Ann. Inst. Statist. Math.
Vol. 40, No. 1, 165-178 (1988)

A CLASS OF MULTIPLE SAMPLE TESTS BASED ON EMPIRICAL COVERAGES

P. W. MIELKE, Jr. and Y. C. YAO

Department of Statistics, Colorado State University, Fort Collins, CO 80523, U.S.A.

(Received March 16, 1987; revised August 25, 1987)

Abstract. A class of multiple sample tests based on empirical coverages is proposed which is a generalization of Greenwood's and Sherman's one-sample goodness-of-fit test statistics. The asymptotic normality of the tests is established by embedding the empirical coverages into a stationary process satisfying the strong mixing condition.

Key words and phrases: Spacing, strong mixing condition, asymptotic normality.

1. Introduction

Let $X_1,\ldots, X_n$ be a random sample from an unknown distribution function F. For a given continuous distribution function F_0, define

$$C_i = F_0(X_{i,n}) - F_0(X_{i-1,n}), \qquad i = 1,\ldots, n + 1 ,$$

$$A_\nu = \sum_{i=1}^{n+1} |C_i - (n + 1)^{-1}|^\nu, \qquad \nu > 0 ,$$

where $X_{1,n}<\cdots<X_{n,n}$ are the order statistics of $X_1,\ldots, X_n$, $X_{0,n}=-\infty$, and $X_{n+1,n}=\infty$. To test for the hypothesis $F=F_0$, Greenwood (1946) and Sherman (1950) proposed the test statistics A_2 and A_1, respectively. Rao and Sethuraman (1975) studied the empirical distribution of the C_i (which are often called coverages or spacings) and derived the asymptotic behavior of A_ν. Some exact percentage points of A_1 and A_2 under $F=F_0$ have been obtained for selected values of n (Rao (1976), Burrows (1979), Currie (1981) and Stephens (1981)).

These one-sample goodness-of-fit tests A_ν can be generalized to the multiple sample case as follows. For $\alpha=1,\ldots, g$, (g: an integer≥ 2), let $X_j^{(\alpha)}$ ($j=1, \ldots, n_\alpha$) be a random sample of size n_α from an unknown distribution function $F^{(\alpha)}$. Set $N=n_1+\cdots+n_g$ and note that the N observations divide the real line into $N+1$ intervals. Let $F_N(x)=(N+1)^{-1}$ (number of observations among the N

pooled observations being $\leq x$) if $x<\infty$, $=1$ if $x=\infty$. Note that F_N is slightly different from the empirical distribution of the pooled sample. Now, define the $n_\alpha+1$ *empirical coverages* associated with the n_α observations of the α-th sample by

$$C_{j/\alpha} = F_N(X_{j/\alpha}) - F_N(X_{j-1/\alpha}), \quad j = 1,\ldots, n_\alpha + 1 , \tag{1.1}$$

where $X_{1/\alpha} < \cdots < X_{n_\alpha/\alpha}$ are the order statistics of $X_1^{(\alpha)},\ldots, X_{n_\alpha}^{(\alpha)}$, $X_{0/\alpha}=-\infty$, and $X_{n_\alpha+1/\alpha}=\infty$. For example, suppose $g=2$, $n_1=2$, $n_2=4$ and $X_{1/1}<X_{1/2}<X_{2/1}<X_{2/2}<X_{3/2}<X_{4/2}$. Then the empirical coverages are $C_{1/1}=1/7$, $C_{2/1}=2/7$, $C_{3/1}=4/7$, $C_{1/2}=C_{2/2}=2/7$ and $C_{3/2}=C_{4/2}=C_{5/2}=1/7$. Under the null hypothesis H_0 that the g samples come from a common continuous distribution function, it can be shown that $C_{j/\alpha}$ has mean $1/(n_\alpha+1)$. We therefore propose to reject H_0 if the statistic

$$B_\nu = \sum_{\alpha=1}^{g} a(n_\alpha/N) \sum_{j=1}^{n_\alpha+1} |C_{j/\alpha} - (n_\alpha + 1)^{-1}|^\nu \tag{1.2}$$

is large where $a:(0, 1)\to R$ is a positive weighting function. Holst and Rao (1980, 1981) considered the two-sample case and derived the asymptotic behavior of the statistic $\Sigma h_{jN}((N+1)C_{j/1}-1)$. Rao and Murthy (1981) proposed a two-sample statistic which is equivalent to B_2 with constant weighting function. But they did not derive the asymptotic distribution.

In the next section, it is shown, under H_0 and Condition (A) (see Theorem 2.1) and by embedding the empirical coverages into a stationary process satisfying the strong mixing condition, that $B_\nu(\nu>1/2)$ is asymptotically normal as $N\to\infty$. While we conjecture that Condition (A) holds in general, this condition is verified only for the special case of $\nu=2$ and $g=2$ in Section 3. We have not been able to obtain the power behavior of B_ν against general alternative hypotheses. Some numerical results are given in Section 4.

