Bernoulli **10**(2), 2004, 221–250

Berry–Esseen and central limit theorems for serial rank statistics via graphs

TATYANA S. TUROVA

Centre for Mathematical Sciences, Lund University, Box 118, Lund S-221 00, Sweden.
E-mail: tatyana@maths.lth.se

We use Stein's method in combination with notions from graph theory to obtain Berry–Esseen and central limit theorems for generalized serial rank statistics.

Keywords: graphs; rank statistics; serial ranks

1. Introduction and statements of main results

Consider the nonparametric test of

$$H_0 : X_1, \dots, X_n \text{ are independent}$$

against

$$H_1 : X_1, \dots, X_n \text{ are serially dependent,}$$

where $X_1, \dots, X_n$ are observations of a time series at times $i = 1, \dots, n$, and under H_0 the random variables $X_1, \dots, X_n$ are also assumed to have a common continuous distribution function. Many of the standard nonparametric test statistics for H_0 against H_1 are based upon the vector of ranks $(R(1), \dots, R(n))$, where $R(i)$ denotes the rank of X_i. Such statistics are typically of one of two basic forms, which we introduce below. First, let $\mathcal{N} := \{1, \dots, n\}$, and for any $0 < k < n$ set

$$\mathcal{N}_k := \{(i_0, \dots, i_k) \in \mathcal{N}^{k+1} : i_j \neq i_l \text{ for any } j \neq l\}. \tag{1.1}$$

Further, for $0 < r < n$, let

$$A = \{a(I) : I \in \mathcal{N}_r\} \tag{1.2}$$

be an array of real constants. Define the *generalized serial rank statistic*

$$W_A := \sum_{i=1}^{n} a(R(i), R((i-1)_{\mathrm{mod}\ n}), \dots, R((i-r)_{\mathrm{mod}\ n})), \tag{1.3}$$

where, for any $-n+1 \leqslant M \leqslant n-1$,

$$M_{\mathrm{mod}\ n} = \begin{cases} M + n, & \text{if } M \leqslant 0, \\ M, & \text{if } M > 0. \end{cases}$$

1350–7265

Many classical serial rank statistics can be put into one of these two general forms. For instance, the Wald–Wolfowitz statistic (Wald and Wolfowitz 1943)

$$W_{A_n} := \sum_{i=1}^{n} a_n(R(i))a_n(R((i-1)_{\mathrm{mod}\ n}))$$

is of the form W_A. Serial rank statistics of a *non-circular* form

$$T_A := \sum_{i=r+1}^{n} a(R(i), R(i-1), \ldots, R(i-r))$$

have been systematically studied by Hallin *et al*. (1985) and Hallin and Puri (1988; 1994). Among other results, they have shown how to construct statistics of this form that are asymptotically locally most powerful against certain ARMA alternatives.

The aim of this paper is to prove a Berry–Esseen theorem for serial rank statistics by exploiting their underlying graph structure, in combination with the method of Stein (1972). The graph structure of such statistics was first disclosed in Haeusler *et al*. (2000). The graph representation is particularly helpful when computing the moments of serial rank statistics. For future reference, we record here two results about the mean and the variance of such statistics. In order to state these results we must introduce some notation. First, note that by definition (1.1), for any $0 \leqslant k < n$,

$$|\mathcal{N}_k| = n(n-1)\ldots(n-k), \tag{1.4}$$

where $|C|$ denotes the cardinality of a set C. Now set

$$\mu_A := \frac{n}{|\mathcal{N}_r|} \sum_{I \in \mathcal{N}_r} a(I), \tag{1.5}$$

and, for all $n > 2r+1$,

$$\sigma_A^2 := \left(\frac{n}{|\mathcal{N}_r|} - \frac{n(n-(2r+1))}{|\mathcal{N}_{2r+1}|}\right) \sum_{I \in \mathcal{N}_r} a^2(I) \tag{1.6}$$

$$+\ 2\sum_{k=1}^{r}\left(\frac{n}{|\mathcal{N}_{r+k}|} - \frac{n(n-(2r+1))}{|\mathcal{N}_{2r+1}|}\right) \sum_{(i_0,\ldots,i_{r+k}) \in \mathcal{N}_{r+k}} a(i_0, \ldots, i_r)a(i_k, \ldots, i_{r+k}).$$

Proposition 1.1. *Under the hypothesis* H_0 *one has*

$$\mathrm{E}W_A = \mu_A, \tag{1.7}$$

and under the additional assumption that

$$\mu_A = 0, \tag{1.8}$$

one also has, for all $n > 2r+1$,

$$\mathrm{E}W_A^2 = \mathrm{var}(W_A) = \sigma_A^2. \tag{1.9}$$

Formula (1.7) is of course trivial, and the proof of (1.9) becomes an easy exercise after we have discovered the underlying graph structure (see formula (2.3) below).

To state our Berry–Esseen theorem we require the following assumption on the array A: for some $0 < K < \infty$ and all $0 < r < n$,

$$K_A := \frac{n}{|\mathcal{N}_r|} \sum_{I \in \mathcal{N}_r} a^2(I) \leqslant K. \tag{1.10}$$

Further, we introduce the notation

$$\beta_A = \sum_{I \in \mathcal{N}_r} |a^3(I)|$$

and

$$n(s) = (n-1) \ldots (n-s).$$

From now on Z will denote a standard normal random variable.

Theorem 1.1. *Under assumptions* (1.8), (1.10) *and*

$$\sigma_A^2 = 1, \tag{1.11}$$

we have, for any fixed $r > 0$ *and* $n > r+1$,

$$\sup_{-\infty<x<\infty} |P\{W_A \leqslant x\} - P\{Z \leqslant x\}| \leqslant c(r, K) \frac{\beta_A}{n(r)}, \tag{1.12}$$

where the constant $c(r, K)$ *does not depend on* n.

A similar theorem holds for the statistics T_A as well, but we omit the details here for the sake of brevity.

The first Berry–Esseen theorem for serial rank statistics was proven by Hallin and Rifi (1997) for the subclass of T_A of the form

$$\overline{T}_{A_n} := (n-r)^{-1/2} \sum_{i=r+1}^{n} a_n^{(1)}(R(i)) a_n^{(2)}(R(i-r)).$$

They showed under a number of regularity and smoothness conditions that

$$\sup_{-\infty<x<\infty} |P\{\overline{T}_{A_n} \leqslant x\} - P\{Z \leqslant x\}| = O(n^{-1/2}).$$

The proof of Hallin and Rifi (1997) was based upon the characteristic function methods of van Zwet (1982) and Does (1982).

In our proof we shall follow closely in the steps of Bolthausen (1984), who developed Stein's method to establish a Berry–Esseen theorem for linear rank statistics. Essential for Bolthausen's proof is a randomization step using a combinatorial argument. His approach has since been applied to other rank-type statistics by Bolthausen and Götze (1993), Loh (1996), and Zhao *et al*. (1997). Bolthausen's (1984) construction also inspired Mason and Turova (2000) to come up with a randomization for the serial rank statistics when $r = 1$.

This is where the statistics' underlying graph structure comes vitally into play. As in the paper by Bolthausen (1984), it relies on the specific features of the $r = 1$ case, which roughly speaking means a 'one-node boundary' in terms of graphs, and allows one to use relatively easy combinatorics in the proof. Therefore, one of our major tasks here is to construct an appropriate coupling for the case $r \geqslant 1$.

Next we state our central limit theorem.

Theorem 1.2. *For any fixed* $r > 0$*, let* $A_n = \{a_n(I) : I \in \mathcal{N}_r\}$*,* $n > r+1$*, be a sequence of arrays of numbers satisfying* (1.8), (1.10) *and* (1.11) *uniformly in* $n > 2r+1$*. Also assume the Lindeberg condition that, for all* $\varepsilon > 0$*,*

$$\lim_{n\to\infty} \frac{n}{|\mathcal{N}_r|} \sum_{|a_n(I)|>\varepsilon} a_n^2(I) = 0.$$

Then $W_{A_n} \xrightarrow{d} Z$ *as* $n \to \infty$*.*

Theorem 1.2 is proved using Stein's method along lines similar to those followed by Schneller (1998) to obtain a central limit theorem for linear rank statistics. It is an extension of the central limit theorem for W_{A_n} obtained by Mason and Turova (2000) for the special case $r = 1$. After the necessary coupling is constructed, as detailed in Section 4, the proof remains much the same as in the case $r = 1$.

Here is a useful sufficient condition for both the Lindeberg condition and (1.10) to hold. Assume that, for all $I \in \mathcal{N}_r$ and $n > r+1$,

$$|a(I)| \leqslant n^{-1/2} g((n+1)^{-1} I),$$

where g is a non-negative measurable function defined on $(0, 1)^{r+1}$, such that, as $n \to \infty$,

$$n^{-r-1} \sum_{I\in\mathcal{N}_r} g^2((n+1)^{-1} I) \to \int_{(0,1)^{r+1}} g^2(u)\mathrm{d}u < \infty.$$

Then it is easy to verify that the Lindeberg condition is satisfied, as well as (1.10) (uniformly in $n > r+1$) for some $K > 0$. If, in addition, we assume that, as $n \to \infty$,

$$n^{-r-1} \sum_{I\in\mathcal{N}_r} g^3((n+1)^{-1} I) \to \int_{(0,1)^{r+1}} g^3(u)\mathrm{d}u := C < \infty,$$

then we obtain that $\beta_A / n(r) \leqslant C/\sqrt{n}$.

It will be shown in the course of the proof of Theorem 1.1 that under its assumptions one always has, for all $n > 2r+1$,

$$\frac{\beta_A}{n(r)} \geqslant \frac{1}{(1+2r)^{3/2}\sqrt{n}}.$$

This says that $O(n^{-1/2})$ is the best rate achievable by Theorem 1.1.

2. The graph structure

2.1. The statistics W_A via graphs

The statistics we introduce above can be written in an alternative way using a graph representation. Let us now introduce some notions from graph theory that we need for our approach.

A *directed graph* H is defined by a set of vertices $V \subseteq \mathcal{N} := \{1, \ldots, n\}$ and by a set of ordered pairs $\{(i_k, j_k) : k = 1, \ldots, N\} \subseteq \mathcal{N}_1$, $N \geqslant 1$, where $i_k, j_k \in V$ and $i_k \neq j_k$ for each $k = 1, \ldots, N$. The pair (i_k, j_k) represents the arc from the vertex i_k to the vertex j_k. Thus the positive integer N denotes the number of arcs in the directed graph. Any set of $k \geqslant 1$ connected arcs in the form $\{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{k-1}, i_k)\}$, with $i_m \neq i_j$ for any $m \neq j$, we call a k-path, or a path, from i_0 to i_k. We say that two vertices are connected in a graph if and only if there is a path in this graph between them. Any set of $k \geqslant 3$ connected arcs in the form $\{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{k-2}, i_{k-1}), (i_{k-1}, i_0)\}$, with $i_m \neq i_j$ for any $m \neq j$, we call a directed cycle on k vertices.

A directed graph whose edges form one directed cycle which passes through every vertex of V is called a (directed) Hamiltonian cycle. For any $n \geqslant 3$ and $V \subseteq \mathcal{N}$, let $\mathcal{H}(V)$ be the set of all Hamiltonian cycles on V. Thus any graph H in $\mathcal{H}(V)$ is defined simply by the set of its arcs, and we shall write

$$H = \{(i_1, i_2), (i_2, i_3), \ldots, (i_{|V|}, i_1)\}, \tag{2.1}$$

where $\{i_1, i_2, \ldots, i_{|V|}\} = V$. In particular, when $V = \mathcal{N}$ we shall use the notation $\mathcal{H}(\mathcal{N}) = \mathcal{H}_n$.

