האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

A PREMUTATION TEST FOR MATCHING

by

LARRY GOLDSTEIN and YOSEF RINOTT

Discussion Paper # 301 November 2002

מרכז לחקר הרציונליות וההחלטות האינטראקטיביות

CENTER FOR RATIONALITY AND INTERACTIVE DECISION THEORY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

A PERMUTATION TEST FOR MATCHING

LARRY GOLDSTEIN* AND YOSEF RINOTT[†]

Abstract. We consider a permutation method for testing whether observations given in their natural pairing exhibits an unusual level of similarity in situations where any two observations may be similar at some unknown baseline level. Under a null hypotheses where there is no distinguished pairing of the observations, a normal approximation, with explicit bounds and rates, is presented for determining approximate critical test levels.

1. Introduction

Schiffman et. al (1978), with statistical assistance by one of the authors[†], studied the influence of a doctor's prior probabilities of diseases on diagnosis. First, each doctor in a sample produced a ranking X of the prevalence, or probability, of various diseases; such a ranking is the result of both common medical knowledge and the particular doctor's personal experience. The doctors were then presented with compatible medical scenarios from which they were to produce a ranked list Y of diagnoses. Rank correlations between X and Y were then computed. To test the hypotheses that a doctors personal experience does not influence his diagnostic rankings a null hypotheses of no correlation is not appropriate since even with no influence of personal experience on diagnosis one would expect that the pair of rankings produced would be correlated due to the influence of common medical knowledge. The absence of a baseline correlation raises the question of how high the rank correlations need to be to justify rejection of the null hypothesis.

For our next example, consider an instructor who wants to know if students are copying from their neighbors in a class where students take an exam while seated in pairs. Given a measure of similarity between two exams X_1 and X_2 we expect exams to be similar even in the absence of copying, due to knowledge in common. Therefore, we want to test if the similarity between seated pairs is unusual relative to some unknown baseline similarity. This example is different from the first in that here a similarity score can be computed for any pair of exams X_i, X_j, whereas in the first example the correlations of interest are those between X and Y.

The situations described also arise, for example, in studying whether the similarity with which husbands and wives rank movies is based not only on the quality of the movie, but also on factors common to husbands and wives such as taste or mutual influence, or in environmental studies for testing whether pairs matched by, for example, a neighborhood within a city, exhibit health problems in common more than what can be explained by health problems prevalent in the city.

Our first example is a specific instance of a problem of the following type. Given pairs of observations $(X_1, Y_1), \ldots, (X_m, Y_m)$, where X_i and Y_i take values in spaces so that a proximity function $c(X, Y)$ is defined, we want to test whether the natural pairing of X_i to Y_i exhibits a significantly higher level of similarity than an unknown baseline level. The null hypotheses that the level of similarity in the natural pairing is the same as the baseline level can be formulated

Received by the editors October 9 2002.
1991 *Mathematics Subject Classification.* Primary 62G10; Secondary 62E20.
Key words and phrases. Similarity, Normal Approximation, Stein's Method.
† Supported by a grant from the Israel Academy of Sciences.

as the hypothesis that the vectors $[(X_1, Y_{\pi(1)}), \ldots, (X_m, Y_{\pi(m)})]$ are identically distributed for all $\pi \in S_m$, the permutation group. Conditioning on $e_{ij} = c(X_i, Y_j)$, a permutation test which compares the value of

$$V_\pi = \sum_{i=1}^{m} e_{i\pi(i)}$$

for $\pi = \mathrm{id}$ the identity, against critical values of the distribution of V_π when π is uniform over S_m, can be used to test the null hypothesis. Many authors (see Bolthausen 1984 and Chen and Ho 1978, and references therein) study the normal approximation for the permutation distribution of V_π, which can be used to obtain approximate critical values. Related work can be also be found in Stein (1986), whose ideas and methods have strongly influenced us.

