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  
aeemtazkeer@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 $Z_1, Z_2, ..., Z_{n-1}$ 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

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

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 $D_m = Z_{mk} - Z_{(m-1)k}$, $m = 1, 2, ..., N'$, $D_{N'+1}^{(k)} = 1 - Z_{N'k}$, 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 $D = (D_1^{(k)}, D_2^{(k)}, ..., D_N^{(k)})$ 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
where \( g(z) = 1 + dt(z) \delta(n) \), \( 0 \leq z \leq 1 \),
\hspace{1cm} (2)

Assume that the order of spacing i.e. \( k \) and \( n \) may tends to infinity jointly. We consider
test based on the statistic
\[ L_N(D) = \sum_{m=1}^{N} g_i\left(nD_{m}^{(k)}\right), \quad i = 1, 2. \]
\hspace{1cm} (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 \( \left\{ D_{m,n}^{(k)} \right\} \). 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 $0 = Z_0 \leq Z_1 \leq \ldots \leq Z_{n-1} \leq Z_n = 1$ be an ordered sample from uniform [0,1] distribution and $T_{m,n}^{(k)}$ their non-overlapping $k$-spacings. Let $f_{m,n}(u)$, $m = 1, 2, \ldots, N$, be measurable functions. Consider the statistics

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

(5)

Let $Z$ and $Z_{m,k}$, $m = 1, 2, \ldots, N$ 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, $\{f_{m,n}(u), m = 1, 2, \ldots, N\}$ 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 $T_m, m = 1, 2, \ldots, n$ it is well known

$$L(T_1, T_2, \ldots, T_n) = L \left( Z_1, Z_2, \ldots, Z_n \mid \sum_{m=1}^{n} Z_m = 0 \right),$$

where $L(X)$ represent the distribution of a random vector $X$ and $Z_1, Z_2, \ldots, Z_n$ are independent random variables of common exponential distribution with mean 1.

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

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

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

(6)

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

$$\sigma_N^2 = \left( 1 - \rho_N^2 \right) Var Q_N(Z)$$

and clearly $Q_N(T) = G_n(T) - A_N$ also, by the well-known inequality $\beta_{3,n} \leq \beta_{3,N}^{1/2}$ 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

$$\sup_{m} \left| P_N(Z) - \Phi(Z) \right| \leq C \beta_{3,N} \quad \text{as} \quad N \to \infty,$$

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

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

**Theorem 2.2**

If

Pak.j.stat.oper.res. Vol.XIV No.4 2018 pp839-851 841
\[
\lim_{N \to \infty} \frac{1}{N} \sigma^2_x > 0 \quad (7)
\]

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

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 Theorem2.1 gets the form

\[
\left| P_N(Z) - \Phi(Z) \right| \leq C_1 N^{-\theta/2} \quad (10)
\]

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

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

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

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 \( \sigma^2_N \).

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

**Theorem 2.3:**

For all \( Z \geq 0 \), \( Z = o\left(\sqrt{N}\right) \)

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

where \( \Lambda_N(u) = \lambda_{0N} + \lambda_{1N} u + \ldots \) is a special Cramer’s type power series; for \( N \) large enough

\[
\left| \lambda_{jN} \right| \leq C < \infty , \ j = 0,1,2,\ldots
\]

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

\[
f_{m,n}(u) = f(u) = g_i(u), \quad i = 1,2, \quad (11)
\]
A Note on Asymptotical Efficiency of the Goodness of Fit Tests Based on Disjoint k-Spacings Statistic

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 \( N \to \infty \). Here

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

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

where \( \gamma \) is the well-known Euler function.

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

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

(12)

here and what follows \( P_j, E_j \) and \( \text{Var}_j \) 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 \( D_{m,n}^{(k)} \), \( m=1, 2, \ldots, N \) 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)

\[
n D_{m,n}^{(k)} = n T_{m,n}^{(k)} \left( 1 - l \left( \frac{m}{N} \right) \delta(n) + O_p \left( \delta^2(n) \right) \right),
\]

(13)

where \( O_p(.) \) 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

\[
\frac{1}{N} \sum_{m=1}^{N} p \left( \frac{m}{N} \right) = \int_{0}^{1} p(u) du + o(1), \quad \text{as} \quad N \to \infty.
\]

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
Then
\[ E_i L_N(Z) = N\left(A(g_i) + B_N(g_i)\right) + o\left(N\delta^2(n)\right), \quad \text{Var}_i L_N(Z) = N\sigma^2(g_i)(1 + o(1)), \] (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
\[ \rho(g_i) = \text{corr}\left(h(Z), Z^2 - 2(k+1)Z + k(k+1)\right) \]

