

Rates of convergence for multivariate normal approximation with applications to dense graphs and doubly indexed permutation statistics

XIAO FANG¹ and ADRIAN RÖLLIN²

¹*Department of Statistics, Stanford University, Stanford, CA 94305-4065, USA.*

E-mail: fangxiao@stanford.edu

²*Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546. E-mail: adrian.roellin@nus.edu.sg*

We provide a new general theorem for multivariate normal approximation on convex sets. The theorem is formulated in terms of a multivariate extension of Stein couplings. We apply the results to a homogeneity test in dense random graphs and to prove multivariate asymptotic normality for certain doubly indexed permutation statistics.

Keywords: dense graph limits; multivariate normal approximation; non-smooth metrics; permutation statistics; random graphs; Stein’s method

1. Introduction

Let W and Z be d -dimensional random vectors, $d \geq 1$, where Z has standard d -dimensional Gaussian distribution. We are concerned with bounding the quantity

$$d_c(\mathcal{L}(W), \mathcal{L}(Z)) = \sup_{A \in \mathcal{A}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)|, \quad (1.1)$$

where \mathcal{A} denotes the collection of all the convex sets in \mathbb{R}^d .

Our main tool is Stein’s method for the multivariate normal distribution, which has already been used to obtain bounds on (1.1), the two main contributions coming from Götze [18] for sums of independent random vectors (see also Bhattacharya and Holmes [6]), and Rinott and Rotar’ [25] for sums of dependent random vectors that allow for a certain decomposition. Most other contributions on multivariate normal approximation via Stein’s method have focused on smooth functions; see, for example, Barbour [3], Goldstein and Rinott [17], Raič [23] and Reinert and Röllin [24].

The main aim of this article is to improve the results of Rinott and Rotar’ [25] in two important ways. First, we remove a logarithmic factor in the error bound of Rinott and Rotar’ [25]. The techniques that allow us to do this are taken from Fang [15] and will yield optimal rates of convergence in some applications. Second, the assumptions made on the dependence by Rinott and Rotar’ [25] do not cover the applications we will discuss here. Instead, we will use a mul-

tivariate generalisation of Stein couplings to achieve the necessary generality. Stein couplings, introduced by Chen and Röllin [11], capture the minimal structural assumption necessary to use Stein’s method for normal approximation.

We will also keep the dependence of the constants on the dimensionality explicit and as small as possible without blowing up the proofs, but we do not pursue optimality in that respect.

The remainder of this article is organised as follows. In Section 2, we will state our main abstract theorem, but we will postpone the (rather technical) proof to Section 4. In Section 3, we will discuss two main applications, one involving permutation statistics and the other a new test for heterogeneity for dense graphs. In Section 5, we will present some standard multivariate Stein couplings for reference.

2. Main results

Stein couplings were introduced by Chen and Röllin [11] in order to unify many of the approaches developed around Stein’s method for normal approximation, such as local approach, size biasing and exchangeable pairs, to name but a few. In the spirit of Chen and Röllin [11], we give a multivariate definition of Stein couplings.

Definition 2.1. *A triple of square integrable d -dimensional random vectors (W, W', G) is called a d -dimensional Stein coupling if*

$$\mathbb{E}\{G^t F(W') - G^t F(W)\} = \mathbb{E}\{W^t F(W)\} \tag{2.1}$$

for all $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$ for which the expectations exist.

Remark 2.2. By choosing $F(w) = e_i$, where e_i is the i th unit vector, it follows from (2.1) that $\mathbb{E}W_i = 0$. Therefore, $\mathbb{E}W = 0$ is a necessary condition for a Stein coupling. Choosing $F(w) = w_j e_i$, it follows that

$$\mathbb{E}\{G(W' - W)^t\} = \text{Cov}(W). \tag{2.2}$$

Throughout this article, $|x|$ denotes the Euclidean norm of $x \in \mathbb{R}^d$, and \mathbb{I}_d denotes the d -dimensional identity matrix. To shorten the formulas somewhat, we will write $\mathbb{E}^W(\cdot)$ to denote conditional expectation $\mathbb{E}(\cdot|W)$.

With this, we can formulate our main result.

Theorem 2.1. *Let (W, W', G) be a d -dimensional Stein coupling. Assume that $\text{Cov}(W) = \mathbb{I}_d$. With $D = W' - W$, suppose that there are positive constants α and β such that*

$$|G| \leq \alpha, \quad |D| \leq \beta. \tag{2.3}$$

Then there is a universal constant C such that

$$\begin{aligned} & d_c(\mathcal{L}(W), \mathcal{L}(Z)) \\ & \leq C(d^{7/4}\alpha\mathbb{E}|D|^2 + d^{1/4}\beta + d^{7/8}\alpha^{1/2}B_1^{1/2} + d^{3/8}B_2 + d^{1/8}B_3^{1/2}), \end{aligned} \tag{2.4}$$

where Z is a d -dimensional standard Gaussian random vector and

$$B_1 = \sqrt{\text{Var } \mathbb{E}^W |D|^2}, \quad B_2 = \sqrt{\sum_{i,j=1}^d \text{Var } \mathbb{E}^W (G_i D_j)},$$

$$B_3 = \sqrt{\sum_{i,j,k=1}^d \text{Var } \mathbb{E}^W (G_i D_j D_k)}.$$

As usual, we can upper bound $\text{Var } \mathbb{E}^W (\cdot)$ by $\text{Var } \mathbb{E}^{\mathcal{F}} (\cdot)$ for any σ -algebra $\mathcal{F} \supset \sigma(W)$. This is a standard trick in Stein’s method and will be used in the applications without further mention.

Note that, if (W, W', G) is a d -dimensional Stein coupling and A is a $m \times d$ matrix, $m \geq 1$, then (AW, AW', AG) is an m -dimensional Stein coupling. In this light, assuming that $\text{Cov}(W) = \mathbb{I}_d$ is a matter of convenience rather than a real restriction. If A is a $d \times d$ matrix, denote by $\|A\|_2$ its operator norm with respect to the Euclidean norm. Noticing that d_c is invariant under linear transformations, we have the following consequence of Theorem 2.1.

Corollary 2.2. *Under the conditions of Theorem 2.1, but now allowing $\text{Cov}(W) = \Sigma$ for any positive definite Σ , there is a universal constant C such that*

$$\begin{aligned} & d_c(\mathcal{L}(\Sigma^{-1/2}W), \mathcal{L}(Z)) \\ &= d_c(\mathcal{L}(W), \mathcal{L}(\Sigma^{1/2}Z)) \\ &\leq C(d^{7/4} \alpha s_2^3 \mathbb{E}|D|^2 + d^{1/4} s_2 \beta + d^{7/8} s_2^{3/2} \alpha^{1/2} B_1^{1/2} + d^{3/8} s_2^2 B_2 + d^{1/8} s_2^{3/2} B_3^{1/2}), \end{aligned} \tag{2.5}$$

where $s_2 = \|\Sigma^{-1/2}\|_2$.

Note that the corollary cannot be expected to be informative if Σ is singular or close to singular. In particular, the W_i need to be standardized so that $\text{Var } W_i, 1 \leq i \leq d$, are all of the same order. The proof of Corollary 2.2 is given in Section 4.

Remark 2.3. If (W, W') is an exchangeable pair of d -dimensional vectors and

$$\mathbb{E}^W (W' - W) = -\Lambda W \tag{2.6}$$

for some invertible $d \times d$ -matrix Λ , then $(W, W', \frac{1}{2}\Lambda^{-1}(W' - W))$ is a Stein coupling and Theorem 2.1 can be applied. In the special case where $\Lambda = \lambda \mathbb{I}_d$, or in other words, if we have

$$\mathbb{E}^W (W' - W) = -\lambda W \tag{2.7}$$

for some $0 < \lambda < 1$, then one can prove a special case of Theorem 2.1 without using exchangeability, but only assuming that $\mathcal{L}(W) = \mathcal{L}(W')$. A sketch of the proof will be given in Section 4. This is analogous to Reinert and Röllin [24], where a result similar to our Theorem 2.1 was obtained for the special case of (2.6), but for a smooth metric, and where also exchangeability was relaxed to equal marginals in the special case of (2.7).

3. Applications

3.1. A confidence interval for dense homogeneous random graphs

One of the basic problems in the statistical analysis of graphs is to test whether the connections between vertices in a graph have arisen ‘completely at random’, or whether there is more structure in the graph. Among several possible null hypotheses, one of the best-studied is the Erdős–Rényi random graph $G(n, p)$, where two vertices are connected with probability p and remain disconnected with probability $1 - p$, independently of all else.

Many test statistics have been analysed in the literature, such as diameter, maximal degree, number of triangles, etc.; see, for example, Pao, Coppersmith and Priebe [22] for a recent overview and simulation studies of the performance of these and other test statistics. Despite the fact that much is known about the behaviour of these test statistics under the null model $G(n, p)$, it seems that little is known, at least theoretically, about how these statistics behave under alternative models, such as heterogeneous models, where the edge probabilities may vary. Here, as a first step, we propose and justify a test that is based on the theory of dense graph limits, and we will show that our test is *consistent*, that is, any deviation from the homogeneous model will eventually be detected (in a sense made precise below).

Theory of dense graph limits

Before we start with the statistical aspect of the problem, we first give a brief introduction to the theory of dense graph limits. We will only discuss those parts of the theory that are necessary for the purpose of our application; we refer to Borgs, Chayes, Lovász, Sós and Vesztergombi [8,9] and Bollobás and Riordan [7] for in-depth discussions. Also, dense graph limit theory is intimately related to the theory of partially exchangeable arrays as studied by Aldous [1]; see Diaconis and Janson [14].

In what follows, all graphs are assumed to be simple, that is, graphs that contain no loops and no multiple edges, and, moreover, we assume that all graphs are undirected. To begin with, we consider non-random graphs. Let F and G be graphs with k , respectively, n vertices. Denote by $\text{inj}(F, G)$ the set of injective graph homomorphisms from F into G , and define

$$t(F, G) = \frac{|\text{inj}(F, G)|}{(n)_k},$$

where $(n)_k := n(n-1) \cdots (n-k+1)$ (note that we follow the notation of Bollobás and Riordan [7]; our t is what Borgs, Chayes, Lovász, Sós and Vesztergombi [8] denote by t_{inj}). The number

$|\text{inj}(F, G)|$ is just the number of copies of F in G multiplied by the number of graph automorphisms of F . Since $|\text{inj}(F, G)| \leq (n)_k$, it is clear that $0 \leq t(F, G) \leq 1$, and we can think of this value as the “density of F in G ”.

Let (G_n) be a sequence of graphs (where $n \geq n_0$ for some unspecified n_0) and for convenience assume that G_n has n vertices. We call this sequence a *dense graph sequence* if the number of edges is of order n^2 . In other words, if K_m denotes the complete graph on m vertices, a graph sequence (G_n) is called dense if $\liminf_{n \rightarrow \infty} t(K_2, G_n) > 0$, and we will in fact mostly consider sequences for which $\lim_{n \rightarrow \infty} t(K_2, G_n)$ exists. Although it would not pose any difficulties to allow the case $\lim_{n \rightarrow \infty} t(K_2, G_n) = 0$ (the “sparse” case), this only leads to degenerate results in the context of dense graph theory, and is therefore excluded for the sake of clarity.

We say that a dense graph sequence (G_n) is convergent if $\lim_{n \rightarrow \infty} t(F, G_n)$ exists for every finite graph F . We can construct a metric d on the set of isomorphism classes of finite graphs, denoted by \mathcal{F} , that quantifies this convergence. Let F_1, F_2, \dots be an arbitrary enumeration of the set of finite graphs. For two graphs G_1 and G_2 , let

$$d(G_1, G_2) = \sum_{i \geq 1} 2^{-i} |t(F_i, G_1) - t(F_i, G_2)|.$$

It turns out that the metric space (\mathcal{F}, d) is not complete. The usual way of constructing the completion of a metric space is to form equivalence classes of sequences that are Cauchy with respect to the metric. However, it turns out that there is a much more natural representation.

Let $\kappa : [0, 1]^2 \rightarrow [0, 1]$ be a measurable and symmetric function; we will call any such function a *standard kernel* (called *graphon* by Borgs, Chayes, Lovász, Sós and Vesztergombi [8]). For any finite graph F with k vertices, let

$$t(F, \kappa) = \int_0^1 \cdots \int_0^1 \prod_{\{i,j\} \in E(F)} \kappa(x_i, x_j) dx_1 \cdots dx_k,$$

where $E(F)$ denotes the edge set of graph F . The quantity $t(F, \kappa)$ can be interpreted the “density of F in κ ”, and we will give a more intuitive representation of $t(F, \kappa)$ involving random graphs later.

