Statistical Theory and Data Analysis

K. Matusita, editor

North-Holland

STATISTICAL THEORY AND DATA ANALYSIS
K. Matusita (Editor)

ON TESTS FOR THE TWO-SAMPLE PROBLEM BASED ON HIGHER ORDER SPACING-FREQUENCIES

S. Rao Jammalamadaka and R.L. Schweitzer

University of California, Santa Barbara, CA

Gordon College, MA

1. *INTRODUCTION*

Let $U_1, U_2, \ldots, U_{N-1}$ and $V_1, \ldots, V_n$ be independent random samples from continuous distribution functions (d.f.) F and G respectively. The two-sample problem consists of testing the hypothesis that these two distributions are the same. As will be apparent soon, we may assume without loss of generality that a probability transformation has been performed on the data so that both distributions have the unit interval [0,1] as their support, and the first sample comes from the uniform distribution on [0,1]. Let

$$G^* = G \circ F^{-1}$$

be the d.f. of the second sample after the transformation. Then the null hypothesis to be tested may be written in the form

$$(1.1) \qquad H_0: \quad G^*(v) = v, \quad 0 \leqslant v \leqslant 1.$$

Let

$$0 \leqslant U_1' \leqslant U_2' \leqslant \ldots \leqslant U_{N-1}' \leqslant 1$$

denote the order statistics from the first sample. Further, define

$$U_0' = 0 \quad \text{and} \quad U_k' = 1 + U_{k-N}' \quad \text{for } k \geqslant N,$$

circularly for convenience. The first order spacings are given by

$$T_k = U'_k - U'_{k-1}, \quad k = 1,\ldots,N;$$

while the m-th order spacings are defined as

$$(1.2) \qquad T_k^{(m)} = U'_{k+m-1} - U'_{k-1}, \quad k = 1,2,\ldots,N.$$

Tests based on first order spacings have been studied extensively for the one-sample goodness-of-fit problem. See for example Pyke [13] and Rao and Sethuraman [17]. Results using higher-order spacings, with $m > 1$, have been obtained by Del Pino [3], Cressie [2] and Kuo and Rao [10].

For $k = 1,\ldots,N$ define the first order *spacing-frequencies* as

$$S_k = \text{the number of } V_j\text{'s in the interval } [U'_{k-1}, U'_k)$$

and the m-th order spacing frequencies by

$$(1.3) \qquad S_k^{(m)} = \sum_{j=0}^{m-1} S_{k+j} = \text{the number of } V_j\text{'s in the interval } [U'_{k-1}, U'_{k+m-1}),$$

where $S_k = S_{k-N}$ for $k > N$.

This paper deals with tests of the two-sample problem based on these m-th order spacing frequencies. Since these numbers $\{S_k^{(m)}\}$ remain invariant under probability transformations, we can assume the distribution of the first sample to be uniform and frame the hypothesis as in (1.1).

Tests based on $\{S_k\}$, the first - order spacing frequencies, for the two-sample problem have been considered by Dixon [4], Godambe [6], Blumenthal [1], Rao [15], Holst and Rao [8], [9] and Rao and Mardia [16].

Since we will be concerned primarily with asymptotics, we take two non-decreasing sequences of positive integers, $\{N_\nu\}$ and $\{n_\nu\}$, and assume throughout that

$$N_\nu, n_\nu \to \infty \quad \text{as } \nu \to \infty,$$

in such a way that the ratio

$$r_\nu = \frac{N_\nu}{n_\nu} \to \rho, \quad 0 < \rho < 1.$$

Note that the spacings, dependent on N_ν, should be labeled as $\{T_{k\nu}^{(m)}\}$. Similarly, the spacing frequencies, dependent on both N_ν and n_ν, should be labeled $\{S_{k\nu}^{(m)}\}$. So we are dealing with triangular arrays of random variables,

$$\{T_{k\nu}^{(m)}, k = 1,\ldots,N_\nu\} \quad \text{and} \quad \{S_{k\nu}^{(m)}, k = 1,\ldots,N_\nu\} \quad \text{for} \quad \nu \geqslant 1.$$

Corresponding to the ν-th such array, let $h_\nu(\cdot)$ and $\{h_{k\nu}(\cdot), k = 1,\ldots,N_\nu\}$ be real-valued functions satisfying some regularity conditions to be specified later. Define

$$Z_\nu = \frac{1}{\sqrt{N_\nu}} \sum_{k=1}^{N_\nu} h_{k\nu}(S_{k\nu}^{(m)}) \quad \text{and} \quad Z_\nu^* = \frac{1}{\sqrt{N_\nu}} \sum_{k=1}^{N_\nu} h(S_{k\nu}^{(m)}),$$

based on the $(N_\nu - 1)$ U-values and the n_ν V-values. Though Z_ν^* is just a special case of Z_ν where the functions $\{h_{k\nu}\}$ do not vary with k, we will distinguish these two cases throughout, since their asymptotic behaviour under local alternatives, is quite different. The Wald-Wolfowitz Run Test and Dixon [4] test are of the type Z_ν^*, whereas the Wilcoxon-Mann-Whitney test is of the form Z_ν. In fact, any linear function of the U-ranks in the combined sample can be expressed as a special case of Z_ν. The dependence on ν will be suppressed to simplify notation except where it is essential for clarity.

A few words about notations: The symbol "$\sim$" stands for "distributed as" while "$\xrightarrow{\mathcal{D}}$" will denote "convergence in distribution". For any sequence of random variables X_n, we write

$$X_n = o_p(g(n)) \quad \text{if } X_n/g(n) \to 0 \text{ in probability}$$

and write

$$X_n = 0_p(g(n))$$

if, for each $\varepsilon > 0$, there is a $K_\varepsilon < \infty$ such that

$$P(|X_n/g(n)| > K_\varepsilon) < \varepsilon$$

for all n sufficiently large. $N(\mu,\Sigma)$ will denote a normal distribution with mean μ and variance-covariance matrix Σ, and $U(0,1)$ the uniform distribution on the unit interval. $\text{Mult}(n;\ p_1,\ldots,p_k)$ will denote a multinomial distribution based on n trials with k cells having probabilities $p_1,\ldots,p_k$ and η will stand for a negative binomial random variable with probability function

$$(1.4)\qquad P(\eta = j) = \binom{m+j-1}{j}\left(\frac{1}{1+\rho}\right)^j\left(\frac{\rho}{1+\rho}\right)^m, \qquad j = 0,1,2,\ldots$$

$\text{Poi}(\lambda)$ will denote a Poisson distribution with parameter λ, and

$$\pi_j(\lambda) = e^{-\lambda}\cdot\frac{\lambda^j}{j!}$$

the probability at j. Finally $\Gamma(m,1)$ will denote a gamma distribution with density

$$(1.5)\qquad f_m(x) = \begin{cases} x^{m-1}e^{-x}/\Gamma(m), & x \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

The three following facts about the spacing frequencies $(S_1,\ldots S_N)$ will be useful, later on.

(a) Under the null hypothesis, (1.1),

$$(1.6)\qquad P(S_1 = j_1, S_2 = j_2, \ldots, S_N = j_N) = \frac{1}{\binom{n+N-1}{n}}$$

for each vector of non-negative integers $(j_1,\ldots,j_N)$ s.t.

$$\sum_{k=1}^{N} j_k = n.$$

(b) If we let $\xi_1,\ldots,\xi_N$ denote independent and identically distributed (i.i.d.) geometric random variables with

$$P[\xi_k = j] = \frac{r}{(1 + r)^{j+1}}, \; j = 0,1,2,\ldots$$

then it is easy to verify that, under the null hypothesis

$$(S_1,\ldots,S_N) \sim (\xi_1,\ldots,\xi_N \mid \sum_{k=1}^{N} \xi_k = n). \tag{1.7}$$

(c) Given the vector of spacings

$$\underset{\sim}{T} = (T_1,T_2,\ldots,T_N),$$

let $X_1,X_2,\ldots,X_N$ be independent Poisson random variables with

$$X_k \sim \text{Poi}(nT_k), \quad k = 1,2,\ldots,N.$$

Then under the null hypothesis

$$(S_1,S_2,\ldots,S_N \mid \underset{\sim}{T}) \sim (X_1,X_2,\ldots,X_N \mid \sum_{i=1}^{N} X_i = n) \tag{1.8}$$

Frequently in this paper it will be convenient to use the notation $\xi_k^{(m)}$, $X_k^{(m)}$, etc. to denote rolling sums defined analogously to those in (1.3) for $S_k^{(m)}$.

