

A Note on Asymptotical Efficiency of the Goodness of Fit Tests Based on Disjoint k-Spacings Statistic

Muhammad Naeem

Mathematics

Deanship of First Combined Year, Umm Al-Qura University Makkah Mukarramah, 21514 KSA

naeemtazkeer@yahoo.com

Abstract

In this paper Pitman's asymptotic efficiency (AE) as well as Kallenberg's intermediate AE of the goodness-of-fit tests based on higher-order non-overlapping spacings is considered. We study log statistic as well as entropy type statistic based on k-spacings when k may tend to infinity as n approaches infinity. It certainly compliments the available results for fixed k and provides more general result. We show that both types of statistics based on higher ordered spacings have higher efficiencies in Pitman's sense compared to their counterparts based on simple spacings. It is also shown that the Kallenberg's intermediate AE of such test coincides with its Pitman's AE, the power of the tests are also discussed.

Keywords: Spacings; Goodness-of-fit; Asymptotic efficiency; Asymptotic power.

1. Introduction

Consider a population with continuous cumulative distribution function (cdf) G and probability density function (pdf) g . We select an increasing order sample $Z'_1, Z'_2, \dots, Z'_{n-1}$ from this population. To show that the distribution of G is the same as that of some known distribution is the classical goodness-of-fit problem which is an important aspect of inferential statistics. A famous procedure adopted by statisticians is to transform the available data to uniform one using the probability integral transformation $Z = U(Z')$. By doing so the support of G is reduced to $[0,1]$ and the known cumulative distribution function reduces to that of a uniform random variable on $[0,1]$. It is, in fact, the problem of testing the null hypothesis

$$H_0: g(z) = 1; 0 \leq z \leq 1, \quad (1)$$

against alternative that g is probability density function of some non-uniform random variable having support on $[0,1]$. Suppose our problem is reduced by the above mentioned construction. One way of testing the goodness-of-fit problem is based on observed frequencies which perform better in detecting differences between the distribution functions. The second type of tests is based on spacings and they are useful to detect differences between the corresponding densities. It is known that comparable test based on k-spacings is better than chi-square test in terms of local power (Jammalamadaka & Tiwari 1987). It certainly provides an edge to the spacings based tests over frequency based tests. Let $Z_0 = 0$ and $Z_n = 1$ the no-overlapping k-spacings are defined as $D_m = Z'_{mk} - Z'_{(m-1)k}$, $m = 1, 2, \dots, N'$, $D_{N'+1}^{(k)} = 1 - Z'_{N'k}$, where integer $k \in [1, n]$, $N' = [n/k]$ is the greatest integer less than or equal to n/k . By putting $N = N'$ if n/k is an integer and $N = N' + 1$ otherwise. Note that $D = (D_1^{(k)}, D_2^{(k)}, \dots, D_N^{(k)})$ and the quantities $D_m^{(k)}$, Z'_m depend on n also but for simplicity of notation the extra suffix is omitted. We are testing hypothesis (1) against the sequence of alternatives

$$H_{1,n}: g(z) = 1 + dl(z)\delta(n), \quad 0 \leq z \leq 1, \quad (2)$$

where $\delta(n) \rightarrow \infty$, as $n \rightarrow \infty$ and $d > 0$ is a distance between $H_{1,n}$ and H_0 , $l(z)$ is a direction of $H_{1,n}$, such that it is continuously differentiable and

$$\int_0^1 l(z) dz = 0, \quad \int_0^1 l^2(z) dz = 1. \quad (3)$$

Assume that the order of spacing i.e. k and n may tends to infinity jointly. We consider test based on the statistic

