Ann. Inst. Statist. Math.
Vol. 41, No. 3, 415–427 (1989)

ASYMPTOTIC NORMALITY OF DOUBLE-INDEXED LINEAR PERMUTATION STATISTICS

DINH TUAN PHAM[1], JOACHIM MÖCKS[2] AND LOTHAR SROKA[2]

[1]*Laboratory TIM 3, CNRS, IMAG. University of Grenoble, B.P. 53X, 38041 Grenoble Cedex, France*
[2]*Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 6900 Heidelberg, FRG*

(Received May 9, 1987; revised October 7, 1988)

Abstract. The paper provides sufficient conditions for the asymptotic normality of statistics of the form $\Sigma a_{ij} b_{R_iR_j}$, where a_{ij} and b_{ij} are real numbers and R_i is a random permutation.

Key words and phrases: Asymptotic normality, linear permutation statistic.

1. Introduction

Let a_{ij} and b_{ij}, $i, j = 1,\dots, n$ be real numbers depending on n. We are interested in statistics of the form $\sum_{i \neq j} a_{ij} b_{R_iR_j}$, where $R_1,\dots, R_n$ is a random permutation of $1,\dots, n$, all permutations being equally likely. The a_{ij} and b_{ij} will be referred to as scores. The statistics are called "double-indexed linear permutation statistics" so as to distinguish them from single-indexed ones, being of the form $\Sigma a_i b_{R_i}$. Double-indexed permutation statistics have been first considered by Daniels (1944) who gave sufficient conditions for their asymptotic normality as $n \to \infty$. Later they appeared in various contexts: as a measure of association between graphs (Friedman and Rafsky (1983)), as a measure of spatial correlation (Cliff and Ord (1981)) and as a multivariate two-sample test statistic (Friedman and Rafsky (1979) and Schilling (1986)). In these contexts, the scores are symmetric ($a_{ij} = a_{ji}$, $b_{ij} = b_{ji}$) while in Daniels' paper they are asymmetric. Therefore, only these two types of scores will be considered here. Note that this restriction is needed for just one set of the scores, since the other one, the a_{ij}, say, may then be symmetrized by $(a_{ij} + a_{ji})/2$, or asymmetrized by $(a_{ij} - a_{ji})/2$, without changing the value of the statistic.

The purpose of the present paper is to find sufficient conditions for the asymptotic normality of the above statistics. We shall show that under certain circumstances, this problem can be reduced to that of a single-indexed linear permutation statistic, which has been extensively studied in the literature (see, for example, Puri and Sen (1971), pp. 70–76). Daniels' conditions will be seen to be sufficient to permit this reduction and also to ensure the asymptotic normality of the corresponding single-indexed statistic. However, this reduction is not possible if row (or column) sums of one set of scores are nearly constant. Such situations may frequently arise when the scores correspond to a sparse graph. This case will be studied in some detail and sufficient conditions for asymptotic normality are provided. Friedman and Rafsky (1983), on the other hand, claimed that Daniels' conditions for asymptotic normality can be weakened to $\Sigma a_{ij} = 0$, $\left(\Sigma a_{ij}a_{ik}a_{il}\right)^2 / \left(\Sigma a_{ij}a_{ik}\right)^3 \to 0$, plus similar conditions on the b_{ij}. However, they did not give an explicit proof for their results and merely stated that they could be obtained by a similar argument as in Daniels (1944). It is not at all clear to us how this can be achieved. Moreover, they applied their conditions to raw scores a_{ij} and b_{ij} for which the condition $\Sigma a_{ij} = 0 = \Sigma b_{ij}$ does not hold. Of course, the scores can be centered to satisfy this condition, but then both the numerator and denominator in Friedman and Rafsky's conditions would change. Centering would make them vanish, if the row sums of the scores are constant, unless the summation is restricted to distinct subscripts. There are situations for which the asymptotic distribution of the statistic is not normal despite Friedman and Rafsky's conditions being satisfied, whether in terms of raw or centered scores (see Section 2). Nevertheless, our results show that under certain conditions, reducing the double-indexed case to the single-indexed one is possible, and then Friedman and Rafsky's conditions suffice for asymptotic normality.

Since the definition of the present statistics does not involve the a_{ii} and b_{ii}, we will assume, for convenience, that they are zero. Further, Σ will denote the summation over all subscripts and Σ' the same summation but restricted to distinct values of them only. Upper case letters will be used to indicate score variables with "randomly permuted" indexes, for example, B_{ij} denotes $b_{R_iR_j}$.