One of the key results of dense graph theory (see, for example, Borgs, Chayes, Lovász, Sós and Vesztergombi [8], Theorem 3.1) is the following. If $t(F, G_n)$ converges for every F , that is, if (G_n) is a Cauchy sequence with respect to d , then there is a standard kernel κ such that $\lim t(F, G_n) = t(F, \kappa)$ for every F . We can therefore say that κ is a *limit of the graph sequence* (G_n) . Analogous to the fact that there are graphs that are isomorphic to each other, there can (and typically will) be several standard kernels representing the same limit. Therefore, an additional step of forming equivalence classes of standard kernels is necessary to obtain the actual completion of the metric space (\mathcal{F}, d) . Since we do not need this we refer again to Borgs, Chayes, Lovász, Sós and Vesztergombi [8,9] on how to characterise these equivalence classes.

So far, all graphs have been non-random. If now (G_n) is a sequence of random graphs defined on a common probability space Ω , we will be interested in statements of the form “ (G_n)

converges to κ almost surely”, meaning that with probability 1, the realisation of a sequence $G_1(\omega), G_2(\omega), \dots$ converges to κ in the sense introduced above. Although it is possible to allow for κ to be random as well, we will only consider fixed κ in what follows.

For a given standard kernel κ , there is an elegant sampling procedure to create random graphs that converge to κ almost surely. Let U_1, U_2, \dots be a sequence of independent random variables that are uniformly distributed on the interval $[0, 1]$. To construct G_n , connect vertices i and j with probability $\kappa(U_i, U_j)$, independently of all other edges. We denote the distribution of the graph G_n obtained in this way by $G(n, \kappa)$ and it is clear that $G(n, p)$ for $0 \leq p \leq 1$ can be identified with $G(n, \kappa)$ for $\kappa \equiv p$, the constant standard kernel. Note that the edges of $G(n, \kappa)$ are conditionally independent given U_1, \dots, U_n , but in general not unconditionally independent. It is now easy to verify that, if $G_n \sim G(n, \kappa)$, then

$$\mathbb{E}t(F, G_n) = t(F, \kappa).$$

Furthermore, we have the following concentration result, which, by Borell–Cantelli, immediately implies that (G_n) converges to κ almost surely.

Lemma 3.1 (Borgs, Chayes, Lovász, Sós and Vesztegombi [8], Lemma 4.4). *If $G_n \sim G(n, \kappa)$ for some standard kernel κ , and if F is a graph on k vertices, then*

$$\mathbb{P}\left[|t(F, G_n) - t(F, \kappa)| > \varepsilon\right] \leq \exp\left(-\frac{\varepsilon^2 n}{4k^2}\right)$$

for every $\varepsilon > 0$.

Remark 3.1. A remark about models that are more general than $G(n, \kappa)$ is in place. It is important to note that dense graph theory is a first order approximation of dense graphs, analogous to the law of large number for random variables. It can be shown that the completion of (\mathcal{F}, d) is compact and therefore, for any dense graph sequence, there must be accumulation points which can be represented by a set \mathcal{K} of standard kernels. So, if one considers graph models that produce dense graphs that allow for more complex dependence between edges, any realisation of a large enough graph from such a model will be close to at least one of the standard kernels from its accumulation points \mathcal{K} . Thus, from this first order point of view, any dependence between the edges becomes irrelevant in the limit, since every $\kappa \in \mathcal{K}$ is also the limit of the model $G(n, \kappa)$. As of yet, there seems to be no established theory of second order fluctuations of dense graphs around their limits that would capture more subtle aspects of such graphs.

Characterisation of homogenous Erdős–Rényi graphs

Recall that K_m denotes the complete graph of size m , and let C_m be the cycle graph of size m . Chung, Graham and Wilson [12] proved the following surprising result, which we shall present reformulated in the language of dense graph limit theory (see Lovász and Szegedy [21] for generalisations of these findings).

Theorem 3.2 (Chung, Graham and Wilson [12], Theorem 1). *If (G_n) is a (non-random) dense graph sequence such that*

$$t(K_2, G_n) \rightarrow p \quad \text{and} \quad t(C_4, G_n) \rightarrow p^4$$

for some $0 < p \leq 1$, then (G_n) converges and the limit is the constant standard kernel $\kappa \equiv p$.

In other words, $\kappa \equiv p$ is the only standard kernel with $t(K_2, \kappa) = p$ and $t(C_4, \kappa) = p^4$, and it is not difficult to show that, if κ is not constant and $t(K_2, \kappa) = p$, then $t(C_4, \kappa) > p^4$. This result suggests that we can use the number of edges and 4-cycles in order to test whether κ is constant or not. Indeed, for $G_n \sim G(n, \kappa)$ with non-constant κ , we should be able to detect a discrepancy between the edge density to the fourth power and 4-cycle density if n is large enough.

However, some care is needed. If G_n is a given graph of size n , define the two statistics

$$T_1(G_n) = \frac{|\text{inj}(K_2, G_n)|}{2}, \quad T_2(G_n) = \frac{|\text{inj}(C_4, G_n)|}{8}.$$

The factors 2 and 8, respectively, are the sizes of the automorphism groups of K_2 and C_4 , respectively. Therefore, T_1 is the number of edges in G_n and T_2 is the number of 4-cycles in G_n . By straightforward calculations we have that, if $G_n \sim G(n, p)$,

$$\text{Var}(T_1(G_n)) = \binom{n}{2} p(1-p), \quad \text{Cov}(T_1(G_n), T_2(G_n)) = 12 \binom{n}{4} p^4(1-p)$$

and

$$\text{Var}(T_2(G_n)) = 3 \binom{n}{4} p^4(1-p)(1+p-13p^2+4np^2+35p^3-24np^3+4n^2p^3).$$

It is clear from this that $\text{Cor}(T_1(G_n), T_2(G_n)) \rightarrow 1$ as $n \rightarrow \infty$, hence, in the limit, the fluctuation of the number of 4-cycles is determined by that of the number of edges in the graph; see Janson and Nowicki [20] for such and more general results. Thus, we cannot use these values directly to construct our test.

Following Janson and Nowicki [20], we can instead consider the density of 4-cycles *corrected* by the edge density (this is essentially the first non-leading term in a Hoeffding-type decomposition for the 4-cycle count). To this end, define the normalised edge count

$$W_1(p, G_n) = \frac{T_1(G_n) - \binom{n}{2} p}{\sigma_1}, \quad \text{with } \sigma_1^2 = \binom{n}{2} p(1-p),$$

and the *corrected* and normalised 4-cycle count

$$W_2(p, G_n) = \frac{T_2(G_n) - 2 \binom{n-2}{2} p^3 T_1(G_n) + 9 \binom{n}{4} p^4}{\sigma_2}$$

with

$$\sigma_2^2 = 3 \binom{n}{4} p^4(1-p)^2(1+2p+(4n-11)p^2);$$

it is easy to see that $\text{Cov}(W_1, W_2) = 0$. In order to motivate the choice of W_1 and W_2 , note that, from Lemma 3.1 and for general κ and $G_n \sim G(n, \kappa)$,

$$\begin{aligned} \frac{W_1(p, G_n)}{n} &\rightarrow \frac{1}{\sqrt{2p(1-p)}} (t(K_2, \kappa) - p), \\ \frac{W_2(p, G_n)}{n^{3/2}} &\rightarrow \frac{1}{4\sqrt{2}p^3(1-p)} (t(C_4, \kappa) - 4p^3t(K_2, \kappa) + 3p^2) \end{aligned}$$

almost surely as $n \rightarrow \infty$, so that $W_1(p, G_n)$ and $W_2(p, G_n)$ can only be expected to be near zero if $\kappa \equiv p$.

Barbour, Karoński and Ruciński [4] use Stein’s method to prove univariate normal approximations of subgraph counts and related statistics, but for quantities such as W_2 they resort to the method of moments. Corresponding multivariate results were obtained by Janson and Nowicki [20] in great generality for incomplete U -statistics using Hoeffding-type decompositions and the methods of moments. For degenerate statistics like W_2 they state that “Stein’s method does not seem to work in that case”.

The reason that W_2 is more difficult to handle is that, if represented as an incomplete U -statistic, many of the summands are uncorrelated (see (3.5) below), which requires more delicate estimates. We note that the arguments of Barbour, Karoński and Ruciński [4] could be, in fact, improved to cover such cases as well.

Theorem 3.3. *Let $G_n \sim G(n, p)$ be a realisation of an Erdős–Rényi random graph on n vertices with edge probability p . Let $W = (W_1(p, G_n), W_2(p, G_n))$ and let Z be a standard bi-variate normal random variable. There is a universal constant C independent of p and n such that*

$$d_c(\mathcal{L}(W), \mathcal{L}(Z)) \leq \frac{C}{p^9(1-p)^3\sqrt{n}}.$$

Theorem 3.3 justifies the following procedure to construct a confidence set for the family of Erdős–Rényi random graphs. Let G_n be a simple graph of size n . Fix $0 < \alpha < 1$ and define the $1 - \alpha$ confidence set as

$$\text{CS}_{1-\alpha}(G_n) = \{0 < p < 1 : W_1^2(p, G_n) + W_2^2(p, G_n) \leq q_{1-\alpha}\},$$

where $q_{1-\alpha}$ is the $1 - \alpha$ quantile of the χ^2 -distribution with 2 degrees of freedom. In words, $\text{CS}_{1-\alpha}(G_n)$ is the set of those p for which G_n is “compatible” with the model $G(n, p)$ at the significance level α . If $\text{CS}_{1-\alpha}(G_n)$ is empty, then G_n is not compatible with any homogeneous Erdős–Rényi random graph model.

For what follows, denote by \mathbb{P}_κ the distribution of G_n under the law $G(n, \kappa)$ and let $\mathbb{P}_p = \mathbb{P}_\kappa$ for $\kappa \equiv p$.

Corollary 3.4. *For any given $0 < p_l < p_u < 1$,*

$$\mathbb{P}_p[p \notin \text{CS}_{1-\alpha}(G_n)] = \alpha + O(n^{-1/2})$$

uniformly in $p \in [p_l, p_u]$ as $n \rightarrow \infty$. Furthermore, if κ is a non-constant standard kernel, and if $n \geq \max\{19, 54q_{1-\alpha}^{1/2}/r_\kappa\}$, we have

$$\mathbb{P}_\kappa[\text{CS}_{1-\alpha}(G_n) \neq \emptyset] \leq 2 \exp\left(-\frac{nr_\kappa^2}{144}\right), \tag{3.1}$$

where

$$r_\kappa^2 = \inf_{0 < p < 1} \left\{ (t(K_2, \kappa) - t(K_2, p))^2 + (t(C_4, \kappa) - t(C_4, p))^2 \right\}$$

(note that $r_\kappa > 0$ from Theorem 3.2 and the discussion thereafter).

Proof. The first part is immediate from Theorem 3.3. For the second part assume that κ is not constant. Consider the points $b(\kappa) = (t(K_2, \kappa), t(C_4, \kappa))$ and $b(p) = (p, p^4)$, and, by slight abuse of notation, $b(n) = (t(K_2, G_n), t(C_4, G_n))$. Using Lemma 3.1, we have

$$\begin{aligned} & \mathbb{P}_\kappa[|b(n) - b(\kappa)| > \varepsilon] \\ & \leq \mathbb{P}_\kappa[|b_1(n) - b_1(\kappa)| > \varepsilon/\sqrt{2}] + \mathbb{P}_\kappa[|b_2(n) - b_2(\kappa)| > \varepsilon/\sqrt{2}] \\ & \leq 2 \exp(-\varepsilon^2 n / 128) \end{aligned} \tag{3.2}$$

for any $\varepsilon > 0$. Now, note that we can write

$$W_1(p, G_n) = \frac{\binom{n}{2}}{\sigma_1} (b_1(n) - b_1(p)), \quad W_2(p, G_n) = \frac{3\binom{n}{4}}{\sigma_2} w(n, p),$$

where

$$w(n, p) = (b_2(n) - b_2(p)) - 4p^3(b_1(n) - b_1(p)).$$

Let

$$\delta = \frac{\sqrt{67} - 4}{306} r_\kappa$$

and define the events

$$\begin{aligned} A_1(p) &= \{|b(n) - b(\kappa)| \leq 8r_\kappa/9, |b_1(n) - b_1(p)| > \delta\}, \\ A_2(p) &= \{|b(n) - b(\kappa)| \leq 8r_\kappa/9, |b_1(n) - b_1(p)| \leq \delta\}. \end{aligned}$$

On one hand, we have

$$W_1(p, G_n)^2 \geq \mathbb{I}[A_1(p)] \left(\frac{\binom{n}{2}\delta}{\sigma_1}\right)^2. \tag{3.3}$$

On the other hand, since $|b(n) - b(\kappa)| \leq 8r_\kappa/9$ implies $|b(n) - b(p)| \geq r_\kappa/9$, we have

$$w(n, p) > \sqrt{(r_\kappa/9)^2 - \delta^2} - 4\delta = \frac{r_\kappa}{18}$$

on $A_2(p)$, and hence

$$W_2(p, G_n)^2 \geq \mathbb{I}[A_2(p)] \left(\frac{3 \binom{n}{4} r_\kappa}{18\sigma_2} \right)^2. \tag{3.4}$$

