APPLICATIONES MATHEMATICAE
25,4 (1999), pp. 359–373

F. CZEKAŁA (Wrocław)

ON TWO TESTS BASED ON DISJOINT m-SPACINGS

Abstract. This paper is concerned with the properties of two statistics based on the logarithms of disjoint m-spacings. The asymptotic normality is established in an elementary way and exact and asymptotic means and variances are computed in the case of uniform distribution on the interval $[0,1]$. This result is generalized to the case when the sample is drawn from a distribution with positive step density on $[0,1]$. Bahadur approximate efficiency of tests based on those statistics is found for such alternatives.

1. Introduction. Let $X_1,\ldots,X_n$ be a sample from the uniform distribution on the interval $[0,1]$. Denote by $X_{(1)},\ldots,X_{(n)}$ the order statistics derived from this sample and define $X_{(0)}=0$, $X_{(n+1)}=1$. Let $m\ge 1$ be fixed. We define the *m-spacings* from the sample $X_1,\ldots,X_n$ as the differences

$$D_{i,n}^{(m)}=X_{(i+m)}-X_{(i)},\quad 0\le i\le n+1-m. \tag{1}$$

Denote by k_n the number of disjoint m-spacings that can be built from a sample of size n:

$$k_n=\left[\frac{n+1}{m}\right], \tag{2}$$

where $[x]$ denotes the integral part of x. Now let $Y_1,\ldots,Y_{n+1}$ be i.i.d. exponential random variables with unit mean. It is known that

$$X_{(i)}\overset{d}{=}\frac{Y_1+\ldots+Y_i}{Y_1+\ldots+Y_{n+1}}, \tag{3}$$

1991 *Mathematics Subject Classification*: 62E15, 62G10, 62G20, 62G30.

Key words and phrases: disjoint m-spacings; limit distributions; step densities; Bahadur approximate efficiency.

Research partially supported by KBN (Committee for Scientific Research, Poland) Grant 2 1033 91 01.

and the equality also holds for the whole vector of order statistics. Set also

$$S_{i,m} = Y_{(i-1)m+1} + \ldots + Y_{im}, \qquad i = 1, \ldots, k_n. \tag{4}$$

We are interested in two statistics based on disjoint, normalized m-spacings:

$$G_{i,n} = \sum_{j=0}^{k_n-1} g_i((n+1)D^{(m)}_{jm,n}), \qquad i = 1, 2, \tag{5}$$

where $g_1(x) = \log x$ and $g_2(x) = x \log x$. The statistic $G_{1,n}$ for $m = 1$ was suggested by Darling (1953). The statistic $G_{2,n}$ (also for $m = 1$) was first considered by Gebert & Kale (1969) and both statistics were derived in a general way by Kale (1969).

2. The exact moments of $G_{1,n}$ and $G_{2,n}$. In this section we compute the exact means and variances of the statistics $G_{1,n}$ and $G_{2,n}$.

LEMMA 1. *Let $Y_1, \ldots, Y_n$ be i.i.d. exponential random variables with unit mean. Let S_k denote the sum of the first k of these variables. Then*

$$\mathrm{E}\{\log S_n\} = \sum_{i=1}^{n-1} \frac{1}{i} - \gamma, \tag{6}$$

$$\mathrm{E}\{S_n \log S_n\} = n\biggl(\sum_{i=1}^{n} \frac{1}{i} - \gamma\biggr), \tag{7}$$

$$\mathrm{Cov}\,\{S_n, \log S_n\} = 1, \tag{8}$$

$$\sigma^2\{\log S_n\} = \sum_{i=n}^{\infty} \frac{1}{i^2}, \tag{9}$$

$$\mathrm{Cov}\,\{\log S_k, \log S_n\} = \sigma^2\{\log S_n\}, \qquad 1 \le k \le n, \tag{10}$$

where γ is Euler's constant.

Proof. The random variable S_n has Erlang's distribution with density

$$g_n(x) = \frac{1}{(n-1)!} x^{n-1} e^{-x}.$$

Hence

$$\mathrm{E}\{\log S_n\} = \frac{1}{(n-1)!} \int_0^{\infty} x^{n-1} e^{-x} \log x \, dx.$$

Denote the integral above by $I(n-1)$. Integrating by parts we obtain $I(n-1) = -n^{-1} + I(n)$. Therefore

$$I(n) = 1 + \frac{1}{2} + \ldots + \frac{1}{n} + I(0). \tag{11}$$

From the theory of the gamma function we know that

$$I(0) = \int_0^\infty e^{-x} \log x \, dx = -\gamma,$$

which together with (11) proves (6). Now (7) results from the fact that

$$\mathrm{E}\{S_n \log S_n\} = \frac{1}{(n-1)!} \int_0^\infty x^{n-1} e^{-x} x \log x \, dx = nI(n).$$

Next we have $\mathrm{E}\{S_n\} = n$ and thus (8) is a consequence of (6) and (7).