2. Asymptotic normality of B_ν under H_0

To derive the asymptotic normality of B_ν under H_0, we assume that $n_\alpha/N = r_\alpha$ ($\alpha=1,\ldots, g$) for some constants r_α with $0<r_\alpha<1$ and $r_1+\cdots+r_g=1$. Let $J(1), J(2), \ldots$ be an independent and identically distributed (iid) sequence with

$$P(J(1) = \alpha) = r_\alpha, \quad \alpha = 1,\ldots, g .$$

Define

$$I_k^{(\alpha)} = \begin{cases} 1, & \text{if} \quad J(k) = \alpha , \\ 0, & \text{otherwise} , \end{cases}$$

$$M_N^{(\alpha)} = \sum_{k=1}^{N} I_k^{(\alpha)} , \tag{2.1}$$

and

$$U_k = a(r_{J(k)})|\tilde{k} - k - r_{J(k)}^{-1}|^{\nu} , \tag{2.2}$$

where $\tilde{k}=\tilde{k}(k)=\inf\{j>k: J(j)=J(k)\}$.

THEOREM 2.1. *Assume that following Condition* (A) *holds.*

Condition (A): $\sigma^2=\mathrm{var}\,(U_1)+2\sum_{k=2}^{\infty}\mathrm{cov}\,(U_1, U_k)-\zeta'\Sigma^{-1}\zeta>0$,

where

$$\zeta' = \left(\sum_{k=1}^{\infty}\mathrm{cov}(U_1, I_k^{(1)}),\ldots, \sum_{k=1}^{\infty}\mathrm{cov}(U_1, I_k^{(g-1)})\right)$$

and Σ is a $(g-1)\times(g-1)$ matrix whose (i, j)-component equals $r_i(1-r_i)$ if $i=j$, $-r_ir_j$ if $i\neq j$. Then for $\nu>1/2$, under H_0, as $N\to\infty$, $\{(N+1)^{\nu}B_{\nu}-NE(U_1)\}/(N^{1/2}\sigma)$ converges to the standard normal distribution.

It should be remarked that the σ^2 in Condition (A) is actually a conditional limit variance (see Lemma 2.3). The evaluation of σ^2 is difficult for general ν and g. In the next section, σ^2 is explicitly given for $\nu=2$ and $g=2$.

LEMMA 2.1. *Let*

$$W_k = U_k + \sum_{\alpha=1}^{g-1} d_{\alpha}I_k^{(\alpha)} ,$$

for some constants $d_1,\ldots, d_{g-1}$. Then the stationary sequence $\{W_k\}$ satisfies the strong mixing condition and the mixing coefficients $\tilde{\alpha}(k)$ satisfy $\tilde{\alpha}(k)\leq D\rho^k$, $k=1, 2,\ldots$ for some $D>0$ and $0<\rho<1$.

(For the definitions of the strong mixing condition and the mixing coefficients, see pp. 305–306 of Ibragimov and Linnik (1971).)

PROOF OF LEMMA 2.1. Denote by $\sigma(X_1,\ldots, X_n)$ the σ-field generated by random variables $X_1,\ldots, X_n$. For $S_1\in\sigma(W_k, k=1,\ldots, l)$ and $S_2\in\sigma(W_k, k=l+m, l+m+1,\ldots)$, S_2 is independent of $S_1\cap E$ where E is the event that for each α $(1\leq\alpha\leq g)$, there exists k such that $l<k<l+m$ and $J(k)=\alpha$. So,

$$\begin{aligned}P(S_1\cap S_2) &= P(S_1\cap S_2|E)P(E) + P(S_1\cap S_2|E^c)P(E^c),\\ P(S_1\cap S_2|E)P(E) &= P(S_1|E)P(S_2|S_1\cap E)P(E) = P(S_2)P(S_1\cap E)\\ &= P(S_1)P(S_2) - P(E^c)P(S_2)P(S_1|E^c).\end{aligned}$$

Hence,

$$\begin{aligned}
&|P(S_1 \cap S_2) - P(S_1)P(S_2)| \\
&\quad = P(E^c)|P(S_1 \cap S_2|E^c) - P(S_2)P(S_1|E^c)| \\
&\quad \le P(E^c) \le \sum_{\alpha=1}^{g}(1 - r_\alpha)^{m-1} \le g(1 - \min_\alpha r_\alpha)^{m-1} .
\end{aligned}$$

This completes the proof. □

LEMMA 2.2. *Assume that Condition* (A) *of the theorem holds. Then*

$$N^{-1/2}\left(\sum_{k=1}^{N} (U_k - E(U_1)),\ M_N^{(1)} - Nr_1, \dots,\ M_N^{(g-1)} - Nr_{g-1}\right)$$

converges in distribution to $(Z_0, Z_1, \dots, Z_{g-1})$ *where the random variables* Z_0, $Z_1, \dots, Z_{g-1}$ *are jointly normal with* $E(Z_\alpha)=0$ $(\alpha=0,\dots,g-1)$, *and*

$$\begin{aligned}
E(Z_0^2) &= \operatorname{var}(U_1) + 2 \sum_{k=2}^{\infty} \operatorname{cov}(U_1, U_k) , \\
E(Z_\alpha^2) &= r_\alpha(1 - r_\alpha), && \alpha = 1,\dots, g - 1 , \\
E(Z_0 Z_\alpha) &= \sum_{k=1}^{\infty} \operatorname{cov}(U_1, I_k^{(\alpha)}), && \alpha = 1,\dots, g - 1 , \\
E(Z_\alpha Z_\beta) &= - r_\alpha r_\beta, && 1 \le \alpha \ne \beta \le g - 1 .
\end{aligned}$$