Setting $A(p) = A_1(p) \cup A_2(p)$ and putting (3.3) and (3.4) together, we obtain

$$\begin{aligned} &W_1(p, G_n)^2 + W_2(p, G_n)^2 \\ &\geq \mathbb{I}[A(p)] r_\kappa^2 \min \left\{ \left(\frac{\binom{n}{4}}{6\sigma_2} \right)^2, \left(\frac{(\sqrt{67} - 4) \binom{n}{2}}{306\sigma_1} \right)^2 \right\} \\ &\geq \mathbb{I}[A(p)] r_\kappa^2 \min \{ 4.3 \cdot 10^{-3} (n-1)_3, 3.7 \cdot 10^{-4} (n)_2 \}. \end{aligned}$$

If $n \geq 19$, we have

$$\min \{ 4.3 \cdot 10^{-3} (n-1)_3, 3.7 \cdot 10^{-4} (n)_2 \} \geq 3.5 \cdot 10^{-4} n^2.$$

Hence, if $n \geq \max \{ 19, (q_{1-\alpha} / (3.5 \cdot 10^{-4} r_\kappa^2))^{1/2} \}$, and using (3.2),

$$\mathbb{P}_\kappa[\text{CS}_{1-\alpha} = \emptyset] \geq \mathbb{P}_\kappa[|b(n) - b(k)| \leq 8r_\kappa/9] \geq 1 - 2 \exp(-nr_\kappa^2/144),$$

which implies (3.1). □

Remark 3.2. Note that (3.1) essentially says that, if the true standard kernel κ is non-constant, the test will eventually detect this for n large enough. It is not clear if this is still true if 4-cycles were to be replaced by triangles. Chung, Graham and Wilson [12], page 361, give an example of non-constant standard kernel κ for which

$$t(K_2, \kappa) = \frac{1}{2}, \quad t(C_3, \kappa) = \frac{1}{8},$$

which also holds for the constant standard kernel $\kappa \equiv 1/2$.

Remark 3.3. To go back to the question posed at the beginning of the section, namely to decide whether a given graph G_n is compatible with any homogenous model $G(n, p)$, $0 < p < 1$, we can formulate this now more precisely as the testing problem

$$H_0: G_n \sim G(n, p) \quad \text{for some } 0 < p < 1$$

against

$$H_1: G_n \sim G(n, \kappa) \quad \text{with } \kappa \not\equiv p \text{ for all } 0 < p < 1.$$

As we have already pointed out in Remark 3.1, from the point of view of first order approximation of dense graph limit theory, the alternative hypothesis is already in its most general form, since the models $G(n, \kappa)$ cover all possible dense graph limits.

We can now define a test $\psi(G_n)$ that rejects the null hypothesis if $C_{1-\alpha}(G_n)$ is empty, that is,

$$\psi(G_n) = \mathbb{I}[W_1^2(p, G_n) + W_2^2(p, G_n) > q_{1-\alpha} \text{ for all } 0 < p < 1].$$

Since

$$\mathbb{P}_p[\psi(G_n) = 1] \leq \mathbb{P}_p[W_1^2(p, G_n) + W_2^2(p, G_n) > q_{1-\alpha}] = \alpha + O(n^{-1/2}),$$

this test has an asymptotic significance level of α or less. Whether the asymptotic significance level is strictly less than or equal to α depends on the asymptotic behaviour of the quantity

$$\inf_{0 < p < 1} \{W_1^2(p, G_n) + W_2^2(p, G_n)\},$$

which cannot be expected to have a χ^2 -distribution. Numerical simulations indicate that the asymptotic significance level of ψ is strictly less than α , but a mathematical proof of this observation eludes us.

Before we prove Theorem 3.3, we need some notation and technical lemmas. For the remainder of this subsection, that is until the end of the proof of Theorem 3.3, we will follow the convention that the elements in an ordered m -tuple (i_1, \dots, i_m) of integers are pairwise different and range from 1 to n , and we will assume the same for sets written as $\{i_1, \dots, i_m\}$, so that $|\{i_1, \dots, i_m\}| = m$ always. For every (i, j, k, l) let

$$\eta_{ijkl} = I_{ij}I_{jk}I_{kl}I_{il} - p^3(I_{ij} + I_{jk} + I_{kl} + I_{il}) + 3p^4,$$

where $I_{ij} = I_{ji}$ is the indicator of the event that there is an edge connecting i and j . Note that between every set of four vertices $\{i, j, k, l\}$, only three essentially different 4-cycles can be spanned, so that, for example, the set of eight 4-tuples

$$\begin{aligned} &\{(i, j, k, l), (j, k, l, i), (k, l, i, j), (l, i, j, k), \\ &(i, l, k, j), (l, k, j, i), (k, j, i, l), (j, i, l, k)\} \end{aligned}$$

represent the same 4-cycle, and hence

$$\begin{aligned} \eta_{ijkl} &= \eta_{jkli} = \eta_{klij} = \eta_{lijk} \\ &= \eta_{ilkj} = \eta_{lkji} = \eta_{kjil} = \eta_{jilk}. \end{aligned}$$

It is also straightforward to verify that, if $\mathcal{V} \subset \{1, \dots, n\}$ is of arbitrary size, then

$$\mathbb{E}\{\eta_{ijkl}(I_{uv})_{u,v \in \mathcal{V}}\} = 0 \tag{3.5}$$

for any (i, j, k, l) with $|\{i, j, k, l\} \cap \mathcal{V}| \leq 2$. From (3.5), we can easily deduce statements about mixed moments. For example, for any (i, j, k, l) and any (u, v) , we have

$$\mathbb{E}\{\eta_{ijkl}I_{uv}\} = 0,$$

or, if $|\{i, j, k, l\} \cap \{u, v, w, m\}| \leq 2$, we have

$$\mathbb{E}\{\eta_{ijkl}\eta_{uvwm}\} = 0.$$

Whenever we will be using such identities (or similar identities with more factors) in the proof, we will only refer to (3.5), since obtaining these covariance formulas from (3.5) is straightforward.

For each $\nu = \{i, j, k, l\}$, let

$$X_{1,\nu} = \frac{1}{\binom{n-2}{2}\sigma_1} (I_{ij} + I_{ik} + I_{il} + I_{jk} + I_{jl} + I_{kl} - 6p),$$

$$X_{2,\nu} = \frac{1}{\sigma_2} (\eta_{ijkl} + \eta_{ijlk} + \eta_{ikjl}),$$

and $X_\nu = (X_{1,\nu}, X_{2,\nu})^t$. Now we can represent W as a sum of locally dependent random vectors, namely

$$W = \sum_{\nu} X_{\nu}, \tag{3.6}$$

where the sum ranges over all subsets $\nu = \{i, j, k, l\}$. To see that (3.6) is the same as in Theorem 3.3, recall that between each set of four vertices $\{i, j, k, l\}$, there can be at most three different 4-cycles, and that, in the definition of $X_{2,\nu}$, one representative of each of them is picked. Furthermore, each edge I_{ij} is over-counted $\binom{n-2}{2}$ times, hence the additional factor $\binom{n-2}{2}^{-1}$ in the definition of $X_{1,\nu}$. It is straightforward to check that

$$\mathbb{E}W = 0, \quad \mathbb{E}\{WW^t\} = \mathbb{I}_2,$$

where \mathbb{I}_m is the m -dimensional unit matrix. Note that X_ν and X_ξ are independent whenever $|\nu \cap \xi| \leq 1$, that is, share at most one vertex. Hence, for each ν , we define the set $A_\nu := \{\xi : |\nu \cap \xi| \geq 2\}$, the ‘neighbourhood’ of X_ν . For given ν , we then have that the collection $(X_\xi)_{\xi \notin A_\nu}$ is independent of X_ν . Therefore, if I is uniformly distributed over all ν ,

$$(W, W', G) := \left(W, W - \sum_{\nu \in A_I} X_\nu, -\binom{n}{4} X_I \right) \tag{3.7}$$

is a Stein coupling (cf. Section 5).

Since the sequence (G_n) starts at some unspecified integer n_0 , we can assume without loss of generality that $n_0 \geq 3$, and, hence, use G_1 and G_2 to denote the first, respectively, second component of the vector G , rather than elements from the random graph sequence (G_n) .

Proof of Theorem 3.3. We apply Theorem 2.1 for the Stein coupling given in (3.7). Let as usual $D = W' - W$. In what follows, C denotes a positive constant independent of p and n , possibly different from line to line. Note first that

$$\sigma_1^2 \geq Cn^2p(1-p), \quad \sigma_2^2 \geq Cn^5p^6(1-p)^2.$$

Hence,

$$|X_\nu| \leq C \left(\frac{1}{n^2\sigma_1} + \frac{1}{\sigma_2} \right) \leq \frac{C}{n^{5/2}p^3(1-p)}$$

and $|A_\nu| \leq Cn^2$, which yields the upper bounds

$$|G| \leq \frac{Cn^{3/2}}{p^3(1-p)} =: \alpha, \quad |D| \leq \frac{C}{p^3(1-p)n^{1/2}} =: \beta. \tag{3.8}$$

The second moment of $|D|$ can be calculated as follows. Noting that $|\xi \cap \xi'| \leq 1$ implies $\mathbb{E}(X_{1,\xi} X_{1,\xi'}) = 0$, and $|\xi \cap \xi'| \leq 2$ implies $\mathbb{E}(X_{2,\xi} X_{2,\xi'}) = 0$, which follows from (3.5), we have

$$\begin{aligned} \mathbb{E}|D|^2 &= \mathbb{E}D_1^2 + \mathbb{E}D_2^2 \\ &= \frac{1}{\binom{n}{4}} \sum_\nu \sum_{\xi, \xi' \in A_\nu} \mathbb{E}(X_{1,\xi} X_{1,\xi'}) + \frac{1}{\binom{n}{4}} \sum_\nu \sum_{\xi, \xi' \in A_\nu} \mathbb{E}(X_{2,\xi} X_{2,\xi'}) \\ &\leq \frac{C}{n^4} \times n^4 \times n^2 \times n^2 \times \frac{1}{n^4 \sigma_1^2} + \frac{C}{n^4} \times n^4 \times n^2 \times n \times \frac{1}{\sigma_2^2} \\ &\leq \frac{C}{n^2 p^6 (1-p)^2}. \end{aligned} \tag{3.9}$$

Define the σ -field $\mathcal{F} = \sigma(G_n)$. Clearly, $\mathcal{F} \supset \sigma(W)$. In the following, we calculate the variances of the conditional expectations in the bound (2.4). First,

$$\begin{aligned} \text{Var}(\mathbb{E}^{\mathcal{F}} G_1 D_1) &\leq \text{Var}\left(\sum_\nu X_{1,\nu} \sum_{\xi \in A_\nu} X_{1,\nu}\right) \\ &= \sum_{\nu, \nu'} \sum_{\xi \in A_\nu, \xi' \in A_{\nu'}} \text{Cov}(X_{1,\nu} X_{1,\xi}, X_{1,\nu'} X_{1,\xi'}) \leq \frac{Cn^{10}}{n^8 \sigma_1^4}, \end{aligned} \tag{3.10}$$

where the last inequality follows from the fact that $\text{Cov}(X_{1,\nu} X_{1,\xi}, X_{1,\nu'} X_{1,\xi'}) \neq 0$ can occur only if $|(\nu \cup \xi) \cap (\nu' \cup \xi')| \geq 2$. By the same argument,

$$\text{Var}(\mathbb{E}^{\mathcal{F}} G_1 D_2) \leq \sum_{\nu, \nu'} \sum_{\xi \in A_\nu, \xi' \in A_{\nu'}} \text{Cov}(X_{1,\nu} X_{2,\xi}, X_{1,\nu'} X_{2,\xi'}) \leq \frac{Cn^{10}}{n^4 \sigma_1^2 \sigma_2^2} \tag{3.11}$$

and

$$\text{Var}(\mathbb{E}^{\mathcal{F}} G_2 D_1) \leq \frac{Cn^{10}}{n^4 \sigma_1^2 \sigma_2^2}. \tag{3.12}$$

In order to bound $\text{Var}(\mathbb{E}^{\mathcal{F}} G_2 D_2)$, we argue that

$$\text{Cov}(X_{2,\nu} X_{2,\xi}, X_{2,\nu'} X_{2,\xi'}) \neq 0 \quad \text{implies} \quad |\nu \cup \xi \cup \nu' \cup \xi'| \leq 9, \tag{3.13}$$

from which we can deduce that

$$\text{Var}(\mathbb{E}^{\mathcal{F}} G_2 D_2) \leq \frac{Cn^9}{\sigma_2^4}. \tag{3.14}$$

To show (3.13), note that the left-hand side implies that

- (i) any intersection of ν , ξ , ν' or ξ' with the union of the other three sets has at least three elements (otherwise we would obtain a contradiction with (3.5)), and
- (ii) at least one of the intersections $\nu \cap \nu'$, $\nu \cap \xi'$, $\xi \cap \nu'$ and $\xi \cap \xi'$ has at least two elements (otherwise $X_{1,\nu'}X_{1,\xi'}$ and $X_{1,\nu}X_{1,\xi}$ would be independent).

