MULTIVARIATE NORMAL APPROXIMATION WITH STEIN'S METHOD OF EXCHANGEABLE PAIRS UNDER A GENERAL LINEARITY CONDITION

By Gesine Reinert* and Adrian Röllin†

University of Oxford

In this paper we establish a multivariate exchangeable pairs approach within the framework of Stein's method to assess distributional distances to potentially singular multivariate normal distributions. By extending the statistics into a higher-dimensional space, we also propose an embedding method which allows for a normal approximation even when the corresponding statistics of interest do not lend themselves easily to Stein's exchangeable pairs approach. To illustrate the method, we provide the examples of runs on the line, the joint count of edges, two-stars and triangles in Bernoulli random graphs, complete U-statistics, and double-indexed permutation statistics.

1. Introduction. Stein's method was first published in Stein (1972) to assess the distance between univariate random variables and the normal distribution. This method has proved particularly powerful in the presence of both local dependence and weak global dependence.

A coupling at the heart of Stein's method for univariate normal approximation is the method of exchangeable pairs, see Stein (1986). Assume that W is a univariate random variable with $\mathbb{E}W = 0$ and $\mathbb{E}W^2 = 1$, and assume that W' is a random variable such that (W, W') makes an exchangeable pair. Assume further that there is a number $\lambda > 0$ such that the conditional expectation of $W' - W$ with respect to W satisfies

$$\mathbb{E}^W(W' - W) = -\lambda W. \tag{1.1}$$

Heuristically, (1.1) can be understood as as linear regression condition. If (W, W') were bivariate normal with correlation ρ, then

$$\mathbb{E}^W W' = \rho W,$$

*Supported in part by MMCOMNET Grant No. FP6-2003-BEST-Path-012999

†Supported by Swiss National Science Foundation project PBZH2–117033

AMS 2000 subject classifications: Primary 60F05; secondary 62E17

Keywords and phrases: Stein's method, multivariate normal approximation, random graphs, U-statistics, Hoeffding projection, exchangeable pairs

and (1.1) would be satisfied with $\lambda = 1 - \rho$. If W was close to normal, then so would be W', and it would not be unreasonable to assume that (1.1) is close to satisfied.

In this spirit, the univariate theorem of Stein (1986) has been extended by Rinott and Rotar (1997). With the same basic setup as in Stein (1986), they generalise (1.1) by assuming that there is a number $\lambda > 0$ and a random variable $R = R(W)$ such that

$$\mathbb{E}^W(W' - W) = -\lambda W + R. \tag{1.2}$$

Note that, unlike Condition (1.1), this is not a condition in the strict sense, as we can define $R := \mathbb{E}^W(W' - W) + \lambda W$ for any λ; however, we always have $\mathbb{E}R = 0$.

One of the results of Rinott and Rotar (1997) is that

$$\begin{aligned} &\sup_x \bigl|\mathbb{P}[W \leqslant x] - \mathbb{P}[Z \leqslant x]\bigr| \\ &\qquad \leqslant \frac{6}{\lambda}\sqrt{\operatorname{Var}\mathbb{E}^W(W'-W)^2} + \frac{6}{\lambda^{1/2}}\sqrt{\mathbb{E}|W'-W|^3} + \frac{19}{\lambda}\sqrt{\operatorname{Var} R}, \end{aligned} \tag{1.3}$$

where Z has standard normal distribution. So clearly, Representation (1.2) is useful only if $\lambda^{-1}\sqrt{\operatorname{Var} R} = \mathrm{o}(1)$. In this case, if λ_1 and λ_2 stem from two different representations (1.2) for which $\lambda_i^{-1}\sqrt{\operatorname{Var} R_i} = \mathrm{o}(1)$ for $i = 1, 2$, then it it easy to see that $|\lambda_1 - \lambda_2|/(\lambda_1 + \lambda_2) = o(1)$; in this sense, λ is asymptotically unique. Rinott and Rotar (1997) then apply bound (1.3) to the number of ones in the anti-voter model, and to weighted U-statistics. Röllin (2008) provides a proof of a variant of (1.3) which does not use exchangeability but only $\mathscr{L}(W') = \mathscr{L}(W)$; in Section 5 we shall discuss this for the multivariate setting in more detail.

Stein's method has been extended to many other distributions, for an overview see for example Reinert (2005). For multivariate normal approximations the method was first adapted by Barbour (1990) and Götze (1991), viewing the normal distribution as the stationary distribution of an Ornstein-Uhlenbeck diffusion, and using the generator of this diffusion as a characterising operator for the normal distribution. Subsequent authors have used this generator approach for multivariate normal approximation with different variants, such as the local approach and the size-biasing approach by Goldstein and Rinott (1996) and Rinott and Rotar (1996), and the zero-biasing approach by Goldstein and Reinert (2005).

The exchangeable pair approach in contrast, while having proved useful in non-normal contexts, see Chatterjee et al. (2005), Chatterjee and Fulman (2006) and Röllin (2007), remained restricted to the one-dimensional setting until very recently. A main stumbling block was that the extension of Condition (1.2) to the multivariate setting is not obvious from the view point of Stein's method.

In Chatterjee and Meckes (2007), this issue was finally addressed. They propose the condition that for all $i = 1, \dots, d$,

$$\mathbb{E}^W(W_i' - W_i) = -\lambda W_i, \tag{1.4}$$

for a fixed number λ, where now $W = (W_1, \dots, W_d)$ and $W' = (W_1', \dots, W_d')$ are identically distributed d-vectors with uncorrelated components an extension to the additional remainder term R was not considered, but would be straightforward). They employ such couplings to bound the distance to the standard multivariate normal distribution. Using the same argument as Röllin (2008), Chatterjee and Meckes (2007) are able to give proofs of their theorems without using exchangeability and apply them successfully to various multivariate applications.

Heuristically, however, if (W, W') were jointly normal, with mean vector 0 and covariance matrix

$$\Sigma_0 = \begin{pmatrix} \Sigma & \tilde{\Sigma} \\ \tilde{\Sigma} & \Sigma \end{pmatrix}, \tag{1.5}$$

then $\mathbb{E}^W W' = \tilde{\Sigma}\Sigma^{-1}W$ (see for example Mardia et al. (1979), p.63, Theorem 3.2.4.), in which case

$$\mathbb{E}^W(W' - W) = -(\mathrm{Id} - \tilde{\Sigma}\Sigma^{-1})W; \tag{1.6}$$

here Id denotes the identity matrix. Again, if (W, W') is approximately jointly normal, then we expect (1.6) to be approximately satisfied. This heuristic leads to the condition that

$$\mathbb{E}^W(W' - W) = -\Lambda W + R \tag{1.7}$$

for an invertible $d \times d$ matrix Λ and a remainder term $R = R(W)$. Even if $\Sigma = \mathrm{Id}$ we would obtain $\Lambda = \mathrm{Id} - \tilde{\Sigma}$, which in general is not diagonal. Hence we argue that (1.7) is not only more general, but also more natural than (1.4).

Different exchangeable pairs will lead to different Λ and R in (1.7); our embedding method suggests suitable decompositions. Indeed, for a specific exchangeable pair (W, W') at hand it is often far from obvious whether this pair will satisfy the linearity condition (1.7) with R of the required small order, unless equal to zero. Consider the case of 2-runs. For a sequence of i.i.d. Bernoulli distributed random variables $\xi_1, \dots, \xi_n$ such that $\mathbb{P}[\xi_1 = 1] = p$, define the centered number of 2-runs

$$V_2 = \sum_{i=1}^{n} \xi_i \xi_{i+1} - np^2$$

where we let $\xi_{n+1} := \xi_1$. The most natural construction of an exchangeable pair in the spirit of Stein (1986) is be to pick uniformly a ξ_i and replace it by an independent copy ξ_i'. Denote by V_2' the resulting number of 2-runs in the new sequence. It is easy to calculate (see Subsection 4.2) that

$$\mathbb{E}^{V_2}(V_2' - V_2) = -\frac{2}{n} V_2 + \frac{2p}{n} \mathbb{E}^{V_2} \sum_{i=1}^{n} (\xi_i - p). \tag{1.8}$$

The conditional expectation on the right hand side of (1.8) is hard to calculate. Furthermore, it has the same order of magnitude as V_2. Also, the weighted U-statistics approach of Rinott and Rotar (1997) (Proposition 1.2) does not yield convergent bounds to the normal distribution. We propose the following approach to this problem. Keeping the above coupling, we define $V_1 := \sum_{i=1}^{n} \xi_i - np$ (and V_1' accordingly) and consider the problem as a 2-dimensional problem $W := \binom{V_1}{V_2}$. Eq. (1.8) now yields $\mathbb{E}^W(V_2' - V_2) = -\frac{2}{n} V_2 + \frac{2p}{n} V_1$, and further calculations reveal that $\mathbb{E}^W(V_1' - V_1) = -\frac{1}{n} V_1$, so that now (1.7) holds with

$$\Lambda = \frac{1}{n} \begin{bmatrix} 1 & 0 \\ -2p & 2 \end{bmatrix}$$

and $R = 0$. Using this embedding into a higher-dimensional setting, the problem now fits into our framework and allows not only for a normal approximation of the primary statistic but for an approximation of the joint distribution of the primary and auxiliary statistics. For this embedding method, the generality of Condition (1.7) is essential, see (4.1) later.

The rest of the article is organised as follows. In the next section we prove an abstract non-singular multivariate normal approximation theorem for smooth test functions, Theorem 2.1. The explicit bound on the distance to the normal distribution is given in terms of the conditional variance, the absolute third moments, and the variance of the remainder term. Proposition 2.9 gives the extension to singular multivariate normal distributions, using Stein's method and the triangle inequality. To illustrate our results, we calculate the example of sums of i.i.d. variables.

Section 3 uses the abstract theorem to obtain a similar result for non-smooth test functions, such as indicators of convex sets. Adapting the approach by Rinott and Rotar (1996) to general multivariate normal approximation, Corollary 3.1 and Corollary 3.3 display how the main terms involved in the error bounds for smooth test functions simply re-appear in the bounds for non-smooth test functions.

Section 4 discusses the above mentioned embedding method and provides as detailed examples runs on the line, the joint counts of edges, two-stars and triangles in a Bernoulli random graph, and complete U-statistics. The latter two examples involve not only auxiliary random variables but also a covariance matrix which is asymptotically singular. While in the last two examples multivariate normal approximations are known, see Janson and Nowicki (1991) for the multivariate graph motif count problem and Lee (1990) for U-statistics, we are not aware of an explicit bound on the distance to the non-standard normal distribution. We also sketch the application to double-indexed permutation statistics, as an example which is not of U-statistic type.

The generality of (1.7) comes at the extra cost that now exchangeability seems almost inevitable. Indeed, in view of Röllin (2008), we were surprised that, in the multivariate setting, the exchangeability condition cannot be removed as easily as in the one-dimensional case. Therefore, the last section discusses the exchangeability condition, Condition (1.7) and their implications. We also propose a possible solution around this problem. Using an approach with a different Stein operator, for which the drift term is allowed to be non-trivial, the exchangeability condition could be removed. But the price to pay would be rather a technical set-up; instead, exchangeability makes the approach in the present article relatively easy to implement.

Standard proofs of auxiliary results are found in Appendix A, whereas details for the examples are in Appendix B.

1.1. *Notation.* Random vectors in $\mathbb{R}^d$ are written in the form $W = (W_1, W_2, \ldots, W_d)^t$, where W_i are $\mathbb{R}$-valued random variables for $i = 1, \ldots, d$. If Σ is a symmetric, non-negative definite matrix, we denote by $\Sigma^{1/2}$ the unique symmetric, non-negative definite square root of Σ. Denote by Id the identity matrix, usually of dimension d. Throughout this article, Z will denote a random variable having standard multivariate normal distribution, also of dimension d.

For ease of presentation we abbreviate the transpose of the inverse of a matrix in the form $\Lambda^{-t} := (\Lambda^{-1})^t$.

Stein's method makes good use of Taylor expansions. For derivatives of smooth functions $h : \mathbb{R}^d \to \mathbb{R}$, we use the notation ∇ for the gradient operator. For the sake of presentation the partial derivatives are abbreviated as $h_i = \frac{\partial}{\partial x_i} h$, $h_{i,j} = \frac{\partial^2}{\partial x_i \partial x_j} h$ unless we would like to emphasise the dependence on the variables.

To derive uniform bounds we shall employ the supremum norm, denoted by $\| \cdot \|$ for both functions and matrices. For a function $h : \mathbb{R}^d \to \mathbb{R}$, we abbreviate $|h|_1 := \sup_i \| \frac{\partial}{\partial x_i} h \|$, $|h|_2 := \sup_{i,j} \| \frac{\partial^2}{\partial x_i \partial x_j} h \|$, and so on, if the corresponding derivatives exist.

2. The distance to multivariate normal distribution in terms of smooth test functions. Firstly we derive a bound on the distance between a multivariate target distribution and a multivariate normal distribution with the same, positive definite covariance matrix. We start by considering smooth test functions; the case of non-smooth test functions will be treated in Section 3.

THEOREM 2.1. *Assume that (W, W') is an exchangeable pair of $\mathbb{R}^d$-valued random variables such that*

$$\mathbb{E}W = 0, \qquad \mathbb{E}WW^t = \Sigma, \tag{2.1}$$

with $\Sigma \in \mathbb{R}^{d \times d}$ symmetric and positive definite. Suppose further that (1.7) is satisfied for an invertible matrix Λ and a $\sigma(W)$-measurable random variable R. Then, if Z has d-dimensional standard normal distribution, we have for every three times differentiable function h,

$$\big|\mathbb{E}h(W) - \mathbb{E}h(\Sigma^{1/2}Z)\big| \leqslant \frac{|h|_2}{4} A + \frac{|h|_3}{12} B + \Big(|h|_1 + \tfrac{1}{2} d \|\Sigma\|^{1/2} |h|_2\Big) C, \tag{2.2}$$

where, with $\lambda^{(i)} := \sum_{m=1}^d |(\Lambda^{-1})_{m,i}|$,

$$A = \sum_{i,j=1}^d \lambda^{(i)} \sqrt{\operatorname{Var} \mathbb{E}^W (W_i' - W_i)(W_j' - W_j)},$$

$$B = \sum_{i,j,k=1}^d \lambda^{(i)} \mathbb{E}\big|(W_i' - W_i)(W_j' - W_j)(W_k' - W_k)\big|,$$

$$C = \sum_{i=1}^d \lambda^{(i)} \sqrt{\operatorname{Var} R_i}.$$

Before we proceed with the proof, we illustrate Theorem2.1 by means of the simple example of sums of i.i.d. random variables and make also some further remarks.

COROLLARY 2.2. *Suppose that $W = (W_1, \dots, W_d)$ is such that, for each i, $W_i = \sum_{j=1}^n X_{i,j}$, where $X_{i,j}, i = 1, \dots, d, j = 1, \dots, n$, are i.i.d. with mean zero and variance $\frac{1}{n}$, so that the covariance matrix $\Sigma = \mathrm{Id}$. Assume further that*

$$\begin{aligned} \mathbb{E}|X_{i,j}|^3 &= \beta n^{-3/2} \text{ for some } \beta < \infty, \\ \mathrm{Var}(X_{i,j}^2) &= \gamma n^{-2} \text{ for some } \gamma < \infty. \end{aligned}$$

Then, for every three times differentiable function h,

$$\left|\mathbb{E}h(W) - \mathbb{E}h(Z)\right| \leqslant \frac{d}{\sqrt{n}}\left(\frac{\sqrt{\gamma}}{4}|h|_2 + \frac{\beta}{6}|h|_3\right).$$

PROOF. We construct an exchangeable pair by choosing a vector I and a summand J uniformly, such that $\mathbb{P}(I = i, J = j) = 1/dn$. If $I = i, J = j$, we replace $X_{i,j}$ by an independent copy $X'_{i,j}$; all other variables remain unchanged. Put

$$W'_I = W_I - X_{I,J} + X'_{I,J};$$

and $W'_k = W_k$ for $k \neq I$; denote by W' the resulting d-vector. Then (W, W') is exchangeable, and, in (1.7),

$$\Lambda = \frac{1}{dn}\,\mathrm{Id}$$

with $R = 0$ and hence $C = 0$. For our bounds we note that $\lambda^{(i)} = dn$. We calculate that

$$\begin{aligned} \mathbb{E}^W(W'_i - W_i)^2 &= \frac{1}{dn}\sum_{\ell=1}^d \mathbf{1}(\ell = i)\sum_{j=1}^n \mathbb{E}^W(X'_{i,j} - X_{i,j})^2 \\ &= \frac{1}{dn} + \frac{1}{dn}\sum_j \mathbb{E}^W X_{i,j}^2. \end{aligned}$$

Thus

$$\mathrm{Var}\,\mathbb{E}^W(W'_i - W_i)^2 \leq \frac{1}{d^2n^2}\sum_j \mathrm{Var}\,X_{ij}^2 \leq \frac{\gamma}{n^3d^2}.$$

Moreover, by construction, for $i \neq k$, almost surely $(W'_i - W_i)(W'_k - W_k) = 0$, and $(W'_i - W_i)(W'_k - W_k)(W'_l - W_l) = 0$, unless $i = k = l$. By assumption,

$$\mathbb{E}|W'_i - W_i|^3 = \frac{1}{dn}\sum_{\ell=1}^d \mathbf{1}(\ell = i)\sum_{j=1}^n \mathbb{E}|X_{i,j} - X'_{i,j}|^3 \leq \frac{2\beta}{dn^{3/2}}.$$

The result now follows directly from Theorem 2.1. ∎

Remark 2.3. Multivariate normal approximations for vectors of sums of i.i.d. random variables have been so intensively studied that there is not enough space to review all the results. The approach most similar to ours is found in Chatterjee and Meckes (2007), where instead of exchanging only one summand, a whole vector would be exchanged. Their results yield

$$|\mathbb{E}h(W)-\Phi h| \quad \leqslant \frac{d^{3/2}\sqrt{\gamma+1}}{2\sqrt{n}}|h|_1+4\frac{d^3\beta}{\sqrt{n}}|h|_2.$$

Due to the different Stein equation used, the dependence on the dimension differs, and the bounds are in terms of different derivatives of the test function. The overall similarity in this special case is apparent.

