A CLASS OF NONPARAMETRIC TESTS FOR INDEPENDENCE IN BIVARIATE POPULATIONS[1]

By S. BHUCHONGKUL

University of California, Berkeley

1. Introduction and summary. This paper is concerned with the problem of testing the independence of two random variables X, Y on the basis of a random sample, $(X_1, Y_1), (X_2, Y_2), \cdots, (X_N, Y_N)$. The joint distribution function H and, consequently, the marginal distribution functions F and G are assumed to be absolutely continuous. The hypothesis to be tested may be stated as $H(x, y) = F(x)G(y)$.

The usual parametric test of this hypothesis is based on the sample correlation coefficient. Several nonparametric tests have been proposed and studied by Kendall [11], Hoeffding [8], [9], Blomqvist [1], and others. Konijn [13] has investigated the asymptotic power properties of some of these tests.

The class of rank tests for independence to be considered in this paper is based on the test statistics of the form

$$T_N = N^{-1}\sum_{i=1}^{N} E_{N,r_i}E'_{N,s_i}Z_{N,r_i}Z'_{N,s_i} \tag{1.1}$$

where $\{E_{N,i}\}$, $\{E'_{N,i}\}$, $i = 1, 2, \cdots, N$, are two sets of constants satisfying certain restrictions to be stated below, and $Z_{N,r_i} = 1$ ($Z'_{N,s_i} = 1$) when $X_i(Y_i)$ is the r_ith (s_ith) smallest of the X's (Y's) and $Z_{N,r_i} = 0$ ($Z'_{N,s_i} = 0$) otherwise.

By taking $E_{N,r_i}(E'_{N,s_i})$ to be the expected value of the r_ith (s_ith) standard normal order statistic from a sample of size N, we get the normal scores test statistic which belongs to the class of c_1-statistics considered by Fisher and Yates [4], Hoeffding [10] and Terry [17]. If we put $E_{N,r_i} = r_i$ and $E_{N,s_i} = s_i$, the resulting test statistic is equivalent to the Spearman rank correlation statistic.

The normal scores test is shown to be (a) the locally most powerful rank test and (b) asymptotically as efficient as the parametric correlation coefficient $\mathfrak{R}$-test for the alternatives (4.1) and (4.2) when the underlying distributions are normal. It is at least as efficient as the $\mathfrak{R}$-test when the alternative belongs to the class (4.1) and when H satisfies the restrictions stated in Theorem 3.

Tables 1 and 2 give the exact null distribution and some critical values of the normal scores statistic for $N \leqq 6$. Table 2 provides a comparison of the t-approximation with the exact distribution.

2. Assumptions and notation. Let $(X_1, Y_1), (X_2, Y_2), \cdots, (X_N, Y_N)$ be N mutually independent pairs of random variables with continuous joint distri-

Received 27 September 1962; revised 27 May 1963.

[1] This paper was prepared with the partial support of the Office of Naval Research, Contract Nonr-222-43. This paper in whole or in part may be reproduced for any purpose of the United States Government.

bution function H. Let F and G be the marginal distribution functions of X and Y, respectively.

Let $F_N(x) = (\text{number of } X_i \leqq x)/N$, $G_N(y) = (\text{number of } Y_i \leqq y)/N$ and $H_N(x, y) = [\text{number of } (X_i, Y_i) \leqq (x, y)]/N$.

The following representation of T_N is equivalent to (1.1):

$$T_N = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J_N[F_N(x)]L_N[G_N(y)]\, dH_N(x, y), \tag{2.1}$$

where $J_N(i/N) = E_{N,i}$ and $L_N(i/N) = E'_{N,i}$.

Let I_N be the interval in which $0 < F_N(x) < 1$ and I'_N be the one in which $0 < G_N(y) < 1$. The symbols K and o_p have the same meaning as in [2].

3. Asymptotic normality.

THEOREM 1. *If*

(1) $J(u) = \lim_{N\to\infty} J_N(u)$ *and* $L(u) = \lim_{N\to\infty} L_N(u)$ *exist for all* $0 < u < 1$ *and are not constant,*

(2) $\int\int_{I_N\times I'_N}\{J_N[F_N(x)]L_N[G_N(y)] - J[F_N(x)]L[G_N(y)]\}\, dH_N(x, y) = o_p(N^{-\frac{1}{2}})$,

