STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

Report S 332

Asymptotic normality in nonparametric methods: part I.

by

Kumar Jogdeo

August 1964

Printed at the Mathematical Centre at Amsterdam,49,2nd Boerhaavestraat.
The Netherlands.

The Mathematical Centre, founded the 11th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Scientific Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

1. INTRODUCTION AND SUMMARY

In the earlier usage of nonparametric tests, the main consideration was given to the fact that the level of significance is preserved even if the assumptions regarding the form of the distribution function were violated. Later, however it was pointed out in several papers of Hodges and Lehmann (1956, 1961), Chernoff and Savage (1958), and others that, contrary to the belief that the nonparametric test looses power by wasting information, it has better efficiency behaviour than the classical tests, asymptotically, at least. The study of finite sample size local efficiency (see J. Klotz (1962)) in fact strengthens this claim further.

The basic tool for studying the asymptotic relative efficiency is Pitman's theorem (see Noether (1954)). However, the fundamental requirement for using this tool is asymptotic normality of the test statistic in the neighbourhood of the hypothesis. The nonparametric test statistics, being functions of ranks which are dependent random variables, the usual central limit theorems cannot be applied directly. In order to remove this difficulty various authors have studied the asymptotic distributions of the nonparametric statistics arising in different situations. Among these, the first important theorem is due to Hoeffding (1948), who proves asymptotic normality of a U-statistic. However, this theorem was not applicable to the rank-score test statistics and a theorem due to Chernoff and Savage (1958) enlarged the class of asymptotically normal nonparametric statistics.

The basic motivation for the latter theorem comes from the fact that, when the sample size becomes large, the dependence between ranks of sample observations, say X_i and X_j, $i \neq j$, is weakened, and, if one is able to separate the independent component from the statistic, then the remainder could be shown to go to zero in probability.

Unfortunately, the particular approach used in the paper of Chernoff and Savage (1958) is not suitable for generalizations or widening the class of asymptotically normal rankscore statistics, the main reason being that the number of higher order terms increases.

Also, for applying the theorem, one has to check a number of regularity conditions. This can be seen from the extensions of the results of Chernoff-Savage (1958) made by Puri (1964) and Bhuchongkul (1964).

A new approach for studying the asymptotic distribution was given by Hájek (1961, 62). In the present paper the same idea is used, namely, the following.

Let $U_1,\ldots,U_N$ be independent identically distributed $R(0,1)$ random variables and let $R_1,\ldots,R_N$ be their respective ranks. Then, the basic result of this paper can be stated briefly as follows. Let $a(\lambda_1,\ldots,\lambda_m)$ be a real valued function with m arguments defined on $(0,1)^m$. Then under certain mild conditions on the function $a(\ldots)$ and the coefficients $b_{\alpha_1,\ldots,\alpha_m}$ it is shown that the statistic

1.1 $$\sum_{\pi} b_{\alpha_1,\ldots,\alpha_m} a\left(\frac{R_{\alpha_1}}{N+1},\ldots,\frac{R_{\alpha_m}}{N+1}\right)$$

has asymptotic normal distribution as $N\to\infty$. Here m is fixed but arbitrary, and $\sum_{\pi}$ denotes the sum over all ordered m-tuples from N.

This is done in three steps in three different sections.

Section I is devoted to inequalities which give suitable upper bounds for the expected value of the square

1.2 $$\left[a\left(\frac{R_{\alpha_1}}{N+1},\ldots,\frac{R_{\alpha_m}}{N+1}\right)-a(U_{\alpha_1},\ldots,U_{\alpha_m})\right]^2 .$$

In section II, it is shown that under certain conditions the statistic obtained by replacing arguments of $a(\ldots)$ in (1.1) by independent observations,

1.3 $$\sum_{\pi} b_{\alpha_1,\ldots,\alpha_m} \quad a(U_{\alpha_1},\ldots,U_{\alpha_m}),$$

is asymptotically equivalent to (1.1).

Although this reduction to (1.3) gives summands having independent components, the summands themselves are not independent.

The final reduction is obtained by taking conditional expectations and then imposing conditions which would guarantee the dominance of the leading term having independent summands. The asymptotic normality then follows from the well known uniform asymptotic negligibility considerations.

A very similar approach can be used for studying the limiting distributions of the statistics of type (1.1), but now involving ranks from more than one sample, where ranking is done separately within samples. Another possibility is that some of the arguments of a(...) in (1.1) are actual observations while others are ranks. This case is also covered in view of the inequality II given in section I.

In part II many problems in the testing of hypotheses are considered, to show that the asymptotic normality holds also in the neighbourhood of the null hypothesis H_o: that the observations are independent and identically distributed.

2. THREE INEQUALITIES

The first step in our approach is to show the equivalence between a class of nonparametric statistics and a corresponding class of statistics composed of independent identically distributed random variables. This will be achieved by three extensions of an inequality due to Hájek (1961, lemma 2.1).

First we prove a lemma to be used later.

Lemma 2.1. Let $\{X_i\}$, $i=1,\ldots,m$, be binomial random variables $B(n,p_i)$, not necessarily independent. Then there exists a constant $K(m)$ depending upon m such that

2.1 $$E|X_1-np_1||X_2-np_2|\cdots|X_m-np_m| \le K(m)\left[\prod_{i=1}^{m}(np_iq_i)\right]^{\frac{1}{2}}$$

where $q_i=1-p_i$.

Proof : Note that if X is a binomial random variable $B(n,p)$ then the central moments of X satisfy the following recurrence relation (see Kendall (1947 vol. I, p. 118):

2.2 $$\mu_{r+1}(X) = pq\left[nr\mu_{r-1}(X) + \frac{d}{dp}\mu_r(X)\right],$$

2.3 $$\mu_{r+1}\left(\frac{X}{n^{\frac{1}{2}}}\right) = pq\left[r\mu_{r-1}\left(\frac{X}{n^{\frac{1}{2}}}\right) + \frac{1}{n^{\frac{1}{2}}}\frac{d}{dp}\mu_r\left(\frac{X}{n^{\frac{1}{2}}}\right)\right].$$

Since

2.4 $$\mu_1\left(\frac{X}{n^{\frac{1}{2}}}\right) = 0, \qquad \mu_2\left(\frac{X}{n^{\frac{1}{2}}}\right) = pq,$$

it can be easily seen that $\frac{1}{(pq)^{\frac{1}{2}}}\mu_r\left(\frac{X}{n^{\frac{1}{2}}}\right)$ is bounded uniformly in n by a constant depending upon r only, and hence

2.5 $$E^{1/r}\left[\frac{X-np}{(npq)^{\frac{1}{2}}}\right]^r \le K(r),$$

or equivalently,

2.6 $$E^{1/r}\left[X-np\right]^r \le K(r) \quad (npq)^{\frac{1}{2}}.$$

Applying Hölder's inequality to the left side of (2.1) and using (2.6) the lemma follows immediately and the proof is terminated.

Define a function ε on an m dimensional cube $(-1,1)^m$:

2.7 $$\varepsilon(x_1,\dots,x_m) = \begin{cases} 1 \text{ if } x_i > 0, \text{ for } i=1,\dots,m; \\ 0 \text{ otherwise.} \end{cases}$$

The following lemma is useful for later applications.

Lemma 2.2 For any real numbers $0 \le Z_{i_1},\dots, Z_{i_m}, j_1/N,\dots,j_m/N, k_1/N,\dots,k_m/N, i_1/N,\dots,i_m/N \le 1$,

$$\{\varepsilon(Z_{i_1}-\frac{j_1}{N},\dots,Z_{i_m}-\frac{j_m}{N}) - \varepsilon(\frac{i_1-j_1}{N},\dots,\frac{i_m-j_m}{N})\}$$
$$\times\{\varepsilon(Z_{i_1}-\frac{k_1}{N},\dots,Z_{i_m}-\frac{k_m}{N}) - \varepsilon(\frac{i_1-k_1}{N},\dots,\frac{i_m-k_m}{N})\}$$
$$\le \{\varepsilon(Z_{i_1}-\frac{\max(j_1,k_1)}{N},\dots,Z_{i_m}-\frac{\max(j_m,k_m)}{N})$$
$$-\varepsilon(\frac{i_1-\max(j_1,k_1)}{N},\dots,\frac{i_m-\max(j_m,k_m)}{N})\}^2.$$

Proof: It suffices to prove that when the right side vanishes the left side is not +1. The right side vanishes only in two ways.

