JOURNAL OF MULTIVARIATE ANALYSIS **56**, 333–350 (1996)
ARTICLE NO. 0017

A Multivariate CLT for Local Dependence with $n^{-1/2} \log n$ Rate and Applications to Multivariate Graph Related Statistics

YOSEF RINOTT*

University of California, San Diego

AND

VLADIMIR ROTAR

Central Economic-Mathematical Institute, Russian Academy of Sciences, Moscow, Russia, and University of California, San Diego

This paper concerns the rate of convergence in the central limit theorem for certain local dependence structures. The main goal of the paper is to obtain estimates of the rate in the multidimensional case. Certain one-dimensional results are also improved by using some more flexible characteristics of dependence. Assuming the summands are bounded, we obtain rates close to those for independent variables. As an application we study the rate of the normal approximation of certain graph related statistics which arise in testing equality of several multivariate distributions. © 1996 Academic Press, Inc.

1. INTRODUCTION

This paper concerns central limit theorems (CLTs) for sums of dependent random vectors, when the dependence structure is described in terms of *dependency neighborhoods*. This type of dependence is given by indicating for every term in the sum a set of other terms on which it "essentially" depends. Such a structure need not be associated with a linear ordering of the summands. This may be compared with more classical CLTs in which the dependence is specified in terms of an ordering (despite the fact that the sum itself is invariant under permutations), such as Markov chains, martingales or various mixing models.

Received December 9, 1994.

MSC 1991 subject classifications: primary; 60F05, 60B12; secondary; 62H15, 04C80.
Key words and phrases: Stein's method, random graphs, multivariate statistics.

* This work was supported in part by NSF Grant DMS 92-05759.

0047-259X/96 $18.00

Dependency neighborhoods were introduced by Stein [21, Corollary 2, p. 110], in the study of normal approximations. Estimates of the rate of such an approximation for nonsmooth functions and bounded random variables were improved in Rinott [17], and a multivariate version for smooth functions only, appears in Goldstein and Rinott [9]. These results prove to be useful in applications in which the dependence is "local" in the sense that the dependency neighborhoods are not too large. This occurs typically for various counts on graphs, where parts of the graphs that are disconnected are almost independent. For such applications, where the present methods are relevant see, for example, [1, 8, 9] and references given there.

The dependence structure in this paper is described in terms of two types of sets, referred to as first-order and second-order dependency neighborhoods. These sets may be random. See Theorem 2.2 and its discussion. In the one-dimensional case, under this dependence structure, we provide a more flexible and somwhat more transparent estimate of the convergence rate in comparison to the above mentioned results, covering the case of nonsmooth functions.

The main goal of this paper is the study of the approximation rates for the same dependence structure in the multidimensional case. In general the convergence rate for dependent random variables may be rather slow. We assume here that the summands are bounded. This assumption enables us to obtain a rate of convergence close to that for independent random variables. The fact that boundedness diminishes the effect of dependence on the convergence rate is well known (see, e.g., [6]).

The proofs in this paper are based on Stein's method. The extension of these methods to nonsmooth functions in the multivariate case is rather nontrivial. We proceed by the methodology developed by Barbour [3] and Götze [10].

The main results and some discussions are given in Section 2. An application to graph related statistics, motivated by hypothesis testing of equality of multivariate distributions, is given in Section 3. The proof of Theorem 2.1 is given in Section 4.

2. Results and Discussion

Let Φ denote the standard normal distribution in R^d, and ϕ the corresponding density function. Given a matrix (or a vector) A, let A^{T} be the transpose, and let $|A|$ the be sum of the absolute values of its components. Using this norm rather than Euclidean norm will simplify some notations and the formulation of the results. This is only a matter of convenience since the norms are equivalent.

Given a function $h: R^d \to R$, set

$$h_\delta^+(x) = \sup\{h(x+y): |y| \leqslant \delta\}, \qquad h_\delta^-(x) = \inf\{h(x+y): |y| \leqslant \delta\},$$
$$\tilde{h}(x; \delta) = h_\delta^+(x) - h_\delta^-(x).$$

Let $\mathscr{H}$ be a class of measurable functions from R^d to R, which are uniformly bounded by some constant assumed to be 1 without loss of generality. Suppose $\mathscr{H}$ satisfies the following properties: for any $h \in \mathscr{H}$ and any $\delta > 0$, the functions $h_\delta^+(x)$ and $h_\delta^-(x)$ are in $\mathscr{H}$, and for any $d \times d$ matrix A and any vector $b \in R^d$, the function $h(Ax+b)$ belongs to $\mathscr{H}$. In addition we assume that for all $\delta > 0$

$$\sup\left\{\int_{R^d} \tilde{h}(x; \delta)\, \Phi(dx): h \in \mathscr{H}\right\} \leqslant a\delta \tag{2.1}$$

for some constant a which depends only on the class $\mathscr{H}$ and the dimension d. Obviously, we may assume $a \geqslant 1$.

The class of indicators of convex sets is known to be such a class (see, e.g., [19, 5]).

Theorem 2.1 below provides a normal approximation for a sum of random vectors in terms of certain decompositions of the sum. This theorem is somewhat formal and should be read together with its natural corollary, Theorem 2.2, which motivates it and is used later in the applications.

Given random vectors $\mathbf{X}_j$ taking values in R^d, we study the proximity of the distribution of the random vector $\mathbf{W} = \sum_{j=1}^n \mathbf{X}_j$ to the normal distribution. We assume throughout that the summands are bounded, that is, $|\mathbf{X}_j| \leqslant B$, $1 \leqslant j \leqslant n$ for some constant B. In particular applications, the constant B, and others below, may be considered to depend on n.

As a measure of the proximity to the standard normal distribution we consider $\sup\{|Eh(\mathbf{W}) - \Phi h|: h \in \mathscr{H}\}$, where $\Phi h = \int_{R^d} h(z)\, \Phi(dz)$.

In the case $d = 1$, the convergence rate of our theorem is better (by $\log n$) than the rate for $d > 1$. For this reason, we state a bound in the univariate case separately.

THEOREM 2.1. *For each $j = 1, ..., n$ assume that we have two representations of $\mathbf{W}$, $\mathbf{W} = \mathbf{U}_j + \mathbf{V}_j$ and $\mathbf{W} = \mathbf{R}_j + \mathbf{T}_j$, such that $|\mathbf{U}_j| \leqslant A_1$, and $|\mathbf{R}_j| \leqslant \mathbf{A}_2$ for constants satisfying $A_1 \leqslant A_2$. Define*

$$\chi_1 = \sum_{j=1}^n E\,|E(\mathbf{X}_j \mid \mathbf{V}_j)|, \qquad \chi_2 = \sum_{j=1}^n E\,|E(\mathbf{X}_j \mathbf{U}_j^{\mathrm{T}}) - E(\mathbf{X}_j \mathbf{U}_j^{\mathrm{T}} \mid \mathbf{T}_j)|,$$
$$\chi_3 = \left| I - \sum_{j=1}^n E(\mathbf{X}_j \mathbf{U}_j^{\mathrm{T}}) \right|, \tag{2.2}$$

where I denotes the identity matrix. Then for $d=1$ there exists a universal constant c such that

$$\sup\{|Eh(W)-\Phi h|: h\in\mathscr{H}\} \leqslant c\{aA_2+n(a+\sqrt{EW^2})\,A_1A_2B+\chi_1+\chi_2+\chi_3\}. \tag{2.3}$$

For $d\geqslant 1$ there exists a constant c depending only on the dimension d such that

$$\sup\{|Eh(\mathbf{W})-\Phi h|: h\in\mathscr{H}\} \leqslant c\{aA_2+naA_1A_2B(|\log A_2B|+\log n) +\chi_1+(|\log A_1B|+\log n)(\chi_2+\chi_3)\}. \tag{2.4}$$

For independent $\mathbf{X}_j$ the natural decomposition is $\mathbf{U}_j=\mathbf{R}_j=\mathbf{X}_j$. In the case of dependence, other decompositions prove to be more effective as shown below. It may seem that the choice of $\mathbf{V}_j$ and $\mathbf{T}_j$ is arbitrary. However, if, for example, the $\mathbf{T}_j$'s are constant, then χ_2 vanishes, but then A_2 may be large. Similarly, if $E\mathbf{W}=0$ and $\operatorname{Var}\mathbf{W}=I$, the choice $\mathbf{V}_j=0$ leads to $\chi_3=0$ but then A_1 may be large. Useful decompositions related to the dependence structure are considered below in detail.