Remark 2.4. Assume that (1.7) is satisfied. What can we then say about the applicability of Theorem 2.1 to $V=AW$, where A is a $k\times d$-matrix with $k\leqslant d$? As the examples of U-statistics and permutation statistics show, we often have that, if (1.7) is satisfied, then it will also be satisfied for lower-dimensional projections (although often with a complicated remainder term R). This is no conincidence. As mentioned already in the Introduction, if W converges to a normal distribution and (W,W') satisfies (1.7), we expect that (W,W') converges jointly to a multivariate normal distribution. Hence, we then also have that (AW,AW') converges jointly to a multi-variate normal distribution with covariance matrix

$$\begin{pmatrix} A\Sigma A^t & A\tilde{\Sigma}A^t \\ A\tilde{\Sigma}A^t & A\Sigma A^t \end{pmatrix},$$

so that, from (1.6), we expect that (1.7) holds in the form

$$\mathbb{E}^{AW}(AW'-AW)=(\mathrm{Id}-A\tilde{\Sigma}A^t(A\Sigma A^t)^{-1})AW+R, \tag{2.3}$$

with R being of the required lower order. However, the example of d-runs shows that things can be more subtle; see Remark 4.3.

Remark 2.5. If we were to normalise the random variables in Theorem 2.1, denoting the normalisation of W by $\hat{W}:=\Sigma^{-1/2}W$ and $\hat{W}'=\Sigma^{-1/2}W'$, then, the conditions of the theorem remain satisfied for $(\hat{W},\hat{W}')$ with $\hat{\Sigma}=\mathrm{Id}$ and $\hat{\Lambda}=\Sigma^{-1/2}\Lambda\Sigma^{1/2}$ as well as $\hat{R}=\Sigma^{-1/2}R$.

Remark 2.6. As a precursor to (1.7), in the context of multivariate zero-biasing, Goldstein and Reinert (2005) use the condition of the form (1.7) for Λ such that

$$\Lambda_{ij}=\begin{cases}\rho & \text{if } i\neq j\\ 1+\rho & \text{if } i=j.\end{cases}$$

After these remarks we proceed to the proof of Theorem 2.1, which is based on the Stein characterization of the normal distribution that $Y \in \mathbb{R}^d$ is a multivariate normal $\mathrm{MVN}(0, \Sigma)$ if and only if

$$\mathbb{E}\{\nabla^t \Sigma \nabla f(Y) - Y^t \nabla f(Y)\} = 0, \quad \text{for all smooth } f : \mathbb{R}^d \to \mathbb{R}. \tag{2.4}$$

We will need the following lemma to prove the theorem; however, see also Remark 2.5, Barbour (1990), Goldstein and Rinott (1996), and Götze (1991). The proof of Lemma 2.7 is routine (see Appendix A).

LEMMA 2.7. *Assume that $h : \mathbb{R}^d \to \mathbb{R}$ has 3 bounded derivatives. Then, if $\Sigma \in \mathbb{R}^{d \times d}$ is symmetric and positive definite, there is a solution $f : \mathbb{R}^d \to \mathbb{R}$ to the equation*

$$\nabla^t \Sigma \nabla f(w) - w^t \nabla f(w) = h(w) - \mathbb{E}h(\Sigma^{1/2} Z), \tag{2.5}$$

which holds for every $w \in \mathbb{R}^d$. The solution f satisfies the bounds

$$\left| \frac{\partial^k f(w)}{\prod_{j=1}^k \partial w_{i_j}} \right| \leqslant \frac{1}{k} \left| \frac{\partial^k h(w)}{\prod_{j=1}^k \partial w_{i_j}} \right| \tag{2.6}$$

for every $w \in \mathbb{R}^d$.

REMARK 2.8. Compared to the main theorem of Chatterjee and Meckes (2007), which only needs the existence of two derivatives, our Theorem 2.1 is more restrictive in the choice of test functions h. This reflects the fact that we make use of Lemma 2.7, which is motivated by Goldstein and Rinott (1996), whereas Chatterjee and Meckes (2007) prove new bounds on the solutions of (2.5), but only for $\Sigma = \mathrm{Id}$; see also Raič (2004) for similar results. The general result of Lemma 2.7, however, allows to work with the unstandardised pair (W, W') which not only usually simplifies the calculations, but also yields more informative bounds if the limiting covariance matrix is singular.

PROOF OF THEOREM 2.1. Our aim is to bound $|\mathbb{E}h(W) - \mathbb{E}h(\Sigma^{1/2} Z)|$ by bounding $|\mathbb{E}\{\nabla^t \Sigma \nabla f(W) - W^t \nabla f(W)\}|$, where f is the solution to the Stein equation (2.5). First we expand $\mathbb{E}W^t \nabla f(W)$. Define the real-valued, anti-symmetric function

$$F(w', w) := \tfrac{1}{2}(w' - w)^t \Lambda^{-t} (\nabla f(w') + \nabla f(w)) \tag{2.7}$$

for $w, w' \in \mathbb{R}^d$, and note that, because of exchangeability, $\mathbb{E}F(W', W) = 0$; see Stein (1986). Thus

$$\begin{aligned} 0 &= \tfrac{1}{2}\mathbb{E}\{(W'-W)^t\Lambda^{-t}(\nabla f(W') + \nabla f(W))\} \\ &= \mathbb{E}\{(W'-W)^t\Lambda^{-t}\nabla f(W))\} \\ &\quad + \tfrac{1}{2}\mathbb{E}\{(W'-W)^t\Lambda^{-t}(\nabla f(W') - \nabla f(W))\} \\ &= \mathbb{E}\{R^t\Lambda^{-t}\nabla f(W)\} - \mathbb{E}\{W^t\nabla f(W)\} \\ &\quad + \tfrac{1}{2}\mathbb{E}\{(W'-W)^t\Lambda^{-t}(\nabla f(W') - \nabla f(W))\}, \end{aligned} \tag{2.8}$$

where we used (1.7) for the last step. Taylor expansion gives

$$\begin{aligned} &(w'-w)^t\Lambda^{-t}(\nabla f(w') - \nabla f(w)) \\ &\quad = \sum_{m,i,j}(\Lambda^{-1})_{m,i}(w'_i - w_i)(w'_j - w_j)\frac{\partial^2 f(w)}{\partial w_m \partial w_j} \\ &\qquad + \sum_{m,i,j,k}(\Lambda^{-1})_{m,i}(w'_i - w_i)(w'_j - w_j)(w'_k - w_k)\tilde{R}_{mjk}, \end{aligned}$$

where

$$|\tilde{R}_{mjk}| \leqslant \frac{1}{2}\left\| \frac{\partial^3 f}{\partial w_m \partial w_j \partial w_k}\right\|. \tag{2.9}$$

Thus in (2.8),

$$\begin{aligned} &\mathbb{E}\{(W'-W)^t\Lambda^{-t}(\nabla f(W') - \nabla f(W))\} \\ &\quad = \sum_{m,i,j}(\Lambda^{-1})_{m,i}\mathbb{E}(W'_i - W_i)(W'_j - W_j)\frac{\partial^2 f(W)}{\partial W_m \partial W_j} \\ &\qquad + \sum_{m,i,j,k}(\Lambda^{-1})_{m,i}(W'_i - W_i)(W'_j - W_j)(W'_k - W_k)\tilde{R}_{mjk}. \end{aligned} \tag{2.10}$$

Now we turn our attention to $\mathbb{E}\nabla^t\Sigma\nabla f(W)$. Note that, because of (2.1), (1.7) and exchangeability,

$$\begin{aligned} \mathbb{E}(W'-W)(W'-W)^t &= \mathbb{E}\{W(W-W')^t\} + \mathbb{E}\{W(W-W')^t\} \\ &= 2\mathbb{E}\{W(\Lambda W - R)^t\} = 2\Sigma\Lambda^t - 2\mathbb{E}(WR^t) =: T. \end{aligned} \tag{2.11}$$

Hence, with T as in (2.11),

$$\begin{aligned} \nabla^t\Sigma\nabla f(w) &= \tfrac{1}{2}\nabla^t T\Lambda^{-t}\nabla f(w) + \nabla^t\mathbb{E}(WR^t)\Lambda^{-t}\nabla f(w) \\ &= \frac{1}{2}\sum_{m,i,j}(\Lambda^{-1})_{m,i}T_{j,i}\frac{\partial^2 f(w)}{\partial w_m \partial w_j} + \sum_{m,i,j}(\Lambda^{-1})_{m,i}\mathbb{E}(W_j R_i)\frac{\partial^2 f(w)}{\partial w_m \partial w_j}. \end{aligned} \tag{2.12}$$

Combining (2.8), (2.10) and (2.12),

$$\begin{aligned}
&|\mathbb{E}\{\nabla^t\Sigma\nabla f(W) - W^t\nabla f(W)\}| \\
&\leqslant \frac{1}{2}\Big|\sum_{m,i,j}\mathbb{E}\Big\{(\Lambda^{-1})_{m,i}[T_{j,i} - \mathbb{E}^W(W_i' - W_i)(W_j' - W_j)]\frac{\partial^2 f(W)}{\partial w_m \partial w_j}\Big\}\Big| \\
&\quad + \frac{1}{2}\Big|\sum_{m,i,j,k}\mathbb{E}\{(\Lambda^{-1})_{m,i}(W_i' - W_i)(W_j' - W_j)(W_k' - W_k)\tilde{R}_{mjk}\}\Big| \\
&\quad + \Big|\sum_{i,m}(\Lambda^{-1})_{m,i}\mathbb{E}\Big\{R_i\frac{\partial f(W)}{\partial w_m}\Big\}\Big| + \Big|\sum_{m,i,j}(\Lambda^{-1})_{m,i}\mathbb{E}(W_jR_i)\mathbb{E}\Big\{\frac{\partial^2 f(W)}{\partial w_m \partial w_j}\Big\}\Big| \\
&\leqslant \frac{|h|_2}{4}\sum_{i,j}\lambda^{(i)}\mathbb{E}|T_{j,i} - \mathbb{E}^W(W_i' - W_i)(W_j' - W_j)| + \frac{|h|_3}{12}B \\
&\quad + |h|_1\sum_i\lambda^{(i)}\mathbb{E}|R_i| + \frac{|h|_2}{2}\sum_{i,j}\lambda^{(i)}\mathbb{E}|W_jR_i|, \qquad (2.13)
\end{aligned}$$

where we used (2.9) to obtain the second inequality, and Lemma 2.7 to obtain the last inequality. From the Cauchy-Schwarz inequality, $\mathbb{E}|R_j| \leqslant \sqrt{\mathbb{E}R_j^2}$ and

$$\mathbb{E}|W_jR_i| \leqslant \sqrt{\mathbb{E}W_j^2\mathbb{E}R_i^2} \leqslant \|\Sigma\|^{1/2}\sqrt{\mathbb{E}R_i^2}.$$

The C-expression in (2.2) now follows from the last two terms of (2.13). Recalling that $\mathbb{E}(W' - W)(W' - W)^t = T$, this proves the first term of (2.2) from the first term of (2.13). □

Sometimes we may wish to assess the distance to a normal distribution for which the covariance matrix Σ_0, while non-negative definite, does not have full rank. Stein's method helps to derive a straightforward bound in this case also. If Σ has full rank, then the Stein characterization (2.4) of the multivariate normal distribution says that, for all f which are solutions of the Stein equation (2.5) for functions $h : \mathbb{R}^d \to \mathbb{R}$ having 3 bounded derivatives,

$$\mathbb{E}\{\nabla^t\Sigma\nabla f(X) - X^t\nabla f(X)\} = 0.$$

We shall show that this characterisation remains valid if the covariance matrix is not of full rank; thus two mean zero multivariate normal distributions can be compared via their covariance matrices. The proof of the following proposition is straightforward and routine (see Appendix A).

PROPOSITION 2.9. *Let X and Y be $\mathbb{R}^d$-valued normal variables with distributions $X \sim \mathrm{MVN}(0,\Sigma)$ and $Y \sim \mathrm{MVN}(0,\Sigma_0)$, where $\Sigma = (\sigma_{i,j})_{i,j=1,\dots,d}$*

has full rank, and $\Sigma_0 = (\sigma^0_{i,j})_{i,j=1,\dots,d}$ *is non-negative definite. Let* $h : \mathbb{R}^d \to \mathbb{R}$ *have 3 bounded derivatives. Then*

$$\bigl|\mathbb{E}h(X) - \mathbb{E}h(Y)\bigr| \leqslant \frac{1}{2}|h|_2 \sum_{i,j=1}^{d} |\sigma_{i,j} - \sigma^0_{i,j}|.$$

Using the triangle inequality and Theorem 2.1 we thus obtain a bound for a normal approximation even for a normal distribution with degenerate covariance matrix. An important example from random graph statistics will be treated in Section 4.

Remark 2.10. In general, as soon as one element of our random vector can be expressed as a linear combination of some other elements of the vector, we cannot expect the matrix Λ to be unique. If $R = 0$, this situation can only occur when the covariance matrix of W does not have full rank (which is excluded in Theorem 2.1). If the covariance matrix Σ of W has full rank, then from $\Lambda_1 W = \Lambda_2 W$ it follows that $\Lambda_1 W W^t = \Lambda_2 W W^t$, and taking expectations, $\Lambda_1 \Sigma = \Lambda_2 \Sigma$. If Σ is invertible, then necessarily $\Lambda_1 = \Lambda_2$.

3. Non-smooth test functions.

We first assume that $\Sigma = \mathrm{Id}$. Following Rinott and Rotar (1996), let Φ denote the standard normal distribution in $\mathbb{R}^d$, and ϕ the corresponding density function. For $h : \mathbb{R}^d \to R$ set

$$\begin{aligned} h^+_\delta(x) &= \sup\{h(x+y) : |y| \le \delta\}, \\ h^-_\delta(x) &= \inf\{h(x+y) : |y| \le \delta\}, \\ \tilde{h}(x,\delta) &= h^+_\delta(x) - h^-_\delta(x). \end{aligned}$$

Let $\mathcal{H}$ be a class of measureable functions $\mathbb{R}^d \to R$ which are uniformly bounded by 1. Suppose that for any $h \in \mathcal{H}$

1. for any $\delta > 0$, $h^+_\delta(x)$ and $h^-_\delta(x)$ are in $\mathcal{H}$,
2. for any $d \times d$ matrix A and any vector $b \in \mathbb{R}^d$, $h(Ax + b) \in \mathcal{H}$,
3. $$\sup_{h \in \mathcal{H}} \left\{ \int_{\mathbb{R}^d} \tilde{h}(x,\delta)\Phi(dx) \right\} \le a\delta \tag{3.1}$$

   for some constant $a = a(\mathcal{H}, \delta)$. Obviously we may assume $a \ge 1$.

The class of indicators of measureable convex sets is such a class; for this class, $a \le 2\sqrt{d}$, see Bolthausen and Götze (1993).

In the same way as in Rinott and Rotar (1996) we can show the following corollary. The presentation differs from Rinott and Rotar (1996) as we make the relationship to the bounds in Theorem 2.1 immediate. The now fairly standard proof is found in Appendix A. We also note forthcoming work by Bhattacharya and Holmes (2007) for a rigorous expostion.

COROLLARY 3.1. *Let W satisfy the conditions of Theorem 2.1, with* $\Sigma = \mathrm{Id}$. *Then, for all $h \in \mathcal{H}$ with $|h| \leq 1$, there exist constants $\gamma = \gamma(d)$ and $a > 1$ such that, with the notation from Theorem 2.1 and* (3.2),

$$\sup_{h\in\mathcal{H}} |\mathbb{E}h(W) - \mathbb{E}h(Z)| \quad \leq \quad \gamma^2\Big(D\log(T^{-1}) + \frac{1}{2}BT^{-1/2} + C + a\sqrt{T}\Big),$$

with

$$T = \frac{1}{a^2}\Big(D + \sqrt{\frac{aB}{2} + D^2}\Big)^2 \text{ and } D = \frac{A}{2} + Cd. \tag{3.2}$$

The constant γ may be different from the constant γ in Lemma A.1.

If A, B and C are $O(n^{-1/2})$, then we would obtain a bound of order $O(n^{-1/4})$. This is poorer than the $n^{-1/2}\log n$ type of bounds obtained in Rinott and Rotar (1996), but Rinott and Rotar (1996) obtain the improved rate by assuming that the random variables are bounded.

Next we generalise the result to arbitrary Σ. Let W have mean vector 0 and variance-covariance matrix Σ. If Λ and R are such that (1.7) is satisfied for W, then $Y = \Sigma^{-1/2}W$ satisfies (1.7) with $\hat{\Lambda} = \Sigma^{-1/2}\Lambda\Sigma^{1/2}$ and $R' = \Sigma^{-1/2}R$. With

$$\hat{\lambda}^{(i)} \quad = \quad \sum_{m=1}^{d} |(\Sigma^{-1/2}\Lambda^{-1}\Sigma^{1/2})_{m,i}|$$

as well as

$$\begin{aligned} A' &= \sum_{i,j} \hat{\lambda}^{(i)} \sqrt{\operatorname{Var}\mathbb{E}^Y \sum_{k,\ell} \Sigma_{i,k}^{-1/2}\Sigma_{j,\ell}^{-1/2}(W_k' - W_k)(W_\ell' - W_\ell)}, \\ B' &= \sum_{i,j,k} \hat{\lambda}^{(i)}\mathbb{E}\Big|\sum_{r,s,t} \Sigma_{i,r}^{-1/2}\Sigma_{j,s}^{-1/2}\Sigma_{k,t}^{-1/2}(W_r' - W_r)(W_s' - W_s)(W_t' - W_t)\Big| \end{aligned}$$

and

$$C' = \sum_{i=1}^{d} \hat{\lambda}^{(i)} \sqrt{\mathbb{E}\Big(\sum_k \Sigma_{i,k}^{-1/2} R_k\Big)^2}, \tag{3.3}$$

we obtain a similar result as before; again the proof is in Appendix A.

COROLLARY 3.2. *Let W be as in Theorem 2.1. Then, for all $h \in \mathcal{H}$ with $|h| \leq 1$, there exist $\gamma = \gamma(d)$ and $a > 1$ such that, with the notation* (3.3), (3.3), *and* (3.3),

$$\sup_{h\in\mathcal{H}} |\mathbb{E}h(W) - \mathbb{E}h(Z)| \quad \leq \quad \gamma^2\Big(-D'\log(T') + \frac{B'}{2\sqrt{T'}} + C' + a\sqrt{T'}\Big),$$

with

$$T' = \frac{1}{a^2}\bigg(D' + \sqrt{\frac{aB'}{2} + D'^2}\bigg)^2 \qquad \text{and} \qquad D' = \frac{A'}{2} + C'd.$$

REMARK 3.3. We could simplify the above bound further, with a coarser bound. Using Minkowski's inequality we have that

$$\operatorname{Var}\sum_{i=1}^{k} X_i \leqslant k^2 \sup_i \operatorname{Var} X_i,$$

and thus obtain the simple estimate

$$\begin{aligned}
&\operatorname{Var}\mathbb{E}^Y \sum_{k,\ell} \Sigma_{i,k}^{-1/2}\Sigma_{j,\ell}^{-1/2}(W'_k - W_k)(W'_\ell - W_\ell)\\
&\qquad \leqslant d^4\|\Sigma^{-1/2}\|^4 \sup_{k,\ell}\operatorname{Var}\mathbb{E}^W\{(W'_k - W_k)(W'_\ell - W_\ell)\}
\end{aligned}$$

and hence

$$A' \quad \leq \quad d^3\|\Sigma^{-1/2}\|^2 \sum_i \hat\lambda^{(i)} \sup_{k,\ell}\sqrt{\operatorname{Var}\mathbb{E}^W\{(W'_k - W_k)(W'_\ell - W_\ell)\}};$$

in B' and C' we could similarly bound $\Sigma_{i,k}^{-1/2}$ by $\|\Sigma^{-1/2}\|$ to obtain a simpler bound. There are however examples, such as the random graph example in Section 4, where $\|\Sigma^{-1/2}\|$ provides a non-informative bound.

