The Annals of Statistics
1984, Vol. 12, No. 3, 1124–1129

ASYMPTOTIC NORMALITY OF A CLASS OF NONLINEAR RANK TESTS FOR INDEPENDENCE

BY SHINGO SHIRAHATA AND KAZUMASA WAKIMOTO

Osaka University and Okayama University

Asymptotic normality of a class of nonlinear rank statistics to test the null hypothesis of total independence of an m-variate population is proved. The rank statistics are generated from $2m$-variate square integrable functions such that they are symmetric and nondegenerate. Some results under contiguous alternatives are also given.

1. Introduction. Let $\mathbf{X}_i = (X_{i1}, \cdots, X_{im})$, $i = 1, \cdots, n$ be a random sample from a population with distribution function $F(\mathbf{x}) = F(x_1, \cdots, x_m)$. Suppose we want to test the null hypothesis H; $F(\mathbf{x}) = \prod_{j=1}^{m} F_j(x_j)$, where F_j is the jth marginal distribution function of F, under the constraint that $m \geq 2$ and F and F_j's are continuous but unknown. Then it is natural to test H based on $\mathbf{R}_i = (R_{i1}, \cdots, R_{im})$, $i = 1, \cdots, n$ where R_{ij} is the rank of X_{ij} among $\{X_{1j}, \cdots, X_{nj}\}$.

The asymptotic normality of linear rank statistics of the form $\sum_{i=1}^{n} J_n(\mathbf{R}_i)$ is investigated by many authors especially when $m = 2$; see Bhuchongkul (1964), Behnen (1971, 1972), Ruymgaart (1974) and Ruymgaart, Shorack and van Zwet (1972). For the case $m > 2$, some rank statistics which are generalizations of product type linear rank statistics are investigated by Puri and Sen (1971), Sinha and Wieand (1977) and Al-Saadi and Young (1981). However, many nonlinear rank statistics are useful in the testing problem of H. Kendall's tau is one of the most useful ones and, in some models, nonlinear rank statistics give locally most powerful rank tests, see Shirahata (1974). Another example $L_n(m)$ is proposed by Wakimoto and Shirahata (1984) where

$$L_n(m) = ½ \sum_{i \neq j=1}^{n} Q_i Q_j \sin | \theta_i - \theta_j |$$

with

$$Q_i = (m + \sum_{j \neq k=1}^{m} \cos\{(R_{ij} - R_{ik})\pi/(n-1)\})^{1/2}$$

and

$$\theta_i = \tan^{-1} (\sum_{j=1}^{m} \sin\{(R_{ij} - 1)\pi/(n-1)\} / \sum_{j=1}^{m} \cos\{(R_{ij} - 1)\pi/(n-1)\}) .$$

The statistic $L_n(m)$ is obtained when the coefficient of concordance is considered based on a graphical representation of the ranked data. For $m = 1$ (though this case is not handled in this paper) certain nonlinear rank statistics also play a role in Beran (1972) for testing randomness against the alternative of serial dependence.

Received June 1983; revised February 1984.
AMS 1980 *subject classifications.* Primary 62E20; secondary 62G10.
Key words and phrases. Asymptotic normality, nondegenerate scores, nonlinear rank test, test for total independence.

In this paper, we consider a class of rank statistics of the form

$$S_n = 2 \sum_{1 \le i \le j \le n} a_n(\mathbf{R}_i, \mathbf{R}_j) \tag{1.1}$$

which includes Kendall's tau and $L_n(m)$ as special cases. In Section 2 we state our main result on the asymptotic normality of S_n under H and in Section 3 the result is proved. In Section 4, an asymptotic property of S_n under contiguous alternatives is given.