Assume now that the left-hand side of (3.13) is true. Since $\xi \in A_\nu$, we have $|\nu \cap \xi| \geq 2$, and hence $|\nu \cup \xi| \leq 6$, and similarly $|\nu' \cup \xi'| \leq 6$. Using (ii), we deduce that one of the three inequalities $|\nu \cup \xi \cap \nu'| \geq 2$, $|\nu \cup \xi \cap \xi'| \geq 2$, or $|\nu' \cup \xi' \cap \nu| \geq 2$ must hold. If the first inequality holds, we obtain $|\nu \cup \xi \cup \nu'| \leq 8$, and, using (i), that $|\nu \cup \xi \cup \nu' \cup \xi'| \leq 9$; the other two inequalities are analogous. This concludes the proof of (3.13).

Collecting the bounds (3.10), (3.11), (3.12) and (3.14), we obtain

$$B_2 \leq \frac{C}{n^{1/2}p^6(1-p)^2}.$$

By similar arguments,

$$\text{Var } \mathbb{E}^{\mathcal{F}}(D_1^2) \leq \frac{C}{n^2\sigma_1^4}, \quad \text{Var } \mathbb{E}^{\mathcal{F}}(D_2^2) \leq \frac{Cn^5}{\sigma_2^4},$$

and, hence,

$$B_1 \leq \frac{C}{n^{5/2}p^6(1-p)^2}.$$

The following bounds can be obtained in a similar fashion, again using (3.5), but we omit the tedious details. We have

$$\begin{aligned} \text{Var } \mathbb{E}^{\mathcal{F}}(G_1 D_1^2) &\leq \frac{Cn^{14}}{n^{12}\sigma_1^6}, & \text{Var } \mathbb{E}^{\mathcal{F}}(G_1 D_1 D_2) &\leq \frac{Cn^{14}}{n^8\sigma_1^4\sigma_2^2}, \\ \text{Var } \mathbb{E}^{\mathcal{F}}(G_1 D_2^2) &\leq \frac{Cn^{13}}{n^4\sigma_1^2\sigma_2^4}, & \text{Var } \mathbb{E}^{\mathcal{F}}(G_2 D_1^2) &\leq \frac{Cn^{14}}{n^8\sigma_1^4\sigma_2^2}, \\ \text{Var } \mathbb{E}^{\mathcal{F}}(G_2 D_1 D_2) &\leq \frac{Cn^{13}}{n^4\sigma_1^2\sigma_2^4}, & \text{Var } \mathbb{E}^{\mathcal{F}}(G_2 D_2^2) &\leq \frac{Cn^{13}}{\sigma_2^6}, \end{aligned}$$

and therefore

$$B_3 \leq \frac{C}{np^9(1-p)^3}.$$

Collecting the bounds on B_1 , B_2 and B_3 , in combination with (3.8) and (3.9), yields the final estimate via (2.4). □

3.2. Joint normality of certain permutation statistics

Let M be a real $n \times n$ matrix and assume that M is anti-symmetric, that is, for each $u, v \in \{1, \dots, n\}$, we have

$$M_{uv} = -M_{vu}.$$

Note that $M_{uu} = 0$. Let π be a permutation of size n , chosen uniformly at random, and consider the statistic

$$W = \sum_{i < j} M_{\pi(i)\pi(j)}. \tag{3.15}$$

Here, sums of the form $\sum_{i < j}$ have to be interpreted as double sums $\sum_{i=1}^{n-1} \sum_{j=i+1}^n$. If it is to be interpreted as a single sum, we will explicitly state the summation index using the notation $\sum_{j:i < j}$.

Permutation statistics of the form (3.15) were considered by Fulman [16] and they are a special case of doubly-indexed permutation statistics

$$\sum_{i,j} a(i, j, \pi(i), \pi(j)) \tag{3.16}$$

with

$$a(i, j, u, v) = \mathbb{I}[i < j]M_{uv}.$$

The reason to study (3.15) is that two important properties of permutations, the number of descents and inversions, can be readily represented in this form. Choosing $M_{u,u+1} = -1$ and $M_{uv} = 0$ for all other $v > u$ (for $v < u$, M_{uv} is defined via anti-symmetry), (3.15) becomes $2\text{Des}(\pi^{-1}) - (n - 1)$, where $\text{Des}(\pi)$ is the number of descents of π ; with $M_{uv} = -1$ for all $u < v$, (3.15) becomes $2\text{Inv}(\pi^{-1}) - \binom{n}{2}$, where $\text{Inv}(\pi)$ is the number of inversions of π .

Using Stein’s method, Zhao, Bai, Chao and Liang [26] prove a general Berry–Esseen type theorem for sums of the form (3.16), but their results do not apply to the number of descents $\text{Des}(\pi)$, which seems to be “too sparse”. In contrast, using a special exchangeable pair, Fulman [16] was able to obtain a rate of convergence of $n^{-1/2}$ for the Kolmogorov metric for both, the number of descents and inversions.

We shall extend Fulman’s results to the multivariate setting. Furthermore, we are able to remove a certain condition on M (present in Fulman’s work), arising from the requirement of exchangeability; cf. Remark 2.3. In addition to extending the exchangeable pair approach by Fulman [16], we also provide a result using the local approach.

Let $M^{(1)}, \dots, M^{(d)}$ be a sequence of real $n \times n$ matrices and assume that each matrix is anti-symmetric. For each r , define $W_r = \sum_{i < j} M_{\pi(i)\pi(j)}^{(r)}$. As in Fulman [16], define

$$A_u^{(r)} = \sum_{v:v>u} M_{uv}^{(r)}, \quad B_u^{(r)} = \sum_{v:v<u} M_{vu}^{(r)}.$$

The mean and covariances of $W = (W_1, \dots, W_d)$ are given in the following lemma.

Lemma 3.5. *We have $\mathbb{E}W = 0$ and*

$$\text{Cov}(W_r, W_s) = \frac{1}{3} \left(\sum_{u < v} M_{uv}^{(r)} M_{uv}^{(s)} + \sum_u (A_u^{(r)} - B_u^{(r)})(A_u^{(s)} - B_u^{(s)}) \right). \tag{3.17}$$

Proof. Both the covariance and the right-hand side of (3.17) are symmetric bilinear forms on the vector space of all anti-symmetric matrices. Moreover, by Lemma 4.3.1 of Fulman [16], both expressions match for $M^{(r)} = M^{(s)}$. Since a symmetric bilinear form is uniquely determined by the corresponding quadratic form, the results follows. \square

With $W = (W_1, \dots, W_d)^t$, we have the following result.

Theorem 3.6. *Let W be as above and let*

$$\beta = \sup_{r,u} \sum_v |M_{uv}^{(r)}|, \quad \beta_2 = \sup_{r,u} \sum_v (M_{uv}^{(r)})^2. \tag{3.18}$$

Assume $\text{Var}(W_r) = 1$ for each $1 \leq r \leq d$. Then, with $\Sigma = \text{Cov}(W)$, there is a positive constant C_d depending only on d , such that

$$d_c(\mathcal{L}(W), \mathcal{L}(\Sigma^{1/2}Z)) \leq C_d \|\Sigma^{-1/2}\|_2^2 (n\beta^3 + n^{1/2}\beta\beta_2^{1/4}). \tag{3.19}$$

Although Theorem 3.6 is widely applicable, it does not yield optimal bounds for the applications discussed below. To this end, we also give a theorem that gives better bounds under the more specific situation where the non-zero entries of $M^{(r)}$ are all near the diagonal and W_1 is the normalised number of inversions.

Theorem 3.7. *Assume the situation of Theorem 3.6. In addition, assume that W_1 is of the specific form*

$$W_1 = \frac{\text{Inv}(\pi) - (1/2)\binom{n}{2}}{\sqrt{(n(n-1)(2n+5))/72}},$$

where $\text{Inv}(\pi)$ is the number of inversions of π . Assume further that there is a positive integer m such that

$$M_{uv}^{(r)} = 0, \quad \text{if } |u - v| > m \text{ and } 2 \leq r \leq d.$$

Then

$$d_c(\mathcal{L}(W), \mathcal{L}(\Sigma^{1/2}Z)) \leq C_{d,m} \|\Sigma^{-1/2}\|_2^2 n\beta^3, \tag{3.20}$$

where $C_{d,m}$ is a positive constant depending only on d and m , and where

$$\beta := \max \left\{ \frac{1}{\sqrt{n}}, \sup_{r,u} \sum_v |M_{uv}^{(r)}| \right\}.$$

Remark 3.4. We will use Corollary 2.2 to prove (3.19) and (3.20). The bounds in (3.19) and (3.20) have fewer terms than the bound in (2.5) because we will make use of the inequality $d \leq \alpha\beta$ for α and β defined in (2.3), which follows from (2.2) and the assumption that $\text{Var}(W_r) = 1$ for each r .

As a corollary of Theorem 3.7, we prove the joint asymptotic normality of the number of descents and inversions of π ; the rate obtained is best possible.

Corollary 3.8. *Let $\text{Des}(\pi)$ and $\text{Inv}(\pi)$ be the number of descents and inversions of π , and let*

$$W = (W_1, W_2)^t = \left(\frac{\text{Inv}(\pi) - (1/2)\binom{n}{2}}{\sqrt{(n(n-1)(2n+5))/72}}, \frac{\text{Des}(\pi) - (n-1)/2}{\sqrt{(n+1)/12}} \right)^t.$$

Then

$$d_c(\mathcal{L}(W), \mathcal{L}(Z)) \leq \frac{C}{\sqrt{n}}$$

for some absolute constant C , where Z is a 2-dimensional standard Gaussian vector.

Proof. Set

$$M_{uv}^{(1)} = \sqrt{\frac{18}{n(n-1)(2n+5)}} \times \begin{cases} -1 & \text{if } v > u, \\ +1 & \text{if } v < u, \\ 0 & \text{otherwise,} \end{cases}$$

and set

$$M_{uv}^{(2)} = \sqrt{\frac{3}{n+1}} \times \begin{cases} -1 & \text{if } v = u + 1, \\ +1 & \text{if } v = u - 1, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, we can take $m = 1$ in Theorem 3.7. Let $\tau = \pi^{-1}$, which is again a uniform random permutation of size n . It can be easily verified that $W_1 = \sum_{i < j} M_{\tau(i)\tau(j)}^{(1)}$ and $W_2 = \sum_{i < j} M_{\tau(i)\tau(j)}^{(2)}$. From Lemma 3.5, $\text{Var}(W_1) = \text{Var}(W_2) = 1$ and $|\text{Cov}(W_1, W_2)| \leq C/n$. Moreover, β as defined in (3.18) is smaller than C/\sqrt{n} . Therefore, the corollary is proved by applying Theorem 3.7. \square

To prove Theorem 3.6, we need the following lemma, the proof of which is straightforward and therefore omitted.

Lemma 3.9. *For $1 \leq r, s, t \leq d$ and β defined in (3.18), we have*

$$\sum_{u_1, \dots, u_6} |M_{u_1 u_2}^{(r)} M_{u_1 u_3}^{(s)} M_{u_4 u_5}^{(r)} M_{u_4 u_6}^{(s)}| \leq n^2 \beta^4, \tag{3.21}$$

$$\sum_{\substack{\{|u_1, u_2, u_3\}|=3, \{|u_4, u_5, u_6\}|=3, \\ \{|u_1, \dots, u_6\}| \leq 5}} |M_{u_1 u_2}^{(r)} M_{u_1 u_3}^{(s)} M_{u_4 u_5}^{(r)} M_{u_4 u_6}^{(s)}| \leq 9n\beta^4, \tag{3.22}$$

$$\sum_{u_1, \dots, u_8} |M_{u_1 u_2}^{(r)} M_{u_1 u_3}^{(s)} M_{u_1 u_4}^{(t)} M_{u_5 u_6}^{(r)} M_{u_5 u_7}^{(s)} M_{u_5 u_8}^{(t)}| \leq n^2 \beta^6, \tag{3.23}$$

$$\sum_{\{|u_1, \dots, u_8\}| \leq 7} |M_{u_1 u_2}^{(r)} M_{u_1 u_3}^{(s)} M_{u_1 u_4}^{(t)} M_{u_5 u_6}^{(r)} M_{u_5 u_7}^{(s)} M_{u_5 u_8}^{(t)}| \leq 22n^2 \beta^4 \beta_2, \tag{3.24}$$

where $\sum_{\{|u_1, \dots, u_k\}| \leq k-1}$ stands for summation over all tuples (u_1, \dots, u_k) for which at least two components are equal.