Note that in the theorem there is no assumption on the mean and variance of $\mathbf{W}$. However, since the distribution of $\mathbf{W}$ is being compared to a standard normal distribution, we obviously have in mind the case where $E\mathbf{W}$ vanishes and $\operatorname{Var}\mathbf{W}$ is the identity matrix I, or approximately so. Since $\mathscr{H}$ is closed under affine transformations, we can always standardize $\mathbf{W}$ (exactly or approximately) before applying the theorem. For $d=1$, when $EW^2=1$, or if EW^2 is bounded by a universal constant, we may simply omit this term from the r.h.s. of (2.3) by absorbing it into c, since we assume $a\geqslant 1$.

Theorem 2.1 appears rather abstract. It was motivated by the following theorem which suggests natural decompositions in terms of the dependency structure. The decomposition is given by sets S_i and $\mathscr{N}_i$ defined below, which may be random.

THEOREM 2.2. *Let S_i and $\mathscr{N}_i$ be subsets of $\{1, ..., n\}$, which in general may be random, such that $i\in S_i\subseteq\mathscr{N}_i$, $i=1, ..., n$. Assume that there exist constants $D_1\leqslant D_2$, such that $\max\{|S_i|: i=1, ..., n\}\leqslant D_1$, and $\max\{|\mathscr{N}_i|: i=1, ..., n\}\leqslant D_2$, where for sets $|\cdot|$ denotes cardinality. Denote*

$$\mathbf{U}_j=\sum_{k\in S_j}\mathbf{X}_k, \qquad \mathbf{V}_j=\mathbf{W}-\mathbf{U}_j,$$
$$\mathbf{R}_j=\sum_{k\in\mathscr{N}_j}\mathbf{X}_k, \qquad \mathbf{T}_j=\mathbf{W}-\mathbf{R}_j, \qquad j=1, ..., n. \tag{2.5}$$

Then for $d=1$ there exists a universal constant c such that

$$\sup\{|Eh(W)-\Phi h|: h\in\mathscr{H}\} \leqslant c\{aD_2B+n(a+\sqrt{EW^2})\,D_1D_2B^3+\chi_1+\chi_2+\chi_3\}, \tag{2.6}$$

where χ_i are defined in (2.2). *For $d\geqslant 1$ there exists a constant c depending only on the dimensional d such that*

$$\sup\{|Eh(\mathbf{W})+\Phi h|: h\in\mathscr{H}\} \leqslant c\{aD_2B+naD_1D_2B^3(|\log B|+\log n) +\chi_1+(|\log B|+\log n)(\chi_2+\chi_3)\}. \tag{2.7}$$

Theorem 2.1 follows from Theorem 2.2 simply by observing that the quantities defined in (2.5) satisfy the assumptions with $A_1=D_1B$, and $A_2=D_2B$. Also, since $\log A_2$ appears in (2.4), it would seem that the quantity $\log D_2$ should appear in (2.7); however, it may be omitted since $D_2\leqslant n$.

Remarks. In general, the dependence structure affects the convergence rate in central limit theorems for dependent random variables through three types of characteristics or conditions. The first type, reflected in χ_1 above, concerns conditional expectations with respect to appropriate fields, which should be small or vanish, as for example in martingales or for exchangeable random variables, where this type of condition appears in terms of correlations. The second type, reflected by χ_2 and χ_3, concerns conditional variances, which should be close in some sense to the unconditional ones. The third and main type reflects special aspects of the dependence structure. Here, this term involves the bound B on the summands, and the bounds D_1 and D_2 on the sizes of the so-called dependency neighborhoods S_i and $\mathscr{N}_i$. A discussion of characteristics of this nature may be found for example in Jacod and Shiryayev [14], Liptser and Shiryayev [16], and Rotar [18]. Next, we consider the above characteristics in more detail.

Formally Theorem 2.2 holds for arbitrary sets S_i and $\mathscr{N}_i$. Clearly, one should choose them so as to minimize the bound given by the theorem, which means that both the χ_i's and the D_i's should be small. In order to clarify the nature of such a choice, assume now that we have standardized $\mathbf{W}$, so that $E\mathbf{W}=0$ and $\operatorname{Var}\mathbf{W}=I$, and consider first the case where for each i there exists S_i such that $\mathbf{X}_i$ is independent of $\{\mathbf{X}_j: j\notin S_i\}$. In this case $\chi_1=0$ and $\chi_3=0$. Suppose also that there exist sets $\mathscr{N}_i$ such that, if k and l are both in S_i, then the collections of variables $\{\mathbf{X}_k,\mathbf{X}_l\}$ and $\{\mathbf{X}_j: j\notin\mathscr{N}_i\}$ are independent. In this case $\chi_2=0$. It may even happen that a stronger kind of independence takes place; that is, the collections $\{\mathbf{X}_j: j\in S_i\}$ and $\{\mathbf{X}_j: j\notin\mathscr{N}_i\}$ are independent. In any case, it is natural to view $\mathscr{N}_i$ as a dependency neighborhood of $\{\mathbf{X}_j: j\in S_i\}$, or a second-order dependency

neighborhood of $\mathbf{X}_i$. In nonexotic situations one may expect $\mathcal{N}_i = \bigcup_{j \in S_i} S_j$ and then $D_2 \leqslant D_1^2$. However, this is not the case in general; for the latter $\mathcal{N}_i$, the present assumptions on the S_i's imply that the above $\mathbf{X}_k$ is independent of $\{\mathbf{X}_j : j \notin \mathcal{N}_i\}$, and the same holds for $\mathbf{X}_l$, but this does not necessarily imply the independence of $\{\mathbf{X}_k, \mathbf{X}_l\}$ and $\{\mathbf{X}_j : j \notin \mathcal{N}_i\}$.

In general, the choice described above may not be feasible, and one may try to achieve an approximation to this situation, that is, to find sets S_i and $\mathcal{N}_i$, "essential dependency sets," such that the χ_i's do not vanish but are small and the D_i's are still not large.

A dependence structure similar to that of Theorem 2.2, with two types of dependency sets, was used before; see, for example, Barbour, Holst, and Janson [4], who considered Poisson approximations, and references there.