For any $r \geqslant 1$ and any $\Gamma \subseteq \mathcal{N}_1$, let $L^r(\Gamma) \subseteq \mathcal{N}_r$ denote the set of all vectors associated with the connected r-paths in Γ, that is,

$$L^r(\Gamma) := \{(j_0, \ldots, j_r) : \{(j_0, j_1), (j_1, j_2), \ldots, (j_{r-1}, j_r)\} \subseteq \Gamma\}. \tag{2.2}$$

In particular, for $r = 1$, we have $L^1(\Gamma) = \Gamma$ for any $\Gamma \subseteq \mathcal{N}_1$. Clearly, for any $H \in \mathcal{H}_n$ there are exactly n different r-paths in H for any $1 \leqslant r < n$, and each path is defined uniquely by the first vertex; for example, if $H = \{(i_1, i_2), (i_2, i_3), \ldots, (i_n, i_1)\}$ then, for any $1 < r < n$, $L^r(H) = \{(i_1, \ldots, i_{r+1}), \ldots, (i_n, i_1, \ldots, i_r)\}$.

Now it is easy to show, using the same idea as in the proof of Mason and Turova (2000), that also in the general case $r > 1$ the following representation also occurs:

$$W_A \stackrel{d}{=} W_A(H) := \sum_{I \in L^r(H)} a(I), \tag{2.3}$$

where H is uniformly distributed on $\mathcal{H}_n$, and the sum runs over all r-paths in H. Formula (2.3) reduces the proof of Proposition 1.1 to straightforward computations.

2.2. Coupling

The key ingredient in the proof of a combinatorial central limit theorem or Berry–Esseen theorem is the construction of the following coupling.

Lemma 2.1. *Let* $n > 2r(r+2) > 0$ *be fixed arbitrarily, and let a random cycle* H *be uniformly distributed on* $\mathcal{H}_n$. *Assume that* $(I_0, I_1, \ldots, I_r)$ *and* $(I_0, J_1, \ldots, J_r)$ *are random vectors distributed independent of* H *and uniformly on* $\mathcal{N}_r$, *so that, conditionally on* $I_0 = i_0 \in \mathcal{N}$, *the vectors* $(I_1, \ldots, I_r)$ *and* $(J_1, \ldots, J_r)$ *are also independent. Then there exist random cycles* H_1 *and* H_2 *such that*

(i) $H \overset{d}{=} H_1 \overset{d}{=} H_2$;
(ii) $(I_0, I_1, \ldots, I_r) \in L^r(H_1)$, $(I_0, J_1, \ldots, J_r) \in L^r(H_2)$;
(iii) $L^r(H_1) \triangle L^r(H_2)$ *and* H *are independent;*
(iv) $(I_0, J_1, \ldots, J_r)$ *and* H_1 *are independent;*
(v) $|H_1 \triangle H_2| < C$ *and* $|H \triangle H_1| < C$ *for some constant* $C = C(r)$ *independent of* n.

Here $\triangle$ *denotes a symmetric difference of two sets.*

We postpone the proof of this lemma to Section 4. As an immediate corollary of formula (2.3) and properties (i)–(iii) from Lemma 2.1 we obtain the following useful result (for the proof of a similar result, see Mason and Turova 2000).

Corollary 2.1. *Let* $n > 2r(r+2) > 0$ *be fixed arbitrarily, and let array* A *satisfy condition* (1.8). *Further, set*

$$W = W_A(H), \qquad W_1 = W_A(H_1), \qquad W_2 = W_A(H_2). \tag{2.4}$$

Then

$$W \overset{d}{=} W_1 \overset{d}{=} W_2 \tag{2.5}$$

and, for any measurable function g,

$$\mathrm{E}Wg(W) = n\mathrm{E}a(I_0, I_1, \ldots, I_r)g(W_1) = n\mathrm{E}a(I_0, J_1, \ldots, J_r)g(W_2). \tag{2.6}$$

2.3. Conditional expectation

Here we shall derive the results necessary for the induction argument in the proof of our Berry–Esseen theorem. Throughout this section we assume $n > (2r+2)C =: \kappa$, with the constant C as in Lemma 2.1.

In what follows we shall denote vectors by letters without indices (e.g. u), or with superscript indices (e.g. u^k). Letters with subscript indices (e.g. u_q or u_q^k), will be reserved for the one-dimensional values only. If $u = (u_1, \ldots, u_i, u_{i+1}, \ldots, u_k)$ and $v = (v_1, \ldots, v_l)$ are two vectors, then we shall use a shorthand notation $(u_1, \ldots, u_i, v, u_{i+1}, \ldots, u_k)$ for the vector $(u_1, \ldots, u_i, v_1, \ldots, v_l, u_{i+1}, \ldots, u_k)$. We shall use bold face to denote the ordered sets of vectors (e.g. $\mathbf{u} = (u^1, \ldots, u^k)$). Let $\{\mathbf{u}\}$ denote, for any set of vectors $\mathbf{u}$, the set of all the components of the vectors of this set.

Now let $\mathbf{v} = (v^0, \ldots, v^t)$, $t \geqslant 0$, be an ordered set of $t+1$ vectors $v^q = (v_0^q, \ldots, v_{p_q}^q)$, $p_q \geqslant 1$, $0 \leqslant q \leqslant t$, fixed arbitrarily but in such a way that $v_l^q \neq v_{l'}^{q'}$ unless $q = q'$ and $l = l'$, $\{\mathbf{v}\} \subset \mathcal{N}$, and

$$|\{\mathbf{v}\}| < \kappa. \tag{2.7}$$

With any vector $i = (i_0, \ldots, i_k) \in \mathcal{N}_k$ where $1 \leqslant k \leqslant n-1$, we shall associate a k-path

$$\Gamma(i) := \{(i_0, i_1), (i_1, i_2), \ldots, (i_{k-1}, i_k)\} \subset \mathcal{N}_1. \tag{2.8}$$

Define also

$$\Gamma(\mathbf{v}) = \bigcup_{i=0}^{t} \Gamma(v^i) \tag{2.9}$$

to be a (non-ordered) set of the paths associated with the vectors of the set $\mathbf{v}$.

Let H be a random cycle uniformly distributed on $\mathcal{H}_n$. Consider the distribution of H conditionally on the event

$$\{\Gamma(\mathbf{v}) \subset H, (v_{p_i}^i, v_0^j) \notin H, 0 \leqslant i, j \leqslant t\}, \tag{2.10}$$

which says that the cycle H passes through every path of the set $\Gamma(\mathbf{v})$, and moreover, the paths of $\Gamma(\mathbf{v})$ are not connected in the cycle H. Let us denote

$$\tilde{\mathcal{N}} = \mathcal{N} \backslash \{\mathbf{v}\} = \mathcal{N} \backslash \{v_0^q, \ldots, v_{p_q}^q, 0 \leqslant q \leqslant t\},$$

and

$$\tilde{\mathcal{N}}_t := \{(j_0, \ldots, j_t) \in \tilde{\mathcal{N}}^{t+1} : j_k \neq j_m \text{ for any } k \neq m\}.$$

For any given cycle $\tilde{H} \in \mathcal{H}(\tilde{\mathcal{N}})$ and vector $x = (x_0, \ldots, x_t) \in \tilde{\mathcal{N}}_t$, construct a new cycle $\tilde{H}_{x,\mathbf{v}} \in \mathcal{H}(\mathcal{N})$ by inserting the path $\Gamma(v^q)$ into the cycle $\tilde{H}$ after the node x_q for each $q = 0, \ldots, t$. To define $\tilde{H}_{x,\mathbf{v}}$ in formal terms, let us introduce for any cycle H the following function determined by the edges of H:

$$E_H(i) = j \text{ if and only if } (i, j) \in H, \qquad \text{for all } i \in \mathcal{N}. \tag{2.11}$$

In this notation, we set

$$\tilde{H}_{x,\mathbf{v}} = \tilde{H} \backslash \{(x_q, E_{\tilde{H}}(x_q)), q = 0, \ldots, t\} \cup \Gamma(\mathbf{v}) \cup \{(x_q, v_0^q), (v_{p_q}^q, E_{\tilde{H}}(x_q)), q = 0, \ldots, t\}. \tag{2.12}$$

Now let $X = (X_0, \ldots, X_t)$ be a random vector and $\tilde{H}$ be a random cycle distributed independently and uniformly on $\tilde{\mathcal{N}}_t$ and $\mathcal{H}(\tilde{\mathcal{N}})$, respectively. Then it is easy to see, in the notation of (2.12), that

$$H|_{\{\Gamma(\mathbf{v}) \subset H, (v_{p_i}^i, v_0^j) \notin H, 0 \leqslant i, j \leqslant t\}} \stackrel{d}{=} \tilde{H}_{X,\mathbf{v}}, \tag{2.13}$$

where $\tilde{H}_{X,\mathbf{v}}|_{X=x} \stackrel{d}{=} \tilde{H}_{x,\mathbf{v}}$. Observe that if $\mathbf{v}' = (v^{\pi(0)}, \ldots, v^{\pi(t)})$, where π is a permutation of the set $\{0, \ldots, t\}$, then, according to our construction,

$$\Gamma(\mathbf{v}') = \Gamma(\mathbf{v}) \tag{2.14}$$

and

$$\tilde{H}_{X,\mathbf{v}'} \stackrel{d}{=} \tilde{H}_{X,\mathbf{v}}. \tag{2.15}$$

We shall now obtain a useful representation for $W_A(H)$ (see (2.3)) conditionally on the event (2.10). First we define, for any vectors $x \in \tilde{\mathcal{N}}_t$ and $j = (j_0, \ldots, j_r) \in \tilde{\mathcal{N}}_r$, a path $\Gamma_{j,x,\mathbf{v}}$ as follows:

(i) if $\{j_0, \ldots, j_{r-1}\} \cap \{x\} = \varnothing$ set, using definition (2.8),

$$\Gamma_{j,x,\mathbf{v}} = \Gamma(j);$$

(ii) otherwise, if $\{j_0, \ldots, j_{r-1}\} \cap \{x\} = \{j_{k_0} = x_{k_0'}, \ldots, j_{k_s} = x_{k_s'}\}$ for some $0 \leqslant k_0 < \ldots < k_s < r$ with $0 \leqslant s < r$, insert the path $\Gamma(v^{k_q'})$ into $\Gamma(j)$ after the node $x_{k_q'}$, for every $q = 0, \ldots, s$, that is, set

$$\Gamma_{j,x,\mathbf{v}} = \Gamma(j_0, \ldots, j_{k_0}, v^{k_0'}, j_{k_0+1}, \ldots, j_{k_1}, v^{k_1'}, \ldots, j_{k_s}, v^{k_s'}, j_{k_s+1}, \ldots, j_s). \tag{2.16}$$

Next, for any array A as in (1.2), we introduce a random array of constants

$$\tilde{A}_{X,\mathbf{v}} := \{\tilde{a}_{X,\mathbf{v}}(j) : j \in \tilde{\mathcal{N}}_r\}$$

such that, for any $x \in \tilde{\mathcal{N}}_t$,

$$\tilde{A}_{X,\mathbf{v}}|_{X=x} = \tilde{A}_{x,\mathbf{v}} := \{\tilde{a}_{x,\mathbf{v}}(j) : j \in \tilde{\mathcal{N}}_r\} \tag{2.17}$$

is a non-random array of constants

$$\tilde{a}_{x,\mathbf{v}}(j) := \begin{cases} \displaystyle\sum_{i=(i_0,\ldots,i_r)\in L^r(\Gamma_{j,x,\mathbf{v}}):i_0\in\{j_0,v_0^q,\ldots,v_{p_q}^q\}} a(i), & \text{if } \{j_0\} \cap \{x\} = x_q, \\ \displaystyle\sum_{i=(i_0,\ldots,i_r)\in L^r(\Gamma_{j,x,\mathbf{v}}):i_0=j_0} a(i), & \text{if } \{j_0\} \cap \{x\} = \varnothing. \end{cases} \tag{2.18}$$

(The last sum contains just one term.) Notice that $\tilde{a}_{x,\mathbf{v}}(j) = a(j)$ unless $\{j_0, \ldots, j_{r-1}\} \cap \{x\} \neq \varnothing$, and in either case the number of terms in the sums in (2.18) is, according to assumption (2.7), at most finite and independent of n, which implies in particular the following uniform bound

$$|\tilde{a}_{x,\mathbf{v}}(i)| \leqslant \kappa_1 \max_{j\in\mathcal{N}_r} |a(j)| \tag{2.19}$$

for some constant κ_1.