1) Both ε terms in the square term are zero in which case one of the first ε terms and one of the last ε terms in the two factors on the left must vanish. However this implies that the left side cannot be +1.
2) Both ε terms on the right side are 1 in which case both the factors on the left zero are.

This completes the proof.

Now, let $U_1,\dots,U_N$ be independent random variables all having rectangular distribution $R(0,1)$. Let R_i, $i=1,\dots,N$, be the ranks of the U_i. Let the order statistic be denoted by $Z_1 < Z_2 < \dots < Z_N$ and thus

2.9 $$U_i = Z_{R_i}.$$

Definition: A collection of N^2 numbers a_{ij} is said to possess Δ-monotonicity if

2.10 $$\Delta_{ij} = (a_{i+1,j+1} - a_{i+1,j} - a_{i,j+1} + a_{i,j}) \geq 0 \text{ for all } (i,j),$$

or

$$\Delta_{ij} \leq 0 \quad \text{for all } (i,j).$$

Consider the function $a(\lambda,\Theta)$ defined on the unit square such that

2.11 $$a(\lambda,\Theta) = a_{i,j} \text{ for } \frac{i-1}{N} < \lambda < \frac{i}{N} \text{ and } \frac{j-1}{N} < \Theta < \frac{j}{N}.$$

Let

2.12 $$a.. = \frac{1}{N^2} \sum_i \sum_j a_{ij} = \int_0^1 \int_0^1 a(\lambda,\Theta)\, d\lambda\, d\Theta,$$

$$\sigma^2 = \frac{1}{N^2} \sum_i \sum_j (a_{ij} - a..)^2 = \int_0^1 \int_0^1 \left[a(\lambda,\Theta) - a..\right]^2 d\lambda\, d\Theta,$$

2.13 $$a_{i\cdot} = \frac{1}{N} \sum_{j=1}^{N} a_{ij},\quad a_{\cdot j} = \frac{1}{N} \sum_{i=1}^{N} a_{ij}.$$

Inequality I. With the above notation if

a) the numbers a_{ij} are Δ-monotone and either

b_1) the sequences $\{a_{i1}\}$, $\{a_{1i}\}$ are monotone in i,

or

b_2) the sequences $\{a_{iN}\}$, $\{a_{Ni}\}$ are monotone in i, then

2.14 $$E\left[a(U_1,U_2) - a\left(\frac{R_1}{N}, \frac{R_2}{N}\right)\right]^2 < k_1 \frac{\max(a_{ij}-a_{..})^2}{(N-1)^{\frac{1}{2}}},$$

<u>where k_1 is a positive constant.</u>

<u>Proof</u>: With $Z_1 < \ldots < Z_N$ fixed, the pair (U_1,U_2) takes $N(N-1)$ values (Z_i,Z_j) with equal probabilities. Thus,

2.15 $$E\left[a(U_1,U_2) - a\left(\frac{R_1}{N}, \frac{R_2}{N}\right)\right]^2$$

$$= \frac{1}{N(N-1)} \quad E \sum_{i \neq j}\sum \left[a(Z_i,Z_j)-a\left(\frac{i}{N}, \frac{j}{N}\right)\right]^2 .$$

Consider the special case of the elementary function ε defined by (2.7),

2.16 $$\varepsilon\left(\lambda - \frac{k}{N}, \theta - \frac{\ell}{N}\right) = \begin{cases} 1 \text{ if } \lambda > \frac{k}{N} \text{ and } \theta > \frac{\ell}{M}, \\ 0 \text{ otherwise.} \end{cases}$$

For given $Z_1 < \ldots < Z_N$ let K and L denote the number of Z_i less than k/N and the number of Z_i less than ℓ/N respectively.
If $K \leq k$ and $L \leq \ell$, then it is obvious that

2.17 $$\varepsilon\left[Z_i - \frac{k}{N}, Z_j - \frac{\ell}{N}\right] - \varepsilon\left[\frac{i-k}{N}, \frac{j-\ell}{N}\right] = \begin{cases} 1 \text{ if } K < i,\ L < j \text{ and either } i < k \text{ or } j \leq \ell, \\ 0 \text{ otherwise.} \end{cases}$$

In general, for any values of K and L it is seen that there are at most $|K-k|\,|L-\ell|+(N-k)\,|L-\ell|+(N-\ell)\,|K-k|$ pairs of (i,j) for which the difference (2.17) is ± 1. Hence,

2.18 $$\sum_i \sum_j \left\{\varepsilon\left[Z_i - \frac{k}{N}, Z_j - \frac{\ell}{N}\right] - \varepsilon\left[\frac{i-k}{N}, \frac{j-\ell}{N}\right]\right\}^2$$

$$\leq |K-k|\,|L-\ell| + (N-k)\,|L-\ell| + (N-\ell)\,|K-k| .$$

Since K and L are binomial random variables,

$$2.19 \quad E\left[\varepsilon(U_1 - \frac{k}{N},\ U_2 - \frac{\ell}{N}) - \varepsilon(\frac{R_1-k}{N} - \frac{R_2-\ell}{N})\right]^2$$

$$\leq \frac{1}{N(N-1)}\left\{E|K-k|\,|L-\ell| + (N-k)E|L-\ell| + (N-\ell)E|K-k|\right\}$$

$$\leq \frac{1}{N(N-1)}\left\{\left[k\,(1-\frac{k}{N})(1-\frac{}{N})\right]^{\frac{1}{2}} + (N-k)\left[\ell(1-\frac{\ell}{N})\right]^{\frac{1}{2}} + (N-\ell)\left[k(1-\frac{k}{N})\right]^{\frac{1}{2}}\right.$$

$$\leq \frac{3}{(N-1)N^{\frac{1}{2}}}\left[N-k\right]^{\frac{1}{2}}\left[N-\ell\right]^{\frac{1}{2}}.$$

With the help of this inequality and the relation between $a(\lambda,\Theta)$ and $\varepsilon(\lambda,\Theta)$ to be stated below the required inequality will follow after some computation. Recalling the definition of Δ_{ij} (see 2.10) it is seen that if,

$$2.20 \qquad b_{ij} = a_{ij} - a_{i1} - a_{1i} + a_{11},$$

then

$$2.21 \qquad b_{NN} = a_{NN} - a_{N1} + a_{1N} + a_{11} = \sum_{k=1}^{N-1}\ \sum_{\ell=1}^{N-1} \Delta_{k\ell};$$

$$b_{NN}^2 = \sum_k \sum_\ell \sum_m \sum_n \Delta_k\ \Delta_{mn}.$$

In general $b(\lambda,\Theta)$ can be expressed as

$$2.22 \qquad b(\lambda,\Theta) = \sum \sum \Delta_{k\ell} \qquad \varepsilon(\lambda - \frac{k}{N},\ \Theta - \frac{\ell}{N}),$$

and hence

$$2.23 \qquad \sum_i \sum_j b_{ij}^2 = \sum_k \sum_\ell \sum_m \sum_n \Delta_{k\ell}\,\Delta_{mn} \sum_i \sum_j \varepsilon(\frac{i-k}{N},\ \frac{j-\ell}{N})\ \varepsilon(\frac{i-m}{N},\ \frac{j-n}{N}).$$

Since

2.24 $$\varepsilon\left(\frac{i-k}{N}, \frac{j-\ell}{N}\right) \varepsilon\left(\frac{i-m}{N}, \frac{j-n}{N}\right) = \begin{cases} 1 & \text{for } i > \max(k,m),\ j > \max(\ell,n) \\ 0 & \text{otherwise,} \end{cases}$$

for fixed (k,ℓ) and (m,n) the number of pairs (i,j) such that the left side of (2.24) is unity, equals $[N-\max(k,m)][N-\max(\ell,n)]$. Hence

