Title
Bahadur efficiencies of spacings tests for goodness of fit

Journal
Annals of the Institute of Statistical Mathematics, 41(3)

ISSN
0020-3157

Authors
Zhou, Xian
Jammalamadaka, S Rao

Publication Date
1989-09-01

DOI
10.1007/bf00050667

Peer reviewed

Ann. Inst. Statist. Math.
Vol. 41, No. 3, 541-553 (1989)

BAHADUR EFFICIENCIES OF SPACINGS TESTS FOR GOODNESS OF FIT

XIAN ZHOU AND S. RAO JAMMALAMADAKA

Statistics and Applied Probability Program, University of California, Santa Barbara, CA 93106, U.S.A.

(Received July 28, 1988; revised November 21, 1988)

Abstract. This paper is concerned with the exact Bahadur efficiencies of spacings statistics. For a general class of statistics based on a fixed number of spacings, the explicit forms of the exact slopes are derived, and it is shown that the sum of the logarithms of spacings is optimal in this class. Some results are extended to the case where the number of spacings increase with the sample size to infinity.

Key words and phrases: Bahadur efficiency, exact slope, large deviation, spacings, goodness of fit.

1. Introduction

Tests based on the observed frequencies as well as those based on spacings provide two basic approaches for the goodness of fit problem. The efficiencies of these tests have been studied in the literature. For example, Sethuraman and Rao (1970), Del Pino (1979), Kuo and Rao (1981), and Jammalamadaka *et al.* (1986) studied the Pitman efficiencies for various spacings tests, while Jammalamadaka and Tiwari (1987) considered the Pitman efficiencies of some spacings tests relative to a chi-square test. Hoeffding (1965) showed the likelihood ratio test based on multinomial frequencies to be optimal, in the Bahadur sense, for a fixed number of cells. Quine and Robinson (1985) studied both Pitman and Bahadur efficiencies for the case when the number of cells increases to infinity. This paper is concerned with the exact Bahadur efficiencies of spacings tests for the two corresponding cases.

Let $X_1,\dots,X_n$ be an ordered sample from a continuous distribution function (d.f.) F. The goodness of fit problem is to test the null hypothesis H_0: $F=F_0$, where F_0 is specified, against the alternative H_1: $F\neq F_0$. By applying the probability integral transformation $x\to F_0(x)$ on all the data, without loss of generality, F_0 can be assumed to be uniform on [0, 1] and F to be supported in [0, 1]. Define spacings by

$$D_i^{(n)} = X_{[n\lambda_i]} - X_{[n\lambda_{i-1}]}, \qquad i = 1,\dots,k ,$$

where $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_{k-1} < \lambda_k = 1$ and $[\cdot]$ denotes the integer part. Write $D^{(n)} = (D_1^{(n)},\dots, D_{k-1}^{(n)})$. We will consider spacings tests that reject H_0 for large values of $J_n(D^{(n)})$ where $J_n(\cdot)$ is a non-negative function defined on the $(k-1)$-dimensional simplex

$$S_{k-1} = \left\{ z = (z_1,\dots,z_{k-1})\colon 0 < z_i < 1,\, i = 1,\dots,k-1;\, \sum_{i=1}^{k-1} z_i < 1 \right\}.$$

Two tests of particular interest are

$$I_k = I_k(\nu^0, D^{(n)}) = \sum_{i=1}^{k} \nu_i^0 \log (\nu_i^0 / D_i^{(n)}) \qquad \text{and}$$

$$Q_k^2 = Q_k^2(D^{(n)}, \nu^0) = \sum_{i=1}^{k} (D_i^{(n)} - \nu_i^0)^2 / \nu_i^0 ,$$

where $\nu_i^0 = \lambda_i - \lambda_{i-1}$, $\nu^0 = (\nu_1^0,\dots,\nu_{k-1}^0)$ and $D_k^{(n)} = 1 - \sum_{i=1}^{k-1} D_i^{(n)}$. The Bahadur efficiency of one test relative to another is defined by the ratio of their "exact slopes" (cf. Section 3).

The organization and results of this paper are as follows: in Section 2, a large deviation result for Dirichlet distributions is derived, which is crucial in finding the Bahadur efficiencies of spacings tests. Section 3 deals with the case of fixed k, and

(i) gives the exact slopes of $J_n(D^{(n)})$ and their explicit forms for I_k and Q_k^2;

(ii) shows that I_k has the highest Bahadur efficiency in a general family and that the Bahadur efficiency of I_k relative to Q_k^2 is strictly greater than one for most alternatives.

In Section 4, we consider the case when k is allowed to increase with n, to infinity. The exact slopes of I_k and Q_k^2 are obtained again and it is shown that the Bahadur efficiency of I_k relative to Q_k^2 is infinity in this case.