$$L_N(D) = \sum_{m=1}^N g_i(nD_m^{(k)}), \quad i = 1, 2. \quad (4)$$

where $g_1(nD) = \log(nD)$, $g_2(nD) = (nD) \log(nD)$. The large value of $L_N(D)$ rejects the hypothesis. Clearly, statistic $L_N(D)$ is symmetric in $\{D_{m,n}^{(k)}\}$. Though symmetric statistics are not efficient in Pitman sense but practically fewer calculations are required in the analysis of such statistics. Tests of the form (4) based on simple spacing, i.e. when $k=1$ and different $g(z)$, have been proposed by many authors (see, for example, (Pyke, 1965) and the references contained therein). Distribution theory of such statistics and their asymptotic efficiencies have been studied, for instance, by (Rao & Sethuraman, 1975) (Holst & Rao, 1981). For $i = 1$ statistics (4) is called log-spacings statistics suggested by (Darling, 1953) with $k=1$ (simple spacing) and was able to obtain its limit distribution as a special case of its introduced formula. For $i = 2$ the statistics (4) is called entropy type statistics and is studied by (Gebert & Kale, 1969) using the formula devised by (Darling, 1953) for the first two moments. The statistics (4) was also studied by (Kale, 1969) in a general way for $k=1$ and $i=1, 2$. The entropy type statistics based on simple spacings is studied by (Jammalamadaka and Tiwari, 1985) using the famous characterization of (Rao and Sethuraman, 1975). For the first time the asymptotic normality of the statistics based on disjoint k -spacings was discussed by (Del Pino, 1979) and has shown that it is more efficient in Pitman sense than simple spacings statistics. We also refer to (Morgan Kuo & Jammalamadaka, 1981) who have studied statistic of the form (4). The entropy statistics based on higher order spacings is studied by (Jammalamadaka and Tiwari, 1987) for fixed k using the characterization of (Del Pino, 1979). The asymptotic normality of statistics equivalent to log-spacings statistics is proved by (Cressie, 1976) for overlapping k -spacings. A statistics based on disjoint k -spacings involve considerably less calculation as compared to that based on overlapping k -spacings. Although statistics based on disjoint k -spacings is less efficient than the one based on overlapping k -spacings but its efficiency can be made even greater than overlapping based statistics by increasing the order of k . In his paper, (Czekala, 1999) studied statistics (4) for fix $k \geq 1$ and calculated Bahadur efficiency for it. The present paper, using the characterization of (Mirakhmedov, 2005), discusses the test of goodness of fit based on (4) with $k \geq 1$ which may increase jointly with n . We show, particularly, that logarithms tests based on higher ordered spacings has higher efficiency in Pitman sense compared to their counterparts based on simple spacings, the (Kallenberg, 1983) intermediate efficiency of statistics (4) is discussed, furthermore, the power of the tests is also studied.

Asymptotic normality of L_N under hypothesis.

We present here two results on asymptotic normality and Cramer's type large deviation theorem obtained by (Mirakhmedov, 2005) and (Mirakhmedov et al. 2011), respectively, for the sum of functions of uniform spacings (i.e. under null hypothesis). Suppose $0 = Z_0 \leq Z_1 \leq \dots \leq Z_{n-1} \leq Z_n = 1$ be an ordered sample from uniform $[0,1]$ distribution and $T_{m,n}^{(k)}$ their non-overlapping k -spacings. Let $f_{m,n}(u)$, $m = 1, 2, \dots, N$, be measurable functions. Consider the statistics

$$G_N(T) = \sum_{m=1}^N f_{m,n}(nT_{m,n}^{(k)}). \quad (5)$$

Let Z and $Z_{m,k}$, $m = 1, 2, \dots, N$ be independent and identically distributed random variables with common density function $\gamma_k(u) = u^{k-1} e^{-u} / \Gamma(k)$, $u > 0$, where $\Gamma(k)$ is well known gamma function. If family of functions, $\{f_{m,n}(u), m = 1, 2, \dots, N\}$ are the same for all m then statistic given in (5) is said to be symmetric. Except this section we will consider symmetric form of our statistic in this article. For simple uniform spacing $T_m, m = 1, 2, \dots, n$ it is well known $L(T_1, T_2, \dots, T_n) = L\left(Z_1, Z_2, \dots, Z_n \mid \sum_{m=1}^n Z_m = 0\right)$, where $L(X)$ represent the distribution of a random vector X and Z_1, Z_2, \dots, Z_n are independent random variables of common exponential distribution with mean 1.

By putting $S_{N,k} = Z_{1,k} + \dots + Z_{N,k}$, $Q_N(Z) = \sum_{m=1}^N f_{m,n}(Z_{m,k})$, $\rho_N = \text{corr}(Q_N, S_{N,k})$,

$$h_{m,n}(Z) = f_{m,n}(Z_{m,k}) - E f_{m,n}(Z_{m,k}) - (Z - k) \rho_N \sqrt{\frac{\text{Var} Q_N}{Nk}}, \quad A_N = \sum_{m=1}^N E f_{m,n}(Z_{m,k}),$$

$$\sigma_N^2 = \sum_{m=1}^N \text{Var} h_{m,n}(Z_{m,k}) \quad (6)$$

$$\beta_{3,n} = \sum_{m=1}^N E \left| \frac{h_{m,n}(Z_{m,k})}{\sigma_N} \right|^3 \quad \text{and} \quad P_N(Z) = P\{Q_N(Z) \leq Z \sigma_N\}. \quad \text{It is to be noted that}$$

$\sigma_N^2 = (1 - \rho_N^2) \text{Var} Q_N(Z)$ and clearly $Q_N(T) = G_N(T) - A_N$ also, by the well-known inequality $\beta_{3,N} \leq \beta_{4,N}^{1/2}$ where $\beta_{4,N}$ can be easily calculated from corresponding statistics, we have the following Theorem.