2.25 $$\sum_i \sum_j b_{ij}^2 = \sum_k \sum_\ell \sum_m \sum_n \Delta_k \Delta_{mn}[N-\max(k,m)][N-\max(\ell,n)].$$

Using these expressions it is seen that

2.26 $$E\left[b(U_1,U_2) - b\left(\frac{R_1}{N}, \frac{R_2}{N}\right)\right]^2$$

$$= \frac{1}{N(N-1)} E \sum_{i \neq j}\sum \left[b(Z_i,Z_j) - b\left(\frac{i}{N}, \frac{j}{N}\right)\right]^2$$

$$= \frac{1}{N(N-1)} \sum_k \sum_\ell \sum_m \sum_n \Delta_{k\ell} \Delta_{mn}$$

$$\times\ E \sum_{i \neq j}\sum \left\{\varepsilon\left(Z_i - \frac{k}{N}, Z_j - \frac{\ell}{N}\right) - \varepsilon\left(\frac{i-k}{N}, \frac{j-\ell}{N}\right)\right\}$$

$$\times \left\{\varepsilon\left(Z_i - \frac{m}{N}, Z_j - \frac{n}{N}\right) - \varepsilon\left(\frac{i-m}{N}, \frac{j-n}{N}\right)\right\}.$$

Applying lemma 2.2 and the equation (2.19) it follows that

2.27 $$E \sum_{i \neq j}\sum \left\{\varepsilon\left(Z_i - \frac{k}{N}, Z_j - \frac{\ell}{N}\right) - \varepsilon\left(\frac{i-k}{N}, \frac{j-\ell}{N}\right)\right\}$$

$$\times \left\{\varepsilon\left(Z_i - \frac{m}{N}, Z_j - \frac{n}{N}\right) - \varepsilon\left(\frac{i-m}{N}, \frac{j-n}{N}\right)\right\}$$

$$\leq E \sum_i \sum_j \left\{\varepsilon\left(Z_i - \frac{\max(k,m)}{N}, Z_j - \frac{\max(\ell, n)}{N}\right)\right.$$

$$\left. - \varepsilon\left(\frac{i-\max(k,m)}{N}, \frac{j-\max(\ell,n)}{N}\right)\right\}^2$$

$$\leq N(N-1)\ E\left[\varepsilon\left(U_1 - \frac{\max(k,m)}{N}, U_2 - \frac{\max(\ell, n)}{N}\right)\right.$$

$$- \varepsilon\left(\frac{R_1 - \max(k,m)}{N}, \frac{R_2 - \max(\ell,n)}{N}\right)\Big]^2$$

$$\leq 3\, N^{\frac{1}{2}}\, [N-\max(k,m)]^{\frac{1}{2}}\, [N-\max(\ell,n)]^{\frac{1}{2}}.$$

Substituting this inequality in (2,26), using (2.21), (2.25) and the fact that $\Delta_{k\ell}\Delta_{mn} \geq 0$ for all k,ℓ,m,n it follows that

2,28 $$E\left[b(U_1,U_2) - b\left(\frac{R_1}{N}, \frac{R_2}{N}\right)\right]^2$$

$$\leq \frac{3}{N^{\frac{1}{2}}(N-1)} \sum\sum\sum\sum \Delta_{k\ell}\,\Delta_{mn}(N-\max(k,m))^{\frac{1}{2}}(N-\max(\ell,n))^{\frac{1}{2}}$$

$$\leq \frac{1}{N(N-1)}\left[\sum\sum\sum\sum \Delta_{k\ell}\,\Delta_{mn}\right]^{\frac{1}{2}}$$

$$\times\left[\sum\sum\sum\sum \Delta_{k\ell}\,\Delta_{mn}(N-\max(k,m))(N-\max(\ell,n))\right]^{\frac{1}{2}}$$

$$\leq \frac{3}{N^{\frac{1}{2}}}\,\frac{|b_{NN}|}{(N-1)}\left[\sum\sum b_{ij}^2\right]^{\frac{1}{2}},$$

which is as same as

2,29 $$E\left[a(U_1,U_2)-a\left(\frac{R_1}{N}, \frac{R_2}{N}\right)-a\left(U_1,\frac{1}{N}\right)+a\left(\frac{R_1}{N}, \frac{1}{N}\right)-a\left(\frac{1}{N}, U_2\right)+a\left(\frac{1}{N}, \frac{R_2}{N}\right)\right]^2$$

$$\leq \frac{3}{N^{\frac{1}{2}}}\,\frac{|a_{NN}-a_{N1}-a_{1N}+a_{11}|}{(N-1)}\left[\sum_i\sum_j (a_{ij}-a_{i1}-a_{1i}+a_{11})^2\right]^{\frac{1}{2}}$$

$$\leq \frac{k_1 \max(a_{ij}-a_{..})^2}{(N-1)^{\frac{1}{2}}}.$$

However,

$$2.30 \quad E\left[a(U_1,U_2) - a(\frac{R_1}{N}, \frac{R_2}{N})\right]^2$$

$$\leq 3E\left[a(U_1,U_2)-(\frac{R_1}{N}, \frac{R_2}{N})-a(U_1, \frac{1}{N})+a(\frac{R_1}{N}, \frac{1}{N})-a(\frac{1}{N}, U_2)+a(\frac{1}{N}, \frac{R_2}{N})\right]^2$$

$$+ 3E\left[a(U_1, \frac{1}{N})-a(\frac{R_1}{N}, \frac{1}{N})\right]^2 + 3E\left[a(\frac{1}{N}, U_2)-a(\frac{1}{N}, \frac{R_2}{N})\right]^2 .$$

Using (2.29) and the inequality of Hájek (1961, lemma 1), it follows that

$$2.31 \quad E\left[a(U_1, U_2) - a(\frac{R_1}{N}, \frac{R_2}{N})\right]^2 \leq \frac{k_2 \max(a_{ij}-a_{..})^2}{(N-1)^{\frac{1}{2}}}$$

$$+ \frac{k_3 \max |a_{i1}-a_{.1}|}{N} \left[\sum_i (a_{i1}-a_{.1})^2\right]^{\frac{1}{2}}$$

$$+ \frac{k_4 \max |a_{1j}-a_{1.}|}{N} \left[\sum_j (a_{1j}-a_{1.})^2\right]^{\frac{1}{2}} \leq \frac{k_1 \max(a_{ij}-a_{..})^2}{(N-1)^{\frac{1}{2}}} .$$

This proves the inequality assuming condition (b_1) of the theorem. To prove the inequality under the condition (b_2) put $a'(\lambda,\Theta) =$ $= -a(1-\lambda, 1-\Theta)$ and observe that

$$2.32 \quad E\left[a(U_1,U_2) - a(\frac{R_1}{N}, \frac{R_2}{N})\right]^2 = E\left[a'(U_1,U_2) - a'(\frac{R_1}{N}, \frac{R_2}{N})\right]^2 .$$

If one proceeds with the numbers a'_{ij}, the inequality (2.14) is obtained merely by noting that $a'_{ij} = -a_{N+1-i,N+1-j}$.