Clearly, according to (2.13) we have

$$W_A(H)|_{\{\Gamma(\mathbf{v})\subset H,(v_{p_i}^i,v_0^j)\notin H,0\leqslant i,j\leqslant t\}} \stackrel{d}{=} W_A(\tilde{H}_{X,\mathbf{v}}). \tag{2.20}$$

Hence taking into account (2.17)–(2.18) we obtain the following representation which will play a key role in the induction argument in the proof of Theorem 1.1:

$$W(H)|_{\{\Gamma(\mathbf{v})\subset H,(v^i_{p_i},v^j_0)\notin H,0\leqslant i,j\leqslant t\}} \stackrel{d}{=} \sum_{J\in L^r(\tilde{H})} \tilde{a}_{X,\mathbf{v}}(J) = W_{\tilde{A}_{X,\mathbf{v}}}(\tilde{H}). \tag{2.21}$$

Lemma 2.2. *Let* $\mathbf{v}$ *be fixed arbitrarily to satisfy* (2.7), *and also let* $x \in \tilde{\mathcal{N}}_t$ *be fixed arbitrarily. Set* $\tilde{A} = \tilde{A}_{x,\mathbf{v}}$, $W = W_A(H)$ *and* $W_{\tilde{A}} = W_{\tilde{A}}(\tilde{H})$. *Under assumptions* (1.8) *and* (1.10), *there exists a constant* $M_r > 0$ *such that, uniformly in* $\mathbf{v}$, x *and* n,

$$|\mathrm{E}W_{\tilde{A}} - \mathrm{E}W| \leqslant M_r \max_{J\in\mathcal{N}_r} |a(J)|, \tag{2.22}$$

and

$$|\mathrm{E}W^2_{\tilde{A}} - \mathrm{E}W^2| \leqslant M_r\left(\max_{J\in\mathcal{N}_r} |a^2(J)| + \max_{J\in\mathcal{N}_r} |a(J)|\right). \tag{2.23}$$

Proof. Notice that $|\tilde{\mathcal{N}}_r| = |\tilde{\mathcal{N}}| \ldots (|\tilde{\mathcal{N}}| - r)$, where

$$n - \kappa < |\tilde{\mathcal{N}}| < n \tag{2.24}$$

by (2.7). We readily see then that, for all $r \leqslant m \leqslant 2r$,

$$\frac{n|\mathcal{N}_m \backslash \tilde{\mathcal{N}}_m|}{|\tilde{\mathcal{N}}_m|} \leqslant D, \tag{2.25}$$

where D is some positive constant depending on r only.

Now consider $W_{\tilde{A}}$. According to (2.18), we have

$$W_{\tilde{A}} = \sum_{J\in L^r(\tilde{H}):\{J_0,\ldots,J_{r-1}\}\cap\{x\}=\varnothing} a(J) + \sum_{J\in L^r(\tilde{H}):\{J_0,\ldots,J_{r-1}\}\cap\{x\}\neq\varnothing} \tilde{a}_{x,\mathbf{v}}(J). \tag{2.26}$$

Obviously,

$$\#\{J \in L^r(\tilde{H}) : \{J_0, \ldots, J_{r-1}\} \cap \{x\} \neq \varnothing\} \leqslant \kappa_2 \tag{2.27}$$

for some $\kappa_2 = \kappa_2(r)$ independent of n, $\mathbf{v}$ and x. Hence, from here, (2.26) and (2.19) we obtain:

$$\left| W_{\tilde{A}} - \sum_{J\in L^r(\tilde{H})} a(J) \right| \leqslant C_1 \max_{J\in\mathcal{N}_r} |a(J)| \tag{2.28}$$

for some positive constant $C_1 = C_1(r)$, which implies that

$$|\mathrm{E}W_{\tilde{A}} - \mathrm{E}W| \leqslant \left| \mathrm{E}\left(\sum_{J\in L^r(\tilde{H})} a(J)\right) - \mathrm{E}W \right| + C_1 \max_{J\in\mathcal{N}_r} |a(J)|. \tag{2.29}$$

Using formula (1.5) together with condition (1.8), we obtain

$$\left| \mathrm{E}\left(\sum_{J \in L^r(\tilde{H})} a(J) \right) - \mathrm{E}W \right| = \left| \frac{|\tilde{\mathcal{N}}|}{|\tilde{\mathcal{N}}_r|} \sum_{J \in \tilde{\mathcal{N}}_r} a(J) \right| \leqslant \frac{n}{|\tilde{\mathcal{N}}_r|} \left| \sum_{J \in \mathcal{N}_r \setminus \tilde{\mathcal{N}}_r} a(J) \right|. \tag{2.30}$$

Employing (2.25), we now readily derive from (2.30) that

$$\left| \mathrm{E} \sum_{J \in L^r(\tilde{H})} a(J) \right| = \left| \mathrm{E} \sum_{J \in L^r(\tilde{H})} a(J) - \mathrm{E}W \right| \leqslant C_2 \max_{J \in \mathcal{N}_r} |a(J)| \tag{2.31}$$

for some positive constant $C_2 = C_2(r)$ uniformly in n. Substituting the last bound into (2.29), we obtain (2.22).

To prove (2.23), consider

$$|\mathrm{E}(W_{\tilde{A}})^2 - \mathrm{E}W^2| \leqslant \left| \mathrm{E}(W_{\tilde{A}})^2 - \mathrm{E}\left(\sum_{J \in L^r(\tilde{H})} a(J) \right)^2 \right| + \left| \mathrm{E}\left(\sum_{J \in L^r(\tilde{H})} a(J) \right)^2 - \mathrm{E}W^2 \right|. \tag{2.32}$$

Set $Y = \sum_{J \in L^r(\tilde{H})} a(J)$, where $|L^r(\tilde{H})| = |\widetilde{N}|$. We see that

$$\left| \mathrm{E}(W_{\tilde{A}})^2 - \mathrm{E}\left(\sum_{J \in L^r(\tilde{H})} a(J) \right)^2 \right| \leqslant \mathrm{E}(W_{\tilde{A}} - Y)^2 + 2\mathrm{E}(|W_{\tilde{A}} - Y|\,|Y|). \tag{2.33}$$

Now by (2.28),

$$|W_{\tilde{A}} - Y| \leqslant C_1 \max_{J \in \mathcal{N}_r} |a(J)|, \tag{2.34}$$

and by the Cauchy–Schwarz inequality,

$$\mathrm{E}|Y| \leqslant \sqrt{\mathrm{var}(Y) + (\mathrm{E}Y)^2} \leqslant \sqrt{\mathrm{var}(Y)} + |\mathrm{E}Y|.$$

Notice that (1.6), in combination with the inequality $|ab| \leqslant (a^2 + b^2)/2$, implies, for all $n > 2r + 1$,

$$\mathrm{E}W_A^2 \leqslant (2r+1) \frac{n}{|\mathcal{N}_r|} \sum_{I \in \mathcal{N}_r} a^2(I). \tag{2.35}$$

Using inequality (2.35), we now obtain

$$\sqrt{\mathrm{var}(Y)} \leqslant \left((2r+1) \frac{n}{|\tilde{\mathcal{N}}_r|} \sum_{I \in \tilde{\mathcal{N}}_r} \left(a(I) - \frac{\mathrm{E}Y}{|\widetilde{\mathcal{N}}|} \right)^2 \right)^{1/2},$$

which by assumption (1.10) and (2.31) is clearly bounded from above by some constant $C_3 = C_3(r)$ uniformly in n.

Gathering together our bounds and recalling (2.31), we now obtain from (2.33) that

$$\left|\mathrm{E}\big(W_{\tilde{A}}\big)^2 - \mathrm{E}Y^2\right| \leqslant C_1 \max_{J\in\mathcal{N}_r} |a(J)| \left[\{C_1 + 2C_2\} \max_{J\in\mathcal{N}_r} |a(J)| + 2C_3\right]. \tag{2.36}$$

Consider now the second summand in (2.32). First, we observe that for any array (1.2) (even without condition (1.8)) and all $n > 2r+1$,

$$\begin{aligned} \mathrm{E}W_A^2 &= \left(\frac{n}{|\mathcal{N}_r|} - \frac{n(n-(2r+1))}{|\mathcal{N}_{2r+1}|}\right) \sum_{I\in\mathcal{N}_r} a^2(I) \\ &\quad + 2\sum_{k=1}^{r} \left(\frac{n}{|\mathcal{N}_{r+k}|} - \frac{n(n-(2r+1))}{|\mathcal{N}_{2r+1}|}\right) \sum_{I=(i_0,\ldots,i_{r+k})\in\mathcal{N}_{r+k}} a(i_0,\ldots,i_r)a(i_k,\ldots,i_{r+k}) \\ &\quad + \frac{n(n-(2r+1))}{|\mathcal{N}_{2r+1}|}\left(\sum_{I\in\mathcal{N}_r} a(I)\right)^2. \end{aligned} \tag{2.37}$$

For further reference we derive from (2.37) a useful representation:

$$\mathrm{E}W_A^2 = \sigma_A^2 + (\mathrm{E}W_A)^2 \frac{n(n-(2r+1))}{|\mathcal{N}_{2r+1}|}\left(\frac{|\mathcal{N}_r|}{n}\right)^2. \tag{2.38}$$

Using formula (2.37) for both of the second moments in the last term of (2.32), we obtain

$$\begin{aligned} \left|\mathrm{E}\left(\sum_{J\in L^r(\tilde{H})} a(J)\right)^2 - \mathrm{E}W^2\right| &\leqslant \left|\frac{n}{|\tilde{\mathcal{N}}_r|}\sum_{I\in\tilde{\mathcal{N}}_r} a^2(I) - \frac{n}{|\mathcal{N}_r|}\sum_{I\in\mathcal{N}_r} a^2(I)\right| \\ &\quad + 2\sum_{k=1}^{r}\left|\frac{n}{|\tilde{\mathcal{N}}_{r+k}|}\sum_{I=(i_0,\ldots,i_{r+k})\in\tilde{\mathcal{N}}_{r+k}} a(i_0,\ldots,i_r)a(i_k,\ldots,i_{r+k})\right. \\ &\quad \left. - \frac{n}{|\mathcal{N}_{r+k}|}\sum_{I=(i_0,\ldots,i_{r+k})\in\mathcal{N}_{r+k}} a(i_0,\ldots,i_r)a(i_k,\ldots,i_{r+k})\right| + d\max_{J\in\mathcal{N}_r}|a(J)|^2 \end{aligned} \tag{2.39}$$

for some positive constant d. Taking into account the bound (2.25), we can apply a similar argument as in (2.30) to every term on the left-hand side of (2.39), except for the last, to obtain:

$$\left|\mathrm{E}\left(\sum_{J\in L^r(\tilde{H})} a(J)\right)^2 - \mathrm{E}W^2\right| \leqslant d_1 \max_{J\in\mathcal{N}_r}|a(J)|^2$$

for some positive constant d_1. This, combined with (2.36) and (2.32), proves (2.23), which finishes the proof of the lemma. □

Next we derive some properties of the statistic $W_{\tilde{A}}$. Under the assumptions of Lemma 2.2, let

$$\bar{\tilde{a}} = \mathrm{E}W_{\tilde{A}}/|\tilde{\mathcal{N}}| = \frac{1}{|\tilde{\mathcal{N}}_r|} \sum_{I \in \tilde{\mathcal{N}}_r} a(I).$$

Proposition 2.1. *In addition to the assumptions of Lemma* 2.2 *and* (1.11)*, assume that*

$$\max_{J \in \mathcal{N}_r} |a(J)| \leqslant \tau,$$

where $\tau > 0$ *satisfies*

$$M_r(\tau^2 + \tau) + M_r^2 \tau^2 \leqslant 1/2. \tag{2.40}$$

Then

$$\mathrm{var}(W_{\tilde{A}}) \geqslant 1/2, \tag{2.41}$$

and there exists a positive constant c, depending only on r, such that

$$\beta_{\tilde{A}} := \sum_{J \in \tilde{\mathcal{N}}_r} \left| \frac{\tilde{a}(J) - \bar{\tilde{a}}}{\sqrt{\mathrm{var}(W_{\tilde{A}})}} \right|^3 \leqslant c\beta_A \tag{2.42}$$

and

$$K_{\tilde{A}} := \frac{|\tilde{\mathcal{N}}|}{|\tilde{\mathcal{N}}_r|} \sum_{J \in \tilde{\mathcal{N}}_r} \left(\frac{\tilde{a}(J) - \bar{\tilde{a}}}{\sqrt{\mathrm{var}(W_{\tilde{A}})}} \right)^2 \leqslant cK_A, \tag{2.43}$$

uniformly in $\mathbf{v}$ *and x.*

The proof of this proposition is based on Lemma 2.2 and a standard application of Jensen's inequality, and therefore we omit it for the sake of brevity.