Turning to B, note that in many applications one starts with a sum of random variables Y_i, such that $\mathrm{Var}(Y_1 + \cdots + Y_n)$ has the oder n (or nI in the multivariate case), and the Y_i's are bounded. Setting $X_i = Y_i/\sqrt{n}$ to apply the theorems, we see that the bound B for the X_i's will have the order $1/\sqrt{n}$. If D_1 and D_2 are bounded, then the part of the bound (2.7) not involving the χ's has the order $n^{-1/2} \log n$.

Clearly, if $E\mathbf{W} = \boldsymbol{\lambda}$, and $\mathrm{Var}\, \mathbf{W} = \Sigma$, we can apply the theorems to $\Gamma^{-1/2}(\mathbf{W} - \boldsymbol{\lambda})$, where Γ equals or approximates Σ. By the invariance of $\mathscr{H}$ under affine transformations, we then obtain a bound on $\sup_{h \in \mathscr{H}} |Eh(\mathbf{W}) - Eh(\Gamma^{1/2}\mathbf{Z} + \boldsymbol{\lambda})|$, where $\mathbf{Z}$ is a standard normal vector in R^d.

We shall consider statistical applications in detail in Section 3. Here we briefly illustrate the dependency structure by an application (with $d = 1$) due to Baldi and Rinott [2]. Some of the rates in the latter paper can be improved by using Theorem 2.2. Consider a random ranking of the $n = 2^m$ vertices of the hypercube $\{0, 1\}^m$. Let Y_i be the indicator of the event that the rank of the ith vertex exceeds that of all the m neighboring vertices (a local maximum). The total number of local maxima is $M = \sum_{i=1}^n Y_i$. It can be shown that if the distance between any two vertices i and j is 3 or more, that is, if the vertices differ in three coordinates or more when viewed as m-vectors of 0's and 1's, then Y_i and Y_j are independent. More generally any collections of $\{Y_j : j \in A_i\}$, $i = 1, 2$, are independent if the distance between the vertex sets A_1 and A_2 is at least 3. Thus, we may choose S_i consisting of the $\binom{m}{2} + m$ vertices of distance at most 2 from i. This results in D_1 which is of the order m^2, or $(\log n)^2$. Setting $\mathcal{N}_i = \bigcup_{j \in S_i} S_j$, we have $D_2 \leqslant D_1^2$ and $\chi_i = 0$, $i = 1, 2, 3$. Here σ_M^{-1} and hence B have the order $(\log n/n)^{1/2}$. For $W = \sum_{i=1}^n (Y_i - EY_i)/\sigma_M$, Theorem 2.2 provides a normal approximation with the rate of $(\log n)^{7.5} n^{-1/2}$.

In the above example a natural linear ordering of the variables does not exist and the dependency neighborhoods are determined by distances on a graph. Note also that the D_i's are not bounded, but increase in n. This

suggests a general scheme, where the dependence structure is described by a graph whose vertices correspond to the random summands, the dependence between two summands is defined as a function of the distance between corresponding vertices, and natural dependency sets arise as "geographical" neighborhoods of vertices. If for example, the dependence as reflected by such a function of the distance decays exponentially, one may choose dependency sets with diameter of the order $c \log n$, and for an appropriate c the characteristics χ_i will have the order $n^{-1/2}$, whereas the constants D_i will have the oder of a power of $\log n$, leading to a convergence rate of $(\log n)^k n^{-1/2}$ for some k.

3. Applications to Graph Related Statistics

Nonrandom Graphs. For the applications described later we need to consider random graphs; however, we start with nonrandom graphs for ease of exposition. Consider a fixed regular graph, with n vertices and vertex degree m. The regularity implies that the number of edges in the graph is $N = nm/2$. Suppose that each vertex is independently assigned one of colors c_i with probability π_i, $i = 1, ..., d$, satisfying $\sum_{i=1}^{d} \pi_i = 1$. Let $\mathbf{W} = (W_1, ..., W_d)$, where W_i, $i = 1, ..., d$, is the number of edges connecting vertices which are both of color c_i; that is, $W_i = \sum_{j=1}^{N} X_{ji}$, where X_{ji} is the indicator of the event that both vertices associated with the edge j have the color c_i. Set $\boldsymbol{\lambda} = E\mathbf{W} = N(\pi_1^2, ..., \pi_p^2)$. The entries of $\Sigma = (\sigma_{ij})$, the covariance matrix of $\mathbf{W}$, satisfy (see [9], hereafter called GR)

$$\begin{aligned} \sigma_{ii} &= \mathrm{Var}(W_i) = N\pi_i^2(1-\pi_i^2) + 2N(m-1)(\pi_i^3 - \pi_i^4); \\ \sigma_{ij} &= \mathrm{Cov}(W_i, W_j) = -N(2m-1)\,\pi_i^2\pi_j^2, \qquad \text{for} \quad i \neq j. \end{aligned} \tag{3.1}$$

Let $L = [\min_{1 \leqslant i \leqslant d} \{\pi_i^2(1-\pi_i)\}]^{-1/2}$. Given a matrix A, let $\|A\|$ denote the maximal absolute value of its entries. It is proved in GR that $\|\Sigma^{-1/2}\| \leqslant N^{-1/2}L$, and, hence, we can apply Theorem 2.2 to $\Sigma^{-1/2}(\mathbf{W} - \boldsymbol{\lambda})$, with the bound on the summands $B = dN^{-1/2}L$. For any edge j we choose a neighborhood S_j consisting of all edges which share a vertex with j. With the natural choice $\mathcal{N}_i = \bigcup_{j \in S_i} S_j$ we have the bounds $D_1 = 2m-1$ and $D_2 = (2m-1)^2$ on the cardinality of S_j and $\mathcal{N}_i$, respectively, and $\chi_1 = \chi_2 = \chi_3 = 0$.

For functions $h \in \mathcal{H}$, using the closure of $\mathcal{H}$ under affine transformations and the facts $m \leqslant n$ and $L \geqslant 1$, we derive from (2.7),

$$\begin{aligned} &\sup\{|Eh(\mathbf{W}) - Eh(\Sigma^{1/2}\mathbf{Z} + \boldsymbol{\lambda})|: h \in \mathcal{H}\} \\ &\qquad \leqslant cam^{3/2}L^3(|\log L| + \log n)n^{-1/2}, \end{aligned} \tag{3.2}$$

where $\mathbf{Z}$ is a d-dimensional standard normal vector, c is a constant depending on the dimension d, and a is defined in (2.1). Such a result, with rate $n^{-1/2}$, but only for smooth functions h was obtained in GR.