Proof of Theorem 3.6. We adopt the construction of W' from Fulman [16]. Let I be uniformly chosen from $\{1, \dots, n\}$ and independently of π . Given I , we define π' as $\pi \circ (I, I + 1, \dots, n)$ where $(I, I + 1, \dots, n)$ denotes the mapping $I \mapsto I + 1 \mapsto \dots \mapsto n \mapsto I$, while keeping the rest identical. As π and π' both are uniformly distributed, W and W' have the same marginal distribution (but are not necessarily exchangeable). Fulman [16] showed that with $\lambda = 2/n$

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

Following Remark 2.3, the bound (2.5) holds with $D = W' - W$ and $G = \frac{1}{2}\lambda^{-1}D = nD/4$ (cf. Section 5). From the construction of W' , we have (cf. Lemma 4.2.1 of Fulman [16])

$$D_r = -2 \sum_{j>I} M_{\pi(I)\pi(j)}^{(r)}$$

for $r \in \{1, \dots, d\}$. By the definition of β in (3.18),

$$|G| \leq C_d n \beta, \quad |D| \leq C_d \beta. \tag{3.25}$$

We first prove that

$$\text{Var} \mathbb{E}^\pi (D_r D_s) \leq \frac{C_d \beta^4}{n}. \tag{3.26}$$

From the construction of W' ,

$$\begin{aligned} & \text{Var} \mathbb{E}^\pi (D_r D_s) \\ &= \text{Var} \left(\frac{4}{n} \sum_{i=1}^n \sum_{\substack{j_1, j_2: \\ j_1, j_2 > i}} M_{\pi(i)\pi(j_1)}^{(r)} M_{\pi(i)\pi(j_2)}^{(s)} \right) \\ &= \frac{16}{n^2} \text{Var} \left(\sum_{i=1}^n \sum_{j>i} M_{\pi(i)\pi(j)}^{(r)} M_{\pi(i)\pi(j)}^{(s)} + \sum_{i=1}^n \sum_{\substack{j_1, j_2: \\ j_1, j_2 > i, j_1 \neq j_2}} M_{\pi(i)\pi(j_1)}^{(r)} M_{\pi(i)\pi(j_2)}^{(s)} \right). \end{aligned}$$

Using antisymmetry, it is not difficult to see that the first double sum in the last line is constant. Hence, we only need to show that

$$\begin{aligned} & \sum_{\substack{\{|i, j_1, j_2\}|=3, \{|k, l_1, l_2\}|=3 \\ \{|i, j_1, j_2\}|=3, \{|k, l_1, l_2\}|=3}} |K_{ij_1 j_2 k l_1 l_2}^{(r,s)}| \\ & := \sum_{\substack{\{|i, j_1, j_2\}|=3, \{|k, l_1, l_2\}|=3}} |\text{Cov}(M_{\pi(i)\pi(j_1)}^{(r)} M_{\pi(i)\pi(j_2)}^{(s)}, M_{\pi(k)\pi(l_1)}^{(r)} M_{\pi(k)\pi(l_2)}^{(s)})| \quad (3.27) \\ & \leq C_d n \beta^4. \end{aligned}$$

We consider the cases $|\{i, j_1, j_2, k, l_1, l_2\}| = 6$ and $|\{i, j_1, j_2, k, l_1, l_2\}| \leq 5$ separately. For the first case, we have

$$\begin{aligned} K_{ij_1 j_2 k l_1 l_2}^{(r,s)} &= \frac{1}{(n)_6} \sum_{\{|u, v_1, v_2, w, z_1, z_2\}|=6} M_{uv_1 v_2 w z_1 z_2}^{(r,s)} - \frac{1}{((n)_3)^2} \sum_{\substack{\{|u, v_1, v_2\}|=3 \\ \{|w, z_1, z_2\}|=3}} M_{uv_1 v_2 w z_1 z_2}^{(r,s)} \\ &= \left(\frac{1}{(n)_6} - \frac{1}{((n)_3)^2} \right) \sum_{\{|u, v_1, v_2, w, z_1, z_2\}|=6} M_{uv_1 v_2 w z_1 z_2}^{(r,s)} \\ &\quad - \frac{1}{((n)_3)^2} \sum_{\substack{\{|u, v_1, v_2\}|=3, \{|w, z_1, z_2\}|=3 \\ \{|u, v_1, v_2, w, z_1, z_2\}| \leq 5}} M_{uv_1 v_2 w z_1 z_2}^{(r,s)}, \end{aligned}$$

where $M_{uv_1 v_2 w z_1 z_2}^{(r,s)} := M_{uv_1}^{(r)} M_{uv_2}^{(s)} M_{wz_1}^{(r)} M_{wz_2}^{(s)}$. By (3.21) and (3.22),

$$|K_{ij_1 j_2 k l_1 l_2}^{(r,s)}| \leq C_d \frac{\beta^4}{n^5}.$$

Next, we consider the case $|\{i, j_1, j_2, k, l_1, l_2\}| \leq 5$. Let $\tilde{\pi}$ be an independent copy of π . Again by (3.21) and (3.22),

$$\begin{aligned} & \sum_{\substack{\{|i, j_1, j_2\}|=3, \{|k, l_1, l_2\}|=3 \\ \{|i, j_1, j_2, k, l_1, l_2\}| \leq 5}} |K_{ij_1 j_2 k l_1 l_2}^{(r,s)}| \\ & \leq \mathbb{E} \left[\sum_{\substack{\{|i, j_1, j_2\}|=3, \{|k, l_1, l_2\}|=3 \\ \{|i, j_1, j_2, k, l_1, l_2\}| \leq 5}} \left(|M_{\pi(i)\pi(j_1)\pi(j_2)\pi(k)\pi(l_1)\pi(l_2)}^{(r,s)}| + |M_{\pi(i)\pi(j_1)\pi(j_2)\tilde{\pi}(k)\tilde{\pi}(l_1)\tilde{\pi}(l_2)}^{(r,s)}| \right) \right] \\ & = \mathbb{E} \left[\sum_{\{|u, v_1, v_2\}|=3, \{|w, z_1, z_2\}|=3} |M_{uv_1 v_2 w z_1 z_2}^{(r,s)}| \left(\mathbf{I}(|\{u, v_1, v_2, w, z_1, z_2\}| \leq 5) \right. \right. \\ & \quad \left. \left. + \mathbf{I}(|\pi^{-1}(\{u, v_1, v_2\}) \cap \tilde{\pi}^{-1}(\{w, z_1, z_2\})| \leq 5) \right) \right] \\ & \leq C_d n \beta^4. \end{aligned}$$

Therefore, we have proved (3.27), and thus (3.26). Again from the construction of W' , we can write

$$\text{Var } \mathbb{E}^\pi(D_r D_s D_t) = \frac{64}{n^2} \sum_{\substack{j_1 > i, j_2 > i, j_3 > i \\ l_1 > k, l_2 > k, l_3 > k}} K_{ij_1 j_2 j_3 k l_1 l_2 l_3}^{(r,s,t)},$$

where

$$K_{ij_1 j_2 j_3 k l_1 l_2 l_3}^{(r,s,t)} := \text{Cov}(M_{\pi(i)\pi(j_1)}^{(r)} M_{\pi(i)\pi(j_2)}^{(s)} M_{\pi(i)\pi(j_3)}^{(t)}, M_{\pi(k)\pi(l_1)}^{(r)} M_{\pi(k)\pi(l_2)}^{(s)} M_{\pi(k)\pi(l_3)}^{(t)}).$$

By the same argument as for $K_{ij_1 j_2 k l_1 l_2}^{(r,s)}$, and using the bounds (3.23) and (3.24) instead of (3.21) and (3.22), we can prove

$$\text{Var } \mathbb{E}^\pi(D_r D_s D_t) \leq C_d \beta^4 \beta_2. \tag{3.28}$$

Applying the bounds (3.25), (3.26) and (3.28) in (2.5) and using $1 \leq C_d n \beta^2$ by Remark 3.4 prove the theorem. \square

Next, we prove Theorem 3.7.

Proof of Theorem 3.7. Let $\tau = \pi^{-1}$. From Diaconis [13], W_1 can be expressed as

$$W_1 = \sum_{u=1}^n \frac{1}{\sqrt{(n(n-1)(2n+5))/72}} \left(\xi_u - \frac{n-u}{2} \right) =: \sum_{u=1}^n X_u^{(1)},$$

where ξ_1 is the minimum number of pairwise adjacent transpositions taking $\tau(1)$ to the first position, ξ_2 is the minimum number of pairwise adjacent transpositions taking $\tau(2)$ to the second position after the first step is done, etc. Because τ is a uniform permutation, $\{\xi_1, \dots, \xi_n\}$ are independent random variables with $\xi_u \sim \text{Uniform}\{0, \dots, n-u\}$ for $1 \leq u \leq n$. For $2 \leq r \leq d$, by the assumption that $M_{uv}^{(r)} = 0$ if $|u-v| > m$,

$$\begin{aligned} W_r &= \sum_{i < j} M_{\pi(i)\pi(j)}^{(r)} = \sum_{\substack{u,v: |u-v| \leq m \\ \pi^{-1}(u) < \pi^{-1}(v)}} M_{uv}^{(r)} \\ &= \sum_{u=1}^n \left(\sum_{v: |u-v| \leq m} M_{uv}^{(r)} \mathbb{I}[\tau(u) < \tau(v)] \right) =: \sum_{u=1}^n X_u^{(r)}. \end{aligned}$$

Let $X_u = (X_u^{(1)}, \dots, X_u^{(d)})^t$. Then $W = \sum_{u=1}^n X_u$. In the above pairwise transposition process, if we know $\{\xi_v : 1 \leq v \leq u+m\}$, then we can reconstruct the positions of $\{\tau(v) : |v-u| \leq m\}$. Observe that the relative order of $\{\tau(v) : |v-u| \leq m\}$ does not depend on $\{\xi_v : 1 \leq v < u-m\}$. Therefore, X_u is measurable with respect to $\{\xi_v : |v-u| \leq m\}$ and W can be viewed as a sum of locally dependent random vectors (cf. Section 5.3) with neighbourhood $A_u = \{u-2m, u+2m\}$ for each $1 \leq u \leq n$. For the Stein coupling (5.2), we have

$$|G| \leq C_{d,m} n \beta, \quad |D| \leq C_{d,m} \beta.$$

Moreover, by the local dependence structure,

$$\begin{aligned} \text{Var } \mathbb{E}^W(G_r D_s) &\leq \text{Var} \left(\sum_{u=1}^n \sum_{v \in A_u} X_u^{(r)} X_v^{(s)} \right) \\ &= \sum_{u=1}^n \sum_{w: |w-u| \leq 6m} \text{Cov} \left(\sum_{v \in A_u} X_u^{(r)} X_v^{(s)}, \sum_{z \in A_w} X_w^{(r)} X_z^{(s)} \right) \\ &\leq C_{d,m} n \beta^4, \end{aligned}$$

where we used the inequality $\text{Cov}(X, Y) \leq (\mathbb{E}X^2 + \mathbb{E}Y^2)/2$. Similarly,

$$\text{Var } \mathbb{E}^W(G_r D_s D_t) \leq C_{d,m} n \beta^6.$$

The bound (3.20) is proved by applying the above bounds to (2.5) and using $1 \leq C_{d,m} n \beta^2$ by Remark 3.4. □

4. Proof of main theorem

For given test function h , we consider the Stein equation

$$\Delta f(w) - w^t \nabla f(w) = h(w) - \mathbb{E}h(Z), \quad w \in \mathbb{R}^d, \tag{4.1}$$

where Δ denotes the Laplacian operator and ∇ the gradient operator. If h is not continuous (like the indicator function of a convex set), then f is not smooth enough to apply Taylor expansion to the necessary degree, so more refined techniques are necessary.

We follow the smoothing technique of Bentkus [5]. Recall that \mathcal{A} is the collection of all convex sets in \mathbb{R}^d . For $A \in \mathcal{A}$, let $h_A(x) = I_A(x)$, and define the smoothed function

$$h_{A,\varepsilon}(w) = \psi \left(\frac{\text{dist}(w, A)}{\varepsilon} \right), \tag{4.2}$$

where $\text{dist}(w, A) = \inf_{v \in A} |w - v|$ and

$$\psi(x) = \begin{cases} 1, & x < 0, \\ 1 - 2x^2, & 0 \leq x < \frac{1}{2}, \\ 2(1 - x)^2, & \frac{1}{2} \leq x < 1, \\ 0, & 1 \leq x. \end{cases} \tag{4.3}$$

Define also

$$A^\varepsilon = \{x \in \mathbb{R}^d : \text{dist}(x, A) \leq \varepsilon\}, \quad A^{-\varepsilon} = \{x \in A : \text{dist}(x, \mathbb{R}^d \setminus A) > \varepsilon\}$$

(note that in general $(A^{-\varepsilon})^\varepsilon \neq A$).