3. Proof of Theorem 1.1

For any array of real numbers

$$B = \{b(I),\ I \in \mathcal{N}_r\}, \tag{3.1}$$

define statistics W_B analogously to (1.3). Further, under the condition

$$\mathrm{var}\ W_B > 0, \tag{3.2}$$

define the numbers

$$\hat{b}(I) = \frac{b(I) - \mathrm{E}W_B/n}{\sqrt{\mathrm{var}\ W_B}}, \qquad I \in \mathcal{N}_r, \tag{3.3}$$

$$\beta_B := \sum_{I \in \mathcal{N}_r} |\hat{b}(I)|^3, \qquad K_B := \frac{n}{|\mathcal{N}_r|} \sum_{I \in \mathcal{N}_r} \hat{b}(I)^2,$$

and a normalized statistic

$$W_{\hat{B}}(H) = \sum_{I \in L^r(H)} \hat{b}(I), \tag{3.4}$$

where $H \in \mathcal{H}_n$ and $\hat{B} = \{\hat{b}(I),\ I \in \mathcal{N}_r\}$.

Without loss of generality we assume that $K \geqslant 1$. Here we closely follow the basic steps of Bolthausen (1984). For any $n > r$, $\gamma > 0$, and $0 < \tau < 1$ fixed arbitrarily, but to be chosen later on, let $\mathcal{B}_n(\gamma, K, \tau)$ be the set of all arrays (3.1) such that the numbers $\hat{b}(I)$ satisfy the following conditions:

$$|\hat{b}(I)| \leqslant \tau \tag{3.5}$$

uniformly in $I \in \mathcal{N}_r$,

$$\beta_B \leqslant \gamma \tag{3.6}$$

and

$$K_B \leqslant K. \tag{3.7}$$

Further, as in Bolthausen (1984), define

$$h_{z,\lambda}(x) = \left(\left(1 + \frac{z - x}{\lambda}\right) \wedge 1\right) \vee 0,$$

$$h_{z,0}(x) = 1_{(-\infty, z]}(x),$$

and let

$$\delta(\lambda, \gamma, K, \tau, n) = \sup\{|\mathrm{E} h_{z,\lambda}(W_{\hat{B}}) - \Phi(h_{z,\lambda})| : z \in \mathbb{R},\ B \in \mathcal{B}_n(\gamma, K, \tau)\}, \tag{3.8}$$

where

$$\Phi(h) = \mathrm{E} h(Z),$$

with Z being a standard normal random variable. We shall also use the notation

$$\delta(\gamma, K, \tau, n) = \delta(0, \gamma, K, \tau, n).$$

For further reference we record the inequality

$$\delta(\gamma, K, \tau, n) \leqslant \delta(\lambda, \gamma, K, \tau, n) + \frac{\lambda}{\sqrt{2\pi}}. \tag{3.9}$$

Set

$$f_{z,\lambda}(x) = \mathrm{e}^{x^2/2} \int_{-\infty}^{x} (h_{z,\lambda}(y) - \Phi(h_{z,\lambda}))\mathrm{e}^{-y^2/2}\,\mathrm{d}y, \tag{3.10}$$

which satisfies the equation

$$f'_{z,\lambda}(x) - x f_{z,\lambda}(x) = h_{z,\lambda}(x) - \Phi(h_{z,\lambda}). \tag{3.11}$$

Observe that

$$|f_{z,\lambda}(x)| \leqslant 1, \qquad |xf_{z,\lambda}(x)| \leqslant 1, \qquad \text{and } |f'_{z,\lambda}(x)| \leqslant 2, \tag{3.12}$$

for all $x \in \mathbb{R}$ uniformly in z, λ. Moreover, using (3.11) and (3.12), one obtains

$$|f'_{z,\lambda}(x+y) - f'_{z,\lambda}(x)| \leqslant |y| \left(1 + 2|x| + \frac{1}{\lambda} \int_0^1 1_{[z,z+\lambda]}(x+sy) \mathrm{d}s \right). \tag{3.13}$$

For the facts (3.9), (3.11), (3.12) and (3.13) refer to Bolthausen (1984).

Notice that both parts of Theorem 1.1 clearly hold whenever, for some fixed $\varepsilon_0 > 0$,

$$\beta_A > \varepsilon_0 n(r). \tag{3.14}$$

Next we shall prove the following result.

Proposition 3.1. *For any* $0 < \tau < 1/2$*, there exist* $\varepsilon_0 = \varepsilon_0(\tau) > 0$*, and positive constants* a_1 *and* a_2 *independent of* τ*, such that, for any array* A *satisfying* (1.8)*,* (1.10)*,* (1.11)*, and* $\beta_A \leqslant \varepsilon_0 n(r)$*, we have, for all* $n > 2r+1$*,*

$$\sup_{-\infty < x < \infty} |P\{W_A \leqslant x\} - P\{Z \leqslant x\}| \leqslant \delta(a_1\beta_A, 8K, \tau, n) + a_2\tau^{-3}\beta_A / n(r). \tag{3.15}$$

Proof. The proof will be inferred from the following lemma. Let $0 < \tau \leqslant 1/2$ be fixed arbitrarily. Set

$$a'(I) = \begin{cases} a(I), & \text{if } |a(I)| \leqslant \tau, \\ 0, & \text{otherwise,} \end{cases} \tag{3.16}$$

and let $A' = \{a'(I), I \in \mathcal{N}_r\}$. Denote further $\mu_{A'} = \mathrm{E} W_{A'}$.

Lemma 3.1. *Under the assumptions of Proposition* 3.1 *we have*

$$P\{W_A \neq W_{A'}\} \leqslant \frac{\tau^{-3}\beta_A}{n(r)}, \tag{3.17}$$

$$|\mu_{A'}| \leqslant \frac{\tau^{-2}\beta_A}{n(r)}, \tag{3.18}$$

$$\mu_{A'}^2 \leqslant \frac{\tau^{-1}\beta_A}{n(r)}, \tag{3.19}$$

$$|\sigma_A^2 - \operatorname{var} W_{A'}| \leqslant \tau^{-2} \frac{C_r \beta_A}{n(r)}, \tag{3.20}$$

for some positive C_r.

Proof. Define

$$Q = \{I : I \in \mathcal{N}_r, |a(I)| > \tau\},$$

for which clearly

$$|Q| \leqslant \tau^{-3}\beta_A. \tag{3.21}$$

Now

$$P\{W_A \neq W_{A'}\} \leqslant nP\{I \in Q\},$$

where I is sampled from $\mathcal{N}_r$ with probability $1/|\mathcal{N}_r|$. Thus, by (3.21),

$$P\{W_A \neq W_{A'}\} \leqslant \frac{n|Q|}{|\mathcal{N}_r|} \leqslant \frac{\tau^{-3}\beta_A}{n(r)},$$

which proves (3.17). Clearly,

$$|\mu_{A'}| = \left| n(r)^{-1} \sum_{I \in Q} a(I) \right| \leqslant \frac{\tau^{-2}\beta_A}{n(r)}, \tag{3.22}$$

from which we derive, for any $0 < \varepsilon_0 < \tau^3$, that $|\mu_{A'}| \leqslant \tau$. These bounds yield (3.18) and (3.19).

To finish the proof of Lemma 3.1 we introduce some notation. Let $\Omega(I)$, $I \in \mathcal{N}_r$, be an indexed class of subsets of $\mathcal{N}_r$, and for $I' \in \mathcal{N}_r$ set

$$\Omega'(I') = \{I \in \mathcal{N}_r : I' \in \Omega(I)\}.$$

Further, let

$$\Omega^* = \max_{I \in \mathcal{N}_r} |\Omega(I)|, \qquad \Omega^{**} = \max_{I' \in \mathcal{N}_r} |\Omega'(I')|, \qquad \overline{\Omega} = \max\{\Omega^*, \Omega^{**}\}.$$

Claim. *With the above notation,*

$$\left| \sum_{I \in Q} a(I) \sum_{I' \in \Omega(I)} a(I') \right| \leqslant \tau^{-2}\beta_A \overline{\Omega}. \tag{3.23}$$

Proof. Notice that

$$\left| \sum_{I \in Q} a(I) \sum_{I' \in \Omega(I)} a(I') \right| \leqslant \left| \sum_{I \in Q} a(I) \sum_{I' \in \Omega(I), I' \notin Q} a(I') \right| + \left| \sum_{I \in Q} a(I) \sum_{I' \in \Omega(I), I' \in Q} a(I') \right| =: S_1 + S_2.$$

By (3.22),

$$S_1 \leqslant \tau\Omega^* \sum_{I \in Q} |a(I)| \leqslant \tau^{-1}\Omega^*\beta_A,$$

and, using the trivial inequality $x^2 + y^2 \geqslant 2|xy|$, we obtain

$$S_2 \leqslant \frac{1}{2} \sum_{I, I' \in Q, I' \in \Omega(I)} a^2(I) + \frac{1}{2} \sum_{I, I' \in Q, I' \in \Omega(I)} a^2(I') \leqslant \frac{1}{2}(\Omega^* + \Omega^{**}) \sum_{I' \in Q} a^2(I') \leqslant \tau^{-1}\overline{\Omega}\beta_A.$$

These bounds prove our claim. □

For each $1 \leqslant k \leqslant r$ and $I = (I_0, \dots, I_r) \in \mathcal{N}_r$, set

$$\Omega_k(I) = \{I' \in \mathcal{N}_r : I' = (I'_0, \dots, I'_r) = (I'_0, \dots, I'_{k-1}, I_k, \dots, I_r)\}.$$

Elementary combinatorics show that, for all $I, I' \in \mathcal{N}_r$ and $1 \leqslant k \leqslant r$,

$$|\Omega_k(I)| \leqslant n^k, \qquad |\Omega'_k(I')| \leqslant n^k.$$

We see then that, by (2.37) and (2.38),

$$\begin{aligned}|\mathrm{E}W_A^2 - \mathrm{E}W_{A'}^2| \leqslant {} & \left(\frac{n}{|\mathcal{N}_r|} - \frac{n(n-(2r+1))}{|\mathcal{N}_{2r+1}|}\right) \sum_{I \in Q} a^2(I) \\ & + 2\sum_{k=1}^{r} \left(\frac{n}{|\mathcal{N}_{r+k}|} - \frac{n(n-(2r+1))}{|\mathcal{N}_{2r+1}|}\right) \left| \sum_{I \in Q} a(I) \sum_{I' \in \Omega_k(I)} a(I') \right| \\ & + (\mathrm{E}W_{A'})^2 \frac{n(n-(2r+1))}{|\mathcal{N}_{2r+1}|} \left(\frac{|\mathcal{N}_r|}{n}\right)^2,\end{aligned}$$

which by r applications of the above claim and by (3.19) is

$$\leqslant \frac{k_r \tau \beta_A}{n(r)} + \frac{2\tau^{-2}\beta_A}{n(r)} \sum_{k=1}^{r} \frac{n^k}{(n-r-1)\dots(n-r-k)} \leqslant \tau^{-2} D_r \frac{\beta_A}{n(r)}, \tag{3.24}$$

where k_r and D_r are finite positive constants independent of n. Now, by (3.19) and (3.24), we obtain that

$$|\sigma_A^2 - \operatorname{var} W_{A'}| \leqslant |\mathrm{E}W_A^2 - \mathrm{E}W_{A'}^2| + \mu_{A'}^2 \leqslant \tau^{-2} D_r \frac{\beta_A}{n(r)} + \tau^{-1} \frac{\beta_A}{n(r)} \leqslant C_r \tau^{-2} \frac{\beta_A}{n(r)},$$

for some $C_r > 0$, which completes the proof of Lemma 3.1. □

Returning to the proof of Proposition 3.1, fix $0 < \tau < 1/2$ in Lemma 3.1 arbitrarily. By (3.20),

$$|1 - \operatorname{var} W_{A'}| \leqslant \tau^{-2} C_r \frac{\beta_A}{n(r)} \leqslant \tau^{-2} C_r \varepsilon_0.$$

