LARRY GOLDSTEIN – YOSEF RINOTT

A Permutation Test For Matching and its Asymptotic Distribution [1]

SUMMARY -
   We consider a permutation method for testing whether observations given in their natural pairing exhibit 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.

Key Words: Similarity, Normal Approximation, Stein's Method

1   Introduction

The work in this paper was motivated by examples such as the following.

Example 1. 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. Statistical thinking, which can be formalized in Bayesian terms, suggests that given a set of symptoms, a doctor's diagnosis or ranking of possible diagnoses should be influenced not only by the symptoms, but also by the disease prevalence at the time of diagnosis. Doctors' information on prevalence may come, for example, from textbooks, articles, and personal experience. The goal of the study was to verify the influence of *personal* prior information or opinion on disease prevalence (henceforth referred to as "personal prior") on diagnosis and help determine whether doctors need to be better educated to take prevalence into account, or if providing them with information on prevalence at the time of diagnosis is useful.

In this study each doctor in a sample produced first a ranking X of the prevalence, or of the probability of various diseases from a given list; such a ranking represents the doctor's personal prior. A compatible medical scenario was then presented to all doctors, and each one of them produced a ranked list Y of possible diagnoses from the same given list. Rank correlations between X and Y for each doctor were then computed. To test the hypotheses that a doctor's personal prior does not influence his diagnostic rankings, a null hypotheses of zero correlation between each doctor's X and Y is not appropriate. Even with no such influence, one

would expect that pairs of rankings would have some nonzero baseline correlation due to the influence of other factors like common medical knowledge. The null hypothesis of interest that there is no influence of personal prior is complex since the baseline correlation is unknown. The presence of an unknown baseline correlation raises the question of how high the within-doctor rank correlations need to be to reject the null hypothesis and assert the claim that there is influence of personal prior on diagnostic rankings.

Correlations are used here as a measure of similarity between ranked lists. Henceforth we will talk about similarity in general, and the approach applies to any measure of similarity or proximity defined on the sample space.

The main focus of this paper is on examples of the following kind:

Example 2. This example is somewhat artificial, but it is simpler and can clarify the issue; it will also help in explaining the example that follows it which is rather similar. 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 exams, we expect any two exams to be similar even in the absence of copying. Common knowledge that all students hopefully have would make their exams similar to a certain, unknown degree. Therefore, we want to test if the similarity between seated pairs is unusual (due to copying) 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.

Example 3. Situations similar to Example 2 arise naturally in environmental and medical studies, where subjects in a given study group are matched (paired) by certain common background of interest, such as having lived in the same neighborhood during a given period, having certain common medical conditions or having certain variables in common (e.g., gender, age, weight, etc.). In order to assess the influence of the background in question on a given set of certain medical conditions (denoted by X_i for subject i), one should test whether matched pairs are more similar than unmatched ones relative to the medical condition being studied. The baseline similarity between unmatched pairs is again unknown, but a certain degree of similarity must certainly exist due to common factors that all subject in the particular study might have. More specifically, suppose we have an even number n of subjects and those indexed by $2i - 1$ and $2i$ form the matched pairs for $i = 1, \ldots, n/2$, and let X_i measure subject i's medical condition. Our goal is to test whether all X_{2i-1} and X_i, which arise from the matched pairs, exhibit more similarity then X_i and X_j from unmatched pairs.

A related testing problem arises in the design of studies involving matching of subjects that are similar by some background criteria in order to reduce variability in other variables of interest. The matching process often requires a great effort. The question of whether it achieves its purpose in producing a higher level of similarity in the variables of interest than would be achieved at random, can be tested as described in this paper.

Example 1 is a specific instance of a problem of the following type. Consider pairs of observations $(X_1, Y_1), \ldots, (X_n, Y_n)$, where X_i and Y_i take values in a space so that a proximity function $c(X, Y)$ is defined. This function may sometimes be obtained as a decreasing function of some metric. However, for the rankings of Example 1, the rank correlation is a relevant proximity function not derived from a metric. We want to test whether the natural pairing of X_i to Y_i exhibits a significantly higher level of proximity or similarity than an unknown baseline level. The null hypotheses that the level of proximity or similarity in the natural pairing is the same as the baseline level can be formulated as the hypothesis that the observations $[(X_1, Y_{\pi(1)}), \ldots, (X_n, Y_{\pi(n)})]$ are identically distributed for all $\pi \in S_n$, the permutation group of n elements. Conditioning on the observed $\{e_{ij} = c(X_i, Y_j)\}$, a permutation test which compares the value of

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

for the special permutation $\pi = \mathrm{id}$ (the identity), against critical values of the distribution of U_π when π is uniform over S_n (the distribution assigning equal probabilities to every element in S_n), can be used to test the null hypothesis.