2. Main result. In order to state our main result, we need the following notations and assumptions.

ASSUMPTION 2.1. The scores $a_n(\mathbf{r}, \mathbf{s})$ used in S_n are symmetric in the sense that $a_n(\mathbf{r}, \mathbf{s}) = a_n(\mathbf{s}, \mathbf{r})$ for all $(\mathbf{r}, \mathbf{s}) \in \Omega \times \Omega$ where

$$\Omega = \{\mathbf{r} = (r_1, \cdots, r_m) \mid \text{ each } r_i \text{ is a positive integer not exceeding } n\}.$$

ASSUMPTION 2.2. If $r_i = s_i$ for at least one i in $\mathbf{r}$ and $\mathbf{s}$, then

$$a_n(\mathbf{r}, \mathbf{s}) = 0.$$

In (1.1), the pairs $(\mathbf{r}, \mathbf{s})$ satisfying the condition of Assumption 2.2 do not appear with probability one. Hence, Assumption 2.2 is only for convenience.

Let us define a function $\phi_n(\mathbf{u}, \mathbf{v})$ on $I_m \times I_m$ by

$$\phi_n(\mathbf{u}, \mathbf{v}) = a_n(\mathbf{r}, \mathbf{s}) \quad \text{if} \quad r_1 - 1 < nu_i \le r_i \quad \text{and} \quad s_i - 1 < nv_i \le s_i \quad \text{for} \quad i = 1, \cdots, m \tag{2.1}$$

where I_m is the m-dimensional unit cube, $\mathbf{u} = (u_1, \cdots, u_m)$ and $\mathbf{v} = (v_i, \cdots, v_m)$ belong to I_m and $(\mathbf{r}, \mathbf{s}) \in \Omega \times \Omega$. Furthermore, put

$$\phi_{1n}(\mathbf{u}) \equiv \int_{I_m} \phi_n(\mathbf{u}, \mathbf{v})\, d\mathbf{v} = a_{1n}(\mathbf{r}) \equiv \sum_{\mathbf{s}\in\Omega} a_n(\mathbf{r}, \mathbf{s})/n^m, \tag{2.2}$$

$$\phi_{1jn}(u_j) \equiv \int_{I_{m-1}} \phi_{1n}(\mathbf{u})\, d\mathbf{u}^{(j)} = a_{1jn}(r_j) \equiv \sum_{\mathbf{r}\in\Omega_j(r_j)} a_{1n}(\mathbf{r})/n^{m-1}, \tag{2.3}$$

for $j = 1, \cdots, m$ and

$$\bar{\phi}_n \equiv \int_{I_m} \phi_{1n}(\mathbf{u})\, d\mathbf{u} = \bar{a}_n \equiv \sum_{\mathbf{r}\in\Omega} a_{1n}(\mathbf{r})/n^m \tag{2.4}$$

with $d\mathbf{u}^{(j)} = du_1, \cdots, du_{j-1}\, du_{j+1}, \cdots, du_m$ and $\Omega_j(k) = \{\mathbf{r} \in \Omega \mid r_j = k\}$ for $\mathbf{u}$ and $\mathbf{r}$ with $r_i - 1 < nu_i \le r_i$, $i = i, \cdots, m$. Note that, by symmetry, we can change the role of $(\mathbf{u}, \mathbf{r})$ with $(\mathbf{v}, \mathbf{s})$ in (2.2) and (2.3). Now, we assume the following.

ASSUMPTION 2.3. There exists a nonconstant square integrable function

$\phi(\mathbf{u}, \mathbf{v})$ defined on $I_m \times I_m$ such that

$$\int_{I_{2m}} (\phi_n(\mathbf{u}, \mathbf{v}) - \phi(\mathbf{u}, \mathbf{v}))^2 \, d\mathbf{u} \, d\mathbf{v} \to 0 \quad \text{as} \quad n \to \infty. \tag{2.5}$$

Let ϕ_1, ϕ_{1j} and $\bar{\phi}$ be defined similarly to ϕ_{1n}, ϕ_{1jn} and $\bar{\phi}_n$, replacing ϕ_n by φ in (2.2), (2.3) and (2.4). From Assumption 2.3 one easily derives

$$\int_{I_m} (\phi_1(\mathbf{u}) - \phi_{1n}(\mathbf{u}))^2 \, d\mathbf{u} \to 0 \quad \text{as} \quad n \to \infty \tag{2.6}$$

and

$$\int_0^1 (\phi_{1j}(u) - \phi_{1jn}(u))^2 \, du \to 0, \quad j = 1, \cdots, m \quad \text{as} \quad n \to \infty. \tag{2.7}$$

Furthermore, it is assumed that the function ϕ is nondegenerate in the sense of Schach (1969):

ASSUMPTION 2.4. The function $\phi_1(\mathbf{u})$ is a nonconstant function.