Inequality II. With the same notation as above, if

a) the numbers a_{ij} are Δ-monotone and either

b_1) the sequence $\{a_{i1}\}$ is monotone in i

or

b_2) the sequence $\{a_{iN}\}$ is monotone in i then

$$2.33 \quad E\left[a(U_1, \frac{R_2}{N}) - a(\frac{R_1}{N}, \frac{R_2}{N})\right]^2 < \frac{k_5 \max(a_{ij}-a)^2}{(N-1)^{\frac{1}{2}}} .$$

Proof: Defining numbers b_{ij} as in (2.20) it is seen that

$$2.34 \quad E\left[b(U_1, \frac{R_2}{N}) - b(\frac{R_1}{N}, \frac{R_2}{N})\right]^2$$

$$= \frac{1}{N(N-1)} E \sum_{i} \sum_{\neq j} \left[b(Z_i, \frac{j}{N}) - b(\frac{i}{N}, \frac{j}{N})\right]^2 .$$

By the same argument as used in the proof of inequality I it follows that for a fixed pair of integers (k,ℓ) the number of pairs (i,j) such that

$$2.35 \quad \left[\varepsilon(Z_i - \frac{k}{N}, \frac{j-\ell}{N}) - \varepsilon(\frac{i-k}{N}, \frac{j-\ell}{N})\right]^2 = 1$$

is equal to $|K-k|$ $(N-\ell)$ and hence

$$2.36 \quad E \sum_{i} \sum_{j} \left[\varepsilon(Z_i - \frac{k}{N}, \frac{j-\ell}{N}) - \varepsilon(\frac{i-k}{N}, \frac{j-\ell}{N})\right]^2$$

$$= (N-\ell)E|K-k| \leq (N-\ell) \left[k(1 - \frac{k}{N})\right]^{\frac{1}{2}} .$$

Expressing b_{ij} in terms of Δ_{ij} and the function ε (see 2.22),

$$2.37 \quad E\left[b(U_1, \frac{R_2}{N}) - b(\frac{R_1}{N}, \frac{R_2}{N})\right]^2$$

$$\leq \frac{1}{N(N-1)} \sum_{k} \sum_{\ell} \sum_{m} \sum_{n} \Delta_{k\ell} \Delta_{mn} E \sum_{i} \sum_{j} \mu_{ijk\ell mn} ,$$

where

$$2.38 \quad \mu_{ijk\ell mn}$$

$$= \left\{\varepsilon(Z_i - \frac{k}{N}, \frac{j-\ell}{N}) - \varepsilon(\frac{i-k}{N}, \frac{j-\ell}{N})\right\} \left\{\varepsilon(Z_i - \frac{m}{N}, \frac{j-n}{N}) - \varepsilon(\frac{i-m}{N}, \frac{j-n}{N})\right\}$$

$$\leq \left\{\varepsilon(Z_i - \frac{\max(k,m)}{N}, \frac{j-\max(\ell,n)}{N}) - \varepsilon(\frac{i-\max(k,m)}{N}, \frac{j-\max(\ell,n)}{N})\right\}^2 .$$

Using (2.36) it follows that

2.39 $$E\left[b(U_1, \frac{R_2}{N}) - b(\frac{R_1}{N}, \frac{R_2}{N})\right]^2$$

$$\leq \sum_k \sum_\ell \sum_m \sum_n \Delta_{k\ell}\,\Delta_{mn} \left\{\frac{N-\max(\ell,n)}{N(N-1)}\right\}\left\{\frac{\max(k,m)}{N}(N-\max(k,n))\right\}^{\frac{1}{2}} .$$

The right side of (2.39) will be increased if we put $\{\max(k,m)/N\} = 1$ and $\{N-\max(\ell,n)\}^{\frac{1}{2}}\{N\}^{-\frac{1}{2}} = 1$. Hence,

2.40 $$E\left[b(U_1, \frac{R_2}{N}) - b(\frac{R_1}{N}, \frac{R_2}{N})\right]^2$$

$$\leq \frac{1}{N^{\frac{1}{2}}(N-1)} \sum_k \sum_\ell \sum_m \sum_n \Delta_{k\ell}\,\Delta_{mn}\left[\left(N-\max(k,m)\right)\left(N-\max(\ell,n)\right)\right]^{\frac{1}{2}}$$

$$\leq \frac{|b_{NN}|}{N^{\frac{1}{2}}(N-1)}\left[\sum_i \sum_j b_{ij}^2\right]^{\frac{1}{2}} .$$

Expressing the b_{ij} in terms of the a_{ij}, the inequality (2.40) can be written as

2.41 $$E\left[a(U_1, \frac{R_2}{N}) - a(\frac{R_1}{N}, \frac{R_2}{N}) - a(U_1, \frac{1}{N}) + a(\frac{R_1}{N}, \frac{1}{N})\right]^2$$

$$\leq \frac{|a_{NN}-a_{1N}-a_{N1}+a_{11}|}{N^{\frac{1}{2}}(N-1)}\left[\sum\sum (a_{ij}-a_{i1}-a_{1i}+a_{11})^2\right]^{\frac{1}{2}} .$$

Using the same procedure as in the proof in inequality I it follows that

2.42 $$E\left[a(U_1, \frac{R_2}{N}) - a(\frac{R_1}{N}, \frac{R_2}{N})\right]^2$$

$$\leq \frac{k_6 \max(a_{ij}-a_{..})^2}{(N-1)^{\frac{1}{2}}} + \frac{k_7 \max|a_{i1}-a_{.1}|}{N}\left[\sum (a_{i1}-a_{.1})^2\right]^{\frac{1}{2}}$$

$$\leq \frac{k_5 \max(a_{ij}-a_{..})^2}{(N-1)^{\frac{1}{2}}} .$$

Inequality III. Let $U_1, U_2, \ldots, U_N$ and $V_1, \ldots, V_N$ be two sets of independent identically distributed uniform random variables on the unit interval (0,1). The two sets are ranked within themselves, and let $R_1, R_2, \ldots, R_N$ and $S_1, \ldots, S_N$ be their respective ranks. Then, with the same notation and assumptions about a_{ij}, as in Inequality I,

$$2.43 \qquad E\left[a(U_1,V_1) - a\left(\frac{R_1}{N}, \frac{S_1}{N}\right)\right]^2 \leq \frac{k_8 \max(a_{ij} - a_{..})^2}{(N-1)^{\frac{1}{2}}} \quad .$$

Proof (indication). With the b_{ij} defined in (2.20) it is seen that

$$2.44 \qquad E\left[b(U_1,V_1) - b\left(\frac{R_1}{N}, \frac{S_1}{N}\right)\right]^2 = \frac{1}{N^2} E \sum_i \sum_j \left[b(Z_i, W_j) - b\left(\frac{i}{N}, \frac{j}{N}\right)\right]^2 ,$$

where $W_1 < \ldots < W_N$ is the ordered statistic corresponding to $V_1, \ldots, V_N$.

Following exactly the same steps as in the proof of inequality I, (2.43) is obtained.

Remark I. Inequalities I, II, III can be generalized for the a(...) functions having any arbitrary but fixed number of arguments. The proofs are along similar lines and lemma 2.1 is useful for such extensions. Also, these three inequalities can be combined into one; however, in this generalization the notation would be very cumbersome and it would be hard to recognize the essential features of the inequalities.

3. ASYMPTOTICALLY EQUIVALENT STATISTICS

Let $\{X_n\}$ and $\{Y_n\}$ be two sequences defined on (Ω, A_n, P_n). $\{X_n\}$ is said be asymptotically equivalent to $\{Y_n\}$ in the quadratic mean if

3.1 $$\frac{E\left[X_n - Y_n\right]^2}{\text{Var } X_n} \to 0, \text{ as } n \to \infty .$$

It can be seen that this is a true equivalence relation. For the sake of brevity the phrase 'in the quadratic mean' will be omitted, and the asymptotic equivalence will be denoted by $X_n \sim Y_n$.

From the above definition it follows that if $\{X_n\}$ converges to a random variable Z in probability then so does $\{Y_n\}$ and if the asymptotic mean and variance of X_n exist and are finite then the asymptotic mean and variance of Y_n exist and are identical to those of X_n.

Now, let c_{ij} be N^2 real numbers, not all of which are equal and a_{ij} be the numbers defined in section 2. These numbers may change with N; however, for the sake of simplicity in notation this dependence is not explicitly shown.