Random Nearest Neighbor Graphs. Consider a sample of n i.i.d. points from an absolutely continuous distribution F in R^k. Let G denote the nearest neighbor graph whose n vertices are these points. This is a directed graph such that from each vertex there is a directed edge pointing to its nearest neighbor (with respect to Euclidean distance, say). As before each vertex is independently assigned one of the colors c_i with probability π_i, $i=1, ..., d$. For a vertex j, let $N(j)$ denote its nearest neighbor. Let $X_{ji}=1$ if the vertices j and $N(j)$ are both assigned the color c_i, and 0 otherwise, $j=1, ..., n$; $i=1, ..., d$. Then $W_i=\sum_{j=1}^n X_{ji}$ counts the number of vertices having the color c_i as well as their nearest neighbors (with mutual nearest neighbors counted twice, once for each vertex). Set $\mathbf{W}=(W_1, ..., W_d)$, and $\boldsymbol{\lambda}=E\mathbf{W}=n(\pi_1^2, ..., \pi_d^2)$. We now calculate the covariance matrix of $\mathbf{W}$, which we denote by Σ_n. For a given realization of the graph G, let $D(j)$ denote the degree of the vertex j, that is, the number of edges pointing to j plus the one edge emanating from j. Let A denote the number of pairs of vertices which are mutual (nearest) neighbors. For the conditional variance of W_i given the graph G we claim that

$$\operatorname{Var}(W_i \mid G)=n\pi_i^2(1-\pi_i^2)+\sum_{j=1}^{n}\binom{D(j)}{2}(\pi_i^3-\pi_i^4)+2A(\pi_i^2-\pi_i^3). \tag{3.3}$$

To see this note that, if j and $N(j)$ are not mutual neighbors, then for the $D(j)$ edges connected to j, say $l_1, ..., l_{D(j)}$ we have $\operatorname{Cov}(X_{l_p i}, X_{l_q i})=\pi_i^3-\pi_i^4$. If j and $N(j)$ are mutual neighbors, then the covariance between X_{ji} and $X_{N(j)i}$ is not $\pi_i^3-\pi_i^4$, but $\pi_i^2-\pi_i^4$ and, therefore, we have to add $2A$ times $\pi_i^2-\pi_i^4-(\pi_i^3-\pi_i^4)=\pi_i^2-\pi_i^3$. Since $\sum_{j=1}^n D(j)=2n$, we can rewrite (3.3) in the form

$$\operatorname{Var}(W_i \mid G)=n(\pi_i^2-\pi_i^3)+\tfrac{1}{2}(\pi_i^3-\pi_i^4)\sum_{j=1}^{n} D^2(j)+2A(\pi_i^2-\pi_i^3). \tag{3.4}$$

Also,

$$\operatorname{Cov}(W_i, W_k \mid G)=-\sum_{j=1}^{n}\binom{D(j)}{2}\pi_i^2\pi_k^2=n\pi_i^2\pi_k^2-\tfrac{1}{2}\pi_i^2\pi_k^2\sum_{j=1}^{n} D^2(j). \tag{3.5}$$

Since $EW_i=E(W_i \mid G)$ for any G, we obtain the variance and covariance by taking expectations of the conditional ones in (3.4) and (3.5). Let $\alpha_n=P(N(N(j))=j)$, that is, the probability that the nearest neighbor of $N(j)$ is j, and set $\beta_n=ED^2(j)$. Both α_n and β_n depend also on F. Note that $2EA=n\alpha_n$. Then

$$\begin{aligned}\operatorname{Var} W_i &= n(\pi_i^2-\pi_i^3)+\tfrac{1}{2}n\beta_n(\pi_i^3-\pi_i^4)+n\alpha_n(\pi_i^2-\pi_i^3),\\ \operatorname{Cov}(W_i, W_k) &= n\pi_i^2\pi_k^2-\tfrac{1}{2}n\beta_n\pi_i^2\pi_k^2, \qquad i\neq k.\end{aligned} \tag{3.6}$$

Let H, J, and K be the diagonal $d \times d$ matrices with ith diagonal entries π_i^3, $\pi_i^2 - \pi_i^4$, and $\pi_i^2 - \pi_i^3$, respectively; let b be a column vector with ith component π_i^2; and let b^{T} denote its transpose. We can write (3.6) as

$$\Sigma_n = \tfrac{1}{2} n(\beta_n - 2)[H - bb^{\mathrm{T}}] + nJ + n\alpha_n K. \tag{3.7}$$

Note that $ED(j) = 2$ and, therefore, $\beta_n \geqslant 4$. It is easy to verify that $[H - bb^{\mathrm{T}}]$ is nonnegative definite, and therefore so is $\Sigma_n - nJ$. Denoting the spectral radius by ρ, we see that

$$\|\Sigma_n^{-1/2}\| \leqslant \rho(\Sigma_n^{-1/2}) \leqslant \rho(n^{-1/2} J^{-1/2}) \leqslant n^{-1/2} M, \tag{3.8}$$

where $M = [\min_{1 \leqslant i \leqslant d} \{\pi_i^2 - \pi_i^4\}]^{-1/2}$. (See, e.g., Horn and Johnson [13] for standard facts on matrices used here and below.) In applying Theorem 2.2 to the vector $\Sigma_n^{-1/2}(\mathbf{W} - \boldsymbol{\lambda})$, this shows that we can take the bound B in Theorem 2.2 to be $B = dn^{-1/2} M$.

Using the fact that the sets S_j in Theorem 2.2 may be random, we choose them depending on the graph G. Specifically, given G, define S_j to consist of j and all vertices which are connected with j by an edge. Note that for any set A, we have $P(\{X_{ji},\ 1 \leqslant i \leqslant d\} \in A \mid G,\ \{X_{li},\ 1 \leqslant i \leqslant d,\ l \notin S_j\} = P(\{X_{ji},\ 1 \leqslant i \leqslant d\} \in A)$. Taking expectations conditioned on $\{X_{li},\ 1 \leqslant i \leqslant d,\ l \notin S_j\}$ we obtain independence of $\{X_{ji},\ 1 \leqslant i \leqslant d\}$ and $\{X_{li},\ 1 \leqslant i \leqslant d,\ l \notin S_j\}$. With similar arguments for $\mathcal{N}_l = \bigcup_{j \in S_l} S_j$, one may conclude that $\chi_1 = \chi_2 = \chi_3 = 0$. It is well known that the degrees in the nearest neighbor graph in R^k are bounded by some constant $K(k)$ which depends on the dimension k, where $K(1) = 2$, $K(2) = 6$, $K(3) = 12$. Estimates are known for all $K(k)$ (see, e.g., [15]). For the sets defined above we have the bounds $D_1 = K(k)$ and $D_2 = K^2(k)$ on the cardinality of S_j ad $\mathcal{N}_l$, respectively.

For any functions $h \in \mathcal{H}$ we obtain, using the closure of $\mathcal{H}$ under affine transformations,

$$\begin{aligned} &\sup\{|Eh(\mathbf{W}) - Eh(\Sigma_n^{1/2}\mathbf{Z} + \boldsymbol{\lambda})|\colon h \in \mathcal{H}\} \\ &\qquad \leqslant caK^3(k)\, M^3(|\log M| + \log n)\, n^{-1/2}, \end{aligned} \tag{3.9}$$

where $\mathbf{Z}$ is a d-dimensional standard normal vector, c is a constant depending on the dimension d, and a is defined in (2.1).

Note that if the color allocation is not done at random and vertices having the same color c_i tend to cluster together, we should expect large values of W_i. This phenomenon was used by Henze [12] (see other references there), who proposed a statistic similar to W_1 for testing equality of two distributions and proved its asymptotic normality.

Our result provides rates and a natural multivariate extension for testing equality of several distributions as we briefly indicate next.

Consider d distributions F_i on R^k, or populations Π_i, $i=1, ..., d$. A sample of size n is obtained by choosing population Π_i with probability π_i, taking a random observation from the chosen population and repeating this procedure independently n times. The nearest neighbor graph whose vertices are the n obtained sample points in R^k is constructed. We color each point by c_i if it is drawn from Π_i. Thus a large W_i indicates that Π_i tends to form clusters, and the vector $\mathbf{W}$ can be used as a test statistic for testing H_0: $F_1 = \cdots = F_d$.