(3) $J_N(1) = o(N^{\frac{1}{2}})$, $L_N(1) = o(N^{\frac{1}{2}})$,

(4) $|J(u)| \leqq K[u(1-u)]^{-\alpha}$, $0 < \alpha < \frac{1}{8}$, $|J'(u)| \leqq K[u(1-u)]^{-1}$, $|J''(u)| \leqq K[u(1-u)]^{-2}$,

where J' *and* J'' *denote the first and second derivatives of* J, *respectively,*

(5) L *satisfies the same conditions in* (4), *then*

$$\lim_{N\to\infty} P\left[\frac{T_N - \mu_N}{\sigma_N} \leqq t\right] = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{t} e^{-x^2/2}\, dx \tag{3.1}$$

uniformly with respect to F, G *and* H, *provided* $\sigma_N \neq 0$; *where*

$$\mu_N = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J[F(x)]L[G(y)]\, dH(x, y) \tag{3.2}$$

and

$$\begin{aligned} N\sigma_N^2 = \operatorname{Var}\Big\{ & J[F(X)]L[G(Y)] \\ & + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\phi_X(u) - F(u)]J'[F(u)]L[G(v)]\, dH(u, v) \\ & + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\phi_Y(v) - G(v)]J[F(u)]L'[G(v)]\, dH(u, v)\Big\}, \end{aligned} \tag{3.3}$$

where $\phi_X(u) = 1$ $(\phi_Y(v) = 1)$ *if* $X \leqq u$ $(Y \leqq v)$ *and is zero otherwise.*

PROOF. $J_N(F_N)L_N(G_N)$ can be written as $[J_N(F_N)L_N(G_N) - J(F_N)L(G_N)] + J(F_N)L(G_N)$ and using Taylor's expansion, we can express $J(F_N)L(G_N)$ as $J(F)L(G) + (F_N - F)J'(F)L(G) + (G_N - G)J(F)L'(G) + (\frac{1}{2})(F_N - F)^2J''[\theta F_N + (1-\theta)F]L[\theta G_N + (1-\theta)G] + (\frac{1}{2})(G_N - G)^2J[\theta F_N + (1-\theta)F]L''[\theta G_N + (1-\theta)G] + (F_N - F)(G_N - G)J'[\theta F_N + (1-\theta)F]L'[\theta G_N + (1-\theta)G]$,

where $0 < \theta < 1$. Finally, noting that dH_N can be written as $d(H_N - H) + dH$, we have

$$T_N = \sum_{i=1}^{3} A_{iN} + \sum_{i=1}^{8} B_{iN}, \tag{3.4}$$

where

$$A_{1N} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J(F)L(G)\, dH_N,$$

$$A_{2N} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (F_N - F)J'(F)L(G)\, dH,$$

$$A_{3N} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (G_N - G)J(F)L'(G)\, dH,$$

$$B_{1N} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (F_N - F)J'(F)L(G)\, d(H_N - H),$$

$$B_{2N} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (G_N - G)J(F)L'(G)\, d(H_N - H),$$

$$B_{3N} = \iint_{I_N \times I_N}{}' [\tfrac{1}{2}(F_N - F)^2]J''[\theta F_N + (1-\theta)F]L[\theta G_N + (1-\theta)G]\, dH_N,$$

$$B_{4N} = \iint_{I_N \times I_N}{}' [\tfrac{1}{2}(G_N - G)^2]J[\theta F_N + (1-\theta)F]L''[\theta G_N + (1-\theta)G]\, dH_N,$$

$$B_{5N} = \iint_{I_N \times I_N}{}' (F_N - F)(G_N - G)J'[\theta F_N + (1-\theta)F]L'[\theta G_N + (1-\theta)G]\, dH_N$$

$$B_{6N} = -\iint_{R\times R - I_N \times I_N}{}' [J(F)L(G) + (F_N - F)J'(F)L(G) + (G_N - G)J(F)L'(G)]\, dH_N,$$

$$B_{7N} = \iint_{R\times R - I_N \times I_N}{}' J_N(F_N)L_N(G_N)\, dH_N,$$

$$B_{8N} = \iint_{I_N \times I_N}{}' [J_N(F_N)L_N(G_N) - J(F_N)L(G_N)]\, dH_N.$$

We shall show that $\sum_{i=1}^{3} A_{iN}$, when properly normalized, has a limiting normal distribution. The B_N terms are shown to be $o_p(N^{-\frac{1}{2}})$ in Section 5.

Using the elementary inequality

$$|ab| \leqq |a|^r/r + |b|^s/s, \qquad r > 1, \quad 1/r + 1/s = 1, \tag{3.5}$$

it can be shown that $\sum_{i=1}^{3} A_{iN}$ is the average of N independent, identically distributed random variables, each having mean μ_N and finite third moment. Hence, (3.1) follows from the Berry-Esseen theorem ([14] pp. 282–288).