To prove (9) we need to calculate

$$\mathrm{E}\{\log^2 S_n\} = \frac{1}{(n-1)!} \int_0^\infty x^{n-1} e^{-x} \log^2 x \, dx.$$

Denote the above integral by $K(n-1)$. Integrating by parts we find that $K(n-1) = -2n^{-1} I(n-1) + K(n)$, which leads to

$$K(n) = K(0) + \left(\sum_{i=1}^n \frac{1}{i} \right)^2 - \sum_{i=1}^n \frac{1}{i^2} - 2\gamma \sum_{i=1}^n \frac{1}{i}.$$

From the theory of the gamma function we know that

$$K(0) = \int_0^\infty e^{-x} \log^2 x \, dx = \frac{\pi^2}{6} + \gamma^2,$$

and thus

$$K(n) = \frac{\pi^2}{6} - \sum_{i=1}^n \frac{1}{i^2} + \left(\sum_{i=1}^n \frac{1}{i} - \gamma \right)^2 = \sum_{i=n+1}^\infty \frac{1}{i^2} + I^2(n).$$

Hence

$$\sigma^2\{\log S_n\} = K(n-1) - I^2(n-1) = \sum_{i=n}^\infty \frac{1}{i^2}.$$

Finally, to prove (10) we need to calculate $\mathrm{E}\{\log S_k \log S_n\}$. We have

$$\mathrm{E}\{\log S_k \log S_n \mid S_n = s\} = \log s \mathrm{E}\{\log S_k \mid S_n = s\}.$$

It is known that $Y_1, \ldots, Y_n$ given $S_n = s$ have the same distribution as spacings from a sample of size $n-1$ from the uniform distribution on the interval $[0, s]$ and correspondingly S_k given $S_n = s$ is distributed as the kth order statistic from this distribution. Now it follows from (3) that we can write

$$\begin{aligned} \mathrm{E}\{\log S_k \mid S_n = s\} &= \mathrm{E}\left\{ \log \left(s \frac{S_k}{S_n} \right) \right\} \\ &= \log s + \mathrm{E}\{\log S_k\} - \mathrm{E}\{\log S_n\} \end{aligned}$$

and we obtain

$$\begin{aligned}\mathrm{E}\{\log S_k \log S_n\} &= \mathrm{E}\{\mathrm{E}\{\log S_k \log S_n \mid S_n\}\} \\ &= \mathrm{E}\{\log^2 S_n + \log S_n(\mathrm{E}\{\log S_k\} - \mathrm{E}\{\log S_n\})\} \\ &= \sigma^2\{\log^2 S_n\} + \mathrm{E}\{\log S_k\}\mathrm{E}\{\log S_n\}.\end{aligned}$$

From the above equality (10) results immediately. This ends the proof of Lemma 1.

Lemma 2. *Let $D_{i,n}^{(m)}$ be the m-spacings defined by* (1). *Then*

$$\mathrm{E}\{\log D_{0,n}^{(m)}\} = -\sum_{i=m}^{n} \frac{1}{i}, \tag{12}$$

$$\sigma^2\{\log D_{0,n}^{(m)}\} = \sum_{i=m}^{n} \frac{1}{i^2}, \tag{13}$$

$$\mathrm{Cov}\{\log D_{0,n}^{(m)}, \log D_{m,n}^{(m)}\} = -\sum_{i=n+1}^{\infty} \frac{1}{i^2}. \tag{14}$$

Proof. According to (3) and (6) we have

$$\mathrm{E}\{\log D_{0,n}^{(m)}\} = \mathrm{E}\{\log S_{1,m}\} - \mathrm{E}\{\log S_{n+1}\} = -\sum_{i=m}^{n} \frac{1}{i}.$$

In the same way, from (3), (10) and (9) we get

$$\begin{aligned}\sigma^2\{\log D_{0,n}^{(m)}\} &= \sigma^2\{\log S_{1,m}\} + \sigma^2\{\log S_{n+1}\} - 2\,\mathrm{Cov}\,\{\log S_{1,m}, \log S_{n+1}\} \\ &= \sigma^2\{\log S_{1,m}\} - \sigma^2\{\log S_{n+1}\} = \sum_{i=m}^{n} \frac{1}{i^2}.\end{aligned}$$

Finally,

$$\begin{aligned}&\mathrm{Cov}\,\{\log D_{0,n}^{(m)}, \log D_{m,n}^{(m)}\} \\ &\quad = \mathrm{Cov}\,\{\log S_{1,m}, \log S_{2,m}\} - \mathrm{Cov}\,\{\log S_{1,m}, \log S_{n+1}\} \\ &\quad\quad - \mathrm{Cov}\,\{\log S_{n+1}, \log S_{2,m}\} + \sigma^2\{\log S_{n+1}\}\end{aligned}$$

and since $S_{1,m}$ and $S_{2,m}$ are independent, $\mathrm{Cov}\,\{\log S_{1,m}, \log S_{2,m}\} = 0$ and (14) follows from (9) and (10).

