The permutation distribution of matrix correlation statistics

A. D. Barbour and Louis H. Y. Chen

Angewandte Mathematik, Universität Zürich
Winterthurerstrasse 190, CH–8057 Zürich, Switzerland
E-mail: adb@amath.unizh.ch
and
Institute for Mathematical Sciences, National University of Singapore
3 Prince George's Park, Singapore 118402
E-mail: lhychen@ims.nus.edu.sg

Many statistics used to test for association between pairs (y_i, z_i) of multivariate observations, sampled from n individuals in a population, are based on comparing the similarity a_{ij} of each pair (i, j) of individuals, as evidenced by the values y_i and y_j, with their similarity b_{ij} based on the values z_i and z_j. A common strategy is to compute the sample correlation between these two sets of values. The appropriate null hypothesis distribution is that derived by permuting the z_i's at random among the individuals, while keeping the y_i's fixed. In this paper, a Berry–Esseen bound for the normal approximation to this null distribution is derived, which is useful even when the matrices a and b are relatively sparse, as is the case in many applications. The proofs are based on constructing a suitable exchangeable pair, a technique at the heart of Stein's method.

1. Introduction

When testing for linear association between the x- and y-values in data consisting of n pairs $(x_1, y_1), \ldots, (x_n, y_n)$ of real numbers, a classical procedure compares the value of Pearson's product moment correlation coefficient

$$r := \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \, \sum_{i=1}^n (y_i - \bar{y})^2}} \tag{1.1}$$

with a critical value derived from its null distribution. In the parametric version, the null distribution is that resulting when $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ are independent random samples from the standard normal distribution. Alternatively, to avoid making such strong assumptions, one can use the permutation null distribution, obtained when the values $x_1, \ldots, x_n$ are paired *at random* with the values $y_1, \ldots, y_n$. This is the distribution of $T := \sum_{i=1}^n x_i y_{\pi(i)}$, where π denotes a uniformly distributed random element of the permutations S_n of the set $[n] := \{1, 2, \ldots, n\}$. The combinatorial central limit theorem of Wald & Wolfowitz (1944) shows that this distribution is asymptotically normal under limiting conditions reminiscent of those for the usual central limit theorem. The same is true for more general statistics T' of the form $T' := \sum_{i=1}^n z(i, \pi(i))$, for z a matrix of reals, as was proved by Hoeffding (1951). Later, in a celebrated application of Stein's method, Bolthausen (1984) established the Berry–Esseen bound

$$d_K(\mathcal{L}(\sigma^{-1}(T' - \mathbb{E}T')), \Phi) \le C\sigma^{-3} n^{-1} \sum_{i=1}^n \sum_{l=1}^n |z^*(i,l)|^3 \qquad (1.2)$$

for a universal constant C, where $\sigma^2 := \mathrm{Var} T'$ and

$$z^*(i,l) := z(i,l) - n^{-1} \left\{ \sum_{j=1}^n z(j,l) + \sum_{k=1}^n z(i,k) \right\} + n^{-2} \sum_{j=1}^n \sum_{k=1}^n z(j,k);$$

here, d_K denotes the Kolmogorov distance between probability distributions,

$$d_K(P,Q) := \sup_x |P\{(-\infty, x]\} - Q\{(-\infty, x]\}|.$$

Specialized to the context of T, this implies that

$$d_K(\mathcal{L}(\sigma^{-1}(T - \mathbb{E}T)), \Phi) \le C\sigma^{-3} n^{-1} \sum_{i=1}^n |\tilde{x}_i|^3 \sum_{i=1}^n |\tilde{y}_i|^3, \qquad (1.3)$$

for the same constant C, where

$$\sigma^2 := \mathrm{Var} T = (n-1)^{-1} \sum_{i=1}^n \tilde{x}_i^2 \sum_{i=1}^n \tilde{y}_i^2,$$

and $\tilde{x}_i := x_i - \bar{x}$, $\tilde{y}_i := y_i - \bar{y}$.

There are other measures of linear association, of which Kendall's τ is an example, which can be viewed mathematically as a 2-dimensional or matrix analogue of Pearson's statistic. However, statistics of this form, which were first systematically studied by Daniels (1944), turn out to have much greater importance when applied as measures of spatial or spatio-temporal association. Their practical introduction began with the work of

Moran (1948) and Geary (1954) in geography and of Knox (1964) and Mantel (1967) in epidemiology; see also the book by Hubert (1987). They typically take the form

$$W := \sum\nolimits_{i,j}' a_{ij} b_{ij},$$

where a_{ij} and b_{ij} are two different measures of closeness of the i'th and j-th observations, which may or may not be related; the null distribution of interest is then that of

$$W := \sum\nolimits_{i,j}' a_{ij} b_{\pi(i)\pi(j)},$$

where π is uniformly distributed on S_n. Here and throughout, the notation $\sum'_{i,j}$ denotes the sum over all ordered pairs of *distinct* indices $i, j \in [n]$; a similar convention is adopted for more indices. In this and the following section, we assume that the matrices a and b are symmetric, as is natural for such applications. Other choices are possible, and Daniels (1944) himself considered anti-symmetric matrices; these, together with other, more general constructions, are covered by the results of Section 3.

Various conditions under which the null distribution of W should be asymptotically normal have been advanced in the literature. The first approach, using the method of moments, is that of Daniels (1944). Another approach involves approximating b_{ij} by $b(n^{-1}i, n^{-1}j)$ for a piecewise Δ-monotone function b on $[0,1]^2$ (see Jogdeo (1968)), and then approximating W by $\sum'_{i,j} a_{ij} b(U_i, U_j)$, where the random variables U_i, $1 \le i \le n$, are independent and uniformly distributed on $[0,1]$. This direction was taken by Shapiro and Hubert (1979), who based their arguments on the work of Jogdeo (1968). Both of these approaches give rise to unnecessarily restrictive conditions; in particular, both require that the leading, linear term in a Hoeffding projection of W should be dominant. This requirement greatly restricts the practical application of the theorems, and alternative conditions, avoiding the difficulty, were proposed in Abe (1969) and in Cliff & Ord (1973). Barton (unpublished manuscript) gave counter-examples to these proposals. He then studied the case in which a is the incidence matrix of a graph and $b_{ij} = b(c(i), c(j))$ is a function of the colours $c(i)$, $c(j)$ assigned to the vertices i and j of the graph, and gave a correct, although still rather restrictive, set of conditions for asymptotic normality in this context.

Satisfactory conditions for asymptotic normality in the general case were given in Barbour & Eagleson (1986a). We begin by stating their Theorem 1, for which some more notation is required. Define

$$A_0 := \{(n)_2\}^{-1} \sum\nolimits'_{i,j} a_{ij}; \quad A_{12} := n^{-1} \sum_{i=1}^{n} \{a_i^*\}^2;$$

$$A_{22} := \{(n)_2\}^{-1} \sum\nolimits'_{i,j} \tilde{a}_{ij}^2 \quad \text{and} \quad A_{13} := n^{-1} \sum_{i=1}^{n} |a_i^*|^3,$$

where $(n)_r$ denotes $n(n-1)\cdots(n-r+1)$,

$$a_i^* := (n-2)^{-1} \sum_{j:j\neq i} (a_{ij} - A_0), \quad \text{and} \quad \tilde{a}_{ij} := a_{ij} - a_i^* - a_j^* - A_0,$$

and make similar definitions for b. Then they showed that

$$d_1(\mathcal{L}(\{\mathrm{Var}W\}^{-1/2}(W - \mathbb{E}W)), \Phi) \leq K_1(\delta_1 + \delta_2) \tag{1.4}$$

for a universal constant K_1. In this bound, the quantities δ_1 and δ_2 are given by

$$\delta_1 := n^4\sigma^{-3} A_{13} B_{13} \quad \text{and} \quad \delta_2 := \sqrt{\frac{A_{22}B_{22}}{nA_{12}B_{12}}}, \tag{1.5}$$

and

$$\sigma^2 := \frac{4n^2(n-2)^2}{n-1} A_{12} B_{12}; \tag{1.6}$$

d_1 denotes the bounded Wasserstein metric: for probability measures P and Q on $\mathbb{R}$,

$$d_1(P,Q) := \sup_{f\in\mathcal{F}} \left| \int f\,dP - \int f\,dQ \right|,$$

where $\mathcal{F}$ is the set of bounded Lipschitz functions satisfying $\|f\| \leq 1$ and having Lipschitz constant at most 1.