2. A large deviation theorem for Dirichlet distributions

For $z \in S_{k-1}$, we always write $z_k = 1 - z_1 - \cdots - z_{k-1}$.

For $\nu \in S_{k-1}$, $\boldsymbol{D}(n\nu)$ will denote the Dirichlet distribution with density on S_{k-1} given by

$$p_n(z \mid \nu) = \frac{\Gamma(n)}{\Gamma(n\nu_1)\cdots\Gamma(n\nu_k)} \prod_{i=1}^{k} z_i^{n\nu_i - 1}$$

and the corresponding probability measure will be denoted by $P_n(A \mid \nu)$ for any measurable subset A of S_{k-1}. Define

$$I_k(v, z) = \sum_{i=1}^{k} v_i \log \frac{v_i}{z_i} \quad \text{for} \quad v, z \in S_{k-1}$$

and

$$I_k(v, A) = \inf_{z \in A} I_k(v, z) \quad \text{for} \quad A \subset S_{k-1} .$$

Let $J_n(z)$ and $J(z)$ be non-negative functions defined on S_{k-1} and $\beta = \sup\{J(z): z \in S_{k-1}\}$ ($\beta \le \infty$). Define for $t \in (0, \beta)$

$$A_n(t) = \{z: J_n(z) \ge t\}, \qquad A(t) = \{z: J(z) \ge t\} .$$

The main result of this section is

THEOREM 2.1. *Assume*
(A1) *$J_n(z) \to J(z)$ uniformly in S_{k-1},*
(A2) *$J(z)$ is continuous and has no local maximum values in S_{k-1}, and*
(A3) *$v^{(n)} \to v$ (i.e., $v_i^{(n)} \to v_i$, $i = 1, \ldots, k$) as $n \to \infty$.*
Then for any $t \in (0, \beta)$

$$\lim_{n \to \infty} -\frac{1}{n} \log P_n\{A_n(t) | v^{(n)}\} = I_k(v, A(t)) .$$

The proof follows from Lemmas 2.1 and 2.2. First note that by Stirling's formula, we have

$$\text{(2.1)} \qquad \sqrt{2\pi n}\left(\frac{n}{e}\right)^n < \Gamma(n+1) < 2\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \quad \text{for} \quad n \ge 1 .$$

LEMMA 2.1. *Let $A_n \subset S_{k-1}$ and $v^{(n)}$, $v \in S_{k-1}$ such that $v^{(n)} \to v$. If $I_k(v, A_n) \to d$ as $n \to \infty$ and if $\forall\, \varepsilon > 0$, $\exists$ a non-empty open set A_ε in S_{k-1} and an integer N such that $I_k(v, A_\varepsilon) < d + \varepsilon$ and $A_\varepsilon \subset A_n$ $\forall\, n > N$, then*

$$\lim_{n \to \infty} -\frac{1}{n} \log P_n(A_n | v^{(n)}) = d .$$

PROOF. Let $\max_{1 \le i \le k} v_i < a < 1$. Then $\exists\, N = N(v, a)$ such that $\forall\, n > N$, $n > \max_{1 \le i \le k} (1/v_i^{(n)})$ and $n^{-1} + a v_i < v_i^{(n)} < a^{-1} v_i$, $i = 1, \ldots, k$. By (2.1) and the equality $\Gamma(x+1) = x\Gamma(x)$, we can obtain $\forall\, n > N$,

$$
\begin{aligned}
p_n(z|v^{(n)}) &\le 2\left(\frac{n}{2\pi}\right)^{(k-1)/2} \prod_{i=1}^{k} (v_i^{(n)})^{1/2-nv_i^{(n)}} z_i^{nv_i^{(n)}-1} \\
&\le 2(n)^{(k-1)/2} \prod_{i=1}^{k} (av_i)^{-nv_i/a} z_i^{anv_i} \qquad (2.2)\\
&\le 2(n)^{(k-1)/2} e^{-anI_k(v,z)} a^{-n/a} \prod_{i=1}^{k} v_i^{nv_i(a-1/a)},
\end{aligned}
$$

where the second inequality holds because $av_i < v_i^{(n)} < 1$ and $nv_i^{(n)} < nv_i/a$, which imply $(v_i^{(n)})^{1/2-nv_i^{(n)}} < (v_i^{(n)})^{-nv_i^{(n)}} < (av_i)^{-nv_i^{(n)}} < (av_i)^{-nv_i/a}$. So

$$
\begin{aligned}
P_n(A_n|v^{(n)}) &= \int_{A_n} p_n(z|v^{(n)})dz \\
&\le 2(n)^{(k-1)/2} e^{-anI_k(v,A_n)} a^{-n/a} \prod_{i=1}^{k} v_i^{nv_i(a-1/a)},
\end{aligned}
$$

$$
\underline{\lim}_{n\to\infty} -\frac{1}{n}\log P_n(A_n|v^{(n)}) \ge ad + \frac{1}{a}\log a - \left(a - \frac{1}{a}\right)\sum_{i=1}^{k} v_i \log v_i .
$$

