CA 92618, USA*
[b] *Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106-3110, USA*

ARTICLE INFO

Article history:
Received 18 June 2010
Received in revised form
6 September 2011
Accepted 6 September 2011
Available online 22 September 2011

Keywords:
Goodness-of-fit tests
Order statistics
Gini mean difference
Pitman asymptotic relative efficiency

ABSTRACT

In this paper, we investigate the asymptotic theory for *U*-statistics based on sample spacings, i.e. the gaps between successive observations. The usual asymptotic theory for *U*-statistics does not apply here because spacings are dependent variables. However, under the null hypothesis, the uniform spacings can be expressed as conditionally independent Exponential random variables. We exploit this idea to derive the relevant asymptotic theory both under the null hypothesis and under a sequence of close alternatives.

The generalized Gini mean difference of the sample spacings is a prime example of a *U*-statistic of this type. We show that such a Gini spacings test is analogous to Rao's spacings test. We find the asymptotically locally most powerful test in this class, and it has the same efficacy as the Greenwood statistic.

1. Introduction

The simple goodness-of-fit problem consists of testing fit to a single fixed distribution for a given data set. In particular, consider a random sample $X_1, X_2, \ldots, X_{n-1}$ with distribution function F defined on the real line $\mathbb{R}$. In the statistical literature, much attention has been devoted to the nonparametric problem of simple goodness-of-fit, namely testing the null hypothesis

$$H_0 : F(x) = F_0(x),$$

where F_0 is a completely specified distribution function.

If F is assumed to be continuous as we shall do, by way of the probability integral transform, the support of F reduces to the unit interval [0, 1], and this also permits us to equate F_0 with the Uniform ([0, 1]) distribution. Thus, the goodness-of-fit problem reduces to one of testing uniformity, i.e. testing the null hypothesis

$$H_0 : F(x) = x \quad \text{for } 0 \le x \le 1.$$

Let $X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}$ denote the sample order statistics with support given on the unit interval [0, 1]. We put $X_{(0)} \equiv 0$ and $X_{(n)} \equiv 1$, so that $0 = X_{(0)} \le X_{(1)} \le X_{(2)} \le \cdots \le X_{(n-1)} \le X_{(n)} = 1$. The *sample spacings* are defined by the random variables

$$D_k = X_{(k)} - X_{(k-1)} \quad \text{for } k = 1, 2, \ldots, n. \tag{1.1}$$

If F is the Uniform ([0, 1]) distribution, as under the null hypothesis, we use the special notation $\{U_k\}$ for the sample observations, and

$$T_k = U_{(k)} - U_{(k-1)} \quad \text{for } k = 1, 2, \ldots, n \tag{1.2}$$

for the *uniform spacings*. Tests based on spacings are studied here for testing the null hypothesis.

* Corresponding author.
E-mail addresses: david.deming.tung@gmail.com (D.D. Tung), rao@pstat.ucsb.edu (S. Rao Jammalamadaka).

0378-3758/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2011.09.007

Most common among spacings tests are symmetric spacings tests, i.e. general test statistics of the form

$$V_n(g) = \frac{1}{n}\sum_{k=1}^{n} g(nD_k), \tag{1.3}$$

and

$$W_n(h) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h(nD_i, nD_j), \tag{1.4}$$

where $g(\cdot)$ is a real-valued function satisfying some regularity conditions and $h : [0,\infty) \times [0,\infty) \to \mathbb{R}$ is a symmetric function satisfying some other regularity conditions. Test statistics of the form $V_n(g)$ are symmetric sum-functions of the sample spacings (e.g. cf. Pyke, 1965; Sethuraman and Rao, 1970; Rao and Sethuraman, 1975), i.e. they are symmetric in $\{D_k\}$, and can also be thought of as first-order U-statistics of the sample spacings. The asymptotic theory for symmetric sum-functions of the spacings has been studied in Sethuraman and Rao (1970) and Rao and Sethuraman (1975) via weak convergence of the empirical spacings process. They show that symmetric sum-functions based on these sample spacings cannot discriminate alternatives converging to the null hypothesis at a rate faster than $n^{-1/4}$ and hence have poor asymptotic performance as compared to say the Kolmogorov–Smirnov test. On the other hand, test statistics of the form $W_n(h)$ are second-order U-statistics of the sample spacings and symmetric in the pairs (D_i, D_j). An important example of such a statistic is the generalized Gini mean difference of the sample spacings, given by

$$G_n(r) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} |nD_i - nD_j|^r = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} |nD_i - nD_j|^r}{n(n-1)}, \quad r > 0, \tag{1.5}$$

which is an average over all pairs of absolute pairwise differences of the sample spacings to the rth power. The special case of $G_n(1)$ is the Gini mean difference spacings test, which was proposed in Jammalamadaka and Goria (2004) for testing goodness-of-fit. There they derive both the exact and asymptotic distribution of $G_n(1)$ under the null hypothesis, and show that it has good performance based on Monte Carlo powers. The special case of $G_n(2)$ will be called the Gini mean squared difference spacings test.

Among spacings tests of the form $V_n(g)$, the most common test statistics are

$$\frac{1}{n}\sum_{k=1}^{n} (nD_k)^2, \tag{1.6}$$

$$\frac{1}{n}\sum_{k=1}^{n} \log(nD_k), \tag{1.7}$$

$$\frac{1}{n}\sum_{k=1}^{n} (nD_k)\log(nD_k), \tag{1.8}$$

$$J_n(r) = \frac{1}{n}\sum_{k=1}^{n} |nD_k - 1|^r, \quad r > 0. \tag{1.9}$$

The test statistics (1.6), (1.7), (1.8) and (1.9) are, respectively, the Greenwood spacings test, Darling's log-spacings test, the Kullback–Leibler divergence (relative entropy) of the spacings, and the generalized Rao's spacings test. The generalized Rao's spacings test $J_n(r)$ is an average of the absolute deviations of the spacings to the rth power, and in a sense is analogous to the generalized Gini mean difference spacings test $G_n(r)$. The special case of $J_n(1)$ is the classical Rao's spacings test (cf. Rao, 1969), which can also be used to test the uniformity of circular data. Note that the special case of $J_n(2)$, corresponds to both the Greenwood statistic based on the sum of squares of the spacings, and also the Gini mean squared difference spacings test $G_n(2)$. The test statistic (1.7) was proposed by Greenwood (1946) and will be called here and throughout the Greenwood statistic. The importance of the Greenwood statistic is somewhat justified in view of the result established in Sethuraman and Rao (1970) that among the class of symmetric sum-functions of the spacings, the Greenwood statistic is the asymptotically locally most powerful (ALMP) test.