The result we want to show is the following theorem.

THEOREM 2.1. *Suppose Assumptions* 2.1–2.4 *hold. Then, under* H, $n^{-3/2}S_n$ *is asymptotically equivalent to*

$$T_n = 2n^{-1/2} \sum_{i=1}^n \varphi(\mathbf{U}_i) + n^{1/2}\bar{a}_n$$

and hence $n^{-3/2}S_n$ *is asymptotically* $N(n^{1/2}\bar{a}_n, \sigma^2)$ *where*

$$\varphi(\mathbf{u}) = \phi_1(\mathbf{u}) - \bar{\phi} - \sum_{j=1}^m (\phi_{1j}(u_j) - \bar{\phi}),$$

$$\mathbf{U}_i = (U_{i1}, \cdots, U_{im}) \equiv (F_1(X_{i1}), \cdots, F_m(X_{im}))$$

and

$$\sigma^2 = 4 \int_{I_m} (\phi_1(\mathbf{u}) - \bar{\phi})^2 \, d\mathbf{u} - 4 \sum_{j=1}^m \int_0^1 (\phi_{1j}(u) - \bar{\phi})^2 \, du.$$

3. Proof of Theorem 2.1. Let $Z_{\mathbf{r}} = 1$ if $\mathbf{R}_i = \mathbf{r}$ for some i and $Z_{\mathbf{r}} = 0$ otherwise. Then the statistic S_n is rewritten as

$$\begin{aligned} S_n &= \sum_{\mathbf{r},\mathbf{s}\in\Omega} a_n(\mathbf{r}, \mathbf{s}) Z_{\mathbf{r}} Z_{\mathbf{s}} \\ &= \sum_{\mathbf{r},\mathbf{s}\in\Omega} b_n(\mathbf{r}, \mathbf{s}) Z_{\mathbf{r}} Z_{\mathbf{s}} + 2n \sum_{\mathbf{r}\in\Omega} (a_{1n}(\mathbf{r}) - \bar{a}_n) Z_{\mathbf{r}} + n^2 \bar{a}_n \\ &\equiv T_{1n} + T_{2n} + n^2 \bar{a}_n \quad \text{(say)} \end{aligned}$$

where

$$b_n(\mathbf{r}, \mathbf{s}) = a_n(\mathbf{r}, \mathbf{s}) - a_{1n}(\mathbf{r}) - a_{1n}(\mathbf{s}) + \bar{a}_n.$$

Thus, in order to prove Theorem 2.1, we may establish the lemmas below where it is shown that $n^{-3/2}T_{1n} = o_p(1)$ and that $n^{-3/2}T_{2n}$ is asymptotically equivalent to $T_n - n^{1/2}\bar{a}_n$. Essential in the proofs is Assumption 2.4, i.e. the nondegeneracy of ϕ in the sense of Schach (1969) who considered both the degenerate and the nondegenerate case in the univariate two-sample problem.

If ϕ is degenerate, $n^{-3/2}T_{1n}$ will not be negligible and it is difficult to derive the asymptotic distribution of S_n.

LEMMA 3.1. *Under the conditions of Theorem* 2.1, *we have*

$$E(T_{1n}) = O(n) \tag{3.1}$$

and

$$E(T_{1n}^2) = O(n). \tag{3.2}$$

PROOF. Let $\psi_n(\mathbf{u}, \mathbf{v})$, $b_{1n}(\mathbf{r})$ and $\psi_{1n}(\mathbf{u})$ be defined similarly to ϕ_n, a_{1n} and ϕ_{1n} with b_n replacing a_n in (2.1), (2.2) and (2.3). Furthermore, let

$$\psi(\mathbf{u}, \mathbf{v}) = \phi(\mathbf{u}, \mathbf{v}) - \phi_1(\mathbf{u}) - \phi_1(\mathbf{v}) + \bar{\phi}$$

and let $\psi_1(\mathbf{u})$ be defined similarly to ϕ_{1n} with ψ replacing ϕ_n in (2.2). Then, clearly, $b_{1n}(\mathbf{r}) = \psi_1(\mathbf{u}) = \psi_{1n}(\mathbf{u}) = 0$ for $\mathbf{u} \in I_m$ and $\mathbf{r} \in \Omega$. Put

$$E_k = \{(\mathbf{r}, \mathbf{s}) \in \Omega \times \Omega \mid r_i = s_i \text{ for exactly } k \text{ coordinates}\}. \tag{3.3}$$