Letting $a \to 1-$ gives

$$
\lim_{n\to\infty} -\frac{1}{n}\log P_n(A_n|v^{(n)}) \ge d . \qquad (2.3)
$$

On the other hand, again by (2.1) we obtain

$$
p_n(z|v^{(n)}) \ge \frac{1}{2}\left(\frac{n}{2\pi}\right)^{(k-1)/2} e^{-nI_k(v^{(n)},z)} \prod_{i=1}^{k} (v_i^{(n)})^{1/2} .
$$

Thus, $\exists\, N_1 = N_1(v)$ such that $\forall\, n > N_1$,

$$
p_n(z|v^{(n)}) \ge e^{-nI_k(v^{(n)},z)} \;\forall\, z \in S_{k-1} . \qquad (2.4)
$$

Let $\varepsilon > 0$, A_ε and N_2 be as in the conditions of the lemma and $d(\varepsilon) = I_k(v, A_\varepsilon)$. Then, $\exists\, z \in A_\varepsilon$ such that $d(\varepsilon) < I_k(v,z) < d(\varepsilon) + \varepsilon$. Thus $B_\varepsilon = A_\varepsilon \cap \{z: d(\varepsilon) < I_k(v,z) < d(\varepsilon)+\varepsilon\}$ is a non-empty open subset of S_{k-1} so that $\int_{B_\varepsilon} dz > 0$. Since $\bar{B}_\varepsilon \subset S_{k-1}$, $I_k(v^{(n)},z) \to I_k(v,z)$ uniformly in $z \in \bar{B}_\varepsilon$. Thus, $\exists\, N_3 = N_3(v,\varepsilon)$ such that $\forall\, n > N_3$, $I_k(v^{(n)},z) < I_k(v,z) + \varepsilon$ $\forall\, z \in B_\varepsilon$ and so $\forall\, n > \max(N_1, N_2, N_3)$, by (2.4) and note that $B_\varepsilon \subset A_\varepsilon \subset A_n$

$$
\begin{aligned}
P_n(A_n|v^{(n)}) &\ge \int_{B_\varepsilon} e^{-nI_k(v^{(n)},z)}dz \ge \int_{B_\varepsilon} e^{-n(I_k(v,z)+\varepsilon)}dz \\
&\ge e^{-n(d(\varepsilon)+2\varepsilon)} \int_{B_\varepsilon} dz .
\end{aligned}
$$

Hence $\overline{\lim}_{n\to\infty} - (1/n) \log P_n(A_n|v^{(n)}) \le d(\varepsilon) + 2\varepsilon < d + 3\varepsilon \;\forall\; \varepsilon > 0$, which together with (2.3), proves the lemma. □

LEMMA 2.2. *If $J(\cdot)$ is continuous and has no local maximum values in S_{k-1}, then for each $v \in S_{k-1}$, $I_k(v, A(t))$ is continuous in $t \in (0, \beta)$.*

PROOF. Fix $t \in (0, \beta)$. By Lemma 4.3 of Hoeffding (1965), $\exists\, z^0 \in \overline{A(t)}$ such that $I_k(v, z^0) = I_k(v, A(t))$. Note that $z^0 \in S_{k-1}$ because $I_k(v, z^0) < \infty$ and $I_k(v, \cdot)$ is continuous. By the conditions on $J(z)$, for $\varepsilon > 0$, $\exists\, z' \in S_{k-1}$ such that $|I_k(v, z') - I_k(v, z^0)| < \varepsilon$ and $\delta = J(z') - J(z^0) > 0$. If $t \le s < \min(t + \delta, \beta)$, then $J(z') = J(z^0) + \delta \ge t + \delta > s$, so that $z' \in A(s)$ and $I_k(v, A(t)) \le I_k(v, A(s)) \le I_k(v, z') < I_k(v, z^0) + \varepsilon = I_k(v, A(t)) + \varepsilon$. This shows the right continuity of $I_k(v, A(t))$ in t. For the left continuity, let $s_n \uparrow t$. Again by Lemma 4.3 of Hoeffding (1965), for each n, $\exists\, z^n \in \overline{A(s_n)}$ such that $I_k(v, z^n) = I_k(v, A(s_n))$. Because $\{z^n\}$ is bounded, there is a convergent subsequence $\{z^{n'}\}$ of $\{z^n\}$. Let z^∞ be the limit of $\{z^{n'}\}$. Then $z^\infty \in S_{k-1}$ since $I_k(v, z^n)$ is bounded by $I_k(v, A(t))$. Thus, $J(z^\infty) = \lim_{n'\to\infty} J(z^{n'}) \ge \lim_{n'\to\infty} s_{n'} = t$. It follows that $z^\infty \in A(t)$ and

$$I_k(v, A(s_{n'})) = I_k(v, z^{n'}) \to I_k(v, z^\infty) \ge I_k(v, A(t)) . \tag{2.5}$$