The asymptotic distribution for second-order U-statistics of the sample spacings under the null hypothesis is studied in the next section. Section 3 deals with their asymptotic behavior under a sequence of close alternatives. Section 4 contains results on the ALMP test for the class of second-order U-statistics of the sample spacings. Section 5 features examples.

2. The asymptotic null distribution

In this section, we obtain the asymptotic distribution for second-order U-statistics of the uniform spacings, i.e. general test statistics of the form

$$W_n(h) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h(nT_i, nT_j), \tag{2.1}$$

under the null hypothesis. Under the null hypothesis, the uniform spacings have the well-known conditional representation

$$(nT_1, nT_2, \ldots, nT_n) \simeq \left(\frac{Z_1}{\overline{Z}_n}, \frac{Z_2}{\overline{Z}_n}, \ldots, \frac{Z_n}{\overline{Z}_n}\right) \simeq (Z_1, Z_2, \ldots, Z_n \,|\, \overline{Z}_n = 1), \tag{2.2}$$

where $Z_1, Z_2, \ldots, Z_n$ are independent Exponential(1) random variables (e.g. see Wilks, 1962, Section 7.7) and $\overline{Z}_n = n^{-1} \sum_{j=1}^{n} Z_j$ denotes their sample average. Here as elsewhere, we use $\simeq$ to denote the distributional equivalence of quantities on the left and right hand sides of the symbol.

There are at least two known approaches to deriving the asymptotic distribution for second-order U-statistics of the uniform spacings under the null hypothesis. One approach is by applying a conditional limit theorem for U-statistics due to Holst (1981, Theorem 6.2). A second approach is by way of the well-known Hoeffding decomposition for U-statistics, which connects the asymptotic theory for U-statistics of the uniform spacings with the asymptotic theory for symmetric sum-functions of the uniform spacings.

Let

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h(Z_i, Z_j), \tag{2.3}$$

be a second-order U-statistic based on the independent Exponential(1) random variables $Z_1, Z_2, \ldots, Z_n$, where the kernel $h : [0,\infty) \times [0,\infty) \to \mathbb{R}$ is a symmetric function with $\mathrm{Var}[h(Z_1, Z_2)] < \infty$. For the case of independent and identically distributed random variables, the Hoeffding decomposition (cf. Lee, 1990, Section 1.6) asserts that a U-statistic of order k is a linear combination of uncorrelated U-statistics of order $1, 2, \ldots, k$. The case for $k=2$ has been most studied and best understood. We state the Hoeffding decomposition for U_n in the following:

Lemma 1. *The Hoeffding decomposition of U_n has the form*

$$U_n = \theta + \frac{2}{n} \sum_{k=1}^{n} h^{(1)}(Z_k) + \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h^{(2)}(Z_i, Z_j), \tag{2.4}$$

where

$$\theta = \mathbb{E}[h(Z_1, Z_2)], \tag{2.5}$$

$$g(t) = \mathbb{E}[h(t, Z_2)] = \mathbb{E}[h(Z_1, Z_2) \,|\, Z_1 = t], \tag{2.6}$$

$$h^{(1)}(t) = g(t) - \theta, \tag{2.7}$$

$$h^{(2)}(z_1, z_2) = h(z_1, z_2) - g(z_1) - g(z_2) + \theta. \tag{2.8}$$

Moreover, the normalized U-statistic

$$\sqrt{n}(U_n - \theta) = \frac{2}{\sqrt{n}} \sum_{k=1}^{n} [g(Z_k) - \theta] + n^{1/2} R_n, \tag{2.9}$$

where

$$n^{1/2} R_n = \frac{2\sqrt{n}}{n(n-1)} \sum_{1 \le i < j \le n} h^{(2)}(Z_i, Z_j) = o_P(1). \tag{2.10}$$

The following result provides identities for the expectations, variances, and covariances of the U-statistic kernels in the Hoeffding decomposition. The proofs of these are relatively straightforward and rely on elementary properties of conditional expectation.

Lemma 2. *Let Z_1, Z_2 and Z_3 be independent Exponential(1) random variables, and let g be defined as in Lemma 1. Then*

$$\mathbb{E}[h(Z_1, Z_2)] = \mathbb{E}[g(Z_1)], \tag{2.11}$$

$$\mathrm{cov}[h(Z_1, Z_2), h(Z_1, Z_3)] = \mathrm{var}[g(Z_1)], \tag{2.12}$$

$$\mathrm{Cov}[h(Z_1,Z_2),Z_1] = \mathrm{Cov}[g(Z_1),Z_1], \tag{2.13}$$

$$\mathrm{Cov}[h(Z_1,Z_2),(Z_1-2)^2+(Z_2-2)^2] = 2\cdot\mathrm{Cov}[g(Z_1),(Z_1-2)^2]. \tag{2.14}$$

The next result states the asymptotic null distribution for symmetric sum-functions of the uniform spacings (cf. Sethuraman and Rao, 1970), and together with Lemmas 1 and 2, these results will help establish the main result of this section. We use the notation $\mathrm{N}_1(\mu,\sigma^2)$ to denote the one-dimensional Normal distribution with mean μ and variance σ^2.

Lemma 3. *Under the null hypothesis, in the limit as* $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}[g(nT_k)-\mathbb{E}g(Z_1)] \xrightarrow{D} \mathrm{N}_1(0,\sigma^2(g)), \tag{2.15}$$

where

$$\sigma^2(g) = \mathrm{Var}[g(Z_1)] - \frac{\mathrm{Cov}^2[g(Z_1),Z_1]}{\mathrm{Var}[Z_1]}. \tag{2.16}$$

By combining Lemmas 1, 2 and 3, we arrive at the main result of this section.

