J. Appl. Prob. **47,** 378–393 (2010)
Printed in England

RANDOM SUBGRAPH COUNTS AND U-STATISTICS: MULTIVARIATE NORMAL APPROXIMATION VIA EXCHANGEABLE PAIRS AND EMBEDDING

GESINE REINERT,* *University of Oxford*
ADRIAN RÖLLIN,** *National University of Singapore*

Abstract

In Reinert and Röllin (2009) a new approach—called the 'embedding method'—was introduced, which allows us to make use of exchangeable pairs for normal and multivariate normal approximations with Stein's method in cases where the corresponding couplings do not satisfy a certain linearity condition. The key idea is to embed the problem into a higher-dimensional space in such a way that the linearity condition is then satisfied. Here we apply the embedding to U-statistics as well as to subgraph counts in random graphs.

Keywords: Multivariate normal distribution; U-statistics; random graph statistics; Stein's method; exchangeable pair

2010 Mathematics Subject Classification: Primary 60F05
Secondary 62E17; 05C80

1. Introduction

Stein's method, first introduced in the 1970s [15], has proved a powerful tool for assessing distributional distances, such as the normal distribution, in the presence of dependence. When considering sums W of random variables, the dependence between these random variables needs to be weak in order for the distance to a normal distribution to be small. For quantifying weak dependence, Stein [16] introduced the method of exchangeable pairs: construct a sum W' such that (W, W') form an exchangeable pair, and such that $\mathrm{E}^W(W' - W)$ is (at least approximately) linear in W. This linearity condition arises naturally when thinking of correlated bivariate normals. The generalisation of this approach to a multivariate setting remained untackled until recently when Chatterjee and Meckes [4] solved the problem in the case of exchangeable vectors (W, W') such that $\mathrm{E}^W(W' - W) = -\lambda W + R$, where λ is a scalar and R is a remainder vector, with $\mathrm{E}\,|R|$ small. This is a rather special case; in [11] we considered the general setting where

$$\mathrm{E}^W(W' - W) = -\Lambda W + R \tag{1}$$

for a matrix Λ and a vector R with small $\mathrm{E}\,|R|$. In a follow-up paper by Meckes [8] the results of [4] and [11] are combined using slightly different smoothness conditions on test functions as compared to [11]; nonsmooth test functions are not treated by Meckes, but the bounds obtained there improve on those from [11] for the example of d-runs with respect to smooth test functions.

Received 11 January 2010; revision received 30 March 2010.
* Postal address: Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK.
Supported in part by BBSRC and EPSRC through OCISB.
** Postal address: Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546. Email address: staar@nus.edu.sg
Supported in part by the Swiss National Science Foundation project PBZH2–117033.

A surprising finding in [11] was that it is often possible to embed a random vector W into a random vector $\hat{W}$ of larger, but still finite, dimension, such that (1) holds with $R = 0$; this embedding does not directly correspond to Hoeffding projections, although it is related to them. Here we explore the embedding method further, by illustrating its use in two important examples. In the first example we consider complete nondegenerate U-statistics, and in the second example we consider the joint count of edges and triangles in Bernoulli random graphs. In both examples the limiting covariance matrix is not of full rank, yet the bounds on the normal approximation are of the expected order.

The paper is organised as follows. In Section 2 we review the theoretical results in [11], giving bounds on the distance to normal distributions under the linearity condition (1), both for smooth test functions and nonsmooth test functions. In Section 3 we discuss the embedding method, and point out a link to Rademacher integrals and chaos decompositions. In Section 4 we illustrate the embedding method for complete nondegenerate U-statistics; the embedding vector contains lower-order U-statistics which are obtained via fixing components. Section 5 gives a normal approximation for the joint counts of the number of edges and the number of triangles in a Bernoulli random graph; to the authors' knowledge, these are the first explicit bounds for this multivariate problem. The embedding method suggests counting the number of 2-stars as well, which makes the results not only more informative but also, surprisingly, easier to derive.

2. Theoretical bounds for a multivariate normal approximation

2.1. Notation

Denote by $W = (W_1, W_2, \ldots, W_d)^\top$ random vectors in $\mathbb{R}^d$, where W_i are $\mathbb{R}$-valued random variables for $i = 1, \ldots, d$. We denote by Σ symmetric, nonnegative definite matrices, and, hence, by $\Sigma^{1/2}$ the unique symmetric square root of Σ. Denote by Id the identity matrix, where we omit the dimension d. Throughout this paper, Z denotes a random variable having standard d-dimensional multivariate normal distribution. We abbreviate the transpose of the inverse of a matrix Λ as $\Lambda^{-\top} := (\Lambda^{-1})^\top$.

For derivatives of smooth functions $h\colon \mathbb{R}^d \to \mathbb{R}$, we use the notation ∇ for the gradient operator. Denote by $\|\cdot\|$ the supremum norm for both functions and matrices. If the corresponding derivatives exist for some function $h\colon \mathbb{R}^d \to \mathbb{R}$, we abbreviate

$$|h|_1 := \sup_i \left\| \frac{\partial}{\partial x_i} h \right\|, \qquad |h|_2 := \sup_{i,j} \left\| \frac{\partial^2}{\partial x_i \partial x_j} h \right\|,$$

and so on.