Lemma 3. *The exact mean and variance of the statistic $G_{1,n}$ are given by*

$$\mathrm{E}\{G_{1,n}\} = k_n\Big(\log(n+1) - \sum_{i=m}^{n} \frac{1}{i}\Big), \tag{15}$$

$$\sigma^2\{G_{1,n}\} = k_n\Big(\sum_{i=m}^{\infty}\frac{1}{i^2} - k_n\sum_{i=n+1}^{\infty}\frac{1}{i^2}\Big). \tag{16}$$

P r o o f. Formula (15) results from the definition of $G_{1,n}$ and Lemma 2. To find the variance of $G_{1,n}$ note that

$$\begin{aligned}\sigma^2\{G_{1,n}\} &= k_n\sigma^2\{\log D_{0,n}^{(m)}\} + k_n(k_n-1)\,\mathrm{Cov}\,\{\log D_{0,n}^{(m)}, \log D_{m,n}^{(m)}\}\\ &= k_n\sigma^2\{\log S_{1,m}\} - k_n^2\sigma^2\{\log S_{n+1}\},\end{aligned}$$

and (13) applied to the last equality gives us (16).

Lemma 4. *Let $l_n = k_n m$. Then*

$$\mathrm{E}\{G_{2,n}\} = l_n\Big(\log(n+1) - \sum_{i=m+1}^{n+1}\frac{1}{i}\Big) \tag{17}$$

$$\begin{aligned}\sigma^2\{G_{2,n}\} = \frac{l_n}{n+2}\bigg\{&(n+1)\Big((m+1)\sum_{i=m+1}^{\infty}\frac{1}{i^2} - (l_n+1)\sum_{i=n+2}^{\infty}\frac{1}{i^2}\Big)\\ &+ (n+1-l_n)\Big(\Big(\sum_{i=m+1}^{n+2}\frac{1}{i}\Big)^2 - 2\sum_{i=m+1}^{n+2}\frac{1}{i} + \frac{1}{(n+2)^2}\Big)\bigg\}.\end{aligned} \tag{18}$$

In the case when $l_n = n+1$, i.e. when the disjoint m-spacings span the whole interval $[0,1]$, (18) *simplifies to*

$$\sigma^2\{G_{2,n}\} = \frac{(n+1)^2}{n+2}\Big((m+1)\sum_{i=m+1}^{\infty}\frac{1}{i^2} - (n+2)\sum_{i=n+2}^{\infty}\frac{1}{i^2}\Big). \tag{19}$$

P r o o f. Formulae (17) and (18) written in a slightly different form were proved in Czekała (1996) as Lemma 5.

3. The asymptotic normality of $G_{1,n}$ and $G_{2,n}$. The asymptotic normality of a wide class of statistics based on disjoint m-spacings was proved by Del Pino (1979) via weak convergence in the space $D[0,\infty]$ of the empirical process of the m-spacings to a Gaussian process. However, the asymptotic normality of statistics based on disjoint m-spacings can be proved quite elementarily using a modified version of a theorem of Proschan & Pyke (1964). This method was also used by Cressie (1976) to prove the asymptotic normality of an equivalent to the statistic $G_{1,n}$ based on overlapping m-spacings. We present the theorem of Proschan & Pyke in a form suitable for dealing with disjoint m-spacings.

Lemma 5. *Let $X_1,\ldots,X_{n+1}$ be i.i.d. random variables and let k_n be defined by* (2). *Assume that* $\mathrm{E}\{|X_1|\} < \infty$ *and set* $\mathrm{E}\{X_1\} = \mu$. *Let $h(x,y)$ be any real function of two variables satisfying the following conditions*:

(i) $h(x,y)$ *is continuously differentiable in* y. *For convenience we denote this derivative by* $h'(x,y)$.

(ii) $\mathrm{E}\{|h'(S_{1,m},\mu)|\} < \infty$, *where* $S_{i,m}$ *are sums of* X_i*'s defined as in* (4). *Set* $\mathrm{E}\{h'(S_{1,m},\mu)\} = C$.

(iii) *For all double sequences* $\{\theta_{ni} : 1 \le i \le k_n,\ n \ge 1\}$ *of random variables for which*

$$\max_{1\le i\le k_n} |\theta_{ni} - \mu| \xrightarrow{P.1} 0 \tag{20}$$

we have

$$\frac{1}{k_n}\sum_{i=1}^{k_n} (h'(S_{i,m},\theta_{ni}) - h'(S_{i,m},\mu)) \xrightarrow{P.1} 0. \tag{21}$$

Assume also that X_1 *and* $h(S_{1,m},\mu)$ *have finite variances* σ_x^2 , σ_h^2 *and covariance* σ_{xh}. *If we denote by* $\overline{X}_n$ *the sample mean of* X_i*'s then the random variable*

$$Z_n = \frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n} (h(S_{i,m},\overline{X}_n) - \mathrm{E}\{h(S_{1,m},\mu)\}) \tag{22}$$

converges in distribution to a normal random variable with zero mean and finite variance σ_z^2 *given by*

$$\sigma_z^2 = \sigma_h^2 + (C^2/m)\sigma_x^2 + (2C/m)\sigma_{xh}. \tag{23}$$

Proof. We have Taylor's expansion

$$h(S_{i,m},\overline{X}_n) = h(S_{i,m},\mu) + (\overline{X}_n - \mu)h'(S_{i,m},\theta_{ni}).$$

Applying this to (22) we can show as in the proof of Theorem 1 of Proschan & Pyke (1964) that

$$Z_n = \frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}\left(h(S_{i,m},\mu) - \mathrm{E}\{h(S_{1,m},\mu)\} + \frac{C}{m}(S_{i,m} - m\mu)\right) + \varepsilon_n,$$

where $\varepsilon_n \to 0$ in probability. Since $S_{i,m}$ are independent random variables the result follows from the ordinary central limit theorem.