Theorem 2.1

There exists a positive constant C such that

$$\sup |P_N(Z) - \Phi(Z)| \leq C \beta_{3,N}, \quad Z \in R \quad \text{as} \quad N \rightarrow \infty$$

where $\Phi(Z)$ is the standard normal distribution.

Theorem 2.1 is the corollary -2 of Mirakhmedov (2005).

Theorem 2.2

If

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sigma_N^2 > 0 \quad (7)$$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=1}^N E \left| h_{m,n} (Z_{m,k}) \right|^3 < \infty \quad (8)$$

then for y such that $0 \leq Z \leq \sqrt{\theta' \ln N}$ where $\theta' \in (0,1)$

$$1 - P_N(Z) = (1 - \Phi(Z))(1 + o(1)) \quad (9)$$

Proof:

Due to conditions (7) and (8), inequality in Theorem 2.1 gets the form

$$|P_N(Z) - \Phi(Z)| \leq C_1 N^{-\theta'/2} \quad (10)$$

Theorem 2.1 is readily proved from inequality (10) if $0 \leq Z \leq C$.

Let

$$C \leq Z \leq \sqrt{\theta' \ln N}$$

then by the well-known inequality

$$\frac{Ze^{-Z^2/2}}{\sqrt{2\pi}(1+Z^2)} < 1 - \Phi(Z) < \frac{e^{-Z^2/2}}{Z\sqrt{2\pi}} \quad \forall \quad Z > 0 \text{ and inequality (10) we have}$$

$$\left| \frac{1 - P_N(Z)}{1 - \Phi(Z)} - 1 \right| \leq C_1 |P_N(Z) - \Phi(Z)| Z \exp\{Z^2/2\} \leq C_2 N^{-1/2} \sqrt{\ln N} N^{\theta'/2} = o(1)$$

It completes the proof of Theorem 2.2.

By virtue of Theorem 2.1 it is clear that random variable G_N has asymptotically normal distribution with expectation A_N and variance σ_N^2 .

It is obvious that for the statistic $G_N(Z)$ Cramer's condition: there exists $H > 0$ such that $E \exp\{H |f_{m,n}(Z_{m,k})|\} < \infty$, is satisfied. Therefore, by Theorem of (Mirakhmedov et al. 2011), it follows

Theorem 2.3:

For all $Z \geq 0, Z = o(\sqrt{n})$

$$P_0 \{G_N(Z) \geq Z\sigma_N \sqrt{N} + NA_N\} = \Phi(-Z) \exp \left\{ -\frac{Z^3}{\sqrt{N}} \Lambda_N \left(-\frac{Z}{\sqrt{N}} \right) \right\} \left(1 + O \left(\frac{Z+1}{\sqrt{N}} \right) \right),$$

where $\Lambda_N(u) = \lambda_{0N} + \lambda_{1N}u + \dots$ is a special Cramer's type power series; for N large enough $|\lambda_{jN}| \leq C < \infty, j = 0, 1, 2, \dots$

The statistic L_N of equation (4) is a special case of (5) with

$$f_{m,n}(u) = f(u) = g_i(u), \quad i = 1, 2, \quad (11)$$

where $g_1(u) = \log u$ and $g_2(u) = u \log u$. Therefore, as a consequence of Theorem 2.1, we get the following Theorem

Theorem 2.4:

Under H_0 the statistics L_N has asymptotically normal distribution with expectation $NA(g_i)$ and variance $N\sigma^2(g_i)$ as $N \rightarrow \infty$. Here

$$A(g_1) = 1 + \frac{1}{2} + \dots + \frac{1}{k-1} - \gamma, \sigma^2(g_1) = \frac{\pi^2}{6} - \sum_{j=1}^k \frac{1}{j^2} - \frac{1}{k} \quad \text{and}$$

$$A(g_2) = k \left\{ 1 + \frac{1}{2} + \dots + \frac{1}{k} \right\}, \sigma^2(g_2) = k(k+1) \left\{ \frac{\pi^2}{6} - \sum_{j=1}^k \frac{1}{j^2} \right\} - k$$

where γ is the well-known Euler function.

From Theorem 2.4 and well known theorem on convergence of moments (Moran 1984) it follows that

$$E_0 L_N(Z) = NA(g_i) + o(N) \text{ and } \text{Var}_0 L_N(Z) = N\sigma^2(g_i)(1 + o(1)), \quad (12)$$

here and what follows P_J, E_J and Var_J denotes the probability, expectation and variance accounted under $H_J, J = 0, 1$.

Asymptotic normality of $L_N(Z)$ under alternatives.