We start by considering smooth test functions.

Theorem 1. (Cf. [11, Theoem 2.1].) *Assume that (W, W') is an exchangeable pair of $\mathbb{R}^d$-valued random variables such that*

$$\mathrm{E}\, W = 0, \qquad \mathrm{E}\, W W^\top = \Sigma,$$

with $\Sigma \in \mathbb{R}^{d \times d}$ symmetric and positive definite. Suppose further that (1) 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,

$$|\mathrm{E} h(W) - \mathrm{E}\, h(\Sigma^{1/2} Z)| \le \frac{|h|_2}{4} A + \frac{|h|_3}{12} B + \left(|h|_1 + \frac{1}{2} d \|\Sigma\|^{1/2} |h|_2 \right) C, \tag{2}$$

https://doi.org/10.1239/jap/1276784898 Published online by Cambridge University Press

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

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

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

$$C = \sum_{i} \lambda^{(i)} \sqrt{\mathrm{E}\, R_i^2}.$$

Note that Theorem 1 is a nonasymptotic result and, therefore, yields bounds for any finite n. These bounds do not require underlying independence of the random variables, although our examples will illustrate that in order to assess the quantities A, B, and C, the explicit structure of the problem needs to be taken into account. With respect to the requirements on the derivatives of the test functions, bounds of the form (2) are comparable with those that could be obtained from Lindeberg's proof of the central limit theorem; see, for example, [3] for the one-dimensional case. Lindeberg's original proof for independent summands allows for a straightforward adaptation to temporal dependence structures, such as martingales; cf., for example, [1] and [14]. Stein's method, however, makes no assumption about the ordering of the index set. In applications such as U-statistics, this ordering is typically introduced 'artificially' (see [7, p. 118ff.] and also [5] for a graph related example), and, hence, bounds of the form of our Theorem 1 appear more natural in this context, as the order of the index set does not influence the bounds. In particular, to obtain optimal bounds in the Kolmogorov metric, Stein's method is typically better suited than a martingale approach in cases where martingales are only an auxiliary construct and are not intrinsic to the particular problem.

The proof of Theorem 1 is based on the Stein characterization of the normal distribution which states that $Y \in \mathbb{R}^d$ is multivariate normal, $\mathrm{MVN}(0, \Sigma)$, if and only if

$$\mathrm{E}\, Y^{\top} \nabla f(Y) = \mathrm{E}\, \nabla^{\top} \Sigma \nabla f(Y) \quad \text{for all smooth } f : \mathbb{R}^d \to \mathbb{R}.$$

In Meckes [8] a different norm for functions and operators is used to obtain a similar result, and the difference in the bounds depending on the chosen norm is illustrated for the example of runs on the line.

Theorem 1 can be extended to allow for covariance matrices which are not full rank, using the triangle inequality in conjunction with the following proposition.

Proposition 1. (Cf. [11, Proposition 2.8].) *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 nonnegative definite. Let $h : \mathbb{R}^d \to \mathbb{R}$ have two bounded derivatives. Then*

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

For nonsmooth test functions, following [12], let Φ denote the standard normal distribution in $\mathbb{R}^d$, and let ϕ be the corresponding density function. For $h : \mathbb{R}^d \to R$, set

$$h_\delta^+(x) = \sup\{h(x+y) : |y| \le \delta\}, \qquad h_\delta^-(x) = \inf\{h(x+y) : |y| \le \delta\},$$
$$\text{and} \quad \tilde{h}(x, \delta) = h_\delta^+(x) - h_\delta^-(x).$$

Let $\mathcal{H}$ be a class of measurable functions $\mathbb{R}^d \to \mathbb{R}$ which are uniformly bounded by 1. Suppose that, for any $h \in \mathcal{H}$ and any $\delta > 0$, $h_\delta^+(x)$ and $h_\delta^-(x)$ are in $\mathcal{H}$; for any $d \times d$ matrix A and any vector $b \in \mathbb{R}^d$, $h(Ax+b) \in \mathcal{H}$; and, for some constant $a = a(\mathcal{H}, \delta)$, $\sup_{h\in\mathcal{H}}\{\int_{\mathbb{R}^d} \tilde{h}(x,\delta)\Phi(\mathrm{d}x)\} \le a\delta$. Obviously, we may assume that $a \ge 1$. The class of indicators of measurable convex sets is a class where $a \le 2\sqrt{d}$; see [2].

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

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

and

$$A' = \sum_{i,j} \hat{\lambda}^{(i)} \sqrt{\operatorname{var} \mathrm{E}^Y \sum_{k,\ell} \Sigma_{i,k}^{-1/2}\Sigma_{j,\ell}^{-1/2}(W_k' - W_k)(W_\ell' - W_\ell)},$$

$$B' = \sum_{i,j,k} \hat{\lambda}^{(i)}\, \mathrm{E}\left|\sum_{r,s,t} \Sigma_{i,r}^{-1/2}\Sigma_{j,s}^{-1/2}\Sigma_{k,t}^{-1/2}(W_r' - W_r)(W_s' - W_s)(W_t' - W_t)\right|,$$

$$C' = \sum_i \hat{\lambda}^{(i)} \sqrt{\mathrm{E}\Big(\sum_k \Sigma_{i,k}^{-1/2} R_k\Big)^2},$$

we have the following result [11].