This result implies that a sequence $W^{(n)}$ of statistics of this form is asymptotically normal if both $\delta_1^{(n)}$ and $\delta_2^{(n)}$ tend to 0. Although the argument is still based on a Hoeffding projection of W, the conditions are both simpler and less restrictive than those previously given, and the error in the approximation, measured in terms of the metric d_1, is bounded by directly calculable quantities. The leading, linear component in the projection is a 1-dimensional statistic to which the Wald–Wolfowitz theorem can be applied with error of order $O(\delta_1)$; the quantity δ_2 accounts for the error involved in neglecting the remaining, quadratic component of W.

The distance (1.4) between $\mathcal{L}(\{\mathrm{Var}W\}^{-1/2}(W - \mathbb{E}W))$ and Φ, measured with respect to the metric d_1, automatically implies a bound with respect to the Kolmogorov distance d_K of order $O(\delta_1^{1/2}+\delta_2^{1/2})$; see, for example, the brief remarks in Erickson (1974, pp. 527,528). Zhao *et al.* (1997), working directly in terms of Kolmogorov distance, proved that

$$d_K(\mathcal{L}(\sigma^{-1/2}(W-\mathbb{E}W)),\Phi) \le K_2(\delta_1+\delta_3), \tag{1.7}$$

where

$$\delta_3 := \sigma^{-3}n^4 A_{23}B_{23} \quad \text{and} \quad A_{23} := n^{-2}\sum\nolimits'_{i,j} |\tilde{a}_{ij}|^3;$$

this bound is the specialization to our problem of their more general theorem giving the rate of convergence to the normal for the statistic $X := \sum_i \sum_j C(i,j;\pi(i),\pi(j))$, where C is an arbitrary 4-dimensional array. For Kolmogorov distance, (1.7) can be asymptotically sharper than the rate deduced from (1.4). Suppose, for instance, that $a_{ij}^{(n)} = a(i/n,j/n)$ and $b_{ij}^{(n)} = b(i/n,j/n)$, $1 \le i,j \le n$, for smooth symmetric functions $a,b : [0,1]^2 \to \mathbb{R}$. Then

$$\begin{aligned}
\frac{1}{n(n-1)}\sum\nolimits'_{i,j} a_{ij}^{(n)} &\sim \int_0^1\int_0^1 a(x,y)\,dx\,dy =: a_0;\\
a_i^{*(n)} &\sim a_1(i/n), \quad \text{where} \quad a_1(x) := \int_0^1 \{a(x,y)-a_0\}\,dy;\\
\tilde{a}_{ij}^{(n)} &\sim a_2(i/n,j/n), \quad \text{where} \quad a_2(x,y) := a(x,y)-a_1(x)-a_1(y)-a_0;\\
A_{12}^{(n)} &\sim \int_0^1 a_1^2(x)\,dx =: A^{[12]};\\
A_{22}^{(n)} &\sim \int_0^1\int_0^1 a_2^2(x,y)\,dx\,dy =: A^{[22]};\\
A_{13}^{(n)} &\sim \int_0^1 |a_1(x)|^3\,dx =: A^{[13]};\\
A_{23}^{(n)} &\sim \int_0^1\int_0^1 |a_2(x,y)|^3\,dx\,dy =: A^{[23]},
\end{aligned}$$

and similar asymptotics hold also for b. Hence, if the $A^{[lm]}$ and $B^{[lm]}$ are all positive, for $l=1,2$ and $m=2,3$, it follows that

$$\begin{aligned}
\delta_1 &\sim n^{-1/2}\frac{A^{[13]}B^{[13]}}{(A^{[12]}B^{[12]})^{3/2}} \asymp n^{-1/2};\\
\delta_2 &\sim n^{-1/2}\left\{\frac{A^{[22]}B^{[22]}}{A^{[12]}B^{[12]}}\right\}^{1/2} \asymp n^{-1/2},
\end{aligned} \tag{1.8}$$

and

$$\delta_3 \sim n^{-1/2} \frac{A^{[23]} B^{[23]}}{(A^{[12]} B^{[12]})^{3/2}} \asymp n^{-1/2}. \tag{1.9}$$

Thus both $\delta_1 + \delta_2$ and $\delta_1 + \delta_3$ are of order $O(n^{-1/2})$, but for d_K the bound (1.4) then only implies a rate of $O(n^{-1/4})$, which is worse than the order $O(n^{-1/2})$ implied by (1.7).

However, as remarked above and as discussed at greater length in Barbour & Eagleson (1986b), these asymptotics are not those relevant in many statistical applications. To illustrate this, we specialize to graph-based measures of association, such as those considered by Knox (1964), Bloemena (1964) and Friedman & Rafsky (1983), and take both $a^{(n)}$ and $b^{(n)}$ to be incidence matrices of graphs. In examples concerning spatial and temporal correlation, increasing n typically means increasing either the area of the study or its timespan; in either case, the number of neighbours of a vertex is unlikely to increase proportionally to n, as the above asymptotics would suggest, but rather more slowly. So suppose that the typical vertex degree in $a^{(n)}$ is of order n^α for some $0 < \alpha \leq 1$, and in $b^{(n)}$ of order n^β for some $0 < \beta \leq 1$. This means that each row and column of A (B) typically contains $O(n^\alpha)$ ($O(n^\beta)$) 1's, all other entries being 0. Then the corresponding asymptotics for the A_{rs} typically become

$$A_{12}^{(n)} \asymp n^{-2+2\alpha}; \quad A_{22}^{(n)} \asymp n^{-1+\alpha}; \quad A_{13}^{(n)} \asymp n^{-3+3\alpha} \quad \text{and} \quad A_{23}^{(n)} \asymp n^{-1+\alpha}, \tag{1.10}$$

though for balanced graphs both $A_{12}^{(n)}$ and $A_{13}^{(n)}$ may be of smaller order; analogous formulae hold for the B_{rs}. Note that these relations agree with the previous asymptotics only when $\alpha = \beta = 1$.

It now follows from (1.10) that

$$\delta_1 \asymp n^{-1/2}; \quad \delta_2 \asymp n^{\frac{1}{2}\{1-(\alpha+\beta)\}}; \quad \delta_3 \asymp n^{\frac{1}{2}\{7-4(\alpha+\beta)\}}.$$

Thus $\delta_1 = o(\delta_2)$ and $\delta_2 = o(\delta_3)$ unless $\alpha = \beta = 1$. Furthermore, $\delta_2^{1/2} = o(\delta_3)$ whenever $\alpha + \beta < 13/7$, so that, even for d_K, the bound (1.4) implies a better result than (1.7) in this range; indeed, the bound in (1.7) tends to zero as $n \to \infty$ only if $\alpha + \beta > 7/4$, whereas that in (1.4) tends to zero so long as $\alpha + \beta > 1$.

In Theorem 2.4, we improve the d_K bounds to the same order $O(\delta_1 + \delta_2)$ as that given in (1.4) for the metric d_1. This then gives a convergence rate which is uniformly better than that given in Zhao *et al.* (1997) whenever $\alpha + \beta < 2$, and is as good when $\alpha + \beta = 2$. Indeed, the quantity δ_1 is the same as that appearing in Bolthausen's Lyapounov third moment bound for approximating the linear component in the projection, and is of order $O(n^{-1/2})$ in all the asymptotics considered above. The quantity δ_2 is the ratio of the standard deviations of the quadratic and linear components, and is a natural measure of the effect of neglecting the quadratic component in comparison with the linear component. Hence nothing essentially better than our theorem is to be hoped for from any argument that is based, as are both ours and that of Zhao *et al.* (1997), on using the linear projection for the approximation and ignoring the quadratic component.

Our argument, motivated by Stein's method, proceeds by way of a concentration inequality, and is much in the spirit of Chen and Shao (2004). The concentration inequality is used to bound the probability that the linear component takes values in certain intervals of random length; this in turn enables the error incurred by neglecting the quadratic component to be controlled. We also use the same technique in the final section, to prove a d_K-bound of analogous order for the error in the normal approximation of $X = \sum_i \sum_j C(i,j;\pi(i),\pi(j))$, the context studied by Zhao *et al.* (1997): see Theorem 3.3. The resulting bounds are again always of asymptotic order at least as good as those of Zhao *et al.* (1997), and are frequently much better.