Then we have

$$\#E_k = \binom{m}{k} n^m (n-1)^{m-k}, \quad k = 0, \cdots, m$$

and

$$Z_{\mathbf{r}} Z_{\mathbf{s}} = 0 \quad \text{if} \quad (\mathbf{r}, \mathbf{s}) \in E_1 \cdots E_{m-1}. \tag{3.4}$$

From (3.4)

$$\begin{aligned} E(T_{1n}) &= n^{-m+1} \textstyle\sum_{\mathbf{r}\in\Omega} b_n(\mathbf{r}, \mathbf{r}) + (n(n-1))^{-m+1} \sum_{(\mathbf{r},\mathbf{s})\in E_0} b_n(\mathbf{r}, \mathbf{s}) \\ &= n \int_{I_m} \psi_n(\mathbf{u}, \mathbf{u})\, d\mathbf{u} + n^{m+1}(n-1)^{-m+1} \int_{I_{2m}} \psi_n(\mathbf{u}, \mathbf{v})\, d\mathbf{u}\, d\mathbf{v} \\ &\quad - \textstyle\sum_{j=0}^{m-1} n^{j+1}(n-1)^{-m+1} \int_{I_{m+j}} \psi_n(\mathbf{u}, \mathbf{v}_{m-j}) d\mathbf{u}\, d\mathbf{v}_{m-j} \end{aligned}$$

where

$$\mathbf{v}_j = \mathbf{v}_j(\mathbf{u}) = \{\mathbf{v} \mid v_i = u_i \text{ for exactly } j \text{ coordiantes}\}.$$

From the degeneracy and the integrability of ψ_n, we have (3.1).

Consider a decomposition of $\Omega \times \Omega \times \Omega \times \Omega$ which is analogous to (3.3). Then, using a similar property to (3.4), we can get an integral representation of $E(T_{1n}^2)$. From the representation and the degeneracy and the square integrability of ψ_n, we can get (3.2). The detailed calculation is very long and is omitted.

LEMMA 3.2. *Under the conditions of Theorem* 2.1, $n^{-3/2}T_{2n}$ *is asymptotically equivalent to* $T_n - n^{1/2}\bar{a}_n$.

PROOF. By definition, we have

$$T_{2n} = 2n \sum_{i=1}^{n} c_n(\mathbf{R}_i)$$

where

$$c_n(\mathbf{r}) = a_{1n}(\mathbf{r}) - \bar{a}_n - \sum_{j=1}^{m} (a_{1jn}(r_j) - \bar{a}_n)$$

since

$$\sum_{\mathbf{r}\in\Omega} (a_{1jn}(r_j) - \bar{a}_n)Z_{\mathbf{r}} = 0.$$

Hence we may show that

$$E\{(\varphi(\mathbf{U}_1) - c_n(\mathbf{R}_1))^2\} = o(1) \tag{3.5}$$

and

$$E\{(\varphi(\mathbf{U}_1) - c_n(\mathbf{R}_1))(\varphi(\mathbf{U}_2) - c_n(\mathbf{R}_2))\} = o(n^{-1}). \tag{3.6}$$

The assertion (3.5) is proved by (2.6), (2.7) and a generalization of Behnen (1972, (3.19)). The assertion (3.6) can be proved by considering a conditional expectation, given $\mathbf{U}^{(1)}, \cdots, \mathbf{U}^{(m)}$ where $\mathbf{U}^{(j)}$ is the order statistics of $(U_{1j}, \cdots, U_{nj})$, and the properties

$$\sum_{\mathbf{r}\in\Omega_j(k)} c_n(\mathbf{r}) = 0, \quad j = 1, \cdots, m; \quad k = 1, \cdots, n$$

and

$$\int_{I_{m-1}} \varphi(\mathbf{u})\, d\mathbf{u}^{(j)} = 0, \quad j = 1, \cdots, m; \quad 0 < u_j < 1.$$

The detailed arguments are omitted because they are easy but laborious.