Theorem 4. *Under the null hypothesis, in the limit as* $n\to\infty$,

$$\sqrt{n}\left(\frac{2}{n(n-1)}\sum_{1\le i<j\le n} h(nT_i,nT_j) - \mathbb{E}[h(Z_1,Z_2)]\right) \xrightarrow{D} \mathrm{N}_1(0,\sigma^2(h)), \tag{2.17}$$

where

$$\sigma^2(h) = 4(\sigma_1^2-\sigma_{12}^2), \tag{2.18}$$

$$\sigma_1^2 = \mathrm{Cov}[h(Z_1,Z_2),h(Z_1,Z_3)], \tag{2.19}$$

$$\sigma_{12}^2 = \frac{\mathrm{Cov}^2[h(Z_1,Z_2),Z_1]}{\mathrm{Var}[Z_1]}. \tag{2.20}$$

Proof. From Lemma 1, and the conditional representation of the uniform spacings (2.2), we have

$$\begin{aligned}\sqrt{n}\left(\frac{2}{n(n-1)}\sum_{1\le i<j\le n} h(nT_i,nT_j) - \mathbb{E}[h(Z_1,Z_2)]\right) &\simeq (\sqrt{n}(U_n-\theta)\,|\,\overline{Z}_n = 1)\\ &\simeq \left(\frac{2}{\sqrt{n}}\sum_{k=1}^{n}[g(Z_k)-\mathbb{E}g(Z_1)]\,\middle|\,\overline{Z}_n = 1\right) + \mathrm{o_P}(1) \simeq \frac{2}{\sqrt{n}}\sum_{k=1}^{n}[g(nT_k)-\mathbb{E}g(Z_1)] + \mathrm{o_P}(1) \xrightarrow{D} \mathrm{N}_1(0,4\sigma^2(g)).\end{aligned}$$

The convergence in distribution to the $\mathrm{N}_1(0,4\sigma^2(g))$ distribution follows from Lemma 3 and Slutsky's Theorem. By Lemma 2, the asymptotic variance

$$4\sigma^2(g) = 4\left(\mathrm{Var}[g(Z_1)] - \frac{\mathrm{Cov}^2[g(Z_1),Z_1]}{\mathrm{Var}[Z_1]}\right) = 4(\sigma_1^2-\sigma_{12}^2) = \sigma^2(h).$$

This completes the proof. □

3. The asymptotic distribution under a sequence of close alternatives

In this section, we derive the asymptotic distribution for second-order *U*-statistics of the sample spacings, i.e. general test statistics of the form

$$W_n(h) = \frac{2}{n(n-1)}\sum_{1\le i<j\le n} h(nD_i,nD_j), \tag{3.1}$$

under a sequence of close alternatives. In order to study asymptotic efficiencies, one needs to obtain the asymptotic distribution of test statistics under a sequence of close alternatives (also called smooth alternatives), which converges to the null hypothesis. Thus, we specify the alternative hypothesis by a sequence of distribution functions $\{F_n(x) : n\ge 1\}$ that converges to the Uniform ([0, 1]) distribution function, which corresponds to the null hypothesis, in the limit as $n\to\infty$.

For symmetric spacings tests, the appropriate sequence of close alternatives (cf. Sethuraman and Rao, 1970; Rao and Sethuraman, 1975) is obtained by letting the distribution function

$$F_n(x) = x + \frac{L_n(x)}{n^{1/4}} \quad \text{for } 0 \le x \le 1, \tag{3.2}$$

where $L_n(0) = L_n(1) = 0$. We further assume that $L_n(x)$ is twice differentiable on the unit interval [0,1] and that there exists a function $L(x)$ which is twice continuously differentiable with $L(0) = L(1) = 0$ and

$$n^{1/4} \sup_{0 \le x \le 1} |L_n(x) - L(x)| = o(1), \tag{3.3}$$

$$n^{1/4} \sup_{0 \le x \le 1} |L_n'(x) - l(x)| = o(1), \tag{3.4}$$

$$n^{1/4} \sup_{0 \le x \le 1} |L_n''(x) - l'(x)| = o(1), \tag{3.5}$$

where $l(x)$ and $l'(x)$ are, respectively, the first and second derivatives of $L(x)$. Note that $L(x) = \int_0^x l(u)\,du$ and $\int_0^1 l(u)\,du = 0$ by the fundamental theorem of calculus.

For completeness and the reader's convenience, we state as our next result the asymptotic distribution for symmetric sum-functions of the spacings (cf. Sethuraman and Rao, 1970) under a sequence of close alternatives.

Lemma 5. *Under the close alternatives* (3.2), *in the limit as* $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} [g(nD_k) - \mathbb{E}g(Z_1)] \xrightarrow{D} N_1(\mu(g), \sigma^2(g)), \tag{3.6}$$

where

$$\mu(g) = \frac{1}{2}\left(\int_0^1 l^2(u)\,du\right) \mathrm{Cov}[g(Z_1), (Z_1 - 2)^2], \tag{3.7}$$

$$\sigma^2(g) = \mathrm{Var}[g(Z_1)] - \frac{\mathrm{Cov}^2[g(Z_1), Z_1]}{\mathrm{Var}[Z_1]}. \tag{3.8}$$

Now let $0 = \xi_0 < \xi_1 < \xi_2 < \cdots < \xi_n < \xi_{n+1} = 1$ form a partition of the unit interval [0, 1], where $\xi_k = k/n + 1$, for $k = 0, 1, 2, \ldots, n+1$. Under the close alternatives, the sample spacings $\{D_k\}$ are related to the uniform spacings $\{T_k\}$ by the relation

$$nD_k = n[F_n^{-1}(U_{(k)}) - F_n^{-1}(U_{(k-1)})] = nT_k + \left(\frac{-l(\xi_k)}{n^{1/4}} + \frac{l^2(\xi_k) + L(\xi_k)l'(\xi_k)}{n^{1/2}}\right)(nT_k) + o_P(n^{-1/2}), \tag{3.9}$$

where $o_P(\cdot)$ is uniform in k. This follows from the mean value theorem for differential calculus and a continuity argument found in Rao and Sethuraman (1975).

We assume that $h : [0,\infty) \times [0,\infty) \to \mathbb{R}$ is a symmetric function with first and second-order partial derivatives. We use the notation $h_x(x,y) = \partial h/\partial x$ and $h_y(x,y) = \partial h/\partial y$ to denote the first partial derivatives of h, and use $h_{xx}(x,y) = \partial^2 h/\partial x^2, h_{yy}(x,y) = \partial^2 h/\partial y^2$ and $h_{xy}(x,y) = \partial^2 h/\partial x \partial y$ to denote the second-order partial derivatives of h.

Using (3.9) and Taylor expansion, we obtain the following convergence result under the close alternatives for a suitably normalized difference between a second-order U-statistic of the sample spacings and the same U-statistic based on uniform spacings. This result will be used to help establish the main result of this section.