In this paper, we focus on situations of the second type, which can be put in the following framework. Given an even number n of paired observations $(X_1, X_2), (X_3, X_4), \ldots, (X_{n-1}, X_n)$, with values in a space so that a proximity function $c(X_i, X_j)$ is defined, we want to test whether the natural pairing of X_{2i-1} with X_{2i} exhibits a significantly higher level of similarity than an unknown baseline level. The null hypotheses that the similarity level of the natural pairing is the same as the baseline level is here formulated as the hypothesis that the vectors $[(X_i, X_{\pi(i)}), i < \pi(i)]$ are identically distributed for all $\pi \in \Pi_n$, where

$$\Pi_n = \{\pi \in S_n : \pi^2 = \mathrm{id}, \pi(i) \neq i \quad \text{for all } i\}. \tag{1}$$

The condition $\pi^2 = \mathrm{id}$ reflects the fact that if i is paired with j then j is paired with i, and the condition $\pi(i) \neq i$ the fact that no i can be paired with itself. The natural pairing corresponds to the permutation $\tilde{\pi} \in \Pi_n$ specified by the conditions $\tilde{\pi}(2i-1) = 2i$ and $(\tilde{\pi})^2 = \mathrm{id}$.

Conditioning on $e_{ij} = c(X_i, X_j)$, we consider the permutation test which compares the value of

$$U_\pi = \sum_{i=1}^{n} e_{i\pi(i)} \tag{2}$$

at $\pi = \tilde{\pi}$ against critical values of the distribution of U_π when π is uniform over Π_n. For the null hypothesis to be true it is sufficient that the $X's$ are exchangeable, but the null hypothesis is complex and does not specify the distribution of U_π nor the baseline similarity. In the absence of a null distribution, the above permutation test seems very natural. By allowing all possible pairings $i \neq j$ in $c(X_i, X_j)$, we obtain higher power than in the previous type of example where pairing was restricted to X's with Y's.

We shall provide a normal approximation, including bounds, rates, and explicit constants, using Stein's method, to the permutation distribution of U_π, in order to determine approximate critical values for the permutation test. The methods used here apply to the permutation distribution of V_π, mutatis mutandis.

Henceforth we suppress the dependence of U_π in (2) on π. Furthermore, for values g_{ij} with $g_{ii} = 0$, we set

$$g_{i+} = \sum_{j=1}^{n} g_{ij}, \quad g_{+j} = \sum_{i=1}^{n} g_{ij}, \quad g_{++} = \sum_{i,j=1}^{n} g_{ij}, \quad \text{and} \quad \overline{g_{i+}} = \frac{1}{n-1} g_{i+}.$$

Note that the terms $e_{i\pi(i)}$ and $e_{\pi(i)i}$ always appear together in the sum U, and we may therefore assume without loss of generality that $e_{ij} = e_{ji}$. The diagonal terms e_{ii} never enter U and we take them to be 0. Given such a collection of numbers e_{ij}, define

$$d_{ij} = \begin{cases} e_{ij} - e_{i+}/(n-2) - e_{+j}/(n-2) + e_{++}/[(n-1)(n-2)] & i \neq j \\ 0 & i = j. \end{cases} \tag{3}$$

Bounds to the normal approximation for the permutation distribution of U are contained in the following theorem. For convenience we assume without further comment that $n \geq 10$.

Theorem 1.1. *Let U be given by (2), π be uniform over Π_n in (1) and*

$$W = \frac{U - EU}{\sqrt{Var(U)}}, \quad \alpha = \max |d_{ij} - d_{kl}|, \quad \text{and} \quad \delta = \sup_{w \in R} |P(W \leq w) - \Phi(w)|,$$

where Φ is the standard normal distribution function. Then

$$EU = e_{++}/(n-1),$$

$$Var(U) = \frac{2}{(n-1)(n-3)} \left((n-2) \sum_{i,k=1}^{n} e_{ik}^2 + \frac{1}{n-1} e_{++}^2 - 2 \sum_{i=1}^{n} e_{i+}^2 \right), \tag{4}$$

and there exist constants c_1, c_2 such that

$$\delta \quad \leq \quad c_1 n^{1/2} \sqrt{\{ \sum_{i,j=1}^{n} d_{ij}{}^4 / (\sum_{i,j=1}^{n} d_{ij}^{\,2})^2 \}} + \frac{c_2 \alpha^3 n^{5/2}}{(\sum_{i,j=1}^{n} d_{ij}^{\,2})^{3/2}}.$$