Lemma 4.1 (Lemma 2.3 of Bentkus [5]). *The function $h_{A,\varepsilon}$ as defined above has the following properties:*

$$(i) \quad h_{A,\varepsilon}(w) = 1 \quad \text{for all } w \in A, \tag{4.4}$$

$$(ii) \quad h_{A,\varepsilon}(w) = 0 \quad \text{for all } w \in \mathbb{R}^d \setminus A^\varepsilon, \tag{4.5}$$

$$(iii) \quad 0 \leq h_{A,\varepsilon}(w) \leq 1 \quad \text{for all } w \in A^\varepsilon \setminus A, \tag{4.6}$$

$$(iv) \quad |\nabla h_{A,\varepsilon}(w)| \leq 2\varepsilon^{-1} \quad \text{for all } w \in \mathbb{R}^d, \tag{4.7}$$

$$(v) \quad |\nabla h_{A,\varepsilon}(v) - \nabla h_{A,\varepsilon}(w)| \leq 8|v - w|\varepsilon^{-2} \quad \text{for all } v, w \in \mathbb{R}^d. \tag{4.8}$$

Lemma 4.2. *For any d -dimensional random vector W ,*

$$d_c(\mathcal{L}(W), \mathcal{L}(Z)) \leq 4d^{1/4}\varepsilon + \sup_{A \in \mathcal{A}} |\mathbb{E}h_{A,\varepsilon}(W) - \mathbb{E}h_{A,\varepsilon}(Z)|. \tag{4.9}$$

Proof. By (2.2) of Bentkus [5], for any $\varepsilon > 0$,

$$\begin{aligned} d_c(\mathcal{L}(W), \mathcal{L}(Z)) &\leq \sup_{A \in \mathcal{A}} |\mathbb{E}h_{A,\varepsilon}(W) - \mathbb{E}h_{A,\varepsilon}(Z)| \\ &\quad + \sup_{A \in \mathcal{A}} \max\{\mathbb{P}(Z \in A^\varepsilon \setminus A), \mathbb{P}(Z \in A \setminus A^{-\varepsilon})\}. \end{aligned}$$

From Ball [2] and Bentkus [5], we have

$$\sup_{A \in \mathcal{A}} \max\{\mathbb{P}(Z \in A^\varepsilon \setminus A), \mathbb{P}(Z \in A \setminus A^{-\varepsilon})\} \leq 4d^{1/4}\varepsilon \tag{4.10}$$

(the dependence on d in (4.9) is optimal; see Bentkus [5]). □

Fix now ε and a convex $A \subset \mathbb{R}^d$. It can be verified directly that with

$$g_{A,\varepsilon}(w, s) = -\frac{1}{2(1-s)} \int_{\mathbb{R}^d} [h_{A,\varepsilon}(\sqrt{1-s}w + \sqrt{s}z) - \mathbb{E}h_{A,\varepsilon}(Z)] \varphi(z) dz,$$

a solution to (4.1) is (cf. Götze [18])

$$f_{A,\varepsilon}(w) = \int_0^1 g_{A,\varepsilon}(w, s) ds, \tag{4.11}$$

where φ is the density function of the d -dimensional standard normal distribution. In what follows, we keep the dependence on A and ε implicit and write $g = g_{A,\varepsilon}$, $f = f_{A,\varepsilon}$ and $h = h_{A,\varepsilon}$. For real-valued functions on \mathbb{R}^d we will write $f_i(x)$ for $\partial f(x)/\partial x_i$, $f_{ij}(x)$ for $\partial^2 f(x)/(\partial x_i \partial x_j)$ and so forth. Also we write $g_i(w, s) = \partial g(w, s)/\partial w_i$ and so on.

Using this notation and the integration by parts formula, we have for $1 \leq i, j, k \leq d$ that

$$\begin{aligned} g_{ij}(w, s) &= -\frac{1}{2s} \int_{\mathbb{R}^d} h(\sqrt{1-s}w + \sqrt{s}z) \varphi_{ij}(z) \, dz \\ &= \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} h_j(\sqrt{1-s}w + \sqrt{s}z) \varphi_i(z) \, dz \end{aligned} \tag{4.12}$$

and

$$\begin{aligned} g_{ijk}(w, s) &= \frac{\sqrt{1-s}}{2s^{3/2}} \int_{\mathbb{R}^d} h(\sqrt{1-s}w + \sqrt{s}z) \varphi_{ijk}(z) \, dz \\ &= \frac{\sqrt{1-s}}{2\sqrt{s}} \int_{\mathbb{R}^d} h_{jk}(\sqrt{1-s}w + \sqrt{s}z) \varphi_i(z) \, dz. \end{aligned} \tag{4.13}$$

Lemma 4.3. For each map $a : \{1, \dots, d\}^k \rightarrow \mathbb{R}$, we have

$$\int_{\mathbb{R}^d} \left(\sum_{i_1, \dots, i_k=1}^d a(i_1, \dots, i_k) \frac{\varphi_{i_1 \dots i_k}(z)}{\varphi(z)} \right)^2 \varphi(z) \, dz \leq k! \sum_{i_1, \dots, i_k=1}^d (a(i_1, \dots, i_k))^2. \tag{4.14}$$

Proof. We will prove that

$$\int_{\mathbb{R}^d} \frac{\varphi_{i_1 \dots i_k}(z)}{\varphi(z)} \frac{\varphi_{j_1 \dots j_k}(z)}{\varphi(z)} \varphi(z) \, dz = \sum_{\pi} \delta_{i_{\pi(1)} j_1} \cdots \delta_{i_{\pi(k)} j_k}, \tag{4.15}$$

where the summation is over all permutations of the set $\{1, \dots, k\}$ and δ is the Kronecker delta. By (4.15),

$$\begin{aligned} \int_{\mathbb{R}^d} \left(\sum_{i_1, \dots, i_k=1}^d a(i_1, \dots, i_k) \frac{\varphi_{i_1 \dots i_k}(z)}{\varphi(z)} \right)^2 \varphi(z) \, dz &= \sum_{\pi} \sum_{i_1, \dots, i_k=1}^d a(i_1, \dots, i_k) a(i_{\pi(1)}, \dots, i_{\pi(k)}) \\ &\leq k! \sum_{i_1, \dots, i_k=1}^d (a(i_1, \dots, i_k))^2. \end{aligned}$$

To prove (4.15), we observe that

$$\begin{aligned} &\int_{\mathbb{R}^d} \frac{\varphi_{i_1 \dots i_k}(z)}{\varphi(z)} \frac{\varphi_{j_1 \dots j_k}(z)}{\varphi(z)} \varphi(z) \, dz \\ &= \frac{\partial^{2k}}{\partial x_{i_1} \cdots \partial x_{i_k} \partial y_{j_1} \cdots \partial y_{j_k}} \Big|_{x=y=0} \int_{\mathbb{R}^d} \frac{\varphi(z+x)}{\varphi(z)} \frac{\varphi(z+y)}{\varphi(z)} \varphi(z) \, dz \\ &= \frac{\partial^{2k}}{\partial w_1 \cdots \partial w_{2k}} \Big|_{x=y=0} e^{(x,y)} \end{aligned}$$

where $w_s = x_{i_s}$ and $w_{s+k} = y_{j_s}$ for $s = 1, 2, \dots, k$. By Faà di Bruno’s formula (see Hardy [19]), the latter expression equals

$$\sum_{P_1, \dots, P_m} \frac{\partial^{|P_1|} \langle x, y \rangle}{\prod_{s \in P_1} \partial w_s} \cdots \frac{\partial^{|P_m|} \langle x, y \rangle}{\prod_{s \in P_m} \partial w_s} \Big|_{x=y=0},$$

where the summation is over all unordered partitions of the set $\{1, 2, \dots, 2k\}$ and $|\cdot|$ denotes the cardinality. However, the summand is non-zero if and only if each P_r is of the form $\{s, t + k\}$ where $i_s = j_t$. This proves (4.15). □

Proof of Theorem 2.1. Fix $A \in \mathcal{A}$ and $\varepsilon > 0$ (to be chosen later) and let $f = f_{A,\varepsilon}$ be the solution to the Stein equation (4.1) with respect to $h = h_{A,\varepsilon}$ as defined by (4.2). Let

$$\kappa := d_c(\mathcal{L}(W), \mathcal{L}(Z)).$$

Adding and subtracting the corresponding terms, we have for $g(w, s) = g_{A,\varepsilon}(w, s)$ in (4.11),

$$\begin{aligned} & \mathbb{E}\{\Delta g(W, s) - W^t \nabla g(W, s)\} \\ &= \mathbb{E}\{G^t \nabla g(W', s) - G^t \nabla g(W, s) - W^t \nabla g(W, s)\} \\ & \quad + \sum_{i,j=1}^d \mathbb{E}\{(\delta_{ij} - G_i D_j) g_{ij}(W, s)\} \\ & \quad - \mathbb{E}\left\{ \sum_{i=1}^d G_i g_i(W', s) - \sum_{i=1}^d G_i g_i(W, s) - \sum_{i,j=1}^d G_i D_j g_{ij}(W, s) \right\} \\ &=: R_0(s) + R_1(s) - R_2(s). \end{aligned}$$

As (W, W', G) is a Stein coupling, clearly $R_0(s) \equiv 0$. Therefore, by (4.1),

$$\mathbb{E}h(W) - \mathbb{E}h(Z) = \int_0^1 (R_1(s) - R_2(s)) \, ds.$$

To estimate $\int_0^1 R_1(s) \, ds$, we consider the cases $\varepsilon^2 < s \leq 1$ and $0 < s \leq \varepsilon^2$ separately. For the first case, we use the first expression of $g_{ij}(w, s)$ in (4.12) and obtain

$$\begin{aligned} \int_{\varepsilon^2}^1 R_1(s) \, ds &= \sum_{i,j=1}^d \mathbb{E} \int_{\varepsilon^2}^1 \left(-\frac{1}{2s}\right) \int_{\mathbb{R}^d} [\mathbb{E}^W(\delta_{ij} - G_i D_j)] \\ & \quad \times [h(\sqrt{1-s}W + \sqrt{s}z) - h(\sqrt{1-s}W)] \varphi_{ij}(z) \, dz \, ds, \end{aligned}$$

where we used the fact that $\int_{\mathbb{R}^d} \varphi_{ij}(z) dz = 0$. By the Cauchy–Schwarz inequality, (4.14) and (2.2),

$$\begin{aligned} & \sum_{i,j=1}^d \mathbb{E} \int_{\mathbb{R}^d} [\mathbb{E}^W(\delta_{ij} - G_i D_j)] [h(\sqrt{1-s}W + \sqrt{s}z) - h(\sqrt{1-s}W)] \varphi_{ij}(z) dz \\ & \leq \left\{ \mathbb{E} \int_{\mathbb{R}^d} \left(\sum_{i,j=1}^d [\mathbb{E}^W(\delta_{ij} - G_i D_j)] \frac{\varphi_{ij}(z)}{\varphi(z)} \right)^2 \varphi(z) dz \right\}^{1/2} \\ & \quad \times \left\{ \int_{\mathbb{R}^d} \mathbb{E} [h(\sqrt{1-s}W + \sqrt{s}z) - h(\sqrt{1-s}W)]^2 \varphi(z) dz \right\}^{1/2} \\ & \leq \sqrt{2} B_2 \left\{ \int_{\mathbb{R}^d} \mathbb{E} [h(\sqrt{1-s}W + \sqrt{s}z) - h(\sqrt{1-s}W)]^2 \varphi(z) dz \right\}^{1/2}. \end{aligned}$$

From the definition of κ and the concentration inequality of the standard d -dimensional Gaussian distribution (cf. (4.10)), we have

$$\begin{aligned} & \mathbb{E} \{ h(\sqrt{1-s}W + \sqrt{s}z) - h(\sqrt{1-s}W) \}^2 \\ & \leq \mathbb{E} \{ \mathbb{I}[\text{dist}(\sqrt{1-s}W, A^\varepsilon \setminus A) \leq \sqrt{s}|z|] \} \\ & \leq \mathbb{P} \{ \mathbb{I}[\text{dist}(\sqrt{1-s}Z, A^\varepsilon \setminus A) \leq \sqrt{s}|z|] \} + 2d_c(\mathcal{L}(W), \mathcal{L}(Z)) \\ & \leq 4d^{1/4} \left(\frac{\varepsilon}{\sqrt{1-s}} + 2\sqrt{\frac{s}{1-s}}|z| \right) + 2\kappa. \end{aligned} \tag{4.16}$$

Using the Cauchy–Schwarz inequality, the bound (4.7), the simple inequality $\sqrt{a_1 + a_2 + a_3} \leq \sqrt{a_1} + \sqrt{a_2} + \sqrt{a_3}$ for $a_1, a_2, a_3 \geq 0$, and $\int_{\mathbb{R}^d} |z|^{1/2} \varphi(z) dz \leq d^{1/4}$, we have

$$\begin{aligned} \left| \int_{\varepsilon^2}^1 R_1(s) ds \right| & \leq C B_2 \int_{\varepsilon^2}^1 \frac{1}{s} \int_{\mathbb{R}^d} \left(d^{1/4} \frac{\varepsilon}{\sqrt{1-s}} + d^{1/4} \sqrt{\frac{s}{1-s}}|z| + \kappa \right)^{1/2} \varphi(z) dz ds \\ & \leq C B_2 (d^{1/8} \varepsilon^{1/2} |\log \varepsilon| + d^{3/8} + \kappa^{1/2} |\log \varepsilon|), \end{aligned}$$

where we used $\int_{\varepsilon^2}^1 \frac{1}{s(1-s)^{1/4}} ds \leq C |\log \varepsilon|$ and $\int_{\varepsilon^2}^1 \frac{1}{s^{3/4}(1-s)^{1/4}} ds \leq C$.