The next lemma was stated and proved in Proschan & Pyke (1964). We only recall it here so that it can be used in verifying condition (iii) of Lemma 5 for our statistics $G_{1,n}$ and $G_{2,n}$.

Lemma 6. *Let* $k(x,y)$ *be any real function of two variables. If* k *satisfies*

$$\lim_{\varepsilon\to 0} \mathrm{E}\{\sup_{|\theta-\mu|<\varepsilon} |k(S_{1,m},\theta) - k(S_{1,m},\mu)|\} = 0, \tag{24}$$

then k *also satisfies condition* (iii) *of Lemma* 5.

THEOREM 1. *The random variable $G_{1,n}$ is asymptotically normal and its asymptotic mean $e_{1,n}$ and variance $\sigma_{1,n}^2$ are given by*

$$e_{1,n} = \frac{n+1}{m}\Big(\sum_{i=1}^{m-1}\frac{1}{i} - \gamma\Big) \tag{25}$$

and

$$\sigma_{1,n}^2 = \frac{n+1}{m}\Big(\sum_{i=m}^{\infty}\frac{1}{i^2} - \frac{1}{m}\Big). \tag{26}$$

Proof. According to (3) we have

$$G_{1,n} \stackrel{d}{=} \sum_{i=1}^{k_n} \log\left(\frac{Y_{(i-1)m+1} + \ldots + Y_{im}}{\overline{Y}_{n+1}}\right).$$

To prove the asymptotic normality of $G_{1,n}$ we can now use Lemma 5. If we take $h(x,y) = \log(x/y)$ then $h'(x,y) = -1/y$ and we only need to check condition (iii). To that end we will use Lemma 6. Because $\mu = 1$ we have

$$|h'(S_{1,m},\theta) - h'(S_{1,m},\mu)| = |1 - 1/\theta|,$$

and it is clear that condition (24) is satisfied. To find the asymptotic variance of $G_{1,N}$ we have to calculate σ_z^2 as defined in Lemma 5. We have

$$\begin{aligned}\sigma_z^2 &= \sigma_h^2 + \frac{C^2}{m}\sigma_x^2 + \frac{2C}{m}\sigma_{xh}\\ &= \sigma^2\{\log S_{1,m}\} + \frac{1}{m^2}\sigma^2\{S_{1,m}\} - \frac{2}{m}\operatorname{Cov}\{S_{1,m}\log S_{1,m}\},\\ &= \sum_{i=m}^{\infty}\frac{1}{i^2} + \frac{1}{m^2}m - \frac{2}{m}.\end{aligned}$$

Since $(n+1)m^{-1}k_n^{-1} \to 1$ it follows that the asymptotic variance of $G_{1,n}$ is equivalent to $\sigma_{1,n}^2$ given by (26). We obtain formula (25) in a similar way computing $\mathrm{E}\{h(S_{1,m},\mu)\}$ by means of Lemma 1.

THEOREM 2. *The random variable $G_{2,n}$ is asymptotically normal and its asymptotic mean $e_{2,n}$ and variance $\sigma_{2,n}^2$ are given by*

$$e_{2,n} = (n+1)\Big(\sum_{i=1}^{m}\frac{1}{i} - \gamma\Big) \tag{27}$$

and

$$\sigma_{2,n}^2 = (n+1)\Big((m+1)\sum_{i=m+1}^{\infty}\frac{1}{i^2} - 1\Big). \tag{28}$$

Proof. We prove this theorem in the same way as Theorem 1. Let $h(x,y) = (x/y)\log(x/y)$. We only need to check condition (24) of Lemma 6 to prove the asymptotic normality of $G_{2,n}$. We have

$$h'(x,y) = -(x/y^2)\log(x/y) - x/y^2$$

and it follows that

$$|h'(S_{1,m},\theta) - h'(S_{1,m},\mu)| \le (|S_{1,m}\log S_{1,m}| + |S_{1,m}|)\left|\frac{1}{\theta^2} - 1\right| + |S_{1,m}|\left|\frac{\log\theta}{\theta^2}\right|.$$