Corollary 1. *Let W be as in Theorem 1. Then, for all $h \in \mathcal{H}$ with $|h| \le 1$, there exist $\gamma = \gamma(d)$ and $a > 1$ such that*

$$\sup_{h\in\mathcal{H}} |\mathrm{E}h(W) - \mathrm{E}\,h(Z)| \le \gamma^2\left(-D'\log(T') + \frac{B'}{2\sqrt{T'}} + C' + a\sqrt{T'}\right),$$

where

$$T' = \frac{1}{a^2}\left(D' + \sqrt{\frac{aB'}{2} + D'^2}\right)^2 \quad \textit{and} \quad D' = \frac{A'}{2} + C'd.$$

Remark 1. We can simplify the above bound further. Using Minkowski's inequality, we have $\operatorname{var}\sum_{i=1}^k X_i \le k^2 \sup_i \operatorname{var} X_i$. Thus, we obtain the simple estimate

$$\begin{aligned}\operatorname{var} \mathrm{E}^Y &\sum_{k,\ell} \Sigma_{i,k}^{-1/2}\Sigma_{j,\ell}^{-1/2}(W_k' - W_k)(W_\ell' - W_\ell)\\ &\le d^4 \|\Sigma^{-1/2}\|^4 \sup_{k,\ell} \operatorname{var} \mathrm{E}^W\{(W_k' - W_k)(W_\ell' - W_\ell)\},\end{aligned}$$

and, hence,

$$A' \le d^3\|\Sigma^{-1/2}\|^2 \sum_i \hat{\lambda}^{(i)} \sup_{k,\ell}\sqrt{\operatorname{var} \mathrm{E}^W\{(W_k' - W_k)(W_\ell' - W_\ell)\}};$$

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 5, where $\|\Sigma^{-1/2}\|$ provides a noninformative bound.

https://doi.org/10.1239/jap/1276784898 Published online by Cambridge University Press

Remark 2. Note that, if (W, W') is exchangeable and (1) is satisfied, we have

$$\mathrm{E}(W' - W)(W' - W)^\top = 2\,\mathrm{E}\, W(\Lambda W)^\top = 2\Sigma\Lambda^\top. \tag{3}$$

On the other hand, if we only have $\mathscr{L}(W) = \mathscr{L}(W')$, we obtain

$$\mathrm{E}(W' - W)(W' - W)^\top = \Lambda\Sigma + \Sigma\Lambda^\top. \tag{4}$$

Hence, to check in an application whether the often tedious calculation of Σ and Λ has been carried out correctly, we can combine (3) and (4) to conclude that, under the conditions of Theorem 1, we must have $\Lambda\Sigma = \Sigma\Lambda^\top$.

3. The embedding method

Assume that an ℓ-dimensional random vector $W_{(\ell)}$ 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 independent and identically distributed (i.i.d.) random variables $\mathbb{X} = (X_1, \dots, X_n)$, we can uniformly choose 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 guarantee that $(W_{(\ell)}, W'_{(\ell)})$ will satisfy condition (1) when R is of the required smaller order or even equal to 0; hence, in this case Theorem 1 would not yield useful bounds.

Surprisingly, it is often possible to extend $W_{(\ell)}$ to a vector $W \in \mathbb{R}^d$ such that we can construct an exchangeable pair (W, W') which satisfies condition (1) with $R = 0$. If we can bound the distance of the distribution $\mathscr{L}(W)$ to a d-dimensional multivariate normal distribution then a bound on the distance of the distribution $\mathscr{L}(W_{(\ell)})$ to an ℓ-dimensional multivariate normal distribution follows immediately.

In order to obtain useful bounds in Theorem 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.

As a first illustration of the method, it was observed in [9] that, for functions which depend on the first d coordinates of an infinite Rademacher sequence, that is, a sequence of symmetric $\{-1, 1\}$ random variables, the natural embedding vector is a vector of Rademacher integrals of lower order. A similar construction works fairly generally as follows. Assume that $F = F(X_1, \dots, X_d)$ is a random variable that depends uniquely on the first d coordinates of a sequence X of i.i.d. mean 0 random variables, with $\mathrm{E}(F) = 0$ and $\mathrm{E}(F^2) = 1$, of the form

$$F = \sum_{n=1}^{d} \sum_{1 \le i_1 < \cdots < i_n \le d} n!\, f_n(i_1, \dots, i_n) X_{i_1} \cdots X_{i_n} =: \sum_{n=1}^{d} J_n(f_n); \tag{5}$$

such representations occur as chaotic decompositions for functionals of Rademacher sequences. A natural exchangeable pair construction is as follows. Pick an index I so that $\mathrm{P}(I = i) = 1/d$ for $i = 1, \dots, d$, independently of $X_1, \dots, X_d$, and if $I = i$, replace X_i by an independent copy X_i^* in all sums in decomposition (5) which involve X_i. Call the resulting expression F', and the corresponding sums $J'_n(f_n)$, $n = 1, \dots, d$. Now choosing $W = (J_1(f_1), \dots, J_d(f_d))$

as the embedding vector, we check that, for all $n = 1, \dots, d$,

$$\begin{aligned}
&\mathrm{E}(J_n'(f_n) - J_n(f_n) \mid W) \\
&\quad = -\frac{1}{d}\sum_{i=1}^{d}\sum_{1\le i_1<\cdots<i_n\le d} \mathbf{1}_{\{i_1,\dots,i_n\}}(i) n!\, f_n(i_1,\dots,i_n)\,\mathrm{E}(X_{i_1}\cdots X_{i_n} \mid W) \\
&\quad = -\frac{n}{d}\, J_n(f_n).
\end{aligned}$$