Assumption (4) is more restrictive than the corresponding one imposed by Chernoff and Savage in [2]. Nevertheless, if we consider $J = F^{-1}$, these conditions are satisfied by several distribution functions F, such as, normal, exponential, logistic and uniform. The inverse of the Cauchy distribution function does not satisfy Assumption (4) in this paper, nor does it satisfy the one in [2] where $\alpha = (\frac{1}{2}) - \delta$ is restricted to the interval $(0, \frac{1}{2})$. The inverse of the distribution function F for which $F(x)$ behaves like $1 - x^{-\frac{4}{3}}$ as x approaches $+\infty$ satisfies Assumption (4) in [2] but not the one in this paper.

Theorem 2 which is an extension of Theorem 2 in [2], shows that Assumptions (1) to (3) in Theorem 1 are likely to be satisfied when we deal with the inverse functions of distribution functions.

THEOREM 2. *If $J_N(i/N)$ and $L_N(i/N)$ are the expectations of the* ith *order statistic of samples of size N from populations whose cumulative distribution functions are the inverse functions of J and L respectively and if Assumptions* (4) *and* (5) *of Theorem* 1 *are satisfied, then*

$$\lim_{N\to\infty} J_N(u) = J(u), \qquad \lim_{N\to\infty} L_N(u) = L(u), \qquad 0 < u < 1, \tag{3.6}$$

$$J_N(1) = o(N^{\frac{1}{2}}), \qquad L_N(1) = o(N^{\frac{1}{2}}), \tag{3.7}$$

$$\iint_{I_N\times I_N} [J_N(F_N)L_N(G_N) - J(F_N)L(G_N)]\, dH_N = o(N^{-\frac{1}{2}}). \tag{3.8}$$

PROOF. Equation (3.6) is well known and (3.7) follows immediately from the result in [2] that $|J_N(1)| \leqq KN^{\alpha}$ and the assumption that $0 < \alpha < \frac{1}{8}$.

To prove (3.8), we write $J_N(F_N)L_N(G_N) - J(F_N)L(G_N)$ as $J_N(F_N)[L_N(G_N) - L(G_N)] + L_N(G_N)[J_N(F_N)] - J(F_N)] - [J_N(F_N) - J(F_N)][L_N(G_N) - L(G_N)]$. The result follows from (7.24) and (7.25) in [2].

4. The normal scores test. Before investigating the power properties of this test, it is necessary to specify the alternatives to be considered. As Konijn remarks in [13], it appears difficult to formulate a class of alternatives which is reasonably wide and reasonably easy to handle mathematically.

We shall consider the class of alternatives under which X and Y are given by

$$X = (1 - \theta)U + \theta Z, \qquad Y = (1 - \theta)V + \theta Z \tag{4.1}$$

where $0 \leqq \theta \leqq 1$ and U, V, Z are independent random variables. The hypothesis to be tested is that $\theta = 0$.

This class is similar to that considered by Konijn in [13] under which

$$X = \lambda_1 U + \lambda_2 V, \qquad Y = \lambda_3 U + \lambda_4 V \tag{4.2}$$

where the λ_i's are real numbers, U and V are independently distributed. The hypothesis for this model states that $\lambda_1 = \lambda_4 = 1$, $\lambda_2 = \lambda_3 = 0$.

It is easy to show that the normal scores test is the locally most powerful rank test against the alternatives of the form (4.1) and (4.2) with the variables

involved being normally distributed or the alternative that X and Y have a bivariate normal distribution with correlation coefficient ρ.

4.1. *Comparison with the correlation coefficient* $\mathfrak{R}$*-test.* Throughout this section and the following section, unless otherwise stated, the alternatives are assumed to be of the form (4.1).

Theorem 1 establishes the asymptotic normality of the normal scores statistic. Now we proceed to find the variance of the statistic T_N under the hypothesis of independence. Under H, we have

$$\begin{aligned}(4.1.1)\qquad &\iint [\phi_X(x) - F(x)]J'[F(x)]L[G(y)]h(x, y)\,dx\,dy \\ &\qquad = -E_0\{L[G_0(Y)]\}\{J[F_0(X)] - E_0\{J[F_0(X)]\}\},\end{aligned}$$

where the functions with subscript 0 are computed under H. A similar expression can be obtained for $\int\int[\phi_Y(y) - G(y)]J[F_0(x)]L'[G_0(y)]h(x, y)\,dx\,dy$.