4. The embedding method and applications.

4.1. *General framework.* Assume that an ℓ-dimensional random variable $W_{(\ell)}$ of interest is given. Often, the construction of an exchangeable pair $(W_{(\ell)}, W'_{(\ell)})$ is straightforward. If, say, $W_{(\ell)} = W_{(\ell)}(\mathbb{X})$ is a function of i.i.d. random variables $\mathbb{X} = (X_1, \ldots, X_n)$, one can choose uniformly an index I from 1 to n, replace X_I by an independent copy X'_I, and define $W'_{(\ell)} := W_{(\ell)}(\mathbb{X}')$, where $\mathbb{X}'$ is now the vector $\mathbb{X}$ but with X_I replaced by X'_I.

In general there is no hope that $(W_{(\ell)}, W'_{(\ell)})$ will satisfy Condition (1.2) with R being of the required smaller order or even equal to zero, so that in this case Theorem 2.1 would not yield useful bounds.

Surprisingly often it is possible, though, to extend $W_{(\ell)}$ to a vector $W \in \mathbb{R}^d$ such that we can construct and exchangeable pair (W, W') which satisfies Condition (1.2) with $R = 0$. If we can bound the distance of the distribution $\mathcal{L}(W)$ to a d-dimensional multivariate normal distribution, a bound on the distance of the distribution $\mathcal{L}(W_{(\ell)})$ to an $\ell-$dimentional multivariate normal distribution follows immediately.

To explain the approach, we turn the problem on its head. Suppose that $W \in \mathbb{R}^d$ is such that we can construct and exchangeable pair (W, W') which satisfies Condition (1.2) with $R = 0$. Rename the first ℓ components to comprise $W_{(\ell)}$, so that

$$W = \begin{bmatrix} W_{(\ell)} \\ W^{(d-\ell)} \end{bmatrix},$$

and $W_{(\ell)} = I_{\ell,0} W$, with

$$I_{\ell,0} = (Id_\ell, 0_{\ell\times(d-\ell)}),$$

$0_{\ell\times(d-\ell)}$ denoting the $\ell \times (d-\ell)$-matrix consisting entirely of 0's. Defining $W'_{(\ell)} = I_{\ell,0} W'$, it follows that $(W_{(\ell)}, W'_{(\ell)})$ forms an exchangeable pair. From (1.2),

$$\begin{aligned} \mathbb{E}^W (W_{(\ell)} - W'_{(\ell)}) &= I_{\ell,0} \mathbb{E}^W (W - W') \\ &= -I_{\ell,0} \Lambda W. \end{aligned}$$

Now decompose the matrix Λ as

$$\Lambda = \begin{bmatrix} \Lambda_{1,1} & \Lambda_{1,2} \\ \Lambda_{2,1} & \Lambda_{2,2} \end{bmatrix},$$

where $\Lambda_{1,1}$ denotes an $\ell \times \ell$ submatrix, $\Lambda_{1,2}$ denotes an $\ell \times (d-\ell)$ submatrix, and so on. Then

$$I_{\ell,0} \Lambda W = \Lambda_{1,1} W_{(\ell)} + \Lambda_{1,2} W^{(d-\ell)},$$

and hence

$$\mathbb{E}^W (W_{(\ell)} - W'_{(\ell)}) = -\Lambda_{1,1} W_{(\ell)} - \Lambda_{1,2} W^{(d-\ell)}.$$

Conditioning on $W_{(\ell)}$ gives that

$$\mathbb{E}^{W_{(\ell)}} (W_{(\ell)} - W'_{(\ell)}) = -\Lambda_{1,1} W_{(\ell)} - \Lambda_{1,2} \mathbb{E}^{W_{(\ell)}} W^{(d-\ell)}.$$

Thus Condition (1.2) is satisfied with

$$R = -\Lambda_{1,2}\mathbb{E}^{W_{(\ell)}} W^{(d-\ell)}. \tag{4.1}$$

If $\Lambda_{1,2} = 0$, then no embedding is required. But if $\Lambda_{1,2} \neq 0$, then the remainder R in (1.2) is a nontrivial linear combination of random variables, and these random variables could serve as embedding vector. In order to obtain useful bounds in Theorem 2.1, the embedding dimension d should not be too large. In the examples below it will be obvious how to choose $W^{(d-\ell)}$ to make the construction work.

While the embedding method is reminiscent of Hoeffding projections for U-statistics, Subsection 4.4 clarifies the difference.

4.2. *Runs on the line.* Let $\mathbb{X} = (\xi_1, \dots, \xi_n)$ be a sequence of independent random variables with distribution Bernoulli(p), $0 < p < 1$, that is $\mathbb{P}[\xi_1 = 1] = 1 - \mathbb{P}[\xi_1 = 0] = p$. For $d > 1$, define the (centered) number of d-runs as

$$V_d := \sum_{m=1}^{n} (\xi_m \xi_{m+1} \cdots \xi_{m+d-1} - p^d),$$

where, for convenience, we assume the torus convention that $\xi_{n+1} \equiv \xi_1$, $\xi_{n+2} \equiv \xi_2$ and so on.

As mentioned in the introduction, if we want to use the obvious construction of an exchangeable pair, the univariate version of exchangeable pairs of Rinott and Rotar (1997) (Proposition 1.2) does not yield convergent bounds of V_d to the standard normal distribution if $d > 1$. However, we can tackle the example with our approach by incorporating the auxiliary variables $V_1, \dots, V_{d-1}$, such that the problem becomes linear in a higher-dimensional setting.

We construct an exchangeable pair $(\mathbb{X}, \mathbb{X}')$, where instead of just one, we resample $d-1$ of the ξ_i. To this end, let I be uniformly distributed over $\{1, \dots, n\}$ and let $\tilde{\xi}_1, \dots, \tilde{\xi}_n$ be independent copies of the ξ_i. Let $\mathbb{X}'$ be the same as $\mathbb{X}$ but with the subsequence $\xi_I, \xi_{I+1}, \dots, \xi_{I+d-2}$ of length $d-1$ replaced by $\xi'_I, \xi'_{I+1}, \dots, \xi'_{I+d-2}$. Clearly $(\mathbb{X}, \mathbb{X}')$ forms an exchangeable pair. Define $V'_i := V_i(\mathbb{X}')$; we have

$$\begin{aligned} V'_i - V_i = &- \sum_{m=I-i+1}^{I+d-2} \xi_m \cdots \xi_{m+i-1} \quad + \sum_{m=I+d-i}^{I+d-2} \xi'_m \cdots \xi'_{I-1} \xi_I \cdots \xi_{m+i-1} \\ &+ \sum_{m=I}^{I+d-i-1} \xi'_m \cdots \xi'_{m+i-1} \quad + \sum_{m=I-i+1}^{I-1} \xi_m \cdots \xi_{I-1} \xi'_I \cdots \xi'_{m+i-1}, \end{aligned} \tag{4.2}$$

where sums $\sum_a^b$ are defined to be zero if $a > b$. Now, (4.2) yields

$$\begin{aligned}\mathbb{E}^W(V_i' - V_i) \\ &= -n^{-1}\big[(d+i-2)V_i - 2pV_{i-1} - 2p^2V_{i-2} - \cdots - 2p^{i-1}V_1\big] \\ &= -n^{-1}\Big[(d+i-2)V_i + 2\sum_{k=1}^{i-1} p^{i-k}V_k\Big].\end{aligned} \tag{4.3}$$

From this representation we see that we may take $V_1, \ldots, V_{d-1}$ as the auxiliary random variables.

Straightforward calculations yield that, for all $i \geqslant j$,

$$\begin{aligned}\mathbb{E}(V_iV_j) &= n\big[(i-j+1)p^i + 2\sum_{l=1}^{j-1} p^{i+j-l} - (i+j-1)p^{i+j}\big] \\ &= np^i(1-p)\sum_{k=0}^{j-1}(i-j+1+2k)p^k.\end{aligned} \tag{4.4}$$

In particular

$$\mathbb{E}V_i^2 = np^i(1-p)\sum_{k=0}^{i-1}(1+2k)p^k, \tag{4.5}$$

which lies in the interval between $np^i(1-p)$ and $np^i(1-p)i^2$. Thus we define the W_i to be the weighted versions

$$W_i := \frac{V_i}{\sqrt{np^i(1-p)}}, \tag{4.6}$$

and from (4.4) we have for general i and j

$$\mathbb{E}(W_iW_j) = p^{\frac{|i-j|}{2}} \sum_{k=0}^{i\wedge j-1}(|i-j|+1+2k)p^k =: \sigma_{i,j}. \tag{4.7}$$

From (4.7) it is clear that the corresponding $\Sigma = (\sigma_{i,j})_{i,j}$ is constant for all n and of full rank. For $p \to 0$, Σ converges to uncorrelated coordinates and for $p \to 1$ to a matrix of rank 1. For applications and further references see Glaz et al. (2001) and Balakrishnan and Koutras (2002). Now, from (4.3) we have

$$\mathbb{E}^W(W_i' - W_i) = -n^{-1}\Big[(d+i-2)W_i + 2\sum_{k=1}^{i-1} p^{\frac{i-k}{2}}W_k\Big].$$

Thus (1.7) is satisfied with $R = 0$ and

$$\Lambda = \frac{1}{n}\begin{bmatrix} d-1 & & & & & \\ -2p^{\frac{1}{2}} & d & & & 0 & \\ \vdots & & \ddots & & & \\ -2p^{\frac{k-1}{2}} & \cdots & -2p^{\frac{1}{2}} & d+k-2 & & \\ \vdots & & & & \ddots & \\ -2p^{\frac{d-1}{2}} & & \cdots & & -2p^{\frac{1}{2}} & 2(d-1) \end{bmatrix}.$$

THEOREM 4.1. *With W defined as in* (4.6)*, $n > 2d-1$ and Σ given through* (4.7)*, we have for three times differentiable functions h that*

$$|\mathbb{E}h(W) - \mathbb{E}h(\Sigma^{1/2}Z)| \leqslant \frac{37d^{7/2}|h|_2}{p^d(1-p)\sqrt{n}} + \frac{10d^5|h|_3}{p^{3d/2}(1-p)^{3/2}\sqrt{n}}.$$

PROOF. Some rough estimates yield that for all $1 \leqslant i,j,k \leqslant d$

$$\lambda^{(i)} \leqslant \frac{15n}{d},$$

$$\operatorname{Var}\mathbb{E}^W(W_i' - W_i)(W_j' - W_j) \leqslant \frac{96d^5}{n^3p^{2d}(1-p)^2},$$

$$\mathbb{E}\big|(W_i' - W_i)(W_j' - W_j)(W_k' - W_k)\big| \leqslant \frac{8d^3}{n^{3/2}p^{3d/2}(1-p)^{3/2}}.$$

Now apply Theorem 2.1. Details can be found in Appendix B. □

REMARK 4.2. Although the bound is quite crude with respect to the dimension and hence mainly of theoretical interest, it is explicit. However, for small values of p or large values of d, Poisson approximation is more appropriate, and in these cases the bounds for normal approximation cannot be expected to be good unless n is very large. We also note that V_d exhibits a local dependence structure and thus also Stein's method using the local approach, such as in Rinott and Rotar (1996), could easily be used; and, of course, there is an abundance of results about m-dependent sequences.

REMARK 4.3. In the case of 2-runs, using the notation of (1.8) and the consequent paragraph, it is not difficult to see that, for any choice of λ and defining $R = R(V_2, V_1) := \sigma^{-1}(\lambda V_2 - \frac{2}{n}V_2 + \frac{2p}{n}V_1)$, we have that $\lambda^{-1}\sqrt{\operatorname{Var} R}$ is at least of order 1 as $n \to \infty$, where $\sigma^2 := \operatorname{Var} V_2$. However, Remark 2.4 suggest that it should nevertheless be possible to choose λ such that, with $\tilde{R} = \tilde{R}(V_2) := \mathbb{E}^{V_2}R = \sigma^{-1}(\lambda V_2 - \frac{2}{n}V_2 + \frac{2p}{n}\mathbb{E}^{V_2}V_1)$, we have $\lambda\sqrt{\operatorname{Var}\tilde{R}} =$

o(1), so that a representation (1.2) could indeed be found with R being of the required small order (and this is supported by numerical simulations). But, as $\mathbb{E}^{V_2}V_1$ is hard to calculate, in this situation the application of the multivariate version (1.7) and Theorem 2.1 is straightforward.

4.3. *An example from random graphs.* Let $G(n,p)$ denote a Bernoulli random graph on n vertices, with edge probabilities p; we assume that $n \geq 4$ and that $0 < p < 1$. Let $I_{i,j} = I_{j,i}$ be the Bernoulli(p)-indicator that edge (i,j) is present in the graph; these indicators are independent.

To test whether in a given network there is a significant number of triangles (or, relatedly, a high degree of clustering), a so-called *conditional uniform graph test* is often employed, see for example Holme (2005). In one form, the edges of the graph are randomised, the number of triangles is counted in such randomised graphs, and the observed number of triangles is compared to the numbers arising from such randomizations. When assessing statistical significance it is hence desirable to know the conditional distribution of the number of triangles (or other graph statistics of interest) given the number of edges. As in real networks the number of vertices may be relatively small, a multivariate normal approximation together with a bound on the distance to the normal would be desirable.

Our interest is hence in the joint distribution of the total number of edges, described by

$$T = \frac{1}{2}\sum_{i,j} I_{i,j} = \sum_{i<j} I_{i,j}$$

and the number of triangles,

$$U = \frac{1}{6}\sum_{i,j,k \text{ distinct}} I_{i,j}I_{j,k}I_{j,k} = \sum_{i<j<k} I_{i,j}I_{j,k}I_{j,k}.$$

Here and in what follows, "i,j,k distinct" is short for "$(i,j,k) : i \neq j \neq k \neq i$"; later we shall also use "i,j,k,ℓ distinct", which is the analogous abbreviation for four indices. Note that

$$\mathbb{E}T = \binom{n}{2}p \quad \text{and } \mathbb{E}U = \binom{n}{3}p^3.$$

Construction of an exchangeable pair

As both T and U are functions of the vector $\mathbb{X} = (I_{i,j}, 1 \leq i < j \leq n)$ of independent, identically distributed edge indicators, we build an exchangeable pair by choosing a potential edge (i,j) uniformly at random, and replacing $I_{i,j}$ by an independent copy $I'_{i,j}$. More formally, pick (I,J) according to

$$\mathbb{P}[I=i, J=j] = \frac{1}{\binom{n}{2}}, \quad 1 \leqslant i < j \leqslant n.$$

If $I=i$, $J=j$ we replace $I_{i,j} = I_{j,i}$ by an independent copy $I'_{i,j} = I'_{j,i}$ and put

$$T' \;=\; T - (I_{I,J} - I'_{I,J}),$$

and

$$U' \;=\; U - \sum_{k:k\neq I,J} (I_{I,J} - I'_{I,J}) I_{J,k} I_{I,k}.$$

Following our approach, conditioning yields

$$\begin{aligned}
\mathbb{E}^{T,U}(T'-T) &= \frac{2}{n(n-1)} \sum_{i<j} \mathbb{E}^{T,U}(I'_{i,j} - I_{i,j} | I=i, J=j) \\
&= p - \frac{2}{n(n-1)} T = \frac{2}{n(n-1)} (\mathbb{E}T - T) \\
&= -\frac{1}{\binom{n}{2}} (T - \mathbb{E}T),
\end{aligned}$$

which depends on T only; but

$$\begin{aligned}
&-\mathbb{E}^{T,U}(U'-U) \\
&\quad= \frac{2}{n(n-1)} \sum_{i<j} \mathbb{E}^{T,U} \sum_{k:k\neq i,j} (I_{i,j} I_{j,k} I_{i,k} - I'_{i,j} I_{j,k} I_{i,k}) \\
&\quad= 3\frac{2}{n(n-1)} U - p \frac{2}{n(n-1)} \mathbb{E}^{T,U} \sum_{i<j,\, k\neq i,j} I_{j,k} I_{i,k}
\end{aligned}$$

depends not only on U but also on the number V of 2-stars,

$$V \;:=\; \frac{1}{2} \sum_{i,j,k \text{ distinct}} I_{i,j} I_{j,k}.$$

We note that

$$\mathbb{E}V = 3\binom{n}{3} p^2.$$

Using our random pair (I, J) we put

$$V' = V - \sum_{k:k\neq I,J} (I_{I,J} - I'_{I,J})(I_{J,k} + I_{I,k}).$$

Including V as auxiliary statistic, we put $W = (T - \mathbb{E}T, V - \mathbb{E}V, U - \mathbb{E}U)$, and $W' = (T' - \mathbb{E}T, V' - \mathbb{E}V, U' - \mathbb{E}U)$. Then (W, W') forms an exchangeable pair, and

$$\begin{aligned}
&-\mathbb{E}^W(V' - V)\\
&= \frac{1}{\binom{n}{2}} \sum_{i<j} \mathbb{E}^W \sum_{k:k\neq i,j} (I_{i,j} - I'_{i,j})(I_{j,k} + I_{i,k})\\
&= \frac{1}{\binom{n}{2}} \mathbb{E}^W \sum_{i<j,\, k\neq i,j} (I_{i,j}I_{j,k} + I_{i,j}I_{i,k}) - p\frac{1}{\binom{n}{2}} \mathbb{E}^W \sum_{i<j,\, k\neq i,j} (I_{j,k} + I_{i,k})\\
&= 2\frac{1}{\binom{n}{2}} V - 2p\frac{1}{\binom{n}{2}}(n-2)T\\
&= -2\frac{1}{\binom{n}{2}}(V - \mathbb{E}V) + 2p\frac{(n-2)}{\binom{n}{2}}(T - \mathbb{E}T),
\end{aligned}$$

where the last equality follows from $\mathbb{E}(V' - V) = 0$. Thus (1.7) is satisfied with $R = 0$ and Λ given by

$$\Lambda = \frac{1}{\binom{n}{2}} \begin{pmatrix} 1 & 0 & 0 \\ -2(n-2)p & 2 & 0 \\ 0 & -p & 3 \end{pmatrix}.$$

As the variances, calculated in Appendix B.3, are not all of the same order, we re-scale our variables, similarly to Janson and Nowicki (1991), as follows. Put

$$T_1 = \frac{n-2}{n^2} T, \quad V_1 = \frac{1}{n^2} V, \quad U_1 = \frac{1}{n^2} U.$$

For these re-scaled variables we re-scale W' as for W to obtain T_1', V_1' and U_1', so that (W_1, W_1') is also exchangeable. The covariance matrix Σ_1 for $W_1 = (T_1 - \mathbb{E}T_1, V_1 - \mathbb{E}V_1, U_1 - \mathbb{E}U_1)$ equals

$$\Sigma_1 = 3\frac{(n-2)\binom{n}{3}}{n^4} p(1-p) \times \begin{pmatrix} 1 & 2p & p^2 \\ 2p & 4p^2 + \frac{p(1-p)}{n-2} & 2p^3 + \frac{p^2(1-p)}{n-2} \\ p^2 & 2p^3 + \frac{p^2(1-p)}{n-2} & p^4 + \frac{p^2(1+p-2p^2)}{3(n-2)} \end{pmatrix}, \quad (4.8)$$

and (1.7) is satisfied with $R = 0$ and Λ_1 given by

$$\Lambda_1 = \frac{1}{\binom{n}{2}} \begin{pmatrix} 1 & 0 & 0 \\ -2p & 2 & 0 \\ 0 & -p & 3 \end{pmatrix}.$$

REMARK 4.4. The observation that the 2-stars form a useful auxiliary statistic can also be found in Janson and Nowicki (1991); there it is related to Hoeffding-type projections.