Thus if $\varepsilon_0 < \tau^2/(4C_r)$, we have $\operatorname{var} W_{A'} > 1/4$. This, together with (3.18), gives, for all $I \in \mathcal{N}_r$,

$$\left| \frac{a'(I) - \mu_{A'}/n}{\sqrt{\operatorname{var} W_{A'}}} \right| \leqslant 2(|a'(I)| + |\mu_{A'}|/n) \leqslant 2(\tau + \tau^{-2}\varepsilon_0).$$

It is easy to check now, using Jensen's inequality, that if $\varepsilon_0 < \tau^3 \min\{1/(4C_r), 1\}$ then

$$A' \in \mathcal{B}_n(64\beta_A, 8K, 4\tau). \tag{3.25}$$

Clearly, now

$$\sup_{-\infty<x<\infty} |P\{W_A \leqslant x\} - P\{Z \leqslant x\}| \leqslant \sup_{-\infty<x<\infty} |P\{W_{A'} \leqslant x\} - P\{Z \leqslant x\}| + P\{W_A \neq W_{A'}\}$$
$$\leqslant \sup_{-\infty<x<\infty} |P\{W_{A'} \leqslant x\} - P\{Z \leqslant x\}| + \tau^{-3}\beta_A/n(r) \tag{3.26}$$

by (3.17). Next note that

$$\begin{aligned}
&\sup_{-\infty<x<\infty} |P\{W_{A'} \leqslant x\} - P\{Z \leqslant x\}| \\
&\quad \leqslant \sup_{-\infty<x<\infty} \left| P\left\{ \frac{W_{A'} - \mu_{A'}}{\sqrt{\operatorname{var} W_{A'}}} \leqslant \frac{x - \mu_{A'}}{\sqrt{\operatorname{var} W_{A'}}} \right\} - P\left\{ Z \leqslant \frac{x - \mu_{A'}}{\sqrt{\operatorname{var} W_{A'}}} \right\} \right| \\
&\qquad + \sup_{-\infty<x<\infty} |P\{Z \leqslant (x - \mu_{A'})/\sqrt{\operatorname{var} W_{A'}}\} - P\{Z \leqslant x\}| \\
&\quad \leqslant \delta(64\beta_A, 8K, 4\tau, n) + \sup_{-\infty<x<\infty} |P\{Z \leqslant (x - \mu_{A'})/\sqrt{\operatorname{var} W_{A'}}\} - P\{Z \leqslant x\}|
\end{aligned} \tag{3.27}$$

by (3.25). Obviously,

$$\sup_{-\infty<x<\infty} |P\{Z \leqslant (x - \mu_{A'})/\sqrt{\operatorname{var} W_{A'}}\} - P\{Z \leqslant x\}| \leqslant C_1(|\mu_{A'}| + |1 - \operatorname{var} W_{A'}|),$$

for some absolute constant C_1, which by (3.18) and (3.20), is readily seen to be

$$\leqslant D\frac{\tau^{-2}\beta_A}{n(r)}, \tag{3.28}$$

for some constant D. Combining inequalities (3.26), (3.27) and (3.28), we obtain (3.15) after an obvious adjustment of the constants. □

In order to complete the proof of Theorem 1.1 it will now be enough to show that, at least for some $0 < \tau < 1/2$ and $c_0 > 0$,

$$\delta(a_1\beta_A, 8K, \tau, n) \leqslant c_0\beta_A/n(r) \tag{3.29}$$

whenever $\beta_A \leqslant \varepsilon_0 n(r)$.

From now on we set $K' = 8K$ and

$$\gamma = a_1\beta_A \leqslant a_1\varepsilon_0 n(r). \tag{3.30}$$

Let $\mathcal{B}_n^0(\gamma, K', \tau)$ denote the subset of $\mathcal{B}_n(\gamma, K', \tau)$ of the arrays B such that

$$\sum_{I\in\mathcal{N}_r} b(I) = 0 \quad \text{and} \quad \sigma_B^2 = 1,$$

in which case $\hat{b}(I) = b(I)$, implying $W_B = W_{\hat{B}}$,

$$\mathrm{E}W_B = 0, \qquad \operatorname{var} W_B = \sigma_B^2 = 1, \tag{3.31}$$

$$\sum_{I\in\mathcal{N}_r}|b(I)|^3=\beta_B\leqslant\gamma \quad \text{and} \quad \frac{n}{|\mathcal{N}_r|}\sum_{I\in\mathcal{N}_r}b^2(I)=K_B\leqslant K'. \tag{3.32}$$

To prove (3.29), assume that $B\in\mathcal{B}_n^0(\gamma, K', \tau)$, and consider

$$W_{B,i}=W_B(H_i), \qquad i=1, 2, \tag{3.33}$$

where H_1 and H_2 are defined by Lemma 2.1 above. Recalling (2.5), we obtain

$$W_B(H)\stackrel{d}{=}W_{B,1}\stackrel{d}{=}W_{B,2}. \tag{3.34}$$

Furthermore, by (2.6)

$$\begin{aligned}
&\mathrm{E}[W_B(H)f_{z,\lambda}(W_B(H))] \qquad (3.35)\\
&\quad= n\mathrm{E}[b(I_0, J_1, \ldots, J_r)f_{z,\lambda}(W_{B,2})]\\
&\quad= n\mathrm{E}[b(I_0, J_1, \ldots, J_r)f_{z,\lambda}(W_{B,1})]+n\mathrm{E}[b(I_0, J_1, \ldots, J_r)(W_{B,2}-W_{B,1})f'_{z,\lambda}(W_B)]\\
&\qquad+n\mathrm{E}\Big[b(I_0, J_1, \ldots, J_r)(W_{B,2}-W_{B,1})\\
&\qquad\times\int_0^1(f'_{z,\lambda}(W_B+(W_{B,1}-W_B)+t(W_{B,2}-W_{B,1}))-f'_{z,\lambda}(W_B))\mathrm{d}t\Big].
\end{aligned}$$

By property (iv) from Lemma 2.1 the vector $(I_0, J_1, \ldots, J_r)$ is independent of $W_{B,1}$, therefore by (2.6) and (3.31), together with (3.34), we have

$$n\mathrm{E}[b(I_0, J_1, \ldots, J_r)f_{z,\lambda}(W_{B,1})]=\mathrm{E}W_{B,1}\mathrm{E}f_{z,\lambda}(W_{B,1})=0.$$

Observe that $W_{B,2}-W_{B,1}$ is a function of $L^r(H_1)\triangle L^r(H_2)$, which, by property (iii) of our coupling, is independent of W_B. Hence, by (2.5), (3.31), (3.34), and property (iv),

$$\begin{aligned}
n\mathrm{E}[b(I_0, J_1, \ldots, J_r)(W_{B,2}-W_{B,1})f'_{z,\lambda}(W_B)]&=n\mathrm{E}[b(I_0, J_1, \ldots, J_r)W_{B,2}]\mathrm{E}f'_{z,\lambda}(W_B)\\
&=\mathrm{var}(W_B)\mathrm{E}f'(W_B)=\mathrm{E}f'_{z,\lambda}(W_B).
\end{aligned}$$

Putting everything together now and using (3.11), we obtain

$$\begin{aligned}
|\mathrm{E}h_{z,\lambda}(W_B)-\Phi(h_{z,\lambda})| &= n\mathrm{E}\Big|b(I_0, J_1, \ldots, J_r)(W_{B,2}-W_{B,1}) \qquad (3.36)\\
&\quad\times\int_0^1(f'_{z,\lambda}(W_B+(W_{B,1}-W_B)+t(W_{B,2}-W_{B,1}))-f'_{z,\lambda}(W_B))dt\Big|,
\end{aligned}$$

which in turn, by (3.13), is

$$\leqslant n\mathrm{E}[|b(I_0, J_1, \ldots, J_r)(W_{B,2} - W_{B,1})|(|W_{B,1} - W_B| + |W_{B,2} - W_{B,1}|)] \tag{3.37}$$
$$+ 2n\mathrm{E}|b(I_0, J_1, \ldots, J_r)(W_{B,1} - W_{B,2})(|W_{B,1} - W_B| + |W_{B,2} - W_{B,1}|)W_B|$$
$$+ n\mathrm{E}\bigg[|b(I_0, J_1, \ldots, J_r)(W_{B,2} - W_{B,1})|(|W_{B,1} - W_B| + |W_{B,2} - W_{B,1}|).$$
$$\times \frac{1}{\lambda}\int_0^1\int_0^1 \mathbf{1}_{[z,z+\lambda]}(W_B + s(W_{B,1} - W_B) + ts(W_{B,2} - W_{B,1}))\mathrm{d}s\,\mathrm{d}t\bigg]$$
$$=: \Delta_1 + \Delta_2 + \Delta_3.$$

Notice that

$$\Delta_1 \leqslant n\mathrm{E}b(I_0, J_1, \ldots, J_r)\sum |b(I^1)b(I^2)|,$$

where the sum runs over $I^1 \in L^r(H_1)\triangle L^r(H) \cup L^r(H_1)\triangle L^r(H_2)$ and $I^2 \in L^r(H_1)\triangle L^r(H_2)$, therefore by property (v) from Lemma 2.1 this sum contains at most a finite number of terms. It is not difficult to derive from here under assumption (3.32) that, for some positive constant c_1 independent of n,

$$\Delta_1 \leqslant nc_1 \sum_{I\in\mathcal{N}_r} \frac{|b(I)|^3}{|\mathcal{N}_r|} \leqslant \frac{c_1\gamma}{n(r)}. \tag{3.38}$$

To bound the remaining two terms in (3.37) we shall use conditioning. Consider

$$\Delta_2 = 2n\mathrm{EE}\{|b(I_0, J_1, \ldots, J_r)(W_{B,1} - W_{B,2})W_B| \tag{3.39}$$
$$\times (|W_{B,1} - W_B| + |W_{B,2} - W_{B,1}|)|L^r(H_1)\triangle L^r(H), L^r(H_1)\triangle L^r(H_2), (I_0, J_1, \ldots, J_r)\}.$$

Notice that the difference $W_{B,1} - W_{B,2}$ is a function of $L^r(H_1)\triangle L^r(H_2)$, while $W_B - W_{B,1}$ is a function of $L^r(H_1)\triangle L^r(H)$. Hence, properties (iv) and (v) allow us to derive from (3.39) that

$$\Delta_2 = 2n\mathrm{E}\{|b(I_0, J_1, \ldots, J_r)(W_{B,1} - W_{B,2})|(|W_{B,1} - W_B| + |W_{B,2} - W_{B,1}|) \tag{3.40}$$
$$\times \mathrm{E}\{|W_B|\,|L^r(H)\triangle L^r(H_1), L^r(H_1)\triangle L^r(H_2), (I_0, J_1, \ldots, J_r)\}\}.$$

Consider the last conditional expectation. Assume that an event

$$\mathcal{A} = \{L^r(H)\backslash L^r(H_1) = \mathcal{U}_1, L^r(H_1)\backslash L^r(H) = \mathcal{U}_1',$$
$$L^r(H_1)\backslash L^r(H_2) = \mathcal{U}_2, L^r(H_2)\backslash L^r(H_1) = \mathcal{U}_2', (I_0, J_1, \ldots, J_r) = (i_0, j_1, \ldots, j_r)\}$$

has a positive probability. Since all the cycles H, H_1 and H_2 are defined on the same set of vertices, and also $(I_0, J_1, \ldots, J_r) \in H_2$, then for any event $\mathcal{A}$ there is a unique non-empty set of vectors, say $\{v^0, \ldots, v^t\}$, with $v^i = (v_0^i, \ldots, v_{p_i}^i)$ and $p_i \geqslant r$ for any $0 \leqslant i \leqslant t$, such that

$$\mathrm{E}\{|W_B(H)|\,|\mathcal{A}\} = \mathrm{E}\left\{|W_B(H)|\,|\bigcup_{q=0}^{t}\Gamma(v^q)\subset H,\,(v^i_{p_i},\,v^j_0)\notin H,\,0\leqslant i,\,j\leqslant t\right\}. \tag{3.41}$$

Notice also that the condition $P\{\mathcal{A}\}>0$ implies $|\{v^0_0,\,\dots,\,v^t_{p_t}\}|<(2r+2)C=\kappa$ according to property (v) of our coupling.