PROOF OF LEMMA 2.2. Let

$$W_k = d_0 U_k + \sum_{\alpha=1}^{g-1} d_\alpha I_k^{(\alpha)} ,$$

for not all zero constants $d_0, \dots, d_{g-1}$. If $d_0=0$, then $W_1, W_2, \dots$ are iid so that $N^{-1/2}(W_1 + \cdots + W_N - NE(W_1))$ converges to the normal distribution with mean 0 and variance

$$(d_1,\dots, d_{g-1})\, \Sigma (d_1,\dots, d_{g-1})' = \operatorname{var}(d_1 Z_1 + \cdots + d_{g-1} Z_{g-1}) .$$

Suppose $d_0 \ne 0$. Without loss of generality, we assume $d_0 = 1$. Thus, since U_k and $I_k^{(\beta)}$ $(k \ge 2)$ are independent of $I_1^{(\alpha)}$,

$$\begin{aligned}
(2.3)\qquad & \operatorname{var}(W_1) + 2 \sum_{k=2}^{\infty} \operatorname{cov}(W_1, W_k) \\
&\quad = \operatorname{var}(U_1) + 2 \sum_{k=2}^{\infty} \operatorname{cov}(U_1, U_k) + 2d'\zeta + d'\Sigma d
\end{aligned}$$

$$= \sigma^2 + (d + \Sigma^{-1}\zeta)'\Sigma(d + \Sigma^{-1}\zeta) > 0 \quad \text{by Condition (A)} ,$$

where $d' = (d_1, \dots, d_{g-1})$. By Lemma 2.1 and Theorem 18.5.3 of Ibragimov and Linnik (1971), $N^{-1/2}(W_1 + \cdots + W_N - NE(W_1))$ converges to the normal distribution with mean 0 and variance equal to (2.3). This variance also equals $\mathrm{var}(Z_0 + d_1 Z_1 + \cdots + d_{g-1} Z_{g-1})$. We have shown that every linear combination of the random variables $N^{-1/2} \sum_{k=1}^{N} (U_k - E(U_1)), \dots, N^{-1/2}(M_N^{(g-1)} - Nr_{g-1})$ converges to a normal distribution so that they converge jointly to $Z_0, \dots, Z_{g-1}$ in distribution. □

LEMMA 2.3. *As $N \to \infty$, for every z,*

$$P\left(N^{-1/2} \sum_{k=1}^{N} (U_k - E(U_1)) \le z \mid M_N^{(\alpha)} = n_\alpha,\ \alpha = 1, \dots, g-1\right)$$

converges to $\Phi(z/\sigma)$ where Φ is the standard normal distribution function.

PROOF OF LEMMA 2.3. It may be remarked that Holst (1979) considered a similar problem and derived conditional limit distributions. However, his results do not apply to our case since U_k are not independent. Intuitively, by Lemma 2.2, the conditional distribution of $N^{-1/2} \sum_{k=1}^{N} (U_k - E(U_1))$ given $M_N^{(\alpha)} = n_\alpha$ $(\alpha = 1, \dots, g-1)$ converges to the conditional distribution of Z_0 given $Z_\alpha = 0$ $(\alpha = 1, \dots, g-1)$, which is normal with mean 0 and variance σ^2. To make this rigorous, some delicate analysis is needed.

In the following, we will consider only the case $g = 2$. The general case can be treated similarly with more complicated notation. Let $L(X)$ and $L(X \mid Y = y)$ denote the distribution of X and the conditional distribution of X given $Y = y$. For each N, let $J_N'(1), \dots, J_N'(N)$ be independent of $J(k)$ $(1 \le k < \infty)$ such that

$$L(J_N'(1), \dots, J_N'(N)) = L(J(1), \dots, J(N) \mid M_N^{(1)} = n_1) .$$

For any fixed positive integer $\xi \le N^{1/2}$, choose, at random, ξ of those n_1 J_N''s that are equal to 1, and change the values of these ξ J_N''s to 2. Denote this new sequence by $J_N''(1), \dots, J_N''(N)$. Clearly,

$$L(J_N''(1), \dots, J_N''(N)) = L(J(1), \dots, J(N) \mid M_N^{(1)} = n_1 - \xi) .$$

Define

$$J_N'(k) = J_N''(k) = J(k) \qquad \text{for} \qquad k \ge N + 1,$$

and

$$U'_k = a(r_{J'_N(k)})|k' - k - r^{-1}_{J'_N(k)}|^\nu, \qquad k \le N, \tag{2.4}$$

$$U''_k = a(r_{J''_N(k)})|k'' - k - r^{-1}_{J''_N(k)}|^\nu, \qquad k \le N, \tag{2.5}$$

where $k'=k'(k)=\inf\{j>k: J'_N(j)=J'_N(k)\}$ and $k''=k''(k)=\inf\{j>k: J''_N(j)=J''_N(k)\}$. Clearly,

$$L(U'_k, k = 1, 2, \ldots) = L(U_k, k = 1, 2, \ldots | M_N^{(1)} = n_1), \tag{2.6}$$

$$L(U''_k, k = 1, 2, \ldots) = L(U_k, k = 1, 2, \ldots | M_N^{(1)} = n_1 - \xi). \tag{2.7}$$