REMARK 4.5. With $n \to \infty$ we obtain as approximating covariance matrix

$$\Sigma_0 \quad = \quad \frac{1}{2}p(1-p) \times \begin{pmatrix} 1 & 2p & p^2 \\ 2p & 4p^2 & 2p^3 \\ p^2 & 2p^3 & p^4 \end{pmatrix}. \tag{4.9}$$

As also observed in Janson and Nowicki (1991), this matrix has rank 1. It is not difficult to see that the maximal diagonal entry of the inverse Σ^{-1} tends to ∞ as $n \to \infty$, so that a uniform bound on the square root of Σ_1^{-1}, as suggested in Remark 3.3, will not be useful.

Our vector of interest is now $W = (T - \mathbb{E}T, V - \mathbb{E}V, U - \mathbb{E}U)$, re-scaled to $W_1 = (T_1 - \mathbb{E}T_1, V_1 - \mathbb{E}V_1, U_1 - \mathbb{E}U_1)$. In Janson and Nowicki (1991), a normal approximation for W_1 is derived, but no bounds on the approximation are given. Using Theorem 2.1 we obtain explicit bounds, as follows.

PROPOSITION 4.6. *Let $W_1 = (T_1 - \mathbb{E}T_1, V_1 - \mathbb{E}V_1, U_1 - \mathbb{E}U_1)$ be the centralised count vector of the number of edges, two-stars and triangles in a Bernoulli(p)-random graph. Let Σ_1 be given as in* (4.8)*. Then, for every three times differentiable function h,*

$$\left|\mathbb{E}h(W) - \mathbb{E}h(\Sigma_1^{1/2}Z)\right| \quad \leqslant \quad \frac{|h|_2}{n}\left(\frac{35}{4} + 9n^{-1}\right) + \frac{8|h|_3}{3n}\left(1 + n^{-1} + n^{-2}\right).$$

Again we do not claim that the constants in the bound are sharp. However, as we have $\binom{n}{2}$ random edges in the model, the order $O(n^{-1})$ of the bound is as expected.

While for simplicity our other bounds are given as expressions which are uniform in p, bounds dependent on p are derived on the way.

PROOF. Here we only give the main bounds; the calculations for the bounds on A and B are in Appendix B.3. The inverse matrix Λ_1^{-1} is easy to calculate; for $\lambda^{(i)} = \sum_{m=1}^{d} |(\Lambda)_1^{-1})_{m,i}|$, for simplicity we use the uniform bound

$$|\lambda^{(i)}| \leq \frac{3}{2}n^2, \quad i = 1, 2, 3.$$

For A in Theorem 2.1 we obtain that

$$A \quad < \quad 35n^{-1} + 36n^{-2},$$

and for B in Theorem 2.1 calculations yield

$$B < \frac{3}{2}n^2 \times 9 \times \frac{64}{27}\left(n^{-3} + n^{-4} + n^{-5}\right) = 32\left(n^{-1} + n^{-2} + n^{-3}\right).$$

Collecting the bounds gives the result. □

Using Proposition 2.9, we also derive a normal approximation for Σ_0 given in (4.9).

COROLLARY 4.7. *Under the assumptions of Proposition 4.6, for every three times differentiable function h,*

$$\begin{aligned}\left|\mathbb{E}h(W) - \mathbb{E}h(\Sigma_0^{1/2}Z)\right| \quad &\leqslant \quad \frac{|h|_2}{2n}\left(44 + 21n^{-1} + 32n^{-2} + 4n^{-3}\right)\\ &\quad + \frac{8|h|_3}{3n}\left(1 + n^{-1} + n^{-2}\right).\end{aligned}$$

PROOF. We employ Proposition 4.6 and Proposition 2.9, with the triangle inequality. A straightforward calculation shows that

$$\left|\frac{3(n-2)\binom{n}{3}}{n^4} - \frac{1}{2}\right| \quad \leq \quad \frac{3}{2}n^{-1} + 2n^{-3}$$

and so

$$\begin{aligned}&\sum_{i,j=1}^{d} |\sigma_{i,j} - \sigma^0_{i,j}|\\ &\leq \quad \left(\frac{3}{2}n^{-1} + 2n^{-3}\right)\left\{1 + 4p + 6p^2 + 4p^3 + p^4\right\}\\ &\qquad + \left(\frac{p(1-p)}{n-2} + 2\frac{p^2(1-p)}{n-2} + \frac{p^2(1-p)(4-p)}{3(n-2)}\right)\left(\frac{3}{2}n^{-1} + 2n^{-3} + 1\right)\\ &< \quad 26n^{-1} + 3n^{-2} + 32n^{-3} + 4n^{-4}.\end{aligned}$$

Here we used the crude bound that $(n-2)^{-1} \leq \frac{3}{2}n^{-1}$. The corollary follows. □

As a consequence of Corollary 4.7, the conditional graph test which fixes the number of edges and then counts the number of triangles would, in the normal regime, yield a degenerate limiting distribution for the number of triangles. As the number of edges is a function of the vertex degrees, the issue also occurs when fixing the vertex degrees while randomising over the edges.

4.4. *Complete non-degenerate U-statistics.* Let $\mathbb{X} = (X_1, \dots, X_n)$ be a sequence of i.i.d. random elements taking values in a space $\mathcal{X}$. Let ψ be a measurable and symmetric function from $\mathcal{X}^d$ to $\mathbb{R}$, and, for each $k = 1, \dots, d$, let

$$\psi_k(x_1, \dots, x_k) := \mathbb{E}\psi(x_1, \dots, x_k, X_{k+1}, \dots, X_d).$$

Assume without loss of generality that $\mathbb{E}\psi(X_1, \dots, X_d) = 0$. For any subset $\alpha \subset \{1, \dots, n\}$ of size k write $\psi_k(\alpha) := \psi_k(X_{i_1}, \dots, X_{i_k})$ where the i_j are the elements of α. Define the statistics

$$U_k := \sum_{|\alpha|=k} \psi_k(\alpha),$$

where $\sum_{E(\alpha)}$ denotes summation over all subsets $\alpha \subset \{1, \dots, n\}$ which satisfy the property E. Then U_d coincides with the usual U-statistics with kernel ψ (note that, in our notation, the normalising constant $\binom{n}{k}^{-1}$ is not included in U_k). Assume that U_d is non-degenerate, that is, $\mathbb{P}[\psi_1(X_1) = 0] < 1$. Put

$$W_k := n^{1/2} \binom{n}{k}^{-1} U_k.$$

It is well known that $\operatorname{Var} W_k \asymp 1$ (see e.g. Lee (1990)). Note also that, as $n \to \infty$, $\Sigma := \mathbb{E}(WW^t)$ will converge to a covariance matrix of rank 1, as we assume non-degeneracy and hence $U_1 = \sum_{i=1}^n \psi_1(X_i)$ will dominate the behaviour of each U_k.

Using an exchangeable pairs coupling, Rinott and Rotar (1997) proved a univariate normal approximation theorem for non-degenerate and degenerate weighted U-statistics with symmetric weight function under fairly mild conditions on the weights. They show that (1.7) is satisfied for the one-dimensional case and a non-trivial remainder term, related to Hoeffding projections of smaller order. However, we will use Theorem 2.1 to obtain a result for the whole vector $(W_1, \dots, W_d)$, where $W_1, \dots, W_{d-1}$ are not the Hoeffding, but related projections and therefore not of smaller order.

Using Stein's method and the approach of decomposable random variables, Raič (2004) proved rates of convergence for vectors of U-statistics where the coordinates are assumed to be uncorrelated (but nevertheless based upon the same sample $X_1, \dots, X_n$). The next theorem can be seen as a complement to Raič's results because, in our case, a normalisation is not appropriate.

Let $X'_1, \dots, X'_n$ be independent copies of $X_1, \dots, X_n$. Define the random variables $\psi'_{j,k}(\alpha)$ analogously to $\psi_k(\alpha)$ but based on the sequence $X_1, \dots, X_{j-1}, X'_j, X_{j+1}, \dots, X_n$. Define the coupling as in Rinott and Rotar (1997), that is, pick uniformly an index J from $\{1, \dots, n\}$ and replace X_J by X'_J, so that $U'_k = \sum_{|\alpha|=k} \psi'_{J,k}(\alpha)$; it is easy to see that (U', U) is exchangeable. Note now that, if $j \notin \alpha$, $\psi'_{j,k}(\alpha) = \psi_k(\alpha)$, and that $\mathbb{E}^{\mathbb{X}}\psi'_{j,k}(\alpha) = \psi_{k-1}(\alpha \setminus \{j\})$ if $j \in \alpha$. Thus

$$\begin{aligned} \mathbb{E}^{\mathbb{X}}(U'_k - U_k) &= \frac{1}{n}\sum_{j=1}^{n} \sum_{\substack{|\alpha|=k,\\ \alpha\ni j}} \mathbb{E}^{\mathbb{X}}\{\psi'_{j,k}(\alpha) - \psi_k(\alpha)\} \\ &= -\frac{k}{n}U_k + \frac{1}{n}\sum_{j=1}^{n} \sum_{\substack{|\alpha|=k,\\ \alpha\ni j}} \psi_{k-1}(\alpha \setminus \{j\}) \\ &= -\frac{k}{n}U_k + \frac{n-k+1}{n} \sum_{|\beta|=k-1} \psi_{k-1}(\beta) \\ &= -\frac{k}{n}U_k + \frac{n-k+1}{n} U_{k-1}. \end{aligned} \tag{4.10}$$

The second-to-last equality follows from the observation that

$$\sum_{\substack{|\alpha|=k,\\ \alpha\ni j}} \psi_{k-1}(\alpha \setminus \{j\}) = \sum_{\substack{|\beta|=k-1,\\ \beta\not\ni j}} \psi_{k-1}(\beta),$$

so that every set β of size $k-1$ appears exactly $n-(k-1)$ times in the corresponding double sum of (4.10). Thus

$$\mathbb{E}^{\mathbb{X}}(W'_k - W_k) = -\frac{k}{n}(W_k - W_{k-1}).$$

Hence, (1.7) is satisfied for $R = 0$ and

$$\Lambda = \frac{1}{n}\begin{bmatrix} 1 & & & & & \\ -2 & 2 & & & 0 & \\ & -3 & 3 & & & \\ & 0 & & \ddots & \ddots & \\ & & & & -d & d \end{bmatrix}.$$

Applying Theorem 2.1 yields the following result.

THEOREM 4.8. *Assume that* $\rho := \mathbb{E}\psi(X_1,\dots,X_d)^4 < \infty$. *With the above notation, we have for every three times differentiable function* h

$$|\mathbb{E}h(W) - \mathbb{E}h(\Sigma^{1/2}Z)| \leqslant \Big(4\rho^{1/2}d^6|h|_2 + \rho^{3/4}d^7|h|_3\Big)\, n^{-\frac{1}{2}}.$$

PROOF. Some rough estimates yield that for all $1 \leqslant i,j,k \leqslant d$

$$\begin{gathered}\lambda^{(i)} \leqslant dn,\\ \operatorname{Var}\mathbb{E}^W(W_i' - W_i)(W_j' - W_j) \leqslant 256\rho d^6 n^{-3},\\ \mathbb{E}\big|(W_i' - W_i)(W_j' - W_j)(W_k' - W_k)\big| \leqslant 8\rho^{3/4}d^3 n^{-3/2}.\end{gathered}$$

Apply now Theorem 2.1. □

REMARK 4.9. Note that Rinott and Rotar (1997) implicitely use the representation

$$\mathbb{E}^W(W_d' - W_d) = -\frac{1}{n}W_d + \frac{1}{n}(dW_{d-1} - (d-1)W_d) =: -\frac{1}{n}W_d + R; \quad (4.11)$$

compare this with their representation (3.3) of the remainder R, for which they show that it is of the required lower order. We can also see this using Hoeffding projections. Denote by $H^{(j)}$ the jth Hoeffding projection of $\binom{n}{k}^{-1}U_k$ (for a definition we refer to Lee (1990)) and recall that the random variables of the sequence $H^{(1)},\dots,H^{(d)}$ are uncorrelated and have strictly decreasing variances of order $n^{-1}, n^{-2},\dots,n^{-d}$ (these are the exact orders, as we assume non-degeneracy). From Theorem 1 of Lee (1990) we have the representation

$$W_k = n^{1/2}\sum_{j=1}^{k}\binom{k}{j}H^{(j)}$$

for each k, based on the same projections $H^{(j)}$ as the conditional expectations ψ_j are the same for all W_k. From this it follows that the random variable $H^{(1)}$ with the largest variance disappears in the remainder R of (4.11). Hence, $\lambda^{-1}\sqrt{\operatorname{Var}R} = \mathrm{O}(n^{-1/2})$.

4.5. *Double-indexed permutation statistics.* Let $a_{i,j,k,l}$, $1 \leqslant i,j,k,l \leqslant n$, be real numbers such that $a_{i,j,k,l} = 0$ whenever $i = j$ but $k \neq j$. Assume that

$$\sum_{i,j,k,l} a_{i,j,k,l} = 0 \quad (4.12)$$

and define

$$V_0 = V_0(\pi) = \sum_{s,t=1}^{n} a_{s,t,\pi(s),\pi(t)},$$

where π is a uniformly drawn random permutations of size n. The asymptotic normality of V_0 was proved by Zhao et al. (1997), generalising the proof of Bolthausen (1984), which is related to the exchangeable pair coupling. Barbour and Chen (2005) used the exchangeable pair coupling to find a non-trivial representation of V_0 of the form (1.2) with a non-zero remainder term R; see their article also for a historical overview.

We will discuss here only the applicability of this example to Theorem 2.1 to illustrate the embedding method, which contrasts with Barbour and Chen (2005) in the sense that, with our approach, again one does not need to find a one-dimensional representation of the form (1.2) but can use directly the multidimensional version (1.7) in a straightforward manner. We also do not bound the error terms because the corresponding calculations are too involved for the purpose of this paper.

Construct now an exchangeable pair as follows. Let I and J be distributed uniformly over $1,\dots,n$ conditioned that $I \neq J$. Define the permutation $\pi' = (\pi(I)\pi(J)) \circ \pi$ so that π' is the permutation where $\pi'(k) = \pi(k)$ for all $k \neq I, J$, and where $\pi'(I) = \pi(J)$ and $\pi'(J) = \pi(I)$. Let now for the sake of a simpler notation $a^{\pi}_{i,j,k,l} := a_{i,j,\pi(k),\pi(l)}$. Defining $W' = W(\pi')$ we have

$$\begin{aligned} V_0' - V_0 = & -\sum_{s=1}^{n} (a^{\pi}_{I,s,I,s} + a^{\pi}_{J,s,J,s} + a^{\pi}_{s,I,s,I} + a^{\pi}_{s,J,s,J}) \\ & + (a^{\pi}_{I,I,I,I} + a^{\pi}_{I,J,I,J} + a^{\pi}_{J,I,J,I} + a^{\pi}_{J,J,J,J}) \\ & + \sum_{s=1}^{n} \Big(a^{\pi}_{I,s,J,s} + a^{\pi}_{J,s,I,s} + a^{\pi}_{s,I,s,J} + a^{\pi}_{s,J,s,I}\Big) \\ & - (a^{\pi}_{I,I,J,J} + a^{\pi}_{I,J,J,I} + a^{\pi}_{J,I,I,J} + a^{\pi}_{J,J,I,I}) \end{aligned}$$

Hence,

$$\begin{aligned} & \mathbb{E}^{\pi}(V_0' - V_0) \\ & \quad = -\frac{1}{n(n-1)} \sum_{i\neq j} \sum_{s=1}^{n} (a^{\pi}_{i,s,i,s} + a^{\pi}_{j,s,j,s} + a^{\pi}_{s,i,s,i} + a^{\pi}_{s,j,s,j}) \\ & \quad\quad + \frac{1}{n(n-1)} \sum_{i\neq j} (a^{\pi}_{i,i,i,i} + a^{\pi}_{i,j,i,j} + a^{\pi}_{j,i,j,i} + a^{\pi}_{j,j,j,j}) \\ & \quad\quad + \frac{1}{n(n-1)} \sum_{i\neq j} \sum_{s=1}^{n} (a^{\pi}_{i,s,j,s} + a^{\pi}_{j,s,i,s} + a^{\pi}_{s,i,s,j} + a^{\pi}_{s,j,s,i}) \\ & \quad\quad - \frac{1}{n(n-1)} \sum_{i\neq j} (a^{\pi}_{i,i,j,j} + a^{\pi}_{i,j,j,i} + a^{\pi}_{j,i,i,j} + a^{\pi}_{j,j,i,i}) \end{aligned}$$

$$
\begin{aligned}
&= -\frac{4}{n}V_0 + \frac{2}{n(n-1)}\sum_{s=1}^{n}\sum_{i\neq j}(a^\pi_{i,s,j,s} + a^\pi_{s,i,s,j}) \\
&\quad + \frac{2}{n(n-1)}\sum_{i\neq j}(a^\pi_{i,i,i,i} + a^\pi_{i,j,i,j}) - \frac{2}{n(n-1)}\sum_{i\neq j}(a^\pi_{i,i,j,j} + a^\pi_{i,j,j,i}) \\
&= \lambda\Big(-\frac{2n-1}{n}V_0 + V_1 + V_2\Big) + R_1 + R_2
\end{aligned}
$$

with $\lambda := 2/(n-1)$ and where

$$
R_1 := \lambda\sum_{i=1}^{n} a^\pi_{i,i,i,i} - \frac{\lambda}{n}\sum_{i,j=1}^{n} a_{i,i,j,j}, \qquad R_2 := -\frac{\lambda}{n}\sum_{i,j=1}^{n} a^\pi_{i,j,j,i},
$$

$$
V_i := \sum_{s=1}^{n} a^{(i)}_{s,\pi(s)} \qquad \text{for } i = 1,2, \text{ where}
$$

$$
a^{(1)}_{s,t} := \frac{1}{n}\sum_{i,j} a_{s,i,t,j}, \qquad a^{(2)}_{s,t} := \frac{1}{n}\sum_{i,j} a_{i,s,j,t}.
$$

Now, for $i = 1,2$,

$$
V_i' - V_i = -a^{(i)}_{I,\pi(I)} - a^{(i)}_{J,\pi(J)} + a^{(i)}_{I,\pi(J)} + a^{(i)}_{J,\pi(I)}
$$

and thus

$$
\begin{aligned}
\mathbb{E}^\pi(V_i' - V_i) &= -\frac{2}{n}V_i + \frac{2}{n(n-1)}\sum_{i\neq j} a^{(i)}_{i,\pi(j)} \\
&= -\lambda V_i + \frac{2}{n(n-1)}\sum_{i,j} a^{(i)}_{i,\pi(j)} \\
&= -\lambda V_i,
\end{aligned}
$$

where the last equality follows from (4.12). Thus, (1.7) holds for $W = (V_0, V_1, V_2)^t$ with

$$
\Lambda = \lambda\begin{pmatrix} \frac{2n-1}{n} & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
$$

and $R = (R_1 + R_2, 0, 0)^t$.