Finally, since $s_n \uparrow t$ implies that $I_k(v, A(s_n))$ is increasing in n and $I_k(v, A(s_n)) \le I_k(v, A(t)) \;\forall\; n$, (2.5) shows that $I_k(v, A(s_n)) \to I_k(v, A(t))$ as $n \to \infty$, which gives the left continuity and completes the proof. □

PROOF OF THEOREM 2.1. Let $t \in (0, \beta)$ and $\varepsilon > 0$. By Lemma 2.2, $\exists\, \delta > 0$ such that $t + \delta < \beta$ and

$$I_k(v, A(t)) - \varepsilon < I_k(v, A(t - \delta)) \le I_k(v, A(t + \delta)) < I_k(v, A(t)) + \varepsilon . \tag{2.6}$$

Since $J_n \to J$ uniformly in S_{k-1}, $\exists\, N$ such that $\forall\, n > N$, $A(t + \delta) \subset A_n(t) \subset A(t - \delta)$, hence $I_k(v, A(t - \delta)) \le I_k(v, A_n(t)) \le I_k(v, A(t + \delta))$. This and (2.6) show that

$$I_k(v, A_n(t)) \to I_k(v, A(t)) . \tag{2.7}$$

Moreover, take $A_\varepsilon = \{z: J(z) > t + \delta\}$. Then A_ε is open and non-empty, and

$$I_k(v, A_\varepsilon) = I_k(v, A(t + \delta)) < I_k(v, A(t)) + \varepsilon . \tag{2.8}$$

Finally, the theorem follows from (2.7), (2.8) and Lemma 2.1. □

3. Bahadur efficiencies of spacings tests with fixed k

In this section, we consider the case when k and λ_i's are fixed and assume $n > \max_{1\le i\le k} (\lambda_i - \lambda_{i-1})^{-1}$ so that $D_i^{(n)} > 0$ with probability one (see Section 1 for notations). It is well-known that under H_0, $D^{(n)}$ has a Dirichlet distribution with parameters $[n\lambda_i] - [n\lambda_{i-1}]$, $i = 1,\dots,k$. If we write $v_i^{(n)} = ([n\lambda_i] - [n\lambda_{i-1}])/n$, $v^{(n)} = (v_1^{(n)},\dots,v_{k-1}^{(n)})$, then $D^{(n)} \sim \boldsymbol{D}(nv^{(n)})$ and $v^{(n)} \to v^0 = (v_1^0,\dots,v_{k-1}^0)$ where $v_i^0 = \lambda_i - \lambda_{i-1}$.

Since $D^{(n)} \to v^0$ (a.s.) under H_0, it is reasonable to reject H_0 when $D^{(n)}$ is too far from v^0. Thus we consider a family $\boldsymbol{F}$ of spacings tests which reject H_0 for large values of $J_n(D^{(n)})$, where $J_n(\cdot)$ is defined on S_{k-1} with properties:

(i) $J_n(z) \ge 0 \ \forall\, z \in S_{k-1}$ and $J_n(z) = 0$ iff $z = v^0$;

(ii) $J_n(z) \to J(z)$ uniformly in S_{k-1} for some J satisfying (A2) of Theorem 2.1.

In particular, we are interested in

$$J_n(z) = I_k(v^0, z) \quad \text{and} \quad J_n(z) = Q_k^2(z, v^0)\ \forall\, n$$

where

$$Q_k^2(z, v) = \sum_{i=1}^{k} (z_i - v_i)^2 / v_i \quad \text{for} \quad z,\ \ v \in S_{k-1}\,.$$

The test $Q_k^2(D^{(n)}, v^0)$ can be thought of as a spacings version of the classical chi-square test, and in fact, both $nQ_k^2(D^{(n)}, v^0)$ and $2nI_k(v^0, D^{(n)})$ have a limiting distribution of χ_{k-1}^2. (But we will not give the proofs here.) The assumption that J has no local maximum values in S_{k-1} is not unusual because, as a measure of the distance between z and v^0, $J(z)$ should increase when z moves farther away from v^0. It is easy to check that $I_k(v^0, z)$ and $Q_k^2(z, v^0)$ both have this property.