With the same notation as in section 2 define

$$c.. = \sum_{i \neq j}\sum c_{ij}/N(N-1),$$

$$S_N = \sum_{i \neq j}\sum c_{ij}a_{R_i\, R_j},$$

3.2 $$S_N{}^1 = \sum_{i \neq j}\sum (c_{ij} - c..)\, a(U_i, \frac{R_j}{N}) + c.. \sum_{i \neq j}\sum a_{ij},$$

$$S_N{}^* = \sum_{i \neq j}\sum c_{ij}\, a_{R_i\, S_j},$$

$$T_N = \sum_{i \neq j}\sum (c_{ij} - c..)\, a(U_i,U_j) + c.. \sum_{i \neq j}\sum a_{ij},$$

$$T_N{}^* = \sum_{i \neq j}\sum (c_{ij} - c..)\, a(U_i,V_j) + c.. \sum_{i \neq j}\sum a_{ij}.$$

Theorem 3.1 *With the same notation and assumptions of inequalities* I, II, III *of section* 2, *if*

a) $$\lim_{N \to \infty} \frac{\max(a_{ij} - a..)^4}{N} = 0,$$

b) $$\lim_{N \to \infty} \frac{\sum \sum (a_{ij} - a..)^2}{N(N - 1)} > 0,$$

then

3.3 $$S_N \sim T_N, \; S_N \sim S_N^1, \; S_N^* \sim T_N^*.$$

Proof. The asymptotic equivalence of S_N and T_N is proved here. The other two can be proved in a similar manner and hence are not considered.

An obvious extension of lemma 2.3 of Hájek, (1961) which is useful here, can be given as in the following.

Let $\{c_{ij}\}$ and $\{d_{ij}\}$ be two sets each having N^2 real numbers and let

3.4 $$c.. = \sum_{i} \sum_{\neq j} c_{ij}/N(N-1), \; d.. = \sum_{i} \sum_{\neq j} d_{ij}/N(N-1).$$

Then

3.5 $$\mathrm{Var} \sum \sum c_{ij} \, d_{R_i, R_j}$$

$$= \frac{1}{N(N-1)} \sum \sum (c_{ij} - c..)^2 \quad \sum \sum (d_{ij} - d..)^2$$

$$\leq \frac{1}{N(N-1)} \sum \sum (c_{ij} - c..)^2 \quad \sum \sum d_{ij}^2.$$

Observing that

3.6 $$S_N - T_N = \sum\sum (c_{ij}-c_{..}) \left[a(U_i,U_j) - a(\frac{R_i}{N}, \frac{R_j}{N})\right]$$

$$= \sum\sum (c_{ij}-c_{..}) \left[a(Z_{R_i},Z_{R_j}) - a(\frac{R_i}{N}, \frac{R_j}{N})\right],$$

3.7 $$E\left[S_N - T_N \mid Z_1,\ldots,Z_N\right] = 0,$$

and using (3.5), it follows that

3.8 $$E\left[S_N - T_N \mid Z_1,\ldots,Z_N\right]^2 = \text{Var}\left[S_N - T_N \mid Z_1,\ldots,Z_N\right]$$

$$\leq \frac{1}{N(N-1)} \sum\sum (c_{ij}-c_{..})^2 \sum\sum \left[a(Z_{R_i},Z_{R_j}) - a(\frac{R_i}{N}, \frac{R_j}{N})\right]^2$$

$$\leq \frac{1}{N(N-1)} \sum\sum (c_{ij}-c_{..})^2 \sum\sum \left[a(U_i,U_j) - a(\frac{R_i}{N}, \frac{R_j}{N})\right]^2 .$$

Taking expectations on both sides and using inequality I,

3.9 $$E\left[S_N - T_N\right]^2 \leq \sum\sum (c_{ij}-c_{..})^2 \, E\left[a(U_1,U_2) - a(\frac{R_1}{N}, \frac{R_2}{N})\right]^2$$

$$\leq \sum\sum (c_{ij}-c_{..})^2 \cdot \frac{k_1 \max(a_{ij}-a_{..})^2}{(N-1)^{\frac{1}{2}}} .$$

From (3.5) it is clear that

3.10 $$\text{Var}\, S_N = \frac{1}{N(N-1)} \sum\sum (c_{ij}-c_{..})^2 \sum\sum (a_{ij}-a_{..})^2,$$

and hence using conditions (a) and (b) it follows that

3.11 $$\frac{E\left[S_N - T_N\right]^2}{\text{Var}\, S_n}$$

$$\leq \frac{k_1 \max(a_{ij}-a_{..})^2}{(N-1)^{\frac{1}{2}}} \; \frac{N(N-1)}{\sum\sum (a_{ij}-a_{..})^2} \to 0 \text{ as } N \to \infty .$$

The proof is terminated.

In order to apply theorem 3.1 to various nonparametric statistics it is essential to find a set of sufficient conditions in terms of the distribution functions which will be used for constructing various rank score tests.

The following lemma states that the uniform integrability condition assures the fulfillment of the conditions of theorem 3.1.

Let $\phi(\lambda_1,\ldots,\lambda_m)$ be a real - valued function defined on the unit hypercube $(0,1)^m$ and let ϕ belong to the space L_p, that is:

3.12 $$\int_0^1 \cdots \int_0^1 \left| \phi(\lambda_1,\ldots,\lambda_m)^p \right. d\lambda_1 \ldots d\lambda_m < \infty .$$

In practice, however, a rank score function is defined on the ranks, or equivalently, on N points i/N, i=1,...,N.

This can be constructed from ϕ in several ways. The function ϕ can be expressed in terms of the distribution functions and conditions on ϕ can be transformed to those on the distributions. Before giving the actual construction we shall state conditions which will make these constructions more meaningful.

Let $\phi_N(\lambda_1,\ldots,\lambda_m)$ be a nondecreasing real-valued step function defined on the unit hypercube $(0,1)^m$ such that ϕ_N is constant over open cubes $\prod_{j=1}^{m} (\frac{i_j-1}{N}, \frac{i_j}{N})$.

For the sake of simplicity we consider the case of m=2.

Lemma 3.1. The conditions a) and b) of theorem 3.1 are satisfied with

3.13 $$a_{ij} = \phi_N(\frac{i}{N}, \frac{j}{N}) ,$$

provided

i) ϕ_N converges pointwise to a nonconstant function ϕ which belongs to L_8 and

ii) the functions ϕ_N are uniformly integrable.

Proof. From uniform integrability of ϕ_N^8,

3.14 $$\frac{\max|a_{ij}|^8}{N^2} = \max_{i,j} \int_{\frac{j-1}{N}}^{\frac{j}{N}} \int_{\frac{i-1}{N}}^{\frac{i}{N}} \phi_N^8(\lambda,\theta)\, d\lambda \, d\theta \to 0 \text{ as } N \to \infty .$$

From the nonconstancy of ϕ and L_p convergence

3.15 $$\frac{\sum\sum (a_{ij}-a_{..})^2}{N(N-1)} = \int\int \left[\phi_N(\lambda,\theta) - \bar{\phi}_N\right]^2 d\lambda \, d\theta$$

$$\to \int\int \left[\phi(\lambda,\theta) - \bar{\phi}\right]^2 d\lambda \, d\theta > 0,$$

as $N \to \infty$

where

3.16 $$\bar{\phi}_N = \int\int \phi_N(\lambda,\theta)\, d\lambda \, d\theta, \text{ and } \bar{\phi} = \int\int \phi(\lambda,\theta)\, d\lambda \, d\theta .$$

In the following some constructions are given, in particular, the extension of lemma 2.2 of Hájek (1961).

Let $\phi(\lambda,\theta)$ be a real-valued function defined on the unit square $(0,1)^2$ and let ϕ belong to the space L_p. Thus,

3.17 $$\int_0^1 \int_0^1 \left|\phi(\lambda,\theta)\right|^p d\lambda \, d\theta < \infty .$$

Define

3.18 $$\phi_N(\lambda,\theta) = \phi\left(\frac{i}{N+1}, \frac{j}{N+1}\right) \text{ for } \frac{i-1}{N} < \lambda \leq \frac{i}{N}, \ \frac{j-1}{N} < \theta \leq \frac{j}{N} .$$

Lemma 3.1. With the above notation if ϕ is monotone in λ and θ then,

i) the functions ϕ_N^k are uniformly integrable for $k=1,\ldots,p$, and

ii) $\lim_{N\to\infty} \int_0^1 \int_0^1 \left|\phi_N(\lambda,\theta) - \phi(\lambda,\theta)\right|^k d\lambda \, d\theta = 0$ for $k=1,\ldots,p$.

Proof. It suffices to show that the assertions hold for k=p. First assume that $\phi(0,0) \geq 0$ and that ϕ is monotone nondecreasing.