In the special case where $a_{ijkl} = b_{ij}c_{kl}$ with $b_{ii} = c_{ii} = 0$ for all i,j,k,l and where (b_{ij}) or (c_{kl}) is symmetric up to a (possibly negative) constant, we have $R_1 = 0$ and $R_2 = \beta\lambda n^{-1}V_0$ for some number β, so that (1.7) holds with a $R = 0$ and a slightly different Λ, which would simplifiy the estimates.

Note that these assumptions hold for example if either (b_{ij}) or (c_{ij}) is the adjacency matrix of an undirected graph containing no self-loops.

Mann-Whitney-Wilcoxon statistic. Let $x_1, \dots, x_{n_x}$ and $y_1, \dots, y_{n_y}$, $n_x + n_y = n$, be independent random samples from unknown distributions F_X and F_Y, respectively. The MWW-statistic is then defined to be the number of pairs (x_i, y_j) such that $x_i < y_j$. Let $\pi(i)$ be the rank of z_i, where $z = (x_1, \dots, x_{n_x}, y_1, \dots, y_{n_y})$ is the combined sample. To test the hypothesis $H_0 : F_X = F_Y$, we may assume that π has uniform distribution. It is easy to see that, defining

$$a_{i,j,k,l} = \begin{cases} +\frac{1}{2} & \text{if } 1 \leqslant i \leqslant n_x,\ n_x + 1 \leqslant j \leqslant n \text{ and } 1 \leqslant k < l \leqslant n, \\ -\frac{1}{2} & \text{if } 1 \leqslant i \leqslant n_x,\ n_x + 1 \leqslant j \leqslant n \text{ and } 1 \leqslant l < k \leqslant n, \\ 0 & \text{else,} \end{cases}$$

V_0 is equivalent to the MWW-statistic (up to a shift). It is well known that $\operatorname{Var} V_0 = n_x n_y (n+1)/12$ (see Mann and Whitney (1947)), so that if, for some $0 < \alpha < 1$, $n_x \asymp \alpha n$ and $n_y \asymp (1-\alpha)n$, respectively, we have $\operatorname{Var} V_0 \asymp n^3$.

Note now that, as $a_{i,i,k,l} = 0$ for all i, k, l and as $\sum_{i,j} a_{i,j,\pi(j),\pi(i)} = -\sum_{i,j} a_{i,j,\pi(i),\pi(j)}$, we have $R_1 = 0$ and $R_2 = -\frac{\lambda}{n} V_0$. Hence, the remainder term C in Theorem 2.1 has the required lower order.

Further, we calculate that $a^{(1)}_{i,j} = \frac{n_y(n-2j+1)}{2n}$ if $1 \leqslant i \leqslant n_x$ and $a^{(1)}_{i,j} = 0$ otherwise, and therefore, using the variance formula for the usual singly indexed permutation statistics (see Hoeffding (1951)),

$$\operatorname{Var} V_1 = \frac{1}{n-1} \sum_{i,j=1}^{n} (a^{(1)}_{i,j} - a^{(1)}_{i,\cdot} - a^{(1)}_{\cdot,j} + a^{(1)}_{\cdot,\cdot})^2 \asymp n^3.$$

The same asymptotic is true for V_2, so that indeed $W = n^{-3/2}(V_0, V_1, V_2)$ with the above coupling and choice of Λ is a good candidate for Theorem 2.1.

5. Some comments on the exchangeability condition. Exchangeability is used twice in the proof of Theorem 2.1, namely in (2.8) and (2.11) In this section we not only discuss the necessity of this condition if one uses the Stein operator of the form in Eq. (2.5), but we also suggest a possible way to avoid exchangeability.

5.1. *Exchangeability and anti-symmetric functions.* In (2.8), we use exchangeability in the spirit of Stein (1986). It has been proved by Röllin

(2008) that in the one-dimensional setting the exchangeability condition can be omitted for normal approximation by replacing the usual anti-symmetric function (2.7) with $F(w, w') = g(w') - g(w)$, where now only equality in distribution is needed to obtain an identity similar to (2.8). Also Chatterjee and Meckes (2007) proved their results with this new function F but under the stronger condition (1.4). However, there seems to be no obvious way to apply the above approach under the more general assumption (1.7) (even with $R = 0$) to remove the exchangeability condition. To see this note that, by multivariate Taylor expansion,

$$g(w') = g(w) + (w' - w)^t \nabla g(w) + \tfrac{1}{2} \nabla^t (w' - w)(w' - w)^t \nabla g(w) + r(w', w), \tag{5.1}$$

where r is the corresponding remainder term of the expansion. Thus (5.1) and (1.7) yield the identity

$$\begin{aligned} 0 &= \mathbb{E} g(W') - \mathbb{E} g(W) \\ &= -\mathbb{E}\{W^t \Lambda^t \nabla g(W)\} + \tfrac{1}{2} \mathbb{E}\{\nabla^t (W' - W)(W' - W)^t \nabla g(W)\} \\ &\quad + \mathbb{E} r(W', W), \end{aligned} \tag{5.2}$$

for any suitable function g. To optimally match (5.2) and the left hand side of (2.5) it is clear that we have to choose g such that the system of partial differential equations

$$\Lambda^t \nabla g = \nabla f \tag{5.3}$$

is satisfied. In the one-dimensional setting of Röllin (2008) and the multivariate setting $\Lambda = \lambda I$ of Chatterjee and Meckes (2007), (5.3) can be solved by setting $g = \lambda^{-1} f$. Indeed (5.3) cannot be solved in general, but (5.3) has a twice continuously partially differentiable solution g for a sufficiently large class of functions f only if $\Lambda = \lambda I$.

5.2. *Exchangeability, the covariance matrix and the* Λ *matrix.* In (2.11), using only equality in distribution instead of exchangeability, we would obtain

$$\mathbb{E}(W' - W)(W' - W)^t = \Lambda \Sigma + \Sigma \Lambda^t. \tag{5.4}$$

It is clear from (2.13) that the canonical choice for the variance structure of the approximating multivariate normal distribution would then be

$$\tfrac{1}{2} \mathbb{E}(W' - W)(W' - W)^t \Lambda^{-t} = \tfrac{1}{2}(\Lambda \Sigma \Lambda^{-t} + \Sigma) =: \tilde{\Sigma}, \tag{5.5}$$

which in the exchangeable setting reduces to Σ, see (2.11). Without exchangeability, however, there seems to be no hope that $\tilde{\Sigma}$ would be symmetric and positive-definite as needed unless further assumptions are made.

LEMMA 5.1. *$\tilde\Sigma = \Sigma$ if and only if $\hat\Lambda := \Sigma^{-1/2}\Lambda\Sigma^{1/2}$ is symmetric.*

PROOF. Note that

$$\begin{aligned}\Lambda\Sigma\Lambda^{-t} &= \Sigma^{1/2}\Sigma^{-1/2}\Lambda\Sigma^{1/2}\Sigma^{1/2}\Lambda^{-t}\Sigma^{-1/2}\Sigma^{1/2} \\ &= \Sigma^{1/2}\hat\Lambda\hat\Lambda^{-t}\Sigma^{1/2}.\end{aligned} \tag{5.6}$$

So, if $\hat\Lambda$ is symmetric then clearly $\tilde\Sigma = \Sigma$. If, on the other hand, $\tilde\Sigma = \Sigma$, then (5.5) and (5.6) imply that $\hat\Lambda\hat\Lambda^{-t} = \mathrm{Id}$. By the uniqueness of the inverse, symmetry of $\hat\Lambda^{-1}$ and hence of $\hat\Lambda$ follows. □

LEMMA 5.2. *If (W, W') is exchangeable then $\hat\Lambda$ is symmetric.*

PROOF. If (W', W) is exchangeable, we have from (2.11) that $\tilde\Sigma = \Sigma$ and hence, by Lemma 5.1, the claim follows. □

5.3. *An approach without exchangeability.* Assume that we have given a pair (W', W) such that $\mathscr{L}(W') = \mathscr{L}(W)$ (not necessarily exchangeable), $\mathbb{E}W = 0$, $\mathbb{E}WW^t = \Sigma$ and such that (1.7) is satisfied for some Λ and small R. According to the Markov process interpretation of Stein's method as introduced by Barbour (1990) and Götze (1991), for assessing the distance between the distribution of W and a multivariate normal distribution $\mathrm{MVN}_d(0, \Sigma)$, we evaluate $\mathbb{E}\mathcal{A}f(W)$, where $\mathcal{A}$ is the generator of a stationary Markov (usually Ornstein-Uhlenbeck) process with stationary distribution $\mathrm{MVN}_d(0, \Sigma)$, and f is the solution of the Stein equation $\mathcal{A}f(x) = h(x) - \mathbb{E}h(Z)$.

It is crucial that the dynamics of the Markov process are similar to the dynamics of the Markov process $(W_t)_{t\geqslant 0}$, defined through the coupling (W', W), namely the continuous time Markov jump process with generator

$$\mathcal{B}f(w) = \mathbb{E}\{f(W')|W = w\} - f(w);$$

see Röllin (2008).

This suggests to take a diffusion X_t which is the solution to the SDE

$$X_t = -\Lambda X_t dt + \sigma dB_t,$$

with initial point X_0, where B_t is a standard d-dimensional Browinan motion. From general theory (see e.g. Karatzas and Shreve (1988), Section 5.6) we have that such a process exists and, if $\mathscr{L}(X_0)$ is Gaussian, then the whole process X_t is Gaussian. If furthermore all of the eigenvalues of Λ have positive real parts then X_t has stationary distribution $\mathrm{MVN}_d(0,\Sigma)$, where Σ is the existing and unique solution to the equation

$$\Lambda\Sigma + \Sigma\Lambda^t = \sigma\sigma^t;$$

compare this with (5.4). Using the infinitesimal generator of this process, we obtain the Stein operator

$$\begin{aligned} \mathcal{A}f(x) &= \tfrac{1}{2}\nabla^t\sigma\sigma^t\nabla f(x) - (\Lambda x)^t\nabla f(x). \\ &= \tfrac{1}{2}\nabla^t(\Lambda\Sigma + \Sigma\Lambda^t)\nabla f(x) - x^t\Lambda^t\nabla f(x). \end{aligned} \tag{5.7}$$

To the best of our knowledge, these operators are new as Stein operators and not comparable to Barbour (1990) because of the non-trivial drift. So, it is straightforward to see that, using (1.7) and, assuming again for simplicity that $R = 0$, Eq. (5.7) would lead to an approximation of the form

$$\begin{aligned} &\bigl|\mathbb{E}h(W) - \mathbb{E}h(\Sigma^{1/2}Z)\bigr| \\ &\quad\leqslant \sum_{i,j}\sqrt{\mathrm{Var}\,\mathbb{E}^W(W_i'-W_i)(W_j'-W_j)}\left\|\frac{\partial^2 f}{\partial w_i\partial w_j}\right\| \\ &\qquad + \sum_{i,j,k}\mathbb{E}\bigl|(W_i'-W_i)(W_j'-W_j)(W_k'-W_k)\bigr|\left\|\frac{\partial^3 f}{\partial w_i\partial w_j\partial w_k}\right\|, \end{aligned}$$

without using exchangeability. Note that the factors corresponding to $\lambda^{(i)}$ in Theorem 2.1 would now appear in the bounds on the derivatives of f, which is the solution to the Stein equation

$$\tfrac{1}{2}\nabla^t(\Lambda\Sigma + \Sigma\Lambda^t)\nabla f(x) - x^t\Lambda^t\nabla f(x) = h(x) - \mathbb{E}h(\Sigma^{1/2}Z),$$

if such a solution exists.

Assume now in addition that $\hat\Lambda = \Sigma^{-1/2}\Lambda\Sigma^{1/2}$ is symmetric. In this case, (5.7) simplifies to

$$\mathcal{A}f(x) = \tfrac{1}{2}\nabla^t\Sigma\Lambda^t\nabla f(x) - x^t\Lambda^t\nabla f(x). \tag{5.8}$$

The construction of this process is not difficult. Decompose $\hat\Lambda = UDU^t$, where U is orthogonal and D diagonal. Let Y_t be a d-dimensional Ornstein-Uhlenbeck diffusion, where the coordinates are independent and such that coordinate i has drift $-d_iy_i$ and diffusion rate $\sqrt{2d_i}$. Then, $X_t = \Sigma^{1/2}UY_t$ is the diffusion to the generator (5.8) with the desired stationary distribution $\mathrm{MVN}_d(0,\Sigma)$. However, note that, if Z_t is a standard Ornstein-Uhlenbeck process with local drift $-\operatorname{Id}$ and diffusion rate $\sqrt{2}$ in each of the coordinates, it is not possible to obtain Y_t (and hence X_t) as a transformation of the form AZ_t, because for any matrix A we have

$$\mathbb{E}^{AZ_t}(AZ_{t+\varepsilon} + AZ_t) = A\mathbb{E}^{AZ_t}\mathbb{E}^{Z_t}(Z_{t+\varepsilon} + Z_t) = -\varepsilon AZ_t + \mathrm{o}(\varepsilon),$$

which is again a process with drift $-\operatorname{Id}$.

Note that the processes $X_t = \Sigma^{1/2}UY_t$ are time-reversible, whereas for non-symmetric $\hat{\Lambda}$ they will in general not be, as the process will then "rotate" around the origin in specific directions, from which the time direction can be deduced.

If however (W', W) is exchangeable, then $\hat{\Lambda}$ is symmetric by Lemma 5.2. Although (5.8) would be the canonical Stein operator in this case, the approach through (2.7) and (2.8) allows us to compare the dynamics of W_t directly with that of the process $\Sigma^{1/2}Z_t$ by exploiting the exchangeability, instead of using the more complicated Stein operator (5.8) of the process X_t.

References.

N. Balakrishnan and M. V. Koutras (2002). *Runs and scans with applications*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], New York.

A. D. Barbour (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84**, 297–322.

A. D. Barbour and L. H. Y. Chen (2005). The permutation distribution of matrix correlation statistics. In *Stein's method and applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 223–245. Singapore University Press.

R. N. Bhattacharya and S. Holmes (2007). An exposition of Götze's paper. *Preprint*.

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

E. Bolthausen and F. Götze (1993). The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21**, 1692–1710.

S. Chatterjee and J. Fulman (2006). Exponential approximation by exchangeable pairs and spectral graph theory. *Preprint*. Available at `www.arxiv.org/math.PR/0605552`.

S. Chatterjee and E. Meckes (2007). Multivariate normal approximation using exchangeable pairs. *Preprint*. Available at `www.arxiv.org/abs/math.PR/0701464`.

S. Chatterjee, P. Diaconis, and E. Meckes (2005). Exchangeable pairs and Poisson approximation. *Probab. Surv.* **2**, 64–106. Available at `www.arxiv.org/abs/math.PR/0411525`.

J. Glaz, J. Naus, and S. Wallenstein (2001). *Scan statistics*. Springer Series in Statistics. Springer-Verlag, New York. ISBN 0-387-98819-X.

L. Goldstein and G. Reinert (2005). Zero biasing in one and higher dimensions, and applications. In *Stein's method and applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 1–18. Singapore University Press.

L. Goldstein and Y. Rinott (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33**, 1–17.

F. Götze (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**, 724–739.

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

P. Holme (2005). Network reachabilty of real-world contact sequences. *Phys. Rev. E* **71**, 046119.

S. Janson and K. Nowicki (1991). The asymptotic distributions of generalised U-statistics with applications to random graphs. *Probab. Theory Related Fields* **90**, 341–375.

I. Karatzas and S. E. Shreve (1988). Brownian motion and stochastic calculus. Graduate Texts in Mathematics, Springer-Verlag, New York.

A. J. Lee (1990). *U-statistics*, volume 110 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York. Theory and practice.

F. Lemeire (1975). Bounds for condition numbers of triangular and trapezoid matrices. *Nordisk Tidskr. Informationsbehandling (BIT)* **15**, 58–64.

W. L. Loh (2007). A multivariate central limit theorm for randomised orthogonal array sampling designs in computer experiments. *Manuscript*.

H. B. Mann and D. R. Whitney (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statistics* **18**, 50–60.

K. V. Mardia, J. T. Kent, and J. M. Bibby (1979). *Multivariate analysis*. Academic Press [Harcourt Brace Jovanovich Publishers], London. ISBN 0-12-471250-9. Probability and Mathematical Statistics: A Series of Monographs and Textbooks.