The exact slope of a test statistic T_n is defined as follows: let H_0 be rejected for large values of T_n. If $-2n^{-1}\log[1 - G_0(T_n)] \xrightarrow{P} s_T$ under H_1, where $G_0(t) = P_0(T_n \le t)$ and P_0 is the null probability measure, then s_T is the exact slope of T_n. The Bahadur efficiency of a test T_n relative to another T_n' is defined by $BE(T_n, T_n') = s_T / s_{T'}$ provided s_T and $s_{T'}$ are not both zeros. The following is a basic theorem for exact slopes:

THEOREM 3.1. (Bahadur (1960)) *If* $T_n \xrightarrow{P} b$ *under* H_1 *and*

$$\lim_{n\to\infty} -\frac{2}{n}\log P_0\{T_n \ge t\} = c(t), \qquad t \in I\,,$$

for some open interval I containing b and for some $c(t)$ which is continuous on I, then $s_T = c(b)$.

Now we are ready to give the exact slopes for tests in the family $\boldsymbol{F}$.

THEOREM 3.2. *Let the alternative be H_1: $F = F_1$ and $v^1 = (v_1^1, \ldots, v_{k-1}^1)$ with $v_i^1 = F_1^{-1}(\lambda_i) - F_1^{-1}(\lambda_{i-1})$. If $v^1 \in S_{k-1}$, $v^1 \neq v^0$ and $J(v^1) \neq \sup J(z)$, then the exact slope of $J_n(D^{(n)})$ for $J_n \in \boldsymbol{F}$*

$$s_J = 2I_k(v^0, A(J(v^1))) > 0, \quad (A(t) = \{z: J(z) \geq t\}) . \tag{3.1}$$

PROOF. First note that under H_1, $D^{(n)} \to v^1$ a.s., hence

$$J_n(D^{(n)}) \to J(v^1) , \tag{3.2}$$

(a.s.) under H_1. Because $D^{(n)} \sim \boldsymbol{D}(nv^{(n)})$ and $v^{(n)} \to v^0$ by Theorem 2.1

$$\lim_{n\to\infty} -\frac{2}{n} \log P_0\{J_n(D^{(n)}) \geq t\} = 2I_k(v^0, A(t)) , \tag{3.3}$$

for $t \in (0, \beta)$ where $\beta = \sup J(z)$. Moreover, by the conditions we have $J(v^1) \in (0, \beta)$ and by Lemma 2.2 $I_k(v^0, A(t))$ is continuous in $t \in (0, \beta)$. Thus (3.1) follows from (3.2), (3.3) and Theorem 3.1. □

THEOREM 3.3. *Let v^1 be as in Theorem 3.2 and s_I denote the exact slope of $I_k(v^0, D^{(n)})$, then $s_I \geq s_J$ for all $J_n \in \boldsymbol{F}$.*

PROOF. By Theorem 3.2, we have $s_J = 2I_k(v^0, A(J(v^1)))$, $(A(t) = \{J \geq t\})$, and in particular, $s_I = 2I_k(v^0, v^1)$ since $I_k(v^0, \{I_k(v^0, z) \geq t\}) = t$. Notice that $t \leq J(v^1)$ implies $v^1 \in A(t)$ and $I_k(v^0, v^1) \geq I_k(v^0, A(t))$. Hence $I_k(v^0, v^1) \geq I_k(v^0, A(J(v^1)))$. That is, $s_I \geq s_J$. □

Remark. Since $s_I > 0$, Theorem 3.3 shows that the Bahadur efficiency of $I_k(v^0, D^{(n)})$ relative to any $J_n \in \boldsymbol{F}$ is $BE(I, J) = s_I / s_J \geq 1$. Hence $I_k(v^0, D^{(n)})$ is the optimal test in the family $\boldsymbol{F}$ (in the Bahadur sense). The following theorem shows that the likelihood ratio test is optimal, in the Bahadur sense, for spacings tests, as for multinomial frequencies.

THEOREM 3.4. *For every $z \in S_{k-1}$,*

$$\lim_{n\to\infty} -\frac{1}{n}\left[\log p_n(z|v^0) - \log\left\{ \sup_v p_n(z|v) \right\}\right] = I_k(v^0, z) .$$

PROOF. Stirling's formula (2.1) and the equality $\Gamma(x+2) = x(x+1)\Gamma(x)$ imply

$$p_n(z|v) \leq Cn^{2k} e^{-nI_k(v,z)} \prod_{i=1}^{k} z_i^{-1} \; \forall \, n \quad \text{and} \quad v \in S_{k-1}, \tag{3.4}$$

where $C > 0$ is a constant, and similar to (2.4) we can obtain that $\exists \, N$ such that $\forall \, n > N$,

$$p_n(z|v^0) \geq e^{-nI_k(v^0,z)} \quad \text{and} \quad p_n(z|z) \geq e^{-nI_k(z,z)} = 1 . \tag{3.5}$$