We need two lemmas in order to study $L_N(Z)$ under alternatives (2). Recall that $D_{m,n}^{(k)}$, $m=1, 2, \dots, N$ are the k -spacings under alternatives (2). Then $D_{m,n}^{(k)}$ can be reduced to uniform spacings $T_{m,n}^{(k)}$, see (Morgan & Jammalamadaka 1981). The corresponding relation presented here as

Lemma 3.1: Under alternative (2)

$$nD_{m,n}^{(k)} = nT_{m,n}^{(k)} \left(1 - l \left(\frac{m}{N} \right) \delta(n) + O_p(\delta^2(n)) \right), \quad (13)$$

where $O_p(\cdot)$ is uniform in m .

The following simple Lemma stated without proof.

Lemma 3.2: Let $p(u)$ defined on $(0,1)$ be continuous except possibly for finitely many u and be bounded in absolute value by an integrable function. Then

$$\frac{1}{N} \sum_{m=1}^N p \left(\frac{m}{N} \right) = \int_0^1 p(u) du + o(1), \text{ as } N \rightarrow \infty.$$

In what follows the functions $h(u)$ (symmetric form) and $g_i(u), i=1,2$ are from (6) and (11) respectively.

Theorem 3.1:

Let

$$k\delta(n) = o(1). \quad (14)$$

Then

$$E_1 L_N(Z) = N(A(g_i) + B_N(g_i)) + o(N\delta^2(n)), \quad \text{Var}_1 L_N(Z) = N\sigma^2(g_i)(1 + o(1)), \quad (15)$$

where $B_N(g_i) = \frac{1}{2}\delta^2(n)\sigma(g_i)\sqrt{2k(k+1)}\rho(g_i)d^2$, and

$$\rho(g_i) = \text{corr}(h(Z), Z^2 - 2(k+1)Z + k(k+1))$$

Proof. Let $E_J L_N(Z) = A_{JN}(g_i)$. Due to (12), Lemma 3.1 and Lemma 3.2 we have

$$A_{JN}(g_i) = N \int_0^1 E g_i \left(Z(1 - dl(u)\delta(n) + O_p(\delta^2(n))) \right) du + o(1) = N \int_0^1 \psi(u; k, \tau) du + o(1), \quad (16)$$

where

$$\tau \equiv \tau(n, u) = dl(u)\delta(n) + O(\delta^2(n)), \quad (17)$$

and

$$\psi(u; k, \tau) = \int_0^\infty g_i(t(1 - \tau(n, u))) \gamma(t) dt.$$

We have

$$\psi(u; k, \tau) = \frac{1}{(1 - \tau)^k} \int_0^\infty g_i(v) \gamma_k(v) \exp\{-v\tau/(1 - \tau)\} dv.$$

According to Laplace's method of asymptotic expansion of integrals (Keener 1987) we can use the following expansion

$$\exp\{-v\tau/(1 - \tau)\} = 1 - \frac{v\tau}{1 - \tau} + \frac{(v\tau)^2}{2(1 - \tau)^2} + O((v\tau)^3).$$

in order to get an asymptotic expansion of the integral in (16). By using the above and

$$(1 - \tau)^k = 1 + k\tau + \frac{1}{2}k(k+1)\tau^2 + O((k\tau)^3),$$

since under (14) $k\tau = O(1)$, we find

$$\begin{aligned} \psi(u; k, \tau) &= \frac{1}{\Gamma(k)} \int_0^\infty g_i(v) v^{k-1} e^{-v} \left(1 - (v-k)\tau + \frac{1}{2}(v^2 - 2(k+1)v + k(k+1))\tau^2 + O((v\tau)^3) \right) dv \\ &= E g_i(Z) - \tau E((Z-k)g_i(Z)) + \frac{1}{2}\tau^2 E(g_i(Z)(Z^2 - 2(k+1)Z + k(k+1))) + O(\tau^3 E g_i(Z) Z^3) \end{aligned}$$

By putting this result into (16), use (17) instead of τ and take into account (3) to get

$$A_{JN}(g_i) = N A(g_i) + \frac{N d^2}{2} \delta^2(n) E(g_i(Z)(Z^2 - 2(k+1)Z + k(k+1))) + O(N \delta^3(n) E h(Z) Z^3) \quad (18)$$

By direct calculations we find

$$\begin{aligned} E(Z^2 - 2(k+1)Z + k(k+1)) &= 0, \\ E(Z-k)(Z^2 - 2(k+1)Z + k(k+1)) &= 0, \\ \text{Var}(Z^2 - 2(k+1)Z + k(k+1)) &= 2k(k+1), \end{aligned}$$

$$E(g_1(Z)) = \sum_{j=1}^k \frac{1}{j} - \gamma - (k)^{-1}, \quad E(g_2(Z)) = k \left(\sum_{j=1}^k \frac{1}{j} - \gamma \right)$$

$$\rho(g_i) = \text{corr}(h(Z), Z^2 - 2(k+1)Z + k(k+1)),$$

$$Eg_i(Z)(Z^2 - 2(k+1)Z + k(k+1)) = \sigma(g_i) \sqrt{k(k+1)} \rho(g_i). \quad (19)$$

The first relation in (15) follows from these and (18). The second relation in (15) can be proved by similar manner. Proof of Theorem 3.1 is completed.