From (3.3) and (4.1.1), we have

$$\begin{aligned}(4.1.2)\qquad N\,\mathrm{Var}_0(T_N) &= \mathrm{Var}_0\{J[F_0(X)]L[G_0(Y)]\} \\ &\quad - E_0^2\{L[G_0(Y)]\}\,\mathrm{Var}_0\{J[F_0(X)]\} - E_0^2\{J[F_0(X)]\}\,\mathrm{Var}_0\{L[G_0(Y)]\}.\end{aligned}$$

Moreover, it is easy to see that if we are dealing with a sequence of alternatives of the forms (4.1) or (4.2) or the bivariate normal form which converges to the hypothesis as $N \to \infty$, then

$$(4.1.3)\qquad \lim_{N\to\infty} \mathrm{Var}_N(T_N)/\mathrm{Var}_0(T_N) = 1.$$

In dealing with the normal scores statistic $\mathfrak{N}_N$, we replace J and L by $J_0 = \Phi^{-1}$, the inverse of the standard normal distribution function and get

$$(4.1.4)\qquad E(\mathfrak{N}_N) = \iint J_0[F(x)]J_0[G(y)]h(x, y)\,dx\,dy$$

and

$$(4.1.5)\qquad \mathrm{Var}_0(\mathfrak{N}_N) = 1/N.$$

In this section, the normal scores test will be compared with the parametric test, $\mathfrak{R}$, based on the sample correlation coefficient,

$$(4.1.6)\qquad R_N = \sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}) \Big/ \left[\sum_{i=1}^{N}(X_i - \bar{X})^2 \sum_{i=1}^{N}(Y_i - \bar{Y})^2\right]^{\frac{1}{2}}.$$

Cramér ([3], pp. 359–366) shows that

$$(4.1.7)\qquad E(R_N) = \rho + O(N^{-1})$$

where ρ is the correlation coefficient of X and Y and

$$(4.1.8)\qquad \mathrm{Var}_0(R_N) = 1/N.$$

Under (4.1) with U, V, Z being independent normal variables, we have

$$E(\mathfrak{N}_N) = \rho. \tag{4.1.9}$$

Hence, the Pitman efficiency of the $\mathfrak{N}$-test with respect to the $\mathfrak{R}$-test in this case is unity.

This conclusion holds whenever the marginal distributions of X and Y are normal. In particular, it holds when the alternatives are of the form (4.2) with U, V being independent normal variables or when X and Y have a bivariate normal distribution with correlation coefficient ρ.

THEOREM 3. *Under* (4.1), *let* f_0, g_0 *and* f_0^* *be the density functions of* U, V *and* Z, *respectively. Let* f_θ *and* g_θ *denote the marginal density functions of* X *and* Y, *and* h_θ *their joint density function. If*

(1) U, V *and* Z *have finite fourth moments,*

(2) f_0, g_0 *and* f_0^* *are positive and continuous on* $(-\infty, \infty)$,

(3) f_0 *and* g_0 *are twice differentiable on* $(-\infty, \infty)$,

$$(4)\quad \frac{d^i}{d\theta^i}\mu_N(\theta)\,|_{\theta=0} = \iint \frac{d^i}{d\theta^i}\{J_0[F_\theta(x)]J_0[G_\theta(y)]h_\theta(x,y)\}_{\theta=0}\,dx\,dy$$

for $i = 1, 2$ *and* $\mu_N(\theta) = E(\mathfrak{N}_N)$ *given by* (4.1.4),

then the asymptotic efficiency of the normal scores test with respect to the correlation coefficient test is given by

$$e_{\mathfrak{N},\mathfrak{R}}(F_0, G_0, F_0^*) = \left\{\int J_0'[F_0(x)]f_0^2(x)\,dx \int J_0'[G_0(y)]g_0^2(y)\,dy\right\}^2. \tag{4.1.10}$$

PROOF. We first note that

$$\left.\frac{d\rho}{d\theta}\right|_{\theta=0} = 0; \qquad \left.\frac{d^2\rho}{d\theta^2}\right|_{\theta=0} = 2, \tag{4.1.11}$$

also,

$$\frac{d}{d\theta}\mu_N(\theta)\,|_{\theta=0} = 0, \tag{4.1.12}$$

and

$$\frac{d^2}{d\theta^2}[\mu_N(\theta)]\,|_{\theta=0} = 2\int J_0'[F_0(x)]f_0^2(x)\,dx \int J_0'[G_0(y)]g_0^2(y)\,dy. \tag{4.1.13}$$

Applying an extension of Pitman's theorem due to Noether [15], we get (4.1.10).