Example: Each of n individuals is assigned two sets of values, one describing them in socio-economic terms, the other in terms of consumption pattern. Thus individual i is assigned values $y_i \in \mathcal{Y}$ and $z_i \in \mathcal{Z}$, where $\mathcal{Y}$ and $\mathcal{Z}$ are typically high dimensional Euclidean spaces. The individuals are then clustered twice, once using the y_i values, and once using the z_i, and it is of interest to know whether the two clusterings are related. To test this, one can set $a_{ij}(b_{ij}) = 1$ if i and j belong to the same $\mathcal{Y}$- ($\mathcal{Z}$-) cluster, and 0 otherwise, and compare the value of the statistic $W := \sum'_{i,j} a_{ij}b_{ij}$ to its null distribution. To fix ideas, suppose that there are $k_{\mathcal{Y}}$ $\mathcal{Y}$-clusters, with sizes roughly uniformly distributed between 1 and $2n/k_{\mathcal{Y}}$. Then

$$A_{12} \asymp k_{\mathcal{Y}}^{-2}, \quad A_{22} \asymp k_{\mathcal{Y}}^{-1}, \quad \text{and} \quad A_{13} \asymp k_{\mathcal{Y}}^{-3}.$$

Assume that analogous relations hold also for the $\mathcal{Z}$-clustering. Then

$$\delta_1 \asymp n^{-1/2} \quad \text{and} \quad \delta_2 \asymp \{n^{-1}k_{\mathcal{Y}}k_{\mathcal{Z}}\}^{1/2}.$$

As n increases, more data being available, it is natural to allow for a more informative analysis, by letting both $k_{\mathcal{Y}} = k_{\mathcal{Y}}^{(n)}$ and $k_{\mathcal{Z}} = k_{\mathcal{Z}}^{(n)}$ increase with n. Then the normal approximation to the null distribution of W is guaranteed by Theorem 2.4 to have error of order $O(\{n^{-1}k_{\mathcal{Y}}^{(n)}k_{\mathcal{Z}}^{(n)}\}^{1/2})$, which is $o(1)$ as long as $k_{\mathcal{Y}}^{(n)} k_{\mathcal{Z}}^{(n)} = o(n)$. This represents a substantial improvement over the order to be obtained from the theorem of Zhao *et al.* (1997), which is $o(1)$ only so long as $k_{\mathcal{Y}}^{(n)} k_{\mathcal{Z}}^{(n)} = o(n^{1/4})$.

Remark: It is nonetheless interesting to note that there is a second d_1-bound given in Barbour & Eagleson (1986a), of more complicated form than those given above, which does not rely on projection, and which establishes asymptotic normality in yet wider circumstances. Under the asymptotics of (1.10), it yields the result

$$d_1(\mathcal{L}(\{\mathrm{Var} W\}^{-1/2}(W - \mathbb{E}W)), \Phi) = O(n^{-\frac{1}{2}\min(\alpha+\beta,1)}), \qquad (1.11)$$

establishing asymptotic normality for *all* choices of $\alpha, \beta > 0$, and not just for $\alpha + \beta > 1$. This then also implies that

$$d_K(\mathcal{L}(\{\mathrm{Var} W\}^{-1/2}(W - \mathbb{E}W)), \Phi) = O(n^{-\frac{1}{4}\min(\alpha+\beta,1)}), \qquad (1.12)$$

of smaller order than $\delta_1 + \delta_2$ if $\alpha + \beta < 3/2$, indicating that the approach by way of projection, while much simpler in construction, loses some precision.

2. Matrix correlation statistics

Let A and B be symmetric $n \times n$ matrices, and define

$$W := W(\pi) := \sum\nolimits_{i,j}' a_{ij} b_{\pi(i)\pi(j)}, \qquad (2.1)$$

where π is a uniform random permutation of $[n] := \{1, 2, \ldots, n\}$ and $\sum\nolimits_{i,j}'$ denotes the sum over pairs (i, j) of *distinct* elements of $[n]$. Thus, in the definition of W, the diagonal elements play no part, so that we may assume that $a_{ii} = b_{ii} = 0$ for all $i \in [n]$. We then note that

$$W = \sum\nolimits_{i,j}' (a_{ij} - A_0)(b_{\pi(i)\pi(j)} - B_0) + \mu,$$

where

$$\mu := \frac{1}{(n)_2} \sum\nolimits_{i,j}' \sum\nolimits_{l,m}' a_{ij} b_{lm};$$

we use the notation from the introduction throughout. We note also that, with a_i^* and $\tilde{a}_{ij}$ defined as before, we have

$$\sum_{i=1}^{n} a_i^* = 0; \quad \sum_{i:i\neq j} \tilde{a}_{ij} = \sum_{j:j\neq i} \tilde{a}_{ij} = 0. \qquad (2.2)$$

With these definitions, and with their analogues for the matrix B, it was shown in Barbour & Eagleson (1986a) that

$$W = \mu + \widetilde{V} + \widetilde{\Delta}, \tag{2.3}$$

where

$$\widetilde{V} := 2(n-2)\sum_{i=1}^{n} a_i^* b_{\pi(i)}^* \quad \text{and} \quad \widetilde{\Delta} := {\sum}'_{i,j} \tilde{a}_{ij}\tilde{b}_{\pi(i)\pi(j)}; \tag{2.4}$$

it was also shown that

$$\sigma^2 := \mathrm{Var}\widetilde{V} = \frac{4n^2(n-2)^2}{n-1} A_{12}B_{12}, \quad \mathrm{Var}\widetilde{\Delta} = \frac{2n(n-1)^2}{n-3} A_{22}B_{22},$$

and that $\mathrm{Cov}\,(\widetilde{V}, \widetilde{\Delta}) = 0$. In this section, we derive a concentration inequality bounding the probability of $V := \sigma^{-1}\widetilde{V}$ belonging to an interval of random length $|\Delta|$, where $\Delta := \sigma^{-1}\widetilde{\Delta}$, and use it to deduce a normal approximation to W. The argument is based on ideas from Stein's method, using an exchangeable pair. Broadly speaking, we show that

$$\mathbb{E}\{V f_\Delta(V)\} \approx \mathbb{E}\{f'_\Delta(V)\},$$

for a function f_Δ which is chosen so that $f'_\Delta(v) \approx \mathbf{1}_{[z,z+\Delta]}(v)$, and which also satisfies $\|f_\Delta\| \approx \frac{1}{2}|\Delta|$: see also Chen & Shao (2004). To avoid trivial exceptions, we assume that $A_{12}B_{12} > 0$; otherwise, $\widetilde{V} = 0$ a.s.

The first step is to construct an exchangeable pair (Stein 1986, p. 2). We do this by defining a pair of independent random variables I and J, each uniformly distributed on $[n]$, which are independent also of π, and then defining π' by

$$\pi'(\alpha) = \begin{cases} \pi(\alpha) & \text{if } \alpha \neq I, J; \\ \pi(J) & \text{if } \alpha = I; \\ \pi(I) & \text{if } \alpha = J. \end{cases} \tag{2.5}$$

We then set

$$V' := 2\sigma^{-1}(n-2)\sum_{i=1}^{n} a_i^* b_{\pi'(i)}^*; \quad \Delta' := \sigma^{-1} {\sum}'_{i,j} \tilde{a}_{ij}\tilde{b}_{\pi'(i)\pi'(j)} :$$

it follows immediately that (V, Δ) and (V', Δ') are exchangeable. We use this to prove the following lemma.