For $k=1,\ldots, N$, U'_k and U''_k are different only if one of the three events $E_{N,i}(k)$ $(i=1, 2, 3)$ occurs, i.e.,

$$\{U'_k \neq U''_k\} \subset \bigcup_i E_{N,i}(k), \tag{2.8}$$

where the event $E_{N,1}(k)$ is that $J'_N(k)=1$, $J''_N(k)=2$; $E_{N,2}(k)$ is that $J'_N(k)=J''_N(k)=1$ and there exists a positive integer $l \le N-k$ such that $J'_N(i)=2$ $(i=k+1,\ldots, k+l-1)$, $J'_N(k+l)=1$, $J''_N(k+l)=2$; $E_{N,3}(k)$ is that $J'_N(k)=2$ and there exists a positive integer $l \le N-k$ such that $J'_N(i)=J''_N(i)=1$ $(i=k+1,\ldots, k+l-1)$, $J'_N(k+l)=1$, $J''_N(k+l)=2$. We have

$$\begin{aligned}
P(E_{N,1}(k)) &= \frac{n_1}{N}\cdot\frac{\xi}{n_1} = \frac{\xi}{N}, \\
P(E_{N,2}(k)) &= \sum_{l=1}^{N-k} \frac{n_1(n_1-1)n_2(n_2-1)\cdots(n_2-l+2)}{N(N-1)\cdots(N-l)}\cdot\frac{(n_1-\xi)\xi}{n_1(n_1-1)} \\
&\le \sum_{l=1}^{N-k}\left(\frac{n_2}{N}\right)^{l-1}\left(\frac{\xi}{n_1}\right) \\
&\le \left(1-\frac{n_2}{N}\right)^{-1}\left(\frac{\xi}{n_1}\right) = \left(\frac{\xi}{N}\right)\left(\frac{N}{n_1}\right)^2,
\end{aligned} \tag{2.9}$$

$$\begin{aligned}
P(E_{N,3}(k)) &= \sum_{l=1}^{N-k} \frac{n_2 n_1(n_1-1)\cdots(n_1-l+1)}{N(N-1)\cdots(N-l)} \\
&\quad\cdot\frac{(n_1-\xi)\cdots(n_1-\xi-l+2)\xi}{n_1(n_1-1)\cdots(n_1-l+1)} \\
&\le \sum_{l=1}^{N-k}\left(\frac{n_1}{N}\right)^{l}\left(\frac{\xi}{n_1}\right) \le \left(\frac{n_1}{N}\right)\left(1-\frac{n_1}{N}\right)^{-1}\left(\frac{\xi}{n_1}\right) = \left(\frac{\xi}{N}\right)\left(\frac{N}{n_2}\right).
\end{aligned} \tag{2.10}$$

So, for some constant $D_1>0$,

$$(2.11)\qquad P\Big(\bigcup_i E_{N,i}(k)\Big) \le D_1(\xi/N)\,.$$

Note that in (2.9) and (2.10), the l-th term of the sum decreases geometrically fast as l increases. Also, note that U_k' and U_k'' grow only polynomially fast in $k'-k$ and $k''-k$(see (2.4) and (2.5)), while $P(k'-k=l)$ and $P(k''-k=l)$ decrease geometrically fast as l increases. So, it can be shown that for some constant $D_2>0$,

$$(2.12)\qquad E\Big\{U_k'\Big|\bigcup_i E_{N,i}(k)\Big\} \le D_2,\ E\Big\{U_k''\Big|\bigcup_i E_{N,i}(k)\Big\} \le D_2\,,$$

for $k=1,\ldots, N$, $\xi\le N^{1/2}$ and large N. So, by (2.8), (2.11) and (2.12),

$$\begin{aligned} E\Big|\sum_{k=1}^{N}(U_k' - U_k'')\Big| &\le \sum_{k=1}^{N} E\Big\{|U_k' - U_k''|\Big|\bigcup_i E_{N,i}(k)\Big\}\, P\Big(\bigcup_i E_{N,i}(k)\Big) \\ &\le 2D_1D_2\xi\,. \end{aligned}$$

That is, for $0<\varepsilon<1$ and $\xi<\varepsilon^2N^{1/2}$ and for large N,

$$(2.13)\qquad P\Big(\Big|\sum_{k=1}^{N}(U_k' - U_k'')\Big| > \varepsilon N^{1/2}\Big) \le 2D_1D_2\xi/(\varepsilon N^{1/2}) \le 2D_1D_2\varepsilon\,.$$

By (2.13), for any fixed z, for $\xi<\varepsilon^2N^{1/2}$ and for large N,

$$\begin{aligned} &P\Big(N^{-1/2}\sum_{k=1}^{N}\{U_k' - E(U_1)\} \le z - \varepsilon\Big) - 2D_1D_2\varepsilon \\ &\quad\le P\Big(N^{-1/2}\sum_{k=1}^{N}\{U_k'' - E(U_1)\} \le z\Big) \\ &\quad\le P\Big(N^{-1/2}\sum_{k=1}^{N}\{U_k' - E(U_1)\} \le z + \varepsilon\Big) + 2D_1D_2\varepsilon\,. \end{aligned}$$

which, by (2.6) and (2.7), is equivalent to

$$\begin{aligned} (2.14)\quad &P\Big(N^{-1/2}\sum_{k=1}^{N}\{U_k - E(U_1)\} \le z - \varepsilon\Big|M_N^{(1)} = n_1\Big) - 2D_1D_2\varepsilon \\ &\quad\le P\Big(N^{-1/2}\sum_{k=1}^{N}\{U_k - E(U_1)\} \le z\Big|M_N^{(1)} = n_1 - \xi\Big) \\ &\quad\le P\Big(N^{-1/2}\sum_{k=1}^{N}\{U_k - E(U_1)\} \le z + \varepsilon\Big|M_N^{(1)} = n_1\Big) + 2D_1D_2\varepsilon\,, \end{aligned}$$