Thus, with $W' = (J_1'(f_1), \dots, J_d'(f_d))$, condition (1) is satisfied, with the matrix $\Lambda = (\lambda_{i,j})_{1\le i,j\le d}$ having all of its off-diagonal entries equal to 0 and its diagonal entries, $\lambda_{n,n}$, equal to n/d for $n = 1, \dots, d$. Note that, although diagonal, the diagonal entries of this Λ are not equal. It is not possible to correct this by simple coordinatewise scaling of W, as this will change Σ only and leave Λ unaffected; see also the discussion in [11, Section 5]. Hence, again, the generality of (1) is essential here.

4. Complete nondegenerate U-statistics

Using the exchangeable pairs coupling, Rinott and Rotar [13] proved a univariate normal approximation theorem for nondegenerate weighted U-statistics with symmetric weight function under fairly mild conditions on the weights. Using the typical coupling, where uniformly a random variable X_i is chosen and replaced by an independent copy, they showed that (1) is satisfied for the one-dimensional case and a nontrivial remainder term, corresponding to Hoeffding projections of smaller order. It should not be difficult (but nevertheless cumbersome) to generalise their result to the multivariate case, where d different U-statistics are regarded based on the same sample of independent random variables, such that (1) is satisfied with $\Lambda = I$ and a nontrivial remainder term, again of lower order; for multivariate approximations of several U-statistics, see also the book by Lee [7]. However, as we want to emphasize the use of Theorem 1 for nondiagonal Λ, we take a different approach.

Let $X_1, \dots, X_n$ be a sequence of i.i.d. random elements taking values in a measure 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

$$\psi_k(X_1, \dots, X_k) := \mathrm{E}(\psi(X_1, \dots, X_d) \mid X_1, \dots, X_k).$$

Assume without loss of generality that $\mathrm{E}\,\psi(X_1, \dots, X_d) = 0$. For any subset $\alpha \subset \{1, \dots, n\}$ of size k, write $\psi_k(\alpha) := \psi(X_{i_1}, \dots, X_{i_k})$, 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. Assume that U_d is nondegenerate, that is, $\mathrm{P}(\psi_1(X_1) = 0) < 1$. Set

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

It is well known that $\operatorname{var} W_k \asymp 1$ (see, e.g. [7, p. 10ff.]). Note also that, as $n \to \infty$, $\Sigma := \mathrm{E}(WW^{\top})$ will converge to a covariance matrix with all entries equal to $\operatorname{var}\psi_1(X_1)$ and which is thus of rank 1, as we assume nondegeneracy and, hence, $U_1 = \sum_{i=1}^n \psi_1(X_i)$ will dominate the behaviour of each U_k.

https://doi.org/10.1239/jap/1276784898 Published online by Cambridge University Press

We can make the connection with Hoeffding projections more explicit. If $H^{(j)}$ denotes the jth Hoeffding projections of $\binom{n}{k}^{-1}U_k$ (for a detailed discussion, we refer the reader to [7, p. 25ff.]), then we have the representation

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

for each $k = 1, \dots, d$. Thus, for complete U-statistics and with the coupling used in Theorem 2, below, the embedding vector for U_d consists of weighted averages of the Hoeffding projections. However, whereas the Hoeffding projections are unique, our embedding is not and will depend on the specific coupling.

Using Stein's method and the approach of decomposable random variables, Raič [10] 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, as in our case, a normalization is not appropriate.

Theorem 2. *With the above notation, and if $\rho := \mathrm{E}\,\psi(X_1, \dots, X_d)^4 < \infty$, we have, for every three times differentiable function h,*

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

Proof. Let $X_1', \dots, X_n'$ be independent copies of $X_1, \dots, X_n$. For any subset $\alpha \subset \{1, \dots, n\}$ of size k, 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 [13], that is, uniformly pick 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. Now note that, if $j \notin \alpha$, $\psi_{j,k}'(\alpha) = \psi_k(\alpha)$, and, with $X = (X_1, \dots, X_n)$, that $\mathrm{E}^X\,\psi_{j,k}'(\alpha) = \psi_{k-1}(\alpha \setminus \{j\})$ if $j \in \alpha$. Thus,

$$\begin{aligned}
\mathrm{E}^X(U_k' - U_k) &= \frac{1}{n}\sum_{j=1}^{n}\sum_{\substack{|\alpha|=k\\ \alpha\ni j}}\mathrm{E}^X(\psi_{j,k}'(\alpha) - \psi_k(\alpha)) \\
&= -\frac{k}{n}U_k + \frac{1}{n}\sum_{j=1}^{n}\sum_{\substack{|\alpha|=k\\ \alpha\ni j}}\psi_{k-1}(\alpha \setminus \{j\}) \\
&= -\frac{k}{n}U_k + \frac{n-k+1}{n}\sum_{|\beta|=k-1}\psi_{k-1}(\beta) \\
&= -\frac{k}{n}U_k + \frac{n-k+1}{n}U_{k-1}, \qquad (6)
\end{aligned}$$

where the third equality follows from the observation that

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

and, thus, in the corresponding double sum of (6), every set β of size $k-1$ appears exactly $n-(k-1)$ times. Thus,