Tests which reduce $\mathbf{W}$ to a univariate statistic and reject H_0, for example, when $\sum_{i=1}^{d} a_i |W_i|^p$ is large, for some p and weights a_i depending on the alternative, or when $\max_{1 \leqslant i \leqslant d} W_i$ is large, may appear natural. More general rejection regions in R^d may arise in connection with specific alternatives. For example, if one suspects a priori that some populations are more likely to differ from the rest than others, then rejection of H_0 if $|W_i| > b_i$ for at least one i, $i=1, ..., d$, for suitable critical values b_i, is a natural choice for a rejection region. The asymptotic distribution of $\mathbf{W}$ under H_0, that is, when $F_1 = \cdots = F_d = F$, is needed to determine critical values for the tests (and, more generally, rejection regions).

Calculations similar to those of Henze [12] show that α_n and β_n of (3.7) converge to finite limits, and therefore there exists a matrix Θ such that $|(1/n)\Sigma_n - \Theta| \to 0$. For a continuous distribution F of the vertex locations, all these limits do not depend on F.

Replacing the covariance matrix Σ_n by $n\Theta$ leads to a nonparametric statistic. More specifically, the statistic $\hat{\mathbf{W}} = n^{-1/2}\Theta^{-1/2}(\mathbf{W} - \boldsymbol{\lambda})$, and its asymptotic distribution do not depend on F. As indicated in (3.11) below, there is a price to pay for the convenience of using the nonparametric statistic $\hat{\mathbf{W}}$: its normal approximation rate seems slower compared to that of $\mathbf{W}$ because of the additional approximation of the covariance matrix.

We apply Theorem 2.2 to the vector $\hat{\mathbf{W}}$. In view of (3.8) we have $\|\Theta^{-1/2}\| \leqslant M$ and therefore we can use the same bound $B = dn^{-1/2}M$ as before.

The covariance matrix of $\hat{\mathbf{W}}$ is not I, and therefore χ_3 does not vanish. Taking into account that instead of $\mathbf{X}_j$ the summands now are $n^{-1/2}\Theta^{-1/2}(\mathbf{X}_j - E\mathbf{X}_j)$, it is easy to verify that $\chi_3 = |I - n^{-1/2}\Theta^{-1/2} \times \Sigma_n n^{-1/2}\Theta^{-1/2}|$. Standard arguments show that

$$\chi_3 = |n^{-1/2}\Theta^{-1/2}(n\Theta - \Sigma_n)\, n^{-1/2}\Theta^{-1/2}| \leqslant |n^{-1/2}\Theta^{-1/2}|^2\, |n\Theta - \Sigma_n|$$

$$\leqslant (d^2 n^{-1/2} M)^2\, |n\Theta - \Sigma_n| = d^4 M^2 \left| \frac{1}{n}\Sigma_n - \Theta \right|. \tag{3.10}$$

For $h \in \mathscr{H}$ we obtain

$$\sup\{|Eh(\mathbf{W}) - Eh(n^{1/2}\Theta^{1/2}\mathbf{Z} + \boldsymbol{\lambda})|: h \in \mathscr{H}\}$$
$$\leqslant caK^3(k)\, M^3(|\log M| + \log n)\left\{n^{-1/2} + \left|\frac{1}{n}\Sigma_n - \Theta\right|\right\}, \qquad (3.11)$$

where $\mathbf{Z}$ is a d-dimensional standard normal vector, c is a constant dependent on the dimension d, and a is defined in (2.1).

The rate in which $|(1/n)\,\Sigma_n - \Theta|$ converges to zero, when the points (vertices) are distributed according to F in R^k, depends on F. A careful examination and adjustments of the calculations in Henze [2] indicate that under reasonable conditions, such as that F is associated with a probability density having a bounded derivative, it is impossible to assert that this rate is of order $n^{-1/2}$. Moreover, the rate becomes slower as the dimension k increases. Thus the rate of normal approximation in (3.11) will generally be determined by the last term.

4. Proof of Theorem 2.1

We denote all constants by c, even when we have in mind different constants in the same equation, as long as they depend only on the dimension d. The method we use is based on the following differential equation due to Barbour [3] and Götze [10],

$$\Delta\Psi(x) - x \cdot \nabla\Psi(x) = h(x) - \Phi h, \qquad x \in R^d, \qquad (4.1)$$

which allows us to evaluate the expectation $Eh(\mathbf{W}) - \Phi h$. In (4.1), Δ is the Laplacian; that is $\Delta\Psi(z) = \operatorname{Tr} \Psi^{(2)}(x)$, where $\Psi^{(2)}(x)$ is the Hessian matrix of second derivatives, and ∇ denotes the gradient.

For $d = 1$, this approach is due to Stein [20, 21]. In this case (4.1) reduces to

$$f'(x) - xf(x) = h(x) - \Phi h, \qquad x \in R, \qquad (4.2)$$

with $f = \Psi'$. The unique bounded solution of (4.2) is given by

$$f(x) = \frac{1}{\phi(x)} \int_{-\infty}^{x} (h(u) - \Phi h)\, \phi(u)\, du. \qquad (4.3)$$

For $d > 1$, a solution for a smoothed version of h is given below, but in both cases we are able to work only with smooth functions. To this end define the following smoothing of h:

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

It is worth noting that $\Phi h_s = \Phi h$ for any s.

A bound on the error which arises from this smoothing is provided by the following version due to Götze [10] of a smoothing lemma of Bhattacharya and Ranga Rao [5].

LEMMA 4.1. *Let Q be a probability measure on R^d. Then there exists a constant $c>0$ which depends only on the dimension d such that for any $0<t<1$,*

$$\sup\left\{\left|\int_{R^d} h d(Q-\Phi)\right| : h\in\mathscr{H}\right\}$$
$$\leqslant c\left[\sup\left\{\left|\int_{R^d}(h-\Phi h)_t\, dQ\right| : h\in\mathscr{H}\right\}+a\sqrt{t}\right],$$

where a is defined in (2.1).

Let us return to (4.1). If h is replaced by h_t, one may verify that Eq. (4.1) has the solution $\Psi_t(x)=-\frac{1}{2}\int_t^1 [h_s(x)-\Phi h](ds/(1-s))$ [10]. Note that $\Psi_t(x)$ depends also on h, but this is suppressed in the national. Let $\Psi_t^{(1)}=\nabla\Psi_t$, and let $\Psi_t^{(2)}(x)$ denote the $d\times d$ Hessian matrix whose pqth entry, denoted by $\Psi^{(2)}_{t(pq)}(x)$, is $\partial^2\Psi_t(x)/(\partial x_p\,\partial x_q)$. It was shown in Götze [10] that for $|h|\leqslant 1$ there exists a universal constant c such that

$$|\Psi_t^{(1)}(\cdot)|<c, \qquad |\Psi_t^{2)}(\cdot)|<c\log(t^{-1}). \tag{4.4}$$

Setting $K_j=\mathbf{X}_j\mathbf{U}_j^{\mathrm{T}}$, a $d\times d$ random matrix, we have by (4.1)

$$Eh_t(\mathbf{W})-\Phi h=E[\Delta\Psi_t(\mathbf{W})-\mathbf{W}\cdot\nabla\Psi_t(\mathbf{W})]=\mathscr{A}-\mathscr{B}-\mathscr{C}+\mathscr{D}, \tag{4.5}$$

where

$$\begin{aligned}
\mathscr{A}&=ETr\left[\Psi_t^{(2)}(\mathbf{W})\left(I-\sum_{j=1}^n K_j\right)\right],\\
\mathscr{B}&=\sum_{j=1}^n E[\mathbf{X}_j\cdot\nabla\Psi_t(\mathbf{V}_j)],\\
\mathscr{C}&=\sum_{j=1}^n E\{\mathbf{X}_j\cdot[\nabla\Psi_t(\mathbf{W})-\nabla\Psi_t(\mathbf{V}_j)-\Psi_t^{(2)}(\mathbf{V}_j)\,\mathbf{U}_j^{\mathrm{T}}]\},\\
\mathscr{D}&=\sum_{j=1}^n ETr\{K_j[\Psi_t^{(2)}(\mathbf{W})-\Psi_t^{(2)}(\mathbf{V}_j)]\}.
\end{aligned} \tag{4.6}$$

The next Lemma is required in order to bound Taylor series remainders arising in the evaluation of the quantities of (4.6).