If, for example, the constants d_{ij} are bounded then α is bounded, and if in addition $\sum_{ij} d_{ij}^2 = O(n^2)$, then, in view of (7), the bound above decays at the rate of $\mathrm{Var}(U)^{-1/2} = n^{-1/2}$. Below a somewhat crude calculation gives the upper bounds of $c_1 \leq 86, c_2 \leq 243$.

2. Proof of Theorem 1.1

We compute the mean and variance of U in Section 2.1, and establish the upper bound on the normal approximation in Section 2.2.

2.1. Mean and Variance of U.

To compute the mean and variance of $U = \sum_{i=1}^{n} e_{i\pi(i)}$, where π is chosen uniformly from Π_n, we have the following Lemma.

Lemma 2.1. *Let g_{ij} satisfy $g_{ii} = 0$ and set*

$$f_{ij} = \begin{cases} g_{ij} - \overline{g_{i+}} & i \neq j \\ 0 & i = j. \end{cases}$$

Then with

$$V = \sum_{i=1}^{n} g_{i\pi(i)}$$

we have

$$Eg_{i\pi(i)} = \overline{g_{i+}} \quad \text{and therefore} \quad EV = \sum_{i=1}^{n} \overline{g_{i+}}$$

and

$$Var(V) = \frac{1}{(n-1)(n-3)} \left((2n-5) \sum_{|\{i,j\}|=2} f_{ij}^2 + \sum_{|\{i,j\}|=2} f_{ij} f_{ji} \right).$$

Proof: Since $\pi(i)$ can be any $j \neq i$ with probability $1/(n-1)$, we have

$$Eg_{i\pi(i)} = \frac{1}{n-1} \sum_{j:j \neq i} g_{ij} = \frac{1}{n-1} g_{i+} = \overline{g_{i+}},$$

and so

$$\begin{aligned}\mathrm{Var}(V) &= \mathrm{Var}\sum_{i=1}^n f_{i\pi(i)} \\ &= \sum_{i=1}^n E f_{i\pi(i)}^2 + \sum_{|\{i,j\}|=2} E(f_{i\pi(i)} f_{j\pi(j)}) \\ &= \frac{1}{n-1}\sum_{|\{i,j\}|=2} f_{ij}^2 + \sum_{|\{i,j\}|=2} E(f_{i\pi(i)} f_{j\pi(j)}).\end{aligned}$$

Now note that the probability is $1/(n-1)$ that $\pi(i) = j$, and therefore that $\pi(j) = i$. If $\pi(i) \neq j$, then $i, j, \pi(i), \pi(j)$ are all distinct, and given any $|\{i,j,k,l\}| = 4$ the probability that $\pi(i) = k$ and $\pi(j) = l$ is $1/[(n-1)(n-3)]$. We therefore have

$$E\sum_{|\{i,j\}|=2} f_{i\pi(i)} f_{j\pi(j)} = \frac{1}{n-1}\sum_{|\{i,j\}|=2} f_{ij}^2 + \frac{1}{(n-1)(n-3)}\sum_{|\{i,j,k,l\}|=4} f_{ik} f_{jl}. \tag{5}$$

The first equality below follows by summing over $l \notin \{i,j,k\}$, and using $f_{jj} = 0$ and $f_{j+} = 0$, and the second in a similar way by summing over $j \notin \{i,k\}$;

$$\begin{aligned}\sum_{|\{i,j,k,l\}|=4} f_{ik} f_{jl} &= \sum_{|\{i,j,k\}|=3} f_{ik}(-f_{ji} - f_{jk}) \\ &= \sum_{|\{i,k\}|=2} f_{ik}(f_{ki} + f_{ik}) \\ &= \sum_{|\{i,k\}|=2} \left(f_{ik} f_{ki} + f_{ik}^2\right).\end{aligned}$$