M. Raič (2004). A multivariate CLT for decomposable random vectors with finite second moments. *J. Theoret. Probab.* **17**, 573–603.

G. Reinert (2005). Three general approaches to Stein's method. In *An introduction to Stein's method*, volume 4 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 183–221. Singapore Univ. Press, Singapore.

Y. Rinott and V. Rotar (1996). A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics. *J. Multivariate Anal.* **56**, 333–350.

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**, 1080–1105.

A. Röllin (2007). Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Prob.* **17**, 1596–1614.

A. Röllin (2008). A note on the exchangeability condition in Stein's method. *Statistics and Probability Letters*, in press.

W. I. Smirnow (1986). *Lehrgang der höheren Mathematik. Teil II.* Hochschulbücher für Mathematik [University Books for Mathematics], 2. VEB Deutscher Verlag der Wissenschaften, Berlin, sixteenth edition. Translated from Russian by Klaus Krienes.

C. Stein (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602, Berkeley, Calif. Univ. California Press.

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

L. Zhao, Z. Bai, C.-C. Chao, and W.-Q. Liang (1997). Error bound in a central limit theorem of double-indexed permutation statistics. *Ann. Statist.* **25**, 2210–2227.

APPENDIX A: PROOFS OF THE LEMMAS AND COROLLARIES

A.1. Proof of Lemma 2.7. Let $Z_s := we^{-s} + \sqrt{1-e^{-2s}}\Sigma^{1/2}Z$, so that $Z_0 = w$ and $Z_\infty = \Sigma^{1/2}Z$. Define the function

$$f(w) = -\int_0^\infty [\mathbb{E}h(Z_s) - \mathbb{E}h(\Sigma^{1/2}Z)]ds$$

for every $w \in \mathbb{R}^d$. Straightforward Taylor expansion of $\mathbb{E}h(Z_s) - \mathbb{E}h(\Sigma^{1/2}Z)$ shows that, for each fixed w, f is well-defined. To show that f is a solution to (2.5), observe that

$$\begin{aligned} h(w) - \mathbb{E}h(\Sigma^{1/2}Z) &= \int_0^\infty \frac{d}{ds}\mathbb{E}h(Z_s)ds = \int_0^\infty \mathbb{E}\frac{d}{ds}h(Z_s)ds \\ &= -\int_0^\infty e^{-s}w^t\mathbb{E}\nabla h(Z_s)ds + \int_0^\infty \frac{e^{-2s}}{\sqrt{1-e^{-2s}}}\mathbb{E}\{(\Sigma^{1/2}Z)^t\nabla h(Z_s)\}ds. \end{aligned}$$

The above interchanging of expectation and differentiation is permissible due to dominated convergence, as $|\nabla h(Z_s)| \le |h_1|$ and $|\{(\Sigma^{1/2}Z)^t\nabla h(Z_s)\}| \le |h_1||\Sigma^{1/2}Z|$ and $\mathbb{E}|\Sigma^{1/2}Z| < \infty$. Noting that

$$w^t\nabla f(w) = \int_0^\infty e^{-s}w^t\mathbb{E}\nabla h(Z_s)ds$$

and

$$\begin{aligned} \nabla^t\Sigma\nabla f(w) &= -\int_0^\infty e^{-2s}\mathbb{E}\{\nabla^t\Sigma\nabla h(Z_s)\}ds \\ &= -\int_0^\infty \frac{e^{-2s}}{\sqrt{1-e^{-2s}}}\mathbb{E}\{(\Sigma^{1/2}Z)^t\nabla h(Z_s)\}ds, \end{aligned}$$

from (2.4), Eq. (2.5) follows. Now we note that

$$\sum_i \frac{\partial}{\partial w_i}h(Z_s) = e^{-s}\sum_i h_i(Z_s) = e^{-s}Dh(Z_s),$$

and similarly for higher total derivatives. If D^kh is bounded, then, by dominated convergence,

$$D^kf(w) = -\int_0^\infty e^{-ks}\mathbb{E}D^kh(Z_s)ds.$$

Taking absolute values and evaluating the integral $\int_0^\infty e^{-ks}ds$ yields (2.6). □

A.2. Proof of Lemma 2.9. We shall show below that (2.4) remains valid if the covariance matrix is not of full rank. Then we have, for $h : \mathbb{R}^d \to \mathbb{R}$ with 3 bounded derivatives and f the solution of the Stein equation (2.5) with Σ,

$$\begin{aligned}
&\big|\mathbb{E}h(X) - \mathbb{E}h(Y)\big| \\
&\qquad = \big|\mathbb{E}\nabla^t\Sigma\nabla f(Y) - \mathbb{E}Y^t\nabla f(Y)\big| \\
&\qquad = \big|\mathbb{E}\nabla^t\Sigma\nabla f(Y) - \mathbb{E}Y^t\nabla f(Y) - (\mathbb{E}\nabla^t\Sigma_0\nabla f(Y) - \mathbb{E}Y^t\nabla f(Y))\big| \\
&\qquad = \big|\mathbb{E}\nabla^t(\Sigma - \Sigma_0)\nabla f(Y)\big| \\
&\qquad \le \sum_{i,j=1}^d |\sigma_{i,j} - \sigma_{i,j}^0|\big|f_{i,j}(Y)\big| \le \frac{1}{2}|h|_2 \sum_{i,j=1}^d |\sigma_{i,j} - \sigma_{i,j}^0|,
\end{aligned}$$

where we used the bound (2.6) for the last step.

To prove the assertion, all that remains to show is that (2.4) remains valid if the covariance matrix Σ is not of full rank. Assume that the rank of Σ is k. Let $\lambda_1, \dots, \lambda_k$ denote the non-zero eigenvalues of Σ. Let $Z \in \mathbb{R}^k$ have $\mathrm{MVN}(0, \Lambda_1)$-distribution, where Λ_1 is the diagonal matrix with entries $\lambda_1, \dots, \lambda_k$ on the diagonal; in particular, the components $Z_1, \dots, Z_k$ are independent. Then there exists a $(d \times k)$-matrix $B = (b_{i,j})_{i=1,\dots,d, j=1,\dots,k}$ such that $B'B = \mathrm{Id}_k$, $\Sigma = B\Lambda_1 B'$, and $Y = BZ$, see for example Theorem 2.5.6 in Mardia et al. (1979). Thus we may employ the one-dimensional Stein equation to obtain that

$$\begin{aligned}
\mathbb{E}Y^t\nabla f(Y) &= \sum_{i=1}^d \sum_{j=1}^k b_{i,j}\mathbb{E}\{Z_j f_i(BZ)\} \\
&= \sum_{i=1}^d \sum_{j=1}^k b_{i,j} \sum_{t=1}^k \lambda_j b_{t,j}\mathbb{E}f_{i,t}(BZ) \\
&= \mathbb{E}\{\nabla^t\Sigma\nabla f(Y)\}.
\end{aligned}$$

This finishes the proof.

A.3. Preliminaries for the proofs of Section 3. For $h \in \mathcal{H}$ define the following smoothing:

$$h_s(x) = \int_{\mathbb{R}^d} h\big(s^{1/2}y + (1-s)^{1/2}x\big)\Phi(dy), \quad 0 < s < 1.$$

We note that $\Phi h_s = \Phi h$ for any s.

A key result is the bound on the error which arises from this smoothing; it was first obtained by Götze as a version of a smoothing lemma by

Bhattacharya and Ranga Rao. We follow the exposition of Rinott and Rotar (1996).

LEMMA A.1. *Let Q be a probability measure on $\mathbb{R}^d$, and let $W \sim Q, Z \sim \Phi$. Then there exists a constant $\gamma > 0$ which depends only on the dimension d such that for $0 < t < 1$,*

$$\sup_{h\in\mathcal{H}} |\mathbb{E}h(W) - \mathbb{E}h(Z)| \quad \leq \quad \gamma \left[\sup_{h\in\mathcal{H}} |\mathbb{E}(h - \Phi h)_t(W)| + a\sqrt{t} \right].$$

The constant a is as in (3.1).

A.4. Proof of Corollary 3.1. Let $0 < t < 1$. If h is replaced by h_t in the multivariate Stein equation (2.5), then this Stein equation has solution

$$\Psi_t(x) = \frac{1}{2} \int_t^1 \frac{h_s(x) - \Phi h}{1 - s}\, ds,$$

and for $|h| \leq 1$ it is shown in Götze (1991) and also in Loh (2007), that there is a constant $\gamma = \gamma(d)$ depending only on the dimension d such that

$$|\Psi_t|_1 \leq \gamma, \quad |\Psi_t|_2 \leq \gamma \log(t^{-1}); \tag{A.1}$$

the γ is in general not equal to the γ in Lemma A.1. Following our proof we obtain, as in (2.13),

$$\begin{aligned} &\big|\mathbb{E}h_t(W) - \mathbb{E}h_t(Z)\big| = \big|\mathbb{E}\{\nabla^t\nabla\Psi_t(W) - W^t\nabla\Psi_t(W)\}\big| \\ &\quad\leq \frac{\gamma}{2}\log(t^{-1})A \\ &\qquad + \frac{1}{2}\sum_{m,i,j} \left|(\Lambda^{-1})_{m,i}\mathbb{E}(W_i' - W_i)(W_j' - W_j)(W_k' - W_k)R_{mjk}\right| \\ &\qquad + \gamma C \left(1 + d\log(t^{-1})\right), \end{aligned} \tag{A.2}$$

with A, B and C as in Theorem 2.1. For the last step we used the same estimates as applied for the remainder term in (2.13), and that $\Sigma = \mathrm{Id}$.

For the remainder term R_{mjk}, in Loh (2007), Lemma 1 (p.20) it is shown that, if $|h| \leq 1$, then there is a constant c_0 (depending only on d) such that, for any finite signed measures Q on $\mathbb{R}^d$,

$$\begin{aligned} &\sup_{1\leq p,q,r\leq d} \left| \int_{\mathbb{R}^d} \frac{\partial^3}{\partial z_p \partial z_q \partial z_r} \Psi_t(z) Q(dz) \right| \\ &\quad\leq \frac{c_0}{\sqrt{t}} \sup_{0\leq s\leq 1, y\in\mathbb{R}^d} \left| \int_{\mathbb{R}^d} h(sv + y) Q(dv) \right|. \end{aligned}$$

Thus we can bound the second term in (A.2) by $\frac{c_0}{2\sqrt{t}}B$. For simplicity we re-label γ as the maximum of γ, γ^2, and γc_0, yielding that

$$\sup_{h\in\mathcal{H}}|\mathbb{E}h(W)-\mathbb{E}h(Z)| \quad\leq\quad \gamma^2\Big(D\log(t^{-1})+\tfrac{1}{2}Bt^{-1/2}+C+a\sqrt{t}\Big),$$

with D from (3.2). The minimum with respect to t is attained for $T = \frac{1}{a^2}\left(D+\sqrt{\frac{aB}{2}+D^2}\right)^2$, which gives the assertion.

A.5. Proof of Corollary 3.2. We standardise $Y=\Sigma^{-1/2}W$. From Condition (2) we have that for any $d\times d$ matrix A and any vector $b\in\mathbb{R}^d$, $h(Ax+b)\in\mathcal{H}$, so in particular $h(\Sigma^{-1/2}x)\in\mathcal{H}$. Hence the above bounds (A.1) can be applied directly. The proof now continues along the lines of the proof of Corollary 3.1, but with the standardised variables, yielding

$$\begin{aligned}
&\big|\mathbb{E}(h-\Phi h)_t(W)\big|\\
&\leq \frac{\gamma}{2}\log(t^{-1})\sum_{i,j}\hat\lambda^{(i)}\sqrt{\operatorname{Var}\mathbb{E}^Y(Y_i'-Y_i)(Y_j'-Y_j)}\\
&\quad+\frac{\gamma}{2\sqrt{t}}\sum_{i,j,k}\hat\lambda^{(i)}\mathbb{E}\left|(Y_i'-Y_i)(Y_j'-Y_j)(Y_k'-Y_k)\right|\\
&\quad+\gamma\sum_i\hat\lambda^{(i)}\Big(\sqrt{\mathbb{E}(\Sigma^{-1/2}R)_i^2}+d\log(t^{-1})\sqrt{\mathbb{E}(\Sigma^{-1/2}R)_i^2}\Big)\\
&=\frac{\gamma}{2}\bigg\{\log(t^{-1})\sum_{i,j}\hat\lambda^{(i)}\sqrt{\operatorname{Var}\mathbb{E}^W\sum_{k,\ell}\Sigma_{i,k}^{-1/2}\Sigma_{j,\ell}^{-1/2}(W_k'-W_k)(W_\ell'-W_\ell)}\\
&\quad+\frac{1}{\sqrt{t}}\sum_{i,j,k}\hat\lambda^{(i)}\mathbb{E}\bigg|\sum_{r,s,t}\Sigma_{i,r}^{-1/2}\Sigma_{j,s}^{-1/2}\Sigma_{k,t}^{-1/2}(W_r'-W_r)(W_s'-W_s)(W_t'-W_t)\bigg|\\
&\quad+\gamma\sum_i\hat\lambda^{(i)}\bigg(\sqrt{\mathbb{E}\Big(\sum_k\Sigma_{i,k}^{-1/2}R_k\Big)^2}+d\log(t^{-1})\sqrt{\mathbb{E}\Big(\sum_k\Sigma_{i,k}^{-1/2}R_k\Big)^2}\bigg)\bigg\}.
\end{aligned}$$

The proof now follows the proof of Corollary 3.1. We omit the details. □

A.6. Details for (5.3)**.** In general, if h is twice continuously partially differentiable, then for all a and for all $i,j=1,\dots,d$, $h_{i,j}(a)=h_{j,i}(a)$. If $\nabla g=\Lambda^{-t}\nabla f$, then, with $B=(b_{i,j})_{i,j=1}^d=\Lambda^{-t}$,

$$g_i(x)=\sum_k b_{i,k}f_k(x),\quad g_j(x)=\sum_\ell b_{j,\ell}f_\ell(x).$$

If g is twice continuously partially differentiable, it follows that

$$\sum_k b_{i,k} f_{k,j}(x) = \sum_\ell b_{j,\ell} f_{\ell,i}(x).$$

For functions f which depend only on one coordinate, say j, we obtain for $i \neq j$ that $b_{i,j} f_{j,j}(x) = 0$, so that the off-diagonal elements of B all have to vanish, giving that

$$b_{i,i} f_{i,j}(x) = b_{j,j} f_{j,i}(x).$$

If f is twice continuously partially differentiable, then it follows that all diagonal elements of B have to be identical, yielding again $B = \lambda I$, where λ is a constant.

APPENDIX B: DETAILS OF THE APPLICATIONS

B.1. Details of the runs example. The following lemma may be useful when the non-diagonal entries of Λ are small compared to the diagonal-entries.

LEMMA B.1. *Assume that Λ is lower triangular and assume that there is $a > 0$ such that $|\Lambda_{i,j}| \leqslant a$ for all $j < i$. Then, with $\gamma := \inf_i |\Lambda_{ii}|$,*

$$\sup_i \lambda^{(i)} \leqslant \frac{(a/\gamma + 1)^{d-1}}{\gamma}$$

PROOF. Note that

$$|V_i' - V_i| \leqslant d + i - 2 \tag{B.1}$$

almost surely.

Write $\Lambda = \Lambda_E \Lambda_D$, where Λ_D is diagonal with the same diagonal as Λ and Λ_E is lower triangular with diagonal entries equal to 1 and $(\Lambda_E)_{i,j} := \Lambda_{i,j}/\Lambda_{j,j}$. Denote by $\|\cdot\|_p$ the usual p-norm for matrices and recall that for any matrix A, $\|A\|_1 = \sup_j \sum_i |A_{i,j}|$. Then,

$$\lambda^{(i)} \leqslant \|\Lambda^{-1}\|_1 \leqslant \|\Lambda_D^{-1}\|_1 \|\Lambda_E^{-1}\|_1.$$

Noting that $|(\Lambda_E)_{i,j}| \leqslant a/\gamma$ for all $j < i$, we have from Lemeire (1975) that

$$\|\Lambda_E^{-1}\|_1 \leqslant (a/\gamma + 1)^{d-1}.$$

Now, as $\|\Lambda_D^{-1}\|_1 = \gamma^{-1}$, the claim follows. □

From (4.2), it is easy to see that for every i and j there is a function $\nu_{i,j}$ such that

$$\mathbb{E}^{\xi}(V_i' - V_i)(V_j' - V_j) = \frac{1}{n}\sum_{m=1}^{n} \nu_{i,j}(\xi_{m-i\vee j+1}, \dots, \xi_{m+d+i\vee j-3}),$$

and $\|\nu_{i,j}\| \leqslant (d+i-2)(d+j-2) \leqslant 4d^2$ from (B.1). Write $\nu_{i,j}(m) := \nu_{i,j}(\xi_{m-i\vee j+1}, \dots, \xi_{m+d+i\vee j-3})$. As $\nu_{i,j}(m)$ and $\nu_{i,j}(m')$ are independent if $|m-m'| \geqslant 3d$, this implies

$$\begin{aligned}
&\operatorname{Var}\mathbb{E}^{W}(W_i' - W_i)(W_j' - W_j)\\
&\quad\leqslant \frac{1}{n^2 p^{i+j}(1-p)^2}\operatorname{Var}\mathbb{E}^{\xi}(V_i' - V_i)(V_j' - V_j)\\
&\quad= \frac{1}{n^4 p^{i+j}(1-p)^2}\sum_{m,m'=1}^{n}\operatorname{Cov}(\nu_{i,j}(m), \nu_{i,j}(m'))\\
&\quad\leqslant \frac{96d^5}{n^3 p^{2d}(1-p)^2}.
\end{aligned}$$

For the second summand in (2.2) we use (B.1) to obtain the simple estimate

$$\begin{aligned}
\mathbb{E}|(W_i' - W_i)(W_j' - W_j)(W_k' - W_k)| &\leqslant \frac{(d+i-2)(d+j-2)(d+k-2)}{n^{3/2}p^{3d/2}(1-p)^{3/2}}\\
&\leqslant \frac{8d^3}{n^{3/2}p^{3d/2}(1-p)^{3/2}}.
\end{aligned}$$

Applying Lemma B.1 to the matrix $n\Lambda$ with $a = 2$ and $\gamma = d-1$, we obtain

$$\lambda^{(i)} \leqslant \frac{n\left(\frac{2}{d-1}+1\right)^{d-1}}{(d-1)} \leqslant \frac{15n}{d}.$$

Combining all estimates with Theorem 2.1 proves the claim.