Lemma 2.1: *We have*

$$\mathbb{E}\{V - V' \,|\, \pi\} = 2n^{-1}V; \qquad \mathbb{E}\{(V-V')^2\} = 4n^{-1}\mathbb{E}V^2 = 4n^{-1};$$
$$\mathbb{E}\{(\Delta - \Delta')^2\} = 8n^{-2}(n-1)\mathbb{E}\Delta^2 = 8n^{-2}(n-1)\delta_2^2,$$

where

$$\tilde\delta_2^2 := \mathbb{E}\Delta^2 = \frac{\mathrm{Var}\widetilde{\Delta}}{\mathrm{Var}\widetilde{V}} = \frac{(n-1)^3}{2n(n-2)^2(n-3)}\,\frac{A_{22}B_{22}}{A_{12}B_{12}} \sim \tfrac12\delta_2^2.$$

Proof: If U and U' are exchangeable, then

$$\mathbb{E}\{(U'-U)(U'+U)\} = 0,$$

which in turn implies that

$$\mathbb{E}\{(U'-U)^2\} = 2\mathbb{E}\{U(U-U')\}. \tag{2.6}$$

Applying this first with $U = V$, we have

$$\begin{aligned}\mathbb{E}\{(V'-V)^2\} &= 2\mathbb{E}\{V\mathbb{E}(V-V'\,|\,\pi)\}\\ &= 4\sigma^{-1}(n-2)\mathbb{E}\{V\mathbb{E}(H(I,J;\pi)\,|\,\pi)\},\end{aligned} \tag{2.7}$$

where

$$\begin{aligned}H(I,J;\pi) &:= h(I,J,\pi(I),\pi(J))\\ &:= a_I^*b_{\pi(I)}^* + a_J^*b_{\pi(J)}^* - a_I^*b_{\pi(J)}^* - a_J^*b_{\pi(I)}^*,\end{aligned}$$

because

$$V - V' = 2\sigma^{-1}(n-2)H(I,J;\pi). \tag{2.8}$$

Now $H(I,I;\pi) = 0$, and

$$\begin{aligned}\mathbb{E}(a_I^*b_{\pi(I)}^*\mathbf{1}\{I\ne J\}\,|\,\pi) &= \frac{n-1}{n^2}\sum_{i=1}^n a_i^*b_{\pi(i)}^* = \frac{(n-1)\sigma V}{2n^2(n-2)};\\ \mathbb{E}(a_I^*b_{\pi(J)}^*\mathbf{1}\{I\ne J\}\,|\,\pi) &= n^{-2}{\sum}'_{i,j}\, a_i^*b_{\pi(j)}^* = \frac{-\sigma V}{2n^2(n-2)},\end{aligned}$$

from (2.2) and (2.4). Hence it follows that

$$\mathbb{E}(H(I,J;\pi)\,|\,\pi) = \sigma V/\{n(n-2)\}, \tag{2.9}$$

which, from (2.8), is equivalent to $\mathbb{E}\{V-V'\,|\,\pi\} = 2n^{-1}V$, and hence, from (2.7), that

$$\mathbb{E}\{(V'-V)^2\} = 4n^{-1}\mathbb{E}V^2,$$

as claimed.

For the second part, apply (2.6) with $U = \Delta$; this gives

$$\mathbb{E}\{(\Delta'-\Delta)^2\} = 2\sigma^{-1}\mathbb{E}\{\Delta\mathbb{E}(K(I,J;\pi)\,|\,\pi)\},$$

where

$$\begin{aligned}K(I,J;\pi) := \sum_{j\neq I,J}&\{\tilde a_{Ij}(\tilde b_{\pi(I)\pi(j)} - \tilde b_{\pi(J)\pi(j)}) + \tilde a_{Jj}(\tilde b_{\pi(J)\pi(j)} - \tilde b_{\pi(I)\pi(j)})\}\\ &+ \sum_{i\neq I,J}\{\tilde a_{iI}(\tilde b_{\pi(i)\pi(I)} - \tilde b_{\pi(i)\pi(J)}) + \tilde a_{iJ}(\tilde b_{\pi(i)\pi(J)} - \tilde b_{\pi(i)\pi(I)})\}\\ &+ \tilde a_{JI}(\tilde b_{\pi(J)\pi(I)} - \tilde b_{\pi(I)\pi(J)}) + \tilde a_{IJ}(\tilde b_{\pi(I)\pi(J)} - \tilde b_{\pi(J)\pi(I)}).\end{aligned}$$

Now

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}\{I\neq J\}\sum_{j:j\neq I,J}\tilde a_{Ij}\tilde b_{\pi(I)\pi(j)}\,\Big|\,\pi\right) &= n^{-2}{\sum}'_{j,l,m}\tilde a_{lj}\tilde b_{\pi(l)\pi(j)}\\ &= n^{-2}(n-2)\sigma\Delta,\end{aligned}$$

where the sum ${\sum}'_{j,l,m}$ is over all *distinct* choices of the indices, and

$$\mathbb{E}\left(\mathbf{1}\{I\neq J\}\sum_{j:j\neq I,J}\tilde a_{Ij}\tilde b_{\pi(J)\pi(j)}\,\Big|\,\pi\right) = n^{-2}{\sum}'_{j,l,m}\tilde a_{lj}\tilde b_{\pi(m)\pi(j)} = -n^{-2}\sigma\Delta,$$

in view of (2.2). Hence

$$\mathbb{E}(K(I,J;\pi)\,|\,\pi) = 4n^{-2}(n-1)\sigma\Delta,$$

and the lemma follows because $\mathbb{E}\Delta^2 = \tilde\delta_2^2$. ∎

Now suppose that δ is chosen such that

$$\tfrac14 n\mathbb{E}\{|V'-V|\min(|V'-V|,\delta)\} \ge \tfrac34; \tag{2.10}$$

this is possible, since $\frac14 n\mathbb{E}\{|V'-V|\min(|V'-V|,\delta)\}$ is a continuous and increasing function of $\delta > 0$, taking all values between 0 and 1. Then we have the following bound.

Lemma 2.2: *For δ such that (2.10) is satisfied, it follows that*

$$\max\{\mathbb{P}[z-|\Delta|\le V\le z],\mathbb{P}[z\le V\le z+|\Delta|]\} \le (1+\sqrt2)\tilde\delta_2 + 2\delta + 2\mathrm{Var}\eta(\delta)$$

for all z, where

$$\eta(\delta) := \tfrac14 n\mathbb{E}\{|V'-V|\min(|V'-V|,\delta)\,|\,\pi\}.$$

Proof: We give the argument for $\mathbb{P}[z \le V \le z + |\Delta|]$. Define the function $f_\Delta = f_\Delta^{(z)}$ by

$$f_\Delta(x) = \begin{cases} -(\frac{1}{2}|\Delta| + \delta) & \text{if } x \le z - \delta; \\ -\frac{1}{2}|\Delta| + x - z & \text{if } z - \delta < x < z + |\Delta| + \delta; \\ \frac{1}{2}|\Delta| + \delta & \text{if } x \ge z + |\Delta| + \delta. \end{cases}$$

Then, by the exchangeability of (V, Δ) and (V', Δ'), we have

$$\mathbb{E}\{(V' - V)[f_\Delta(V') + f_{\Delta'}(V)]\} = 0;$$

adding $2\mathbb{E}\{(V - V')f_\Delta(V)\}$ to each side and multiplying by $n/4$, this gives

$$\begin{aligned} &\tfrac{1}{2}n\mathbb{E}\{(V - V')f_\Delta(V)\} \\ &\quad = \tfrac{1}{4}n\mathbb{E}\{(V' - V)[f_\Delta(V') - f_\Delta(V)]\} + \tfrac{1}{4}n\mathbb{E}\{(V' - V)[f_{\Delta'}(V) - f_\Delta(V)]\}. \end{aligned} \tag{2.11}$$

The left-hand side can be rewritten as in Lemma 2.1 as

$$\begin{aligned} \tfrac{1}{2}n\mathbb{E}\{(V - V')f_\Delta(V)\} &= \tfrac{1}{2}n\mathbb{E}\{f_\Delta(V)\mathbb{E}(V - V' \,|\, \pi)\} \\ &= \mathbb{E}\{Vf_\Delta(V)\}, \end{aligned}$$

this last from Lemma 2.1, implying in turn that

$$\begin{aligned} \left|\tfrac{1}{2}n\mathbb{E}\{(V - V')f_\Delta(V)\}\right| &\le \tfrac{1}{2}\mathbb{E}|V\Delta| + \delta\mathbb{E}|V| \\ &\le \tfrac{1}{2}\tilde{\delta}_2 + \delta, \end{aligned} \tag{2.12}$$

from the Cauchy–Schwarz inequality, from the definition of $\tilde{\delta}_2$, and because $\mathbb{E}V^2 = 1$.