Lemma 6. *Under the close alternatives* (3.2), *in the limit as* $n \to \infty$,

$$\frac{2\sqrt{n}}{n(n-1)} \sum_{1 \le i < j \le n} [h(nD_i, nD_j) - h(nT_i, nT_j)] \xrightarrow{\mathbb{P}} \frac{1}{2}\left(\int_0^1 l^2(u)\,du\right) \cdot \mathrm{Cov}[h(Z_1, Z_2), (Z_1 - 2)^2 + (Z_2 - 2)^2]. \tag{3.10}$$

Proof. Using (3.9) in a two-dimensional Taylor expansion of $h(nD_i, nD_j)$ around $h(nT_i, nT_j)$ gives

$$\begin{aligned} h(nD_i, nD_j) - h(nT_i, nT_j) &= \left(\frac{-l(\xi_i)}{n^{1/4}} + \frac{l^2(\xi_i) + L(\xi_i)l'(\xi_i)}{n^{1/2}}\right)(nT_i) h_x(nT_i, nT_j) + \left(\frac{-l(\xi_j)}{n^{1/4}} + \frac{l^2(\xi_j) + L(\xi_j)l'(\xi_j)}{n^{1/2}}\right)(nT_j) h_y(nT_i, nT_j) \\ &\quad + \left(\frac{-l(\xi_i)}{n^{1/4}} + \frac{l^2(\xi_i) + L(\xi_i)l'(\xi_i)}{n^{1/2}}\right) \cdot \left(\frac{-l(\xi_j)}{n^{1/4}} + \frac{l^2(\xi_j) + L(\xi_j)l'(\xi_j)}{n^{1/2}}\right) \cdot (nT_i)(nT_j)\, h_{xy}(nT_i, nT_j) \\ &\quad + \frac{1}{2}\left(\frac{-l(\xi_i)}{n^{1/4}} + \frac{l^2(\xi_i) + L(\xi_i)l'(\xi_i)}{n^{1/2}}\right)^2 (nT_i)^2 h_{xx}(nT_i, nT_j) \\ &\quad + \frac{1}{2}\left(\frac{-l(\xi_j)}{n^{1/4}} + \frac{l^2(\xi_j) + L(\xi_j)l'(\xi_j)}{n^{1/2}}\right)^2 (nT_j)^2 h_{yy}(nT_i, nT_j) + o_P(n^{-1/2}). \end{aligned} \tag{3.11}$$

By summing over all $i<j$, we have

$$\begin{aligned}
&\frac{2\sqrt{n}}{n(n-1)}\sum_{1\le i<j\le n}[h(nD_i,nD_j)-h(nT_i,nT_j)]\\
&\quad=-\frac{2n^{1/4}}{n(n-1)}\sum_{1\le i<j\le n}l(\xi_i)(nT_i)h_x(nT_i,nT_j)-\frac{2n^{1/4}}{n(n-1)}\sum_{1\le i<j\le n}l(\xi_j)(nT_j)h_y(nT_i,nT_j)\\
&\qquad+\frac{2}{n(n-1)}\sum_{1\le i<j\le n}[l^2(\xi_i)+L(\xi_i)l'(\xi_i)](nT_i)h_x(nT_i,nT_j)+\frac{2}{n(n-1)}\sum_{1\le i<j\le n}[l^2(\xi_j)+L(\xi_j)l'(\xi_j)](nT_j)h_y(nT_i,nT_j)\\
&\qquad+\frac{2}{n(n-1)}\sum_{1\le i<j\le n}l(\xi_i)l(\xi_j)(nT_i)(nT_j)h_{xy}(nT_i,nT_j)+\frac{1}{n(n-1)}\sum_{1\le i<j\le n}l^2(\xi_i)(nT_i)^2h_{xx}(nT_i,nT_j)\\
&\qquad+\frac{1}{n(n-1)}\sum_{1\le i<j\le n}l^2(\xi_j)(nT_j)^2h_{yy}(nT_i,nT_j)+o_P(1). \qquad (3.12)
\end{aligned}$$

The composite trapezoid rule asserts that there exists a number $c\in(0,1)$ for which

$$\int_0^1 l(u)\,du=\frac{1}{n+1}\sum_{k=1}^{n}l(\xi_k)-\frac{l''(c)}{12(n+1)^2}.$$

Since we have

$$\lim_{n\to\infty}n^{1/4}\left(\frac{1}{n+1}\sum_{k=1}^{n}l(\xi_k)-\int_0^1 l(u)\,du\right)=\lim_{n\to\infty}\frac{n^{1/4}\cdot l''(c)}{12(n+1)^2}=0,$$

then the first two terms on the RHS of (3.12) converge in probability to zero.

Since $(nT_i,nT_j)\xrightarrow{D}(Z_1,Z_2)$, as $n\to\infty$, observe that both

$$\frac{2}{n(n-1)}\sum_{1\le i<j\le n}[l^2(\xi_i)+L(\xi_i)l'(\xi_i)](nT_i)h_x(nT_i,nT_j)\xrightarrow{\mathbb{P}}\mathbb{E}[Z_1h_x(Z_1,Z_2)]\int_0^1[l^2(x)+L(x)l'(x)]\,dx=0,$$

and

$$\frac{2}{n(n-1)}\sum_{1\le i<j\le n}[l^2(\xi_j)+L(\xi_j)l'(\xi_j)](nT_j)h_y(nT_i,nT_j)\xrightarrow{\mathbb{P}}\mathbb{E}[Z_2h_y(Z_1,Z_2)]\int_0^1[l^2(y)+L(y)l'(y)]\,dy=0,$$

because from integration by parts

$$\int_0^1 L(u)l'(u)\,du=-\int_0^1 l^2(u)\,du.$$

Observe also that

$$\frac{2}{n(n-1)}\sum_{1\le i<j\le n}l(\xi_i)l(\xi_j)(nT_i)(nT_j)h_{xy}(nT_i,nT_j)\xrightarrow{\mathbb{P}}\left(\int_0^1\int_0^1 l(x)l(y)\,dx\,dy\right)\cdot\mathbb{E}[Z_1Z_2h_{xy}(Z_1,Z_2)]=0,$$

because

$$\int_0^1\int_0^1 l(x)l(y)\,dx\,dy=\left(\int_0^1 l(x)\,dx\right)\left(\int_0^1 l(y)\,dy\right)=0.$$

Moreover, we have

$$\frac{1}{n(n-1)}\sum_{1\le i<j\le n}l^2(\xi_i)(nT_i)^2h_{xx}(nT_i,nT_j)\xrightarrow{\mathbb{P}}\frac{1}{2}\left(\int_0^1 l^2(x)\,dx\right)\cdot\mathbb{E}[Z_1^2h_{xx}(Z_1,Z_2)],$$

and

$$\frac{1}{n(n-1)}\sum_{1\le i<j\le n}l^2(\xi_j)(nT_j)^2h_{yy}(nT_i,nT_j)\xrightarrow{\mathbb{P}}\frac{1}{2}\left(\int_0^1 l^2(y)\,dy\right)\cdot\mathbb{E}[Z_2^2h_{yy}(Z_1,Z_2)],$$

with

$$\mathbb{E}[Z_1^2h_{xx}(Z_1,Z_2)+Z_2^2h_{yy}(Z_1,Z_2)]=\mathrm{Cov}[h(Z_1,Z_2),(Z_1-2)^2+(Z_2-2)^2].$$

This completes the proof. □

By combining Theorem 4 with Lemma 6, we arrive at the main result of this section.