The functions of θ above are continuous and equal to 0 at $\theta = 1$ and using this it is easy to see that condition (24) holds. To prove (27) note that according to (22) the asymptotic mean of $G_{2,n}$ is equal to $k_n \mathrm{E}\{h(S_{1,m},\mu)\}$, which is asymptotically equivalent to $(n+1)m^{-1}\mathrm{E}\{h(S_{1,m}\mu)\}$. Thus we have

$$e_{2,n} = \frac{n+1}{m}\mathrm{E}\{S_{1,m}\log S_{1,m}\} = \frac{n+1}{m}I(m),$$

which proves (27). To prove (28) we need to calculate σ_z^2 given by (23). To that end we first have to find σ_h^2, σ_{xh} and C as defined in Lemma 5. We have

$$\begin{aligned} \sigma_h^2 &= \mathrm{E}\{S_{1,m}^2\log^2 S_{1,m}\} - \mathrm{E}^2\{S_{1,m}\log S_{1,m}\} \\ &= m(m+1)K(m+1) - (mI(m))^2, \end{aligned} \tag{29}$$

where $K(m+1)$ and $I(m)$ are as in the proof of Lemma 1. We get similarly

$$\begin{aligned} \sigma_{xh} &= \mathrm{E}\{S_{1,m}\log S_{1,m}\} - \mathrm{E}\{S_{1,m}\}\mathrm{E}\{S_{1,m}\log S_{1,m}\} \\ &= m(m+1)I(m+1) - m^2 I(m), \end{aligned} \tag{30}$$

and

$$C = \mathrm{E}\{h'(S_{1,m},\mu)\} = \mathrm{E}\{-S_{1,m}\log S_{1,m} - S_{1,m}\} = -m(I(m)+1). \tag{31}$$

Putting (29)–(31) together we obtain the following expression for σ_z^2:

$$\sigma_z^2 = m(m+1)\sum_{i=m+1}^{\infty}\frac{1}{i^2} - m,$$

which proves (28) as the asymptotic variance of $G_{2,n}$ is equal to $k_n\sigma_z^2$ which is asymptotically equivalent to $(n+1)m^{-1}\sigma_z^2$.

The asymptotic normalizing constants for $G_{2,n}$ were also found by Jammalamadaka & Tiwari (1986) using the results of Del Pino (1979).

4. The asymptotic normality of $G_{1,n}$ and $G_{2,n}$ for some alternatives. In this section we find the asymptotic distribution of $G_{1,n}$ and $G_{2,n}$ for fixed alternatives with positive step densities on $[0,1]$. Let $\underline{X} = (X_1, X_2, \ldots, X_n, \ldots)$ be a sequence of random variables which are used to form $G_{i,n}$. We can write $G_{i,n} = \phi_i(n,\underline{X})$. To simplify the notation we assume that $\phi_i(0, \underline{X}) = 0$. Now let $k \geq 1$ be a fixed integer and fix $0 = x_0 < x_1 < \ldots < x_k = 1$. Define

$$I_1 = [x_0, x_1), \ \ldots, \ I_{k-1} = [x_{k-2}, x_{k-1}), \quad I_k = [x_{k-1}, x_k].$$

The length of the interval I_i will be denoted by d_i and its indicator function by $\mathbf{1}_{I_i}(x)$. Let $f_i > 0$, $i = 1, \ldots, k$, be fixed numbers such that $\sum_{i=1}^{k} f_i d_i = 1$. These numbers together with the intervals I_i define a step density f:

$$f(x) = \sum_{i=1}^{k} f_i \, \mathbf{1}_{I_i}(x). \tag{32}$$

Define $p_i = f_i d_i$. We have $\sum p_i = 1$ and hence there exist numbers $0 = x'_0 < x'_1 < \ldots < x'_k = 1$ such that the intervals I'_i, defined similarly to I_i, have lengths p_i. There also exists a vector $(\underline{U}, \underline{Y}^1, \ldots, \underline{Y}^k)$ of random elements such that:

(a) the coordinates of this vector are stochastically independent,

(b) $\underline{U} = (U_1, U_2, \ldots)$ is a sequence of independent random variables uniformly distributed on $[0, 1]$,

(c) $\underline{Y}^i = (Y_1^i, Y_2^i, \ldots)$, for $i = 1, \ldots, k$, are sequences of independent random variables with uniform distribution on I_i.

We can now define a sequence $\underline{Z} = (Z_1, Z_2, \ldots)$ of independent random variables with density f:

$$Z_n = \sum_{i=1}^{k} \mathbf{1}_{I'_i}(U_n) Y_n^i, \qquad n \geq 1. \tag{33}$$

Denote by $N_{i,n}$ the number of random variables $Z_1, \ldots, Z_n$ taking values in the interval I_i, i.e.

$$N_{i,n} = \sum_{j=1}^{n} \mathbf{1}_{I_i}(Z_j). \tag{34}$$

It is easy to see that $N_{i,n} = \sum_{j=1}^{n} \mathbf{1}_{I'_i}(U_j)$ and thus the sequence of vectors $(N_{1,n}, \ldots, N_{k,n})$ is independent of $(\underline{Y}^1, \ldots, \underline{Y}^k)$. In the sequel we will write N_i instead of $N_{i,n}$ for brevity.

The asymptotic normality of $G_{2,n}$ in the case of fixed alternative distributions with step densities was proved in Czekała (1996). This result is presented in Theorem 3. The asymptotic distribution of $G_{2,n}$ under a sequence of alternatives converging to the uniform distribution was found by Jammalamadaka & Tiwari (1986).