The formula for $\mathrm{Var}(V)$ now follows by collecting terms. □

Writing for the moment U_d and U_e for the values of U based on d_{ij} and e_{ij} respectively, we have

$$U_d = U_e - \frac{e_{++}}{n-1}.$$

In order to see the above relation between U_d and U_e, sum (3) over i with $i \neq j$, and use symmetry to yield $e_{+j} = e_{j+}$, and obtain

$$d_{+j} = e_{+j} - [e_{++} - e_{j+}]/(n-2) - (n-1)e_{+j}/(n-2) + e_{++}/(n-2) = 0,$$

so that

$$d_{i+} = d_{+j} = d_{++} = 0. \tag{6}$$

Since the distribution of U_d is a simple translation of that for U_e we study $U_d = U$; henceforth we suppress the d.

Applying Lemma 2.1 with $g_{ij} = d_{ij}$, since $d_{i+} = 0$ we have $f_{ij} = d_{ij}$ and therefore

$$EU = 0;$$

using also $d_{ij} = d_{ji}$,

$$\mathrm{Var}(U) = \frac{2(n-2)}{(n-1)(n-3)}\sum_{i,j=1}^n d_{ij}^2. \tag{7}$$

In terms of the symmetric but otherwise arbitrary values e_{ij} which may not satisfy (6), the variance in (4) is obtained by substituting (3) into (7).

2.2. **Normal Approximation Upper Bound.** We apply the following theorem, which is a special case of (1.10) of Theorem 1.2 of Rinott and Rotar (1997), when $R = 0$, using (1.12). The latter is based on Stein's method (Stein 1986, pg 35), with an improvement on the rates under some condition.

Theorem 2.1. *Let* (W, W^*) *be exchangeable with* $EW = 0$ *and* $EW^2 = 1$ *such that for* $0 < \lambda < 1$ *we have*

$$E(W^*|W) = (1 - \lambda)W. \tag{8}$$

If

$$|W^* - W| \le A \tag{9}$$

for a constant A*, then*

$$\begin{aligned} \delta = \sup_{w \in R} &|P(W \le w) - \Phi(w)| \\ \le \ & \frac{12}{\lambda}\sqrt{Var\{E[(W^* - W)^2|W]\}} + \sqrt{\frac{2}{\pi}}\frac{(48 + \sqrt{32})A^3}{\lambda}. \end{aligned}$$

We shall apply Theorem 2.1 to $W = U/\sigma$, but for convenience we first describe the coupling and compute the relevant quantities in terms of U. Given a permutation π chosen uniformly from Π_n construct the permutation π^* in Π_n by choosing I, J distinct and uniformly, and imposing $\pi^*(I) = J$ (and therefore $\pi^*(J) = I$), and $\pi^*(\pi(I)) = \pi(J)$ (and therefore $\pi^*(\pi(J)) = \pi(I)$) and fixing the values of $\pi^*(k) = \pi(k)$ for $k \notin \{I, J, \pi(I), \pi(J)\}$. With $U = \sum_i d_{i\pi(i)}$, let $U^* = \sum_i d_{i\pi^*(i)}$.

To verify (8), first note that

$$U^* - U \quad = \quad 2\left(d_{IJ} + d_{\pi(I)\pi(J)} - (d_{I\pi(I)} + d_{J\pi(J)})\right), \tag{10}$$

where the factor 2 accounts for the symmetry $d_{ij} = d_{ji}$.

Letting C be the event that $J \neq \pi(I)$, we have $(U^* - U) = (U^* - U)\mathbf{1}_C$ and therefore

$$E((U^* - U)|U) = E((U^* - U)\mathbf{1}_C|U).$$

For the first two terms in (10), recalling $d_{++} = 0$, and using that (I, J) is independent of π and equals any of the $n(n-1)$ pairs (i, j) for which $i \neq j$,

$$\begin{aligned} E(d_{IJ}\mathbf{1}_C|\pi) &= \frac{1}{n(n-1)} \sum_{|\{i,j\}|=2} d_{ij}\mathbf{1}(j \neq \pi(i)) \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\sum_{j:j\neq i} d_{ij} - d_{i\pi(i)} \right) \\ &= \frac{-1}{n(n-1)} U, \end{aligned}$$

and similarly for the term $d_{\pi(I)\,\pi(J)}$, as $(\pi(I), \pi(J))$ has the same distribution as (I, J).