Theorem 7. *Under the close alternatives* (3.2), *in the limit as* $n\to\infty$,

$$\sqrt{n}\left(\frac{2}{n(n-1)}\sum_{1\le i<j\le n} h(nD_i,nD_j)-\mathbb{E}[h(Z_1,Z_2)]\right)\xrightarrow{D}\mathrm{N}_1(\mu(h),\sigma^2(h)), \tag{3.13}$$

where

$$\mu(h)=\frac{1}{2}\left(\int_0^1 l^2(u)\,du\right)\cdot\mathrm{Cov}[h(Z_1,Z_2),(Z_1-2)^2+(Z_2-2)^2], \tag{3.14}$$

$$\sigma^2(h)=4(\sigma_1^2-\sigma_{12}^2), \tag{3.15}$$

$$\sigma_1^2=\mathrm{Cov}[h(Z_1,Z_2),h(Z_1,Z_3)], \tag{3.16}$$

$$\sigma_{12}^2=\frac{\mathrm{Cov}^2[h(Z_1,Z_2),Z_1]}{\mathrm{Var}[Z_1]}. \tag{3.17}$$

Proof. Observe that

$$\sqrt{n}\left(\frac{2}{n(n-1)}\sum_{1\le i<j\le n} h(nD_i,nD_j)-\mathbb{E}[h(Z_1,Z_2)]\right)=\sqrt{n}\left(\frac{2}{n(n-1)}\sum_{1\le i<j\le n} h(nT_i,nT_j)-\mathbb{E}[h(Z_1,Z_2)]\right)$$
$$+\frac{2\sqrt{n}}{n(n-1)}\sum_{1\le i<j\le n}[h(nD_i,nD_j)-h(nT_i,nT_j)]\xrightarrow{D}\mathrm{N}_1(\mu(h),\sigma^2(h)).$$

The convergence in distribution follows from Slutsky's Theorem, where the first term on the RHS converges to the $\mathrm{N}_1(0,\sigma^2(h))$ distribution by Theorem 4, and the second term converges in probability to $\mu(h)$ by Lemma 6. This completes the proof. □

4. The asymptotically locally most powerful test

Recall that $\mu(g)$ and $\sigma^2(g)$ denote the asymptotic mean and asymptotic variance corresponding to the general test statistic

$$V_n(g)=\frac{1}{n}\sum_{k=1}^{n} g(nD_k),$$

under the sequence of close alternatives. Here it is assumed that $V_n(g)$ has been normalized to have asymptotic mean zero and finite variance under the null hypothesis. The *Pitman asymptotic relative efficiency* (ARE) of $V_n(g_1)$ relative to $V_n(g_2)$ is given by

$$\mathrm{ARE}(g_1,g_2)=\left(\frac{e^2(g_1)}{e^2(g_2)}\right)^2=\frac{\left(\frac{\mu^2(g_1)}{\sigma^2(g_1)}\right)^2}{\left(\frac{\mu^2(g_2)}{\sigma^2(g_2)}\right)^2}. \tag{4.1}$$

The quantity

$$e^2(g)=\frac{\mu^2(g)}{\sigma^2(g)} \tag{4.2}$$

is called the *efficacy* of the test $V_n(g)$. A test with maximum efficacy is the asymptotically locally most powerful (ALMP) test. In order to find the ALMP test, for symmetric sum-functions of the spacings, against the close alternatives, one needs to find a function $g(\cdot)$ which maximizes

$$e(g)=\frac{(\int_0^1 l^2(u)\,du)\,\mathrm{Cov}[g(Z_1),(Z_1-2)^2]}{2\{\mathrm{Var}[g(Z_1)]-\mathrm{Cov}^2[g(Z_1),Z_1]\}^{1/2}}. \tag{4.3}$$

As mentioned before, the importance of the Greenwood statistic is somewhat justified by the next two results, which were established in Sethuraman and Rao (1970). The Greenwood statistic is the ALMP test among the class of symmetric sum-functions of the spacings.

Lemma 8. *The functional $e(g)$ is maximized by taking $g(t)=t^2$, which in turn gives*

$$\max e(g)=\int_0^1 l^2(u)\,du. \tag{4.4}$$

Lemma 9. *For symmetric sum-functions of the spacings, the asymptotically locally most powerful (ALMP) test of the null hypothesis against the sequence of close alternatives is to reject the null hypothesis when*

$$\sum_{k=1}^{n}(nD_k)^2 > C(\alpha),$$

where the critical value $C(\alpha)$ is determined by the level of significance α. The asymptotic distribution of this optimal statistic under the sequence of close alternatives (3.2) *is given by*

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}[(nD_k)^2-2]\xrightarrow{D}\mathrm{N}_1\left(2\left(\int_0^1 l^2(u)\,du\right),4\right). \tag{4.5}$$

The asymptotic distribution under the null hypothesis is obtained by taking $l(u)=0$ in the above.

Recall that $\mu(h)$ and $\sigma^2(h)$ denote the asymptotic mean and asymptotic variance corresponding to the general test statistic

$$W_n(h)=\frac{2}{n(n-1)}\sum_{1\le i<j\le n}h(nD_i,nD_j),$$

under the sequence of close alternatives. It is assumed that $W_n(h)$ has been normalized to have asymptotic mean zero and finite variance under the null hypothesis. The Pitman ARE of $W_n(h_1)$ relative to $W_n(h_2)$ is given by

$$\mathrm{ARE}(h_1,h_2)=\left(\frac{e^2(h_1)}{e^2(h_2)}\right)^2=\frac{\left(\frac{\mu^2(h_1)}{\sigma^2(h_1)}\right)^2}{\left(\frac{\mu^2(h_2)}{\sigma^2(h_2)}\right)^2}, \tag{4.6}$$

and the quantity

$$e^2(h)=\frac{\mu^2(h)}{\sigma^2(h)} \tag{4.7}$$

is the efficacy of the test $W_n(h)$.