THEOREM 3. *If $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. random variables with density $f > 0$ given by* (32) *then*

$$\frac{G_{2,n} - e_{2,n}(f)}{\sigma_{2,n}(f)} \xrightarrow{d} N(0,1), \tag{35}$$

where

$$e_{2,n}(f) = (n+1)\Big(\sum_{i=1}^{m} \frac{1}{i} - \gamma + \mathrm{E}\{f(X_1)^{-1} \log (f(X_1)^{-1})\}\Big), \tag{36}$$

$$\sigma^2_{2,n}(f) = (n+1)\Big(\mathrm{E}\{f(X_1)^{-2}\}(m+1) \sum_{i=m+1}^{\infty} \frac{1}{i^2} - 1\Big). \tag{37}$$

As the proof of asymptotic normality for the statistic $G_{1,n}$ is very similar we will omit most details and state only basic lemmas that are required. Similar lemmas with complete proofs can be found in Czekała (1996).

LEMMA 7. *The statistic $\phi_1(n, \underline{Z})$ has asymptotically the same distribution as*

$$\sum_{i=1}^{k} \phi_1(N_i, \underline{Y}^i) + \frac{1}{m}\Big(n \log n - \sum_{i=1}^{k} N_i \log N_i\Big). \tag{38}$$

LEMMA 8. *Let $V_{i,n}$, $i = 1, \ldots, k$, $n \ge 0$, be random variables defined as follows*:

$$V_{i,n} = \sigma_{1,n}^{-1}\Big(\phi_1(n, \underline{Y}^i) - \frac{n+1}{m} \log d_i - e_{1,n}\Big),$$

where $e_{1,n}$ and $\sigma_{1,n}$ are defined by (25) *and* (26). *Then*

$$\Big(V_{1,N_1}, \ldots, V_{k,N_k}, \frac{N_1 - np_1}{\sqrt{n}}, \ldots, \frac{N_k - np_k}{\sqrt{n}}\Big) \xrightarrow{d} (V_1, \ldots, V_k, W_1, \ldots, W_k), \tag{39}$$

where the V_i are independent and normally $N(0,1)$ distributed random variables, the vector $(W_1, \ldots, W_k)$ is independent of $(V_1, \ldots, V_k)$ and has the multivariate normal distribution $N(0, \Sigma)$, where $\Sigma = [\sigma_{i,j}]$ and

$$\sigma_{i,j} = \begin{cases} -p_i p_j & \text{for } i \ne j, \\ p_i - p_i^2 & \text{for } i = j, \end{cases} \qquad i, j = 1, \ldots, k.$$

LEMMA 9. *For N_i defined by* (34),

$$\frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{k} N_i \log \frac{d_i}{N_i} - \sum_{i=1}^{k} np_i \log \frac{nd_i}{p_i}\Big) \xrightarrow{d} \sum_{i=1}^{k} W_i \log \frac{d_i}{p_i}, \tag{40}$$

where W_i's are defined in Lemma 8.

Now we can state and prove the following theorem.

THEOREM 4. *If $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. random variables with density $f > 0$ given by* (32) *then*

$$\frac{G_{1,n} - e_{1,n}(f)}{\sigma_{1,n}(f)} \xrightarrow{d} N(0,1), \tag{41}$$

where

$$e_{1,n}(f) = \frac{n+1}{m}\Big(\sum_{i=1}^{m-1}\frac{1}{i} - \gamma + \mathrm{E}\{\log(f(X_1)^{-1})\}\Big), \tag{42}$$

$$\sigma^2_{1,n}(f) = \frac{n+1}{m^2}\Big(m\sum_{i=m}^{\infty}\frac{1}{i^2} - 1 + \sigma^2\{\log(f(X_1)^{-1})\}\Big). \tag{43}$$

Proof. For brevity, denote $m\sum_{i=m}^{\infty} i^{-2} - 1$ by C_m^2. Using Theorem 1 and Lemmas 7–9 it can be shown that

$$\frac{1}{\sqrt{n+1}}\bigg(\phi_1(n,Z) - \frac{n+1}{m}\Big(\sum_{i=1}^{m-1}\frac{1}{i} - \gamma + \log n + \sum_{i=1}^{k} p_i \log\frac{nd_i}{p_i}\Big)\bigg) \xrightarrow{d} \frac{C_m}{m}\sum_{i=1}^{k}\sqrt{p_i}\,V_i + \frac{1}{m}\sum_{i=1}^{k} W_i \log\frac{d_i}{p_i}. \tag{44}$$

The right-hand side of (44) is normally distributed with mean zero and variance

$$\frac{C_m^2}{m^2} + \frac{1}{m^2}\sigma^2\{\log(f(X_1)^{-1})\}.$$

If we divide both sides of (44) by the square root of this value and note that

$$\mathrm{E}\{\log(f(X_1)^{-1})\} = \int_0^1 f(x)\log(f(x)^{-1})\,dx = \sum_{i=1}^{k} f_i d_i \log f_i^{-1} = \sum_{i=1}^{k} p_i \log\frac{d_i}{p_i}$$

we get the assertion.