2. Reduction to the single-indexed case

The main idea is to decompose the space of scores in a way similar to the analysis of variance. Let E denote the space of square matrices of order n with vanishing diagonal. Let E_0, E_1 and E_2 be the subspaces of E consisting of matrices with constant off-diagonal elements, with off-diagonal elements of the general form $c_i + d_j$, $\Sigma c_i = \Sigma d_j = 0$, and with vanishing row and column sums, respectively. Since only symmetric and asymmetric scores are considered, we introduce the subspaces E_k^+, E_k^- of E_k, $k = 0, 1, 2$, consisting of symmetric and asymmetric matrices, respectively. Clearly,

E_0, E_1 and E_2 are orthogonal to each other (identifying E with $\mathbb{R}^{n(n-1)}$), and so are the corresponding symmetric or asymmetric subspaces. Put

$$a'_{ij} = a_{ij} - \frac{1}{n(n-1)} \Sigma a_{kl} \quad \text{if} \quad i \neq j, \qquad = 0 \quad \text{otherwise}, \tag{2.1}$$
$$a'_{i+} = \sum_j a'_{ij}, \qquad a'_{+j} = \sum_i a'_{ij},$$

so that $\Sigma a'_{ij} = \Sigma a'_{i+} = a'_{+j} = 0$. Define, in the symmetric case,

$$a^*_{ij} = a'_{ij} - \frac{1}{n-2}(a'_{i+} + a'_{+j}) \quad \text{if} \quad i \neq j, \qquad = 0 \quad \text{otherwise}. \tag{2.2}$$

Then, since $\Sigma a'_{+j} = 0$,

$$\sum_j a^*_{ij} = a'_{i+} - \frac{n-1}{n-2} a'_{i+} - \frac{1}{n-2} \sum_{j \neq i} a'_{+j} = 0,$$

and hence, the matrix (a^*_{ij}) belongs to E_2^+. Thus, any symmetric matrix (a_{ij}) in E can be written as a sum of elements of E_0^+, E_1^+ and E_2^+:

$$a_{ij} = \frac{1}{n(n-1)} \Sigma a_{kl} + \frac{1}{n-2}(a'_{i+} + a'_{+j}) + a^*_{ij}, \qquad i \neq j.$$

Similarly, in the asymmetric case, the matrix (a^*_{ij}) defined by

$$a^*_{ij} = a_{ij} - \frac{1}{n}(a'_{i+} + a'_{+j}) \quad \text{if} \quad i \neq j, \qquad = 0 \quad \text{otherwise}, \tag{2.2'}$$

belongs to E_2^-. Thus, any asymmetric matrix (a_{ij}) in E can be written as a sum of elements of E_1^- and E_2^-:

$$a_{ij} = \frac{1}{n}(a'_{i+} + a'_{+j}) + a^*_{ij}, \qquad i \neq j,$$

since $a_{ij} = a'_{ij}$.

Define b'_{ij}, b^*_{ij} and b'_{i+}, b'_{+j} in the same way. It follows that in the symmetric case,

$$\Sigma a_{ij} B_{ij} = \Sigma a^*_{ij} B^*_{ij} + \frac{2}{n-2}\left(\Sigma a'_{i+} B'_{i+}\right) + \text{const}. \tag{2.3}$$

and in the asymmetric case,

(2.3′) $$\Sigma a_{ij} B_{ij} = \Sigma a_{ij}^* B_{ij}^* + \frac{2}{n} \left(\Sigma a'_{i+} B'_{i+} \right).$$

Let T_n, U_n and V_n denote the left-hand side and the first and second terms of the right-hand side of (2.3) or (2.3′), respectively, and let t_n^2, u_n^2 and v_n^2 be their variances. One has $EU_n = EV_n = 0$ and

$$(T_n - ET_n)/v_n = (U_n/u_n)(u_n/v_n) + (V_n/v_n).$$

Thus, if $u_n/v_n \to 0$, the variance of $(T_n - ET_n)/v_n$, which is t_n^2/v_n^2 tends to 1 and hence, $(T_n - ET_n)/t_n$ has the same limiting distribution as V_n/v_n. Note that V_n is a single-indexed linear permutation statistic. Similarly, if $v_n/u_n \to 0$, $(T_n - ET_n)/t_n$ has the same limiting distribution as U_n/u_n.

The variance of the linear permutation statistic $\Sigma a'_{+i} B'_{+i}$ is well known (see, for example, Puri and Sen (1971), p. 73). We have, in the symmetric case,

(2.4) $$v_n^2 = 4\left(\Sigma a'^2_{i+}\right)\left(\Sigma b'^2_{i+}\right)/[(n-2)^2(n-1)],$$

and in the asymmetric case,

(2.4′) $$v_n^2 = 4\left(\Sigma a'^2_{i+}\right)\left(\Sigma b'^2_{i+}\right)/[n^2(n-1)].$$

In order to compute u_n^2, observe that $E(B_{ij} B_{kl})$ does not depend on the particular indexes i, j, k and l, but only on whether and how many of them are tied. Hence,

$$u_n^2 = \frac{2}{n(n-1)} \left(\Sigma a_{ij}^{*2}\right)\left(\Sigma b_{ij}^{*2}\right) + \frac{4}{n(n-1)(n-2)} \left(\Sigma' a_{ij}^* a_{ik}^*\right)\left(\Sigma' b_{ij}^* b_{ik}^*\right)$$
$$+ \frac{1}{n(n-1)(n-2)(n-3)} \left(\Sigma' a_{ij}^* a_{kl}^*\right)\left(\Sigma' b_{ij}^* b_{kl}^*\right).$$