In order to find the ALMP test, for U-statistics of the spacings, against the sequence of close alternatives, we need to find a function h which maximizes the functional

$$e(h)=\frac{(\int_0^1 l^2(u)\,du)\cdot\mathrm{Cov}[h(Z_1,Z_2),(Z_1-2)^2+(Z_2-2)^2]}{4\{\mathrm{Cov}[h(Z_1,Z_2),h(Z_1,Z_3)]-\mathrm{Cov}^2[h(Z_1,Z_2),Z_1]\}^{1/2}}. \tag{4.8}$$

Lemma 10. *The functional $e(h)$ is maximized by taking the symmetric function $h(z_1,z_2)=(z_1-z_2)^2$, which in turn gives*

$$\max e(h)=\int_0^1 l^2(u)\,du, \tag{4.9}$$

which corresponds to that of the Greenwood statistic.

Proof. It is enough to find a function h which maximizes the numerator in (4.8). By the Cauchy–Bunyakovsky–Schwarz inequality, we have

$$e(h)=\frac{(\int_0^1 l^2(u)\,du)\cdot\mathrm{Cov}[h(Z_1,Z_2),(Z_1-2)^2+(Z_2-2)^2]}{4\{\mathrm{Cov}[h(Z_1,Z_2),h(Z_1,Z_3)]-\mathrm{Cov}^2[h(Z_1,Z_2),Z_1]\}^{1/2}}\le\frac{(\int_0^1 l^2(u)\,du)\sqrt{\mathrm{Var}[h(Z_1,Z_2)]}\sqrt{\mathrm{Var}[(Z_1-2)^2+(Z_2-2)^2]}}{4\{\mathrm{Cov}[h(Z_1,Z_2),h(Z_1,Z_3)]-\mathrm{Cov}^2[h(Z_1,Z_2),Z_1]\}^{1/2}}. \tag{4.10}$$

The inequalities become equalities if and only if $h(z_1,z_2)=a[(z_1-2)^2+(z_2-2)^2]+b$, for some real numbers $a\neq 0$ and b. In this particular case, the functional $e(h)$ attains the upper bound in (4.10), i.e.

$$e(h)=\frac{(\int_0^1 l^2(u)\,du)a\cdot\mathrm{Var}[(Z_1-2)^2+(Z_2-2)^2]}{4\sqrt{a^2\cdot\mathrm{Var}(Z_1-2)^2-a^2\cdot\mathrm{Cov}^2[(Z_1-2)^2,Z_1]}}=\frac{(\int_0^1 l^2(u)\,du)2a\cdot\mathbb{E}[Z_1^2+Z_2^2]}{4\sqrt{4a^2}}=\int_0^1 l^2(u)\,du.$$

On the other hand, since $\mathrm{Cov}[Z_1Z_2,(Z_1-2)^2+(Z_2-2)^2]=0$, then the maximum of the functional $e(h)$ is attained by taking $h(z_1,z_2)=(z_1-z_2)^2$, and directly from (4.8) we have

$$\max e(h)=\frac{(\int_0^1 l^2(u)\,du)\cdot \mathrm{Cov}[(Z_1-Z_2)^2,(Z_1-2)^2+(Z_2-2)^2]}{4\{\mathrm{Cov}[(Z_1-Z_2)^2,(Z_1-Z_3)^2]-\mathrm{Cov}^2[(Z_1-Z_2)^2,Z_1]\}^{1/2}}=\frac{(\int_0^1 l^2(u)\,du)\cdot 2\cdot \mathbb{E}[Z_1^2+Z_2^2]}{8}=\int_0^1 l^2(u)\,du.$$

This completes the proof. □

By combining Theorem 7 and Lemma 10, we have the following

Theorem 11. *For U-statistics of the spacings, the asymptotically locally most powerful (ALMP) test of the null hypothesis against the sequence of close alternatives is to reject the null hypothesis when*

$$\sum_{1\le i<j\le n}(nD_i-nD_j)^2>C(\alpha),$$

where the critical value $C(\alpha)$ is determined by the level of significance α. The asymptotic distribution of this optimal statistic under the sequence of close alternatives (3.2) *is given by*

$$\sqrt{n}\left(\frac{2}{n(n-1)}\sum_{1\le i<j\le n}(nD_i-nD_j)^2-2\right)\xrightarrow{D}\mathrm{N}_1\left(4\left(\int_0^1 l^2(u)\,du\right),16\right). \tag{4.11}$$

The asymptotic distribution under the null hypothesis is obtained by taking $l(u)=0$ in the above.

5. Some examples

To illustrate the results of this paper, specifically Theorems 7 and 11, we present several noteworthy examples, and obtain their asymptotic distributions under the sequence of close alternatives (3.2). We also compare the efficacies of the generalized Gini mean difference spacings test and the generalized Rao's spacings test.

Example (*Gini Mean Difference Spacings Test*). Under the null hypothesis of uniformity, the Gini mean difference spacings test

$$G_n(1)=\frac{2}{n(n-1)}\sum_{1\le i<j\le n}|nD_i-nD_j|=\frac{\sum_{i=1}^n\sum_{j=1}^n|nD_i-nD_j|}{n(n-1)}\simeq\frac{2S_{n-1}}{n-1}, \tag{5.1}$$

where $S_{n-1}=\sum_{k=1}^{n-1}U_k$ is the sum of $(n-1)$ independent Uniform ([0, 1]) random variables. The probability distribution of S_{n-1} is known as the Irwin–Hall Uniform sum distribution, and was first derived by P.S. Laplace in 1814 (cf. Wilks, 1962; Feller, 1971, Theorem 1, I.9).