The paper is organized as follows: In Section 2, we consider test statistics of the form Z_ν, and derive their asymptotic distribution under the null hypothesis. In Section 3 we obtain the distribution of these statistics under an appropriate sequence of local alternatives. By computing the Pitman Asymptotic Relative Efficiencies (ARE's) we show that among this class, the asymptotically Locally Most Powerful (LMP) test is a linear combination of the spacing-frequencies. Section 4 deals with the symmetric statistic Z^*_ν and its asymptotic distribution. By considering alternatives which converge to the null at the rate of $n^{-\frac{1}{4}}$, the asymptotically LMP test is derived to be that based on sum of squares of the spacing-frequencies. The results of this paper extend those of Holst and Rao [8] to the m-th order spacing-frequencies where $m \geq 1$.

2. NONSYMMETRIC TESTS UNDER THE NULL HYPOTHESIS

We consider the class of statistics of the form

$$Z_\nu = N^{-\frac{1}{2}} \sum_1^N h_k(S_k^{(m)})$$

where the functions $\{h_k\}$ satisfy the following

ASSUMPTION A. The real-valued functions $\{h_k(\cdot)\}$ defined on $\{0,1,2,\ldots\}$ satisfy Assumption A if they are of the form

$$h_k(j) = h(\frac{k}{N+1}, j) \quad \text{for } k = 1,2,\ldots,N,\ j = 0,1,2,\ldots,$$

for some function $h(u,j)$ defined on $(0,1) \times \{0,1,2,\ldots\}$ with the properties:

(i) $h(u,j)$ is continuous in u, except perhaps for finitely many u, and the set of discontinuities, if any, is independent of j;

(ii) $h(u,j)$ is not of the form $c \cdot j + h(u)$ where $h(\cdot)$ is a function on $[0,1]$ and $c \in R$;

(iii) $\int_0^1 E(h^4(u,\eta))du < \infty$, where η has the negative binomial distribution (1.4).

We may add, without loss of generality,

(iv) $E(\frac{1}{\sqrt{N}} \sum_1^N h_k(S_k^{(m)})) = 0$ under the null hypothesis, (1.1).

Recall that by the representation given in (1.7) we have

$$Z_\nu = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} h_k(\xi_k^{(m)}) \mid \left(\sum_1^N \xi_k = n \right)$$

where $\xi_1, \ldots, \xi_N$ are i.i.d. geometric random variables. The following theorem on the asymptotic distribution of Z_ν can be obtained as a special case of Theorem 2 of Holst [7], p. 553, by taking $(\xi_k, h_k(\xi_k))$ in place of (X_k, Y_k) in that theorem.

THEOREM 2.1. If $M, N \to \infty$ such that (hereafter abbreviated s.t.)

$$\frac{M}{N} \to \gamma, \quad 0 < \gamma \leqslant 1,$$

and there exists some $\gamma_0 < 1$ s.t. $\gamma_0 \leqslant \gamma \leqslant 1$ implies that

$$\begin{bmatrix} \sum_{k=1}^{M-m} h_k(\xi_k^{(m)})/\sqrt{N} \\ \sum_{k=1}^{M} (\xi_k - \frac{1}{r})\, r/\sqrt{N(1+r)} \end{bmatrix} \overset{\mathcal{D}}{\to} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} A_\gamma & B_\gamma \\ B_\gamma & \gamma \end{bmatrix} \right),$$

where A_γ and B_γ are constants s.t.

$$A_\gamma \to A_1, \quad B_\gamma \to B_1 \quad \text{as } \gamma \to 1^-,$$

then

$$\left(\frac{1}{\sqrt{N}} \sum_{k=1}^{N} h_k(S_k^{(m)}) \right) \overset{\mathcal{D}}{\to} N(0, A_1 - B_1^2). \quad \square$$

To establish the asymptotic normality required in Theorem 2.1, we need only verify the following Liapunov type condition (2.1) for m-dependent sequences. (cf. Orey [12])

PROPOSITION 2.1. Let $\{X_k : k \geqslant 1\}$ be a sequence of m-dependent random variables with zero means, and for some $\delta > 0$, let

$$E(X_k^{2+\delta}) = \alpha_k < \infty, \quad k = 1,2,\ldots .$$

Let σ_n^2 be the variance of $(X_1 + \cdots + X_n)$. If

$$(2.1) \qquad (\sum_1^n \alpha_k)/\sigma_n^{2+\delta} \to 0 \quad \text{as} \quad n \to \infty,$$

then

$$\frac{X_1 + \cdots + X_n}{\sigma_n} \xrightarrow{\mathcal{D}} N(0,1). \quad \square$$

This gives us an easily verified condition that is sufficient for Theorem 2.1 to hold, using which we prove

THEOREM 2.2. If the functions $\{h_k(\cdot)\}$ satisfy Assumption A, and if $M,N \to \infty$ s.t. $\frac{M}{N} \to \gamma$, $0 < \gamma \leqslant 1$, then

$$\begin{bmatrix} \frac{1}{\sqrt{N}} \sum_1^{M-m} h_k(\xi_k^{(m)}) \\ \frac{1}{\sqrt{N}} \sum_1^{M} (\xi_k - \frac{1}{r}) \frac{r}{\sqrt{1+r}} \end{bmatrix} \xrightarrow{\mathcal{D}} N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} A_\gamma & B_\gamma \\ B_\gamma & \gamma \end{bmatrix})$$

where A_γ, B_γ are constants s.t. $A_\gamma \to A_1$ and $B_\gamma \to B_1$ as $\gamma \to 1^-$.

Proof: The joint asymptotic normality is established by showing that condition (2.1) holds for the $(m - 1)$ dependent random sequence defined (for any fixed real numbers t and s) by

$$X_k = t \cdot h_k(\xi_k^{(m)}) + s.(\xi_k - \frac{1}{r}) r/\sqrt{1 + r} .$$

The argument is very similar to that of Corollary 2.1 of Holst and Rao [8]. Because of condition (iii) in Assumption A, the term

$$N^{\delta/2} \sum_{k=1}^{M-m} E[N^{-\frac{1}{2}} X_k]^{2+\delta} = N^{-1} \sum_{k=1}^{M-m} E(X_k^{2+\delta})$$

tends to a finite limit as does

$$\mathrm{Var}[N^{-\frac{1}{2}} \sum_{k=1}^{M-m} X_k].$$

Thus in the ratio

$$\frac{\sum_{1}^{M-m} E[N^{-\frac{1}{2}} X_k]^{2+\delta}}{\mathrm{Var}[\sum_{1}^{M-m} N^{-\frac{1}{2}} X_k]^{1+\delta/2}},$$

the numerator converges to zero while the denominator remains bounded away from zero so that Condition (2.1) is satisfied.

All that remains then is the calculation of A_γ and B_γ. Now

$$A_\gamma = \lim_{N\to\infty} (N^{-\frac{1}{2}} \sum_{1}^{M-m} h_k(\xi_k^{(m)}))$$

$$= \sum_{k=-m+1}^{m-1} \int_0^\gamma \mathrm{Cov}(h(u,\xi_1^{(m)}), h(u,\xi_{1+k}^{(m)}))du$$

and

$$B_\gamma = \lim_{N\to\infty} \mathrm{Cov}(\sum_{1}^{M-m} h_k(\xi_k^{(m)})/\sqrt{N}, \sum_{1}^{M} (\xi_k - \frac{1}{r}) r/\sqrt{N(1+r)})$$

$$= \lim_{N\to\infty} \frac{r}{N\sqrt{1+r}} \sum_{k=1}^{M-m} \mathrm{Cov}(h_k(\xi_k^{(m)}), \xi_k^{(m)})$$

$$= \frac{\rho}{\sqrt{1+\rho}} \int_0^\gamma \mathrm{Cov}(h(u,\eta),\eta)du,$$

where

$$\rho = \lim_{\nu\to\infty} \frac{N_\nu}{n_\nu}$$

and η has the negative binomial distribution (1.4). We see finally then that

$$A_1 - B_1^2 = \sum_{-m+1}^{m-1} \int_0^1 \mathrm{Cov}(h(u,\xi_1^{(m)}), h(u,\xi_{1+k}^{(m)}))du$$

$$- (\int_0^1 \mathrm{Cov}(h(u,\eta),\eta)du)^2 \frac{\rho^2}{1+\rho} . \square$$