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

Hence, (1) is satisfied for $R = 0$ and

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

with lower triangular Λ^{-1} such that, if $l \le k$,

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

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

$$\lambda^{(l)} \le dn. \tag{7}$$

Now define $\eta_{j,k}(\alpha) := \psi'_{j,k}(\alpha) - \psi_k(\alpha)$. Then we have, for every $k, l = 1, \dots, d$,

$$\mathrm{E}^{X,X'}((U'_k - U_k)(U'_l - U_l)) = \frac{1}{n}\sum_{j=1}^{n}\Bigg(\sum_{\substack{|\alpha|=k,\, |\beta|=l \\ \alpha\cap\beta\ni j}} \eta_{j,k}(\alpha)\eta_{j,l}(\beta)\Bigg) \tag{8}$$

and

$$\begin{aligned} &\mathrm{E}(\mathrm{E}^{X,X'}((U'_k - U_k)(U'_l - U_l)))^2 \\ &\quad = \frac{1}{n^2}\sum_{i,j=1}^{n}\sum_{\substack{|\alpha|=k,\, |\beta|=l \\ \alpha\cap\beta\ni i}}\sum_{\substack{|\gamma|=k,\, |\delta|=l \\ \gamma\cap\delta\ni j}} \mathrm{E}(\eta_{i,k}(\alpha)\eta_{i,l}(\beta)\eta_{j,k}(\gamma)\eta_{j,l}(\delta)). \end{aligned} \tag{9}$$

Now note that, if the sets $\alpha \cup \beta$ and $\gamma \cup \delta$ are disjoint (which can only happen if $i \ne j$),

$$\mathrm{E}(\eta_{i,k}(\alpha)\eta_{i,k}(\beta)\eta_{j,l}(\gamma)\eta_{j,l}(\delta)) = \mathrm{E}(\eta_{i,k}(\alpha)\eta_{i,k}(\beta))\,\mathrm{E}(\eta_{j,l}(\gamma)\eta_{j,l}(\delta)),$$

due to independence. The variance of (8), that is (9) minus the square of the expectation of (8), contains only summands where $\alpha \cup \beta$ and $\gamma \cup \delta$ are not disjoint. Recall now that $\rho = \mathrm{E}\,\psi(X_1, \dots, X_d)^4$. Bounding all the nonvanishing terms simply by 32ρ, it only remains to count the number of nonvanishing terms. Thus,

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

where the equality is just a split of the sum over j into the cases $j \in \alpha \cup \beta$ and $j \notin \alpha \cup \beta$.

https://doi.org/10.1239/jap/1276784898 Published online by Cambridge University Press

In the former case we automatically have $(\alpha \cup \beta) \cap (\gamma \cup \delta) \neq \varnothing$. It is now not difficult to see that

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

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

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

we further have

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

where we have also used the fact that

$$\binom{n-|\alpha\cup\beta|}{k-1} \ge \binom{n-k-l+1}{k-1}.$$

The following statements are straightforward to prove:

$$\binom{n-1}{k-1}\binom{n}{k}^{-1} = \frac{k}{n}, \tag{10}$$

$$\binom{n-k-l+1}{k-1}\binom{n}{k}^{-1} \ge \frac{k}{n}\left(\frac{n-2k-l+3}{n}\right)^k \ge \frac{k}{n}\left(1 - \frac{k(2k+l-3)}{n}\right). \tag{11}$$

Thus, from (10),

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

From (10) and (11),

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

Thus, for all k and l,

$$\begin{aligned}\operatorname{var}\mathrm{E}^W(W_k'-W_k)(W_l'-W_l) &\le \operatorname{var}\mathrm{E}^{X,X'}(W_k'-W_k)(W_l'-W_l)\\ &\le \frac{256\rho d^6}{n^3}.\end{aligned} \tag{12}$$

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

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

https://doi.org/10.1239/jap/1276784898 Published online by Cambridge University Press

using (16), below; hence, along with (10),

$$\begin{aligned}
\mathrm{E}\,|(W_i' - W_i)(W_k' - W_k)(W_l' - W_l)| &\le \max_{m=i,k,l} \mathrm{E}\,|W_m' - W_m|^3 \\
&\le 8\rho^{3/4} n^{3/2} \max_{m=i,k,l} \binom{n}{m}^{-3} \binom{n-1}{m-1}^{3} \\
&\le 8\rho^{3/4} d^3 n^{-3/2}. \qquad (13)
\end{aligned}$$

Applying Theorem 1 with the estimates (7), (12), and (13) proves the claim.

Remark 3. Using the operator norm as used by Meckes [8], we would be able to achieve a bound of $n\log(d+1)$ instead of (7), but using bounds for the total derivatives of the test functions, $\sup_{x\in\mathbb{R}^k}\|D^r h(x)\|_{\mathrm{op}}$, instead of bounds for $|h|_r$.