5. Bahadur approximate efficiencies. In this section we use the previous results to examine the Bahadur approximate ARE of tests based on $G_{i,n}$ for various $m \geq 1$. For $m = 1$ the Bahadur approximate efficiency of some tests based on spacings, including $G_{1,n}$ and $G_{2,n}$, was computed by Bartoszewicz (1995). This type of asymptotic efficiency of two sequences of tests was introduced by Bahadur (1960). It is defined as the quotient of the approximate slopes of these tests. The approximate slope can be defined for a sequence of tests that satisfies some regularity conditions. Such a sequence is called by Bahadur (1960) a standard sequence. If we are given a family of probability measures $\{P_\theta : \theta \in \Theta\}$ and test the hypothesis H that $\theta \in \Theta_0 \subset \Theta$ then a sequence $\{T_n\}$ of real-valued statistics is called a *standard sequence* (for testing H) if the following three conditions are satisfied:

I. There exists a continuous probability distribution function F such that, for each $\theta \in \Theta_0$,

$$\lim_{n\to\infty} P_\theta(T_n \le x) = F(x) \quad \text{for every } x.$$

II. There exists a constant a, $0 < a < \infty$, such that

$$\log(1 - F(x)) = -\frac{ax^2}{2}(1 + o(1)) \quad \text{as } x \to \infty.$$

III. There exists a function b on $\Theta - \Theta_0$, with $0 < b < \infty$, such that, for each $\theta \in \Theta - \Theta_0$,

$$\lim_{n\to\infty} P_\theta\left(\left|\frac{T_n}{\sqrt{n}} - b(\theta)\right| > x\right) = 0 \quad \text{for every } x > 0.$$

The *approximate slope* $c(\theta)$ is defined for a standard sequence $\{T_n\}$ as

$$c(\theta) = \begin{cases} 0 & \text{for } \theta \in \Theta_0, \\ a(b(\theta))^2 & \text{for } \theta \in \Theta - \Theta_0. \end{cases}$$

If $\theta \in \Theta - \Theta_0$ and $\{T_{1,n}\}$ and $\{T_{2,n}\}$ are two standard sequences we denote Bahadur's approximate efficiency of $\{T_{1,n}\}$ with respect to $\{T_{2,n}\}$ as $e_B(\{T_{1,n}\}, \{T_{2,n}\}, \theta)$. In our case we set

$$T_{i,n}^{(m)} = (G_{i,n} - e_{i,n})/\sigma_{i,n}, \quad i = 1, 2. \tag{45}$$

According to Theorems 1 and 2 condition I is satisfied with F being the distribution function of the standard normal distribution. Bahadur (1960) showed that for such F condition II is satisfied with $a = 1$. It remains to check condition III. We have

$$T_{i,n}^{(m)} = \frac{G_{i,n} - e_{i,n}(f)}{\sigma_{i,n}(f)} \cdot \frac{\sigma_{i,n}(f)}{\sigma_{i,n}} + \frac{e_{i,n}(f) - e_{i,n}}{\sigma_{i,n}}.$$

For brevity, set

$$R(m) = m \sum_{i=m}^{\infty} \frac{1}{i^2}, \quad m \ge 1.$$

Then from Theorems 1 and 4 we obtain

$$\frac{T_{1,n}^{(m)}}{\sqrt{n}} \xrightarrow{P_f} \frac{\mathrm{E}\{\log(f(X_1)^{-1})\}}{\sqrt{R(m) - 1}} \tag{46}$$

and from Theorems 2 and 3,

$$\frac{T_{2,n}^{(m)}}{\sqrt{n}} \xrightarrow{P_f} \frac{\mathrm{E}\{f(X_1)^{-1}\log(f(X_1)^{-1})\}}{\sqrt{R(m+1) - 1}}. \tag{47}$$

It is well known that if f is not identically 1 then

$$\mathrm{E}\{\log(f(X_1)^{-1})\} = -\int_0^1 \log(F(X))F(X)\,DX < 0$$

and

$$\mathrm{E}\{f(X_1)^{-1}\log(f(X_1)^{-1})\} = -\int_0^1 \log(f(x))\,dx > 0.$$

Thus we have shown that condition III is satisfied with $b_1(f)$ and $b_2(f)$ equal to the right-hand side of (46) and (47) respectively, and therefore we have proved the following theorem.

THEOREM 5. *Let $f > 0$ be a step density defined by* (32) *but not identically equal to* 1 *and let $\underline{X} = (X_1, X_2, \ldots, X_n, \ldots)$ be a sequence of i.i.d. random variables with density f. If $\{T_{i,n}^{(m)}\} = \{T_{i,n}^{(m)}(\underline{X})\}$, $i = 1, 2$, are the sequences of tests defined by* (45) *then their Bahadur approximate slopes $c_i^{(m)}(f)$ are given by*

$$c_1^{(m)}(f) = \frac{\mathrm{E}^2\{\log(f(X_1)^{-1})\}}{R(m)-1}, \tag{48}$$

$$c_2^{(m)}(f) = \frac{\mathrm{E}^2\{f(X_1)^{-1}\log(f(X_1)^{-1})\}}{R(m+1)-1}. \tag{49}$$

Both $c_1^{(m)}(f)$ and $c_2^{(m)}(f)$ depend on the same function $R(m)$. The basic properties of this function that are of interest to us are given in the next lemma.