These results are summed up in the following corollary:

COROLLARY 2.1. If $n, N \to \infty$ s.t. $\frac{N}{n} \to \rho$, $0 < \rho < 1$, and the functions $\{h_k\}$ satisfy Assumption A, then under the null hypothesis (1.1),

$$\frac{1}{\sqrt{N}} \sum_1^N h_k(S_k^{(m)}) \xrightarrow{\mathcal{D}} N(0,\sigma^2),$$

where

$$(2.2) \qquad \sigma^2 = \sum_{-m+1}^{m-1} \int_0^1 \mathrm{Cov}(h(u,\xi_1^{(m)}), h(u,\xi_{1+k}^{(m)}))du$$

$$- (\int_0^1 \mathrm{Cov}(h(u,\eta),\eta)du)^2 \frac{\rho^2}{1+\rho} . \square$$

Specializing to the symmetric case Z_ν^* we obtian:

COROLLARY 2.2. Under the null hypothesis (1.1), if $h(j)$ defined for $j = 0,1,2,...$ is non-linear in j and $E\, h^4(\eta) < \infty$ when η has distribution (1.4), then

$$N^{-\frac{1}{2}} \sum_{k=1}^N [h(S_k^{(m)}) - E\, h(\eta)] \xrightarrow{\mathcal{D}} N(0,\sigma^2)$$

as $N \to \infty$, where

$$\sigma^2 = \sum_{-m+1}^{m-1} \mathrm{Cov}(h(\xi_1^{(m)}), h(\xi_{1+j}^{(m)})) - \frac{\rho^2}{1+\rho} \cdot \mathrm{Cov}^2(h(\eta),\eta). \square$$

3. *NON-SYMMETRIC TESTS UNDER CLOSE ALTERNATIVES*

Given the U-observations, the probability that a V-observation will fall in the interval $[U'_{k-1}, U'_{k+m-1})$ is

$$T_k^{(m)} = U'_{k+m-1} - U'_{k-1}$$

under the null hypothesis. More generally, it is

$$D_k^{(m)} = G(U'_{k+m-1}) - G(U'_{k-1})$$

when the alternative G holds. It is clear that the conditional distribution of the spacing frequencies is

$$(3.1) \qquad (S_1, S_2, \ldots, S_N) \,|\, (U_1, \ldots, U_N) \sim \text{Mult.}(n;\ D_1, D_2, \ldots, D_N),$$

under the alternative G.

We will study the asymptotic behaviour of the statistic Z_ν under the sequence of alternatives

$$(3.2) \qquad G_N(v) = v + \frac{L_N(v)}{\sqrt{N}}, \quad 0 < v < 1.$$

The function $L_N(v)$ and its derivatives

$$\ell_N(v) = \frac{d}{dv} L_N(v), \ \ell'_N(v) = \frac{d^2}{dv^2} L_N(v)$$

satisfy the following regularity condition:

ASSUMPTION B.

$$L_N(0) = L_N(1) = 0,$$

and there exists a continuous function $L(v)$ s.t. for $0 \leq v \leq 1$,

$$L_N(v) = \sqrt{N}[G_N(v) - v] \to L(v) \quad \text{as } N \to \infty.$$

Further

$$\ell_N(v), \quad \ell_N'(v), \quad \ell(v) = \frac{d}{dv} L(v), \quad \ell'(v) = \frac{d^2}{dv^2} L(v)$$

all exist and are continuous with

$$\sup_{0 \leq v \leq 1} |\ell_N'(v) - \ell'(v)| = o(1).$$

Under this assumption we have

$$nD_k^{(m)} = n(G_N(U_{k+m-1}') - G_N(U_{k-1}'))$$

$$= n(U_{k+m-1}' - U_{k-1}' + \frac{L_N(U_{k+m-1}') - L_N(U_{k-1}')}{\sqrt{N}})$$

$$= nT_k^{(m)}(1 + \frac{L_N(U_{k+m-1}') - L_N(U_{k-1}')}{U_{k+m-1}' - U_{k-1}'} \frac{1}{\sqrt{N}})$$

$$= nT_k^{(m)}(1 + \frac{\ell(\frac{k}{N+1})}{\sqrt{N}}) + o_p(N^{-\frac{1}{2}}),$$

where $o_p(\cdot)$ is uniform in k.

Now analogous to the representation (1.8) for the spacing frequencies, we may write that conditional on the vector $\underset{\sim}{T}$, Z_ν is

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{N} h_k(S_k^{(m)}) | \underset{\sim}{T} \sim \frac{1}{\sqrt{N}} \sum_{k=1}^{N} h_k(Y_k^{(m)}) | (\sum_{1}^{N} Y_k = n)$$

where the Y_k's are independent Poisson random variables, with

$$Y_k \sim \text{Poi}(nD_k).$$

Note then that

$$Y_k^{(m)} = \sum_{j=0}^{m-1} Y_{k+j} \sim \text{Poi}(nD_k^{(m)}),$$

and that conditional on $\underset{\sim}{T}$, these $\{Y_k^{(m)}\}$ form an $(m-1)$-dependent sequence. Recall also that under the null hypothesis, we use the corresponding sequences

$$X_k \sim \mathrm{Poi}(nT_k) \quad \text{and} \quad X_k^{(m)} \sim \mathrm{Poi}(nT_k^{(m)}).$$

The following two lemmas are necessary to derive the distribution of Z_ν under the alternatives $\{G_N\}$

LEMMA 3.1. Let $M, N \to \infty$ s.t. $\frac{M}{N} \to \gamma$, $0 < \gamma < 1$. Then

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{M-m} E(h_k(Y_k^{(m)}) \,|\, D_k^{(m)}) - \frac{1}{\sqrt{N}} \sum_{k=1}^{M-m} E(h_k(X_k^{(m)}) \,|\, T_k^{(m)}) \overset{P}{\to} A(\gamma),$$

where

$$A(\gamma) = \frac{\rho}{1+\rho} \int_0^\gamma \ell(u)\,\mathrm{Cov}(h(u,\eta),\eta)\,du.$$

Proof: Define the function

$$g(r,x) = \sum_{j=0}^{\infty} h(r,j)\pi_j(x).$$

This function is continuous and has continuous derivatives of all orders with respect to its second argument since a Poisson random variable has finite moments of all orders. Note that

$$g\left(\frac{k}{N+1}, nD_k^{(m)}\right) = \sum_{j=0}^{\infty} h\left(\frac{k}{N+1}, j\right)\pi_j(nD_k^{(m)})$$

$$= E(h_k(Y_k^{(m)}) \,|\, D_k^{(m)})$$

and similarly

$$g\left(\frac{k}{N+1}, nT_k^{(m)}\right) = E(h_k(X_k^{(m)}) \,|\, T_k^{(m)}).$$

Using a Taylor expansion in the second argument around $nT_k^{(m)}$, we obtain

$$g\left(\frac{k}{N+1}, nD_k^{(m)}\right) = g\left(\frac{k}{N+1}, nT_k^{(m)}\left[1 + \frac{\ell\left(\frac{k}{N+1}\right)}{\sqrt{N}}\right] + o_p\left(\frac{1}{\sqrt{N}}\right)\right)$$

$$= g\left(\frac{k}{N+1}, nT_k^{(m)}\right)$$

$$+ \frac{\ell\left(\frac{k}{N+1}\right)}{\sqrt{N}} nT_k^{(m)} g_x\left(\frac{k}{N+1}, nT_k^{(m)}\right) + o_p\left(\frac{1}{\sqrt{N}}\right)$$

where g_x denotes the partial derivative with respect to the second argument.