For the case $0 < s \leq \varepsilon^2$, we use the second expression of $g_{ij}(w, s)$ in (4.12), the Cauchy–Schwarz inequality, the bound (4.7) and (4.14), and obtain

$$\begin{aligned} & \left| \int_0^{\varepsilon^2} R_1(s) ds \right| \\ & = \left| \sum_{i,j=1}^d \mathbb{E} \int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \int_{\mathbb{R}^d} [\mathbb{E}^W(\delta_{ij} - G_i D_j)] h_j(\sqrt{1-s}W + \sqrt{s}z) \varphi_i(z) dz ds \right| \end{aligned}$$

$$\begin{aligned} &\leq \varepsilon \left| \mathbb{E} \int_{\mathbb{R}^d} \frac{2}{\varepsilon} \sum_{i=1}^d \left\{ \sum_{j=1}^d [\mathbb{E}^W (\delta_{ij} - G_i D_j)]^2 \right\}^{1/2} \frac{\varphi_i(z)}{\varphi(z)} \varphi(z) \, dz \right| \\ &\leq 2\mathbb{E} \left\{ \int_{\mathbb{R}^d} \left\{ \sum_{i=1}^d \left\{ \sum_{j=1}^d [\mathbb{E}^W (\delta_{ij} - G_i D_j)]^2 \right\}^{1/2} \frac{\varphi_i(z)}{\varphi(z)} \right\}^2 \varphi(z) \, dz \right\}^{1/2} \\ &\leq 2B_2, \end{aligned}$$

where the factor ε in the first inequality comes from $\int_0^{\varepsilon^2} \frac{1}{2\sqrt{s}} \, ds \leq \varepsilon$. Therefore,

$$\left| \int_0^1 R_1(s) \, ds \right| \leq CB_2(d^{1/8}\varepsilon^{1/2}|\log \varepsilon| + d^{3/8} + \kappa^{1/2}|\log \varepsilon|).$$

In order to estimate $\int_0^1 R_2(s) \, ds$, let U and V be independent random variables distributed uniformly on $[0, 1]$. Then

$$R_2(s) = \mathbb{E} \sum_{i,j,k=1}^d U G_i D_j D_k g_{ijk}(W + UV D, s).$$

We again consider the cases $\varepsilon^2 < s \leq 1$ and $0 < s \leq \varepsilon^2$ separately.

For the first case, we use the first expression of $g_{ijk}(w, s)$ in (4.13) and obtain

$$\begin{aligned} \int_{\varepsilon^2}^1 R_2(s) \, ds &= \sum_{i,j,k=1}^d \mathbb{E} \int_{\varepsilon^2}^1 \frac{\sqrt{1-s}}{2s^{3/2}} \int_{\mathbb{R}^d} [h(\sqrt{1-s}W + \sqrt{s}z + \sqrt{1-s}UV D) \\ &\quad - h(\sqrt{1-s}W + \sqrt{s}z)] U G_i D_j D_k \varphi_{ijk}(z) \, dz \, ds \\ &\quad + \sum_{i,j,k=1}^d \mathbb{E} \int_{\varepsilon^2}^1 \frac{\sqrt{1-s}}{2s^{3/2}} \int_{\mathbb{R}^d} h(\sqrt{1-s}W + \sqrt{s}z) \\ &\quad \times U [\mathbb{E}^W (G_i D_j D_k) - \mathbb{E}(G_i D_j D_k)] \varphi_{ijk}(z) \, dz \, ds \\ &\quad + \sum_{i,j,k=1}^d \mathbb{E} \int_{\varepsilon^2}^1 \frac{\sqrt{1-s}}{2s^{3/2}} \int_{\mathbb{R}^d} [h(\sqrt{1-s}W + \sqrt{s}z) \\ &\quad - h(\sqrt{1-s}Z + \sqrt{s}z)] U \mathbb{E}(G_i D_j D_k) \varphi_{ijk}(z) \, dz \, ds \\ &\quad + \sum_{i,j,k=1}^d \mathbb{E} \int_{\varepsilon}^1 U \mathbb{E}(G_i D_j D_k) g_{ijk}(Z, s) \, dz \, ds \\ &=: R_{2,1,1} + R_{2,1,2} + R_{2,1,3} + R_{2,1,4}, \end{aligned}$$

where Z is an independent d -dimensional standard Gaussian random vector. Now, it is straightforward to verify that for any $u, v, w, z \in \mathbb{R}^d$

$$\sum_{i,j,k=1}^d u_i v_j w_k \varphi_{ijk}(z) = -u^t z v^t z w^t z \varphi(z) + (u^t v w^t z + u^t w v^t z + v^t w u^t z) \varphi(z). \quad (4.17)$$

In bounding $\int_{\varepsilon^2}^1 R_2(s) ds$, the integration with respect to s is bounded by $\int_{\varepsilon^2}^1 \frac{1}{s^{3/2}} ds \leq C\varepsilon^{-1}$. From (4.17) and the boundedness condition (2.3),

$$\begin{aligned} |R_{2,1,1}| &\leq \mathbb{E} \int_{\varepsilon^2}^1 \frac{\sqrt{1-s}}{2s^{3/2}} \int_{\mathbb{R}^d} \mathbb{1}[\text{dist}(\sqrt{1-s}W + \sqrt{s}z, A^\varepsilon \setminus A) \leq \sqrt{1-s}\beta] \\ &\quad \times U \left| \sum_{i,j,k=1}^d G_i D_j D_k \varphi_{ijk}(z) \right| dz ds \\ &\leq \alpha \int_{\varepsilon^2}^1 \frac{\sqrt{1-s}}{4s^{3/2}} \int_{\mathbb{R}^d} \mathbb{E} \{ \mathbb{1}[\text{dist}(\sqrt{1-s}W + \sqrt{s}z, A^\varepsilon \setminus A) \leq \sqrt{1-s}\beta] \\ &\quad \times [\mathbb{E}|D|^2 + (\mathbb{E}^W |D|^2 - \mathbb{E}|D|^2)] \} (3|z| + |z|^3) \varphi(z) dz ds \\ &\leq Cd^{3/2} \alpha \mathbb{E}|D|^2 \varepsilon^{-1} (\kappa + d^{1/4}(\beta + \varepsilon)) + Cd^{3/2} \varepsilon^{-1} \alpha B_1, \end{aligned}$$

where in the last inequality we used a similar recursive inequality as (4.16) as well as $\int_{\mathbb{R}^d} |z|^3 \varphi(z) dz \leq d^{3/2}$.

From the Cauchy–Schwarz inequality and (4.14),

$$|R_{2,1,2}| \leq C\varepsilon^{-1} B_3.$$

From (4.17) and a recursive inequality as (4.16),

$$|R_{2,1,3}| \leq C(\kappa\varepsilon^{-1} + d^{1/4}) \mathbb{E}(|G||D|^2).$$

For $R_{2,1,4}$, observe that

$$\begin{aligned} \mathbb{E}g(Z + w, s) &= -\frac{1}{2(1-s)} \int_{\mathbb{R}^d} [\mathbb{E}h(\sqrt{1-s}(Z + w) + \sqrt{s}z) - \mathbb{E}h(Z)] \varphi(z) dz \\ &= -\frac{1}{2(1-s)} \int_{\mathbb{R}^d} h(\sqrt{1-s}w + z) \varphi(z) dz + \frac{1}{2(1-s)} \mathbb{E}h(Z) \\ &= -\frac{1}{2(1-s)} \int_{\mathbb{R}^d} h(x) \varphi(x - \sqrt{1-s}w) dx + \frac{1}{2(1-s)} \mathbb{E}h(Z). \end{aligned}$$

Differentiating and evaluating at $w = 0$, we obtain

$$\mathbb{E}g_{ijk}(Z, s) = \frac{\sqrt{1-s}}{2} \int_{\mathbb{R}^d} h(x) \varphi_{ijk}(x) dx.$$

Now with (4.17),

$$|R_{2,1,4}| \leq C\mathbb{E}(|G||D|^2).$$

For the case $0 < s \leq \varepsilon^2$, we use the second expression of g_{ijk} in (4.13). From (4.8) and $|\sum_{i=1}^d G_i \varphi_i(z)| \leq \alpha|z|\varphi(z)$,

$$\begin{aligned} & \left| \int_0^{\varepsilon^2} R_2(s) \, ds \right| \\ &= \left| \sum_{i,j,k=1}^d \mathbb{E} \int_0^{\varepsilon^2} \frac{\sqrt{1-s}}{2s^{1/2}} \int_{\mathbb{R}^d} h_{jk}(\sqrt{1-s}W + \sqrt{s}z + \sqrt{1-s}UV D) U G_i D_j D_k \varphi_i(z) \, dz \, ds \right| \\ &\leq \frac{8\alpha}{\varepsilon^2} \mathbb{E} \int_0^{\varepsilon^2} \frac{\sqrt{1-s}}{2s^{1/2}} \int_{\mathbb{R}^d} \mathbb{I}[\text{dist}(\sqrt{1-s}W + \sqrt{s}z, A^\varepsilon \setminus A) \leq \sqrt{1-s}\beta] \\ &\quad \times [|D|^2 + (\mathbb{E}^W |D|^2 - \mathbb{E}|D|^2)] |z|\varphi(z) \, dz \, ds \\ &\leq Cd^{1/2}\alpha(\mathbb{E}|D|^2(\kappa\varepsilon^{-1} + d^{1/4}\beta\varepsilon^{-1} + d^{1/4}) + \varepsilon^{-1}\sqrt{\text{Var}\mathbb{E}^W|D|^2}), \end{aligned}$$

where in the last inequality we used a similar recursive inequality as (4.16) as well as $\int_{\mathbb{R}^d} |z|\varphi(z) \, dz \leq d^{1/2}$ and $\int_0^{\varepsilon^2} \frac{1}{2s^{1/2}} \, ds \leq \varepsilon$.

Therefore,

$$\left| \int_0^1 R_2(s) \, ds \right| \leq C(d^{3/2}\alpha\mathbb{E}|D|^2\varepsilon^{-1}(\kappa + d^{1/4}(\beta + \varepsilon)) + d^{3/2}\varepsilon^{-1}\alpha B_1 + \varepsilon^{-1}B_3).$$

Collecting the bounds and using the smoothing inequality (4.9), we obtain the following recursive inequality

$$\begin{aligned} \kappa &\leq C\{d^{3/2}\alpha\mathbb{E}|D|^2\varepsilon^{-1}(\kappa + d^{1/4}(\beta + \varepsilon)) + d^{3/2}\varepsilon^{-1}\alpha B_1 \\ &\quad + \varepsilon^{-1}B_3 + B_2(d^{3/8} + d^{1/8}\varepsilon^{1/2}|\log \varepsilon| + \sqrt{\kappa}|\log \varepsilon|)\} + 4d^{1/4}\varepsilon. \end{aligned} \tag{4.18}$$

Let

$$\varepsilon = 2Cd^{3/2}\alpha\mathbb{E}|D|^2 + \beta + d^{5/8}\alpha^{1/2}B_1^{1/2} + d^{1/8}B_2 + d^{-1/8}B_3^{1/2}$$

with the same constant C as in (4.18). The theorem is proved by solving the recursive inequality for κ and observing that as long as ε is smaller than an absolute constant, $\varepsilon^{1/2}|\log \varepsilon| \leq C$ and $\kappa^{1/2}|\log \varepsilon| \leq Cd^{1/8}$, the latter follows by solving the recursive inequality for κ by upper bounding $\sqrt{\kappa}$ in (4.18) by 1. □