Note that $\Sigma' a_{ij}^* a_{ik}^* = -\Sigma a_{ij}^{*2}$ since $a_{i1}^* + \cdots + a_{in}^* = 0$. Similarly, $\Sigma' a_{ij}^* a_{kl}^* = \Sigma a_{ij}^{*2} + \Sigma a_{ij}^* a_{ji}^*$. Thus, in the symmetric case,

(2.5) $$u_n^2 = \frac{2}{n(n-3)} \left(\Sigma a_{ij}^{*2}\right)\left(\Sigma b_{ij}^{*2}\right),$$

and in the asymmetric case,

(2.5′) $$u_n^2 = \frac{2}{(n-1)(n-2)} \left(\Sigma a_{ij}^{*2}\right)\left(\Sigma b_{ij}^{*2}\right).$$

From the above arguments, one thus gets the following result.

LEMMA 2.1. *Under the condition*

$$\text{(R0)} \qquad n\left[\left(\Sigma a_{ij}^{*2}\right)\left(\Sigma b_{ij}^{*2}\right)\right]/\left[\left(\Sigma a_{i+}'^{2}\right)\left(\Sigma b_{i+}'^{2}\right)\right] \to 0 \quad as \quad n \to \infty ,$$

the asymptotic distribution of $T_n - ET_n$, *where* $T_n = \Sigma a_{ij} B_{ij}$, *is the same as that of* $(2/n)\Sigma a_{i+}' B_{i+}'$.

Note that by (2.2) and (2.2′) and the Pythagorean theorem, Σa_{ij}^{*2} may be obtained by $\Sigma a_{ij}'^{2} - 2\left(\Sigma a_{i+}'^{2}\right)/(n-2)$ in the symmetric case or $\Sigma a_{ij}'^{2} - 2\left(\Sigma a_{i+}'^{2}\right)/n$ in the asymmetric case. Hence, a stronger condition, but easier to check, is obtained from (R0) by just replacing a_{ij}^{*}, b_{ij}^{*} with a_{ij}', b_{ij}'.

In many applications, one of the score set, the b_{ij} say, satisfies

$$\left(\Sigma b_{ij}^{*2}\right)/\left(\Sigma b_{i+}'^{2}\right) = O(n^{-1}), \quad \left(\Sigma |b_{i+}'|^{r}\right)/\left(\Sigma b_{i+}'^{2}\right)^{r/2} = O(n^{1-r/2}) .$$

For example, this holds when $b_{ij} = b_i b_j$ (for $i \neq j$), where b_i is the indicator function of a subset of m elements of $\{1,\ldots,n\}$ with n/m, $n/(n-m)$ bounded. Then (R0) is equivalent to

$$\text{(R1)} \qquad \left(\Sigma a_{ij}'^{2}\right)/\left(\Sigma a_{i+}'^{2}\right) \to 0 ,$$

and the single-indexed linear permutation statistic $\Sigma a_{i+} B_{i+}$ is asymptotically normal if $\left(\Sigma |a_{i+}'|^{r}\right)/\left(\Sigma a_{i+}'^{2}\right)^{r/2} \to 0$ for some $r > 2$ (Hoeffding (1951)). Note that if positive raw scores a_{ij} are used in place of centered scores a_{ij}', the last condition, with $r = 3$, is just Friedman and Rafsky's (1983) condition. Since $a_{i+}' = a_{i+} - \Sigma a_{i+}/n$ by (2.1), implying $\Sigma a_{i+}'^{2} = \Sigma a_{i+}^{2} - \left(\Sigma a_{i+}\right)^{2}/n$, it can be seen from the triangular inequality that one may replace centered scores by raw scores if $\limsup \left(\Sigma a_{i+}\right)^{2}/\left(n \Sigma a_{i+}^{2}\right) < 1$. Under the last condition, (R1) is equivalent to $\left(\Sigma a_{ij}^{2}\right)/\left(\Sigma a_{i+}^{2}\right) \to 0$, being a mild requirement.

Remarks. (1) Since $\Sigma' a_{ij}' a_{ik}' = \Sigma a_{i+}'^{2} - \Sigma a_{ij}'^{2}$, Daniels' conditions, namely $\liminf_{n\to\infty} \left(\Sigma' a_{ij}' a_{ik}'\right)/(n^{3} \max a_{ij}'^{2}) > 0$ and the same for the b_{ij}, imply $\left(\Sigma a_{ij}'^{2}\right)/\left(\Sigma a_{i+}'^{2}\right) = O(n^{-1})$ and the same for the b_{ij}. Hence, (R0) with a_{ij}', b_{ij}' in place of a_{ij}^{*}, b_{ij}^{*}, is fulfilled. Daniels' conditions are also sufficient to ensure the asymptotic normality of $\Sigma a_{i+}' B_{i+}'$ (compare Puri and Sen (1971), Chapter 3.4). It is clear that they are by far too strong, since only (R0) and a further mild condition on the a_{i+}' and b_{i+}' suffice to ensure asymptotic normality.