It should be noted that if $F_0 = G_0$, $e_{\mathfrak{N},\mathfrak{R}}(F_0, G_0, F_0^*)$ is the square of the corresponding 2-sample efficiency.

Chernoff and Savage have shown in [2] that $\int J_0'[F_0(x)]f_0^2(x)\,dx \geqq 1$ and it is

equal to 1 only if $F_0 = \Phi$. Hence, $e_{\mathfrak{N},\mathfrak{R}}(F_0, G_0, F_0^*) \geqq 1$ and the equality holds only if F_0 and G_0 are Φ.

Under (4.2) with U and V having the same distribution function F_0, we can apply another extension of Pitman's theorem due to Konijn [13], and get

$$(4.1.14) \quad e_{\mathfrak{N},\mathfrak{R}}(F_0) = \left\{\int xJ_0[F_0(x)]f_0(x)\,dx \int J_0'[F_0(x)]f_0^2(x)\,dx\right\}^2.$$

Regarding x as a function of J_0 as in [2], we can write

$$(4.1.15) \quad e_{\mathfrak{N},\mathfrak{R}}(F_0) = \left\{xJ_0\varphi(J_0)\,dJ_0 \int \left[\varphi(J_0) \Big/ \left(\frac{dx}{dJ_0}\right)\right] dJ_0\right\}^2.$$

Subject to the restrictions that F_0 has mean 0 and variance 1, $\int\int x(J_0)J_0\varphi(J_0)$ $[dx/dJ_0^*]\varphi(J_0^*)\,dJ_0\,dJ_0^*$ does not attain its minimum at the point $x(J_0) = J_0$, i.e., when x is the identity function ([5]), pp. 528–552). Consequently, $e_{\mathfrak{N},\mathfrak{R}}(F_0)$ need not be $\geqq 1$ for all F_0.

4.2. *Comparison with the rank correlation test.* The rank correlation test is based on the statistic

$$(4.2.1) \quad T_{0N} = [12/N(N^2 - 1)]\sum_{i=1}^{N} (R_i - \tfrac{1}{2}(N + 1))(S_i - \tfrac{1}{2}(N + 1))$$

where R_i and S_i are the ranks of X_i and Y_i, respectively. Except for some constants, T_{0N} is the same as

$$(4.2.2) \quad W_N = \iint J_N[F_N(x)]J_N[G_N(y)]\,dH_N(x, y) = \frac{1}{N^3}\sum_{i=1}^{N} R_i S_i$$

when $J_N(i/N) = i/N$.

Since J is the identity in this case, we have

$$(4.2.3) \quad E(W_N) = \iint F(x)G(y)\,dH(x, y)$$

and

$$(4.2.4) \quad \mathrm{Var}_0(W_N) = 1/(144N).$$

In this section it is not necessary to assume the finiteness of the fourth moments of U, V and Z. However, we still have to assume that Z has finite second moment. So, without any loss of generality, let Z have zero mean and unit variance. The following computations are made under assumptions (2), (3) and (4) of Theorem 3.

Let $\nu_N(\theta)$ denote $E(W_N)$ under (4.1). Then, under the above assumptions, we have

$$(4.2.5) \quad \frac{d}{d\theta}[\nu_N(\theta)]\,|_{\theta=0} = 0,$$

and

$$\frac{d^2}{d\theta^2}[\nu_N(\theta)]\,|_{\theta=0} = 2\int f_0^2(x)\,dx \int g_0^2(y)\,dy. \tag{4.2.6}$$

Hence, the Pitman efficiency of the rank correlation test with respect to the normal scores test is

$$e_{W,\mathfrak{N}}(F_0, G_0, F_0^*) = \left\{ \frac{12\int f_0^2(x)\,dx \int g_0^2(y)dy}{\int J_0'[F_0(x)]f_0^2(x)\,dx \int J_0'[G_0(y)]g_0^2(y)\,dy} \right\}^2 \tag{4.2.7}$$

which for $F_0 = G_0 = F$ is the square of the expression for $e_{W,\mathfrak{N}}(F)$ in the 2-sample problem obtained by Hodges and Lehmann in [7].

4.3. *Exact null distribution.* Assume that $X_1 < X_2 < \cdots < X_N$. Then, the normal scores statistic can be written as

$$\mathfrak{N}_N = N^{-1}\sum_{i=1}^{N} (EZ_{N,i})(EZ_{N,s_i})\phi_{N,s_i}, \tag{4.3.1}$$

where $\phi_{N,s_i} = 1$ if Y_i has rank s_i among the Y's and it is zero otherwise.