Thus

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{M-m} E(h_k(Y_k^{(m)}) \mid D_k^{(m)})$$

$$= \frac{1}{\sqrt{N}} \sum_{k=1}^{M-m} g\left(\frac{k}{N+1}, nD_k^{(m)}\right)$$

$$= \frac{1}{\sqrt{N}} \sum_{1}^{M-m} g\left(\frac{k}{N+1}, nT_k^{(m)}\right)$$

$$+ \frac{1}{N} \sum_{1}^{M-m} \ell\left(\frac{k}{N+1}\right) nT_k^{(m)} g_x\left(\frac{k}{N+1}, nT_k^{(m)}\right) + o_p(1).$$

Notice that

$$E\left(nT_k^{(m)} g_x\left(\frac{k}{N+1}, nT_k^{(m)}\right)\right)$$

$$= E\left(\sum_{j=0}^{\infty} h\left(\frac{k}{N+1}, j\right) \pi_j(nT_k^{(m)})[j - nT_k^{(m)}] \mid nT_k^{(m)}\right)$$

We now need the two following facts (i) the distribution of

$$NT_k^{(m)} = r_\nu \cdot nT_k^{(m)}$$

converges to that of a $\Gamma(m,1)$ random variable, say S and (ii) if

$$S \sim \Gamma(m,1) \quad \text{and} \quad (\eta \mid S = s) \sim \mathrm{Poi}(S/\rho),$$

then η has the negative binomial distribution defined in (1.4). Thus the expectation given above tends to

$$\mathrm{Cov}(h(\frac{k}{N+1}, \eta), \eta) \cdot \frac{\rho}{1+\rho} \ .$$

Now by the law of large numbers and Lemma 2.1 of Holst and Rao [8] on convergence of sums to an integral, we have as $N \to \infty$

$$\frac{1}{N} \sum_{k=1}^{M-m} \ell(\frac{k}{N+1}) n T_k^{(m)} g_x(\frac{k}{N+1}, n T_k^{(m)})$$

$$\overset{P}{\to} \frac{\rho}{1+\rho} \int_0^{\gamma} \ell(u) \mathrm{Cov}(h(u,\eta),\eta)\, du.$$

$$= A(\gamma). \ \square$$

LEMMA 3.2. Let $M,N \to \infty$ s.t. $\frac{M}{N} \to \gamma$, $0 < \gamma < 1$. Then for any real numbers s,t the ratio

$$\frac{E(\exp\{ \sum_1^{M-m} [\frac{is}{\sqrt{N}}(h_k(Y_k^{(m)}) - E(h_k(Y_k^{(m)}) \mid \underset{\sim}{D}))] + \sum_1^M \frac{it}{\sqrt{n}}(Y_k - nD_k)\} \mid \underset{\sim}{D})}{E(\exp\{ \sum_1^{M-m} [\frac{is}{\sqrt{N}}(h_k(X_k^{(m)}) - E(h_k(X_k^{(m)}) \mid \underset{\sim}{T}))] + \sum_1^M \frac{it}{\sqrt{n}}(Y_k - nT_k)\} \mid \underset{\sim}{T})}$$

(3.3)

converges to 1 in probability.

Proof: Let

$$\mu_k = E(h_k(Y_k^{(m)}) \mid \underset{\sim}{D}).$$

Using a Taylor expansion for the numerator in (3.3), it is seen to be bounded by

$$1 + iE(\{ \sum_1^{M-m} \frac{s}{\sqrt{N}} (h_k(Y_k^{(m)}) - \mu_k) + \sum_1^M \frac{t}{\sqrt{n}} (Y_k - nD_k)\} | \underset{\sim}{D})$$

$$- \frac{1}{2} E[\{ \sum_1^{M-m} \frac{s}{\sqrt{N}} (h_k(Y_k^{(m)}) - \mu_k) + \sum_1^M \frac{t}{\sqrt{n}} (Y_k - nD_k)\}^2 | \underset{\sim}{D}]$$

$$+ \frac{1}{6} E[| \sum_1^{M-m} \frac{s}{\sqrt{N}} (h_k(Y_k^{(m)}) - \mu_k) + \sum_1^M \frac{t}{\sqrt{n}} (Y_k - nD_k)|^3 | \underset{\sim}{D}].$$

It is clear that the second term of this expression is zero. To deal with the third term, we expand those terms involving $\{\underset{\sim}{D}\}$ as expressions involving $\{\underset{\sim}{T}\}$ using Taylor series, as in the proof of Lemma 3.1. After considerable simplification, the numerator in (3.3) can be shown to be

$$1 - \frac{1}{2N} \{ \sum_{j=-m+1}^{M-m} \sum_{k=1}^{M-m} s^2 [E(h_k(X_k^{(m)}) h_{k+j}(X_{k+j}^{(m)}) | \underset{\sim}{T})$$

$$- E(h_k(X_k^{(m)}) | \underset{\sim}{T}) E(h_{k+j}(X_{k+j}^{(m)}) | \underset{\sim}{T})]$$

$$+ t^2 \rho \sum_1^M (nT_k)$$

$$+ 2st\sqrt{\rho} \sum_1^{M-m} [E(h_k(X_k^{(m)}) X_k^{(m)} | \underset{\sim}{T})$$

$$- nT_k^{(m)} E(h_k(X_k^{(m)}) | \underset{\sim}{T})]\}$$

$$+ o_p(\frac{1}{\sqrt{N}}).$$

Next we consider the denominator in (3.3). The analog of (3.4) for the denominator is identical except that $X_k^{(m)}$ is substituted for $Y_k^{(m)}$,

$$\mu_k' = E(h_k(X_k^{(m)}) | \underset{\sim}{T})$$

for μ_k and $nT_k^{(m)}$ for $nD_k^{(m)}$, $\underset{\sim}{T}$ for $\underset{\sim}{D}$ throughout. As before, the second term is seen to be equal to zero, and the remainder term can be shown to converge to zero in probability. The resultant expression is the same as (3.5). Thus the ratio (3.3) converges in probability to 1. □

As noted before, under these alternatives, we have

$$E(\exp(\frac{it}{\sqrt{N}}\sum_{1}^{M-m} h_k(S_k^{(m)})|\underset{\sim}{D}))$$

$$= E(\exp(\frac{it}{\sqrt{N}}\sum_{1}^{M-m} h_k(Y_k^{(m)}))|\sum_{1}^{N} Y_k = n).$$

Applying Theorem 1 of Holst [7] we obtain

$$E(\exp(\frac{it}{\sqrt{N}}\sum_{1}^{M-m} h_k(S_k^{(m)})|\underset{\sim}{D}))$$

$$= (2\pi P\{\sum_{1}^{N} Y_k = n|\underset{\sim}{D}\})^{-1}$$

$$\times \int_{-\pi}^{\pi} E(\exp(\frac{it}{\sqrt{N}}\sum_{1}^{M-m} h_k(Y_k^{(m)}) + is\sum_{1}^{N}(Y_k - nD_k)|\underset{\sim}{D}))ds.$$

Since

$$\sum_{1}^{N} Y_k|\underset{\sim}{D} \sim \text{Poi}(n)$$

for any vector $\underset{\sim}{D}$, it follows that

$$P\{\sum_{1}^{N} Y_k = n|\underset{\sim}{D}\} = \frac{e^{-n}n^n}{n!} = (2\pi n)^{-\frac{1}{2}}\exp(o(1)),$$

by Stirling's formula. Thus, we have

$$(3.6)\qquad E(\exp(\frac{it}{\sqrt{N}}\sum_{1}^{M-m} h_k(S_k^{(m)})))$$

$$= E(E(\exp(\frac{it}{\sqrt{N}}\sum_{1}^{M-m} h_k(S_k^{(m)})|\underset{\sim}{D})))$$

(Contd)

(Contd) $$= \frac{1}{\sqrt{2\pi}} \exp(o(1))$$

$$\times \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} E\{E(\exp(\frac{it}{\sqrt{N}} \sum_{1}^{M-m} h_k(Y_k^{(m)}) + \frac{is}{\sqrt{n}} \sum_{1}^{M} (Y_k - nD_k)) | \underset{\sim}{D})$$

$$\times E(\exp(\frac{is}{\sqrt{n}} \sum_{M+1}^{N} (Y_k - nD_k)) | \underset{\sim}{D})\} ds.$$

Combining these results, we have

THEOREM 3.1. Let $n, N \to \infty$ s.t. $\frac{N}{n} \to \rho$, $0 < \rho < 1$. Let the functions $\{h_k\}$ satisfy Assumption A. Let η denote a negative binomial random variable as in (1.4). For the sequence of alternatives (3.2) satisfying Assumption B, we have

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{N} h_k(S_k^{(m)}) \xrightarrow{\mathcal{D}} N(\mu, \sigma^2)$$

where

$$\mu = [\int_0^1 \ell(u) \text{Cov}(h(u,\eta),\eta) du] \frac{\rho}{1+\rho}$$

and

$$\sigma^2 = \sum_{j=-m+1}^{m-1} \int_0^1 \text{Cov}(h(u,\xi_1^{(m)}), h(u,\xi_{1+j}^{(m)})) du$$

$$- (\int_0^1 \text{Cov}(h(u,\eta),\eta) du)^2 \frac{\rho^2}{1+\rho} .$$

Proof: In (3.6) above it is clear that the integrand is dominated by the function $f_\nu(s)$ defined as