(2) The above derivations crucially depend on the fact that $u_n/v_n \to 0$.

If this is not true, asymptotic normality may not hold. For example, suppose n is even and take $a_{ij} = 1$ if $i \neq j$ and are both in I or I', $= 0$ otherwise, where I is a subset of $\{1,\dots,n\}$ with $n/2$ elements and I' is the complement of I. The row sums of these scores are constant, and the centered scores, for $i \neq j$, are $a'_{ij} = n/[2(n-1)]$ if i, j are both in I or I', $= -(n/2-1)/(n-1)$ otherwise. Thus, $a'_{ij} = a_i a_j + 1/[2(n-1)]$, $i \neq j$, where $a_i = 2^{-1/2}$ or $-2^{-1/2}$ according to whether i is in I or I'. Define b_{ij} and b_i in the same way with respect to another subset J of $\{1,\dots,n\}$. Then $\Sigma a_{ij} B_{ij} = (\Sigma a_i B_i)^2 - \Sigma a_i^2 B_i^2 + (1/4)n/(n-1)$. By using the results on the asymptotic normality of the single-indexed permutation statistic, it can be seen that $4/n$ times the above random variable is asymptotically distributed like $Z^2 - 1$, Z being a standard normal variate. Thus, asymptotic normality does not hold in this case. Note that the a_{ij}, as well as a'_{ij} satisfy Friedman and Rafsky's (1983) conditions (for centered scores, summation over distinct subscripts is assumed, since otherwise an indeterminate expression 0/0 arises). Note that (R0) is not satisfied here, since $a'_{i+} = b'_{i+} = 0$ for all i.

(3) Shapiro and Hubert (1979) have derived conditions for the asymptotic normality of the permutation statistic by relating it to a statistic of the form $W = \Sigma' d_{ij} a(X_i, X_j)$, where the X_i are independent random variables. Their conditions correspond to the case $u_n/v_n \to 0$ (condition AE of their Lemma 2.1) and are rather strong (conditions A3–A4). It is surprising to us that they did not center the scores d_{ij}. Centering seems to be indispensable, since the variance of the permutation statistic $\Sigma' d_{ij} A_{ij}$, but not that of W, is zero if the d_{ij} are constant. The proof of their Theorem 3.1 would not be valid if the d_{ij} were constant (or almost so) and not centered.

(4) It can be easily seen that the covariance of $\Sigma a^*_{ij} B^*_{ij}$ with $\Sigma a'_{i+} B'_{i+}$ vanishes. This immediately yields the variance of $\Sigma a_{ij} B_{ij}$:

$$(2.6)\quad t_n^2 = 4\left(\Sigma a'^2_{i+}\right)\left(\Sigma b'^2_{i+}\right)/[(n-2)^2(n-1)] + 2\left(\Sigma a^{*2}_{ij}\right)\left(\Sigma b^{*2}_{ij}\right)/[n(n-3)],$$

in the symmetric case and

$$(2.6')\quad t_n^2 = 4\left(\Sigma a'^2_{i+}\right)\left(\Sigma b'^2_{i+}\right)/[n^2(n-1)] + 2\left(\Sigma a^{*2}_{ij}\right)\left(\Sigma b^{*2}_{ij}\right)/[(n-1)(n-2)],$$

in the asymmetric case.

3. The irreducible case

We first consider the case where the ratio of variances $v_n/u_n \to 0$. In this case, $t_n/u_n \to 1$ and the statistic $(T_n - ET_n)/t_n$ has the same limiting distribution as U_n/u_n. We may drop the sign $*$ in b^*_{ij} by adding the condition that row and column sums of these scores vanish. Then, a^*_{ij} may be replaced by a_{ij} as well, since this does not alter the value of the statistic.

For ease of reading, proofs are relegated to the end of this section.

THEOREM 3.1. *Suppose that the scores satisfy*

(A1) $$\max\Big(\sum_j |a_{ij}|\Big) = O\big[\big(\Sigma a_{ij}^2\big)/(n \max |a_{ij}|)\big],$$

(A2) $$\big(\Sigma a_{ij}^2\big)/(n^2 \max a_{ij}^2) \to 0 \quad \textit{as} \quad n \to \infty,$$

(B0) $$\sum_j b_{ij} = 0 \quad \textit{for all} \quad i,$$

(B1) $$\Sigma |b_{ij}|^r/n^2 = O\big[\big(\Sigma b_{ij}^2/n^2\big)^{r/2}\big], \qquad r = 3, 4, \ldots .$$