In carrying out the test, we reject H if $\mathfrak{N}_N$ is too large or too small or if its absolute value is too large, according to the alternative under consideration.

The computations for Table 1 are based on the following:

$$P\left\{\mathfrak{N}_N = \frac{1}{N}\sum_{i=1}^{N} (EZ_{N,i})(EZ_{N,s_i})\,|\,H\right\} = \frac{1}{N!}, \tag{4.3.2}$$

where $(s_1, s_2, \cdots, s_N)$ is a permutation of $(1, 2, \cdots, N)$, and

$$EZ_{N,i} = -EZ_{N,N+1-i} \tag{4.3.3}$$

with $EZ_{N,(N+1)/2} = 0$ when N is odd.

The values $EZ_{N,i}$ used in the computations are obtained from Table 28 in [6]. For $N > 6$, it is advisable to use more accurate values of $EZ_{N,i}$ which can be found in [16].

For $N > 6$, the distinct values that the statistic can take on are too numerous to compute without the aid of an electric computer. An approximation to the exact null distribution is, therefore, desirable. Terry has pointed out in [17] that the t-distribution with $N - 2$ degrees of freedom provides a good approximation to the null distribution of $r(N-2)^{\frac{1}{2}}/(1-r^2)^{\frac{1}{2}}$, where, in our case,

$$r = \sum_{i=1}^{N} (EZ_{N,i})(EZ_{N,s_i}) \Big/ \sum_{i=1}^{N} (EZ_{N,i})^2. \tag{4.3.4}$$

An idea of the accuracy of this approximation is obtained from Table 2. For $N = 6$, the third column of the table gives the approximate values obtained from the t-tables.

TABLE 1

Exact null distribution of $\mathfrak{N}_N$

$P_c = N!\ P\{\mathfrak{N}_N \geqq c\}$

c	P_c	c	P_c	c	P_c	c	P_c
$N = 2$		$N = 6$		2.273188	96	.980468	235
.636192	1	4.116514	1	2.269953	98	.933757	239
$N = 3$		3.953298	2	2.173593	102	.924993	243
1.431432	1	3.922914	4	2.123064	106	.883228	247
.715716	3	3.729314	5	2.095428	110	.855592	251
$N = 4$		3.725889	7	2.017453	118	.824328	253
2.294100	1	3.581938	11	1.998188	122	.806133	257
1.941264	2	3.562673	13	1.958553	124	.786868	261
1.758276	4	3.532289	15	1.921093	128	.777617	265
1.222452	5	3.404178	17	1.879328	132	.740157	269
.970632	9	3.335264	18	1.848064	133	.727968	273
.535824	11	3.257289	22	1.823853	137	.681257	277
0.0	13	3.210578	24	1.773717	141	.661992	279
$N = 5$		3.191313	28	1.764953	143	.658757	283
3.195188	1	3.172048	29	1.704993	147	.511868	291
2.950163	3	3.063689	33	1.685728	149	.472233	293
2.748964	5	3.016978	35	1.646093	157	.465157	297
2.503939	7	3.013553	37	1.626828	165	.452968	305
2.460113	9	2.982289	39	1.580117	169	.406257	309
2.302740	10	2.916313	43	1.567928	171	.386992	313
2.215088	11	2.866664	47	1.548853	175	.328092	317
2.173279	15	2.839218	51	1.452493	179	.309897	319
1.842619	17	2.788689	53	1.374328	183	.309017	321
1.727055	21	2.691842	54	1.371093	187	.278633	323
1.683229	25	2.663813	58	1.351828	191	.259368	331
1.396395	27	2.591664	60	1.333633	193	.240103	333
1.352569	31	2.544953	64	1.305117	197	.236868	337
1.107544	35	2.525688	68	1.274733	201	.212657	341
1.021909	39	2.498242	70	1.255468	205	.199408	342
.906345	41	2.486053	74	1.246217	207	.140508	346
.862519	43	2.467858	75	1.177493	211	.081608	348
.575685	47	2.398064	79	1.158228	215	.062343	352
.531859	51	2.388813	83	1.118593	219	.034897	356
.446224	53	2.351353	87	1.111517	223	.015632	360
.245025	57	2.304642	88	1.099328	227		
.201199	59	2.292453	92	1.039368	231		
0.0	61						

TABLE 2

Critical values of $\mathfrak{N}_N$

$P_c^* = P\{\mathfrak{N}_N \geqq c\}$

c	P_c^*	c	P_c^*	Approximation to P_c^*
$N = 4$		$N = 5$		
2.294100	.041667	2.215088	.091667	
1.941264	.083333	2.173279	.125000	
		$N = 6$		
$N = 5$		3.725889	.009722	.006544
3.195188	.008333	3.335264	.025000	.025306
2.950163	.025000	3.016978	.048611	.048745
2.748964	.041667	3.013553	.051389	.049046
2.503939	.058333	2.498242	.097222	.100719
		2.486053	.102778	.102127

5. Higher order terms. In the proof that the B_N terms are $o_p(N^{-\frac{1}{2}})$, the following result is used throughout.