The uniform integrability of the functions ϕ_N^p will be proved by the sucessive application of an inequality of Hájek (1961, lemma 2.1) and the Fubini theorem.

Consider the function

3.19 $$\xi_N(\lambda,\theta) = \phi\left(\frac{i}{N+1},\theta\right) \quad \text{for} \quad \frac{i-1}{N} \leq \lambda < \frac{i}{N},$$

and an open rectangle $R \subset (0,1)^2$. It is seen from a construction in the above lemma of Hájek that

3.20 $$\int_R\int \xi_N^p(\lambda,\theta) \leq \int_R\int \phi^p(\tfrac{3}{4},\theta)\, d\theta\ \lambda + 4 \iint_{B_1} \phi^p(\lambda,\theta)\, d\lambda\ d\theta$$

where B_1 is a rectangle and the Lebesgue measure $\mu(B_1) = \mu(R)$. Defining now

3.21 $$\phi_N(\lambda,\theta) = \xi_N\left(\lambda, \frac{j}{N+1}\right) \text{ for } \frac{j-1}{N} \leq \theta < \frac{j}{N},$$

and applying the inequality (3.20) to (3.21) it follows that

3.22 $$\int_R\int \phi_N^p(\lambda,\theta)\, d\lambda\ d\theta$$

$$\leq \int_R\int \xi_N^p(\lambda,\tfrac{3}{4})\, d\lambda\ d\theta + 4 \iint_{B_1} \xi_N^p(\lambda,\theta)\, d\lambda\ d\theta$$

$$\leq \mu(R)\, \phi^p(\tfrac{3}{4},\tfrac{3}{4}) + 4 \iint_{B_1} \phi^p(\lambda,\tfrac{3}{4})\, d\lambda\ d\theta$$

$$+ 4 \iint_{B_1} \phi^p(\tfrac{3}{4},\theta)\, d\lambda\ d\theta + 16 \iint_{B_2} \phi^p(\lambda,\theta)\, d\lambda\ d\theta,$$

where B_1, B_2 are rectangles and $\mu(B_2) = \mu(B_1) = \mu(R)$. For the consideration of uniform integrability, the upper bound given in (3.22) for any arbitrary rectangle $R \subset (0,1)^2$ is sufficent and this completes the first assertion.

The second assertion follows from the L_p convergence theorem. To remove the restriction $\phi(0,0) \geq 0$, observe that a function $\phi(\lambda,\theta)$ which is nondecreasing in λ and θ can be expressed as

3.23 $$\phi(\lambda,\theta) = \phi^{+}(\lambda,\theta) - \phi^{*}(1-\lambda, 1-\theta),$$

where $\phi^{+}(\lambda,\theta)$ is the positive part of $\phi(\lambda,\theta)$ and $\phi^{*}(\lambda,\theta)$ is the negative part of the function ϕ^{o} where

3.24 $$\phi^{o}(\lambda,\theta) = \phi(1-\lambda, 1-\theta),$$

and ϕ^{*} is nondecreasing and nonnegative.

Expressing $\phi(\lambda,\theta)$ as in (3.23) it can be seen that the assertions follow for the corresponding ϕ_N functions. Lastly, if a function is monotone nonincreasing the multiplication by -1 gives us the same results. This completes the proof.

Another way of constructing a ϕ_N function from ϕ is:

3.25 $$\phi_N(\lambda,\theta)=N^2 \int_{\frac{i-1}{N}}^{\frac{i}{N}} \int_{\frac{i-1}{N}}^{\frac{j}{N}} \phi(\lambda,\theta)\, d\lambda \; d\theta \; ,$$

$$\text{for } \frac{i-1}{N} < \lambda \leq \frac{i}{N}, \; \frac{j-1}{N} < \theta \leq \frac{j}{N} .$$

It is clear that the functions ϕ_N defined by (3.25) can be replaced in lemma 3.1.

With the help of these ϕ_N functions rank score statistics can be constructed and these can be seen to be equivalent to statistics involving independent uniform random variables.

Following is a typical example of the function ϕ which can be constructed from an absolutely continuous distribution function F, whose first two derivates f and f' exist:

3.26 $$\phi(\lambda) = -\frac{f'[F^{-1}(\lambda)]}{f[F^{-1}(\lambda)]}, \quad 0 < \lambda < 1.$$

The scope of application of the above theory can be widened by the following considerations.

The condition of Δ-monotonicity can be weakened considerably. Suppose the set of numbers $\{a_{ij}\}$ or $\{\phi_N(\frac{i}{N+1}, \frac{j}{N+1})\}$ can be expressed as a linear combination of sets satisfying Δ-monotonicity, say

3.27 $$a_{ij} = a_{ij}^{(1)} + a_{ij}^{(2)} + \dots + a_{ij}^{(k)} \quad \text{for } i,j = 1,\dots,N,$$

where $\{a_{ij}^{(\ell)}\}$, $\ell = 1,\dots,k$; satisfy the Δ-monotonicity condition, but the set $\{a_{ij}\}$ does not. The asymptotic equivalence considered in theorem 3.1 can be proved very easily by expressing the statistics as a linear combination and applying c_r-inequality.

The monotonicity condition of the ϕ function can be weakened by the same consideration of linear combinations as above. As far as application is concerned, the function ϕ should be expressible as a linear combination of a finite number of monotone functions and the set of numbers $\{a_{ij}\}$ satisfying (3.17) as piecewise Δ-monotone.

The above discussion, theorem 3.1, lemma 3.1 and 3.2 lead to the following:

Theorem 3.2 *Let c_{ij} be N^2 numbers not all of which are equal and let*

$$S_N = \sum\sum c_{ij}\, \phi_N\left(\frac{R_i}{N+1}, \frac{R_j}{N+1}\right),$$

$$S_N^1 = \sum\sum (c_{ij} - c_{..})\, \phi_N\left(U_i, \frac{R_j}{N+1}\right) + c_{..} \sum\sum \phi_N\left(\frac{j}{N+1}, \frac{j}{N+1}\right),$$

$$T_N = \sum\sum (c_{ij} - c_{..})\, \phi_N(U_i, U_j) + c_{..} \sum\sum \phi_N\left(\frac{i}{N+1}, \frac{j}{N+1}\right),$$

$$S_N^* = \sum\sum c_{ij}\, \phi_N\left(\frac{R_i}{N+1}, \frac{S_j}{N+1}\right),$$

$$T_N^* = \sum\sum (c_{ij} - c_{..})\, \phi_N(U_i, V_j) + c_{..} \sum \phi_N\left(\frac{i}{N+1}, \frac{j}{N+1}\right),$$

where $U_1,\ldots,U_N,V_1,V_2,\ldots,V_N$ are independent uniform random variables on (0,1) and $(R_1,\ldots,R_N)$, $(S_1,\ldots,S_N)$ are the ranks among $(U_1,\ldots,U_N)$ and $(V_1,\ldots,V_N)$ respectively. If

i) ϕ_N is piecewise Δ-monotone,

ii) ϕ_N is obtained from a function ϕ which belongs to L_8 and is piecewise monotone,

iii) ϕ_N satisfies either (3.18) or (3.25) or the conditions (i) and (ii) of lemma 3.1 with k=8 then

3.28 $$S_N \sim S_N^1 , \ S_N \sim T_N, \ S_N^* \sim T_N^* .$$

4. ASYMPTOTIC NORMALITY

The results of section 3 reduce the problem of finding the asymptotic distributions of the rank score statistics S_N, S_N^1 and S_N^* to the simpler one of finding asymptotic distributions of T_N and T_N^*.

In the following, the asymptotic normality of the statistic T_N is considered (that of T_N^* follows along similar lines).

The statistic

4.1 $$T_N = \sum_{\alpha_1} \cdots \sum_{\alpha_m} (b_{\alpha_1,\ldots,\alpha_m} - \bar{b})\, \phi\,(U_{\alpha_1},\ldots, U_{\alpha_m}) + \bar{b} \sum_{\alpha_1} \cdots \sum_{\alpha_m} \phi\,\left(\frac{\alpha_1}{N+1}, \ldots, \frac{\alpha_m}{N+1}\right)$$

has the same form as the Hoeffding (1948) U-statistic except for the coefficients. For studying the conditions for asymptotic normality, the same method as that adopted by Hoeffding (1948) will be used.