(3.7) $$f_\nu(s) = E\left| E(\exp(\frac{is}{\sqrt{n}} \sum_{M+1}^{N} (Y_k - nD_k)) | \underset{\sim}{D}) \right|$$ (Contd)

(Contd)

$$= E\left|\exp\left(n\left(1 - \sum_{1}^{M} T_j - \frac{L_N\left(\sum^{M} T_j\right)}{\sqrt{N}}\right)\left(e^{-is/\sqrt{n}} - 1 - \frac{is}{\sqrt{n}}\right)\right)\right|$$

$$\to f(s) = e^{-(1-\gamma)s^2/2} \quad \text{as } \gamma \to \infty.$$

It is clear that $f(s)$ is integrable, and so by the Lebesgue Dominated Convergence Theorem we have as $\nu \to \infty$,

$$\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} f_\nu(s)\,ds \to \int_{-\infty}^{\infty} f(s)\,ds. \tag{3.8}$$

Application of the Extended Lebesgue Dominated Convergence Theorem (cf. C.R. Rao [14] p. 136) gives, along with Lemmas 3.1, 3.2 and equations (3.6), (3.7) and (3.8) that

$$\lim_{\nu\to\infty} E\left(\exp\left(\frac{it}{\sqrt{N}} \sum_{1}^{M-m} h_k(S_k^{(m)})\right)\right)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(itA(\gamma))$$

$$\times \lim E\left[E\left\{\exp\left(\frac{it}{\sqrt{n}} \sum_{1}^{M-m} h_k(X_k^{(m)}) + \frac{is}{\sqrt{n}} \sum_{1}^{M} (X_k - nT_k)\right) \middle| \underset{\sim}{T}\right\}\right.$$

$$\left.\times E\left(\exp\left(\frac{is}{\sqrt{n}} \sum_{M+1}^{N} (X_k - nT_k)\right) \middle| \underset{\sim}{T}\right)\right]ds$$

$$= \exp(itA(\gamma)) \lim E\left(\exp\left(\frac{it}{\sqrt{N}} \sum_{1}^{M} h_k(S_k^{(m)\prime})\right)\right)$$

where

$$(S_1', S_2', \ldots, S_N') \,|\, \underset{\sim}{T} \sim \text{Mult}(n; T_1, T_2, \ldots, T_N).$$

The result now follows from Corollary 2.1. □

Next, we take up the topic of the asymptotically most powerful test. The Pitman asymptotic relative efficiency (ARE) of one test statistic of level α relative to another is defined as the limit of the inverse ratio of sample sizes required to achieve the same limiting power at a sequence of alternatives converging to the null hypothesis. If a particular test has limiting power in the open interval $(\alpha,1)$, then a measure of its rate of convergence, called its "efficacy", can be computed. If we let $\mu(h)$ and $\sigma^2(h)$ denote the asymptotic mean and variance of the test statistic $Z_\nu(h)$ based on a function $h(\cdot,\cdot)$, the efficiency of the test, under certain general regularity conditions, including asymptotic normality of the test statistic, is

$$e_h^2 = \frac{\mu^2(h)}{\sigma^2(h)}$$

(see Fraser [5]). The ARE of $Z_\nu(h_1)$ relative to $Z_\nu(h_2)$ can be calculated as

$$\left(\frac{\mu^2(h_1)}{\sigma^2(h_1)} \right) / \left(\frac{\mu^2(h_2)}{\sigma^2(h_2)} \right) .$$

The test with maximum efficacy has asymptotically maximum local power.

Thus for tests of the form:

$$\text{Reject } H_0 \text{ for } \frac{1}{\sqrt{N}} \sum_{k=1}^{N} h_k(S_k^{(m)}) > C,$$

we would want to find the function h which maximizes

$$(3.9) \qquad e_h = \int_0^1 \ell(u)\operatorname{Cov}(h(u,\eta),\eta)\,du \, \frac{\rho}{1+\rho}$$
$$\times \Big[\sum_{-m+1}^{m-1} \int_0^1 \operatorname{Cov}(h(u,\xi_1^{(m)}),h(u,\xi_{1+j}^{(m)}))\,du$$
$$- \Big(\int_0^1 \operatorname{Cov}(h(u,\eta),\eta)\,du\Big)^2 \frac{\rho^2}{1+\rho} \Big]^{-\frac{1}{2}}.$$

LEMMA 3.3. The value of e_h in equation (3.9) is maximized by taking

$$h(u,j) = \ell(u)\cdot j,$$

with the resulting maximum value being

$$\max e_h = [\int_0^1 \ell^2(u)du]^{\frac{1}{2}}[1 + \rho]^{-\frac{1}{2}}$$

under alternatives of the form (3.2) satisfying Assumption B.

Proof: Consider a non-degenerate statistic

$$W_\nu(h) = \sum_{k=1}^{N} h(\frac{k}{N+1}, S_k^{(m)})$$

with

$$\mathrm{Var}(W(h)) = \sigma_h^2.$$

It is easily verified that if

$$h_2(t) = \alpha h_1(t) + \beta$$

where $\alpha \neq 0$ and β are real numbers, then

$$e_{h_1} = e_{h_2}.$$

In particular we may take

$$\alpha = [\mathrm{Var}(h_1(t))]^{-\frac{1}{2}},$$

so that

$$\mathrm{Var}(h_2(t)) = 1.$$

Therefore we may assume without loss of generality that

$$\sigma_h^2 = 1.$$

Thus, we consider the class M of all h satisfying Assumption A with σ_h^2 defined as in (2.2) assumed to be 1.

Then for any $h \in M$

$$e_h = \int_0^1 \ell(u)\,\mathrm{Cov}(h(u,\eta),\eta)\,du \,\frac{\rho}{1+\rho} . \tag{3.10}$$

By the Cauchy-Schwarz inequality,

$$\mathrm{Cov}(h(u,\eta),\eta) \leqslant \sqrt{\mathrm{Var}(h(u,\eta))}\ \sqrt{\mathrm{Var}(\eta)}$$

with equality if and only if

$$h(u,j) = a(u)\cdot j$$

where a(u) is continuous on [0,1]. Thus e_h in (3.10) attains its maximum when

$$h(u,j) = a(u)\cdot j$$

is such that

$$a(u)\cdot\ell(u) \geqslant 0 \quad \text{for } 0 \leqslant u \leqslant 1.$$

Since

$$\sum_{-m+1}^{m-1} \mathrm{Cov}(\xi_1^{(m)},\xi_{1+k}^{(m)}) = \mathrm{Cov}(\xi_1^{(m)}, \sum_{-m+1}^{m-1}\sum_{j=0}^{m-1} \xi_{1+k+j})$$

$$= \sum_{j=0}^{m-1} \mathrm{Cov}(\xi_1^{(m)}, \sum_{k=-m+1}^{m-1} \xi_{1+k+j})$$

$$= \sum_{j=0}^{m-1} \mathrm{Cov}(\eta,\eta)$$

$$= \sum_{j=0}^{m-1} m(\frac{1+\rho}{\rho^2}) = m^2\,\frac{1+\rho}{\rho^2} ,$$

$$e_h = \frac{\int_0^1 \ell(u)a(u)\,\frac{m(1+\rho)}{\rho^2}\,du\,\frac{\rho}{1+\rho}}{\left[\int_0^1 a^2(u)du\,\frac{m^2(1+\rho)}{\rho^2} - \left(\int_0^1 a(u)du\right)^2\frac{\rho^2}{1+\rho}\left[\frac{m(1+\rho)}{\rho^2}\right]^2\right]^{\frac{1}{2}}}$$

$$= \frac{\int_0^1 \ell(u)a(u)du}{\left[\int_0^1 a^2(u)du - \left(\int_0^1 a(u)du\right)^2\right]^{\frac{1}{2}}}\,[1+\rho]^{-\frac{1}{2}}$$

$$= \mathrm{Cor}(\ell(U),a(U))\cdot[\mathrm{Var}(\ell(U))]^{\frac{1}{2}}[1+\rho]^{-\frac{1}{2}},$$

where $\mathrm{Cor}(X,Y)$ denotes the correlation coefficient between the random variables X and Y. From this it is seen that e_h is maximized by taking

$$a(u) = \ell(u), \quad 0 \leqslant u \leqslant 1.$$

Further,

$$\max e_h = \left(\int_0^1 \ell^2(u)du\right)^{\frac{1}{2}}[1+\rho]^{-\frac{1}{2}}. \quad \square$$