Now we can use the results of Lemma 2.2, taking into account that $0<\tau<1$ in (3.5), in order to obtain the bound

$$\mathrm{E}\{|W_B|\,|\mathcal{A}\}\leqslant M,$$

where the constant $M>0$ is independent of n and $\mathcal{A}$. This, together with (3.40) and the definition of $\mathcal{A}$, gives us

$$\Delta_2\leqslant 2M\Delta_1\leqslant\frac{2Mc_1\gamma}{n(r)}. \tag{3.42}$$

Finally, we shall find an upper bound for Δ_3. Set

$$\begin{aligned}\alpha(\lambda,\,\gamma,\,K',\,\tau,\,n)\\ &=\sup\left\{P\left\{W_B(H)\in[z,\,z+\lambda]\,\middle|\,\bigcup_{q=0}^{t}\Gamma(v^q)\subset H,\,(v^i_{p_i},\,v^j_0)\notin H,\,0\leqslant i,\,j\leqslant t\right\}:\right.\\ &\qquad\left. r<|\{v^0_0,\,\dots,\,v^t_{p_t}\}|<\kappa,\,z\in\mathbb{R},\,B\in\mathcal{B}^0_n(\gamma,\,K',\,\tau)\right\}.\end{aligned}$$

Analogously to (3.40)–(3.42) we obtain

$$\Delta_3\leqslant\frac{1}{\lambda}\Delta_1\alpha(\lambda,\,\gamma,\,K',\,\tau,\,n). \tag{3.43}$$

Let us fix $0<\tau<1/2$ in (3.5) such that condition (2.40) is satisfied. Hence, from now on τ is a positive constant depending on r only. Then using (2.21) and the results of Proposition 2.1, we obtain the bound

$$\alpha(\lambda,\,\gamma,\,K',\,\tau,\,n)\leqslant\sup\{P\{W_{B_{n-m}}\in[z,\,z+\lambda]\}:z\in\mathbb{R},\,B\in\mathcal{B}'_{n-m},\,r+1\leqslant m\leqslant\kappa\}, \tag{3.44}$$

where $\mathcal{B}'_{n-m}$ denotes the set of all arrays of real numbers

$$B_{n-m}=\{b(I),\,I=(I_0,\,\dots,\,I_r),\,I_i\in\{1,\,\dots,\,n-m\},\,I_i\neq I_j\text{ for any }i\neq j\},$$

such that

$$\mu_{B_{n-m}}=0,\qquad \mathrm{var}(W_{B_{n-m}})\geqslant 1/2,\qquad \beta_{B_{n-m}}\leqslant c\gamma,\qquad K_{B_{n-m}}\leqslant cK'. \tag{3.45}$$

Notice that, for any $B_{n-m}\in\mathcal{B}'_{n-m}$,

$$\sup_z P\{W_{B_{n-m}}\in[z,\,z+\lambda]\}\leqslant\sup_z P\{W_{\hat{B}_{n-m}}\in[z,\,z+2\lambda]\}, \tag{3.46}$$

where we use the notation $\hat{B}$ as in (3.4). Assuming now that ε_0 is chosen sufficiently small, we can again use (3.15), taking into account (3.45) and (3.30), to infer that the last term in (3.46) is

$$\leqslant 2\delta(a'c\gamma, 8cK', \tau, n-m) + \frac{2a_2'\tau^{-3}c\gamma}{(n-m)(r)} + \frac{2\lambda}{\sqrt{2\pi}}, \tag{3.47}$$

where τ is the same as fixed above, a' and a_2' are some positive constants, and $(n-m)(r) := (n-m-1)\dots(n-m-r)$. Combining (3.44), (3.46), and (3.47), we now obtain

$$\alpha(\lambda, \gamma, K', \tau, n) \leqslant 2 \max_{1\leqslant m\leqslant\kappa} \delta(a'c\gamma, 8cK', \tau, n-m) + a_3\frac{\gamma}{n(r)} + \frac{2\lambda}{\sqrt{2\pi}}$$

for some $a_3 > 0$. Substituting the last bound and (3.38) into (3.43), we obtain

$$\Delta_3 \leqslant c_3\frac{\gamma}{n(r)}\left(1 + \frac{\gamma}{\lambda n(r)} + \frac{1}{\lambda}\max_{1\leqslant m\leqslant\kappa} \delta(a'c\gamma, 8cK', \tau, n-m)\right) \tag{3.48}$$

for some constant $c_3 > 0$. Combining (3.48), (3.38), (3.42) with (3.37) and (3.36), and taking into account (3.8) and (3.9), we obtain the inequality

$$\delta(\gamma, K', \tau, n) \leqslant c_4\frac{\gamma}{n(r)}\left(1 + \frac{\gamma}{\lambda n(r)} + \frac{1}{\lambda}\max_{1\leqslant m\leqslant\kappa} \delta(a'c\gamma, 8cK', \tau, n-m)\right) + \frac{\lambda}{\sqrt{2\pi}} \tag{3.49}$$

for some constant $c_4 > 0$.

Without loss of generality we can assume from now on that $c > 1$ and $a' > 1$. Choosing

$$\lambda = \frac{24a'c^2c_4\gamma}{n(r)},$$

we derive from (3.49) that for some constant c_5 independent of n, and for all $K' \geqslant 1$,

$$\delta(\gamma, K', \tau, n) \leqslant c_5\frac{\gamma K'}{n(r)} + \frac{1}{24c^2a'}\max_{1\leqslant m\leqslant\kappa} \delta(a'c\gamma, 8cK', \tau, n-m).$$

Next observe that (2.35), in combination with the Cauchy–Schwarz inequality, shows that

$$(\beta_A)^{2/3}|\mathcal{N}_r|^{1/3} \geqslant \sum_{I\in\mathcal{N}_r} |a(I)|^2 \geqslant n(r)\sigma_A^2/(2r+1) = n(r)/(2r+1), \tag{3.50}$$

which says that for all $n \geqslant n_0$, for some n_0, we have $\gamma = a'\beta_A \geqslant 1$. Further, for all $n \geqslant n_1$, for some n_1,

$$\max_{1\leqslant m\leqslant\kappa} \frac{n(r)}{(n-m)(r)} \leqslant \frac{3}{2}.$$

This implies, for all $n > 2\kappa \vee n_0 \vee n_1$,

$$\frac{n(r)\delta(\gamma, K', \tau, n)}{\gamma K'} \leqslant c_5 + \frac{1}{2}\max_{1\leqslant m\leqslant\kappa} \frac{(n-m)(r)\delta(a'c\gamma, 8cK', \tau, n-m)}{a'c\gamma 8cK'},$$

which, in turn, implies, since $c > 1$, $1 \leqslant \gamma \leqslant a_1\varepsilon_0 n(r)$ and $K' \geqslant 1$, that for all n sufficiently large,

$$\sup_{\gamma \geqslant 1, K \geqslant 1} \frac{n(r)\delta(\gamma, K, \tau, n)}{\gamma K} \leqslant c_5 + \frac{1}{2} \max_{1 \leqslant m \leqslant \kappa} \sup_{\gamma \geqslant 1, K \geqslant 1} \frac{(n-m)(r)\delta(\gamma, K, \tau, n-m)}{\gamma K}.$$

This last inequality readily implies (3.29), which by (3.15) finishes the proof of Theorem 1.1 for W_A. □

4. Proof of Lemma 2.1

4.1. Main graph construction

Throughout this section we assume that $n > 2r(r+2)$.

Given a cycle H and vectors $(I_0, I_1, \dots, I_r)$ and $(I_0, J_1, \dots, J_r)$, let us construct cycles H_1 and H_2. First of all, notice that each of the latter cycles should be some modification of the cycle H which preserves all but a finite (in n) number of the edges of the cycle H in order to possess property (v). Obviously, in order to have property (ii) the cycle H_1 should pass through the path $(I_0, I_1, \dots, I_r)$, while H_2 should incorporate the path $(I_0, J_1, \dots, J_r)$. That would be a fairly easy task to accomplish, namely to change a finite number of the edges of the cycle H in order to pass through a given r-path. However, the problem becomes much more complicated when we try to achieve property (iii) as well. To see where the difficulty comes from, let us assume that we have two cycles H_1 and H_2 satisfying (ii) and (v), which means that $H_1 \triangle H = B_1$ and $H_2 \triangle H = B_2$, where B_1 and B_2 are finite (i.e. bounded uniformly in n) non-empty sets of edges. Hence, the set $L^r(H_1) \triangle L^r(H_2)$ might contain those r-paths of H which have at least one vertex in common with the vertices in B_1 or B_2. This clearly violates property (iii), which states that $L^r(H_1) \triangle L^r(H_2)$ and H are independent. Our way to overcome this obstacle is to introduce auxiliary random elements (denoted by V below) which will take care of the r-boundaries of the set $\{I, J\} := \{I_0, I_1, \dots, I_r\} \cup \{J_1, \dots, J_r\}$ in the graphs H_1 and H_2. An r-boundary of a vertex v in a cycle is called the set of all the vertices of this cycle reachable from v along a connected path of at most r (non-directed) edges. We call an r-boundary of a set the union of the r-boundaries of the vertices in this set without the vertices of the set itself. Our aim now is to construct cycles H_1 and H_2 so that the r-boundaries of the set $\{I, J\}$ in either of these cycles are independent of H.

Let us outline our strategy. First we define, independently of H, a new random element $V = V(I, J)$ to represent an r-boundary of $\{I, J\}$ in the cycle H_1. Further, we determine a set $\mathbf{U}$ of paths through which cycle H_1 should pass in order to have the given boundary of the set $\{I, J\}$. Finally, we will introduce an algorithm to modify a cycle H as little as possible, in order to obtain a new cycle $G(H, \mathbf{U})$ which passes through a given collection of paths. We will conclude by proving that this algorithm produces random cycles with required properties (i)–(v).

For any $m \geqslant 1$ and any vector $u = (u_1, \dots, u_m)$, we shall denote by $\{u\} = \{u_1, \dots, u_m\}$ the set of values of its coordinates. Also, for any array of vectors $\mathbf{U} = \{u^1, \dots, u^m\}$ we denote $\{\mathbf{U}\} = \cup_{i=0}^{m} \{u^i\}$.

Let $\{I = (I_0, \ldots, I_r)$ and $J = (J_1, \ldots, J_r)$ be any fixed realizations of the corresponding random vectors, and let h be a cycle passing through $\Gamma(I)$. Write

$$h = \{(I_0, I_1), \ldots, (I_r, h_{r+1}), \ldots, (h_{n-1}, I_0)\}. \tag{4.1}$$

Consider the r-boundary of the set $\{I, J\}$ in this cycle. Define

$$k(I, J) := |\{J\}\backslash\{I\}| \tag{4.2}$$

to be the number of components of J which are not in the set $\{I\}$.

In the simplest situation when $k(I, J) = 0$, that is, when $\{J\} \subset \{I\}$, the r-boundary of $\{I, J\}$ in a cycle (4.1) is clearly $\{h_{n-r}, \ldots, h_{n-1}, h_{r+1}, \ldots, h_{2r}\}$. Observe that the location of this r-boundary, that is, the indices $\{n-r, \ldots, n-1, r+1, \ldots, 2r\}$, is determined uniquely and independent of this particular cycle passing through I. Thus in the case $k(I, J) = 0$ we define $V = V(I, J)$ to be a random vector with $2r$ components indexed as the vertices of the r-boundary of $\{I, J\}$,

$$V = V^0 = (V^0_{r+1}, \ldots, V^0_{2r}, V^0_{n-r}, \ldots, V^0_{n-1}), \tag{4.3}$$

and distributed uniformly over the set $\{(x_1, \ldots, x_{2r}) : x_i \in \mathcal{N}\backslash\{I\}, x_i \neq x_j\}$.