As in other sections, for the sake of simplicity, ϕ functions with two arguments are considered. The cases of symmetric and non-symmetric ϕ are treated separately. For some special values of $b_{\alpha_1,\ldots,\alpha_m}$, an example is cited where some well known limit theorems for dependent random variables can be applied.

Case I: Symmetric ϕ .

The statistic T_N in (4.1) becomes

4.2 $$T_N = \sum_{i} \sum_{\neq j} (c_{ij} - \bar{c})\, \phi\,(U_i, U_j) + \bar{c} \sum_{i} \sum_{\neq j} \phi\,\left(\frac{i}{N+1}, \frac{j}{N+1}\right).$$

Without loss of generality assume that

4.3 $$\int_0^1 \int_0^1 \phi\,(\lambda,\theta)\, d\lambda \;\, d\theta = 0$$

Here

4.4 $$\phi(\lambda,\theta) = \phi(\theta,\lambda),$$

and hence

4.5 $$\text{Var } T_N = \zeta_1 \left[\sum_{i} \neq \sum_{j} \neq \sum_{k} (b_{ij}b_{ik} + b_{ij}b_{ki} + b_{ij}b_{jk} + b_{ij}b_{kj}) \right] + \zeta_2 \left[\sum_{i} \neq \sum_{j} (b_{ij}^{\,2} + b_{ij}\, b_{ji}) \right],$$

where

4.6 $$\zeta_1 = E\, \phi(U_i, U_j)\, \phi\, (U_i, U_k),$$

$$\zeta_2 = E\, \phi^2(U_i, U_j),\ b_{ij} = c_{ij} - \bar{c}, \quad \text{for} \quad i \neq j \neq k.$$

Let the conditional expectation, for fixed U_i, be written as

4.7 $$\phi_1(U_i) = E^{U_i}\, \phi\, (U_i, U_j) = E^{U_i}\, \phi\, (U_j, U_i).$$

Then

4.8 $$E\ \ \phi_1(U_i) = E\, \phi(U_i, U_j) = 0,$$

$$\text{Var}\ \ \phi_1(U_i) = E\ \ \phi_1^{\,2}(U_i) = E\ E^{U_i}\ \phi\, (U_i, U_j)\ E^{U_i} \phi\, (U_i, U_k)$$

$$= E\, \phi(U_i, U_j)\ \phi\ (U_i, U_k) = \zeta_1.$$

Let

4.9 $$V_N = \sum_i \phi_1(U_i) \sum_{j(\neq i)} (b_{ij} + b_{ji}) = \sum_i B_i\, \phi_1(U_i),$$

where

4.10 $$B_i = \sum_{j(\neq i)} (b_{ij} + b_{ji}).$$

Then it follows that

4.11 $$\text{Var } V_N = \zeta_1 \sum_i \left[\sum_{j(\neq i)} b_{ij} + b_{ji} \right]^2 .$$

In the following, the conditions under which $T_N \sim V_N$ are studied. Let

4.12 $$B_N = \sum_i \underset{\neq}{} \sum_j \underset{\neq}{} \sum_k (b_{ij}b_{ik} + b_{ij}b_{ki} + b_{ij}b_{jk} + b_{ij}b_{ki}),$$

$$C_N = \sum_{i \neq j} \sum (b_{ij})^2, \quad D_N = \sum_{i \neq j} \sum b_{ij}\, b_{ji} .$$

Then the expressions for variances can be written as

4.13 $$\text{Var } T_N = B_N \; \zeta_1 + (C_N + D_N)\, \zeta_2,$$

4.14 $$\text{Var } V_N = (2C_N + 2D_N + B_N)\, \zeta_1,$$

4.15 $$\text{Covar } (T_N, V_N)$$

$$= E \left[\sum_i \phi_1(U_i) \sum_{j(\neq i)} (b_{ij}+b_{ji}) \right] \left[\sum_{i \neq j} \sum b_{ij}\, \phi(U_i, U_j) \right]$$

$$= E \left[\sum_i \phi_1(U_i) \sum_{j \neq i} (b_{ij}+b_{ji}) \right] \left[\sum_i E^{U_i} \sum_{j(\neq i)} b_{ij}\, \phi(U_i, U_j) \right]$$

$$= E\, V_N^2 = (2C_N + 2D_N + B_N)\, \zeta_1 .$$

Hence

4.16 $$E(V_N - T_N)^2 = \text{Var } T_N + \text{Var } T_N - 2 \text{ Covar } (V_N, T_N)$$

$$= \text{Var } T_N - \text{Var } V_N = (\zeta_2 - 2\, \zeta_1)(C_N + D_N).$$

If

4.17 $$\zeta_1 \neq 0 \quad \text{and} \quad \frac{C_N + D_N}{B_N} \to 0,$$

then from (4.13), (4.14) and (4.16) it follows that $V_N \sim T_N$. It can be seen that the number of terms in the expression of B_N is of higher order compared to C_N and D_N and (4.17) will be satisfied if the b_{ij} are of the same order.

Using the fact that $V_N \sim T_N$ and applying results of Hájek (1961) to the statistic V_N the theorem stated below follows immediately.

Theorem 4.1. If the function ϕ is symmetric in its arguments, the functions ϕ and ϕ_N satisfy the conditions of theorem 3.2, and

i) $\zeta_1 \neq 0$,

ii) $\dfrac{C_N + D_N}{B_N} \to 0$,

iii) $\lim\limits_{N\to\infty} \dfrac{\max_i B_i^2}{\sum_i B_i^2} = 0$,

then the statistics S_N, S_N^1 of section 3 have an asymptotic normal distribution with mean zero and variance $E\,T_N^2$.

Case II: Nonsymmetric ϕ.

Let

4.18 $$\Psi_1(U_i) = E^{U_i}\,\phi(U_i,U_j),$$

$$\Psi_2(U_i) = E^{U_i}\,\phi(U_j,U_i),$$

and

4.19 $$\zeta_{11} = E\,\phi(U_i,U_j)\,\phi(U_i,U_k) = E\,\Psi_1^2(U_i),$$

$$\zeta_{12} = E\,\phi(U_i, U_j)\,\phi(U_k, U_i) = E\,\phi(U_j, U_i)\,\phi(U_i, U_k) = E\,\Psi_1(U_i)\,\Psi_2(U_i),$$

$$\zeta_{13} = E\,\phi(U_j, U_i)\,\phi(U_k, U_i) = E\,\Psi_2^{\,2}(U_i),$$

$$\zeta_{21} = E\,\phi^2(U_i, U_j), \quad \zeta_{22} = E\,\phi(U_i, U_j)\,\phi(U_j, U_i)$$

and

4.20 $$B_{1N} = \sum_{i \neq j \neq k}\sum\sum b_{ij}b_{ik}, \quad B_{2N} = \sum_{i \neq j \neq k}\sum\sum (b_{ij}b_{ki} + b_{ji}+b_{ij}),$$

$$B_{3N} = \sum_{i \neq j \neq k}\sum\sum b_{ji}\, b_{ki} \,.$$

With this notation it is readily seen that

4.21 $$\text{Var } T_N = \zeta_{11} B_{1N} + \zeta_{12} B_{2N} + \zeta_{13} B_{3N} + \zeta_{21} C_N + \zeta_{22} D_N \,.$$

Let

4.22 $$W_N = \sum_i \Psi_1(U_i) \sum_{j(\neq i)} b_{ij} + \sum_i \Psi_2(U_i) \sum_{j(\neq i)} b_{ji}$$

$$= \sum_i c_{iN}\, \Psi_1(U_i) + \sum_i d_{iN}\, \Psi_2(U_i),$$

where

4.23 $$c_{iN} = \sum_{j(\neq i)} b_{ij}, \quad d_{iN} = \sum_{j(\neq i)} b_{ji} \,.$$

Hence,

4.24 $$\text{Var } W_N = \zeta_{11} \sum c_{iN}^{\,2} + \zeta_{13} \sum d_{iN}^{\,2} + 2\,\zeta_{12} \sum c_{iN}d_{iN}$$

$$= \zeta_{11}\left[C_N + B_{1N}\right] + \zeta_{13}\left[C_N + B_{3N}\right] + \zeta_{12}\left[B_{2N} + 2D_N\right],$$

4.25 $$\text{Covar } (W_N\,,\, T_N) = \text{Var } W_N,$$

4.26 $$E(W_N - T_N)^2 = \text{Var } T_N - \text{Var } W_N = C_N(\zeta_{21} - \zeta_{11} - \zeta_{13}) + D_N(\zeta_{22} - 2\zeta_{12}),$$

and

4.27 $$\frac{E(W_N-T_N)^2}{\text{Var } W_N} = \frac{C_N(\zeta_{21}-\zeta_{11}-\zeta_{13}) + D_N(\zeta_{22}-2\zeta_{12})}{\zeta_{11}(C_N+B_{1N}) + \zeta_{13}(C_N+B_{3N}) + \zeta_{12}(B_{2N}+2D_N)} .$$

Note that the number of terms in B_{1N}, B_{2N} and B_{3N} is of higher order than that in C_N and D_N.