Then, $\big(\Sigma a_{ij} B_{ij}\big)/\big[2\big(\Sigma a_{ij}^2\big)\big(\Sigma b_{ij}^2\big)/n^2\big]^{1/2}$ *converges in distribution to a standard normal variate.*

Note that Cliff and Ord (1981) gave a similar result, but without an explicit proof. Conditions (B0) and (B1) reduce to those of Cliff and Ord if b_{ij} is of the form $b_i b_j$ for $i \neq j$. Our conditions (A1) and (A2) are weaker than their conditions on the a_{ij}, namely $\max_i\Big(\sum_j a_{ij}\Big)$ bounded and $\Sigma a_{ij}^2/n \to \gamma > 0$ ($a_{ij} \geq 0$ and sum to n). Indeed, from the last conditions, $\max a_{ij} \leq \max_i\Big(\sum_j a_{ij}\Big)$, which is bounded, and $\max a_{ij} \geq \big(\Sigma a_{ij}^2\big)/\big(\Sigma a_{ij}\big)$ which tends to $\gamma > 0$, hence, (A1) and (A2) hold.

As a related result, we mention Theorem 4.1.2 of Bloemena (1964), which gives a sufficient condition for the conditional asymptotic normality of the two sample test statistic of Schilling (1986) and Henze ((1988), Propositions 2.1 and 2.2).

Conditions (A1) and (A2) suffice for most applications. The first one requires that the row sums are roughly evenly distributed, the second one that the scores are "sparse", in the sense that most of them are zero or small with respect to their maximum absolute value. This guarantees to exclude that a_{ij} takes the form of a product $a_i a_j$, which, as was seen earlier, can lead to a non-normal limiting distribution.

By the argument given at the beginning of this section, the above result still holds if in (A1) and (A2) the a_{ij} are replaced by centered scores a'_{ij} or reduced scores a^*_{ij}. Call the corresponding conditions (A′1), (A′2) and (A*1), (A*2). Since $\Sigma a'^2_{ij} = \Sigma a_{ij}^2 - \big(\Sigma a_{ij}\big)^2/[n(n-1)]$ and $\big|\Sigma a_{ij}/n\big| < \max\Big(\sum_j |a_{ij}|\Big)$, it can be seen that (A1) and (A2) imply (A′1) and (A′2). Similarly, (A′1) and (A′2) imply (A*1) and (A*2). The converse may not hold. Thus, (A*1) and (A*2) are weaker than (A1) and (A2), but in practice the latter conditions are easier to check. Similarly, it would be convenient to have conditions on the centered scores and not on the reduced scores, as required by (B0). Thus, consider instead of (B0)

(B′0) $$\Sigma b_{ij} = 0 ,$$

meaning that the b_{ij} are centered. Since (A′1) and (A′2) already imply $(\Sigma a_{i+}'^2)/(n\Sigma a_{ij}^{*2}) \to 0$, from (2.4) and (2.5) or (2.4′) and (2.5′), the condition $v_n/u_n \to 0$ holds if

(B′1) $$\Sigma b_{i+}^2 = O(\Sigma b_{ij}^2) .$$

From (2.2) or (2.2′) with the b's in place of the a's and using the triangular inequality (in the l^r norm), it can be seen that condition (B1) holds for b_{ij}^* if it holds for b_{ij} and $\Sigma|b_{i+}|^r/n = O[(\Sigma b_{ij}^2)^{r/2}]$. Put $x_i = |b_{i+}|/(\Sigma b_{i+}^2)^{1/2}$, then $x_i \le 1$ and hence, $\Sigma x_i^r \le \Sigma x_i^2 = 1$, giving $\Sigma|b_{i+}|^r \le (\Sigma b_{i+}^2)^{r/2}$. Thus, (B′1) implies $\Sigma|b_{i+}|^r = O[(\Sigma b_{ij}^2)^{r/2}]$. *Hence, in Theorem* 3.1, (B0) *can be replaced by* (B′0) *and* (B′1).

The general case where the ratio u_n/v_n neither converges to 0 nor to infinity is difficult to treat. Nevertheless, under rather strong conditions on the scores a_{ij}, asymptotic normality still holds.

THEOREM 3.2. *Suppose that the scores satisfy* (B′0) *and*

(A′0) $$\Sigma a_{ij} = 0 ,$$

(A′1) $$\max\left(\sum_j |a_{ij}|\right) = O(\max |a_{ij}|) ,$$

(A′2) $$\liminf (\Sigma a_{ij}^2)/(n \max a_{ij}^2) > 0 ,$$

(B″1) $$\limsup (\Sigma b_{i+}^2/n)/(\Sigma b_{ij}^2) < 1 .$$

Then, $(\Sigma a_{ij} B_{ij})/w_n$, *where*

$$w_n^2 = 4(\Sigma' a_{ij} a_{ik})(\Sigma' b_{ij} b_{ik})/n^3 + 2(\Sigma a_{ij}^2)(\Sigma b_{ij}^2)/n^2 ,$$

converges in distribution to a standard normal variate.