The permutation distribution of U_π for π uniform over S_n was studied in numerous other statistical contexts. For a seminal reference which contains both theory and applications see Wald and Wolfowitz (1944). More recent articles which in turn contain further references include the following. In connection with linear rank statistics, Ho and Chen (1978) and Bolthausen (1984) computed bounds on the rate of convergence to normality. Bickel and van Zwet (1978) give more background and results on linear rank statistics for two-sample problems, including an Edgeworth expansion for a special case of (1). Diaconis, Graham and Holmes (1999) discuss similar statistics and also some subsets of permutations related to tests for independence. Kolchin and Chistyakov (1973) discuss the permutation distribution for the subset of permutations with one cycle. Below we discuss a rather different subset of permutations, in which the number of cycles is maximal. For general theory on permutation tests see, e.g, Pesarin (2001) and references

therein. Related work on normal approximations can be also be found in Stein (1986), whose ideas and methods have strongly influenced us and other authors. We will give a brief indication of some basic ideas of Stein's method.

In this paper we focus on situations as described in Examples 2 and 3 where all pairings can be compared. We consider 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 special 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 special pairing is the same as the baseline level is here formulated as the hypothesis that the variables $[(X_i, X_{\pi(i)}), i < \pi(i)]$ are identically distributed for all $\pi \in \Pi_n$ where

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

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 special pairing which we suspect may show a high similarity level 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 the set of values $\{e_{ij} = c(X_i, X_j)\}$ we consider the permutation test which compares the value of U_π at the special permutation $\pi = \tilde{\pi}$ against critical values of the distribution of U_π when π is uniform over Π_n.

The proposed two tests discussed above appear similar, as in both tests the observed similarity related to a special permutation is compared to critical values computed against a null distribution induced by the uniform distribution over a space of relevant permutations. For the first test that space is S_n and the special permutation is the identity (which matches X_i with Y_i), and for the second test the space of permutations is Π_n and the special permutation is $\tilde{\pi}$ (which matches X_{2i-1} with X_i). We henceforth discuss only the second case and study the permutation distribution relative to Π_n. The methods used here apply to the permutation distribution over the whole of S_n mutatis mutandis.

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.

We shall provide a normal approximation to the permutation distribution of U_π of (1) including bounds, rates, and explicit constants in order to determine approximate critical values for the permutation test.

4

Henceforth we suppress the dependence of U_π in (1) 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

$$d_{ij} = \begin{cases} e_{ij} - \frac{e_{i+}}{(n-2)} - \frac{e_{+j}}{(n-2)} + \frac{e_{++}}{(n-1)(n-2)} & i \neq j \\ 0 & i = j. \end{cases} \quad (2)$$

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 *Let U be given by (1), π be uniform over Π_n 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

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

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

and there exist constants c_1, c_2 such that

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

If, for example, the constants d_{ij} are bounded then α is bounded and $\sum_{ij} d_{ij}^2 = O(n^2)$, so in view of (6), the bound above decays at the rate of $Var(U)^{-1/2} = n^{-1/2}$. Below a somewhat crude calculation gives the upper bounds of $c_1 \leq 86, c_2 \leq 243$.

2 Proof of Theorem 1

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 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

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

6

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

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

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. $\square$

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 (2) 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

$$d_{i+} = d_{+j} = d_{++} = 0. \quad (5)$$

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

$$EU = 0;$$

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

$$\mathrm{Var}(U) = \frac{2(n-2)}{(n-1)(n-3)} \sum_{i,j=1}^{n} d_{ij}^2. \quad (6)$$

7

In terms of the symmetric but otherwise arbitrary values e_{ij} which may not satisfy (5), the variance in (3) is obtained by substituting (2) into (6).

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

$$E(W^*|W) = (1 - \lambda)W. \tag{7}$$

If

$$|W^* - W| \leq A \tag{8}$$

for a constant A, then

$$\delta = \sup_{w \in R} |P(W \leq w) - \Phi(w)|$$

$$\leq \frac{12}{\lambda} \sqrt{Var\{E[(W^* - W)^2|W]\}} + \sqrt{\frac{2}{\pi}} \frac{(48 + \sqrt{32})A^3}{\lambda}. \tag{9}$$

We briefly indicate the idea behind the proof of a theorem of this type. This discussion can serve as some introduction to Stein's method for the interested reader, but it is not necessary for the rest of the paper.