(3.4) and (3.5) show that $\forall \, n > N$,

$$\frac{p_n(z|v^0)}{\sup_v p_n(z|v)} \leq \frac{p_n(z|v^0)}{p_n(z|z)} \leq Cn^{2k} e^{-nI_k(v^0,z)} \prod_{i=1}^{k} z_i^{-1} ,$$

and

$$\frac{p_n(z|v^0)}{\sup_v p_n(z|v)} \geq e^{-nI_k(v^0,z)} (Cn^{2k})^{-1} \prod_{i=1}^{k} z_i .$$

Hence

$$\lim_{n\to\infty} -\frac{1}{n} \log \frac{p_n(z|v^0)}{\sup_v p_n(z|v)} = I_k(v^0, z) . \qquad \square$$

THEOREM 3.5. *Let v^1 be as in Theorem 3.2. The exact slope of $Q_k^2(D^{(n)}, v^0)$ is given by*

$$s_Q = (1 - v_{\min}^0) \log \frac{1}{a} + v_{\min}^0 \log \frac{1}{b} ,$$

where $v_{\min}^0 = \min_{1\leq i\leq k} v_i^0$ and

$$a = 1 - \left[\frac{v_{\min}^0}{1 - v_{\min}^0} Q_k^2(v^1, v^0) \right]^{1/2}, \qquad b = 1 + \left[\frac{1 - v_{\min}^0}{- v_{\min}^0} Q_k^2(v^1, v^0) \right]^{1/2} .$$

The proof of Theorem 3.5 is similar to that of Theorem 8.1 of Hoeffding (1965). The details are omitted here. For the same reason, the following theorem is also stated without proof.

THEOREM 3.6. *Let $BE(I, Q^2)$ denote the Bahadur efficiency of*

$I_k(\nu^0, D^{(n)})$ relative to $Q_k^2(D^{(n)}, \nu^0)$. Then $BE(I, Q^2) > 1$ unless

$$\nu_i^1/\nu_i^0 = \begin{cases} b & \text{if} \quad \nu_i^0 = \nu_{\min}^0 , \\ a & \text{otherwise} , \end{cases} \tag{3.6}$$

where a, b are as in Theorem 3.5.

Remark. For $F_1 \neq F_0$, (3.6) generally does not hold and we may choose λ_i's to let (3.6) fail (for example, let ν_i^1/ν_i^0 take more than two different values). Hence Theorem 3.5 states that $I_k(\nu^0, D^{(n)})$ is basically more efficient than $Q_k^2(\nu^0, D^{(n)})$ (in the Bahadur sense).

4. The Bahadur efficiencies of spacings tests when $k \to \infty$

In this section, we allow k to increase with n to infinity, but take the partition λ_i's in a particular way as

$$\lambda_i = i/k \qquad i = 0, 1, \ldots, k ,$$

so that $\lambda_i - \lambda_{i-1} = 1/k$ for $i = 1, \ldots, k$. Moreover, we assume, without loss of generality, that $m = n/k$ takes integer values so that

$$\nu_i^{(n)} = ([n\lambda_i] - [n\lambda_{i-1}])/n = \lambda_i - \lambda_{i-1} = 1/k, \qquad i = 1, \ldots, k .$$

We are interested in the Bahadur efficiencies of the tests $I_k(\nu^{(n)}, D^{(n)})$ and $Q_k^2(D^{(n)}, \nu^{(n)})$. Let H_1: $F = F_1$ satisfy the following assumptions:

ASSUMPTIONS 4.1.
(i) $G_1 = F_1^{-1}$ and $g_1 = G_1'$ exist on $[0, 1]$;
(ii) g_1 is continuous on $[0, 1]$;
(iii) $0 < g_1(y) < \infty \ \forall \ y \in [0, 1]$.

Let k vary with n, in such a way as

$$k = k(n) = cn^q(1 + o(1)), \qquad c > 0, \quad 0 < q < 1 . \tag{4.1}$$

THEOREM 4.1. *If H_1 satisfies Assumptions 4.1 and $k = k(n)$ is as given in (4.1), then the exact slopes of $I_k(\nu^{(n)}, D^{(n)})$ and $Q_k^2(D^{(n)}, \nu^{(n)})$ are $s_I = \int_0^1 -2 \log g_1(y) dy$ and $s_Q = 0$, respectively. Hence, $BE(I, Q^2) = \infty$.*

We will prove two lemmas first.