The probability density function of S_{n-1} has the form

$$f_{S_{n-1}}(s)=\frac{1}{(n-2)!}\sum_{k=0}^{n-1}\binom{n-1}{k}(-1)^k(s-k)_+^{n-2}\cdot I(0<s<n-1),$$

and can be derived via the Fourier inversion formula, and Cauchy's integral formula from complex analysis. It follows that $G_n(1)$ has probability density function

$$f_{G_n(1)}(y)=\frac{n-1}{2(n-2)!}\sum_{k=0}^{n-1}\binom{n-1}{k}(-1)^k((n-1)y/2-k)_+^{n-2}\cdot I(0<y<2).$$

Let the kernel $h(z_1,z_2)=|z_1-z_2|$, so that $h_{xx}(z_1,z_2)=h_{yy}(z_1,z_2)=2\delta(z_1-z_2)$, where $\delta(\cdot)$ is the Dirac delta function. Then $\mathbb{E}[h(Z_1,Z_2)]=\mathbb{E}|Z_1-Z_2|=\Gamma(2)=1$. The asymptotic mean

$$\begin{aligned}\mu(h)&=\frac{1}{2}\left(\int_0^1 l^2(u)\,du\right)\cdot\mathrm{Cov}[|Z_1-Z_2|,(Z_1-2)^2+(Z_2-2)^2]=\frac{1}{2}\left(\int_0^1 l^2(u)\,du\right)\cdot\mathbb{E}[(Z_1^2+Z_2^2)\cdot 2\delta(Z_1-Z_2)]\\&=\left(\int_0^1 l^2(u)\,du\right)\int_0^\infty\left(\int_0^\infty[u^2+v^2]e^{-u}\delta(u-v)\,du\right)e^{-v}\,dv=\left(\int_0^1 l^2(u)\,du\right)\cdot 2\int_0^\infty v^2e^{-2v}\,dv=\frac{1}{2}\int_0^1 l^2(u)\,du.\end{aligned}$$

The asymptotic variance

$$\sigma^2(h)=4\left(\mathrm{Cov}[|Z_1-Z_2|,|Z_1-Z_3|]-\frac{\mathrm{Cov}^2[|Z_1-Z_2|,Z_1]}{\mathrm{Var}[Z_1]}\right)=1/3.$$

Under the close alternatives, in the limit as $n\to\infty$,

$$\sqrt{n}(G_n(1)-1)\xrightarrow{D}\mathrm{N}_1\left(\frac{1}{2}\left(\int_0^1 l^2(u)\,du\right),1/3\right). \tag{5.2}$$

Example (*Gini Mean Squared Difference Spacings Test*). From Theorem 11, under a sequence of close alternatives, the Gini mean squared difference spacings test

$$G_n(2)=\frac{2}{n(n-1)}\sum_{1\le i<j\le n}(nD_i-nD_j)^2=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(nD_i-nD_j)^2, \tag{5.3}$$

is the ALMP test for U-statistics of the spacings. This also means that the $G_n(2)$ is the best test among the generalized Gini mean difference spacings tests $G_n(r), r>0$. Moreover, by Lemma 10, the Gini mean squared difference spacings test has efficacy

$$e^2(h)=\left(\int_0^1 l^2(u)\,du\right)^2,$$

which is the same as that of the classical Greenwood statistic.

Let the kernel $h(z_1,z_2)=(z_1-z_2)^2$, so that $h_{xx}(z_1,z_2)=h_{yy}(z_1,z_2)=2$. Then $\mathbb{E}[h(Z_1,Z_2)]=\mathbb{E}(Z_1-Z_2)^2=\Gamma(3)=2$. The asymptotic mean

$$\mu(h)=\frac{1}{2}\left(\int_0^1 l^2(u)\,du\right)\cdot\mathrm{Cov}[(Z_1-Z_2)^2,(Z_1-2)^2+(Z_2-2)^2]=\frac{1}{2}\left(\int_0^1 l^2(u)\,du\right)\cdot 2\cdot\mathbb{E}[(Z_1^2+Z_2^2)]=4\int_0^1 l^2(u)\,du.$$

The asymptotic variance

$$\sigma^2(h)=4\left(\mathrm{Cov}[(Z_1-Z_2)^2,(Z_1-Z_3)^2]-\frac{\mathrm{Cov}^2[(Z_1-Z_2)^2,Z_1]}{\mathrm{Var}[Z_1]}\right)=16.$$

Under the close alternatives, in the limit as $n\to\infty$,

$$\sqrt{n}(G_n(2)-2)\xrightarrow{D}\mathrm{N}_1\left(4\left(\int_0^1 l^2(u)\,du\right),16\right). \tag{5.4}$$

Example (*Kullback–Leibler Divergence Spacings Test*). Under the close alternatives, the Kullback–Leibler divergence measure of the spacings has the asymptotic distribution,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}[(nD_k)\log(nD_k)+(\gamma-1)]\xrightarrow{D}\mathrm{N}_1\left(\frac{1}{2}\left(\int_0^1 l^2(u)\,du\right),\frac{\pi^2}{3}-3\right). \tag{5.5}$$

Here, the constant $\gamma=\mathbb{E}[-\log(Z_1)]=0.57721\ldots$ is the famous Euler–Mascheroni constant.

As a toy example, take $h(z_1,z_2)=\log(z_1^{z_1}\cdot z_2^{z_2})=z_1\log z_1+z_2\log z_2$. Then under the close alternatives, in the limit as $n\to\infty$,

$$\sqrt{n}\left(\frac{2}{n(n-1)}\sum_{1\le i<j\le n}\log[(nD_i)^{nD_i}\cdot(nD_j)^{nD_j}]+2(\gamma-1)\right)\xrightarrow{D}\mathrm{N}_1\left(\left(\int_0^1 l^2(u)\,du\right),4\left(\frac{\pi^2}{3}-3\right)\right). \tag{5.6}$$

However, there really is not anything new, because this U-statistic is simply a linear transformation of the Kullback–Leibler divergence statistic, i.e.

$$\frac{2}{n(n-1)}\sum_{1\le i<j\le n}\log[(nD_i)^{nD_i}\cdot(nD_j)^{nD_j}]=\frac{2}{n}\sum_{k=1}^{n}(nD_k)\log(nD_k).$$

Example (*Log-Spacings U-Statistic*). Under the close alternatives, Darling's log-spacings test has the asymptotic distribution,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}[\log(nD_k)+\gamma]\xrightarrow{D}\mathrm{N}_1\left(-\frac{1}{2}\left(\int_0^1 l^2(u)\,du\right),\frac{\pi^2}{6}-1\right). \tag{5.7}$$

To avoid the trivial example where the U-statistic is a linear transformation of Darling's log-spacings test, we let $h(z_1,z_2)=\log(z_1+z_2)$ so that $h_{xx}(z_1,z_2)=h_{yy}(z_1,z_2)=-(Z_1+Z_2)^{-2}$. We have $\mathbb{E}[h(Z_1,Z_2)]=\mathbb{E}[\log(Z_1+Z_2)]=1-\gamma$. The asymptotic mean

$$\mu(h)=\frac{1}{2}\left(\int_0^1 l^2(u)\,du\right)\cdot\mathrm{Cov}[\log(Z_1+Z_2),(Z_1-2)^2+(Z_2-2)^2]=\frac{1}{2}\left(\int_0^1 l^2(u)\,du\right)\cdot\mathbb{E}\left[\frac{-(Z_1^2+Z_2^2)}{(Z_1+Z_2)^2}\right]=-\frac{1}{3}\int_0^1 l^2(u)\,du.$$