5. Edge and triangle counts in Bernoulli random graphs

Typical summaries for random graphs are the degree distribution and the number of triangles, as a proxy for the clustering coefficient in a random graph, which is the expected ratio of the number of triangles over the number of 2-stars a randomly chosen vertex is involved with. Conditional uniform graph tests are based on fixing the degree distribution and randomising over the edges, conditional on keeping the degree distribution fixed. Our next example shows that even when fixing only the number of edges, not even the degree distribution, under a normal asymptotic regime, the number of triangles, or the number of 2-stars, is already asymptotically determined. Let $G(n, p)$ denote a Bernoulli random graph on n vertices, with edge probabilities p; we assume that $n \ge 4$ 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. Our interest is 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,

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

Here and in what follows, 'i, j, k distinct' is short for '$(i, j, k)\colon i \ne j \ne k \ne i$'; later we will also use 'i, j, k, ℓ distinct', which is the analogous abbreviation for four indices.

In view of the embedding method, we also include the auxiliary statistic related to the number of 2-stars,

$$V := \frac{1}{2}\sum_{i,j,k \text{ distinct}} I_{i,j}I_{j,k} = \sum_{i<j<k} (I_{i,j}I_{j,k} + I_{i,j}I_{i,k} + I_{j,k}I_{i,k}).$$

We note that

$$\mathrm{E}\,T = \binom{n}{2}p, \qquad \mathrm{E}\,V = 3\binom{n}{3}p^2, \quad \text{and} \quad \mathrm{E}\,U = \binom{n}{3}p^3.$$

https://doi.org/10.1239/jap/1276784898 Published online by Cambridge University Press

With some calculation, we find that the variances are not all of the same order. Hence, we rescale our variables (cf. [6]), setting

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

For these rescaled variables, the covariance matrix Σ_1 for $W_1 = (T_1 - \mathrm{E}\,T_1, V_1 - \mathrm{E}\,V_1, U_1 - \mathrm{E}\,U_1)$ equals

$$\Sigma_1 = \frac{3(n-2)}{n^4}\binom{n}{3}p(1-p)\begin{pmatrix} 1 & 2p & p^2 \\ 2p & 4p^2 + \dfrac{p(1-p)}{n-2} & 2p^3 + \dfrac{p^2(1-p)}{n-2} \\ p^2 & 2p^3 + \dfrac{p^2(1-p)}{n-2} & p^4 + \dfrac{p^2(1+p-2p^2)}{3(n-2)} \end{pmatrix}. \tag{14}$$

Remark 4. With $n \to \infty$, we obtain as the approximating covariance matrix

$$\Sigma_0 = \frac{1}{2}p(1-p)\begin{pmatrix} 1 & 2p & p^2 \\ 2p & 4p^2 & 2p^3 \\ p^2 & 2p^3 & p^4 \end{pmatrix}. \tag{15}$$

As also observed in [6], 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 1, will not be useful.

Janson and Nowicki [6] derived a normal limit for W_1, but no bounds on the approximation were given. Using Theorem 1, we obtain explicit bounds, as follows.

Proposition 2. *Let $W_1 = (T_1 - \mathrm{E}\,T_1, V_1 - \mathrm{E}\,V_1, U_1 - \mathrm{E}\,U_1)$ be the centralized count vector of the number of edges, two-stars, and triangles in a Bernoulli(p) random graph. Let Σ_1 be given as in (14). Then, for every three times differentiable function h,*

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

While we do not claim that the constants in the bound are sharp, 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. In this example, we were not able to obtain any improvement on the bounds using the operator bounds [8].

Proof of Proposition 2. The proof consists of two stages. Firstly, we construct an exchangeable pair; it will turn out that $R = 0$ in (1) and, hence, C in Theorem 1 will vanish. In the second stage we bound the terms A and B in Theorem 1.

Construction of an exchangeable pair. Our vector of interest is now $W = (T - \mathrm{E}\,T, V - \mathrm{E}\,V, U - \mathrm{E}\,U)$, rescaled to $W_1 = (T_1 - \mathrm{E}\,T_1, V_1 - \mathrm{E}\,V_1, U_1 - \mathrm{E}\,U_1)$. 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

$$\mathrm{P}(I = i,\ J = j) = \binom{n}{2}^{-1}, \qquad 1 \le i < j \le n.$$

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

$$\begin{aligned} T' &= T - (I_{I,J} - I'_{I,J}), \\ V' &= V - \sum_{\{k:\, k \neq I,J\}} (I_{I,J} - I'_{I,J})(I_{J,k} + I_{I,k}), \\ U' &= U - \sum_{\{k:\, k \neq I,J\}} (I_{I,J} - I'_{I,J}) I_{J,k} I_{I,k}. \end{aligned}$$

Set $W' = (T' - \mathrm{E}\, T, V' - \mathrm{E}\, V, U' - \mathrm{E}\, U)$. Then (W, W') forms an exchangeable pair. We rescale W' as W to obtain T'_1, V'_1, and U'_1, so that (W_1, W'_1) is also exchangeable.