LEMMA 4.2. *Let* $\mathbf{W}$, $\mathbf{V}$, *and* $\mathbf{U}$ *be any random vectors in* R^d *satisfying* $\mathbf{W}=\mathbf{V}+\mathbf{U}$, *and let* Y *be any random variable. Suppose* $|\mathbf{U}|\leqslant C_1$ *and* $|Y|\leqslant C_2$, *where* C_1 *and* C_2 *are numbers. Set* $\gamma=\sup\{|Eh(\mathbf{W})-\Phi h|\colon h\in\mathscr{H}\}$, *and let* $\Psi^{(3)}_{t(pqr)}$ *denote the third partial derivative of* Ψ_t *with respect to the indicated indices* p, q, *and* r. *Then there exists a constant* c *which depends only on the dimension* d, *such that for any* $0\leqslant\tau\leqslant 1$ *and* $h\in\mathscr{H}$,

$$|EY\Psi^{(3)}_{t(pqr)}(\mathbf{V}+\tau\mathbf{U})|\leqslant cC_2(\gamma/\sqrt{t}+aC_1/\sqrt{t}+a\,|\log t|), \tag{4.7}$$

where a *is defined in* (2.1).

Proof of Lemma 4.2. Substitution and differentiation yield the formula

$$\Psi^{(3)}_{t(pqr)}(x)=c\int_t^1\frac{(1-s)^{1/2}}{s^{3/2}}\,ds\int_{R^d}h(\sqrt{s}\,z+\sqrt{1-s}\,x)\,\phi^{(3)}_{(pqr)}(z)\,dz,$$

where here $c=\frac{1}{2}$, but in the sequel c will stand for any positive constant which may depend on the dimension d. Observe that $\int_{R^d}\phi^{(3)}_{(pqr)}(z)\,dz=0$, since this integral can be written as $(\partial^3/\partial x_p\,\partial x_q\,\partial x_r)\int_{R^d}\phi(z+x)\,dz$ at $x=0$, and the last integral equals the constant 1. This fact will be used in (4.8) below in the third equality, where the added term vanishes. Abbreviating $\phi^{(3)}$ for $\phi^{(3)}_{(pqr)}$, we have

$$\begin{aligned}
&|EY\Psi^{(3)}_{t(pqr)}(\mathbf{V}+\tau\mathbf{U})|\\
&=\left|c\int_t^1\frac{(1-s)^{1/2}}{s^{3/2}}\,ds\int_{R^d}EYh(\sqrt{s}\,z+\sqrt{1-s}\,(\mathbf{V}+\tau\mathbf{U}))\,\phi^{(3)}(z)\,dz\right|\\
&=\left|c\int_t^1\frac{(1-s)^{1/2}}{s^{3/2}}\,ds\int_{R^d}EYh(\sqrt{1-s}\,\mathbf{W}-\sqrt{1-s}(1-\tau)\mathbf{U}+\sqrt{s}\,z)\,\phi^{(3)}(z)\,dz\right|\\
&=\left|c\int_t^1\frac{(1-s)^{1/2}}{s^{3/2}}\,ds\int_{R^d}EY\{h(\sqrt{1-s}\,\mathbf{W}-\sqrt{1-s}\,(1-\tau)\mathbf{U}+\sqrt{s}\,z)\right.\\
&\quad\left.-h(\sqrt{1-s}\,\mathbf{W}-\sqrt{1-s}\,(1-\tau)\mathbf{U})\}\,\phi^{(3)}(z)\,dz\right|\\
&\leqslant c\int_t^1\frac{1}{s^{3/2}}\,ds\,C_2\int_{R^d}E\{\sup_{|u|\leqslant C_1+\sqrt{s}\,|z|}h(\sqrt{1-s}\,\mathbf{W}+u)\\
&\quad-\inf_{|u|\leqslant C_1+\sqrt{s}\,|z|}h(\sqrt{1-s}\,\mathbf{W}+u)\}\,|\phi^{(3)}(z)|\,dz\\
&=cC_2\int_t^1\frac{1}{s^{3/2}}\,ds\int_{R^d}E\tilde{h}(\sqrt{1-s}\,\mathbf{W};\,C_1+\sqrt{s}\,|z|)\,|\phi^{(3)}(z)|\,dz.
\end{aligned}\tag{4.8}$$

Let $\mathbf{Z}$ indicate an independent d-variate standard normal variable. By adding and subtracting the same term, the last quantity of (4.8) equals

$$cC_2 \int_t^1 \frac{1}{s^{3/2}}\, ds \int_{R^d} E\{\tilde{h}(\sqrt{1-s}\,\mathbf{W};\, C_1+\sqrt{s}\,|z|) - \tilde{h}(\sqrt{1-s}\,\mathbf{Z};\, C_1+\sqrt{s}\,|z|) + \tilde{h}(\sqrt{1-s}\,\mathbf{Z};\, C_1+\sqrt{s}\,|z|)\}\ |\phi^{(3)}(z)|\, dz. \tag{4.9}$$

It is easy to see that by the definition of $\tilde{h}$, for any $\varepsilon > 0$,

$$\begin{aligned} &|E\{\tilde{h}(\sqrt{1-s}\,\mathbf{W};\varepsilon) - \tilde{h}(\sqrt{1-s}\,\mathbf{Z};\varepsilon)\}| \\ &\quad \leqslant |E\{h_\varepsilon^+(\sqrt{1-s}\,\mathbf{W}) - E\{h_\varepsilon^+(\sqrt{1-s}\,\mathbf{Z})\}| \\ &\qquad + |E\{h_\varepsilon^-(\sqrt{1-s}\,\mathbf{W}) - E\{h_\varepsilon^-(\sqrt{1-s}\,\mathbf{Z})\}|. \end{aligned} \tag{4.10}$$

By the assumptions on the class $\mathscr{H}$ and the definition of γ we see that for any $\varepsilon > 0$ the expression in (4.10) is bounded by 2γ. As $\int_t^1 (1/s^{3/2})\, ds \leqslant c/\sqrt{t}$, we conclude that for some c,

$$\int_t^1 \frac{1}{s^{3/2}}\, ds \int_{R^d} |E\{\tilde{h}(\sqrt{1-s}\,\mathbf{W};\, C_1+\sqrt{s}\,|z|) - \tilde{h}(\sqrt{1-s}\,\mathbf{Z};\, C_1+\sqrt{s}\,|z|)\}\ \phi^{(3)}(z))\}\ \phi^{(3)}(z)|\, dz \leqslant c\gamma/\sqrt{t}. \tag{4.11}$$

Recording this fact we now study the last term of (4.9). Note that by (2.1)

$$E\tilde{h}(\sqrt{1-s}\,\mathbf{Z};\, C_1+\sqrt{s}\,|z|) \leqslant a(C_1+\sqrt{s}\,|z|).$$

Therefore,

$$\begin{aligned} &\int_t^1 \frac{1}{s^{3/2}}\, ds \int_{R^d} E\tilde{h}(\sqrt{1-s}\,\mathbf{Z};\, C_1+\sqrt{s}\,|z|)\ |\phi^{(3)}(z)|\, dz \\ &\quad \leqslant a \int_t^1 \frac{1}{s^{3/2}}\, ds \int_{R^d} (C_1+\sqrt{s}\,|z|)\ |\phi^{(3)}(z)|\, dz \\ &\quad \leqslant ca(C_1/\sqrt{t} + |\log t|). \end{aligned} \tag{4.12}$$

Lemma 4.2 now follows.