Summing up the results of Lemma 3.3 and Theorem 3.1, we obtain:

THEOREM 3.2. If the sequence of alternatives satisfies Assumption B, then the asymptotically most powerful (AMP) test of the null hypothesis against the alternatives (3.2) is to reject H_0 when

$$\text{(3.11)} \qquad T^* = \sum_{k=1}^{N} \ell\left(\frac{k}{N+1}\right)S_k^{(m)} > C,$$

where C is a constant determined by the significance level α. The asymptotic distribution of this optimal statistic is given by

$$\text{(3.12)} \qquad \frac{1}{\sqrt{N}}\sum_{k=1}^{N} \ell\left(\frac{k}{N+1}\right)\left(S_k^{(m)} - \frac{mn}{N}\right) \xrightarrow{\mathcal{D}} N(0,\sigma^2)$$

under H_0, with

$$\sigma^2 = \frac{m^2(1+\rho)}{\rho^2}\int_0^1 \ell^2(u)\,du, \tag{3.13}$$

while under the alternatives (3.2)

$$\frac{1}{\sqrt{N}}\sum_{k=1}^{N} \ell\left(\frac{k}{N+1}\right)\left(S_k^{(m)} - \frac{mn}{N}\right) \xrightarrow{\mathcal{D}} N\left(\frac{m}{\rho}\int_0^1 \ell^2(u)\,du, \sigma^2\right). \quad \square \tag{3.14}$$

REMARK 3.1. This AMP statistic T* in (3.11) has the same efficacy as the corresponding statistic for m = 1 derived in Holst and Rao [8]. Thus for finite m, if one were to use the AMP test, there would be no gain in considering m-spacing frequencies with $m > 1$. For applications of non-symmetric tests, the reader may refer to Holst and Rao [8].

4. *SYMMETRIC TESTS BASED ON SPACING FREQUENCIES*

This section deals with the class of statistics symmetric in $\{S_1^{(m)}, S_2^{(m)}, \ldots, S_N^{(m)}\}$, i.e., the class of statistics of the form

$$Z_\nu^* = \sum_{k=1}^{N} h(S_k^{(m)}) \tag{4.1}$$

for some given function h(·) satisfying certain regularity assumptions. Such tests are rotationally invariant, and thus are useful for the problem of testing the equivalence of two distributions on a circle. It is clear that Z_ν^* is a special case of the statistic Z_ν discussed in the previous section, and its asymptotic distribution under the null hypothesis is given by Corollary 2.2. However, since

$$\int_0^1 \ell(u)\,du = 0,$$

it follows as a consequence of Theorem 3.1 that the asymptotic distribution of Z_ν^* under the sequence of alternatives (3.2) coincides with that under the null hypothesis. Thus, symmetric statistics of the type (4.1) have no power against such close alternatives. In order to make efficiency comparisons, we consider here the more distant alternatives

$$G_N^*(v) = v + \frac{L_N(v)}{N^{\frac{1}{4}}}, \quad 0 \leqslant v \leqslant 1. \tag{4.2}$$

with

$$L_N(v) = N^{\frac{1}{4}}(G_N(F^{-1}(u)) - u).$$

For this symmetric case, we will make the slightly stronger assumption on L_N:

ASSUMPTION B*. Assume L_N is twice differentiable on [0,1], and that there is a function L(u), $0 \leqslant u \leqslant 1$, which is twice continuously differentiable and has the properties

$$L(0) = L(1) = 0 \quad \text{and} \quad N^{\frac{1}{4}} \sup_{0 \leqslant u \leqslant 1} |L_N''(u) - \ell'(u)| = o(1),$$

where ℓ, ℓ' denote the first and second derivatives of L, respectively.

Note that for such smooth alternatives satisfying B* the following also hold:

$$N^{\frac{1}{4}} \sup_{0 \leqslant u \leqslant 1} |L_N(u) - L(u)| = o(1)$$

and

$$N^{\frac{1}{4}} \sup_{0 \leqslant u \leqslant 1} |L_N'(u) - \ell(u)| = o(1).$$

Also, here

$$\begin{aligned} nD_k^{(m)} &= n(G_N(U'_{k+m-1}) - (G_N(U'_{k-1})) \\ &= n\left(U'_{k+m-1} - U'_{k-1} + \frac{L_N(U'_{k+m-1}) - L_N(U'_{k-1})}{N^{\frac{1}{4}}}\right) \quad \text{(Contd)} \end{aligned}$$

(Contd) $$= nT_k^{(m)}(1 + \ell(k/N + 1)N^{-\frac{1}{4}}) + o_p(N^{-\frac{1}{2}})$$

where $o_p(\cdot)$ is uniform in k.

Let $W_0, W_1, \ldots, W_{n-1}$ be i.i.d. exp(1) r.v.s. with pdf e^{-w} for $w \geqslant 0$. Define the rotating partial sums (of m terms at a time)

$$W_k^{(m)} = \sum_{j=0}^{m-1} W_{k+j}, \qquad k = 0,1,\ldots,n-1$$

with the convenient notation

$$W_j = W_{j-n} \quad \text{for} \quad j \geqslant n.$$

Let S stand for a $\Gamma(m,1)$ r.v. with pdf as in (1.5). From the representation (1.8), the conditional mean under the alternatives is given by

$$\mu_\nu(\underset{\sim}{D}) = E(\sum_{k=1}^{N} h(S_k^{(m)}) | \underset{\sim}{D}) = \sum_{k=1}^{N} (\sum_{j=0}^{\infty} h(j)\pi_j(nD_k^{(m)})).$$

This is of the form

$$\sum_{1}^{N} g(nD_k^{(m)})$$

where

$$g(t) = \sum_{j=0}^{\infty} h(j)\pi_j(t).$$

This function g(t) satisfies the condition

(4.3) $$g(t) \leqslant C_1(t^{C_2} + 1)$$

for some non-negative constants C_1 and C_2 if h(j) satisfies a similar condition

$$h(j) \leqslant d_1(j^{d_2} + 1) \tag{4.4}$$

for nonnegative constants d_1 and d_2. Now we utilize a result (Theorem 4.2) of Kuo and Rao [10] on the asymptotic distribution of the statistic

$$\mu_\nu(\underset{\sim}{D}) = \sum_{1}^{N} g(nD_k^{(m)}),$$

based on the m-spacings. Observe that the condition (4.4) on $h(\cdot)$, which implies condition (4.3) on $g(\cdot)$, satisfies the Assumption II required there. (See their Remark 1 immediately following Assumption II)

THEOREM 4.1. (Kuo and Rao [10]). Let the sequence of alternatives (4.2) satisfy Assumption B* and let $h(\cdot)$ satisfy condition 4.4. Then under the alternatives (4.2),

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{N} [g(nD_k^{(m)}) - E(g(\frac{S}{\rho}))] \xrightarrow{\mathcal{D}} N(\mu,\sigma^2),$$

where

$$\mu = (\int_0^1 \ell^2(u)\,du)\,\mathrm{Cov}(g(\frac{S}{\rho}),(S - m - 1)^2)/2$$

and

$$\sigma^2 = \sum_{-m+1}^{m-1} \mathrm{Cov}(g(\frac{W_0^{(m)}}{\rho}),g(\frac{W_k^{(m)}}{\rho})) - (\mathrm{Cov}(g(\frac{S}{\rho}),S))^2. \ \square$$

As a consequence, we obtain:

COROLLARY 4.2. Let $h(\cdot)$ be any function, non-linear in the integers satisfying (4.4) and let

$$\mu_\nu(\underset{\sim}{D}) = E(\sum_{k=1}^{N} h(S_k^{(m)})\,|\,\underset{\sim}{D}) = \sum_{k=1}^{N} g(nD_k^{(m)}),$$

and let

$$\mu_\nu = E(\sum_{k=1}^{N} h(S_k^{(m)}) | \underset{\sim}{T}) = \sum_{k=1}^{N} g(nT_k^{(m)})$$

be as defined before. Then

$$\frac{1}{\sqrt{N}} (\mu_\nu(\underset{\sim}{D}) - E\mu_\nu) \overset{\mathcal{D}}{\to} N(b,c),$$

where

$$(4.6) \qquad b = (\int_0^1 \ell^2(u)du) \, Cov(\sum_{j=0}^{\infty} h(j)\pi_j(\frac{S}{\rho}), (S - m - 1)^2)/2$$

and

$$(4.7) \qquad c = \sum_{-m+1}^{m-1} Cov(\sum_{j=0}^{\infty} h(j)\pi_j(\frac{W_0^{(m)}}{\rho}), \sum_{j=0}^{\infty} h(j)\pi_j(\frac{W_k^{(m)}}{\rho}))$$

$$- (Cov(\sum_{j=0}^{\infty} h(j)\pi_j(\frac{S}{\rho}), S))^2. \quad \square$$

The next two lemmas are necessary for obtaining the asymptotic distribution of Z* under the alternatives.