LEMMA 4.1. *Under the conditions of Theorem 4.1,*

$$\lim_{n\to\infty} -\frac{1}{n} P_0\{I_k(\nu^{(n)}, D^{(n)}) \geq t\} = t . \tag{4.2}$$

PROOF. Let $p_n(z|\nu)$ and $P_n(A|\nu)$ be as in Section 2. By (2.2),

$$\begin{aligned} p_n(z|\nu^{(n)}) &\leq 2\left(\frac{n}{2\pi}\right)^{(k-1)/2} \prod_{i=1}^{k} \left(\frac{1}{k}\right)^{1/2-n/k} z_i^{n/k-1} \\ &\leq 2n^k k^k e^{-(n-k)I_k(\nu^{(n)}, z)} . \end{aligned}$$

Hence for $A_n = \{z\colon I_k(\nu^{(n)}, z) \geq t\}$, since $I_k(\nu^{(n)}, A_n) = t$,

$$\begin{aligned} &\varliminf_{n\to\infty} -\frac{1}{n} \log P_n(A_n|\nu^{(n)}) \\ &\geq \varliminf_{n\to\infty} \left[\left(1-\frac{k}{n}\right) I_k(\nu^{(n)}, A_n) - \frac{k}{n}\log(nk) - \frac{1}{n}\log 2\right] = t . \end{aligned} \tag{4.3}$$

On the other hand, Jensen's inequality gives $I_k(\nu^{(n)}, z) \geq I_2(1/k, z_1)$, where z_1 is the first coordinate of z. It is not hard to see that $I_2(1/k, z_1)$ is increasing in $z_1 \in (1/k, 1)$ and $\exists\, x_k \in (1/k, 1)$ such that $I_2(1/k, x_k) = t$. Thus by (2.1),

$$\begin{aligned} P_0\{I_k(\nu^{(n)}, D^{(n)}) \geq t\} &\geq P_0\{I_2(\nu^{(n)}, D_1^{(n)}) \geq t\} = P_0\{D_1^{(n)} \geq x_k\} \\ &= \frac{\Gamma(n)}{\Gamma(m)\Gamma(n-m)} \int_{x_k}^{1} z_1^{m-1}(1-z_1)^{n-m-1} dz_1 \\ &\geq \frac{1}{4\sqrt{2\pi}} \frac{1}{n} \left[m\left(1-\frac{1}{k}\right)\right]^{1/2} e^{-nt} . \end{aligned}$$

Hence

$$\varlimsup_{n\to\infty} -\frac{1}{n} \log P_0(I_k(\nu^{(n)}, D^{(n)}) \geq t) \leq t . \tag{4.4}$$

Since $D^{(n)} \sim \boldsymbol{D}(n\nu^{(n)})$ under H_0, (4.3) and (4.4) prove the lemma. □

LEMMA 4.2. *Suppose*

(i) H_1 *satisfies Assumptions* 4.1 *with corresponding* G_1 *and* g_1;

(ii) h *is a continuous function in* $(0, \infty)$;

(iii) $n/m^j \to 0$ *for some* $j \geq 2$.

Then under H_1,

$$\frac{1}{k} \sum_{i=1}^{k} h(kD_i^{(n)}) \xrightarrow{P} \int_0^1 h(g_1(y))dy .$$

PROOF. Let $U_1 < U_2 < \cdots < U_{n-1}$ be an ordered sample from a uniform distribution on [0, 1] and $T_i^{(n)} = U_{im} - U_{im-m}$. Then by the Mean Value Theorem, $kD_i^{(n)} \sim g_1(\tilde{U}_{im})\, kT_i^{(n)}$ where $U_{im-m} < \tilde{U}_{im} < U_{im}$. It can be shown that under the conditions of the lemma

$$\max_{1\le i\le k} |kT_i^{(n)} - 1| \xrightarrow{P} 0 \quad \text{and} \quad \max_{1\le i\le k} \left| \tilde{U}_{im} - \frac{i}{k} \right| \xrightarrow{P} 0 ,$$

which imply

$$\begin{aligned}\max_{1\le i\le k} \left| kD_i^{(n)} - g_1\left(\frac{i}{k}\right)\right| &\sim \max_{1\le i\le k} \left| g_1(\tilde{U}_{im})kT_i^{(n)} - g_1\left(\frac{i}{k}\right)\right| \\ &\le \max_{1\le i\le k} \left| g_1(\tilde{U}_{im}) - g_1\left(\frac{i}{k}\right)\right| (kT_i^{(n)}) \\ &\quad + \max_{1\le i\le k} \left| g_1\left(\frac{i}{k}\right)\right| \left| kT_i^{(n)} - 1 \right| \xrightarrow{P} 0 .\end{aligned}$$

It follows that under H_1,

$$\left| \frac{1}{k} \sum_{i=1}^{k} h(kD_i^{(n)}) - \frac{1}{k} \sum_{i=1}^{k} h\left(g_1\left(\frac{i}{k}\right)\right)\right| \xrightarrow{P} 0 . \tag{4.5}$$