Remark. Under (A′0) and (B′0), one has $\Sigma' a_{ij} a_{ik} = \Sigma a_{i+}^2 - \Sigma a_{ij}^2$ and a similar equality for the b_{ij}. Hence,

$$w_n^2 = [4(n-2)/n^4]\left(\Sigma a_{i+}^2\right)\left(\Sigma b_{i+}^2\right)$$
$$+ (2/n^2)\left(\Sigma a_{ij}^2 - \frac{2}{n}\Sigma a_{i+}^2\right)\left(\Sigma b_{ij}^2 - \frac{2}{n}\Sigma b_{i+}^2\right) + (4/n^3)\left(\Sigma a_{ij}^2\right)\left(\Sigma b_{ij}^2\right).$$

Conditions (A′1), (A′2) and (B″1) imply that the last term in the above right-hand side is negligible with respect to the second. It can then be seen that w_n is equivalent to t_n, as given by (2.6) or (2.6′).

PROOF OF THEOREM 3.1. The method of proof consists of showing that the moments of the random variable $\left(\Sigma a_{ij} B_{ij}\right)/(\|a\|\,\|b\|/n)$ where $\|a\| = \left(\Sigma a_{ij}^2\right)^{1/2}$, $\|b\| = \left(\Sigma b_{ij}^2\right)^{1/2}$, tend to those of a normal variate with zero mean and variance 2. We shall assume that the scores are symmetric. The proof for asymmetric scores needs only minor modifications to account for the fact that $\Sigma a_{ij} a_{ji} = -\|a\|^2$ and not $\|a\|^2$ as in the symmetric case. For convenience we shall normalize the a_{ij} so that $\max |a_{ij}| = 1$, and use some terminology of graph theory. A graph G may be defined as a finite collection of pairs (μ, ν) of integers, called edges. The integers occurring in those pairs are called nodes. We allow graphs with multiple edges, meaning that the above pairs need not be distinct. The notation $|G|$ will denote the number of nodes in G, which, for convenience, will be labelled as $1,\dots,|G|$. For any partition of the set $\{1,\dots,r\}\times\{1,2\}$, one may associate a graph with r edges (distinct or not) as follows. For $k = 1,\dots,r$, let P_{μ_k} and P_{ν_k} be the sets of the partition containing the points $(k,1)$ and $(k,2)$, respectively, then the k-th edge of the associated graph is $(\mu,\nu)_k = (\mu_k, \nu_k)$. Call $G(r)$ the set of all graphs constructed this way. For a graph G, we put

$$\Sigma(a, G) = \sum_{i_1,\dots,i_{|G|}}' \prod_{(\mu,\nu)\in G} a_{i_\mu i_\nu},$$

and similarly for $\Sigma(b, G)$. Then it can be seen that the moment of order r of $\Sigma a_{ij} B_{ij}$ is the sum of terms of the form

$$M(G) = E\left(\prod_{(j,k)\in G} B_{i_j i_k}\right)\Sigma(a, G) = \frac{(n-|G|)!}{n!}\Sigma(b, G)\,\Sigma(a, G),$$

over all graphs G in $G(r)$. Note that because of the condition $a_{ii} = 0$, we only need to consider graphs where no node is linked to itself.

Let ν be a terminal node of G, in the sense that there is just one (single) edge linked to it. This edge could be of the form (μ,ν) or (ν,μ), $\mu \neq \nu$. We shall consider only the first case since the argument for the second case is similar. Then, from the fact that column sums of the b_{ij} are zero,

$$\Sigma(b, G) = -\sum_{1 \le \lambda \le |G|, \lambda \ne \mu, \nu} \Sigma(b, G_\lambda),$$

where G_λ is the graph obtained from G by replacing the edges (μ, ν) by (λ, ν). Repeating this argument, it is seen that $\Sigma(b, G)$ can be written as a sum of a number (not depending on n) of terms of the form $\Sigma(b, G')$, where G' is a graph obtained from G by successively replacing an edge linking to a terminal node by another edge linking to a different node in the graph, until there are no terminal nodes left. If t is the number of terminal nodes in G, then clearly G' can have at most $|G| - (t{:}2)$ nodes where $(t{:}2)$ denotes the smallest integer greater than $t/2$. On the other hand, by Hölder's inequality, with the nodes of G' relabelled as $1, \ldots, |G'|$,

$$\left| \Sigma(b, G') \right| \le \prod_{(\mu,\nu) \in G'} \left(\sum_{i_1, \ldots, i_{|G'|}} |b_{i_\mu i_\nu}|^r \right)^{1/r},$$