First note that a random variable W has the standard normal distribution if and only if

$$Ef'(W) = EWf(W) \tag{10}$$

holds for all continuous and piecewise continuously differentiable functions f, for which the expectations in (10) exist. This motivates the differential equation (12) in the lemma below.

Set $\Phi h = Eh(Z)$, where Z is standard normal, and h is a function for which the expectation exists. Also, for a real valued h, let $||h||$ denote the sup norm, that is, $||h|| = \sup_x |h(x)|$. The lemma below is elementary, though the bounds in (13) require some calculations.

Lemma 2 *Let h be a bounded piecewise continuously differentiable real valued function. The function*

$$f(w) = e^{w^2/2} \int_{-\infty}^{w} [h(x) - \Phi h]e^{-x^2/2}dx \tag{11}$$

8

solves the (first order linear) differential equation

$$f'(w) - wf(w) = h(w) - \mathbf{\Phi}h, \tag{12}$$

and

(a) $||f|| \leq \sqrt{2\pi}||h||$, (b) $||f'|| \leq 2||h||$, (c) $||f''|| \leq 2||h'||$. (13)

Now, exchangeability of (W, W^*) and (7) directly imply

$$E\{Wf(W)\} = \frac{E\{(W^* - W)[f(W^*) - f(W)]\}}{2\lambda}.$$

Together with (12) this implies

$$Eh(W) - \mathbf{\Phi}h = Ef'(W) - \frac{E\{(W^* - W)[f(W^*) - f(W)]\}}{2\lambda}. \tag{14}$$

The first term in the Taylor expansion of $f(W^*) - f(W)$ is $(W^* - W)f'(W)$, and the r.h.s. of (14) is bounded by $\frac{1}{2\lambda}E\{f'(W)[2\lambda - (W^* - W)^2]\}$ plus a remainder term which we now ignore. By (7) we have $E(W^* - W)^2 = 2\lambda$, and (13 b) and the Cauchy Schwarz inequality readily yield

$$E\{f'(W)[2\lambda - (W^* - W)^2]\} \leq 2||h||\sqrt{\mathrm{Var}\{E[(W^* - W)^2|W]\}}.$$

Using (14), an approximation of an indicator function of a half line by a smooth h yields a term similar to the first term on the r.h.s. of (9), and calculation of the remainder in the above Taylor expansion yields the second. To obtain the precise bound (9), a certain induction and further calculations are needed. $\square$

We shall apply Theorem 2 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 (7), first note that

$$U^* - U = 2\left(d_{IJ} + d_{\pi(I)\pi(J)} - (d_{I\pi(I)} + d_{J\pi(J)})\right), \tag{15}$$

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 (15), 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 (15):

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

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

Now we consider the first term in the bound in Theorem 2; since $\mathcal{F}(U) \subset \mathcal{F}(\pi)$,

$$\operatorname{Var}\{E[(U^* - U)^2|U]\} \leq \operatorname{Var}\{E[(U^* - U)^2|\pi]\}. \tag{16}$$

10

From (15),

$$E((U^* - U)^2|\pi) = 4E\left([(d_{IJ} + d_{\pi(I)\pi(J)}) - (d_{I\pi(I)} + d_{J\pi(J)})]^2|\pi\right). \quad (17)$$

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 $E(d_{IJ}^2|\pi)$ 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, (17) equals

$$\frac{8}{n}\sum_{i=1}^{n}d_{i\pi(i)}^2 + \frac{8}{n(n-1)}\left(\sum_{|\{i,j\}|=2}d_{ij}d_{\pi(i)\pi(j)} + \sum_{|\{i,j\}|=2}d_{i\pi(i)}d_{j\pi(j)}\right). \quad (18)$$

We may write (18) 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 (16), we now need to compute the variance of (18) 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 1 with $g_{ij} = d_{ij}^2$ to obtain

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

11

For the second term above, by Cauchy-Schwarz,

$$|\sum_{|\{i,j\}|=2} f_{ij}f_{ji}| \leq \sqrt{\sum_{|\{i,j\}|=2} f_{ij}^2 \sum_{|\{i,j\}|=2} f_{ji}^2} = \sum_{|\{i,j\}|=2} f_{ij}^2. \qquad (19)$$

Collecting terms we conclude

$$\mathrm{Var}(A_1) \leq \frac{2(n-2)}{n^2(n-1)(n-3)} \sum_{|\{i,j\}|=2} f_{ij}^2 \leq \frac{3}{n^3} \sum_{|\{i,j\}|=2} d_{ij}^4$$

for $n \geq 8$.