But since $(1/k) \sum_{i=1}^{k} h(g_1(i/k)) \to \int_0^1 h(g_1(y))dy$, (4.5) proves Lemma 4.2. □

PROOF OF THEOREM 4.1. Take $h(x) = -\log x$. Then Lemma 4.2 yields that under H_1,

$$I_k(v^{(n)}, D^{(n)}) = \frac{1}{k} \sum_{i=1}^{k} - \log (kD_i^{(n)}) \xrightarrow{P} \int_0^1 - \log g_1(y)dy > 0 ,$$

and from Lemma 4.1 we get $s_I = \int_0^1 - 2 \log g_1(y)dy$. On the other hand,

$$\begin{aligned}P_0\{Q_k^2(D^{(n)}, v^{(n)}) \ge t\} &= P_0\left\{ \sum_{i=1}^{k} (kD_i^{(n)} - 1)^2 \ge kt \right\} \\ &\ge P_0\left\{ D_1^{(n)} \ge \left(\frac{1+t}{k}\right)^{1/2} \right\} \\ &= \frac{\Gamma(n)}{\Gamma(m)\Gamma(n-m)} \int_{\sqrt{(t+1)/k}}^{1} x^{m-1}(1-x)^{n-m-1}dx\end{aligned}$$

$$\geq \left(\frac{1+t}{k}\right)^{m/2} \frac{1}{n}\left[1-\left(\frac{1+t}{k}\right)^{1/2}\right]^n .$$

Thus $\forall\, t \geq 0$,

$$(4.6)\quad \begin{aligned} 0 &\leq -\frac{1}{n}\log P_0\{Q_k^2(D^{(n)}, \nu^{(n)}) \geq t\} \\ &\leq \frac{1}{2k}\log\frac{k}{1+t} + \frac{1}{n}\log n - \log\left[1-\left(\frac{1+t}{k}\right)^{1/2}\right] \to 0 \end{aligned}$$

as $n \to \infty$ and $k \to \infty$. Moreover, by Lemma 4.2 we see that under H_1,

$$Q_k^2(D^{(n)}, \nu^{(n)}) = \frac{1}{k}\sum_{i=1}^{k}(kD_i^{(n)} - 1)^2 \xrightarrow{P} \int_0^1 [g_1(y) - 1]^2 dy > 0 .$$

Hence (4.6) and Theorem 3.1 show that $s_Q = 0$. □

5. Conclusion

The exact Bahadur efficiency of the test statistic based on spacings $I_k = \sum_{i=1}^{k} \nu_i^0 \log(\nu_i^0 / D_i^{(n)})$ relative to its competitor $Q_k^2 = \sum_{i=1}^{k}(D_i^{(n)} - \nu_i^0)^2/\nu_i^0$ is shown to be greater than 1 for finite k (cf. Theorem 3.5) and equals infinity if k is allowed to increase with n (cf. Theorem 4.1). This contrasts with the results that for fixed m, the test Q_k^2, sometimes called the Greenwood statistic, has the highest asymptotic relative efficiency in the Pitman sense. See, for instance, Rao and Kuo (1984).

References

Bahadur, R. R. (1960). Stochastic comparison of tests, *Ann. Math. Statist.*, **31**, 276–295.

Del Pino, G. E. (1979). On the asymptotic distribution of k-spacings with applications to goodness of fit tests, *Ann. Statist.*, 7, 1058–1065.

Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions, *Ann. Math. Statist.*, **36**, 369–408.

Jammalamadaka, S. R. and Tiwari, R. C. (1987). Efficiencies of some disjoint spacings tests relative to a chi-square test, *New Perspectives in Theoretical and Applied Statistics*, (eds. M. L. Puri, J. Wilaplana and W. Wertz), 311–318, Wiley, New York.

Jammalamadaka, S. R., Zhou, X. and Tiwari, R. C. (1986). Asymptotic efficiencies of spacings tests for goodness of fit, Tech. Report No. 4, Statistics Program, UCSB.

Kuo, M. and Rao, J. S. (1981). Limit theory and efficiencies for tests based on higher order spacings, *Statistics—Applications and New Directions*, Proceedings of the Golden Jubilee Conference of the Indian Statistical Institute, 333–352, Statistical Publishing Society, Calcutta.

Quine, M. P. and Robinson, J. (1985). Efficiencies of chi-square and likelihood ratio goodness-of-fit tests, *Ann. Statist.*, **13**, 795–802.

Rao, J. S. and Kuo, M. (1984). Asymptotic results on the Greenwood statistic and some of its generalizations, *J. Roy. Statist. Soc. Ser. B*, **46**, 228–237.

Sethuraman, J. and Rao, J. S. (1970). Pitman efficiencies of tests based on spacings, *Nonparametric Techniques in Statistical Inference*, 405–416, Cambridge Univ. Press, Cambridge.