From Theorem 2.1 and Lemma 3.1 taking into account (19) immediately we have the following

Theorem 3.2:

Under alternatives (2) the random variable $(L_N(Z) - NA) / \sigma(g_i) \sqrt{N}$ has asymptotically normal distribution with expectation $\sqrt{N} B_N(g_i) / \sigma(g_i)$ and variance 1.

Note that

$$\sqrt{N} B_N(g_i) / \sigma(g_i) = \delta^2(n) \sqrt{\frac{n(k+1)}{2}} \rho(g_i) d^2, \quad (20)$$

where

$$\rho(g_i) = \frac{k^{2i-3}}{2(k+1)\sigma_0^2(g)}. \quad (21)$$

Asymptotic Efficiency.

We investigate here the asymptotic efficiency (AE) of the goodness of fit tests based on statistics of type (4). There are several approaches to define AE of tests which differ by the conditions imposed on the asymptotic behaviors of ω , the power β and the sequence of alternatives $H_{n,1}$. The most common is the Pitman's approach. The alternatives (2)

with $\delta(n) = ((k+1)n)^{-1/4}$ are Pitman's alternative. The Pitman asymptotic relative efficiency (ARE) of a test relative to another test is defined to be the limit of the inverse ratio of the sample sizes required to obtain the same limiting power at a sequence of alternative converging to the hypothesis. The limiting power should be a value between the limiting sizes ω and the maximum power 1, in order that it can give information about the power of the test. When this converges to a number in $(\omega, 1)$ then a measure to the rate of this convergence called the efficacy can be computed. Following the idea of (Fraser, 1957) we define efficacy of the test based on statistic, say, L as

$$e(L) = \frac{\mu_L^4}{\sigma_L^2} \quad (22)$$

Under Bahadur's approach the power $\beta < 1$ and the alternative H_1 are fixed, more precisely, H_1 does not approach H_0 and the test is characterized by the rate of decrease of size ω . Similarly, one can fix ω and H_1 and measure the performance of the test by

the rate of convergence of β to 1 this is Hodges Lehman approach. Finally, one can consider two intermediate settings: (i) $0 < \beta < 1$ is fixed while $\omega \rightarrow 0$ and $H_1 \rightarrow H_0$ "not too fast". (ii) $\omega > 0$ is fixed while $\beta \rightarrow 1$ and $H_1 \rightarrow H_0$ "not too fast". These situations give rise to the concept of intermediate asymptotic efficiencies (IAE) due to (Kallenberg, 1983) the reader is also suggested to go through paper (Ingol, 1999). Sometimes, ω -IAE is used in the first case i.e. intermediate between Pitman's and Bahadur's settings and β -IAE in the second case i.e. intermediate between Pitman's and Hodges-Lehman's settings (Ivchenko & Mirakhmedov, 1995). Investigation of the AE of the test is based on the probabilistic limit theorems for the test statistics. The asymptotic normality result of proved by (Mirakhmedov, 2005) and Cramer's type large deviation theorem of (Mirakhmedov et al. 2011) of section (Asymptotic Normality) serve the purpose in this paper. Let $\beta_N(g_i)$ be the power of L_N with size $\omega > 0$ and $\varsigma_N(L_N) = \sqrt{N}(A_{1N}(g_i) - A_0(g_i)) / \sigma_0(g_i)$, $u_\omega = \Phi^{-1}(1 - \omega)$.

Theorem 4.1:

Let H_1 be specified by (2) with $\delta(n) = (n(k+1))^{-1/4}$ Then

(i) the critical region of test L_N is $\{Z : Z \geq u_\omega \sqrt{N} \sigma_0(g_i) + NE_0(g_i(Z))\}$ and the asymptotic power is

$$\Phi\left(\frac{d^2}{\sqrt{2}} \rho(k; g_i) - u_\omega\right) \quad \text{where} \quad \rho^2(k; g_i) = \frac{k^{i-3}}{2(k+1)\sigma_0^2(g_i)} \quad \text{and}$$

$$\sigma_0^2(g_i) = \text{Var}(g_i(Z)) - k^{-1}(\text{cov}(g_i(Z), Z))^2, \quad \Phi \text{ is the standard normal distribution function and } u_\omega = \Phi^{-1}(1 - \omega)$$

(ii) Asymptotically most powerful test is the Greenwood test based on statistics $V_N^2 = \sum_{m=1}^N Z_{m,k}^2$ for which the critical region is $\{Z : Z \geq u_\omega \sqrt{2n(k+1)} + n(k+1)\}$ and the asymptotic power is $\Phi\left(\frac{d^2}{\sqrt{2}} - u_\omega\right)$.