In the case $k(I, J) = k \geqslant 1$ we first introduce a random vector $(\xi_1, \ldots, \xi_k)$ uniformly distributed on the array

$$\Omega_k := \{(u_1, \ldots, u_k) : r + 1 \leqslant u_1 < \ldots < u_k \leqslant n-1, u_i \neq u_j\}, \tag{4.4}$$

to specify later on the locations of the values of $\{J\}\backslash\{I\}$ in a cycle H_1. More exactly, write

$$\{J\}\backslash\{I\} = \{J_{s_1}, \ldots, J_{s_k}\}, \qquad \text{with } s_1 < \ldots < s_k. \tag{4.5}$$

Then in a cycle H_1 written as (4.1), we shall have

$$h_{\xi_l} = J_{s_l} \tag{4.6}$$

for any $1 \leqslant l \leqslant k$.

Now we are ready to define $V(I, J)$ in the general case. Conditionally on I, J such that $k(I, J) = k \geqslant 0$, define a random vector

$$V|_{k(I,J)=k} = (V^0, \ldots, V^k), \tag{4.7}$$

so that, conditionally on $\xi = (\xi_0, \xi_1, \ldots, \xi_k, \xi_{k+1})$ with

$$\xi_0 \equiv r, \ \xi_{k+1} \equiv n-1, \quad \text{and} \quad (\xi_1, \ldots, \xi_k) \in \Omega_k \text{ if } k \geqslant 1,$$

one has, for all $0 \leqslant l \leqslant k$,

$$V^l = \begin{cases} (V^l_{\xi_l+1}, \ldots, V^l_{\xi_l+r}, V^l_{\xi_{l+1}-r}, \ldots, V^l_{\xi_{l+1}-1}), & \text{if } \xi_{l+1} - \xi_l > 2r, \\ (V^l_{\xi_l+1}, \ldots, V^l_{\xi_{l+1}-1}), & \text{if } 1 < \xi_{l+1} - \xi_l \leqslant 2r, \\ \varnothing, & \text{if } \xi_{l+1} - \xi_l = 1 \end{cases}$$

(assuming $\varnothing$ makes no contribution to (4.7)), and the distribution of vector V is uniform on

$$\mathcal{V}(m) := \{(u_1, \ldots, u_m) \in (\mathcal{N}\backslash\{I, J\})^m : u_i \neq u_j \text{ for any } i \neq j\} \tag{4.8}$$

with

$$m = \sum_{l=0}^{k} ((\xi_{l+1} - \xi_l - 1) \wedge 2r). \tag{4.9}$$

Notice that the indices of all V^l will correspond to a location of the r-boundary of the set $\{I, J\}$ in the cycle H_1, that is, if H_1 is written in the form (4.1) we shall have

$$V_i^l = h_i \tag{4.10}$$

for all possible i and l.

Let us now gather the random variables I, J and V into one vector (recall also (4.5))

$$\begin{array}{ll} (I_0, I_1, \ldots, I_r, V^0, J_{s_1}, V^1, J_{s_2}, \ldots, V^{k-1}, J_{s_k}, V^k), & \text{if } k = k(I, J) > 0, \\ (I_0, I_1, \ldots, I_r, V^0), & \text{otherwise.} \end{array} \tag{4.11}$$

Since by our construction the indices of the components of vector (4.11) will represent the location of the set $\{I, J\}$ and its r-boundary in the cycle H_1, we can also determine the corresponding connected paths as follows. In the simplest case $k(I, J) = 0$ after a cyclic permutation of (4.11) we obtain, taking into account (4.3), a vector

$$\mathbf{U} = U^0 = (V^0_{n-r}, \ldots, V^0_{n-1}, I_0, I_1, \ldots, I_r, V^0_{r+1}, \ldots, V^0_{2r}), \tag{4.12}$$

which by its definition (recall (4.10)) will correspond to the directed $3r$-path in the cycle H_1 written as (4.1).

In the general case when $k(I, J) > 0$ we see that the only consecutive vertices in the vector (4.11) which might not be neighbours in the cycle H_1 (recall (4.6) and (4.10)), are the vertices $V^l_{\xi_l + r}$ and $V^l_{\xi_{l+1} - r}$ if $\xi_{l+1} - r > \xi_l + r$. Observe that, for any $n > 2r(r+2)$, the set

$$\{l_0, \ldots, l_M\} = \{0 \leqslant l \leqslant k : \xi_{l+1} - \xi_l - 1 \geqslant 2r\} \tag{4.13}$$

is non-empty, and we can assume $0 \leqslant l_0 < \ldots < l_M \leqslant k$, where $0 \leqslant M \leqslant k$. Now let vector $\mathbf{U}$ be a cyclic permutation of the vector (4.11), call it

$$\mathbf{U} = \mathbf{U}(I, J, \xi, V) = (U^0, \ldots, U^M)$$

with $U^i = (U^i_0, \ldots, U^i_{p_i})$, $i = 0, \ldots, M$, such that

$$U^i_0 = V^l_{\xi_{l+1} - r} \qquad \text{for } l = l_i,\ i = 0, \ldots, M.$$

According to this definition $\mathbf{U} \in \mathcal{N}_{r+k+m}$, and its components $U^0, \ldots, U^M$ will correspond to the connected components of the set $\{I, J\}$ together with a given boundary $V(I, J, \xi)$ in the cycle H_1. Clearly, vector $p = (p_0, \ldots, p_M)$ is a deterministic function of I, J and ξ. We shall write

$$p = p(I, J, \xi), \qquad M = M(I, J, \xi) \tag{4.14}$$

for further reference. Notice that by their definition $p_i \geqslant 1$ for all $0 \leqslant i \leqslant M$, and

$$|\{\mathbf{U}\}| = M + \sum_{i=0}^{M} p_i = r + k + m \leqslant 2r(r+2). \tag{4.15}$$

Now for any array

$$\mathbf{u} = (u^0, \ldots, u^M), \qquad \text{with } u^q = (u_0^q, \ldots, u_{p_q}^q),\ 0 \leqslant q \leqslant M, \tag{4.16}$$

such that $\{P\{\mathbf{U} = \mathbf{u}\} > 0$, we shall define a transformation

$$G(\mathbf{u}, \cdot) : \mathcal{H}_n \to \mathcal{H}_n \tag{4.17}$$

as follows. Let $H \in \mathcal{H}_n$ be fixed arbitrarily.

(i) If $\Gamma(\mathbf{u}) \subset H$, that is, if H passes through every path of $\Gamma(\mathbf{u})$ (see definition (2.9)), set

$$G(\mathbf{u}, H) = H.$$

(ii) Otherwise we construct the cycle $G(\mathbf{u}, H)$ iteratively, by adding new edges to the path $\Gamma(u^0)$. (The reader might find it helpful to draw a picture.) We shall denote by G_k, $k \leqslant n$, the current k-path. We start with

$$G_{p_0} = \Gamma(u^0). \tag{4.18}$$

Then from the last vertex of this path, which is $u_{p_0}^0$, we draw a new edge to the first vertex of H which we missed starting at u_0^0. To be precise, we choose this vertex according to the following procedure. First we define recursively, using definition (2.11),

$$E_H^1(i) = E_H(i), \qquad E_H^k(i) = E_H^1(E_H^{k-1}(i)), \qquad k > 1, \tag{4.19}$$

for all $i \in \{1, \ldots, n\}$, so that $E_H^k(i)$ is the end vertex of the k-path from vertex i along the cycle H. Now, for any $w \in \mathcal{N}$, we can define a number

$$\mu(w) = \mu(w, H, \mathbf{u}) := \min\left\{ m \geqslant 1 : E_H^m(w) \notin \bigcup_{q=0}^{M} (\{u^q\} \backslash \{u_0^q\}) \right\}, \tag{4.20}$$

which is the length of the shortest path in H from vertex w to the vertex outside the set $\cup_{q=0}^{M}(\{u^q\}\backslash\{u_0^q\})$. Now we choose

$$u'_{p_{0+1}} := E_H^{\mu(u_0^0)}(u_0^0) \tag{4.21}$$

and then set

$$G_{p_0+1} := G_{p_0} \cup \{(u_{p_0}^0, u'_{p_0+1})\}.$$

Algorithm. *Assume that we have constructed a k-path*

$$G_k := \{(u'_0, u'_1), (u'_1, u'_2), \ldots, (u'_{k-1}, u'_k)\},$$

for some $p_0 + 1 \leqslant k < n$. *Then there are two following cases.*

Case A. If the last vertex of the current path coincides with the first vertex of one of the paths in $\Gamma(\mathbf{u})$, we add the corresponding path to our current path. More precisely, if $u'_k = u^l_0$ for some $1 \leqslant l \leqslant M$ then we set

$$G_{k+p_l+1}(\mathbf{u}, H) := G_k(\mathbf{u}, H) \cup \Gamma(u^l) \cup \left\{ \left(u^l_{p_l}, E_H^{\mu(u^l_0)}(u^l_0) \right) \right\}, \tag{4.22}$$

and proceed again from the beginning of the algorithm.

Case B. Otherwise, if $u'_k \notin \cup_{q=0}^{M} \{u^q_0\}$, choose

$$u'_{k+1} := E_H^{\mu(u'_k)}(u'_k) \tag{4.23}$$

and set

$$G_{k+1} := G_k \cup \{(u'_k, u'_{k+1})\}.$$

Then proceed again from the beginning of the algorithm.
Due to the fact that H is a cycle, this procedure leads to a construction of a unique cycle $G_n = G_{n-1} \cup \{(u'_{n-1}, u'_n)\}$ where $u'_n = u^0_0$. Finally, set $G(\mathbf{u}, H) := G_n$.

Now let H be a random cycle uniformly distributed on $\mathcal{H}_n$ and independent of the random vectors I, J, ξ, V defined above. Also let

$$\mathbf{U} = \mathbf{U}(I, J, \xi, V) \equiv \mathbf{U}((I_0, I_1, \ldots, I_r), (J_1, \ldots, J_r), \xi, V), \tag{4.24}$$

$$\mathbf{U}' = \mathbf{U}((I_0, J_1, \ldots, J_r), (I_1, \ldots, I_r), \xi, V),$$

and define

$$H_1 = G(\mathbf{U}, H), \qquad H_2 = G(\mathbf{U}', H). \tag{4.25}$$

It follows obviously that

$$\mathbf{U} \overset{d}{=} \mathbf{U}', \tag{4.26}$$

which implies

$$H_1 \overset{d}{=} H_2. \tag{4.27}$$

4.2. Properties of the transformation $G(\mathbf{u}, \cdot)$

We begin with Property (ii). It follows immediately from (4.18) and (4.22) that, for any set $\mathbf{u}$ as in (4.16) and for any $H \in \mathcal{H}_n$,

$$G(\mathbf{u}, H) \in \mathcal{H}(\mathbf{u}) := \{H \in \mathcal{H}_n : \Gamma(\mathbf{u}) \in H\}. \tag{4.28}$$

Since the set of paths $\mathbf{U}$ contains a path $\Gamma(I_0, I_1, \ldots, I_r)$ while $\mathbf{U}'$, contains a path $\Gamma(I_0, J_1, \ldots, J_r)$ property (ii) follows by the definition (4.25) and equality (4.28).

Properties (iii) and (v) also follow immediately by construction.