We now return to (4.6) and start with the term $\mathscr{C}$. Let X_{jp} and U_{jp} denote the pth components of $\mathbf{X}_j$ and $\mathbf{U}_j$, respectively. For $j = 1, ..., n$, Taylor expansion of $\nabla\Psi_t(\mathbf{W})$ centered at $\mathbf{V}_j$ shows that $\mathscr{C}$ is equal to

$$\sum_{j=1}^n E \int_0^1 (1-\tau) \sum_{p=1}^d \sum_{q=1}^d \sum_{r=1}^d \Psi^{(3)}_{t(pqr)}(\mathbf{V}_j + \tau\mathbf{U}_j)\, X_{jp}\, U_{jq}\, U_{jr}\, d\tau. \tag{4.13}$$

We apply Lemma 4.2 for each j, with $\mathbf{U}=\mathbf{U}_j$ and $Y=U_{jp}U_{jq}X_{jr}$, and obtain

$$|\mathscr{C}| \leqslant cnA_1^2 B(\gamma/\sqrt{t} + aA_1/\sqrt{t} + a\,|\log t|), \tag{4.14}$$

where again $\gamma = \sup\{|Eh(\mathbf{W}) - \Phi h|\colon h \in \mathscr{H}\}$.

Next consider the term $\mathscr{D}$ in (4.6). With the notation defined above, a first-order Taylor expansion yields

$$\Psi^{(2)}_{t(pq)}(\mathbf{W}) - \Psi^{(2)}_{t(pq)}(\mathbf{V}_j) = \sum_{r=1}^{d} \int_0^1 \Psi^{(3)}_{t(pqr)}(\mathbf{V}_j + \tau\mathbf{U}_j)\, U_{jr}\, d\tau. \tag{4.15}$$

The term $\mathscr{D}$ is obtained from (4.15) by multiplying by the entries of K_j, and it is easy to see from the definition of K_j that this leads to a term which is similar to the term of (4.13) and thus $|\mathscr{D}|$ is bounded by the r.h.s. of (4.14), possibly with a different constant.

Next note that $\mathscr{B}$ of (4.6) satisfies $\mathscr{B} = \sum_{j=1}^{n} E[\nabla\Psi_t(\mathbf{V}_j)\cdot E(\mathbf{X}_j \mid \mathbf{V}_j)]$. By (4.4) the components of $\nabla\Psi_t(\mathbf{V}_j)$ are uniformly bounded, implying that for some $c>0$

$$|\mathscr{B}| \leqslant c \sum_{j=1}^{n} \sum_{p=1}^{d} E\,|E(X_{jp} \mid \mathbf{V}_j)|. \tag{4.16}$$

Finally, we consider the term $\mathscr{A}$ from (4.6). With $\delta_{pq}=1$ if $p=q$ and 0 otherwise, we have

$$\begin{aligned}
&\operatorname{Tr}\left[\Psi^{(2)}_t(\mathbf{W})\left(I - \sum_{j=1}^{n} K_j\right)\right] \\
&\quad= \sum_{p=1}^{d}\sum_{q=1}^{d} \Psi^{(2)}_{t(pq)}(\mathbf{W})\left(\delta_{pq} - \sum_{j=1}^{n} X_{jq}U_{jp}\right) \\
&\quad= \sum_{p=1}^{d}\sum_{q=1}^{d} \Psi^{(2)}_{t(pq)}(\mathbf{W})\left[\delta_{qp} - \sum_{j=1}^{n} E(X_{jq}U_{jp})\right. \\
&\qquad\left. + \sum_{j=1}^{n} E(X_{jq}U_{jp}) - \sum_{j=1}^{n} X_{jq}U_{jp}\right].
\end{aligned} \tag{4.17}$$

By (4.4), $|\Psi^{(2)}_{t(pq)}(\mathbf{W})| < c\log(t^{-1})$ for all $p, q = 1, \ldots, d$, and $0<t<1$. Therefore we obtain for the first two terms on the r.h.s. of (4.17)

$$\begin{aligned}
&E\left|\sum_{p=1}^{d}\sum_{q=1}^{d} \Psi^{(2)}_{t(pq)}(\mathbf{W})\left[\delta_{qp} - \sum_{j=1}^{n} E(X_{jq}U_{jp})\right]\right| \\
&\qquad\leqslant c\,|\log t| \sum_{p=1}^{d}\sum_{q=1}^{d}\left|\delta_{pq} - \sum_{j=1}^{n} E(X_{jp}U_{jp})\right|.
\end{aligned} \tag{4.18}$$

We rewrite the expression involving the last two terms in (4.17) in the form

$$\sum_{j=1}^{n}\sum_{p=1}^{d}\sum_{q=1}^{d}\{\Psi^{(2)}_{t(pq)}(\mathbf{W})-\Psi^{(2)}_{t(pq)}(\mathbf{T}_j)+\Psi^{(2)}_{t(pq)}(\mathbf{T}_j)\}[E(X_{jq}U_{jp})-X_{jq}U_{jp}]. \tag{4.19}$$

Taylor expansion of $\Psi^{(2)}_{t(pq)}(\mathbf{W})-\Psi^{(2)}_{t(pq)}(\mathbf{T}_j)$ and Lemma 4.2 applied for each j with $\mathbf{U}=\mathbf{R}_j$ and $Y=R_{jr}X_{jq}U_{jp}$ imply

$$\sum_{j=1}^{n}\sum_{p=1}^{d}\sum_{q=1}^{d}|E\{[\Psi^{(2)}_{t(pq)}(\mathbf{W})-\Psi^{(2)}_{t(pq)}(\mathbf{T}_j)][E(X_{jq}U_{jp})-X_{jq}U_{jp}]\}| \leqslant cnA_1A_2B(\gamma/\sqrt{t}+aA_2/\sqrt{t}+a|\log t|). \tag{4.20}$$

Returning to (4.19), we use (4.4) to bound the last term as

$$\sum_{j=1}^{n}\sum_{p=1}^{d}\sum_{q=1}^{d}|E\Psi^{(2)}_{t(pq)}(\mathbf{T}_j)[E(X_{jq}U_{jp})-X_{jp}U_{jp}]| \leqslant c|\log t|\sum_{j=1}^{n}\sum_{p=1}^{d}\sum_{q=1}^{d}E|E(X_{jp}U_{jq})-E(X_{jp}U_{jp}\mid\mathbf{T}_j)|. \tag{4.21}$$