The second term on the right-hand side of (2.11) is easily bounded by

$$\begin{aligned} \left|\tfrac{1}{4}n\mathbb{E}\{(V' - V)[f_{\Delta'}(V) - f_\Delta(V)]\}\right| &\le \tfrac{1}{8}n\mathbb{E}\{|V' - V|\,||\Delta'| - |\Delta||\} \\ &\le \tfrac{1}{8}n\mathbb{E}\{|V' - V|\,|\Delta' - \Delta|\} \\ &\le \tilde{\delta}_2/\sqrt{2}, \end{aligned} \tag{2.13}$$

by Lemma 2.1. For the first, we write

$$\begin{aligned} \tfrac{1}{4}n\mathbb{E}\{(V' - V)[f_\Delta(V') - f_\Delta(V)]\} &= \tfrac{1}{4}n\mathbb{E}\left\{(V' - V)\int_0^{V'-V} f'_\Delta(V + t)\,dt\right\} \\ &= \mathbb{E}\left\{\int_{-\infty}^{\infty} f'_\Delta(V + t)M(t)\,dt\right\}, \end{aligned}$$

where

$$M(t) := \tfrac{1}{4}n(V' - V)[\mathbf{1}\{V' - V > t \ge 0\} - \mathbf{1}\{V' - V < t \le 0\}] \ge 0$$

a.s. for all t. Recalling the definitions of f_Δ and $\eta(\delta)$, it thus follows that

$$\begin{aligned}
&\tfrac14 n\mathbb{E}\{(V'-V)[f_\Delta(V')-f_\Delta(V)]\}\\
&\ge \mathbb{E}\left\{\int_{|t|\le\delta}\mathbf{1}\{z-\delta\le V+t\le z+|\Delta|+\delta\}M(t)\,dt\right\}\\
&\ge \mathbb{E}\left\{\mathbf{1}\{z\le V\le z+|\Delta|\}\int_{|t|\le\delta}M(t)\,dt\right\}\\
&= \tfrac14 n\mathbb{E}\big\{\mathbf{1}\{z\le V\le z+|\Delta|\}|V'-V|\min(|V'-V|,\delta)\big\}\\
&= \mathbb{E}\big\{\eta(\delta)\mathbf{1}\{z\le V\le z+|\Delta|\}\big\}.
\end{aligned}$$

Writing $\eta(\delta)=\mathbb{E}\eta(\delta)+\{\eta(\delta)-\mathbb{E}\eta(\delta)\}$, and noting that $\mathbb{E}\eta(\delta)\ge 3/4$, by choice of δ, this implies that

$$\begin{aligned}
&\tfrac14 n\mathbb{E}\{(V'-V)[f_\Delta(V')-f_\Delta(V)]\} \qquad (2.14)\\
&\quad\ge \tfrac34\mathbb{P}[z\le V\le z+|\Delta|]-\big|\mathbb{E}\big\{(\eta(\delta)-\mathbb{E}\eta(\delta))\mathbf{1}\{z\le V\le z+|\Delta|\}\big\}\big|.
\end{aligned}$$

But now, by Cauchy's inequality applied to the quantities $\sqrt2|\eta(\delta)-\mathbb{E}\eta(\delta)|$ and $\mathbf{1}\{z\le V\le z+|\Delta|\}/\sqrt2$,

$$\begin{aligned}
&\big|\mathbb{E}\big\{(\eta(\delta)-\mathbb{E}\eta(\delta))\mathbf{1}\{z\le V\le z+|\Delta|\}\big\}\big|\\
&\qquad\le \tfrac12\big\{\tfrac12\mathbb{P}[z\le V\le z+|\Delta|]+2\mathrm{Var}\eta(\delta)\big\},
\end{aligned}$$

and the lemma follows from (2.11), (2.12), (2.13) and (2.14). ■

The next step is to choose a suitable value of δ, and then to bound $\mathrm{Var}\eta(\delta)$. For the first, note that $\min(x,\delta)\ge x-x^2/(4\delta)$ in $x\ge0$, so that

$$\begin{aligned}
&\tfrac14 n\mathbb{E}\{|V'-V|\min(|V'-V|,\delta)\}\\
&\quad\ge \tfrac14 n\mathbb{E}\left\{|V'-V|(|V'-V|-|V'-V|^2/(4\delta))\right\}\\
&\quad= 1-\frac{n}{16\delta}\mathbb{E}|V'-V|^3\ge\frac34
\end{aligned}$$

if $\delta\ge\frac14 n\mathbb{E}|V'-V|^3$. Now, from (2.8),

$$\begin{aligned}
\mathbb{E}|V'-V|^3 &= 8\sigma^{-3}(n-2)^3\mathbb{E}|H(I,J;\pi)|^3\\
&\le 8(n-2)\sigma^{-3}\sum\nolimits_{i,j}'\mathbb{E}|h(i,j,\pi(i),\pi(j))|^3\\
&\le 8n^{-1}\sigma^{-3}\sum\nolimits_{i,j}'\sum\nolimits_{l,m}'|h(i,j,l,m)|^3. \qquad (2.15)
\end{aligned}$$

Then, from the definition of h,

$$|h(i,j,l,m)|^3\le 16\{|a_i^*b_l^*|^3+|a_j^*b_m^*|^3+|a_i^*b_m^*|^3+|a_j^*b_l^*|^3\}, \qquad (2.16)$$

implying that

$$\mathbb{E}|V' - V|^3 \le 8n^{-1}\sigma^{-3}\, 64(n-1)^2\, nA_{13}\, nB_{13};$$

hence (2.10) is satisfied with the choice

$$\delta = 128\delta_1, \quad \text{where} \quad \delta_1 := n^4\sigma^{-3}A_{13}B_{13}. \tag{2.17}$$

The bound on $\mathrm{Var}\eta(\delta)$ follows from the next lemma.

Lemma 2.3: *Suppose that $n \ge 4$ and that $d(i,j,r,s)$, $1 \le i,j,r,s \le n$, are such that $\sum_{r,s}' d(i,j,r,s) = 0$, for all $1 \le i,j \le n$. Then it follows, setting $X_{ij} := d(i,j,\pi(i),\pi(j))$ for π uniformly distributed on S_n, that*

$$\mathrm{Var}\left\{\sum\nolimits_{i,j}' X_{ij}\right\} \le 8n^3 D_2(1 + 5n^{-1}),$$

where

$$D_2 := \{(n)_2\}^{-2} \sum\nolimits_{i,j}' \sum\nolimits_{r,s}' d^2(i,j,r,s).$$

Proof: The conditions on $d(i,j,r,s)$ imply that $\mathbb{E}X_{ij} = 0$ for all i,j, so that

$$\mathrm{Var}\left\{\sum\nolimits_{i,j}' X_{ij}\right\} = \sum\nolimits_{i,j}' \sum\nolimits_{l,m}' \mathbb{E}\{X_{ij}X_{lm}\}.$$

We now note that, for i,j,l,m all distinct,

$$\mathbb{E}\{X_{ij}X_{lm}\} = \frac{1}{(n)_4} \sum\nolimits_{r,s,t,u}' d(i,j,r,s)d(l,m,t,u),$$

where, as usual, the sum runs over all ordered quadruples of *distinct* elements of $[n]$. But now, because of the conditions on $d(i,j,r,s)$, we have

$$\mathbb{E}\{X_{ij}X_{lm}\} = \frac{1}{(n)_4} \sum\nolimits_{r,s}' d(i,j,r,s)$$
$$\times\left\{-\sum_{u:u\ne r,s} d(l,m,r,u) - \sum_{u:u\ne s,r} d(l,m,s,u) - \sum_{t:t\ne r} d(l,m,t,r) - \sum_{t:t\ne s} d(l,m,t,s)\right\},$$

so that, summing over all distinct choices of i,j,l,m, and also recalling the inequality $|xy| \le (x^2+y^2)/2$, we have

$$\begin{aligned}
&\Big|{\sum}'_{i,j,l,m} \mathbb{E}\{X_{ij}X_{lm}\}\Big| \\
&\le \frac{1}{(n)_4} {\sum}'_{i,j} {\sum}'_{l,m} {\sum}'_{r,s} \\
&\quad \times \Bigg\{ \sum_{u:u\ne r} |d(i,j,r,s)d(l,m,r,u)| + \sum_{u:u\ne s} |d(i,j,r,s)d(l,m,s,u)| \\
&\qquad + \sum_{t:t\ne r} |d(i,j,r,s)d(l,m,t,r)| + \sum_{t:t\ne s} |d(i,j,r,s)d(l,m,t,s)| \Bigg\} \\
&\le \frac{4}{(n)_4} n(n-1)^2 \sum_{r=1}^{n} {\sum}'_{i,j} \sum_{s:s\ne r} d^2(i,j,r,s) \\
&= \frac{4n^3(n-1)^4}{(n)_4} D_2. \qquad (2.18)
\end{aligned}$$