for all positive integers $\xi<\varepsilon^2N^{1/2}$ and for large N. Therefore,

$$(2.15)\qquad P\Big(N^{-1/2}\sum_{k=1}^{N}\{U_k - E(U_1)\} \le z\Big| - \varepsilon^2 < N^{-1/2}(M_N^{(1)} - Nr_1) < 0\Big)$$

is also bounded from below and above by the left-most term and the right-most term of (2.14), respectively. By Lemma 2.2, letting $N \to \infty$ and then $\varepsilon \to 0+$, (2.15) converges to

$$\lim_{\varepsilon \to 0+} P(Z_0 \le z \mid -\varepsilon^2 < Z_1 < 0) = P(Z_0 \le z \mid Z_1 = 0) = \Phi(z/\sigma) ,$$

and so

$$\overline{\lim_{\varepsilon \to 0+}}\ \overline{\lim_{N \to \infty}}\ P\left(N^{-1/2} \sum_{k=1}^{N} \{U_k - E(U_1)\} \le z - \varepsilon \,\middle|\, M_N^{(1)} = n_1\right) \le \Phi(z/\sigma)$$

$$\le \underline{\lim}_{\varepsilon \to 0+}\ \underline{\lim}_{N \to \infty}\ P\left(N^{-1/2} \sum_{k=1}^{N} \{U_k - E(U_1)\} \le z + \varepsilon \,\middle|\, M_N^{(1)} = n_1\right) . \tag{2.16}$$

Since (2.16) holds for all z and since $\Phi(z/\sigma)$ is continuous in z, we have

$$\lim_{N \to \infty} P\left(N^{-1/2} \sum_{k=1}^{N} \{U_k - E(U_1)\} \le z \,\middle|\, M_N^{(1)} = n_1\right) = \Phi(z/\sigma) . \qquad \square$$

PROOF OF THEOREM 2.1. Define, for $k = 1, \dots, N$,

$$V_{k,N} = a(r_{J(k)}) \left| \tilde{k} - k - \frac{N+1}{n_{J(k)} + 1} \right|^{v}$$

$$V'_{k,N} = \begin{cases} V_{k,N} , & \text{if} \quad \tilde{k} \le N \\ a(r_{J(k)}) \left| (N+1) - k - \dfrac{N+1}{n_{J(k)} + 1} \right|^{v} , & \text{if} \quad \tilde{k} > N , \end{cases}$$

where $\tilde{k} = \tilde{k}(k) = \inf\{j > k : J(j) = J(k)\}$. Also define, for $\alpha = 1, \dots, g$,

$$w_{\alpha,N} = a(r_\alpha) \left| \tilde{\alpha} - \frac{N+1}{n_\alpha + 1} \right|^{v} ,$$

where $\tilde{\alpha} = \tilde{\alpha}(\alpha) = \inf\{j : J(j) = \alpha\}$. Since, conditional on $M_N^{(\alpha)} = n_\alpha$ $(\alpha = 1, \dots, g-1)$, all of the $N! / \prod_{\alpha=1}^{g} n_\alpha!$ permutations of n_1 1's, $\dots$, n_g g's are equally likely, and since under H_0, all of the $N! / \prod_{\alpha=1}^{g} n_\alpha!$ possible allocations of the N measurements to the g samples of sizes $n_1, \dots, n_g$ are equally likely, we have

$$\begin{aligned} &L((N+1)^v B_v \mid H_0) \\ &\quad = L\left(\sum_{\alpha=1}^{g} w_{\alpha,N} + \sum_{k=1}^{N} V'_{k,N} \,\middle|\, M_N^{(\alpha)} = n_\alpha,\ \alpha = 1, \dots, g-1 \right) . \end{aligned} \tag{2.17}$$

It is easily seen that

$$(2.18)\quad L\left(\left\{\sum_{\alpha=1}^{g} w_{\alpha,N}+\sum_{k=1}^{N} V'_{k,N}\right\}-\sum_{k=1}^{N} V_{k,N}\middle| M_N^{(\alpha)}=n_\alpha,\ \alpha=1,\dots,g-1\right)$$
$$=O_p(1)\ .$$

By (2.17), (2.18) and Lemma 2.3, it suffices to show that

$$(2.19)\quad L\left(\sum_{k=1}^{N} U_k-\sum_{k=1}^{N} V_{k,N}\middle| M_N^{(\alpha)}=n_\alpha,\ \alpha=1,\dots,g-1\right)=o_p(N^{1/2})\ .$$

Since the r_α are fixed and k and $\tilde{k}(k)$ are integers, $\tilde{k}-k\neq r_{J(k)}^{-1}$ implies that $\tilde{k}-k-r_{J(k)}^{-1}$ is bounded away from 0. Clearly,

$$r_\alpha^{-1}-(N+1)/(n_\alpha+1)=O(N^{-1})\ ,$$

so that for large N, for $\tilde{k}-k\neq r_{J(k)}^{-1}$,

$$(2.20)\ |U_k-V_{k,N}|\le a(r_{J(k)})\nu|D|^{\nu-1}|r_{J(k)}^{-1}-(N+1)/(n_{J(k)}+1)|$$
$$\le a(r_{J(k)})2\nu|\tilde{k}-k-r_{J(k)}^{-1}|^{\nu-1}|r_{J(k)}^{-1}-(N+1)/(n_{J(k)}+1)|\ ,$$

where D is between $\tilde{k}-k-r_{J(k)}^{-1}$ and $\tilde{k}-k-(N+1)/(n_{J(k)}+1)$.