The asymptotic variance

$$\sigma^2(h) = 4\left(\mathrm{Cov}[\log(Z_1+Z_2),\log(Z_1+Z_3)] - \frac{\mathrm{Cov}^2[\log(Z_1+Z_2),Z_1]}{\mathrm{Var}[Z_1]}\right) = \frac{4\pi^2}{3} - 13.$$

Under the close alternatives, in the limit as $n\to\infty$,

$$\sqrt{n}\left(\frac{2}{n(n-1)}\sum_{1\le i<j\le n}\log(nD_i+nD_j)+(\gamma-1)\right) \xrightarrow{D} \mathrm{N}_1\left(-\frac{1}{3}\left(\int_0^1 l^2(u)\,du\right), \frac{4\pi^2}{3}-13\right). \tag{5.8}$$

Example (*Gini vs. Rao*). We compare the efficacies of the generalized Gini mean difference spacings test

$$G_n(r) = \frac{2}{n(n-1)}\sum_{1\le i<j\le n}|nD_i - nD_j|^r = \frac{\sum_{i=1}^n\sum_{j=1}^n|nD_i-nD_j|^r}{n(n-1)}, \quad r>0,$$

and the generalized Rao's spacings test

$$J_n(r) = \frac{1}{n}\sum_{k=1}^n |nD_k - 1|^r, \quad r>0.$$

It will be convenient to define the *modified efficacy* of a test as

$$e_M^2(\cdot) = \frac{e^2(\cdot)}{\left(\int_0^1 l^2(u)\,du\right)^2}. \tag{5.9}$$

Table 1 lists the modified efficacies of $G_n(r)$ and $J_n(r)$ with respect to various choices of $r>0$, and also the Pitman ARE of $G_n(r)$ relative to $J_n(r)$. It is seen that the generalized Gini mean difference spacings test $G_n(r)$ has better efficacy, and is more Pitman efficient than the generalized Rao's spacings test $J_n(r)$, except for the case $r=2$, when both spacings tests $G_n(2)$ and $J_n(2)$ correspond to the classical Greenwood statistic. Table 2 summarizes our aforementioned examples.

6. Conclusion

We derived the general asymptotic theory for second-order *U*-statistics based on spacings both under the null hypothesis and under a sequence of close alternatives, and found the Gini mean squared difference test is the ALMP test in this class and it has the same efficacy as the Greenwood statistic based on the sum of squares. On the basis of Pitman asymptotic relative efficiency, the generalized Gini mean difference test is asymptotically more efficient than the

Table 1
Modified efficacies for the generalized Rao's spacings test and generalized Gini mean difference spacings tests with Pitman ARE.

r	Generalized Rao	Generalized Gini	Pitman ARE
1	0.572654	0.75	1.715291
3/2	0.892135	0.946889	1.126515
2	1	1	1
5/2	0.93921	0.96137	1.047745
3	0.818649	0.867857	1.123831
4	0.550562	0.615384	1.249337

Table 2
Some examples of symmetric spacings tests.

Test statistic	Mean	Variance	$e_M^2(\cdot)$
$\sqrt{n}(G_n(1)-1)$	$\frac{1}{2}\int_0^1 l^2(u)\,du$	1/3	3/4
$\sqrt{n}(G_n(2)-2)$	$4\int_0^1 l^2(u)\,du$	16	1
$\sqrt{n}\left(\frac{2\sum_{1\le i<j\le n}\log(nD_i+nD_j)}{n(n-1)}+(\gamma-1)\right)$	$-\frac{1}{3}\int_0^1 l^2(u)\,du$	$\frac{4\pi^2}{3}-13$	0.697
$\frac{1}{\sqrt{n}}\sum_{k=1}^n[(nD_k)\log(nD_k)+(\gamma-1)]$	$\frac{1}{2}\int_0^1 l^2(u)\,du$	$\frac{\pi^2}{3}-3$	0.862
$\frac{1}{\sqrt{n}}\sum_{k=1}^n[\log(nD_k)+\gamma]$	$-\frac{1}{2}\int_0^1 l^2(u)\,du$	$\frac{\pi^2}{6}-1$	0.388
$\frac{1}{\sqrt{n}}\sum_{k=1}^n[\lvert nD_k-1\rvert-2e^{-1}]$	$e^{-1}\int_0^1 l^2(u)\,du$	$\frac{8e-20}{e^2}$	0.573

generalized Rao's test. Extension of these ideas to *U*-statistics based on higher-order spacings and to two-sample problems involving "spacing-frequencies" (see, e.g. Holst and Rao, 1980) will be investigated elsewhere.

References

Feller, W., 1971. An Introduction to Probability Theory and its Applications, Vol. II. John Wiley and Sons.
Greenwood, M., 1946. The statistical study of infectious diseases. Journal of the Royal Statistical Society, Series A (General) 109, 85–110.
Holst, L., 1981. Some conditional limit theorems in exponential families. The Annals of Probability 9 (5), 818–830.
Holst, L., Rao, J.S., 1980. Asymptotic theory for some families of two-sample nonparametric statistics. Sankhya: The Indian Journal of Statistics, Series A 42 (Parts 1 & 2), 19–52.
Jammalamadaka, S.R., Goria, M.N., 2004. A test of goodness of fit based on Gini's index of spacings. Statistics and Probability Letters 68, 177–187.
Lee, A.J., 1990. U-Statistics. Statistics: Textbooks and Monographs. Marcel Dekker.
Pyke, R., 1965. Spacings. Journal of the Royal Statistical Society, Series B (Methodological) 27 (3), 395–449.
Rao, J.S., 1969. Some Contributions to the Analysis of Circular Data. Ph.D. Thesis, Indian Statistical Institute, Calcutta, India.
Rao, J.S., Sethuraman, J., 1975. Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors. The Annals of Statistics 3 (2), 213–299.
Sethuraman, J., Rao, J.S., 1970. Pitman efficiencies of tests based in spacings. In: Puri, M.L. (Ed.), Nonparametric Techniques in Statistical Inference, Cambridge University Press, pp. 405–415.
Wilks, S.S., 1962. Mathematical Statistics. John Wiley and Sons.