r being the number of edges of G', the same as that of G. The sum in the above parentheses equals $n^{|G'|} \Sigma |b_{il}|^r / n^2$, which is $O(n^{|G'|} \|b\|^r / n^r)$ by (B1). Hence, $\Sigma(b, G) = O(n^{|G| - (t:2) - r} \|b\|^r)$. If $r = 2p$ and G is a union of p disjoint subgraphs $G_1, \ldots, G_p$ of the form $\{(\mu, \nu), (\nu, \mu)\}$ or $\{(\mu, \nu), (\mu, \nu)\}$, one further has $\Sigma(b, G) = \prod_{q=1}^{p} \Sigma(b, G_q) + o(\|b\|^{2p}) = \|b\|^{2p}[1 + o(1)]$. Indeed, the remainder term is a sum of terms of the form $\Sigma(b, G')$ with G' having the same number of edges as G but $|G'| < |G|$. Call $G^*(2p)$ the set of all such graphs G. We shall show below that for all graphs G with r edges and t terminal nodes, $n^{-(t:2)} \Sigma(a, G) / \|a\|^r$ tends to 1, if r is even and $G \in G^*(r)$, and to 0 otherwise. (Note that in the first case, $|G| = r$ and $t = 0$ necessarily.) This would imply that $M(G)/(\|a\| \, \|b\| / n)^r \to 0$ unless r is even and $G \in G^*(r)$, in which case the limit is 1. Now, the number of elements of $G^*(2p)$ is 2^p times the number of ways one can pick p pairs of balls from an urn containing $2p$ balls, that is, $(2p)!/p!$. This equals precisely the $2p$-th moment of a normal variate with mean zero and variance 2, and Theorem 3.1 is then proved.

We now show the above assertion. Observe that if G is the union of two disjoint subgraphs G' and G'', say, then $\Sigma(|a|, G) \le \Sigma(|a|, G') \, \Sigma(|a|, G')$ where $\Sigma(|a|, G)$ is defined in the same way as $\Sigma(a, G)$ with $|a_{ij}|$ in place of a_{ij}. Thus, we only need to show that for any connected graph G, that is a graph which cannot be decomposed as a union of disjoint subgraphs, with t terminal nodes and r edges, $n^{(t:2)} \Sigma(|a|, G) / \|a\|^r \to 0$, unless $G \in G^*(2p)$. Consider first the case that G is a tree, that is a connected graph without cycles, where a cycle is a sequence of edges of the form $(i_1, i_2), \ldots, (i_{k-1}, i_k), (i_k, i_1)$. Then, $\Sigma(|a|, G) = O[(n \|a\|^2)^{|G|-1}] = O(\|a\|/n)^{|G|-2} \|a\|^{|G|}$. This is seen by summing first with respect to the subscript i_ν, ν being a terminal node of G, showing that the above left-hand side is bounded by

$\max_j \left(\sum_i |a_{ij}| \right) \Sigma(|a|, G')$, G' being the subtree of G without the edge linking to v, then using (A1) and repeating the argument. The above result still holds for a connected graph, since by suppressing redundant edges, we get a tree with the same number of nodes and $\Sigma(|a|, G)$ can only decrease. It follows that for a connected graph,

$$n^{-(t:2)} \Sigma(|a|, G)/\|a\|^r = O[n^{-(t:2)} \|a\|^{|G|-r} (\|a\|/n)^{|G|-2}] .$$

But $2(|G| - t) + t \leq 2r$, or equivalently $|G| - r \leq t/2$, hence the above right-hand side, by (A2) tends to 0 as n goes to infinity, unless $t = 0$, $|G| = r = 2$. This desired result follows from the fact that a general graph can always be written as a union of disjoint connected subgraphs.

PROOF OF THEOREM 3.2. The argument is similar to that given in the proof of Theorem 3.1 and we shall use the same notation. Note that if the a_{ij} satisfy (A′1) and (A′2), then a'_{ij} also satisfy the same conditions. Thus, we may replace a_{ij} by a'_{ij}, which for convenience will be denoted again by a_{ij}, and will be normalized such that $\max |a_{ij}| = 1$.