Turning now to property (i), let $H_0 \in \mathcal{H}_n$ be fixed arbitrarily. Consider the probability function

$$P\{G(\mathbf{U},\, H) = H_0\} = \sum_{\mathbf{u}} P\{G(\mathbf{u},\, H) = H_0\} P\{\mathbf{U} = \mathbf{u}\} \tag{4.29}$$

$$= \sum_{M,\, p=(p_0,\ldots,p_M)} \sum_{\mathbf{u}=(u^0,\ldots,u^M)} \frac{\#\{h \in \mathcal{H}_n : G(\mathbf{u},\, h) = H_0\}}{|\mathcal{H}_n|}$$

$$\times\, P\{\mathbf{U} = \mathbf{u} | M(I,\, J,\, \xi) = M,\; p(I,\, J,\, \xi) = p\} P\{M(I,\, J,\, \xi) = M,\; p(I,\, J,\, \xi) = p\},$$

with functions $M(I,\, J,\, \xi)$ and $p(I,\, J,\, \xi)$ defined in (4.14). Let M and p be fixed arbitrarily but so that $P\{M(I,\, J,\, \xi) = M,\; p(I,\, J,\, \xi) = p\} > 0$. It is obvious that, conditionally on $M(I,\, J,\, \xi) = M$ and $p(I,\, J,\, \xi) = p$, the vector of the components of the array $\mathbf{U}$ is uniformly distributed on $\mathcal{N}_q$ with $q = M + \sum_{i=0}^{M} p_i$ (see (4.15)), which implies

$$P\{\mathbf{U} = \mathbf{u} | M(I,\, J,\, D) = M,\; p(I,\, J,\, D) = p\} = \frac{1}{n(n-1)\ldots(n-q)}. \tag{4.30}$$

Let us define the following set of arrays of $M + 1$ vectors of given cardinalities:

$$\mathcal{N}_q^{(M,p)} = \{\mathbf{u} = (u^0,\, \ldots,\, u^M) : u^i = (u_0^i,\, \ldots,\, u_{p_i}^i),\, (u_0^0,\, \ldots,\, u_{p_M}^M) \in \mathcal{N}_q\}.$$

Observe that, due to symmetry, the number $|\mathcal{H}(\mathbf{u})|$ for any $\mathbf{u} \in \mathcal{N}_q^{(M,p)}$ depends only on the values of M and p but not on a particular choice of $\mathbf{u}$. Also due to symmetry we have, for any fixed $H_0 \in \mathcal{H}(\mathbf{u})$,

$$\#\{h \in \mathcal{H}_n : G(\mathbf{u},\, h) = H_0\} = \frac{|\mathcal{H}_n|}{|\mathcal{H}(\mathbf{u})|}, \tag{4.31}$$

which implies

$$\sum_{\mathbf{u} \in \mathcal{N}_q^{(M,p)}} \#\{h \in \mathcal{H}_n : G(\mathbf{u},\, h) = H_0\} = \#\{\mathbf{v} \in \mathcal{N}_q^{(M,p)} : \Gamma(\mathbf{v}) \in H_0\} \frac{|\mathcal{H}_n|}{|\mathcal{H}(\mathbf{u})|}. \tag{4.32}$$

We shall use the following relation:

$$|\mathcal{H}(\mathbf{u})| = (n - (q+1))! \#\{\mathbf{v} \in \mathcal{N}_q^{(M,p)} : \Gamma(\mathbf{v}) \in H_0\}/n. \tag{4.33}$$

This holds due to the simple observation that, for any fixed position of $M + 1$ different p_l-paths with a total number of $q + 1$ vertices in a cycle, there are $(n - (q+1))!$ ways to place the remaining $n - (q+1)$ vertices. The factor $1/n$ refers to n cyclic permutations.

Combining (4.33) and (4.32) with (4.30), we can now reduce (4.29) to

$$P\{G(\mathbf{U},\, H) = H_0\} = \frac{1}{|\mathcal{H}_n|} \sum_{M,p} P\{M(I,\, J,\, D) = M,\; p(I,\, J,\, D) = p\} = \frac{1}{|\mathcal{H}_n|}. \tag{4.34}$$

This proves

$$H \stackrel{d}{=} G(\mathbf{U},\, H) \equiv H_1, \tag{4.35}$$

which, together with (4.27), implies property (i).

Finally, we consider property (iv). Let $(i_0, j_1, \ldots, j_r) \in \mathcal{N}_r$ and $H_0 \in \mathcal{H}_n$ be fixed arbitrarily. Using definitions (4.25) and (4.24), consider

$$P\{H_1 \equiv G(\mathbf{U}, H) = H_0 | I_0 = i_0, J = j = (j_1, \ldots, j_r)\} \tag{4.36}$$
$$= P\{G(\mathbf{U}(I, j, \xi, V), H) = H_0 | I_0 = i_0, J = j\}.$$

By (4.28) we have

$$P\{G(\mathbf{u}, H) = H_0\} = 0,$$

unless $\mathbf{u}$ is such that $\Gamma(\mathbf{u}) \subset H_0$. Hence it is easy to see that

$$P\{G(\mathbf{U}(I, j, \xi, V), H) = H_0 | I_0 = i_0, J = j\} = 0,$$

unless I is an r-path in H_0 from the vertex i_0, that is, $I = \bar{I} := (i_0, E^1_{H_0}(i_0), \ldots, E^r_{H_0}(i_0))$. Write

$$H_0 = \{(i_0, E^1_{H_0}(i_0)), \ldots, (E^r_{H_0}(i_0), x_{r+1}), \ldots, (x_{n-2}, x_{n-1}), (x_{n-1}, i_0)\}.$$

Observe that the values H_0, $\bar{I}$ and j determine $\bar{\xi}$ and $\overline{V}$ uniquely such that $\Gamma(\mathbf{U}(\bar{I}, j, \xi, V)) \subset H_0$ if and only if $\xi = \bar{\xi}$ and $V = \overline{V}$. Indeed, following the definition of $\mathbf{U}$, we obtain $\bar{\xi} = (\bar{\xi}_0, \bar{\xi}_1, \ldots, \bar{\xi}_k, \bar{\xi}_{k+1})$ with $k = |\{j\} \setminus \{\bar{I}\}|$, such that $\bar{\xi}_0 = r$, $\bar{\xi}_{k+1} = n - 1$, and in the case $k > 0$ we have $\bar{\xi}_1 < \ldots < \bar{\xi}_k$ with $\{\bar{\xi}\} = \{l : x_l \in \{j\} \setminus \{\bar{I}\}\}$. Also, we see that $\overline{V}$ is simply the r-boundary of the set $\{j, \bar{I}\}$ in the cycle H_0. Hence, setting $\overline{\mathbf{u}} = \mathbf{U}(\bar{i}, j, \bar{\xi}, \overline{V})$, by (4.36) we have

$$P\{G(\mathbf{U}(I, j, \xi, V), H) = H_0 | I_0 = i_0, J = j\} \tag{4.37}$$
$$= P\{G(\overline{\mathbf{u}}, H) = H_0\} P\{(I_1, \ldots, I_r) = (E^1_{H_0}(i_0), \ldots, E^r_{H_0}(i_0)) | I_0 = i_0\}$$
$$\times P\{\xi = \bar{\xi}, V = \overline{V} | I = \bar{I}, J = j\}.$$

Recall that the distribution of I is uniform on $\mathcal{N}_r$. Hence,

$$P\{(I_1, \ldots, I_r) = (E^1_{H_0}(i_0), \ldots, E^r_{H_0}(i_0)) | I_0 = i_0\} = \frac{1}{(n-1) \ldots (n-r)}. \tag{4.38}$$

Taking into account (4.31), as well as the fact that given $I = \bar{I}$ and $J = j$ with $k(\bar{I}, j) = k > 0$ the distribution of ξ is uniform on Ω_k, and conditionally on $\xi = \bar{\xi}$ the vector V is uniformly distributed on $\mathcal{V}(m)$ (see (4.4) and (4.8)), we derive from (4.37) and (4.38) that

$$P\{G(\mathbf{U}(I, j, \xi, V), H) = H_0 | I_0 = i_0\} = \frac{1}{|\mathcal{H}(\overline{\mathbf{u}})| \, |\Omega_k| \, |\mathcal{V}(m)|} \, \frac{1}{(n-1) \ldots (n-r)}. \tag{4.39}$$

Also, setting $|\Omega_0| \equiv 1$, it is easy to see that the same formula holds in the case $k = 0$ as well. Observe now that for $\mathcal{H}(\bar{I}) := \{H \in \mathcal{H}_n : \Gamma(\bar{I}) \in H\}$ we have

$$|\mathcal{H}(\bar{I})| = |\mathcal{H}(\overline{\mathbf{u}})| \, |\Omega_k| \, |\mathcal{V}(m)|, \tag{4.40}$$

which follows from the following facts:

1.
$\Gamma(\bar{I}) \subset \Gamma(\bar{\mathbf{u}})$;

2. $|\Omega_k|$ equals the number of possible ways to label the nodes of the set $\{j\}\backslash\{\bar{I}\}$, choosing labels from $\mathcal{N}\backslash\{\bar{I}\}$;
3. $|\mathcal{V}(m)|$ is the number of all possible ways to label the nodes of the r-boundary of the set $\{j\} \cup \{\bar{I}\}$ choosing labels from $\mathcal{N}\backslash(\{j\} \cup \{\bar{I}\})$.

Substituting (4.40) into (4.39), we obtain

$$P\{G(\mathbf{U}(I, j, \xi, V), H) = H_0 | I_0 = i_0\} = \frac{1}{|\mathcal{H}(\bar{I})|} \frac{1}{(n-1)\dots(n-r)}. \tag{4.41}$$

It is easy to compute, for any $i \in \mathcal{N}_r$, that $|\mathcal{H}(i)| = (n - r - 1)!$, which, together with (4.41), implies

$$P\{G(\mathbf{U}(I, j, \xi, V), H) = H_0 | I_0 = i_0\} = \frac{1}{(n-1)!} = \frac{1}{|\mathcal{H}_n|} = P\{H_1 = H_0\}, \tag{4.42}$$

where the last equality is due to (4.34). This, when substituted into (4.36), yields property (iv). This completes the proof of Lemma 2.1.

Acknowledgement

I thank David M. Mason who introduced me to this problem and the related results years ago, and who generously communicated the proof of Proposition 3.1. I also thank the editor, W.R. van Zwet, for his help with a reference, and the referees for useful comments.

The research for this paper was supported by the Swedish Natural Science Research Council.

References

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

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

Does, R.J.M.M. (1982) Berry–Esseen theorems for simple linear rank statistics under the null-hypothesis. *Ann. Statist.*, **4**, 108–156.

Haeusler, E., Mason, D.M. and Turova, T.S. (2000) A study of serial ranks via random graphs. *Bernoulli*, **6**, 541–570.

Hallin, M. and Puri, M.L. (1988) Optimal rank-based procedures for time series analysis: testing an ARMA model against other ARMA models. *Ann. Statist.*, **16**, 402–432.

Hallin, M. and Puri, M.L. (1994) Aligned rank tests for linear models with autocorrelated error terms. *J. Multivariate Anal.*, **50**, 175–237.

Hallin, M. and Rifi, K. (1997) A Berry–Esseen theorem for serial rank statistics. *Ann. Inst. Statist. Math.*, **49**, 777–799.

Hallin, M., Ingenbleek, J.-F. and Puri, M.L. (1985) Linear serial rank tests for randomness against ARMA alternatives. *Ann. Statist.*, **13**, 1156–1181.

Loh, W.L. (1996) On Latin hypercube sampling. *Ann. Statist.*, **24**, 2058–2080.

Mason, D.M. and Turova, T.S. (2000) Motoo's combinatorial central limit theorem for serial rank statistics. In J. Jurečková (ed.), *Proceedings of Workshop on Modern Statistical Inference, Prague, Czech Republic, August 1998*, Special issue of *J. Statist. Plann. Inference*, **91**, 427–440.

Schneller, W. (1988) A short proof of Motoo's combinatorial central limit theorem using Stein's method. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **78**, 2490–252.

Stein, C. (1972) A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In L. LeCam, J. Neyman and E.L. Scott (eds), *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 2, pp. 583–602. Berkeley: University of California Press.

van Zwet, W.R. (1982) On the Edgeworth expansion for the simple linear rank statistic. In B.V. Gnedenko, M.L. Puri and J. Vincze (eds), *Nonparametric Statistical Inference*, Vol. II, Colloq. Math. Soc. János Bolyai 32, pp. 889–909. Amsterdam: North-Holland.

Wald, A. and Wolfowitz, J. (1943) An exact test for randomness in the nonparametric case based on serial correlation. *Ann. Math. Statist.*, **14**, 378–388.

Zhao, L.-C., Bai, Z.-D., 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.

Received October 2001 and revised June 2003