**Proof.** Let \( E_i L_N(Z) = A_{iN}(g_i). \) Due to (12), Lemma 3.1 and Lemma 3.2 we have
\[ A_{iN}(g_i) = \int_0^1 E_{g_i}\left(Z\left(1 - dl(u)\delta(n) + O_p(\delta^2(n))\right)\right)du + o(1) = \int_0^1 \psi(u;k,\tau)du + o(1), \] (16)
where
\[ \tau \equiv \tau(n,u) = dl(u)\delta(n) + O\left(\delta^2(n)\right), \]
and
\[ \psi(u;k,\tau) = \gamma(t)(1 - \tau(n,u)))\gamma(t)dt. \]
We have
\[ \psi(u;k,\tau) = \frac{1}{(1-\tau)^k} \int_0^\infty g_i(v)\gamma_k(v)\exp\left\{-v\tau/(1-\tau)\right\}dv. \]
According to Laplace’s method of asymptotic expansion of integrals (Keener 1987) we can use the following expansion
\[ \exp\left\{-v\tau/(1-\tau)\right\} = 1 - \frac{v\tau}{1-\tau} + \frac{(v\tau)^2}{2(1-\tau)^2} + O\left((v\tau)^3\right), \]
in order to get an asymptotic expansion of the integral in (16). By using the above and
\[ (1-\tau)^k = 1 + k\tau + \frac{1}{2}k(k+1)\tau^2 + O\left((k\tau)^3\right), \]
since under (14) \( k\tau = O(1), \) we find
\[ \psi(u;k,\tau) = \frac{1}{\Gamma(k)} \int_0^\infty g_i(v)v^{k-1}e^{-v}\left(1 - \Upsilon - \frac{v\tau}{1-\tau} + \frac{(v\tau)^2}{2(1-\tau)^2} + O\left((v\tau)^3\right)\right)dv \]
\[ = Eg(Z) - \tau E\left((Z-k)g_i(Z)\right) + \frac{1}{2}\tau^2E\left(g_i(Z)\left(Z^2 - 2(k+1)Z + k(k+1)\right)\right) + O\left(\tau^3Eg_i(Z)Z^3\right) \]
By putting this result into (16), use (17) instead of \( \tau \) and take into account (3) to get
\[ A_{iN}(g_i) = NA(g_i) + \frac{N\delta^2(n)}{2}E\left(g_i(Z)\left(Z^2 - 2(k+1)Z + k(k+1)\right)\right) + O\left(N\delta^3(n)Eh(Z)Z^3\right) \] (18)
By direct calculations we find
\[ E\left(Z^2 - 2(k+1)Z + k(k+1)\right) = 0, \]
\[ E\left(Z - k\right)\left(Z^2 - 2(k+1)Z + k(k+1)\right) = 0, \]
\[ \text{Var}\left(Z^2 - 2(k+1)Z + k(k+1)\right) = 2(k+1), \]
A Note on Asymptotical Efficiency of the Goodness of Fit Tests Based on Disjoint k-Spacings Statistic

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

\[ \rho(g_i) = \text{corr} \left( h(Z), Z^2 - 2(k+1)Z + k(k+1) \right), \]

\[ E g_i(Z) \left( Z^2 - 2(k+1)Z + k(k+1) \right) = \sigma(g_i) \sqrt{k(k+1)} \rho(g_i). \quad (19) \]

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

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

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 \(\omega\), the power \(\beta\) and the sequence of alternatives \(H_{\alpha,1}\). 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 \(\omega\) 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

\[ e(L) = \frac{\mu_L^4}{\sigma_L^4} \quad (22) \]

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 \(\omega\). Similarly, one can fix \(\omega\) and \(H_1\) and measure the performance of the test by
the rate of convergence of $\beta$ to 1 this is Hodges Lehman approach. Finally , one can consider two intermediate settings : (i) $0 < \beta < 1$ is fixed while $\omega \to 0$ and $H_1 \rightarrow H_0$ not too fast". (ii) $\omega > 0$ is fixed while $\beta \to 1$ 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 (Inglot, 1999). Sometimes, $\omega-\text{IAE}$ is used in the first case i.e. intermediate between Pitman's and Bahadur's settings and $\beta-\text{IAE}$ 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 $\sigma_N(L_N)=\sqrt{N}(A_N(g_i)-A_0(g_i))/\sigma_0(g_i), u_\omega=\Phi^{-1}(1-\omega)$.

**Theorem 4.1:**

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

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

$$
\Phi\left(\frac{d^2}{\sqrt{2}} \rho(k;g_i) - u_\omega \right)
$$

where $\rho^2(k;g_i) = \frac{k^{i-3}}{2(k+1)\sigma_0^2(g_i)}$ and $\sigma_0^2(g_i) = \text{Var}(g_i(Z)) - k^{-1} \left( \text{cov}(g_i(Z), Z) \right)^2$. $\Phi$ 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 $V_N^2 = \sum_{m=1}^{N} Z_{m,k}$ for which the critical region is $\left\{Z: Z \geq u_\omega \sqrt{2N(k+1)} + n(k+1) \right\}$ 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 $E_N^c(Z) = \frac{L_N(Z) - NA_0(g_i)}{\sqrt{N} \sigma_0(g_i)}$, by Theorem 2.4