LEMMA 10. *The function $R(m)$, $m \geq 1$, is strictly decreasing and*

$$\lim_{m\to\infty} R(m) = 1. \tag{50}$$

Proof. Let $Q(m) = \sum_{i=m}^{\infty} i^{-2}$. We have

$$R(m+1) = R(m) + Q(m+1) - m^{-1}$$

and it suffices to show that $Q(m+1) < m^{-1}$ for all $m \geq 1$. Set $a_m = m^{-1} + \sum_{i=1}^{m} i^{-2}$. We have

$$a_m - a_{m+1} = m^{-1}(m+1)^{-1} - (m+1)^{-2} > 0$$

and thus the sequence a_m is strictly decreasing. Since $\lim a_m = \pi^2/6$, we get

$$\frac{\pi^2}{6} < m^{-1} + \sum_{i=1}^{m} i^{-2} \quad \text{for all } m \geq 1, \tag{51}$$

Since

$$Q(m+1) = \frac{\pi^2}{6} - \sum_{i=1}^{m} i^{-2}$$

the above inequality is equivalent to $Q(m+1) < m^{-1}$, which proves the monotonicity of $R(m)$. The proof of (50) can be found in Cressie (1976).

From Theorem 5 we get the following formulae for Bahadur's approximate efficiencies:

$$e_B(\{T_{1,n}^{(m_1)}\}, \{T_{1,n}^{(m_2)}\}, f) = \frac{R(m_2) - 1}{R(m_1) - 1}, \tag{52}$$

$$e_B(\{T_{2,n}^{(m_1)}\}, \{T_{2,n}^{(m_2)}\}, f) = \frac{R(m_2+1) - 1}{R(m_1+1) - 1}, \tag{53}$$

$$e_B(\{T_{1,n}^{(m_1)}\}, \{T_{2,n}^{(m_2)}\}, f) = \frac{\mathrm{E}^2\{\log(f(X_1)^{-1})\}}{\mathrm{E}^2\{f(X_1)^{-1}\log(f(X_1)^{-1})\}} \cdot \frac{R(m_2+1) - 1}{R(m_1) - 1}. \tag{54}$$

It is interesting to note that the efficiencies in (52) and (53) do not depend on f. It follows from (52), (53) and Lemma 10 that when the size of the spacings increases to infinity then the efficiency of tests based on $G_{1,n}$ or $G_{2,n}$ increases to infinity independently of any particular alternative. Now let $f(x)$ be any step density and $f_\alpha(x) = 1 + \alpha(f(x) - 1)$, $\alpha > 0$. Then it is easy to show that

$$\lim_{\alpha \to 0} e_B(\{T_{1,n}^{(m_1)}\}, \{T_{2,n}^{(m_2)}\}, f_\alpha) = \frac{R(m_2+1) - 1}{R(m_1) - 1}.$$

Therefore, for step alternatives "close" to 1 and $m_1 = m_2$ the test based on $G_{1,n}$ is less efficient than that based on $G_{2,n}$. Both tests become equally efficient if we take $m_1 = m_2 + 1$.

References

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

J. Bartoszewicz (1995), *Bahadur and Hodges–Lehmann approximate efficiencies of tests based on spacings*, Statist. Probab. Lett. 23, 211–220.

N. Cressie (1976), *On the logarithms of high-order spacings*, Biometrika 63, 343–355.

F. Czekała (1996), *Normalizing constants for a statistic based on logarithms of disjoint spacings*, Appl. Math. (Warsaw) 23 (4), 405–416.

D. A. Darling (1953), *On a class of problems relating to the random division of an interval*, Ann. Math. Statist. 24, 239–253.

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

J. R. Gebert and B. K. Kale (1969), *Goodness of fit tests based on discriminatory information*, Statist. Hefte 3, 192–200.

S. R. Jammalamadaka and R. C. Tiwari (1986), *Efficiencies of some disjoint spacings tests relative to a* χ^2 *test*, in: M. L. Puri, J. Vilaplana and W. Wertz (eds.), *New Perspectives in Theoretical and Applied Statistics*, Wiley, New York, 311–318.

B. K. Kale (1969), *Unified derivation of tests of goodness of fit based on spacings*, Sankhyā Ser. A 31, 43–48.

F. Proschan and R. Pyke (1964), *Asymptotic normality of certain test statistics of exponentiality*, Biometrika 51, 253–255.

Franciszek Czekała
Mathematical Institute
University of Wrocław
Pl. Grunwaldzki 2/4
50-384 Wrocław, Poland
E-mail: czekala@math.uni.wroc.pl

Received on 25.9.1997