Calculation of Λ. For the conditional expectations $\mathrm{E}^W(W' - W)$, firstly we have

$$\begin{aligned} \mathrm{E}^W(T'_1 - T_1) &= \frac{2(n-2)}{n^3(n-1)} \sum_{i<j} \mathrm{E}^W(I'_{i,j} - I_{i,j} \mid I = i,\ J = j) \\ &= \frac{n-2}{n^2} p - \frac{2(n-2)}{n^3(n-1)} T \\ &= -\binom{n}{2}^{-1} (T_1 - \mathrm{E}\, T_1). \end{aligned}$$

Furthermore,

$$\begin{aligned} -\mathrm{E}^W(V'_1 - V_1) &= \frac{1}{n^2}\binom{n}{2}^{-1} \sum_{i<j} \mathrm{E}^W \sum_{k:\, k \neq i,j} (I_{i,j} - I'_{i,j})(I_{j,k} + I_{i,k}) \\ &= 2\frac{1}{n^2}\binom{n}{2}^{-1} V - 2p\frac{1}{n^2}\binom{n}{2}^{-1}(n-2)T \\ &= -2\binom{n}{2}^{-1}(V_1 - \mathrm{E}\, V_1) + 2p\binom{n}{2}^{-1}(T_1 - \mathrm{E}\, T_1), \end{aligned}$$

where the last equality follows from $\mathrm{E}(V'_1 - V_1) = 0$. Similarly,

$$-\mathrm{E}^W(U'_1 - U_1) = -3\binom{n}{2}^{-1}(U_1 - \mathrm{E}\, U_1) + p\binom{n}{2}^{-1}(V_1 - \mathrm{E}\, V_1).$$

Using our rescaling, (1) is satisfied with $R = 0$ and Λ given by

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

Bounding A. The inverse matrix Λ^{-1} is easy to calculate; for simplicity, we will apply the uniform bound for $\lambda^{(i)} = \sum_{m=1}^{d} |(\Lambda^{-1})_{m,i}|$,

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

The bounding of the conditional variances is somewhat laborious. The conditional variances involving $T' - T$ can be calculated exactly. As $I_{i,j}^2 = I_{i,j}$,

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

so that with $\operatorname{var} T$ given through (14), $\operatorname{var}(\mathrm{E}^W(T'-T)^2) = \binom{n}{2}^{-1}(1-2p)^2 p(1-p)$ and

$$\operatorname{var}(\mathrm{E}^W(T_1'-T_1)^2) = \frac{(n-2)^4}{n^8}\binom{n}{2}^{-1}(1-2p)^2 p(1-p) < n^{-6},$$

where we have used the fact that $p(1-p) \le \frac{1}{4}$ for all p. Thus,

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

Similarly,

$$\mathrm{E}^W(T'-T)(U'-U) = \binom{n}{2}^{-1}(pV + 3(1-2p)U);$$

hence,

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

Straightforward calculations show that

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

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

$$\begin{aligned}&\operatorname{var}\mathrm{E}^W(V'-V)^2\\ &\quad\le 5\binom{n}{2}^{-2}\bigg\{p^2(2n-8+4pn^2-4pn)^2\operatorname{var}(T) + 4(np-10p+2)^2\operatorname{var}(V)\\ &\qquad + 36(1-2p)^2\operatorname{var}(U) + 16p^2\operatorname{var}(\tilde{T}^2)\\ &\qquad + (1-2p)^2\operatorname{var}\bigg(\sum_{i,j,k,\ell\text{ distinct}}\mathrm{E}^W I_{i,j}I_{i,k}(I_{i,\ell}+I_{j,\ell})\bigg)\bigg\},\end{aligned}$$

where we have used the fact that in general $\operatorname{var}\sum_{i=1}^k X_i \le k\sum_{i=1}^k \operatorname{var} X_i$. Here, the variances for T, V, and U are given through (14).

We have the bounds

$$p^2(2n - 8 + 4pn^2 - 4pn)^2 \operatorname{var}(T) \le \frac{27}{64}\binom{n}{2} n^2(n+2)^2,$$
$$4(np - 10p + 2)^2 \operatorname{var}(V) \le \frac{16}{27} n^3(n-1)(n-2)(n+1),$$
$$36(1-2p)^2 \operatorname{var}(U) \le \frac{81}{256} n(n-1)(n-2)(3n+2),$$
$$16p^2 \operatorname{var} \tilde{T}^2 \le \frac{27}{32} n^3(n-1),$$
$$\operatorname{var} \sum_{i \ne j} \sum_{\{k:\, k \ne i,j\}} \sum_{\{\ell:\, \ell \ne i,j,k\}} \mathrm{E}^W\, I_{i,j} I_{i,k}(I_{i,\ell} + I_{j,\ell}) \le 3n^2 \binom{n}{4}.$$

We also have

$$\mathrm{E}^W(V' - V)(U' - U) = \binom{n}{2}^{-1} \Bigg(2pV + 6(1-2p)U + p \sum_{i,j,k,\ell \text{ distinct}} I_{i,k} I_{k,j} I_{i,\ell} + (1-2p) \sum_{i,j,k,\ell \text{ distinct}} I_{i,j} I_{i,k} I_{i,\ell} I_{\ell,j}\Bigg).$$