Now consider the third term on the right hand side in (10):

$$E(d_{I\pi(I)}\mathbf{1}_C|\pi) = \frac{1}{n(n-1)} \sum_{|\{i,j\}|=2} d_{i\pi(i)}\mathbf{1}(j \neq \pi(i))$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} d_{i\pi(i)} \sum_{j:j\neq i} \mathbf{1}(j \neq \pi(i)) = \frac{1}{n(n-1)} \sum_{i=1}^{n} d_{i\pi(i)} \sum_{j:j\notin\{i,\pi(i)\}} 1$$

$$= \frac{n-2}{n(n-1)} \sum_{i=1}^{n} d_{i\pi(i)} = \frac{n-2}{n(n-1)} U.$$

By symmetry the same is true for the term $d_{J\pi(J)}$.

Collecting terms and using $\mathcal{F}(U) \subset \mathcal{F}(\pi)$, where $\mathcal{F}(X)$ denotes the sigma field generated by the random variable X, we have

$$E(U^* - U|U) = \frac{-2}{n(n-1)}(2 + 2(n-2))U = -\frac{4}{n}U.$$

Thus (8) holds with $\lambda = 4/n$.

Now we consider the first term in the bound in Theorem 2.1; since $\mathcal{F}(U) \subset \mathcal{F}(\pi)$,

$$\text{Var}\{E[(U^* - U)^2|U]\} \leq \text{Var}\{E[(U^* - U)^2|\pi]\}. \tag{11}$$

From (10),

$$E((U^* - U)^2|\pi) = 4E\left([(d_{IJ} + d_{\pi(I)\pi(J)}) - (d_{I\pi(I)} + d_{J\pi(J)})]^2|\pi\right). \tag{12}$$

When we expand the square we get the following types of terms; (i) the square terms from the first group of parentheses, (ii) mixed terms formed by taking one term from the first group with one term from the second, (iii) the square terms from the second group, (iv) mixed terms between values in the first group, and (v) mixed terms between values in the second group.

(i) The value of the conditional expectation for the square term $\mathrm{E}(d_{IJ}^2|\pi)$ clearly does not depend on π, and therefore contributes a constant value which does not affect the variance. The same is true for $E(d_{\pi(I)\pi(J)}^2|\pi)$ because as (I, J) range over all possible distinct pairs with equal probability so do $(\pi(I), \pi(J))$.

(ii) Terms such as $E(d_{IJ}d_{I\pi(I)}|\pi)$, evaluate to zero. In this particular case take expectation over J first and use $d_{i+} = 0$.

By tallying the contributions from terms (iii),(iv), and (v), we conclude that, up to an additive constant not depending on π, and therefore not affecting the variance, (12) equals

$$4\left(\frac{2}{n}\sum_{i=1}^{n} d_{i\pi(i)}^2 + \frac{2}{n(n-1)}\sum_{|\{i,j\}|=2} d_{ij}d_{\pi(i)\pi(j)} + \frac{2}{n(n-1)}\sum_{|\{i,j\}|=2} d_{i\pi(i)}d_{j\pi(j)}\right). \tag{13}$$

We may write (13) as $8(A_1 + A_2 + A_3)$ where

$$A_1 = \frac{1}{n}\sum_{i=1}^{n} d_{i\pi(i)}^2, \qquad A_2 = \frac{1}{n(n-1)}\sum_{|\{i,j\}|=2} d_{ij}d_{\pi(i)\pi(j)}$$

$$\text{and} \quad A_3 = \frac{1}{n(n-1)}\sum_{|\{i,j\}|=2} d_{i\pi(i)}d_{j\pi(j)}.$$