LEMMA 4.3. Let h be any function, non-linear in the integers satisfying (4.4) and suppose that the sequence of alternatives (4.2) satisfies Assumption B*. Then with probability one,

$$K_\nu(\underset{\sim}{D}) = E(exp(\frac{it}{\sqrt{N}} \sum_{1}^{N} [h(S_k^{(m)}) - \mu(\underset{\sim}{D})]) | \underset{\sim}{D}) \to exp(\frac{-dt^2}{2})$$

for all real t, where

$$(4.8) \qquad d = \sum_{-m+1}^{m-1} E(Cov(h(\xi_1^{(m)}), h(\xi_{1+k}^{(m)}) | W_1^{(j)}, W_{1+j}^{(m-j)}, W_{1+m}^{(j)})$$

$$- \rho[E(\sum_{j=0}^{\infty} h(j)(j - \frac{S}{\rho})\pi_j(\frac{S}{\rho}))]^2.$$

Proof: Recall the multinomial representation (1.8). Applying again Theorem 2 of Holst [7], (p. 553), with (X_k, Y_k) of that theorem replaced by Poisson r.v.s, we obtain a result similar to our Theorem 2.1 (see also Theorem 2.1 of Holst and Rao [8]). Thus we need to establish the required joint asymptotic normality, calculate A_q and B_q and show that

$$d = A_1 - B_1^2.$$

Let $\{Y_i\}$ be independent Poisson r.v.s with

$$Y_i \sim \text{Poi}(nD_k).$$

The joint asymptotic normality of

$$N^{-\frac{1}{2}} \sum_{1}^{M-m} h(Y_k^{(m)}) \quad \text{and} \quad N^{-\frac{1}{2}} \sum_{1}^{M} Y_k$$

is easily established by verifying the Liapunov condition (2.1) for the (m - 1) dependent r.v.s

$$Y_k = [a \cdot h(Y_k^{(m)}) + bY_k]$$

for real numbers a and b. Now,

$$A_q = \lim_{N\to\infty} N^{-1} \sum_{j=-m+1}^{m-1} \sum_{k=1}^{M-m} \text{Cov}(h(Y_k^{(m)}), h(Y_{k+j}^{(m)})).$$

Using Taylor expansions on the individual terms of this sum as in the proof of Lemmas 3.1 and 3.2 and taking the limit, we obtain:

$$A_q = \sum_{j=-m+1}^{m-1} (qE(\text{cov}(h(\xi_1^{(m)}), h(\xi_{1+j}^{(m)}) \mid W_1^{(j)}, W_{1+j}^{(m-j)}, W_{1+m}^{(j)}))),$$

and

$$B_q = \sqrt{\rho}q \cdot E(\sum_{j=0}^{\infty} h(j)(j - S/\rho)\pi_j(S/\rho)),$$

which tend to:

$$A_1 = \sum_{-m+1}^{m-1} E(\mathrm{Cov}(h(\xi_1^{(m)}), h(\xi_{1+j}^{(m)}) | W_1^{(j)}, W_{1+j}^{(m-j)}, W_{1+m}^{(j)}))$$

and

$$B_1 = \sqrt{\rho} E(\sum_{j=0}^{\infty} h(j)(j - \frac{S}{\rho})\pi_j(\frac{S}{\rho}))$$

as $q \to 1^-$. From this it can be checked that

$$d = A_1 - B_1^2$$

is as in (4.8).

LEMMA 4.4.

$$(4.9) \qquad c + d = \sum_{j=-m+1}^{m-1} \mathrm{Cov}(h(\xi_1^{(m)}), h(\xi_{1+j}^{(m)})) - \frac{\rho^2}{1 + \rho} [\mathrm{Cov}(h(\eta), \eta)]^2.$$

Proof: From equations (4.6) and (4.8) above, we have

$$(4.10) \qquad c + d = \sum_{-m+1}^{m-1} \mathrm{Cov}(\sum_{j=0}^{\infty} h(j)\pi_j(\frac{W_1}{\rho}), \sum_{j=0}^{\infty} h(j)\pi_j(\frac{W_{1+k}}{\rho}))$$

$$- (\mathrm{Cov}(\sum_{j=0}^{\infty} h(j)\pi_j(\frac{S}{\rho}), S))^2$$

$$+ \sum_{k=-m+1}^{m-1} E(\mathrm{Cov}(h(\xi_1^{(m)}), h(\xi_{1+k}^{(m)}) | W_1^{(j)}, W_{1+k}^{(m-j)}, W_{1+m}^{(j)}))$$

$$- \rho[E(\sum_{j=0}^{\infty} h(j)(j - \frac{S}{\rho})\pi_j(\frac{S}{\rho})))]^2.$$

The first and third terms in (4.10), both summations on k from -m+1 to m-1 can be combined as

$$\sum_{k=-m+1}^{m-1} \mathrm{Cov}(h(\xi_1^{(m)}), h(\xi_{1+k}^{(m)})).$$

The second term of (4.10) is

$$[\mathrm{Cov}(\sum_{j=0}^{\infty} h(j)\pi_j(\frac{S}{\rho}), S)]^2$$

$$= \{\int_0^{\infty} \sum_{j=0}^{\infty} h(j)\pi_j(\frac{S}{\rho})(s - m)\frac{e^{-j}s^{m-1}}{(m-1)!}ds\}^2$$

$$= \{\frac{\rho}{1+\rho}\sum_{j=0}^{\infty} h(j)\binom{m+j-1}{j}(\frac{1}{1+\rho})^j(\frac{\rho}{1+\rho})^m[j - \frac{m}{\rho}]\}^2$$

$$= \frac{\rho^2}{(1+\rho)^2}[\mathrm{Cov}(h(\eta),\eta)]^2,$$

where η has the negative binomial distribution, (1.4). Similarly it can be shown that the last term in (4.10)

$$\rho[E(\sum_{j=0}^{\infty} h(j)(j - \frac{S}{\rho})\pi_j(\frac{S}{\rho}))]^2 = \frac{\rho^3}{(1+\rho)^2}[\mathrm{Cov}(h(\eta),\eta)]^2.$$

Combining these results one obtains equation (4.9). □

Now we can prove the major result of this section.

THEOREM 4.2. Suppose that the function $h(\cdot)$, non-linear in the integers, satisfies (4.4), and that the sequence of alternatives (4.2) satisfies B*. Then

$$\frac{1}{\sqrt{N}}\sum_{k=1}^{N}(h(S_k^{(m)}) - Eh(\eta)) \xrightarrow{\mathcal{D}} N(\mu,\sigma^2),$$

where

$$\mu = (\int_0^1 \ell^2(u)du)\mathrm{Cov}(h(\eta), \eta^2 - \eta - \frac{2(m+1)\eta}{\rho})/2(\frac{\rho}{1+\rho})^2,$$

and

$$\sigma^2 = \sum_{j=-m+1}^{m-1} \mathrm{Cov}(h(\xi_1^{(m)}), h(\xi_{1+j}^{(m)})) - \frac{\rho^2}{(1+\rho)} [\mathrm{Cov}(h(\eta),\eta)]^2$$

Proof: Using conditional expectation, we may write the characteristic function

$$E[\exp(\frac{it}{\sqrt{N}} \sum_{k=1}^{N} (h(S_k^{(m)}) - E(h(\eta))))]$$

$$= E_{\underset{\sim}{D}} E(\exp(\frac{it}{\sqrt{N}} \sum_{k=1}^{N} (h(S_k^{(m)}) - Eh(\eta)))\,|\,\underset{\sim}{D})$$

$$= E_{\underset{\sim}{D}}(J_\nu(\underset{\sim}{D}) K_\nu(\underset{\sim}{D})),$$

where

$$J_\nu(\underset{\sim}{D}) = \exp(\frac{it}{\sqrt{N}} [\mu_\nu(\underset{\sim}{D}) - E\mu_\nu]),$$

and

$$K_\nu(\underset{\sim}{D}) = E(\exp(\frac{it}{\sqrt{N}} \sum_{k=1}^{N} [h(S_k^{(m)}) - \mu_\nu(\underset{\sim}{D})])\,|\,\underset{\sim}{D}).$$

From Corollary 4.2 it follows that

$$\frac{1}{\sqrt{N}} [\mu_\nu(\underset{\sim}{D}) - E(\mu_\nu)] \overset{\mathcal{D}}{\to} N(b,c)$$

with b and c defined as in (4.6) and (4.7), respectively. So $J_\nu(\underset{\sim}{D})$ converges in distribution. By Lemma 4.3,

$$K_\nu(\underset{\sim}{D}) \to \exp(-(\frac{dt^2}{2}))$$

with probability one, where d is as given in equation (4.8).