For $\tilde{k}-k=r_{J(k)}^{-1}$,

$$(2.21)\quad |U_k-V_{k,N}|\le a(r_{J(k)})|r_{J(k)}^{-1}-(N+1)/(n_{J(k)}+1)|^{\nu}=O(N^{-\nu})\ .$$

For $2^{-1}<\nu<1$, since

$$\sup\{|\tilde{k}-k-r_{J(k)}^{-1}|^{\nu-1}:\tilde{k}-k\neq r_{J(k)}^{-1}\}\le\sup_{1\le\alpha\le g}\sup_{j\neq r_\alpha^{-1}}|j-r_\alpha^{-1}|^{\nu-1}<\infty\ ,$$

it follows from (2.20) and (2.21) that

$$\left|\sum_{k=1}^{N}(U_k-V_{k,N})\right|=N\{O(N^{-1})+O(N^{-\nu})\}=O(N^{1-\nu})\ ,$$

proving (2.19) for $2^{-1}<\nu<1$. For $\nu\ge1$, by (2.20) and (2.21), for large N,

$$(2.22)\quad \left|\sum_{k=1}^{N}(U_k-V_{k,N})\right|\le\sum_{k=1}^{N} a(r_{J(k)})|r_{J(k)}^{-1}-(N+1)/(n_{J(k)}+1)|^{\nu}$$
$$+2\nu\sum_{k=1}^{N}\{a(r_{J(k)})|\tilde{k}-k-r_{J(k)}^{-1}|^{\nu-1}\}$$
$$\cdot|r_{J(k)}^{-1}-(N+1)/(n_{J(k)}+1)|$$
$$\le\{\sup_\alpha a(r_\alpha)\}\sum_{k=1}^{N} O(N^{-\nu})$$

$$+ 2\nu \sum_{k=1}^{N} \{U_k + a(r_{J(k)})\} \, O(N^{-1})$$
$$= O(N^{1-\nu}) + O(N^{-1}) \sum_{k=1}^{N} U_k + O(1) .$$

Since by Lemma 2.3,

$$L\left(\sum_{k=1}^{N} U_k \middle| M_N^{(\alpha)} = n_\alpha, \, \alpha = 1,\ldots, g-1 \right) = O_p(N) ,$$

it follows from (2.22) that

$$L\left(\sum_{k=1}^{N} (U_k - V_{k,N}) \middle| M_N^{(\alpha)} = n_\alpha, \, \alpha = 1,\ldots, g-1 \right) = O_p(1) ,$$

proving (2.19) for $\nu \geq 1$. □

3. The case of B_2 with $g=2$

Condition (A) of Theorem 2.1 is needed for applying the central limit theorem for strongly mixing sequences, i.e., Theorem 18.5.3 of Ibragimov and Linnik (1971). We conjecture that this condition is satisfied if the weights $a(r_\alpha)$ are not all zeroes. In this section, we compute σ^2 for the case of $\nu=2$ and $g=2$ and show that Condition (A) holds in this special case. However, we are not able to verify this condition for general ν and g.

For $\nu=2$ and $g=2$,

$$\sigma^2 = \mathrm{var}\,(U_1) + 2 \sum_{k=2}^{\infty} \mathrm{cov}(U_1, U_k) - r_1^{-1} r_2^{-1} \left\{ \sum_{k=1}^{\infty} \mathrm{cov}\,(U_1, I_k^{(1)}) \right\}^2 . \tag{3.1}$$

Let $\mu = E(U_1)$ and note that $r_2 = 1 - r_1$. For ease of notation, let $a_1 = a(r_1)$ and $a_2 = a(r_2)$. We now evaluate separately the three terms on the right-hand side of (3.1). For $\alpha = 1, 2$,

$$E(U_1 I_1^{(\alpha)}) = \sum_{i=2}^{\infty} a_\alpha (i - 1 - r_\alpha^{-1})^2 r_\alpha^2 (1 - r_\alpha)^{i-2}$$
$$= a_\alpha r_\alpha^{-1} (1 - r_\alpha) ,$$
$$\mu = E(U_1) = E(U_1 I_1^{(1)}) + E(U_1 I_1^{(2)}) = \sum_{\alpha=1}^{2} a_\alpha r_\alpha^{-1} (1 - r_\alpha) ,$$
$$E(U_1^2) = \sum_{\alpha=1}^{2} \sum_{i=2}^{\infty} a_\alpha^2 (i - 1 - r_\alpha^{-1})^4 r_\alpha^2 (1 - r_\alpha)^{i-2}$$
$$= \sum_{\alpha=1}^{2} a_\alpha^2 r_\alpha^{-3} \{(1 - r_\alpha) + 7(1 - r_\alpha)^2 + (1 - r_\alpha)^3\} ,$$