If indices among i,j,l,m are equal, the argument does not use cancellation in the d-sums. First, for $l=i$, we have

$$\mathbb{E}\{X_{ij}X_{im}\} = \frac{1}{(n)_3} {\sum}'_{r,s,u} d(i,j,r,s)d(i,m,r,u),$$

so that

$$\begin{aligned}
&\Big|{\sum}'_{i,j,m} \mathbb{E}\{X_{ij}X_{im}\}\Big| \qquad (2.19) \\
&\le \frac{1}{(n)_3} \sum_{i=1}^{n} \sum_{j:j\ne i} \sum_{m:m\ne i} \sum_{r=1}^{n} \sum_{s:s\ne r} \sum_{u:u\ne r} |d(i,j,r,s)d(i,m,r,u)| \\
&\le \frac{1}{(n)_3} \sum_{i=1}^{n} \sum_{r=1}^{n} (n-1)^2 \sum_{j:j\ne i} \sum_{s:s\ne r} d^2(i,j,r,s) \\
&= \frac{n^2(n-1)^4}{(n)_3} D_2. \qquad (2.20)
\end{aligned}$$

The same bound also holds for the three remaining arrangements with one pair of equal indices. Finally, we have

$${\sum}'_{i,j} \mathbb{E}\{X_{ij}^2\} = {\sum}'_{i,j} {\sum}'_{r,s} \frac{1}{(n)_2} d^2(i,j,r,s) = \frac{n^2(n-1)^2}{(n)_2} D_2, \qquad (2.21)$$

and the same bound follows also for ${\sum}'_{i,j} \mathbb{E}\{X_{ij}X_{ji}\}$.

Collecting the results in (2.18) – (2.21), we have shown that

$$\mathrm{Var}\left\{\sum\nolimits_{i,j}' X_{ij}\right\} \le D_2\left\{\frac{4n^3(n-1)^4}{(n)_4} + \frac{4n^2(n-1)^4}{(n)_3} + \frac{2n^2(n-1)^2}{(n)_2}\right\},$$

and the lemma follows by using elementary computations. ∎

Now $\eta(\delta)$ is almost of the form considered in the previous lemma:

$$\eta(\delta) = \tfrac{1}{4}n^{-1}\sum\nolimits_{i,j}' w(i,j,\pi(i),\pi(j)),$$

where

$$w(i,j,r,s) := 4\sigma^{-2}(n-2)^2|h(i,j,r,s)|\min\left\{|h(i,j,r,s)|, \frac{\sigma\delta}{2(n-2)}\right\}.$$

Hence we can apply Lemma 2.3 to calculate $\mathrm{Var}\eta(\delta)$ by defining

$$d(i,j,r,s) := \tfrac{1}{4}n^{-1}\left\{w(i,j,r,s) - (n)_2^{-1}\sum\nolimits_{l,m}' w(i,j,l,m)\right\},$$

noting that then

$$\begin{aligned} D_2 &\le \{(n)_2\}^{-2}\sum\nolimits_{i,j}'\sum\nolimits_{r,s}' w^2(i,j,r,s) \\ &\le \left(\frac{(n-2)\delta}{2n\sigma}\right)^2\{(n)_2\}^{-2}\sum\nolimits_{i,j}'\sum\nolimits_{r,s}' h^2(i,j,r,s) \\ &= \left(\frac{n-2}{n-1}\right)^2\frac{\delta^2}{\sigma^2}A_{12}B_{12}, \end{aligned}$$

giving

$$\mathrm{Var}\eta(\delta) \le 2\delta^2(1+5n^{-1})n/(n-1) \le 2\delta^2(1+8n^{-1}), \qquad (2.22)$$

uniformly in $n \ge 4$. This completes the preparation.

Theorem 2.4: *For W defined as in (2.1), we have*

$$\sup_z |I\!P[W-\mu \le \sigma z] - \Phi(z)| \le (2+C)\delta + 12\delta^2 + (1+\sqrt{2})\tilde{\delta}_2,$$

where $\delta := 128\delta_1$, $\delta_1 = n^4\sigma^{-3}A_{12}B_{12}$, $\tilde{\delta}_2$ is as defined in Lemma 2.1, Φ denotes the standard normal distribution function, and C is the constant in Bolthausen's (1984) theorem.

Proof: We use the chain of inequalities

$$\begin{aligned} -\mathbb{P}[z-|\Delta|<V\le z] &= -\mathbb{P}[V\le z]+\mathbb{P}[\sigma^{-1}(W-\mu)-\Delta\le z-|\Delta|] \\ &\le \mathbb{P}[\sigma^{-1}(W-\mu)\le z]-\mathbb{P}[V\le z] \\ &\le \mathbb{P}[\sigma^{-1}(W-\mu)-\Delta\le z+|\Delta|]-\mathbb{P}[V\le z] \\ &= \mathbb{P}[z\le V\le z+|\Delta|], \end{aligned}$$

to give, for all z,

$$|\mathbb{P}[\sigma^{-1}(W-\mu)\le z]-\mathbb{P}[V\le z]|\le(1+\sqrt{2})\tilde{\delta}_2+2\delta+12\delta^2,$$

from Lemma 2.2 and (2.22), using the choice $\delta=128\delta_1$ from (2.17). This, together with Bolthausen's bound (1.3)

$$\sup_z|\mathbb{P}[V\le z]-\Phi(z)|\le C\delta_1,$$

completes the proof. ■

Remark: Note that our approximation, in common with that of Zhao *et al.* (1997), normalizes W with σ, and not with $\sqrt{\mathrm{Var}W}$. However, because $\mathrm{Var}\widetilde{\Delta}/\mathrm{Var}\widetilde{V}=\tilde{\delta}_2^2$, we could use $\sqrt{\mathrm{Var}W}$ instead, without affecting the order of the error.

3. General arrays

We now consider the more general statistic

$$X:=\sum_{i=1}^{n}\sum_{j=1}^{n}C(i,j;\pi(i),\pi(j)) \tag{3.1}$$

studied by Zhao *et al.* (1997), where C is an arbitrary 4-dimensional array. The first step is to split the sum into a linear and a quadratic part, as in (2.3) and (2.4). Noting that the terms with $i=j$ fall naturally into the linear part, we adopt a notation which treats them separately. We define 'off-diagonal' sums, in which replacing an index at any position by a '+' sign denotes summation over all indices at this position which, combined with the other index of the (first or last) pair, do not result in a pair of identical indices: thus

$$C(+,j;l,m):=\sum_{i:i\ne j}C(i,j;l,m);\quad C(+,+;l,m):={\sum}'_{i,j}C(i,j;l,m)$$

$$\text{and}\quad C(i,+;+,m):=\sum_{j:j\ne i}\sum_{l:l\ne m}C(i,j;l,m).$$

We then define the 'diagonal' sums, in which replacing a pair of indices by '(++)' indicates summation over all pairs of equal values: thus

$$C((++);l,m) := \sum_{i=1}^{n} C(i,i;l,m) \text{ and } C((++);(++)) := \sum_{i=1}^{n}\sum_{l=1}^{n} C(i,i;l,l).$$