From (4.27) it follows that if

4.28 $$\zeta_{11} \neq 0, \quad \zeta_{12} \neq 0, \quad \zeta_{13} \neq 0, \text{ and } \quad \frac{C_N + D_N}{B_N} \to 0 \quad \text{as} \quad N \to \infty,$$

then

$$W_N \sim T_N .$$

The conditions under which the statistic W_N has asymptotic normal distribution will become clear by the following lemma.

Lemma 4.1. Let $\xi_1, \ldots, \xi_m$ be m piecewise monotone real-valued functions defined on the unit interval (0,1). Let $U_1,\ldots,U_N$ be independent R(0,1) random variables. For every set of nonnegative constants $p_1,\ldots,p_m$ and $q_1,\ldots,q_m$ if the set of coefficients b_{ij}, i=1,...,N; j=1,...,m, are such that

4.29 $$\lim_{N\to\infty} \frac{\max\limits_{1 \leq i \leq N} (p_1 b_{i1} + \ldots + p_m b_{im})^2}{\sum_i (p_1 b_{i1} + \ldots + p_m b_{im})^2} = 0,$$

and if

4.30 $$\infty > \int_0^1 \left[q_1 \xi_1(\lambda) + \ldots + q_m \xi_m(\lambda)\right]^2 d\lambda > 0,$$

then the statistic

4.31 $$T_N = \sum_{i=1}^{N} \left[b_{i1} \; \xi_1(U_i)+\ldots+b_{im} \; \xi_m(U_i) \right]$$

has an asymptotic normal distribution.

Proof: From (4.29) and (4.30) it is clear that theorem 4.1 of Hájek (1961) can be applied and the asymptotic normality of

4.32 $$T_N = \sum_i (p_1 b_{i1}+\ldots+p_m b_{im})(q_1 \; \xi_1(U_i)+\ldots+q_m \; \xi_m(U_i))$$

follows. However the set of constants $p_1,\ldots,p_m$, $q_1,\ldots,q_m$ being arbitrary, the joint normality of

4.33 $$\sum b_{i1} \; \xi_1(U_i),\ldots, \sum b_{im} \; \xi_m(U_i)$$

and hence that of T_N follows.

Applying this lemma to the statistic W_N, the following theorem can be stated:

Theorem 4.2. With the previous notation, if for every set of constants p_1, p_2, p_3, p_4

i) $\zeta_{11} \neq 0$, $\zeta_{12} \neq 0$, $\zeta_{13} \neq 0$,

ii) $$\frac{C_N + D_N}{B_N} \to 0,$$

iii) $$\lim_{N\to\infty} \frac{\max\limits_{1 \leq i \leq n} (p_1 c_{iN} + p_2 d_{iN})^2}{\sum (p_1 c_{iN} + p_2 d_{iN})^2} = 0$$

iv) $$\infty > \int_0^1 \left[p_3 \int_0^1 \phi(\lambda,\theta) \, d\lambda + p_4 \int_0^1 \phi(\theta,\lambda)\lambda \right]^2 d\theta > 0,$$

then the statistics T_N, S_N and S_N^1 (see 4.2), and theorem 3.2) have the same asymptotic normal distribution with mean zero and variance given by (4.21).

Special Cases

In the following, two examples are quoted where the coefficients c_{ij} take values 0 or 1. In these cases, the results of section 3 together with some well known limit theorems can be applied directly.

a) Bhuchongkul (1964) studied a class of tests for testing independence in bivariate populations. The test statistic was of the form

4.34 $$U_N = \sum_{i=1}^{N} \phi_N \left(\frac{R_i}{N}\right) \phi_N \left(\frac{S_i}{N}\right),$$

where R_i is the rank of X_i among $X_1,\ldots,X_N$; S_i is the rank of Y_i among $Y_1,\ldots,Y_N$; ϕ_N satisfies the conditions mentioned in section 3; and $(X_1, Y_1),\ldots,(X_N, Y_N)$ is a random sample from a bivariate population with an absolutely continuous distribution function. In case the X_i and Y_i are independent, it follows from section 3 that the statistic U_N is asymptotically equivalent to

4.35 $$U_N^1 = \sum_{i=1}^{N} \phi \left[F(X_i)\right] \phi \left[G(Y_i)\right],$$

where F and G are the marginal distribution functions. The summands of (4.35) are independent and the standard methods of central limit theorems are applicable.

b) Consider a statistic

4.36 $$V_N = \sum_{i=1}^{N-1} \phi_N \left[\frac{R_i}{N}\right] \phi_N \left[\frac{R_{i+1}}{N}\right],$$

where R_i is the rank of X_i among $X_1,\ldots,X_N$. If the random variables X_i are mutually independent and identically distributed with an absolutely continuous distribution function F, then it is seen from section 3 that V_N is asymptotically equivalent to

4.37 $$V_N^1 = \sum_{i=1}^{N-1} \phi \left[F(X_i)\right] \phi \left[F(X_{i+1})\right].$$

The asymptotic normality of V_N^1 can be proved by using a theorem of Hoeffding and Robbins (1948). The statistic V_N plays an important role in testing serial correlation between successive observations and is studied by the author (1964).

typ: E.J.

REFERENCES

[1] BHUCHONGKUL, S. (1964). A class of nonparametric tests for independence in bivariate populations. Ann. Math. Statist. **35** 138-149.

[2] CHERNOFF, H. and SAVAGE, I.R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. Ann. Math. Statist. **29** 972-994.

[3] HÁJEK, JAROSLAV (1961). Some extensions of the Wald-Wolfowitz-Noether theorem. Ann. Math. Statist. **32** 506-523.

[4] HÁJEK, JAROSLAV (1962). Asymptotically most powerful rank tests. Ann. Math. Statist. **33** 1124-1147.

[5] HODGES, J.L. AND LEHMANN, E.L. (1956). The efficiency of some nonparametric competitors of the t-test. Ann. Math. Statist. **27** 324-335.

[6] HODGES, J.L. AND LEHMANN, E.L. (1961). Comparison of the normal scores and Wilcoxon tests. Proc. Fourth Berkeley Symp. Math. Statist. Prob. 307-317.

[7] HOEFFDING, WASSILY (1948). A class of statistics with asymptotically normal distribution. Ann. Math. Statist. **19** 293-325.

[8] HOEFFDING, WASSILY AND ROBBINS, HERBERT (1948). The central limit theorem for dependent variables. Duke Math. J. **15** 773-780.

[9] JOGDEO, KUMAR (1964). A class of nonparametric tests for serial correlation between successive observations. (Unpublished).

[10] KLOTZ, JEROME (1963). Small sample power and efficiency for the one sample Wilcoxon and normal scores test. Ann. Math. Statist. **34** 624-632.

[11] NOETHER, G.E. (1954). On a theorem of Pitman. Ann. Math. Statist. **25** 514-522.

[12] PURI, M.L. (1964). Asymptotic efficiency of a class of c-sample tests. Ann. Math. Statist. **35** 102-121.