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

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 $\beta_N$ as given by

$$\beta_N = P_1 \left\{ L_N(Z) > c \right\} = P_1 \left\{ \frac{L_N(Z) - NA_i(g_i)}{\sqrt{N}\sigma_0(g_i)} > \frac{c - NA_i(g_i)}{\sqrt{N}\sigma_0(g_i)} \right\}$$

$$= P_1 \left\{ \frac{L_N(Z) - NA_i(g_i)}{\sqrt{N}\sigma_0(g_i)} > u_{\alpha} - \zeta_N \right\}$$

$$= \Phi(\zeta_N - u_{\alpha}) + o(1)$$

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

Since $E(Z) = Var(Z) = k$, $\rho(V^2_N) = 1$, $E_o(V^2_N) = k(k+1)$ and $Var_o(V^2_N) = 2k(k+1)$

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

Now

$$e_N(L_N(Z)) = -\log P_1 \left\{ L_N(Z) \geq NE_i L_N(Z) \right\}$$

$$= -\log P_1 \left\{ \frac{E_k(Z) \geq \frac{NE_k(L_N(Z))}{\sqrt{N} Var_o(L_N(Z))}}{NE_i L_N(Z)} \right\}$$

$$= -\log \Phi(\zeta_N(L_N)) + o(\zeta^2_N(L_N))$$

For our statistics, pitman’s alternatives and due to the relation $-\log(-Z) = 2^{-1} Z^2 (1 + o(1))$ we get $e^o_N(L_N(Z)) = \sigma_0^{-2} (g_i) k^{2i-3} (8(k+1))^{-1}$ and $e^o_N(V^2_N) = (4)^{-1}$ it yields

$$ARE(L_N, V^2_N) = \frac{e^o_N(L_N)}{e^o_N(V^2_N)} = \frac{k^{2i-3}}{2(k+1)} \frac{\sigma_0^{-2}(g_i)}{\sigma_0^{-2}(g_i)} = \rho^2(k; g_i).$$

This completes the proof of the Theorem 3

**Theorem 4.2:**

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

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

(ii) $E\left( \exp\left\{ H \left| g_i(Z) \right| \right\} \right) < \infty$ for some $H > 0$ then in the family $P_{ad}$

$$\frac{e^o_N(L_N)}{e^o_N(V^2_N)} = \frac{d^4 k^{2i-3} \sigma_0^{-2} (g_i)}{2(k+1)} (1 + o(1)).$$
**Proof:** For statistics \( L_n(Z) \) we have \(-\zeta_N(L_N(Z)) = \left( \sqrt{N} \delta^2(n) d^2 \sigma_0^1(g_i) \right)\). So
\[
e_N^\omega(L_N(Z)) = -\text{Log} \Phi(-\zeta_N(L_N(Z))) + o(\zeta_N(L_N(Z)))
= \frac{1}{2} N \delta^4(n) d^4 \sigma_0^{-2}(g_i) (1 + o(1))
\]
Which can be written as
\[
\frac{e_N^\omega(L_N(Z))}{n(k + 1) \delta^4(n)} = \frac{d^4 k^{2i-3} \sigma_0^{-2}(g_i)}{2(k + 1)} (1 + o(1))
= d^4 \rho^2(k; g_i) (1 + o(1)), \quad j = 1, 2.
\]
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 \( e_n^\beta = -\text{Log} P_1\left\{L_N(Z) \leq N \epsilon_0 L_N(Z) \right\} \) 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
\[
P_1\left\{L_N(D_{m,n}) < u \right\} = P_0\left\{\hat{L}_N(T) < u \right\}
\]
Where
\[
\hat{L}_N(T) = \sum_{i=1}^N \hat{g}_{i,1}(nt_{m,n}) \quad \hat{g}_{i,1}(x) = g_i(x(1-\tau(n,u))).
\]
Recall \( \tau \equiv (n,u) = d \lambda(u) \delta(n) + O(\delta^2(n)) \).

We have
\[
E \exp\left\{ H_{g_i}(Z) \right\} = E \exp\left\{ H_{g_i}(Z(1-\tau)) \right\} = \frac{1}{1-\tau} \int_0^\infty \exp\left\{ H_{g_i} - \frac{u}{1-\tau} \right\} du
\]
\[
= \frac{1}{1-\tau} \int_0^\infty \exp\left\{ H_{g_i}(u) - \frac{u\tau}{1-\tau} \right\} du \leq \frac{1}{1-\tau} \int_0^\infty \exp\left\{ H_{g_i}(u) - u \right\} du \leq 2E \exp\left\{ H_{g_i}(Z) \right\}.
\]

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
$e^0_N(g_i)$. The corresponding calculations show that for $e^0_N(g_i)$ the assertions of Theorems 4.1 and Theorem 4.2 are still true i.e. $e^0_N(g_i)$ and $e^w_N(g_i)$ 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)**

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

From the table we see that the value of $\rho^2(k;g_i)$ increases as $k$ increases; although $\lim_{k\to\infty}\rho^2(k;g_i)=1$ 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 $k \to \infty$.

References

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