(iii) Asymptotic relative efficiencies of L_N w.r.t. Greenwood test is $\rho^2(k; g_i)$.

Proof: By putting $\ell_N^0(Z) = \frac{L_N(Z) - NA_0(g_i)}{\sqrt{N}\sigma_0(g_i)}$, by Theorem 2.4

$$P_0\{L_N(Z) > c\} = P_0\left\{\ell_N^0(Z) > \frac{c - NA_0(g_i)}{\sqrt{N}\sigma_0(g_i)}\right\} = 1 - \Phi\left\{\frac{c - NA_0(g_i)}{\sqrt{N}\sigma_0(g_i)}\right\} + o(1)$$

$$= \Phi(-u_\omega) + o(1)$$

where $c = u_\omega \sqrt{N}\sigma_0(g_i) + NA_0(g_i)$, the first relation in (i) is proved.

By using the above results, Theorem 2.4 and Theorem 3.2 we get power β_N as given by

$$\begin{aligned}\beta_N &= P_1\{L_N(Z) > c\} = P_1\left\{\frac{L_N(Z) - NA_1(g_i)}{\sqrt{N}\sigma_0(g_i)} > \frac{c - NA_1(g_i)}{\sqrt{N}\sigma_0(g_i)}\right\} \\ &= P_1\left\{\frac{L_N(Z) - NA_1(g_i)}{\sqrt{N}\sigma_0(g_i)} > u_\omega - \varsigma_N\right\} \\ &= \Phi(\varsigma_N(g_i) - u_\omega) + o(1)\end{aligned}$$

Where for our statistics $\varsigma_N(L_N) = \frac{n(k+1)}{2} \delta^2(n) \rho^2(k; g_i) d^2(1+o(1))$.

Since $E(Z) = \text{Var}(Z) = k$, $\rho(V_N^2) = 1$, $E_0(V_N^2) = k(k+1)$ and $\text{Var}_0(V_N^2) = 2k(k+1)$

We have $\varsigma_N(V_N^2) = \sqrt{\frac{n(k+1)}{2}} \delta^2(n) d^2(1+o(1))$. From this we can easily find out the rejection region and power which is given by $\beta_N = \Phi(\varsigma_N(V_N^2) - u_\omega) + o(1)$. This for the Pitman's alternatives i.e. $\delta(n) = (n(k+1))^{-1/4}$ yields $\beta_N = \Phi\left(\frac{d^2}{\sqrt{2}} - u_\omega\right)$. Thus, part (ii) is proved.

Now

$$\begin{aligned}e_N(L_N(Z)) &= -\text{Log } P_1\{L_N(Z) \geq NE_1 L_N(Z)\} \\ &= -\text{Log } P_1\left\{\frac{L_N(Z) - NE_0(L_N(Z))}{\sqrt{N}\text{Var}_0(L_N(Z))} \geq \frac{NE_1(L_N(Z)) - NE_0(L_N(Z))}{\sqrt{N}\text{Var}_0(L_N(Z))}\right\} \\ &= -\text{Log } \Phi(\varsigma_N(L_N)) + o(\varsigma_N^2(L_N)).\end{aligned}$$

For our statistics, pitman's alternatives and due to the relation $-\text{Log}(-Z) = 2^{-1} Z^2(1+o(1))$ we get $e_N^\omega(L_N(Z)) = \sigma_0^{-2}(g_i) k^{2i-3} (8(k+1))^{-1}$ and $e_N^\omega(V_N^2) = (4)^{-1}$ it yields

$$\text{ARE}(L_N, V_N^2) = \frac{e_N^\omega(L_N)}{e_N^\omega(V_N^2)} = \frac{k^{2i-3}}{2(k+1)\sigma_0^2(g_i)} = \rho^2(k; g_i).$$

This completes the proof of the Theorem 3

Theorem 4.2:

Let H_1 be specified by (2) and $\delta(n) \rightarrow 0$, $\sqrt{nk}\delta^2(n) \rightarrow \infty$. If

- (i) $E|g_i(Z)|^{2+\epsilon} < \infty$ for some $\epsilon > 0$ then in the family of alternatives P_{\log} or
- (ii) $E\left(\exp\{H|g_i(Z)|\}\right) < \infty$ for some $H > 0$ then in the family P_{al}

$$\frac{e_N^\omega(L_N)}{e_N^\omega(V_N^2)} = \frac{d^4 k^{2i-3} \sigma_0^{-2}(g_i)}{2(k+1)} (1+o(1)).$$