Proof of Corollary 2.2. We apply Theorem 2.1 to the Stein coupling

$$(\Sigma^{-1/2}W, \Sigma^{-1/2}W', \Sigma^{-1/2}G).$$

The first two terms in the bound (2.5) are obtained by $|\Sigma^{-1/2}G| \leq s_2|G|$ and $|\Sigma^{-1/2}D| \leq s_2|D|$. For the last three terms, we first observe that for a fixed $d \times d$ orthogonal matrix $U = (U_1, \dots, U_d)^t$, a d -dimensional random vector V and a random variable X ,

$$\begin{aligned} \sum_{i=1}^d \text{Var} \mathbb{E}^W \{(UV)_i X\} &= \sum_{i=1}^d \text{Var} \mathbb{E}^W \{U_i^t V X\} = \sum_{i=1}^d \text{Var} \{U_i^t \mathbb{E}^W \{V X\}\} \\ &= \sum_{i=1}^d U_i^t \text{Cov}(\mathbb{E}^W \{V X\}) U_i = \text{Tr}(U \text{Cov}(\mathbb{E}^W \{V X\}) U^t) \\ &= \text{Tr}(\text{Cov}(\mathbb{E}^W \{V X\})) = \sum_{i=1}^d \text{Var} \{\mathbb{E}^W \{V_i X\}\}. \end{aligned}$$

Therefore, B_1, B_2 and B_3 remain unchanged if we replace G and D by UG and UD . Next, we write $\Sigma^{-1/2} = U\Lambda U^t$ where U is an orthogonal matrix and Λ is a diagonal matrix whose components are bounded by s_2 by definition. Finally, the last three terms in the bound (2.5) are obtained by

$$\begin{aligned} \sum_{i=1}^d \text{Var} \mathbb{E}^W \{(\Sigma^{-1/2}V)_i X\} &= \sum_{i=1}^d \text{Var} \mathbb{E}^W \{(U\Lambda U^t V)_i X\} \\ &= \sum_{i=1}^d \text{Var} \mathbb{E}^W \{(\Lambda U^t V)_i X\} \leq s_2^2 \sum_{i=1}^d \text{Var} \mathbb{E}^W \{(U^t V)_i X\} \\ &= s_2^2 \sum_{i=1}^d \text{Var} \mathbb{E}^W \{V_i X\}. \end{aligned} \quad \square$$

Sketch of the proof for Remark 2.3. Let U and V be uniform on $[0, 1]$, independent of each other and all else. Under the conditions of Remark 2.3, we have from Taylor expansion that

$$\begin{aligned} 0 &= \lambda^{-1} \mathbb{E} \{f(W') - f(W)\} \\ &= \lambda^{-1} \mathbb{E} \sum_{i=1}^d (W'_i - W_i) f_i(W) + \lambda^{-1} \mathbb{E} \sum_{i,j=1}^d U D_i D_j f_{ij}(W + UV D) \\ &= -\mathbb{E} \sum_{i=1}^d W_i f_i(W) + \mathbb{E} \sum_{i,j=1}^d G_i D_j f_{ij}(W) \\ &\quad - 2\mathbb{E} \sum_{i,j=1}^d U G_i D_j (f_{ij}(W) - f_{ij}(W + UV D)). \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}\{\Delta f(W) - W^t \nabla f(W)\} &= \sum_{i,j=1}^d \mathbb{E}\{(\delta_{ij} - G_i D_j) f_{ij}(W)\} \\ &\quad + 2\mathbb{E} \sum_{i,j=1}^d U G_i D_j (f_{ij}(W) - f_{ij}(W + UVD)) \\ &=: R'_1 - R'_2. \end{aligned}$$

The quantity R'_1 is the same as $\int_0^1 R_1(s) ds$ in the proof of Theorem 2.1. The quantity R'_2 contains an additional integration step as compared to $\int_0^1 R_2(s) ds$ of Theorem 2.1, but can be bounded in very much the same way (up to different constants). \square

5. Some Stein couplings

In this section, we describe some known coupling constructions as multivariate Stein couplings for reference.

5.1. Multivariate exchangeable pairs

Chatterjee and Meckes [10] and Reinert and Röllin [24] introduced the exchangeable pairs method for random vectors, which are instances of Stein couplings. Assume that (W, W') is an exchangeable pair of d -dimensional random vectors such that

$$\mathbb{E}^W(W' - W) = -\Lambda W \tag{5.1}$$

for some invertible $(d \times d)$ -matrix Λ . It is straightforward to check that

$$(W, W', G) := (W, W', \frac{1}{2}\Lambda^{-1}(W' - W))$$

is a Stein coupling.

Assume $\text{Var}(W) = \Sigma$ is positive definite. Let $\Sigma^{1/2}$ be the unique positive-definite root of Σ , and let $\Sigma^{-1/2}$ be its corresponding unique inverse. It was shown by Reinert and Röllin [24] that exchangeability of (W, W') implies symmetry of $\hat{\Lambda} = \Sigma^{-1/2}\Lambda\Sigma^{1/2}$. Let therefore O be an orthonormal matrix and let L be a positive diagonal matrix such that $\hat{\Lambda} = OLO^t$. Define $\hat{W} = O^t\Sigma^{-1/2}W$, $\hat{W}' = O^t\Sigma^{-1/2}W'$. It follows from (5.1) that

$$\mathbb{E}^{\hat{W}}(\hat{W}' - \hat{W}) = -L\hat{W}.$$

We could therefore – in principle – restrict ourselves to (W, W') that are uncorrelated with (5.1) being true for diagonal Λ . However, it is often much easier to work with the unstandardized W , as $\Sigma^{-1/2}$ and O are typically difficult to calculate in practice.

5.2. Multivariate size bias couplings

This coupling was considered by Goldstein and Rinott [17]. Let Y be a non-negative d -dimensional random vector with mean μ and covariance matrix Σ . For each $i = \{1, \dots, d\}$, let Y^i be defined on the same probability space as Y and have Y -size biased distribution in direction i , that is,

$$\mathbb{E}\{Y_i f(Y)\} = \mu_i \mathbb{E}f(Y^i)$$

for all functions f such that the expectations exist. Let K be uniformly distributed over $\{1, 2, \dots, d\}$, independent of all else, and let e_K be the d -dimensional unit vector in direction K . Then

$$(W, W', G) := (Y - \mu, Y^K - \mu, d\mu_K e_K)$$

is a Stein coupling.

5.3. Local dependence

A refined version of this dependence was considered by Rinott and Rotar' [25]. Let $(X_i)_{i \in \mathcal{I}}$ be a collection of centered d -dimensional random vectors for some finite index set \mathcal{I} . For each $i \in \mathcal{I}$, assume there is a set $A_i \subset \mathcal{I}$ such that X_i is independent of $(X_j)_{j \in A_i^c}$. Let I be uniformly distributed on \mathcal{I} , independent of all else. Then

$$(W, W', G) := \left(\sum_{i \in \mathcal{I}} X_i, \sum_{i \in \mathcal{I} \setminus A_I} X_i, -nX_I \right) \tag{5.2}$$

is a Stein coupling.

We have the following corollary of Theorem 2.1 for the Stein coupling (5.2). The proof is straightforward and therefore omitted here.

Corollary 5.1. *Let $(X_i)_{i \in \mathcal{I}}$ be a collection of centered d -dimensional random vectors for some finite index set \mathcal{I} with cardinality n . For each $i \in \mathcal{I}$, assume there are sets $A_i \subset B_i \subset \mathcal{I}$ such that X_i is independent of $(X_j)_{j \in A_i^c}$ and $(X_j)_{j \in A_i}$ is independent of $(X_j)_{j \in B_i^c}$. Assume further that*

$$|X_i| \leq \beta, \quad \left| \{j \in \mathcal{I} : (A_j \cap B_i) \cup (A_i \cap B_j) \neq \emptyset\} \right| \leq c_d,$$

where c_d is a constant only depending on d and $|\cdot|$ denotes cardinality. Then we have

$$d_c(\mathcal{L}(W), \mathcal{L}(Z)) \leq C_d \beta^3 n$$

for some constant C_d only depending on d .

Under the condition of the above corollary, the result in Rinott and Rotar' [25] yields the bound

$$d_c(\mathcal{L}(W), \mathcal{L}(Z)) \leq C_d \beta^3 n \log n,$$

which has an additional logarithmic factor.

Note that if the summands are locally dependent, but highly uncorrelated, that is, if $\mathbb{E}(X_i X_j^t) = 0$ “for many” $j \in A_i$, it seems difficult to obtain informative bounds from Rinott and Rotar’ [25]. For example, if one tries to apply their Theorem 2.1 to dense random graphs in Section 3.1, in order for χ_1 in their (2.2) to be small, U_j in their Theorem 2.1 has to be of order n^2 (recall n is the number of vertices in Theorem 3.3), this makes A_1 in their Theorem 2.1 too large for their bound (2.3) to converge to 0. In contrast, our Theorem 2.1 can yield informative bounds in such cases.

Acknowledgments

We are grateful to Louis Chen for fruitful discussions that lead to this research and to Persi Diaconis for the suggestion to study the joint distribution of inversions and descents in random permutations. We also thank the two anonymous referees for their detailed comments which led to many improvements, in particular the correction of an error in Lemma 3.9 in an earlier version of the manuscript.

X. Fang was supported by NUS Research Grants C-389-000-010-101 and C-389-000-012-101. A. Röllin was partially supported by NUS Research Grants R-155-000-098-133 and R-155-000-124-112.

References

- [1] Aldous, D.J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. [MR0637937](#)
- [2] Ball, K. (1993). The reverse isoperimetric problem for Gaussian measure. *Discrete Comput. Geom.* **10** 411–420. [MR1243336](#)
- [3] Barbour, A.D. (1990). Stein’s method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322. [MR1035659](#)
- [4] Barbour, A.D., Karoński, M. and Ruciński, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47** 125–145. [MR1047781](#)
- [5] Bentkus, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. [MR1965117](#)
- [6] Bhattacharya, R.N. and Holmes, S. (2010). An exposition of Götze’s estimation of the rate of convergence in the multivariate central limit theorem. Available at [arXiv:1003.4254](#).
- [7] Bollobás, B. and Riordan, O. (2009). Metrics for sparse graphs. In *Surveys in Combinatorics 2009. London Mathematical Society Lecture Note Series* **365** 211–287. Cambridge: Cambridge Univ. Press. [MR2588543](#)
- [8] Borgs, C., Chayes, J.T., Lovász, L., Sós, V.T. and Vesztegombi, K. (2008). Convergent sequences of dense graphs I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851. [MR2455626](#)
- [9] Borgs, C., Chayes, J.T., Lovász, L., Sós, V.T. and Vesztegombi, K. (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math. (2)* **176** 151–219. [MR2925382](#)
- [10] Chatterjee, S. and Meckes, E. (2008). Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 257–283. [MR2453473](#)
- [11] Chen, L.H.Y. and Röllin, A. (2010). Stein couplings for normal approximation. Preprint. Available at [arXiv:1003.6039](#).

- [12] Chung, F.R.K., Graham, R.L. and Wilson, R.M. (1989). Quasi-random graphs. *Combinatorica* **9** 345–362. [MR1054011](#)
- [13] Diaconis, P. (1988). *Group Representations in Probability and Statistics. Institute of Mathematical Statistics Lecture Notes – Monograph Series* **11**. Hayward, CA: IMS. [MR0964069](#)
- [14] Diaconis, P. and Janson, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** 33–61. [MR2463439](#)
- [15] Fang, X. (2012). Multivariate, combinatorial and discretized normal approximations by Stein’s methods. Ph.D. thesis, National Univ. Singapore.
- [16] Fulman, J. (2004). Stein’s method and non-reversible Markov chains. In *Stein’s Method: Expository Lectures and Applications. Institute of Mathematical Statistics Lecture Notes – Monograph Series* **46** 69–77. Beachwood, OH: IMS. [MR2118603](#)
- [17] Goldstein, L. and Rinott, Y. (1996). Multivariate normal approximations by Stein’s method and size bias couplings. *J. Appl. Probab.* **33** 1–17. [MR1371949](#)
- [18] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. [MR1106283](#)
- [19] Hardy, M. (2006). Combinatorics of partial derivatives. *Electron. J. Combin.* **13** 1–13 (electronic). [MR2200529](#)
- [20] Janson, S. and Nowicki, K. (1991). The asymptotic distributions of generalized U -statistics with applications to random graphs. *Probab. Theory Related Fields* **90** 341–375. [MR1133371](#)
- [21] Lovász, L. and Szegedy, B. (2011). Finitely forcible graphons. *J. Combin. Theory Ser. B* **101** 269–301. [MR2802882](#)
- [22] Pao, H., Coppersmith, G.A. and Priebe, C.E. (2011). Statistical inference on random graphs: Comparative power analyses via Monte Carlo. *J. Comput. Graph. Statist.* **20** 395–416. [MR2847801](#)
- [23] Raič, M. (2004). A multivariate CLT for decomposable random vectors with finite second moments. *J. Theoret. Probab.* **17** 573–603. [MR2091552](#)
- [24] Reinert, G. and Röllin, A. (2009). Multivariate normal approximation with Stein’s method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37** 2150–2173. [MR2573554](#)
- [25] 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. [MR1379533](#)
- [26] Zhao, L., Bai, Z., Chao, C.-C. and Liang, W.-Q. (1997). Error bound in a central limit theorem of double-indexed permutation statistics. *Ann. Statist.* **25** 2210–2227. [MR1474091](#)

Received June 2012 and revised April 2014