In view of (11), we now need to compute the variance of (13) with respect to a uniform $\pi \in \Pi_n$. We have

$$\text{Var}(8(A_1 + A_2 + A_3)) \leq 8^2 \cdot 3\left(\text{Var}(A_1) + \text{Var}(A_2) + \text{Var}(A_3)\right).$$

To calculate $\text{Var}(A_1)$, apply Lemma 2.1 with $g_{ij} = d_{ij}^2$ to obtain

$$\text{Var}(A_1) = \frac{1}{n^2(n-1)(n-3)} \left((2n-5) \sum_{|\{i,j\}|=2} f_{ij}^2 + \sum_{|\{i,j\}|=2} f_{ij} f_{ji} \right).$$

For the second term above, by Cauchy-Schwarz,

$$|\sum_{|\{i,j\}|=2} f_{ij} f_{ji}| \le \sqrt{\sum_{|\{i,j\}|=2} f_{ij}^2 \sum_{|\{i,j\}|=2} f_{ji}^2} = \sum_{|\{i,j\}|=2} f_{ij}^2. \tag{14}$$

Collecting terms we conclude

$$\text{Var}(A_1) \quad \le \quad \frac{2(n-2)}{n^2(n-1)(n-3)} \sum_{|\{i,j\}|=2} f_{ij}^2 \le \frac{3}{n^3} \sum_{|\{i,j\}|=2} d_{ij}^2$$

for $n \ge 8$.

We now turn to $\text{Var}(A_2)$. With

$$\mathcal{I} = \{\mathbf{I} = (i,j,k,l,\pi(i),\pi(j),\pi(k),\pi(l)) : i \ne j, k \ne l, \pi \in \Pi_n\},$$

it can be shown that when π is uniform over Π_n, the probability of a given $\mathbf{I} \in \mathcal{I}$ satisfying $|\mathbf{I}| = s$ is

$$P(\mathbf{I}) = \frac{1}{[n]_s}, \quad s \in \{2,4,6,8\}, \quad \text{where} \quad [n]_s = (n-1)(n-3)\cdots(n-s+1).$$

For $\mathbf{I} = (i,j,k,l,i',j',k',l') \in \mathcal{I}$ set $d_{\mathbf{I}} = d_{ij} d_{kl} d_{i'j'} d_{k'l'}$. We then have

$$\begin{aligned} \text{Var}(\sum_{|\{i,j\}|=2} d_{ij} d_{\pi(i)\pi(j)}) \quad &\le \quad \sum_{|\{i,j\}|=2 |\{k,l\}|=2} d_{ij} d_{kl} E(d_{\pi(i)\pi(j)} d_{\pi(k)\pi(l)}) \\ &= \quad \sum_{\mathbf{I} \in \mathcal{I}} \frac{1}{[n]_{|\mathbf{I}|}} d_{\mathbf{I}} = \sum_{s \in \{2,4,6,8\}} \frac{1}{[n]_s} \sum_{\mathbf{I} \in \mathcal{I}(s)} d_{\mathbf{I}}, \end{aligned} \tag{15}$$

where $\mathcal{I}(s)$ are all those $\mathbf{I} \in \mathcal{I}$ with $|\mathbf{I}| = s$.

Consider first the case of $s = 8$. Since $d_{k'+} = 0$, summing over $l' \notin \{i, j, k, l, i', j'\}$ we have

$$\begin{aligned} \sum_{\mathbf{I} \in \mathcal{I}(8)} d_{\mathbf{I}} &= \sum_{\mathbf{I} \in \mathcal{I}(8)} d_{ij} d_{kl} d_{i'j'} d_{k'l'} \\ &= \quad - \sum_{|\{i,j,k,l,i',j',k'\}|=7} \sum_{l' \in \{i,j,k,l,i',j'\}} d_{ij} d_{kl} d_{i'j'} d_{k'l'}. \end{aligned} \tag{16}$$