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

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

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

$$\mathrm{Var}(\sum_{|\{i,j\}|=2} d_{ij}d_{\pi(i)\pi(j)}) \leq \sum_{|\{i,j\}|=2|\{k,l\}|=2} d_{ij}d_{kl}E(d_{\pi(i)\pi(j)}d_{\pi(k)\pi(l)})$$

$$= \sum_{\mathbf{I}\in\mathcal{I}} \frac{1}{[n]_{|\mathbf{I}|}}d_{\mathbf{I}} = \sum_{s\in\{2,4,6,8\}} \frac{1}{[n]_s} \sum_{\mathbf{I}\in\mathcal{I}(s)} d_{\mathbf{I}}, \qquad (20)$$

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

$$\sum_{\mathbf{I}\in\mathcal{I}(8)} d_{\mathbf{I}} = \sum_{\mathbf{I}\in\mathcal{I}(8)} d_{ij}d_{kl}d_{i'j'}d_{k'l'}$$

$$= -\sum_{|\{i,j,k,l,i',j',k'\}|=7} \sum_{l'\in\{i,j,k,l,i',j'\}} d_{ij}d_{kl}d_{i'j'}d_{k'l'}. \qquad (21)$$

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

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

12

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

where the latter bound holds for $n \geq 10$ 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 (18) and the above discussion that
$$\operatorname{Var}\{E[(U^* - U)^2 | U]\} \leq (8^2 \cdot 3)\frac{17}{n^3} \sum_{i \neq j} d_{ij}^4. \tag{22}$$

We now apply Theorem 2 to $W = U/\sigma$, $W^* = U^*/\sigma$. From (6) we conclude that
$$\operatorname{Var}(U) = \sigma^2 \geq \frac{2}{n} \sum_{|\{i,j\}|=2} d_{ij}^2.$$

It follows from (22),
$$\operatorname{Var}\{E[(W^* - W)^2 | W]\} \leq \frac{8^2 \cdot 3 \cdot 17}{4n} \sum_{|\{i,j\}|=2} d_{ij}{}^4 / (\sum_{|\{i,j\}|=2} d_{ij}^2)^2. \tag{23}$$

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

13

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

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

Applying Theorem 2 with this A, $\lambda = 4/n$ and using expression (23), we have

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

3 Acknowledgements

References

P. BICKEL AND W.R. VAN ZWET (1978). Asymptotic expansions for the power of distribution free tests in the two-sample problem. *Ann. Statistics*, **6, 4**, 937–1004.

E. BOLTHAUSEN (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete*, **66, 3**, 379–386.

P. DIACONIS, R. GRAHAM, AND S. HOLMES (1999). Statistical problems involving permutations with restricted positions. *State of the art in probability and statistics IMS Lecture Notes Monogr. Ser.*, **36**, 195–222.

S. T. HO AND L. H. Y. CHEN (1978). An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probability*, **6, 2**, 231–249.

V.F. KOLCHIN AND V.P. CHISTYAKOV (1973). On a combinatorial limit theorem. *Theory Probability Appl.*, **18**, 728–739.

F. PESARIN (2001). *Multivariate permutation tests: with applications in biostatistics*. John Wiley & Sons.

Y. RINOTT AND V. ROTAR (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.*, **7, 4**, 1080–1105.

A. SCHIFFMAN, S. COHEN, R. NOWIK, AND D. SELLINGER (1978). Initial diagnostic hypotheses: factors which may distort physicians judgement. *Orgnanizational Behavior and Human Performance*, **21**, 305–315.

C. STEIN (1986). *Approximate computation of expectations*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA.

A. WALD AND J. WOLFOWITZ (1944). Statistical tests based on permutations of the observations. *Ann. Math Statistics*, **15**, 358–372.

Larry Goldstein
Department of Mathematics
University of Southern California, DRB 155
Los Angeles, CA (USA)
larry@math.usc.edu

Yosef Rinott
Department of Statistics
Hebrew University
Jerusalem, Israel
rinott@mscc.huji.ac.il