Proof: For statistics $L_n(Z)$ we have $-\zeta_N(L_N(Z)) = (\sqrt{N}\delta^2(n)d^2\sigma_0^1(g_i))$

So

$$\begin{aligned} e_N^\omega(L_N(Z)) &= -\text{Log } \Phi(-\zeta_N(L_N(Z))) + o(\zeta_N(L_N(Z))) \\ &= \frac{1}{2}N\delta^4(n)d^4\sigma_0^{-2}(g_i)(1+o(1)) \end{aligned}$$

Which can be written as

$$\begin{aligned} \frac{e_N^\omega(L_N(Z))}{n(k+1)\delta^4(n)} &= \frac{d^4k^{2i-3}\sigma_0^{-2}(g_i)}{2(k+1)}(1+o(1)) \\ &= d^4\rho^2(k; g_i)(1+o(1)), \quad j=1, 2. \end{aligned}$$

This completes the proof.

The statistic $L_N(Z)$ satisfies the condition (ii) of Theorem 4.2, which is known as Cramer's condition, whereas the Greenwood statistics $V_N^2(Z)$ does not satisfies that condition. Due to this fact and the asymptotic normality theorems stated above we argue that Greenwood test, satisfying condition (i) of Theorem 4.2, is still most efficient in the weak IAE (commonly known as Kallenberg's IAE) sense i.e. in the family of $P_{1/6}$ alternatives while the log-spacings statistics $L_N(Z)$, satisfying Cramer's condition, is much more efficient in strong IAE sense i.e. in the family of alternatives P_{all} except $P_{1/6}$ as compared to Greenwood statistics. These improvements in the efficiency properties of Greenwood test are discussed in detail in the unpublished PhD thesis of the author.

Remarks. The IAE considered above is somewhere between pitman's and Bahadur's efficiency. We may consider IAE between Pitman's and Hodges-Lehmann's by taking asymptotic value of $e_n^\beta = -\text{Log } P_1\{L_N(Z) \leq NE_0L_N(Z)\}$ as a measure of efficiency i.e. in this case we study the rate of convergence of the second type error. Such approach was introduced by (Ivchenko & Mirakhmedov, 1995). The analysis of $B-IAE$ proceeds along the same lines as $\omega-IAE$. Due to Theorem 3.1

$$P_1\{L_N(D_{m,n}^{(k)}) < u\} = P_0\{\hat{L}_N(T) < u\}$$

Where

$$\hat{L}_N(T) = \sum_{m=1}^N \hat{g}_i(nT_{m,n}^{(k)}), \quad \hat{g}_i(x) = g_i(x(1-\tau(n,u))).$$

Recall $\tau \equiv \tau(n, u) = dl(u)\delta(n) + O(\delta^2(n))$.

We have

$$\begin{aligned} E \exp\{H\hat{g}_i(Z)\} &= E \exp\{Hg_i(Z(1-\tau))\} = \frac{1}{1-\tau} \int_0^\infty \exp\left\{Hg - i(u) - \frac{u}{1-\tau}\right\} du \\ &= \frac{1}{1-\tau} \int_0^\infty \exp\left\{Hg_i(u) - u - \frac{u\tau}{1-\tau}\right\} du \leq \frac{1}{1-\tau} \int_0^\infty \exp\{Hg_i(u) - u\} du \leq 2E \exp\{Hg_i(Z)\}. \end{aligned}$$

Thus the conditions of Theorem 2.1 and 2.2 stated in section (Asymptotic Normality) are satisfied and large deviation result of Theorem 2.3 can be used for the analysis of

$e_N^\beta(g_i)$. The corresponding calculations show that for $e_N^\beta(g_i)$ the assertions of Theorems 4.1 and Theorem 4.2 are still true i.e. $e_N^\beta(g_j)$ and $e_N^\omega(g_i)$ are asymptotically equivalent.

Conclusions

In the Table below the values of $\rho^2(k; g_i)$ are given for the different values of k . By using the relations in (21) we get the ARE of statistics $L_N(Z)$. Results are tabulated as under. The table is obtained from simulation through Mat lab

Table 1 comparison of correlation coefficient for different order (k)

K	$\rho^2(k; g_1)$	$\rho(k; g_2)$
1	0.3750	0.8571
2	0.5714	0.9000
3	0.6750	0.9231
4	0.7385	0.9375
5	0.7813	0.9474
10	0.8798	0.9706
20	0.9368	0.9844
50	0.9739	0.9935
100	0.9868	0.9967
150	0.9912	0.9978