B.2. Details of the *U*-statistics example. As Λ is lower triangular, so is Λ^{-1} and, if $l \leqslant k$,

$$(\Lambda^{-1})_{k,l} = n/l,$$

thus, for $l = 1, \dots, d$,

$$\lambda^{(l)} \leqslant dn. \tag{B.2}$$

Define now $\eta_{j,k}(\alpha) := \psi_{j,k}'(\alpha) - \psi_k(\alpha)$. Then we have for every $k, l = 1, \dots, d$,

$$\mathbb{E}^{X,X'}\{(U_k' - U_k)(U_l' - U_l)\} = \frac{1}{n}\sum_{j=1}^{n}\Bigg(\sum_{\substack{|\alpha|=k,|\beta|=l,\\ \alpha\cap\beta\ni j}} \eta_{j,k}(\alpha)\eta_{j,l}(\beta)\Bigg) \tag{B.3}$$

and

$$\begin{aligned}
&\mathbb{E}\big(\mathbb{E}^{X,X'}\{(U'_k - U_k)(U'_l - U_l)\}\big)^2 \\
&\qquad = \frac{1}{n^2}\sum_{i,j=1}^{n}\sum_{\substack{|\alpha|=k,|\beta|=l,\\ \alpha\cap\beta\ni i}}\sum_{\substack{|\gamma|=k,|\delta|=l,\\ \gamma\cap\delta\ni j}}\mathbb{E}\{\eta_{i,k}(\alpha)\eta_{i,l}(\beta)\eta_{j,k}(\gamma)\eta_{j,l}(\delta)\}.
\end{aligned} \tag{B.4}$$

Note now that, if the sets $\alpha\cup\beta$ and $\gamma\cup\delta$ are disjoint (which can only happen if $i \neq j$),

$$\mathbb{E}\{\eta_{i,k}(\alpha)\eta_{i,k}(\beta)\eta_{j,l}(\gamma)\eta_{j,l}(\delta)\} = \mathbb{E}\{\eta_{i,k}(\alpha)\eta_{i,k}(\beta)\}\mathbb{E}\{\eta_{j,l}(\gamma)\eta_{j,l}(\delta)\} \tag{B.5}$$

due to independence. The variance of (B.3), that is (B.4) minus the square of the expectation of (B.3), contains only summands where $\alpha\cup\beta$ and $\gamma\cup\delta$ are not disjoint. Recall now that $\rho = \mathbb{E}\psi(X_1,\dots,X_d)^4$. Bounding all the non-vanishing terms simply by 32ρ, it only remains to count the number of non-vanishing terms. Thus,

$$\begin{aligned}
&\operatorname{Var}\mathbb{E}^{X,X'}(U'_k - U_k)(U'_l - U_l) \\
&\quad \leqslant \frac{1}{n^2}\sum_{i,j=1}^{n}\sum_{\substack{|\alpha|=k,|\beta|=l,\\ \alpha\cap\beta\ni i}}\sum_{\substack{|\gamma|=k,|\delta|=l,\\ \gamma\cap\delta\ni j,(\gamma\cup\delta)\cap(\alpha\cup\beta)\neq\emptyset}} 32\rho \\
&\quad = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{|\alpha|=k,|\beta|=l,\\ \alpha\cap\beta\ni i}}\Bigg(\sum_{j\in\alpha\cup\beta}\sum_{\substack{|\gamma|=k,|\delta|=l,\\ \gamma\cap\delta\ni j}} 32\rho + \sum_{j\notin\alpha\cup\beta}\sum_{\substack{|\gamma|=k,|\delta|=l,\\ \gamma\cap\delta\ni j,(\gamma\cup\delta)\cap(\alpha\cup\beta)\neq\emptyset}} 32\rho\Bigg) \\
&\quad =: A_{k,l} + B_{k,l},
\end{aligned}$$

where the equality is just a split of the sum over j into the cases whether or not $j \in \alpha\cup\beta$. In the former case we automatically have $(\alpha\cup\beta)\cap(\gamma\cup\delta) \neq \emptyset$. It is now not difficult to see that

$$A_{k,l} \leqslant \frac{32\rho(k+l-1)}{n}\binom{n-1}{k-1}^2\binom{n-1}{l-1}^2.$$

Noting that, for fixed j, k, l, α and β,

$$\begin{aligned}
&\{|\gamma| = k, |\delta| = l \,:\, \gamma\cap\delta\ni j, (\gamma\cup\delta)\cap(\alpha\cup\beta)\neq\emptyset\} \\
&\qquad = \{|\gamma| = k, |\delta| = l \,:\, \gamma\cap\delta\ni j\} \\
&\qquad\quad \setminus \{|\gamma| = k, |\delta| = l \,:\, \gamma\cap\delta\ni j, (\gamma\cup\delta)\cap(\alpha\cup\beta) = \emptyset\},
\end{aligned}$$

we further have

$$\begin{aligned}
B_{k,l} &\leqslant \frac{32\rho(n-1)}{n}\binom{n-1}{k-1}\binom{n-1}{l-1}\times \\
&\qquad \times\left\{\binom{n-1}{k-1}\binom{n-1}{l-1} - \binom{n-k-l+1}{k-1}\binom{n-k-l+1}{l-1}\right\},
\end{aligned}$$

where we also used that $\binom{n-|\alpha\cup\beta|}{k-1} \geqslant \binom{n-k-l+1}{k-1}$. The following statements are straightforward to prove:

$$\binom{n-1}{k-1}\binom{n}{k}^{-1} = \frac{k}{n}, \tag{B.6}$$

$$\binom{n-k-l+1}{k-1}\binom{n}{k}^{-1} \geqslant \frac{k}{n}\left(\frac{n-2k-l+3}{n}\right)^k \geqslant \frac{k}{n}\left(1-\frac{k(2k+l-3)}{n}\right). \tag{B.7}$$

Thus, from (B.6),

$$n^2\binom{n}{k}^{-2}\binom{n}{l}^{-2}A_{k,l} \leqslant \frac{32\rho(k+l-1)k^2l^2}{n^3} \leqslant \frac{64\rho d^5}{n^3}. \tag{B.8}$$

From (B.6) and (B.7),

$$n^2\binom{n}{k}^{-2}\binom{n}{l}^{-2}B_{k,l} \leqslant \frac{32\rho k^2l^2\big(k(2k+l-3)+l(k+2l-3)\big)}{n^3} \leqslant \frac{192\rho d^6}{n^3}.$$

Thus, for all k and l,

$$\begin{aligned}\operatorname{Var}\mathbb{E}^W(W_k'-W_k)(W_l'-W_l) &\leqslant \operatorname{Var}\mathbb{E}^{X,X'}(W_k'-W_k)(W_l'-W_l)\\ &\leqslant \frac{256\rho d^6}{n^3}.\end{aligned} \tag{B.9}$$

Notice further that for any $m = 1,\dots,d$,

$$\begin{aligned}\mathbb{E}|U_m'-U_m|^3 &= \frac{1}{n}\sum_{j=1}^n \mathbb{E}\Big|\sum_{\substack{|\alpha|=|\beta|=|\gamma|=m\\ \alpha\cap\beta\cap\gamma\ni j}} \eta_{j,m}(\alpha)\eta_{j,m}(\beta)\eta_{j,m}(\gamma)\Big|\\ &\leqslant 8\rho^{3/4}\binom{n-1}{m-1}^3,\end{aligned}$$

using (B.22); hence, along with (B.6),

$$\begin{aligned}\mathbb{E}|(W_i'-W_i)(W_k'-W_k)(W_l'-W_l)| &\leqslant \max_{m=i,k,l}\mathbb{E}|W_m'-W_m|^3\\ &\leqslant 8\rho^{3/4}n^{3/2}\max_{m=i,k,l}\binom{n}{m}^{-3}\binom{n-1}{m-1}^3\\ &\leqslant 8\rho^{3/4}d^3n^{-3/2}.\end{aligned} \tag{B.10}$$

Applying Theorem 2.1 with the estimates (B.2), (B.9) and (B.10) proves the claim.

B.3. Details of the random graph example.

B.3.1. *Calculation of the covariance matrix.* To calculate the covariance matrix Σ, we put

$$\tilde{I}_{i,j} = I_{i,j} - p$$

as the centralised edge indicator, and similarly we centralise

$$\begin{aligned}
\tilde{T} &= \sum_{i<j} \tilde{I}_{i,j}, \\
\tilde{V} &= \frac{1}{2} \sum_{i,j,k \text{ distinct}} \tilde{I}_{i,j}\tilde{I}_{j,k} = \sum_{i<j<k}, (\tilde{I}_{i,j}\tilde{I}_{j,k} + \tilde{I}_{i,j}\tilde{I}_{i,k} + \tilde{I}_{j,k}\tilde{I}_{i,k}), \\
\tilde{U} &= \sum_{i<j<k} \tilde{I}_{i,j}\tilde{I}_{j,k}\tilde{I}_{i,k}.
\end{aligned}$$

Then, by independence, all these quantities have mean zero.

For the variances, the expectation of the product of centralised indicators vanish unless all the centralised indicators involved are raised to an even power. Hence

$$\operatorname{Var}\tilde{T} = \binom{n}{2} p(1-p), \tag{B.11}$$

$$\operatorname{Var}\tilde{V} = 3\binom{n}{3} p^2(1-p)^2, \tag{B.12}$$

$$\operatorname{Var}\tilde{U} = \binom{n}{3} p^3(1-p)^3. \tag{B.13}$$

Moreover, for the same reason, all covariances between the centralised variables vanish. Expressing T, V and U, we have $\tilde{T} = T - \mathbb{E}T$ so that

$$T = \tilde{T} + \mathbb{E}T = \tilde{T} + \binom{n}{2} p \tag{B.14}$$

and

$$\operatorname{Var} T = \binom{n}{2} p(1-p) = 3\binom{n}{3} \frac{1}{n-2} p(1-p).$$

Next,

$$\begin{aligned}
\tilde{V} &= \sum_{i<j<k} (\tilde{I}_{i,j}\tilde{I}_{j,k} + \tilde{I}_{i,j}\tilde{I}_{i,k} + \tilde{I}_{j,k}\tilde{I}_{i,k}) \\
&= V - 2p \sum_{i<j<k} (I_{i,j} + I_{j,k} + I_{i,k}) + 3p^2 \binom{n}{3}.
\end{aligned}$$

Now

$$\sum_{i<j<k}(I_{i,j}+I_{j,k}+I_{i,k}) = (n-2)T.$$

Hence

$$\tilde V = V-2p(n-2)T+3p^2\binom{n}{3}$$

so that

$$V = \tilde V+2(n-2)p\tilde T+3\binom{n}{3}p^2. \tag{B.15}$$

As $\tilde V$ and $\tilde T$ are uncorrelated, this gives that

$$\begin{aligned}\operatorname{Var} V &= \operatorname{Var}\tilde V+4(n-2)^2p^2\operatorname{Var}(\tilde T)\\ &= 3\binom{n}{3}p^2(1-p)\{1-p+4(n-2)p\}.\end{aligned}$$

For U, we have

$$\begin{aligned}\tilde U &= \sum_{i<j<k}\tilde I_{i,j}\tilde I_{j,k}\tilde I_{i,k}\\ &= \sum_{i<j<k}\{I_{i,j}I_{j,k}I_{i,k}-p(I_{i,j}I_{j,k}+I_{i,j}I_{i,k}+I_{j,k}I_{i,k})\\ &\qquad\qquad +p^2(I_{i,j}+I_{j,k}+I_{i,k})-p^3\}\\ &= U-pV+p^2(n-2)T-p^3\binom{n}{3}.\end{aligned}$$

Using the above expressions (B.14) and (B.15) for T and V we obtain

$$U=\tilde U+p\tilde V+p^2(n-2)\tilde T+p^3\binom{n}{3}. \tag{B.16}$$

This gives for the variance

$$\begin{aligned}\operatorname{Var} U &= \operatorname{Var}(\tilde U)+p^2\operatorname{Var}(\tilde V)+(n-2)^2p^4\operatorname{Var}(\tilde T)\\ &= \binom{n}{3}p^3(1-p)\left\{(1-p)^2+3p(1-p)+3(n-2)p^2\right\}.\end{aligned}$$

We can now also calculate the covariances. Again we use that the centralised variables are uncorrelated to obtain

$$\begin{aligned}\mathrm{Cov}(T,V) &= \mathrm{Cov}\left(\tilde{T}, \tilde{V} + 2(n-2)p\tilde{T}\right)\\ &= 2(n-2)p\,\mathrm{Var}(\tilde{T})\\ &= 6\binom{n}{3}p^2(1-p).\end{aligned}$$

Similarly, $\mathrm{Cov}(T,U) = 3\binom{n}{3}p^3(1-p)$, and, lastly, $\mathrm{Cov}(V,U) = 3\binom{n}{3}p^3(1-p)\left(1-p+2(n-2)p\right)$. With the notation $\bar{n} = n-2$ we obtain the variance-covariance matrix

$$3\binom{n}{3}p(1-p) \times \begin{pmatrix} \frac{1}{\bar{n}} & 2p & p^2 \\ 2p & p\left(4\bar{n}p+1-p\right) & p^2\left(2\bar{n}p+1-p\right) \\ p^2 & p^2\left(2\bar{n}p+1-p\right) & p^2\left\{\bar{n}p^2 + \frac{1}{3}(1+p-2p^2)\right\} \end{pmatrix}, \tag{B.17}$$

and re-scaling yields the variance-covariance matrix (4.8).

B.3.2. *Bounding A.* As mentioned in the sketch of the proof of Proposition 4.6, for simplicity we use the uniform bound

$$|\lambda^{(i)}| \leq \frac{3}{2}n^2, \quad i = 1,2,3.$$

The conditional variances involving $T' - T$ can be calculated exactly. As $I_{i,j}^2 = I_{i,j}$,

$$\begin{aligned}\mathbb{E}^W(T'-T)^2 &= \frac{1}{\binom{n}{2}}\sum_{i<j}\mathbb{E}^W(I'_{i,j} - I_{i,j})^2\\ &= \frac{1}{\binom{n}{2}}\sum_{i<j}\{p - p\mathbb{E}^W I_{i,j} + (1-p)\mathbb{E}^W I_{i,j}\}\\ &= p + (1-2p)\frac{1}{\binom{n}{2}}T,\end{aligned}$$

so that, with $\mathrm{Var}\,T$ given in (B.17),

$$\mathrm{Var}(\mathbb{E}^W(T'-T)^2) = \frac{1}{\binom{n}{2}}(1-2p)^2p(1-p)$$

and

$$\mathrm{Var}(\mathbb{E}^W(T'_1 - T_1)^2) = \frac{(n-2)^4}{n^8\binom{n}{2}}(1-2p)^2p(1-p) < n^{-6},$$

where we used that $p(1-p) \leq 1/4$ for all p. Thus

$$\sqrt{\mathrm{Var}(\mathbb{E}^W(T_1'-T_1)^2))} < n^{-3}.$$

Next,

$$\begin{aligned}
&\mathbb{E}^W(T'-T)(V'-V)\\
&\quad= -\frac{1}{\binom{n}{2}} \sum_{i<j,\, k\neq i,j} \mathbb{E}^W(I_{i,j}'-I_{i,j})^2(I_{i,k}+I_{j,k})\\
&\quad= -\frac{1}{\binom{n}{2}} \sum_{i<j,\, k\neq i,j} \mathbb{E}^W\{p(I_{j,k}+I_{i,k})+(1-2p)(I_{i,j}I_{j,k}+I_{i,j}I_{i,k})\}\\
&\quad= \frac{1}{\binom{n}{2}}(-2(n-2)pT-2(1-2p)V).
\end{aligned}$$

So here we can also calculate the variance of the conditional expectation explicitly. With (B.15),

$$\begin{aligned}
&\mathbb{E}^W(T'-T)(V'-V)\\
&= -\frac{2}{\binom{n}{2}}\left((n-2)p\tilde{T}+(n-2)\binom{n}{2}p^2+(1-2p)\tilde{V}+2(n-2)p(1-2p)\tilde{T}\right.\\
&\qquad \left. +3\binom{n}{3}p^2(1-2p)\right),
\end{aligned}$$

so that

$$\begin{aligned}
&\mathrm{Var}\,\mathbb{E}^W(T'-T)(V'-V)\\
&\quad= \frac{4(n-2)}{\binom{n}{2}}p(1-p)\Big\{(n-2)p^2(3-4p)^2+(1-2p)^2p(1-p)\Big\} < 4,
\end{aligned}$$

where we used that $p^3(1-p) \leq \frac{27}{256}$ and that $n \geq 4$. Thus

$$\sqrt{\mathrm{Var}\,\mathbb{E}^W(T'-T)(V'-V)} < 2n^{-3}.$$

Similarly,

$$\begin{aligned}
&\mathbb{E}^W(T'-T)(U'-U)\\
&\quad= \frac{1}{\binom{n}{2}} \sum_{i<j,\, k\neq i,j} \{p\mathbb{E}^W I_{j,k}I_{i,k}+(1-2p)\mathbb{E}^W I_{i,j}I_{j,k}I_{i,k}\}\\
&\quad= \frac{1}{\binom{n}{2}}(pV+3(1-2p)U).
\end{aligned}$$

Using (B.15) and (B.16) we obtain

$$\begin{aligned}
&\mathbb{E}^W(T'-T)(U'-U)\\
&\quad= \frac{1}{\binom{n}{2}}\Big(3(1-2p)\tilde{U} + p(4-6p)\tilde{V} + (n-2)p^2(5-6p)\tilde{T}\\
&\qquad\qquad + 6\binom{n}{3}p^3(1-p)\Big).
\end{aligned}$$

Thus we calculate that

$$\begin{aligned}
&\operatorname{Var}\mathbb{E}^W(T'-T)(U'-U)\\
&= \frac{n-2}{\binom{n}{2}}p^3(1-p)\left(3(1-2p)^2(1-p)^2\right.\\
&\qquad\qquad \left. + p(1-p)(4-6p)^2 + (n-2)p^2(5-6p)^2\right).
\end{aligned}$$

Using that $p(5-6p) \le \frac{25}{24}$ and $p^3(1-p) \le \frac{27}{256}$, again we obtain

$$\sqrt{\operatorname{Var}\mathbb{E}^W(T'-T)(U'-U)} < n^{-3}.$$

For $\operatorname{Var}\mathbb{E}^W(V'-V)^2$ we introduce the notation

$$N_i = \sum_{j:j\ne i} I_{i,j}, \qquad M_{i,j} = \sum_{k:k\ne i,j} I_{i,k}I_{k,j}. \tag{B.18}$$