With this notation, we set

$$\begin{aligned}
&\widetilde{C}(i,j;l,m) := C(i,j;l,m) \\
&- \frac{n-1}{n(n-2)}\{C(i,+;l,m) + C(+,j;l,m) + C(i,j;l,+) + C(i,j;+,m)\} \\
&- \frac{1}{n(n-2)}\{C(+,i;l,m) + C(j,+;l,m) + C(i,j;+,l) + C(i,j;m,+)\} \\
&+ \frac{1}{(n-1)(n-2)}\{C(+,+;l,m) + C(i,j;+,+)\} \\
&+ \left(\frac{n-1}{n(n-2)}\right)^2 C[1] \;+\; \frac{n-1}{n^2(n-2)^2}C[2] \;+\; \frac{1}{n^2(n-2)^2}C[3] \\
&- \frac{C[4]}{n(n-2)^2} \;-\; \frac{C[5]}{n(n-1)(n-2)^2} \;+\; \frac{C(+,+;+,+)}{(n-1)^2(n-2)^2}
\end{aligned}$$

and

$$\begin{aligned}
&2(n-2)c^*(i,l) := \frac{n-1}{n(n-2)}\{C(i,+;l,+) + C(+,i;+,l)\} \\
&+ \frac{1}{n(n-2)}\{C(i,+;+,l) + C(+,i;l,+)\} \\
&- \frac{1}{n(n-2)}\{C(i,+;+,+) + C(+,i;+,+) + C(+,+;l,+) + C(+,+;+,l)\} \\
&+ \frac{2}{n^2(n-2)}C(+,+;+,+) \;+\; C(i,i;l,l) \\
&- n^{-1}\{C(i,i;(++)) + C((++);l,l)\} + n^{-2}C((++);(++)),
\end{aligned}$$

where

$$\begin{aligned}
C[1] &:= C(i,+;l,+) + C(i,+;+,m) + C(+,j;l,+) + C(+,j;+,m); \\
C[2] &:= C(+,i;l,+) + C(+,i;+,m) + C(j,+;l,+) + C(j,+;+,m) \\
&\quad + C(i,+;+,l) + C(i,+;m,+) + C(+,j;+,l) + C(+,j;m,+); \\
C[3] &:= C(+,i;+,l) + C(+,i;m,+) + C(j,+;+,l) + C(j,+;m,+); \\
C[4] &:= C(i,+;+,+) + C(+,j;+,+) + C(+,+;l,+) + C(+,+;+,m)
\end{aligned}$$

and

$$C[5] \;:=\; C(+,i;+,+) + C(j,+;+,+) + C(+,+;+,l) + C(+,+;m,+).$$

Note that then

$$\widetilde{C}(+,j;l,m) = \widetilde{C}(i,+;l,m) = \widetilde{C}(i,j;+,m) = \widetilde{C}(i,j;l,+) = 0 \qquad (3.2)$$

and that

$$\sum_{i=1}^{n} c^*(i,l) = \sum_{l=1}^{n} c^*(i,l) = 0. \qquad (3.3)$$

With these definitions, X can be decomposed into linear and quadratic parts:

$$X = \mu + \widetilde{V} + \widetilde{\Delta}, \qquad (3.4)$$

where

$$\begin{aligned} \mu &:= \frac{1}{n(n-1)} C(+,+;+,+) \; + \; \frac{1}{n} C((++);(++)); \\ \widetilde{V} \; &= \; \widetilde{V}(\pi) := 2(n-2) \sum_{i=1}^{n} c^*(i,\pi(i)); \\ \widetilde{\Delta} \; &= \; \widetilde{\Delta}(\pi) := \sum\nolimits'_{i,j} \, \widetilde{C}(i,j;\pi(i),\pi(j)). \end{aligned} \qquad (3.5)$$

This decomposition is the analogue of that given in (2.3) and (2.4), and is slightly different from the version given in Zhao *et al.* (1997); here, the terms $\widetilde{V}$ and $\widetilde{\Delta}$ are uncorrelated. We write

$$C_{12} := n^{-2} \sum_{i=1}^{n} \sum_{l=1}^{n} \{c^*(i,l)\}^2; \; C_{22} := \{(n)_2\}^{-2} \sum\nolimits'_{i,j} \sum\nolimits'_{l,m} \{\widetilde{C}(i,j;l,m)\}^2,$$

assuming that $C_{12} > 0$, to avoid trivial exceptions.

Lemma 3.1: *With the definitions above, we have*

$$\mathbb{E}\widetilde{V} = \mathbb{E}\widetilde{\Delta} = \mathbb{E}\{\widetilde{V}\widetilde{\Delta}\} = 0; \quad \sigma^2 := \mathbb{E}\widetilde{V}^2 = \frac{4n^2(n-2)^2}{n-1} C_{12}$$

and

$$\mathbb{E}\widetilde{\Delta}^2 \leq \frac{2n(n-1)^2}{n-3} C_{22}.$$

Proof: We indicate the calculations for $\mathbb{E}\widetilde{\Delta}^2$, which are the most complicated. First, for i, j, l, m distinct, using (3.2), we have

$$\begin{aligned}
&\mathbb{E}\{\widetilde{C}(i,j;\pi(i),\pi(j))\widetilde{C}(l,m;\pi(l),\pi(m))\} \\
&\quad= \frac{1}{(n)_4}\sum\nolimits'_{r,s,t,u} \widetilde{C}(i,j;r,s)\widetilde{C}(l,m;t,u) \\
&\quad= -\frac{1}{(n)_4}\sum\nolimits'_{r,s,t} \widetilde{C}(i,j;r,s)\{\widetilde{C}(l,m;t,r)+\widetilde{C}(l,m;t,s)\} \\
&\quad= \frac{1}{(n)_4}\sum\nolimits'_{r,s} \widetilde{C}(i,j;r,s)\{\widetilde{C}(l,m;s,r)+\widetilde{C}(l,m;r,s)\}.
\end{aligned}$$

Then, adding over i, j, l, m distinct and again using (3.2), it follows that

$$\begin{aligned}
&\left|\sum\nolimits'_{i,j,l,m} \mathbb{E}\{\widetilde{C}(i,j;\pi(i),\pi(j))\widetilde{C}(l,m;\pi(l),\pi(m))\}\right| \\
&= \left|\frac{1}{(n)_4}\sum\nolimits'_{l,m}\sum\nolimits'_{r,s}\{\widetilde{C}(l,m;r,s)+\widetilde{C}(m,l;r,s)\}\{\widetilde{C}(l,m;s,r)+\widetilde{C}(l,m;r,s)\}\right| \\
&\le \frac{4n(n-1)}{(n-2)(n-3)}\, C_{22},
\end{aligned}$$

the last line following because $|xy| \le \frac{1}{2}(x^2+y^2)$. Similar calculations for the cases in which the pairs (i,j) and (l,m) have one index in common give a total contribution of at most $4n(n-1)(n-2)^{-1}C_{22}$, and the terms with $(l,m) = (i,j)$ and $(l,m) = (j,i)$ yield at most $2n(n-1)C_{22}$. Adding these three elements gives the stated bound for $\mathbb{E}\widetilde{\Delta}^2$. ∎

We now write $V = V(\pi) := \sigma^{-1}\widetilde{V}(\pi)$ and $\Delta = \Delta(\pi) := \sigma^{-1}\widetilde{\Delta}(\pi)$, and set

$$\tilde{\delta}_2^2 := \mathrm{Var}\Delta = \sigma^{-2}\mathbb{E}\widetilde{\Delta}^2 \le \frac{(n-1)^3}{2n(n-2)^2(n-3)}\,\frac{C_{22}}{C_{12}}.$$

We then define π' as in (2.5), and write $V' = V(\pi')$ and $\Delta' = \Delta(\pi')$, giving the exchangeable pair (V, Δ), (V', Δ').

Lemma 3.2: *With the above definitions, we have* $\mathbb{E}\{V - V' \mid \pi\} = 2n^{-1}V$, $\mathbb{E}\{(V - V')^2\} = 4n^{-1}$, $\mathbb{E}|V - V'|^3 \le 512n^3\sigma^{-3}C_{13}$ *and then finally* $\mathbb{E}\{(\Delta - \Delta')^2\} \le 8n^{-1}\mathbb{E}\Delta^2 = 8n^{-1}\tilde{\delta}_2^2$, *where*

$$C_{13} := n^{-2}\sum_{i=1}^{n}\sum_{l=1}^{n}|c^*(i,l)|^3.$$

Proof: Much as in the proof of Lemma 2.1, we have

$$V' - V = 2\sigma^{-1}(n-2) \times\{c^*(I,\pi(J)) - c^*(I,\pi(I)) + c^*(J,\pi(I)) - c^*(J,\pi(J))\}\mathbf{1}\{I \neq J\},$$

giving

$$\begin{aligned}
&\mathbb{E}\{V - V' \,|\, \pi\} \\
&= \frac{2(n-2)}{n^2\sigma}\sum\nolimits'_{i,j}\{c^*(i,\pi(i)) - c^*(i,\pi(j)) + c^*(j,\pi(j)) - c^*(j,\pi(i))\} \\
&= \frac{2(n-2)}{n^2\sigma}\left\{\sum_{i=1}^n c^*(i,\pi(i)) + (n-1)\sum_{i=1}^n c^*(i,\pi(i)) + n\sum_{j=1}^n c^*(j,\pi(j))\right\} \\
&= 2n^{-1}V,
\end{aligned}$$

as required, by way of (3.3); the formula for $\mathbb{E}\{(V-V')^2\}$ now follows from (2.6). The argument for $\mathbb{E}|V-V'|^3$ then matches that given in (2.15) and (2.16), leading to the expression

$$\mathbb{E}|V - V'|^3 \le 8n^{-1}\sigma^{-3} \times 16\sum\nolimits'_{i,j}\sum\nolimits'_{l,m}\{|c^*(i,m)|^3 + |c^*(i,l)|^3 + |c^*(j,l)|^3 + |c^*(j,m)|^3\},$$

from which the result follows.