Combining Lemma 4.1, and (4.6), (4.14), (4.16), (4.18), (4.20), and (4.21), and noting that since $A_1\leqslant A_2$, the term in the r.h.s. of (4.14) may be ignored, being smaller than that of (4.20), we obtain

$$\begin{aligned}\gamma \leqslant{}& cnA_1A_2B\gamma/\sqrt{t}+cnaA_1A_2B(A_2/\sqrt{t}+|\log t|)\\ &+c\sum_{j=1}^{n}\sum_{p=1}^{d}E|E(X_{jp}\mid\mathbf{V}_j)|\\ &+c|\log t|\left\{\sum_{p=1}^{d}\sum_{q=1}^{d}\left|\delta_{pq}-\sum_{j=1}^{n}E(X_{jp}U_{jq})\right|\right.\\ &\left.+\sum_{j=1}^{n}\sum_{p=1}^{d}\sum_{q=1}^{d}E|E(X_{jp}U_{jq})-E(X_{jp}U_{jq}\mid\mathbf{T}_j)|\right\}+ca\sqrt{t}.\end{aligned} \tag{4.22}$$

The choice $\sqrt{t}=2cnA_1A_2B$, provided it is less than 1, and simple manipulations yield (2.4), after observing that the last term in (4.22) is of lower order than the second term and may be ignored. If for the above choice $t>1$, then the theorem is trivial.

For the case $d=1$, a better bound, that is (2.3), may be obtained with some extra work which we now sketch. The improvement is achieved by eliminating the terms $\log t$ in (4.22) for $d=1$. Indeed, we may use (4.2) with $h=h_t$, and the solution $f=\Psi_t'$ has the explicit representation (4.3). By Lemma 3 of Stein [21, p. 25], f and f' are bounded, and we immediately see that the term $\log t$, arising in (4.18) and (4.21) is avoided in the present case since now $\Psi_t^{(2)}=f'$.

The term $\log t$ arises also in Lemma 4.2 and, consequently, in (4.20). We now show that it can be avoided in the case $d=1$. Using f'' for $\Psi_t^{(3)}$ and the relation $f''(x)=f(x)+xf'(x)+h_t'(x)$, we recalculate the l.h.s. of (4.7) and obtain

$$EY\Psi_t^{(3)}(V+\tau U)=EY\{f(V+\tau U)+(V+\tau U)f'(V+\tau U)+h_t'(V+\tau U)\}. \tag{4.23}$$

Starting with the first and last terms in (4.23), we have

$$\begin{aligned}
&|EY\{h_t'(V+\tau U)+f(V+\tau U)\}| \\
&\quad=\left|EY\left\{\frac{(1-t)^{1/2}}{t^{1/2}}\int_R h(\sqrt{t}\,z+\sqrt{(1-t)}(V+\tau U))\,\phi'(z)\,dz+f(V+\tau U)\right\}\right| \\
&\quad\leqslant E\left|\frac{Y}{\sqrt{t}}\int_R h(\sqrt{t}\,z+\sqrt{(1-t)}(V+\tau U))\,\phi'(z)\,dz\right|+cC_2,
\end{aligned} \tag{4.24}$$

where the last inequality uses the fact that f is bounded. To estimate the integral on the r.h.s. of (4.24) we follow the same logic as in (4.8)–(4.12); the calculations are simpler because we do not integrate with respect to s. We conclude that

$$|EY\{h_t'(V+\tau U)+f(V+\tau U)\}|\leqslant cC_2(\gamma/\sqrt{t}+aC_1/\sqrt{t}+a+1). \tag{4.25}$$

Returning to the middle term in (4.23), we have, using the boundedness of f' and the Cauchy–Schwarz inequality,

$$\begin{aligned}
|EY(V+\tau U)f'(V+\tau U)|&=|EY(W-(1-\tau)U)f'(V+\tau U)| \\
&\leqslant c[\sqrt{EY^2}\sqrt{EW^2}+E|YU|]\leqslant cC_2(\sqrt{EW^2}+C_1).
\end{aligned} \tag{4.26}$$

This leads to the one-dimensional version of Lemma 4.2:

$$|EY\Psi_t^{(3)}(V+\tau U)|\leqslant cC_2(\gamma/\sqrt{t}+aC_1/\sqrt{t}+a+1+\sqrt{EW^2}+C_1). \tag{4.27}$$

Since $a\geqslant 1$ and $t<1$, the r.h.s. of (4.27) may be reduced to $cC_2(\gamma/\sqrt{t}+aC_1/\sqrt{t}+a+\sqrt{EW^2})$. As this version of Lemma 4.2 differs from the original lemma only in the absense of $\log t$ and the addition of $\sqrt{EW^2}$, we obtain (2.3) by repeating the above arguments.

REFERENCES

[1] AVRAM, F., AND BERTSIMAS, D. (1993). On central limit theorems in geometrical probability. *Ann. Appl. Probab.* **3** 1033–1046.

[2] BALDI, P., AND RINOTT, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.* **17** 1646–1650.

[3] BARBOUR, A. D. (1990). Stein's method for diffusion approximations. *Probab. Theory Rel. Fields* **84** 297–322.

[4] BARBOUR, A. D., HOLST, L., AND JANSON, S. (1992). *Poisson Approximation*. Oxford Univ. Press, Oxford.

[5] BHATTACHARYA, R. N., AND RANGA RAO, R. (1986). *Normal Approximation and Asymptotic Expansion*. Krieger, Melbourne, FL.

[6] BOLTHAUSEN, E. (1982). Exact convergence rates in some martingale central limit theorems. *Ann. Probab.* **10** 672–688.

[7] BOLTHAUSEN, E., AND GÖTZE, F. (1993). The rate of convergence of multivariate sampling statistics. *Ann. Statist.* **21** 1692–1710.

[8] DEMBO, A., AND RINOTT, Y. (1994). Some examples of normal approximations by Stein's method. To appear in *Proceedings of IMA workshop on random discrete structures* (D. Aldous and R. Pemantle, Eds.). Springer-Verlag, Berlin/New York.

[9] GOLDSTEIN, L., AND RINOTT, Y. (1994). On multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.*, to appear.

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

[11] HENZE, N. (1987). On the fraction of random points with specified nearest-neighbour interrelations and degree of attraction. *Adv. Appl. Probab.* **19** 873–895.

[12] HENZE, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *Ann. Statist.* **16** 772–783.

[13] HORN, R. A., AND JOHNSON, C. A. (1985). *Matrix Analysis*. Cambridge Univ. Press, Cambridge.

[14] JACOD, J., AND SHIRYAYEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer-Verlag, New York.

[15] LEECH, J., AND SLOANE, N. J. A. (1971). Sphere packings and error-correcting codes. *Canad. J. Math.* **23** 718–745.

[16] LIPTSER, R. S., AND SHIRYAYEV, A. N. (1989). *Theory of Martingales*. Kluwer Academic, Boston.

[17] RINOTT, Y. (1994). On normal approximation rates for certain sums of dependent random variables. *J. Comput. Appl. Math.* **55** 135–143.

[18] ROTAR, V. I. (1995). *Probability Theory*. World Scientific, New York.

[19] SAZONOV, V. V. (1968). On the multi-dimensional central limit theorem. *Sankhyā Ser. A* **30** 181–204.

[20] STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. Sixth Berkeley Symp. Math. Statist. Probab.*, Vol. 2, pp. 583–602. Univ. California Press, Berkeley.

[21] STEIN, C. (1986). *Approximate Computation of Expectations*. IMS, Hayward, CA.