$$(3.2)\qquad \begin{aligned} \operatorname{var}(U_1) &= E(U_1^2) - \mu^2 \\ &= \sum_{\alpha=1}^{2} a_\alpha^2 r_\alpha^{-3}\{(1-r_\alpha) + 7(1-r_\alpha)^2 + (1-r_\alpha)^3\} \\ &\quad - \left\{\sum_{\alpha=1}^{2} a_\alpha r_\alpha^{-1}(1-r_\alpha)\right\}^2 , \end{aligned}$$

$$(3.3)\qquad \begin{aligned} \operatorname{cov}(U_1, I_1^{(1)}) &= E(U_1 I_1^{(1)}) - \mu r_1 \\ &= a_1 r_1^{-1} r_2 - r_1 \sum_{\alpha=1}^{2} a_\alpha r_\alpha^{-1}(1-r_\alpha) , \end{aligned}$$

$$(3.4)\qquad \begin{aligned} \operatorname{cov}(U_1, I_k^{(1)}) &= E(U_1 I_k^{(1)}) - r_1 E(U_1), \quad \text{for} \quad k \ge 2 \\ &= \Bigg\{ \sum_{\alpha=1}^{2}\sum_{i=2}^{k-1} a_\alpha (i-1-r_\alpha^{-1})^2 r_\alpha^2 (1-r_\alpha)^{i-2} r_1 \\ &\quad + a_1(k-1-r_1^{-1})^2 r_1^2 (1-r_1)^{k-2} \\ &\quad + \sum_{i=k+1}^{\infty} a_2 (i-1-r_2^{-1})^2 r_2^2 (1-r_2)^{i-2} \Bigg\} \\ &\quad - r_1 \sum_{\alpha=1}^{2}\sum_{i=2}^{\infty} a_\alpha (i-1-r_\alpha^{-1})^2 r_\alpha^2 (1-r_\alpha)^{i-2} . \end{aligned}$$

$$(3.5)\qquad \begin{aligned} \sum_{k=2}^{\infty} \operatorname{cov}(U_1, I_k^{(1)}) &= a_1 \sum_{k=2}^{\infty} (k-1-r_1^{-1})^2 r_1^2 (1-r_1)^{k-2} \\ &\quad + a_2 \sum_{k=2}^{\infty}\sum_{i=k+1}^{\infty} (i-1-r_2^{-1})^2 r_2^2 (1-r_2)^{i-2} \\ &\quad - r_1 \sum_{\alpha=1}^{2}\sum_{k=2}^{\infty}\sum_{i=k}^{\infty} a_\alpha (i-1-r_\alpha^{-1})^2 r_\alpha^2 (1-r_\alpha)^{i-2} \\ &= a_1 r_1^{-1} r_2 + a_2 r_2^{-2}(r_1 + 2r_1^2) - r_1 \sum_{\alpha=1}^{2} a_\alpha r_\alpha^{-2} (2(1-r_\alpha) \\ &\quad + (1-r_\alpha)^2) , \end{aligned}$$

$$(3.6)\qquad \begin{aligned} \sum_{k=2}^{\infty} \operatorname{cov}(U_1, U_k) &= \sum_{k=2}^{\infty} \{E(U_1 U_k) - \mu^2\} \\ &= \sum_{k=2}^{\infty} \Bigg[\sum_{\alpha=1}^{2}\sum_{i=2}^{k-1} E\{U_1 U_k \mid J(1) = J(i) = \alpha, J(j) \ne \alpha, \\ &\qquad 2 \le j \le i-1\} P\{J(1) = J(i) = \alpha, J(j) \ne \alpha, \\ &\qquad 2 \le j \le i-1\} - \mu^2 \Bigg] \\ &\quad + \sum_{k=2}^{\infty}\sum_{\alpha=1}^{2} E\{U_1 U_k \mid J(1) = J(k) = \alpha, J(j) \ne \alpha, \\ &\qquad 2 \le j \le k-1\} P\{J(1) = J(k) = \alpha, J(j) \ne \alpha, \\ &\qquad 2 \le j \le k-1\} \\ &\quad + \sum_{k=2}^{\infty}\sum_{\alpha=1}^{2} E\{U_1 U_k \mid J(1) = \alpha, J(j) \ne \alpha, 2 \le j \le k\} \\ &\quad \times P\{J(1) = \alpha, J(j) \ne \alpha, 2 \le j \le k\} . \end{aligned}$$

Since U_k is independent of the event $\{J(1)=J(i)=\alpha, J(j)\neq\alpha, 2\leq j\leq i-1\}$ for $i<k$, the first term on the right-hand side of (3.6) equals

$$\sum_{k=2}^{\infty}\left[\sum_{\alpha=1}^{2}\sum_{i=2}^{k-1} a_\alpha(i-1-r_\alpha^{-1})^2\mu r_\alpha^2(1-r_\alpha)^{i-2}-\mu^2\right]$$
$$= -\mu\sum_{k=2}^{\infty}\sum_{\alpha=1}^{2}\sum_{i=k}^{\infty} a_\alpha(i-1-r_\alpha^{-1})^2 r_\alpha^2(1-r_\alpha)^{i-2}$$
$$= -\mu\sum_{\alpha=1}^{2} a_\alpha r_\alpha^{-2}(2(1-r_\alpha)+(1-r_\alpha)^2)\ .$$