Let $S_{N\epsilon} = \{x:F(x)[1 - F(x)] > (\eta_\epsilon)/N\}$ and $S'_{N\epsilon} = \{y:G(y)[1 - G(y)] > (\eta_\epsilon)/N\}$. Then, for any $\epsilon > 0$, we can choose η_ϵ independently of F, G and N such that

$$P\{X_i \varepsilon S_{N_\epsilon}, Y_i \varepsilon S'_{N_\epsilon}, i = 1, 2, \cdots, N\} \geqq 1 - \epsilon. \tag{5.1}$$

The asymptotic negligibility of B_{7N} and B_{8N} follows immediately from Assumptions (1), (2) and (3) of Theorem 1. One can show that B_{6N} is $o_p(N^{-\frac{1}{2}})$ by applying (5.1), Assumptions (4) and (5) of Theorem 1 and by observing that $|(i/N) - F(X_i)|\{F(X_i)[1 - F(X_i)]\}^{-1}$ is bounded in probability where $X_1 < X_2 < \cdots < X_N$.

Applying the basic inequality (3.5), we get the asymptotic negligibility of B_{3N}, B_{4N} and B_{5N} as an extension of the results in [2].

However, to show that $N^{\frac{1}{2}}B_{1N} \to_P 0$, one cannot extend the method used in [2]. First we write $N^{\frac{1}{2}}B_{1N}$ as

$$N^{\frac{1}{2}}B_{1N} = B_{11N} + B_{12N} + B_{13N} \tag{5.2}$$

where

$$B_{11N} = \int_a^b \int_c^d N^{\frac{1}{2}}(F_N - F)J'(F)L(G)\, d(H_N - H),$$

$$B_{12N} = \iint_{R\times R-[a,b]\times[c,d]} N^{\frac{1}{2}}(F_N - F)J'(F)L(G)\, dH_N,$$

$$B_{13N} = \iint_{R\times R-[a,b]\times[c,d]} N^{\frac{1}{2}}(F_N - F)J'(F)L(G)\, dH.$$

We shall have the desired result if we can show that for N sufficiently large,

$$P\{|B_{11N}| > \epsilon\} < \epsilon \tag{5.3}$$

and for sufficiently small a, c and sufficiently large b, d, the following hold:

$$P\{|B_{12N}| > \epsilon\} < \epsilon, \tag{5.4}$$

$$P\{|B_{13N}| > \epsilon\} < \epsilon. \tag{5.5}$$

The proof of (5.11) is similar to the proof of the Helley-Bray lemma ([14], pp. 180–181).

Let $g^{(N)}(x, y) = N^{\frac{1}{2}}[F_N(x) - F(x)]J'[F(x)]L[G(y)]$. We approximate $g^{(N)}(x, y)$ by simple functions as follows.

Divide $[a, b]$ into $a = x_{m1} < x_{m2} < \cdots < x_{m,k_m+1} = b$, and $[c, d]$ into $c = y_{m1} < y_{m2} < \cdots < y_{m,l_m+1} = d$ such that $\sup_k(x_{m,k+1} - x_{mk}) \to 0$ and $\sup_k(y_{m,k+1} - y_{mk}) \to 0$ as $m \to \infty$.

Define

$$g_m^{(N)}(x, y) = \sum_{i=1}^{k_m} \sum_{j=1}^{l_m} g^{(N)}(x_{m,i+1}, y_{m,j+1}) I_{ij}^{(m)}(x, y),$$

where $I_{ij}^{(m)}(x, y) = 1$ if $x_{m,i} < x \leqq x_{m,i+1}$ and $y_{m,j} < y \leqq y_{m,j+1}$ and otherwise it is zero. Then

$$|B_{11N}| \leqq C_{11N} + C_{12N} + C_{13N} \tag{5.6}$$

where

$$C_{11N} = \left| \int_a^b \int_c^d [g^{(N)}(x, y) - g_m^{(N)}(x, y)] \, dH_N(x, y) \right|,$$

$$C_{12N} = \left| \int_a^b \int_c^d g_m^{(N)}(x, y) \, dH_N(x, y) - \int_a^b \int_c^d g_m^{(N)}(x, y) \, dH(x, y) \right|,$$

$$C_{13N} = \left| \int_a^b \int_c^d [g^{(N)}(x, y) - g_m^{(N)}(x, y)] \, dH(x, y) \right|.$$