The corresponding formula for $\Delta' - \Delta$ is

$$\begin{aligned}
\sigma(\Delta' - \Delta) = &\sum_{j\neq I,J}\{\widetilde{C}(I,j;\pi(J),\pi(j)) - \widetilde{C}(I,j;\pi(I),\pi(j))\} \\
&+ \sum_{i\neq I,J}\{\widetilde{C}(i,I;\pi(i),\pi(J)) - \widetilde{C}(i,I;\pi(i),\pi(I))\} \\
&+ \sum_{j\neq I,J}\{\widetilde{C}(J,j;\pi(I),\pi(j)) - \widetilde{C}(J,j;\pi(J),\pi(j))\} \\
&+ \sum_{i\neq I,J}\{\widetilde{C}(i,J;\pi(i),\pi(I)) - \widetilde{C}(i,J;\pi(i),\pi(J))\} \\
&+ \widetilde{C}(I,J;\pi(J),\pi(I)) - \widetilde{C}(I,J;\pi(I),\pi(J)) \\
&+ \widetilde{C}(J,I;\pi(I),\pi(J)) - \widetilde{C}(J,I;\pi(J),\pi(I)).
\end{aligned}$$

The expectation conditional on π, using (3.2), now yields

$$\mathbb{E}\{\Delta - \Delta' \,|\, \pi\} = 4n^{-2}(n-1)\Delta + 2n^{-2}\Delta - 2n^{-2}\Delta^*,$$

where $\Delta^* := \sigma^{-1} \sum'_{i,j} \widetilde{C}(i,j;\pi(j),\pi(i))$. Hence, from (2.6), it follows that

$$\begin{aligned}\mathbb{E}\{(\Delta' - \Delta)^2\} &= 2\mathbb{E}\{\Delta(\Delta - \Delta')\} \\ &= 4n^{-2}\mathbb{E}\left\{(2n-1)\Delta^2 - \Delta\Delta^*\right\} \;\le\; 8n^{-1}\mathbb{E}\Delta^2,\end{aligned}$$

completing the proof. ■

The rest of the argument is exactly as for Theorem 2.4, with Lemma 2.2 and its proof remaining true in this context, with the new definitions of V, V' and $\tilde{\delta}_2$, with $h(i,j,r,s)$ replaced by $c^*(i,r) - c^*(i,s) + c^*(j,s) - c^*(j,r)$, and with $\delta = 128\delta_1$ being a suitable choice for δ, now with the new definition $\delta_1 := n^4\sigma^{-3}C_{13}$. This leads to the following theorem.

Theorem 3.3: *For X as defined in (3.1), μ as in (3.5) and σ as in Lemma 3.1, we have*

$$\sup_z |\mathbb{P}[X - \mu \le \sigma z] - \Phi(z)| \le (2 + C)\delta + 12\delta^2 + (1 + \sqrt{2})\tilde{\delta}_2,$$

where $\delta := 128\delta_1$, $\delta_1 := n^4\sigma^{-3}C_{13}$, $\tilde{\delta}_2^2 = \mathbb{E}\Delta^2 \le \frac{2n(n-1)^2C_{22}}{(n-3)\sigma^2}$, and C is the constant in Bolthausen's (1984) theorem.

As before, δ_1 is the order of the error bound in Bolthausen's Berry–Esseen bound for the normal approximation to the linear component $\widetilde{V}$, and $\tilde{\delta}_2$ represents the error incurred by neglecting $\widetilde{\Delta}$. In the bound given by Zhao *et al.* (1997), our $\tilde{\delta}_2$ is replaced by a term of order much like

$$\delta_3 := \sigma^{-3} \sum\nolimits'_{i,j} \sum\nolimits'_{l,m} |\widetilde{C}(i,j;l,m)|^3,$$

the difference lying only in the slight difference between their decomposition and ours. Now, by Hölder's inequality, we have $\sigma^3\delta_3\{(n)_2\}^{-2} \ge C_{22}^{3/2}$ and $\delta_1 \ge n^{-1/2}$. Hence

$$\delta_3 \ge \{(n)_2\}^2\sigma^{-3}C_{22}^{3/2} \asymp n\tilde{\delta}_2^3,$$

implying that δ_3 is of larger order than $\tilde{\delta}_2$ if $\tilde{\delta}_2 \gg n^{-1/2}$. Thus the bound given in Theorem 3.3 is always asymptotically of at least as small an order as that of Zhao *et al.* (1997), and is of strictly smaller order whenever $\tilde{\delta}_2 \gg \delta_1$.

Acknowledgement: ADB gratefully acknowledges support from the Institute for Mathematical Sciences of the National University of Singapore, while part of this work was completed.

References

1. O. Abe (1969) A central limit theorem for the number of edges in the random intersection of two graphs. *Ann. Math. Statist.* **40**, 144–151.
2. A. D. Barbour & G. K. Eagleson (1986a) Random association of symmetric arrays. *Stoch. Analysis Applics* **4**, 239–281.
3. A. D. Barbour & G. K. Eagleson (1986b) Tests for space time clustering. In: *Stochastic spatial processes*, Ed. P. Tautu, Lecture Notes in Mathematics **1212**, 42–51: Springer, Berlin.
4. A. R. Bloemena (1964) *Sampling from a graph.* Mathematisch Centrum, Amsterdam.
5. E. Bolthausen (1984) An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. verw. Geb.* **66**, 379–386.
6. L. H. Y. Chen & Q.-M. Shao (2004) Normal approximation for non-linear statistics using a concentration inequality approach. Preprint.
7. A. D. Cliff & J. K. Ord (1973) *Spatial autocorrelation.* Pion, London.
8. H. E. Daniels (1944) The relation between measures of correlation in the universe of sample permutations. *Biometrika* **33**, 129–135.
9. R. V. Erickson (1974) L_1 bounds for asymptotic normality of m-dependent sums using Stein's technique. *Ann. Probab.* **2**, 522–529.
10. J. H. Friedman & L. C. Rafsky (1983) Graph theoretic measures of multivariate association and prediction. *Ann. Statist.* **11**, 377–391.
11. R. C. Geary (1954) The contiguity ratio and statistical mapping. *The Incorporated Statistician* **5**, 115–145.
12. W. Hoeffding (1951) A combinatorial central limit theorem. *Ann. Math. Statist.* **22**, 558–566.
13. L. J. Hubert (1987) *Assignment methods in combinatorial data analysis.* Marcel Dekker, New York.
14. K. Jogdeo (1968) Asymptotic normality in nonparametric methods. *Ann. Math. Statist.* **39**, 905–922.
15. G. Knox (1964) Epidemiology of childhood leukaemia in Northumberland and Durham. *Brit. J. Prev. Soc. Med.* **18**, 17–24.
16. N. Mantel (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220.
17. P. A. P. Moran (1948) The interpretation of statistical maps. *J. Roy. Statist. Soc. B* **10**, 243–251.
18. C. P. Shapiro and L. J. Hubert (1979) Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions. *Ann. Statist.* **7**, 788–794.
19. C. Stein (1986) *Approximate computation of expectations.* IMS Lecture Notes – Monograph Series, vol. 7, Hayward, CA.
20. A. Wald & J. Wolfowitz (1944) Statistical tests based on permutation of the observations. *Ann. Math. Statist.* **15**, 358–372.
21. L. Zhao, Z. Bai, C.-C. Chao & W.-Q. Liang (1997) Error bound in a central limit theorem of double-indexed permutation statistics. *Ann. Statist.* **25**, 2210–2227.