Now,

$$\operatorname{var} \sum_{i,j,k,\ell \text{ distinct}} I_{i,k} I_{k,j} I_{i,\ell} < \binom{n}{4} n^2$$

and

$$\operatorname{var} \sum_{i,j,k,\ell \text{ distinct}} I_{i,j} I_{i,k} I_{i,\ell} I_{j,\ell} < \binom{n}{4}\left(\frac{1}{256} + \frac{1}{16}\binom{n}{2}\right),$$

so that

$$\sqrt{\operatorname{var} \mathrm{E}^W(V_1' - V_1)(U_1' - U_1)} < n^{-3} + 11n^{-4}.$$

Finally,

$$\mathrm{E}^W(U' - U)^2 = \frac{1}{2}\binom{n}{2}^{-1} \Bigg\{2pV + 6(1-2p)U + p \sum_{i,j,k,\ell \text{ distinct}} \mathrm{E}^W\, I_{i,k} I_{k,j} I_{i,\ell} I_{\ell,j} + (1-2p) \sum_{i,j,k,\ell \text{ distinct}} \mathrm{E}^W\, I_{i,j} I_{i,k} I_{k,j} I_{i,\ell} I_{\ell,j}\Bigg\}.$$

With

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

and

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

https://doi.org/10.1239/jap/1276784898 Published online by Cambridge University Press

we obtain

$$\sqrt{\operatorname{var}(\mathrm{E}^W(U_1'-U_1)^2)} < 5n^{-3} + 2n^{-4}.$$

Collecting these bounds we obtain

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

Bounding B. We use the generalized Hölder inequality

$$\mathrm{E}\prod_{i=1}^{3}|X_i| \le \prod_{i=1}^{3}\{\mathrm{E}\,|X_i|^3\}^{1/3} \le \max_{i=1,2,3}\mathrm{E}\,|X_i|^3. \tag{16}$$

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

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

Similarly,

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

so that

$$\mathrm{E}\,|V_1'-V_1|^3 < \tfrac{64}{27}(n^{-3}+n^{-4}+n^{-5}).$$

Lastly,

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

so that

$$\mathrm{E}\,|U_1'-U_1|^3 < \tfrac{54}{256}(n^{-3}+n^{-4}+n^{-5}).$$

Thus, for B, we have

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

Collecting the bounds gives the result.

Remark 5. Had we not introduced V, conditioning would yield

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

The expression $\sum_{i<j,\,k\neq i,j}\mathrm{E}^{T,U}\,I_{j,k}I_{j,k}$ would result in a nonlinear remainder term R in (1). The introduction of V not only avoids this remainder term, indeed $R=0$ in (1), but also yields a more detailed result. This observation that the 2-stars form a useful auxiliary statistic can also be found in [6]; there it is related to Hoeffding-type projections.

https://doi.org/10.1239/jap/1276784898 Published online by Cambridge University Press

Using Proposition 1, we also obtain a normal approximation for Σ_0 given in (15).

Corollary 2. *Under the assumptions of Proposition 2, for every three times differentiable function h,*

$$|\mathrm{E}h(W) - \mathrm{E}\, h(\Sigma_0^{1/2} Z)| \le \frac{|h|_2}{2n}(44 + 21n^{-1} + 32n^{-2} + 4n^{-3}) + \frac{8|h|_3}{3n}(1 + n^{-1} + n^{-2}).$$

Acknowledgement

The authors would like to thank the anonymous referee for helpful comments on the manuscript.

References

[1] Bolthausen, E. (1982). Exact convergence rates in some martingale central limit theorems. *Ann. Prob.* **10,** 672–688.

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

[3] Chatterjee, S. (2006). A generalization of the Lindeberg principle. *Ann. Prob.* **34,** 2061–2076.

[4] Chatterjee, S. and Meckes, E. (2008). Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Amer. J. Prob. Math. Statist.* **4,** 257–283.

[5] Janson, S. and Luczak, M. J. (2008). Susceptibility in subcritical random graphs. *J. Math. Phys.* **49,** 125207, 23pp.

[6] Janson, S. and Nowicki, K. (1991). The asymptotic distributions of generalized U-statistics with applications to random graphs. *Prob. Theory Relat. Fields* **90,** 341–375.

[7] Lee, A. J. (1990). *U-Statistics* (Statist. Text. Monogr. **110**). Marcel Dekker, New York.

[8] Meckes, E. (2009). On Stein's method for multivariate normal approximation. In *High Dimensional Probability V: The Luminy Volume*, eds C. Houdré *et al.*, Institute of Mathematical Statistics, Beachwood, OH, pp. 153–178.

[9] Nourdin, I., Peccati, G. and Reinert, G. (2009). Stein's method and stochastic analysis of Rademacher functionals. Preprint. Available at http://arxiv.org/abs/0810.2890v3.

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

[11] Reinert, G. and Röllin, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Prob.* **37,** 2150–2173.

[12] Rinott, Y. and Rotar′, V. (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.

[13] Rinott, Y. and Rotar′, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Prob.* **7,** 1080–1105.

[14] Rinott, I. and Rotar′, V. I. (1998). Some estimates for the rate of convergence in the CLT for martingales. I. *Theory Prob. Appl.* **43,** 604–619.

[15] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, Vol. II, University California Press, pp. 583–602.

[16] Stein, C. (1986). *Approximate Computation of Expectations* (Inst. Math. Statist. Lecture Notes Monogr. Ser. **7**). Institute of Mathematical Statistics, Hayward, CA.