4. Asymptotic properties under contiguous alternatives. For linear rank statistics and their extended versions, the asymptotic normality under alternatives is proved by assuming some special models or by supposing that the alternative is fixed. See Konijn (1956) and the works referred to in Section 1. However, there are no commonly used nonparametric models for $m > 2$ and nonlinear statistics are very hard to be investigated under fixed alternatives. Therefore, in this section, we give a theorem on contiguous alternatives without assuming special distribution functions.

Consider a sequence of alternatives H_n; $F(\mathbf{x}) = F_n(\mathbf{x})$. Here F_n does not belong to H and H_n is supposed to be contiguous to H in the sense of Hájek and Šidák (1967). Then, since $n^{-3/2}S_n$ is asymptotically a sum of independent and identically distributed random variables under H, it satisfies the conditions of Behnen and Neuhaus (1975) and we can get the following theorem.

THEOREM 4.1. *Suppose that Assumptions* 2.1–2.4 *are satisfied, then* $n^{-3/2}S_n$ *is, under* H_n, *asymptotically* $N(n^{1/2}\bar{a}_n + k_n\sigma, \sigma^2)$ *where* σ^2 *is given in Theorem* 2.1

and

$$k_n = n^{1/2}E_n(h_n(Z) - E(h_n(Z))),$$

$$h_n(z) = \begin{cases} z & \text{if } |z| < n^{1/6} \\ 0 & \text{otherwise,} \end{cases}$$

$$Z = 2\,\varphi(\mathbf{U}_1)/\sigma$$

and where E_n and E are calculated under H_n and H, respectively.

REFERENCES

AL-SAADI, S. D. and YOUNG, D. H. (1981). Critical values of some nonparmetric statistics used for tests of total stochastic independence. *J. Statist. Comput. Simulation* **12** 217–224.

BEHNEN, K. (1971). Asymptotic optimality and ARE of certain rank-order tests under contiguity. *Ann. Math. Statist.* **42** 325–329.

BEHNEN, K. (1972). A characterization of certain rank-order tests with bounds for the asymptotic relative efficiency. *Ann. Math. Statist.* **43** 1839–1851.

BEHNEN, K. and NEUHAUS, G. (1975). A central limit theorem under contiguous alternatives. *Ann. Statist.* **3** 1349–1353.

BERAN, R. J. (1972). Rank spectral processes and tests for serial dependence. *Ann. Math. Statist.* **43** 1749–1766.

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

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic, New York.

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

PURI, M. L. and SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis.* Wiley, New York.

RUYMGAART, F. H. (1974). Asymptotic normality of nonparametric tests for independence. *Ann. Statist.* **2** 892–910.

RUYMGAART, F. H., SHORACK, G. R. and ZWET, W. R. VAN. (1972). Asymptotic normality of nonparametric tests for independence. *Ann. Math. Statist.* **43** 1122–1135.

SCHACH, S. (1969). The asymptotic distribution of some non linear functions of the two-sample rank vector. *Ann. Math. Statist.* **40** 1011–1020.

SHIRAHATA, S. (1974). Locally most powerful rank tests for independence. *Bull. Math. Statist.* **16** 11–21.

SINHA, B. K. and WIEAND, H. S. (1977). Multivariate nonparametric tests for independence. *J. Multivariate Anal.* **7** 572–583.

WAKIMOTO, K. and SHIRAHATA, S. (1984). A coefficient of concordance based on the chart of linked lines. Unpublished manuscript.

DEPARTMENT OF APPLIED MATHEMATICS
FACULTY OF ENGINEERING SCIENCE
OSAKA UNIVERSITY
TOYONAKA, OSAKA 560, JAPAN

DEPARTMENT OF MATHEMATICS AND STATISTICS
COLLEGE OF LIBERAL ARTS AND SCIENCE
OKAYAMA UNIVERSITY
OKAYAMA 700, JAPAN