From the table we see that the value of $\rho^2(k; g_i)$ increases as k increases; although $\lim_{k \rightarrow \infty} \rho^2(k; g_i) = 1$ but the table reveals that enough efficiency is achieved even for $n=20$. Thus we found that

1. The test based on statistics $L_N(Z)$, can detect alternatives (2) at a distance $\delta(n) \geq (n(k+1))^{-1/4}$ and these statistics have higher efficiency in Pitman's and Kallenberg's sense compared to their counterparts based on simple spacings.
2. Asymptotic efficiency of test $L_N(Z)$ is defined by the value of $\rho^2(k; g_i)$ which is correlation coefficient between statistic $L_n(Z)$ and Greenwood's statistic.
3. The Pitman's and Kallenberg's efficiencies coincide and are defined by $\rho^2(k; g_i)$.
4. Among the tests defined in (5), those satisfying Cramer's condition are more efficient compared to those do not satisfy Cramer's condition.
5. The $w-IAE$ and $\beta-IAE$ are asymptotically equivalent.
6. It can be observed that entropy statistics is more efficient than log-statistics for small values of n but for large n the efficiency of both statistics are nearly equivalent.

7. The table reveals that for large enough k the Greenwood statistics is not the only most efficient statistics rather logarithm statistics are as efficient as Greenwood statistics as $k \rightarrow \infty$.

References

1. Cressie, N. (1976). On the logarithms of high -order spacings. *Biometrika*. 63: 343-355.
2. Czekala, F. (1999). On two tests based on disjoint m-spacings. *Applicationes Mathematicae*. 25(4): 359-373.
3. Darling D.A. (1953). On a class of problems relating to the random division of an interval. *Annals Mathematical Statistics*. 24: 239-253.
4. Del Pino. (1979). On the asymptotic distribution of k-spacing with application to goodness of fit tests. *Annals of Statistics*. 7: 1058-1065.
5. Fraser, D.A.S. *Nonparametric methods in Statistics*. John Wiley, New York, 1957.
6. Gebert, J.B. and Kale, B.K. (1969). Goodness of fit test based on discriminatory information. *Statistische Hefte* .10:192-200.
7. Holst, L. and Rao, J.S.(1981). Asymptotic spacing theory with application to two sample problem. *Canadian Journal of Statistics*. 9: 603-610.
8. Inglot, T. (1999). Generalized intermediate efficiency of goodness-of-fit tests. *Mathematical Methods in Statistics*. 8: 487-509.
9. Ivchenko GI & Mirakhmedov Sh M. (1999). Large deviations and intermediate efficiency of decomposable statistics in a multinomial scheme. *Mathematical Methods in Statistics*. 4(3):294-311.
10. Jammalamadaka, S.R. and Tiwari, R.C. (1985). Asymptotic Comparison of Three tests for the Goodness of fit. *Journal of Statistical Planning and Inference*. 2: 295-304.
11. Jammalamadaka SR & Tiwari, R.C. (1987). Efficiency of some disjoint spacings tests relative to a chi-square test. *Perspectives and New Directions in Theoretical and Applied Statistics*, eds. J. Vilaplana and W. Wertz, New York, John Wiley: 311-317.
12. Kale, B.K. (1969). Unified derivation of tests of goodness of fit based on spacings. *Sankhya Ser. A* 31:43-48.
13. Kallenberg, W.C.M. (1983). Intermediate efficiency. *Annals of Statistics*. 11: 170-182.
14. Keener, J.P. *Principles of Applied Mathematics. Transformation and Approximation*.USA:1987.
15. Mirakhmedov Sh M. (2005). Lower estimation of the remainder term in CLT for a sum of the functions of k-Spacings. *Statistics and Probability Letters*. 73: 411-424.

16. Mirakhmedov, Sh. M and Jammalamadaka, S. R. (2013). Higher- order expansions and efficiencies of tests based on the spacings. *Journal of Non-parametric Statistics*. 1-21
17. Mirakhmedov, Sh. M, ; Tirmisi, S. I. and Naeem, M. (2011). Cramer's type large deviation theorem for a sum of higher ordered spacings. *Metrika*. 74: 33-54.
18. Morgan, K. and Jammalamadaka, S.R. (1981). Limit theory and efficiencies for tests based on higher order spacings. *Proceedings of the Indian Stat. Inst. Golden Jubilee Inter. Conference Statistics: Application and New Directions*. 333-352.
19. Moran, P.A.P. An introduction to probability theory. Oxford Univ. Press: 1984.
20. Pyke, R. (1965). Spacings. *Journal of Royal Statistical Society Ser. B* (27): 395-449.
21. Rao, J. S. and Sethuraman, J. (1975). Weak convergence of empirical distribution functions of random variables subject to perturbation and scale factors. *Annals of Statistics*. 299-313.