Similarly, after some tedious algebraic manipulations, the second and third terms equal, respectively,

$$\sum_{\alpha=1}^{2} a_\alpha^2 r_\alpha^{-3}\,(1-r_\alpha)^2$$

and

$$\sum_{\alpha=1}^{2} a_1a_2[(3+r_\alpha^2)+(1-r_\alpha+r_\alpha^2)(1-(1-r_\alpha)+(1-r_\alpha)^2)r_\alpha^{-1}(1-r_\alpha)^{-1}]\ .$$

From (3.2), (3.3), (3.5) and (3.6),

$$\sigma^2 = 4a_1^2r_1^{-3}r_2^2 + 4a_2^2r_1^2r_2^{-3} + 8a_1a_2\ ,$$

which is positive if $a_1^2+a_2^2>0$. So, from the theorem, under H_0,

$$(3.7)\qquad [(N+1)^2B_2 - N(a_1r_1^{-1}r_2 + a_2r_1r_2^{-1})]/[N(4a_1^2r_1^{-3}r_2^2 + 4a_2^2r_1^2r_2^{-3} + 8a_1a_2)]^{1/2}$$

converges to the standard normal distribution.

4. Some remarks

Remark 4.1. If one has a specific alternative in mind, then the weighting function a should be chosen to maximize the power against the alternative. However, since we are unable to obtain the asymptotic behavior of B_v against general alternatives, the problem of finding the optimal weighting function remains open.

Remark 4.2. While only the special case of B_2 with $g=2$ was considered in Section 3, the same techniques, with much more complicated calculation, can be used to derive the asymptotic mean and variance of B_v with any even number v and any g. But, we are not able to deal with the case of non-even v.

Remark 4.3. It can be shown that

$$E\left[\left(C_{j/\alpha} - \frac{1}{n_\alpha + 1}\right)^2 \middle| H_0\right] = \frac{2(N+1) - n_\alpha}{(N+1)(n_\alpha+1)(n_\alpha+2)} - \frac{1}{(n_\alpha+1)^2}.$$

So, for $g=2$,

$$E(B_2|H_0) = a_1\left\{\frac{2(N+1) - n_1}{(N+1)(n_1+2)} - \frac{1}{n_1+1}\right\} + a_2\left\{\frac{2(N+1) - n_2}{(N+1)(n_2+2)} - \frac{1}{n_2+1}\right\}. \tag{4.1}$$

By (3.7), it is easily seen that

$$[(N+1)^2 B_2 - (N+1)^2 E(B_2|H_0)]/[N(4a_1^2 r_1^{-3} r_2^2 + 4a_2^2 r_1^2 r_2^{-3} + 8a_1 a_2)]^{1/2} \tag{4.2}$$

is also asymptotically standard normal. Rao and Murthy (1981) presented some limited simulation results for a statistic equivalent to B_2 with $g=2$ and $a\equiv 1$. From their results, we obtained the (estimated) 95th percentiles of B_2 for several pairs of (n_1, n_2). We then used the two normal approximations (3.7) and (4.2) to estimate the probability that B_2 exceeds the 95th percentile. The results are given in Table 1. It appears that approximation (4.2) is better than (3.7). Both approximations are poor either when n_1 and n_2 are small or when the ratio n_1/n_2 is far from 1. In addition, they both tend to overestimate the upper tail probabilities.

Table 1. Normal approximations (3.7) and (4.2).

n_1	9	9	9	9	19	19	19	19	49
n_2	9	19	49	99	19	49	99	199	49
(3.7)	0.253	0.177	0.180	0.276	0.161	0.102	0.116	0.170	0.064
(4.2)	0.147	0.098	0.098	0.163	0.099	0.060	0.069	0.105	0.043

Acknowledgements

We would like to thank the referees for their valuable comments and for bringing several relevant papers to our attention.

References

Burrows, P. M. (1979). Selected percentage points of Greenwood's statistic, *J. Roy. Statist. Soc. Ser. A*, **142**, 256–258.

Currie, I. D. (1981). Further percentage points of Greenwood's statistic, *J. Roy. Statist. Soc. Ser. A*, **144**, 360–363.

Greenwood, M. (1946). The statistical study of infectious diseases, *J. Roy. Statist. Soc. Ser. A*, **109**, 85–103.

Holst, L. (1979). Two conditional limit theorems with applications, *Ann. Statist.*, **7**, 551–557.

Holst, L. and Rao, J. S. (1980). Asymptotic theory for some families of two-sample nonparametric statistics, *Sankhyā Ser. A*, **42**, 19–52.

Holst, L. and Rao, J. S. (1981). Asymptotic spacings theory with applications to the two-sample problem, *Canad. J. Statist.*, **9**, 79–89.

Ibragimov, I. A. and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff Publishing, Groningen, The Netherlands.

Rao, J. S. (1976). Some tests based on arc lengths for the circle, *Sankhyā Ser. B*, **38**, 329–338.

Rao, J. S. and Murthy, V. K. (1981). A two-sample nonparametric test based on spacing-frequencies, *Proc. of Int. Statist. Inst. Contributed Papers Volume* (43rd session), 223–227.

Rao, J. S. and Sethuraman, J. (1975). Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors, *Ann. Statist.*, **3**, 299–313.

Sherman, B. (1950). A random variable related to the spacing of sample values, *Ann. Math. Statist.*, **21**, 339–361.

Stephens, M. A. (1981). Further percentage points of Greenwood's statistic, *J. Roy. Statist. Soc. Ser. A*, **144**, 364–366.