Applying Cauchy Schwarz to each of the six terms in the inner sum, the absolute value of the expression is bounded by

$$6(n-2)_5 \sum_{|\{i,j\}|=2} d_{ij}^4,$$

where

$$(n)_s = n(n-1)\cdots(n-s+1).$$

For $s \in \{2,4,6\}$ apply Cauchy Schwarz to

$$\sum_{\mathbf{I} \in \mathcal{I}(s)} d_{ij} d_{kl} d_{i'j'} d_{k'l'}$$

to obtain the bound

$$(n-2)_{s-2} \sum_{|\{i,j\}|=2} d_{ij}^4.$$

Therefore

$$\begin{aligned}\mathrm{Var}(A_2) &\le \frac{1}{(n(n-1))^2}\left(\frac{6(n-2)_5}{[n]_8}+\sum_{s\in\{2,4,6\}}\frac{(n-2)_{s-2}}{[n]_s}\right)\sum_{|\{i,j\}|=2} d_{ij}^4\\ &\le \frac{7}{n^3}\sum_{|\{i,j\}|=2} d_{ij}^4,\end{aligned}$$

where the latter bound holds for $n \ge 10$ and follows by elementary calculations.

Although A_3 and A_2 are not identically distributed, it is easy to see that the variance of A_3 can be bounded in exactly the same manner.

We obtain from (13) and the above discussion that

$$\mathrm{Var}\{E[(U^*-U)^2|U]\} \le (8^2\cdot 3)\,\frac{17}{n^3}\sum_{i\ne j} d_{ij}^4. \tag{17}$$

We now apply Theorem 2.1 to $W = U/\sigma$, $W^* = U^*/\sigma$. From (7) we conclude that

$$\mathrm{Var}(U) = \sigma^2 \ge \frac{2}{n}\sum_{|\{i,j\}|=2} d_{ij}^2.$$

It follows from (17),

$$\mathrm{Var}\{E[(W^*-W)^2|W]\} \le \frac{8^2\cdot 3\cdot 17}{4n}\sum_{|\{i,j\}|=2} d_{ij}{}^4/(\sum_{|\{i,j\}|=2} d_{ij}^2)^2. \tag{18}$$

With

$$\alpha = \max|d_{ij} - d_{kl}|,$$

we have $|U^* - U| \le 4\alpha$, and hence

$$\begin{aligned}|W^*-W| &\le \frac{1}{\sigma}|U^*-U| \le \frac{4\alpha}{\sigma}\\ &\le \frac{4\alpha\sqrt{n}}{\sqrt{2\sum_{|\{i,j\}|=2} d_{ij}^2}} = A.\end{aligned}$$

Applying Theorem (2.1) with this A, $\lambda = 4/n$ and using expression (18), we have

$$\delta \le 86n^{1/2}\sqrt{\{\sum_{|\{i,j\}|=2} d_{ij}{}^4/(\sum_{|\{i,j\}|=2} d_{ij}^2)^2\}} + \frac{243\alpha^3 n^{5/2}}{(\sum_{|\{i,j\}|=2} d_{ij}^2)^{3/2}}.$$

References

[1] Bolthausen, E. (1984) An estimate of the remainder in a combinatorial central limit theorem. Z. Wahrsch. Verw. Gebiete 66 379-386.

[2] Ho, Soo Thong; Chen, Louis H. Y. (1978) An L_p bound for the remainder in a combinatorial central limit theorem. Ann. Probability 6, 231–249.

[3] Pesarin, F. (2001) Multivariate permutation tests. With applications in biostatistics. John Wiley & Sons, Ltd., Chichester.

[4] Schiffman, A., Cohen, S., Nowik, R., and Sellinger, D. (1978) Initial Diagnostic hypotheses: factors which may dostort physicians judgement. Orgnanizational Behaveior and Human Performance, 21 305-315.

[5] Stein, C. (1986) Approximate computation of expectations. Institute of Mathematical Statistics Lecture Notes-Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA.

*Department of Mathematics, University of Southern California., † Department of Statistics Hebrew University, Jerusalem
E-mail address: larry@paradox.usc.edu, rinott@mscc.huji.ac.il