Combining these two results, the product $J_\nu(\underset{\sim}{D})K_\nu(\underset{\sim}{D})$ is seen to converge in distribution with probability one. Since

$$\left|J_\nu(\underset{\sim}{D})K_\nu(\underset{\sim}{D})\right| \leqslant 1,$$

this also implies convergence of the moments, so

$$E_{\underset{\sim}{D}}(J_\nu(\underset{\sim}{D})K_\nu(\underset{\sim}{D})) \to \exp(ibt - \frac{(c + d)t^2}{2})$$

By the continuity theorem, a straightfoward calculation, and Lemma 4.4, the result now follows. □

Next, we will find the asymptotically most powerful test of the form Z^*_ν; i.e., the one with the maximum efficacy against a specific sequence of the alternatives (4.2) that satisfies Assumption B*. For tests of the form:

$$\text{Reject } H_0 \text{ for } \frac{1}{\sqrt{N}} \sum_{k=1}^{N} h(S_k^{(m)}) > C,$$

we want to find the function h(·) which maximizes

$$(4.11) \qquad e_h = \frac{(\int_0^1 \ell^2(u)du)\,\mathrm{Cov}(h(\eta), \eta^2 - \eta - \frac{2(m+1)\eta}{\rho})/2}{[\sum_{k=-m+1}^{m-1} \mathrm{Cov}(h(\xi_1^{(m)}), h(\xi_{1+k}^{(m)})) - \frac{\rho^2}{1+\rho}[\mathrm{Cov}(h(\eta), \eta)]^2]^{\frac{1}{2}}} \times \left(\frac{\rho}{1+\rho}\right)^2$$

The following result is established by methods similar to those of Lemma 3.3.

LEMMA 4.5. The value of e_h in (4.11) is maximized by taking

$$h(x) = x^2,$$

with resulting maximum value,

$$\frac{\left(\int_0^1 \ell^2(u)\,du\right)}{(1+\rho)} \sqrt{\frac{3m(m+1)}{4m+2}} \,. \; \square$$

Combining the result of Theorem 4.2 with Lemma 4.5 we obtain

THEOREM 4.3. The asymptotically locally most powerful test of the null hypothesis (1.1) against the sequence of alternatives (4.2) satisfying Assumption B*, is to reject H_0 when

$$T^* = \sum_{k=1}^{N} (S_k^{(m)})^2 > C, \tag{4.12}$$

where C is a constant determined by the significance level α. The asymptotic distribution of T* is given by

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{N} \left[(S_k^{(m)})^2 - \frac{m(m+1+\rho)}{\rho^2} \right] \xrightarrow{\mathcal{D}} N(0,\sigma^2) \tag{4.13}$$

under H_0, with

$$\sigma^2 = \frac{2m(1+\rho)^2(2m+1)(m+1)}{3\rho^4}, \tag{4.14}$$

while under the alternatives (4.2),

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{N} \left[(S_k^{(m)})^2 - \frac{m(m+1+\rho)}{\rho^2} \right] \xrightarrow{\mathcal{D}} N\left(\left(\int_0^1 \ell^2(u)\,du\right) \frac{m(m+1)}{\rho^2}, \sigma^2\right). \; \square \tag{4.15}$$

REMARK 4.1. It is important to note that the optimal statistic, T*, is independent of the particular sequence of alternatives chosen, but that its power is not.

REMARK 4.2. From the equation

$$e_h = \frac{\left(\int_0^1 \ell^2(u)\,du\right)}{1+\rho} \left[\frac{3m}{4} + \frac{3}{8} - \frac{3}{16m} + \frac{3}{32m^2} + o\left(\frac{1}{m^2}\right) \right]^{\frac{1}{2}}$$

for the statistic T*, it is clear that increasing m increases the efficacy, as does decreasing

$$\rho = \lim_{\nu \to \infty} \frac{N_\nu}{n_\nu} .$$

For m > 1, these tests have higher asymptotic efficiency compared to Dixon's test which corresponds to m = 1 (cf. Dixon, [4]).

REMARK 4.3. For given significance level α, let z_α be such that

$$\Phi(z_\alpha) = 1 - \alpha,$$

where

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

is the standard normal distribution function. Then the AMP test of level α is given explicitly by:

$$\text{Reject } H_0 \text{ if } \quad Z_N^* = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} [(S_k^{(m)})^2 - \frac{m(m+1+\rho)}{\rho^2}] > C_\alpha$$

where

$$C_\alpha = z_\alpha [\frac{2m(2m+1)(m+1)(1+\rho)^2}{3\rho^4}]^{\frac{1}{2}} .$$

The asymptotic power of this test is seen to be

$$P_{H_1}\{Z_N^* > C_\alpha\} = \Phi(z_\alpha + (\int_0^1 \ell^2(u)du)[\frac{3m(m+1)}{4m+2}]^{\frac{1}{2}}/(1+\rho)).$$

REFERENCES

[1] Blumenthal, S., The Asymptotic Normality of Two Test Statistics Associated with the Two-Sample Problem. Ann. Math. Statist. 34 (1963) 1513-1523.

[2] Cressie, N., On the Logarithms of High-Order Spacings. Biometrika 63 (1977) 343-355.

[3] DelPiño, G.E., On the Asymptotic Distribution of k-Spacings with Applications to Goodness of Fit Tests. Ann. Statist. 7 (1979) 1058-1066.

[4] Dixon, W.J., A Criterion for Testing the Hypothesis that Two Samples Are from the Same Population. Ann. Math. Statist. 11 (1940) 199-204.

[5] Fraser, D.A.S., Nonparametric Methods in Statistics (Wiley, New York, 1957).

[6] Godambe, V.P., On the Two Sample Problem: A Heuristic Method for Constructing Tests. Ann. Math. Statist. 32 (1961) 1091-1107.

[7] Holst, L., Two Conditional Limit Theorems with Applications. Ann. Statist. 7 (1979) 551-557.

[8] Holst, L. and Rao, J.S., Asymptotic Two-Sample Theory Based on Spacing-Frequencies. Sankhyā, Ser. A 42 (1980) 19-52.

[9] Holst, L. and Rao, J.S., Asymptotic Spacings Theory with Applications to the Two-Sample Problem. Can. J. Statist. 9 No.1 (1981).

[10] Kuo, M. and Rao, J.S., Limit Theory and Efficiencies for Tests Based on Higher Order Spacings, in Statistics: Applications and New Directions, Proceedings of the Golden Jubilee Conference of the Indian Statistical Institute (Statistical Publishing Society, Calcutta, 1982) (to appear).

[11] LeCam, L., Une théorème sur la division d'un intervalle par des points au hasard. Publ. Inst. Statist. Univ. Paris 7 (1958) 7-16.

[12] Orey, S., A Central Limit Theorem for m-Dependent Random Variables. Duke Math. J. 25 (1958) 543-546.

[13] Pyke, R., Spacings. J. Roy. Statist. Soc. Ser. B 7 (1965) 395-449.

[14] Rao, C.R., Linear Statistical Inference and Its Applications (Second Edition) (Wiley, New York, 1973).

[15] Rao, J.S., Some Tests Based on Arc-Length for the Circle. Sankhyā Ser. B 38 (1976) 329-338.

[16] Rao, J.S. and Mardia, K.V., Pitman Efficiencies of Some Two-Sample Nonparametric Tests, in Recent Developments in Statistical Inference and Data Analysis (North-Holland, New York, 1980) 247-254.

[17] Rao, J.S. and Sethuraman, J.S., Weak Convergence of Empirical Distribution Functions of Random Variables Subject to Perturbations and Scale Factors. Ann. Statist. 3 (1975) 299-313.