Since $N^{\frac{1}{2}} \sup_{-\infty<x<\infty} |F_N(x) - F(x)|$ has a nondegenerate limiting distribution [12], it follows that for any $\epsilon > 0$, there exists a K such that

$$P\{N^{\frac{1}{2}} |F_n(x) - F(x)| \leqq K\} \geqq 1 - \epsilon. \tag{5.7}$$

Hence, with probability greater than $1 - \epsilon$, $g^{(N)}(x, y)$ is bounded on $[a, b] \times [c, d]$, and it follows that $C_{12N} \to_P 0$ since $H_N(x, y) \to H(x, y)$ almost surely.

It is not hard to show that $E\,|C_{11N}|$ and $E\,|C_{13N}|$ converge to 0 uniformly in N as $m \to \infty$. Consequently, C_{11N} and C_{13N} converge in probability to 0 uniformly in N as $m \to \infty$ and (5.3) follows.

We can prove (5.4) by showing that

$$E|B_{12N}| \leqq K \iint_{S_{N_\epsilon} \times S'_{N_\epsilon} - [a,b] \times [c,d]} [F(1 - F)]^{-\frac{1}{2}} \left[1 + \frac{(1 - 2F)}{NF}\right]^{\frac{1}{2}} \cdot [G(1 - G)]^{-\alpha} \, dH. \tag{5.8}$$

Since, on S_{N_ϵ}, $F(1 - F) > K/N$, we see that the above integral over the set $S_{N_\epsilon} \times S'_{N_\epsilon}$ is finite. Hence $E\,|B_{12N}| \to 0$ as $a, c \downarrow \infty$ and $b, d \uparrow \infty$.

Furthermore, it is easy to show that

$$E|B_{13N}| \leqq \iint_{R \times R - [a,b] \times [c,d]} [F(1 - F)]^{-\frac{1}{2}} [G(1 - G)]^{-\alpha} \, dH \tag{5.9}$$

which converges to 0 as $a, c \downarrow -\infty$, $b, d \uparrow \infty$ and (5.5) follows. Analogously, one can show that $B_{2N} = o_p(N^{-\frac{1}{2}})$.

6. Acknowledgment. I would like to express my deepest gratitude to Professor E. L. Lehmann whose guidance and encouragement made this work possible.

REFERENCES

[1] BLOMQVIST, N. (1950). On a measure of dependence between two random variables. *Ann. Math. Statist.* **21** 593–600.

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

[3] CRAMÉR, HARALD (1954). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[4] FISHER, R. A. and YATES, F. (1957). *Statistical Tables for Biological, Agricultural and Medical Research*, (3rd ed.). Hafner, New York.

[5] FORSYTH, A. R. (1927). *Calculus of Variations*. Cambridge Univ. Press.

[6] HARTLEY, H. O. and PEARSON, E. S. (ed.) (1956). *Biometrika Tables of Statisticians* **1**. Cambridge Univ. Press.

[7] HODGES, J. L., JR. and LEHMANN, E. L. (1961). Comparison of the normal scores and Wilcoxon tests. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.

[8] HOEFFDING, WASSILY (1948). A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.* **19** 293–325.

[9] HOEFFDING, WASSILY (1948). A non-parametric test of independence. *Ann. Math. Statist.* **19** 546–557.

[10] HOEFFDING, WASSILY (1951). 'Optimum' nonparametric tests. *Proc. Second Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.

[11] KENDALL, M. G. (1955). *Rank Correlation Methods*, (2nd ed.). Griffin, London.

[12] KOLMOGOROV, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari.* **4** 83–91.

[13] KONIJN, H. S. (1956). On the power of certain tests for independence in bivariate populations. *Ann. Math. Statist.* **27** 300–323.

[14] LOÈVE, MICHEL (1960). *Probability Theory*, (2nd ed.). Van Nostrand, Princeton, New Jersey.

[15] NOETHER, G. E. (1955). On a theorem of Pitman. *Ann. Math. Statist.* **26** 64–68.

[16] TEICHROEW, D. (1956). Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution. *Ann. Math. Statist.* **27** 410–426.

[17] TERRY, MILTON E. (1952). Some rank order tests which are most powerful against specific parameteric alternatives. *Ann. Math. Statist.* **23** 346–366.