Then

$$T = \frac{1}{2}\sum_i N_i, \tag{B.19}$$

$$V = \frac{1}{2}\sum_{i\ne j} M_{i,j} = \frac{1}{2}\sum_{i\ne j} I_{i,j}N_i - T = \frac{1}{2}\sum_i N_i^2 - T, \tag{B.20}$$

$$U = \frac{1}{6}\sum_{i\ne j} I_{i,j}M_{i,j}. \tag{B.21}$$

We have

$$\begin{aligned}
&\mathbb{E}^W(V'-V)^2\\
&= \frac{1}{\binom{n}{2}}\sum_{i<j}\mathbb{E}^W(I_{i,j}-I'_{i,j})^2\left(N_j+N_i-2I_{i,j}\right)^2\\
&= \frac{1}{2\binom{n}{2}}\sum_{i\ne j}\left\{p\mathbb{E}^W\left(N_j+N_i-2I_{i,j}\right)^2 + (1-2p)\mathbb{E}^W I_{i,j}\left(N_j+N_i-2I_{i,j}\right)^2\right\}
\end{aligned}$$

$$= \frac{1}{2\binom{n}{2}}\bigg\{p\mathbb{E}^W\Big(4(n-2)(V+T) - 8T + 8T^2 - 16V\Big)$$

$$+ (1-2p)\mathbb{E}^W\Big(2\sum_{i\neq j} I_{i,j}N_i^2 - 8T + 2\sum_{i\neq j} I_{i,j}N_iN_j - 16V\Big)\bigg\},$$

where we used (B.19) and (B.20) for the last equation. Note that

$$\sum_i N_i^2 \quad = \quad \sum_i \sum_{j:j\neq i}\sum_{k:k\neq i} I_{i,j}I_{i,k} = 2T + 2V$$

and

$$\sum_{i\neq j} N_iN_j = 4T^2 - \sum_i N_i^2 = 4T^2 - 2T - 2V$$

as well as

$$\sum_{i\neq j} I_{i,j}N_i^2 = \sum_{i,j,k,\ell\text{ distinct}} I_{i,j}I_{i,k}I_{i,\ell} + 6V + 2T,$$

and

$$\begin{aligned}\sum_{i\neq j} I_{i,j}N_iN_j &= \sum_{i,j,k\text{ distinct}}\ \sum_{\ell:\ell\neq i,j} I_{i,j}I_{i,k}I_{j,\ell} + 2\sum_{i\neq j} I_{i,j}\sum_{k:k\neq i,j} I_{i,k} + \sum_{i\neq j} I_{i,j}\\ &= \sum_{i,j,k,\ell\text{ distinct}} I_{i,j}I_{i,k}I_{j,\ell} + 4V + 6U + 2T,\end{aligned}$$

so that

$$\begin{aligned}&\mathbb{E}^W(V'-V)^2\\ &\quad= \frac{1}{\binom{n}{2}}\bigg\{2p(n-4)T + 2V(np-10p+2) + 6(1-2p)U + 4pT^2\\ &\qquad\qquad + (1-2p)\sum_{\substack{i,j,k,\ell\\ \text{distinct}}} \mathbb{E}^W I_{i,j}I_{i,k}(I_{i,\ell}+I_{j,\ell})\bigg\}.\end{aligned}$$

With the notation $\tilde{T}$ for the centralised variable, we have that

$$\begin{aligned}&\operatorname{Var}\mathbb{E}^W(V'-V)^2\\ &\quad= \frac{1}{\binom{n}{2}^2}\operatorname{Var}\bigg\{p(2n-8+4pn^2-4pn)T + 2V(np-10p+2) + 6(1-2p)U\\ &\qquad\qquad + 4p\tilde{T}^2 + (1-2p)\sum_{\substack{i,j,k,\ell\\ \text{distinct}}} \mathbb{E}^W I_{i,j}I_{i,k}(I_{i,\ell}+I_{j,\ell})\bigg\}\end{aligned}$$

$$
\begin{aligned}
\leq\ & 5\frac{1}{\binom{n}{2}^2}\Big\{p^2(2n-8+4pn^2-4pn)^2\operatorname{Var}(T)+4(np-10p+2)^2\operatorname{Var}(V) \\
& +36(1-2p)^2\operatorname{Var}(U)+16p^2\operatorname{Var}(\tilde{T}^2) \\
& +(1-2p)^2\operatorname{Var}\Big(\sum_{\substack{i,j,k,\ell \\ \text{distinct}}}\mathbb{E}^W I_{i,j}I_{i,k}(I_{i,\ell}+I_{j,\ell})\Big)\Big\},
\end{aligned}
$$

where we used that in general $\operatorname{Var}\sum_{i=1}^k X_i \leq k\sum_{i=1}^k \operatorname{Var} X_i$ and (B.11) for the last inequality. Here, the variances for T, V and U are given in (B.17). To simplify the expression, we use that $p^3(1-p) \leq 27/256$ to bound

$$
p^2(2n-8+4pn^2-4pn)^2\operatorname{Var}(T) \leq \frac{27}{64}\binom{n}{2}n^2(n+2)^2.
$$

Similarly, we bound with $p^2(1-p) \leq 4/27$ and $n \geq 4$

$$
4(np-10p+2)^2\operatorname{Var}(V) \leq \frac{16}{27}n^3(n-1)(n-2)(n+1),
$$

and

$$
36(1-2p)^2\operatorname{Var}(U) \leq \frac{81}{256}n(n-1)(n-2)(3n+2).
$$

We note that $\mathbb{E}\tilde{I}_{i,j}\tilde{I}_{u,v}\tilde{I}_{s,t}\tilde{I}_{k,\ell} = 0$ unless either all pairs of indices are the same, or the product is made up of two distinct index pairs only. Hence

$$
\begin{aligned}
\operatorname{Var}\tilde{T}^2 &= \sum_{i<j}\sum_{u<v}\sum_{s<t}\sum_{k<\ell}\mathbb{E}\tilde{I}_{i,j}\tilde{I}_{u,v}\tilde{I}_{s,t}\tilde{I}_{k,\ell} \\
&< n^2\binom{n}{2}p(1-p),
\end{aligned}
$$

giving

$$
16p^2\operatorname{Var}\tilde{T}^2 \leq \frac{27}{32}n^3(n-1).
$$

For the last variance term, we use that conditional variances can be bounded by unconditional variances, giving

$$
\begin{aligned}
& \operatorname{Var}\sum_{i\neq j}\sum_{k:k\neq i,j}\sum_{\ell:\ell\neq i,j,k}\mathbb{E}^W I_{i,j}I_{i,k}(I_{i,\ell}+I_{j,\ell}) \\
\leq\ & \operatorname{Var}\sum_{\substack{i,j,k,\ell \\ \text{distinct}}} I_{i,j}I_{i,k}(I_{i,\ell}+I_{j,\ell})
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{\substack{i,j,k,\ell\\ \text{distinct}}} \operatorname{Var} I_{i,j} I_{i,k}(I_{i,\ell}+I_{j,\ell})\\
&\quad + \sum_{i,j,k,\ell \text{ distinct}} \ \sum_{r,s,t,u \text{ distinct}} \mathbf{1}((i,j,k,\ell) \neq (r,s,t,u))\\
&\quad \times \mathbf{1}(|\{i,j,k,\ell\} \cap \{r,s,t,u\}| \geq 2)\\
&\quad \times \operatorname{Cov}(I_{i,j}I_{i,k}(I_{i,\ell}+I_{j,\ell}), I_{r,s}I_{r,t}(I_{r,u}+I_{s,u})))\\
&\leq 2\binom{n}{4}\left(p^3(1-p^3) + 4\binom{4}{2}\binom{n}{2}p^2(1-p^4)\right)\\
&< 3n^2\binom{n}{4}.
\end{aligned}
$$

Here we used the independence of the edge indicators. For the last bound we employed that $p^3(1-p^3) \leq 1/4$, that $p^2(1-p^4) \leq (\sqrt{3}-1)/3$, and that $n \geq 4$. Collecting the variances and using that $n \geq 4$,

$$
\begin{aligned}
&\operatorname{Var}(\mathbb{E}^W(V'-V)^2)\\
&\leq 5\frac{1}{\binom{n}{2}^2}\left\{\frac{27}{64}\binom{n}{2}n^2(n+2)^2 + \frac{16}{27}n^3(n-1)(n-2)(n+1)\right.\\
&\qquad\qquad \left. + \frac{81}{256}n(n-1)(n-2)(3n+2) + \frac{27}{32}n^3(n-1) + 3n^2\binom{n}{4}\right\}\\
&< 33n^2.
\end{aligned}
$$

This gives that $\sqrt{\operatorname{Var}(\mathbb{E}^W(V_1'-V_1)^2)} < 6n^{-3}$.

For $\mathbb{E}^W(V'-V)(U'-U)$, we have

$$
\begin{aligned}
&\mathbb{E}^W(V'-V)(U'-U)\\
&\quad = \frac{1}{\binom{n}{2}} \sum_{i<j} \mathbb{E}^W(I_{i,j}-I'_{i,j})^2(N_i+N_j-2I_{i,j})M_{i,j}\\
&\quad = \frac{1}{2\binom{n}{2}} \sum_{i\neq j}\Big\{p\mathbb{E}^W(N_i+N_j-2I_{i,j})M_{i,j}\\
&\qquad\qquad\qquad + (1-2p)\mathbb{E}^W I_{i,j}(N_i+N_j-2I_{i,j})M_{i,j}\Big\}.
\end{aligned}
$$

Recall (B.21), so that

$$
\begin{aligned}
&\mathbb{E}^W(V'-V)(U'-U)\\
&\quad = \frac{1}{\binom{n}{2}}\left(p\sum_{i\neq j}\mathbb{E}^W N_i M_{i,j} - 6(1-p)U + (1-2p)\sum_{i\neq j}\mathbb{E}^W I_{i,j}N_iM_{i,j}\right).
\end{aligned}
$$

Now

$$\begin{aligned}\sum_{i\neq j} N_i M_{i,j} &= 2V + 6U + \sum_{\substack{i,j,k,\ell\\ \text{distinct}}} I_{i,k}I_{k,j}I_{i,\ell}, \text{ and}\\ \sum_{i\neq j} I_{i,j}N_i M_{i,j} &= 12U + \sum_{\substack{i,j,k,\ell\\ \text{distinct}}} I_{i,j}I_{i,k}I_{i,\ell}I_{\ell,j},\end{aligned}$$

so that

$$\begin{aligned}&\mathbb{E}^W(V'-V)(U'-U)\\ &= \frac{1}{\binom{n}{2}}\Bigg(2pV + 6(1-2p)U + p\sum_{\substack{i,j,k,\ell\\ \text{distinct}}} I_{i,k}I_{k,j}I_{i,\ell}\\ &\qquad\qquad + (1-2p)\sum_{\substack{i,j,k,\ell\\ \text{distinct}}} I_{i,j}I_{i,k}I_{i,\ell}I_{\ell,j}\Bigg).\end{aligned}$$

Furthermore, as before,

$$\operatorname{Var}\sum_{i,j,k,\ell \text{ distinct}} \mathbb{E}^W I_{i,k}I_{k,j}I_{i,\ell} < \binom{n}{4}n^2.$$

Similarly as for (B.22),

$$\begin{aligned}\operatorname{Var}\sum_{\substack{i,j,k,\ell\\ \text{distinct}}} \mathbb{E}^W I_{i,j}I_{i,k}I_{i,\ell}I_{j,\ell} &\le \operatorname{Var}\sum_{\substack{i,j,k,\ell\\ \text{distinct}}} I_{i,j}I_{i,k}I_{i,\ell}I_{j,\ell}\\ &\le \binom{n}{4}\left(p^4(1-p^4) + 6\binom{n}{2}p^2(1-p^6)\right)\\ &< \binom{n}{4}\left(\frac{1}{256} + \frac{1}{16}\binom{n}{2}\right).\end{aligned}$$

As $p<1$, we obtain that

$$\begin{aligned}&\operatorname{Var}\mathbb{E}^W(V'-V)(U'-U)\\ &< 4\frac{1}{\binom{n}{2}^2}\Bigg\{12\frac{27}{256}\binom{n}{3}(16(n-2)+1) + 9n + 9)\\ &\qquad + \binom{n}{4}n^2 + \binom{n}{4}\left(\frac{1}{256} + \frac{1}{16}\binom{n}{2}\right)\Bigg\}\\ &< n^2 + 108\end{aligned}$$

so that

$$\sqrt{\mathrm{Var}\left(\mathbb{E}^W(V_1' - V_1)(U_1' - U_1)\right)} \quad < \quad n^{-3} + 11n^{-4}.$$

Finally,

$$\mathbb{E}^W(U' - U)^2 \quad = \quad \frac{1}{2\binom{n}{2}} \sum_{i \neq j} \left(p\mathbb{E}^W M_{i,j}^2 + (1-2p)\mathbb{E}^W I_{i,j} M_{i,j}^2 \right).$$

We have that

$$M_{i,j}^2 = \sum_{k:k\neq i,j} \sum_{\ell:\ell\neq i,j} I_{i,k} I_{k,j} I_{i,\ell} I_{\ell,j} = M_{i,j} + \sum_{k:k\neq i,j} \sum_{\ell:\ell\neq i,j,k} I_{i,k} I_{k,j} I_{i,\ell} I_{\ell,j},$$

and

$$I_{i,j} M_{i,j}^2 \quad = \quad I_{i,j} M_{i,j} + \sum_{k:k\neq i,j} \sum_{\ell:\ell\neq i,j,k} I_{i,j} I_{i,k} I_{k,j} I_{i,\ell} I_{\ell,j},$$

so that

$$\begin{aligned} &\mathbb{E}^W(U' - U)^2 \\ &\quad = \quad \frac{1}{2\binom{n}{2}} \Bigg\{ 2pV + 6(1-2p)U + p \sum_{\substack{i,j,k,\ell \\ \text{distinct}}} \mathbb{E}^W I_{i,k} I_{k,j} I_{i,\ell} I_{\ell,j} \\ &\qquad\qquad + (1-2p) \sum_{\substack{i,j,k,\ell \\ \text{distinct}}} \mathbb{E}^W I_{i,j} I_{i,k} I_{k,j} I_{i,\ell} I_{\ell,j} \Bigg\}. \end{aligned}$$

As for (B.22), we obtain

$$\mathrm{Var} \sum_{\substack{i,j,k,\ell \\ \text{distinct}}} \mathbb{E}^W I_{i,k} I_{k,j} I_{i,\ell} I_{j,\ell} \leq \binom{n}{4} \left(p^4(1-p^4) + 6\binom{n}{2} p^2(1-p^6) \right)$$

and

$$\mathrm{Var} \sum_{\substack{i,j,k,\ell \\ \text{distinct}}} \mathbb{E}^W I_{i,j} I_{i,k} I_{k,j} I_{i,\ell} I_{j,\ell} \leq \binom{n}{4} \left(p^5(1-p^5) + 6\binom{n}{2} p^2(1-p^8) \right).$$

Again using our variance inequalities, we thus obtain that

$$\begin{aligned}
&\mathrm{Var}\left(\mathbb{E}^W(U'-U)^2\right)\\
&\leq \frac{1}{\binom{n}{2}^2}\Bigg\{3\binom{n}{3}p^3(1-p)\Big(4p(4(n-2)p+1-p)\\
&\qquad +36(1-2p)^2((n-2)p^2+\frac{1}{3}(4-5p+p^2))\Big)\\
&\qquad +p^2\binom{n}{4}\left(p^4(1-p^4)+6\binom{n}{2}p^2(1-p^6)\right)\\
&\qquad +(1-2p)^2\binom{n}{4}\left(p^5(1-p^5)+6\binom{n}{2}p^2(1-p^8)\right)\Bigg\}\\
&\leq 22+2n^2,
\end{aligned}$$

so that

$$\sqrt{\mathrm{Var}\left(\mathbb{E}^W(U'-U)^2\right)} < 5n^{-3}+2n^{-4}.$$

Collecting these bounds we obtain for A in Theorem 2.1 that

$$A \quad < \quad 35n^{-1}+36n^{-2}.$$

B.3.3. *Bounding B.* We use the generalised Hölder inequality

$$\mathbb{E}\prod_{i=1}^{3}|X_i| \leq \prod_{i=1}^{3}\{\mathbb{E}|X_i|^3\}^{\frac{1}{3}} \leq \max_{i=1,2,3}\mathbb{E}|X_i|^3. \qquad \text{(B.22)}$$

Firstly,

$$\mathbb{E}|T'-T|^3 = \frac{1}{\binom{n}{2}}\sum_{i<j}\mathbb{E}|I_{i,j}-I'_{i,j}|^3 = 2p(1-p) < \frac{1}{2},$$

so that

$$\mathbb{E}|T_1'-T_1|^3 \quad = \quad \frac{(n-2)^3}{n^6}2p(1-p) < \frac{1}{2}n^{-3}.$$

Similarly,

$$\begin{aligned}
&\mathbb{E}|V'-V|^3\\
&= \frac{1}{\binom{n}{2}}\sum_{i<j}\mathbb{E}|I_{i,j}-I'_{i,j}|^3\sum_{k,\ell,s:k,\ell,s\neq i,j}(I_{j,k}+I_{i,k})(I_{j,\ell}+I_{i,\ell})(I_{j,s}+I_{i,s})\\
&= 2p(1-p)(n-2)\times\\
&\quad\times\left(8p^2+2p(1-p)+2(n-3)(2p^2+2p^3)+8(n-3)(n-4)p^3\right),
\end{aligned}$$

so that

$$\mathbb{E}|V_1' - V_1|^3 \quad < \quad \frac{64}{27}\left(n^{-3} + n^{-4} + n^{-5}\right).$$

Lastly,

$$\begin{aligned}
&\mathbb{E}|U' - U|^3 \\
&= \frac{1}{\binom{n}{2}} \sum_{i<j} \mathbb{E}|I_{i,j} - I'_{i,j}|^3 \sum_{k:k\neq i,j} \sum_{\ell:\ell\neq i,j} \sum_{s:s\neq i,j} I_{j,k}I_{i,k}I_{j,\ell}I_{i,\ell}I_{j,s}I_{i,s} \\
&= 2p(1-p)(n-2)\left(p^2 + (n-3)p^4 + (n-3)(n-4)p^6\right),
\end{aligned}$$

so that

$$\mathbb{E}|U_1' - U_1|^3 \quad < \quad \frac{54}{256}\left(n^{-3} + n^{-4} + n^{-5}\right).$$

Thus for B in Theorem 2.1 we have

$$B < \frac{3}{2}n^2 \times 9 \times \frac{64}{27}\left(n^{-3} + n^{-4} + n^{-5}\right) = 32\left(n^{-1} + n^{-2} + n^{-3}\right).$$

Collecting the bounds gives the result.

Department of Statistics
University of Oxford
1 South Parks Road
Oxford OX1 3TG
United Kingdom