We need to show that the moments of $\left(\Sigma a_{ij} B_{ij}\right)/w_n$ tend to those of a standard normal variate. The moment of order r of $\Sigma a_{ij} B_{ij}$ is the sum of $M(G) = [(n - |G|)!/n!] \, \Sigma(a, G) \, \Sigma(b, G)$ over all graphs G in $G(r)$. Call an edge isolated if it does not share any node with others. By the same argument as in the proof of Theorem 3.1, $\Sigma(b, G)$ can be written as a sum of terms of the form $\Sigma(a, G')$ where G' is obtained from G by replacing each isolated edge by another edge sharing a common node with it, and $\Sigma(b, G') = O(n^{|G'|-r} \|b\|^r)$. Note that if G has s isolated edges, then $|G'|$ can not exceed $|G| - (s{:}2)$, hence, $\Sigma(a, G) = O(n^{|G'|-(s:2)-r} \|b\|^r)$. On the other hand, by (A″1), $\Sigma(|a|, G) = O(n)$ for any connected graph, and hence, if G is the union of c disjoint connected subgraphs, then $\Sigma(a, G) = O(n^c)$. However, if s among them are actually isolated edges, then, as above, $\Sigma(a, G)$ can be written as a sum of terms of the form $\Sigma(a, G')$ with G' having no isolated edges and at most $c - (s{:}2)$ connected disjoint subgraphs. From this and (A′2), $\Sigma(|a|, G)/\|a\|^r = O(n^{c-(s:2)-r/2})$, r being the number of edges of G. But $2(c - s) + s \leq r$ or equivalently $c \leq r/2 + s/2$, with equality if and only if all connected subgraphs of G have at most 2 edges. Therefore, $M(G) = o[(\|a\| \, \|b\|/n)^r]$ unless G is a union of disjoint connected subgraphs each having exactly two edges. Suppose that this is the case and let $G_1, \ldots, G_p$ denote the corresponding subgraphs of G. Then, $\prod_{q=1}^{p} \Sigma(b, G_q) - \Sigma(b, G)$ and $\prod_{q=1}^{p} \Sigma(a, G_q) - \Sigma(a, G)$ are sums of terms of the form $\Sigma(b, G')$ and $\Sigma(a, G')$, respectively, where G' is a graph obtained from G by grouping at least one set of nodes from different G_q and replacing the edges linking them by the ones linking the resulting nodes. By the above results, the first sum is $o(n^{|G|-2b}\|b\|^{2p})$, and since G' can have at most $p-1$ disjoint connected subgraphs, the second sum is $o(\|a\|^{2p})$. Hence, $M(G)=\prod_{q=1}^{p}\Sigma(a,G_q)\Sigma(b,G_q)/n^{|G|_q}+o(w_n^{2p})$, since $\Sigma(b,G)=O(n^{|G|-2p}\|b\|^{2p})$ and $\Sigma(a,G)=O(\|a\|^{2p})$.

Thus, we have shown that the moments of odd order r of $\Sigma' a_{ij}B_{ij}$ are $o(w_n^r)$ and the moments of even order $r=2p$ are given by the sum of $\prod_{k=1}^{q}\Sigma(a,G_q)\Sigma(b,G_q)/n^{|G|_q}+o(w_n^{2p})$ over all sets of p disjoint connected graphs each having two edges, with union in $G(2p)$. To compute this sum, note that such subgraphs can have two or three nodes; those with two nodes can have two different forms: $\{(\mu,\nu),(\nu,\mu)\}$ or $\{(\mu,\nu),(\mu,\nu)\}$, and those with three nodes can have four different forms: $\{(\mu,\nu),(\mu,\lambda)\}$ or $\{(\mu,\nu),(\nu,\lambda)\}$ or $\{(\mu,\nu),(\lambda,\nu)\}$ or $\{(\mu,\nu),(\lambda,\mu)\}$. Now, each graph in $G(2p)$ can be identified with a partition of $\{1,\ldots,2p\}\times\{1,2\}$ as explained at the beginning of the proof of Theorem 3.1, and the above subgraphs can be identified with a partition of $\{p_1,p_2\}\times\{1,2\}$ where p_1,p_2 is a pair of integers in $\{1,\ldots,2p\}$. Thus, the number of ways one can choose l,m subgraphs as above, with two and three nodes, respectively, amounts to $2^l4^m(2p)!/(2^pl!m!)$. It follows that the considered sum is $[(2p)!/(2^pp!)]w_n^{2p}$, up to a term $o(w_n^{2p})$. But the last expression is precisely the $2p$-th moment of a normal variate with mean zero and variance w_n^2. The proof is completed.

Acknowledgement

The authors would like to thank a referee for bringing our attention to references Bloemena (1964) and Henze (1988).

References

Bloemena, A. R. (1964). Sampling from a graph, *Mathematical Center Tract* 2, Mathematisch Centrum, Amsterdam.

Cliff, A. D. and Ord, J. K. (1981). *Spatial Processes*; *Models and Applications*, Pion, London.

Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations, *Biometrika*, **33**, 129–135.

Friedman, J. and Rafsky, L. C. (1979). Multivariate generalisation of the Wald-Wolfowitz and Smirnov two-sample test, *Ann. Statist.*, **7**, 697–717.

Friedman, J. and Rafsky, L. C. (1983). Graph-theoretic measure of multivariate association and prediction, *Ann. Statist.*, **11**, 377–391.

Henze, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences, *Ann. Statist.*, **16**, 772–783.

Hoeffding, W. (1951). A combinatorial central limit theorem, *Ann. Statist.*, **22**, 558–566.

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

Schilling, M. F. (1986). Multivariate two-sample test based on nearest neighbors, *J. Amer. Statist. Assoc.*, **81**, 799-806.

Shapiro, C. P. and Hubert, L. (1979). Asymptotic normality of permutation statistic derived from weighted sum of bivariate functions, *Ann. Statist.*, **7**, 788-794.