Orthogonal decomposition of symmetric functions defined on random permutations

M. Bloznelis

Abstract

Let T denote a real function defined on random subsets of a given family of finite sets. The random variable T is decomposed into the sum of the linear, the quadratic, the cubic etc. parts which are mutually uncorrelated. Applications of this decomposition to the asymptotics of the probability distribution of T (as the sizes of random subsets and of finite sets increase) are discussed.

1 Introduction

Let $\mathcal{X}_k = \{x_{k,1}, \dots, x_{k,N_k}\}$, $k = 1, \dots, h$, be non-intersecting finite sets. Given an integer $0 < n_k < N_k$, let $\mathbb{X}_k$ denote a random subset of $\mathcal{X}_k$ of size n_k which is uniformly distributed over the class of n_k-subsets of $\mathcal{X}_k$. That is, for arbitrary subset $\mathcal{A}_k \subset \mathcal{X}_k$ of size $|\mathcal{A}_k| = n_k$ we have $P\{\mathbb{X}_k = \mathcal{A}_k\} = \binom{N_k}{n_k}^{-1}$. We assume that random subsets $\mathbb{X}_1, \dots, \mathbb{X}_h$ are independent. Given a real function t defined on h-tuples $(\mathcal{A}_1, \dots, \mathcal{A}_h)$ of subsets, introduce the random variable

$$T = t(\mathbb{X}_1, \dots, \mathbb{X}_h). \tag{1}$$

The main object of the present study is the orthogonal decomposition of T: we represent T by the sum of the linear, the quadratic, the cubic etc. parts which are mutually uncorrelated,

$$T = \mathbf{E}T + L + Q + \dots. \tag{2}$$

Here

$$L = \sum_k L_k, \qquad L_k = \sum_{x \in \mathbb{X}_k} g_k(x)$$

denotes the linear part of T and $Q = \sum_{k \le r} Q_{k,r}$ denotes the quadratic part of T,

$$Q_{k,r} = \sum_{x \in \mathbb{X}_k} \sum_{y \in \mathbb{X}_r} g_{k,r}(x,y) \quad \text{for} \quad k < r, \quad \text{and} \quad Q_{k,k} = \sum_{\{x,y\} \subset \mathbb{X}_k} g_{k,k}(x,y).$$

The summands L_k, $Q_{k,r}$ are uncorrelated. The real functions g_k and $g_{k,r}$ as well as those defining higher order nonlinear parts of the decomposition (2) are specified in (4) below.

The orthogonal decomposition provides a useful tool for the analysis of distributional properties of T and its asymptotics as $n = n_1 + \ldots + n_h \to \infty$. Orthogonal decomposition of statistics which are functions of *independent* random variables were studied and applied in a number of papers (Hoeffding [11], Rubin and Vitale [17], Efron and Stein [8], van Zwet [19], Bentkus, Götze and van Zwet [2], etc.) In a combinatorial context this type of decomposition was used by Janson and Nowicki [14], Janson [13], de Jong [7].

In the present paper we construct orthogonal decomposition in the case where the underlying random variables (=elements of random subsets) are dependent. Zhao and Chen [20] and Bloznelis and Götze [4] used orthogonal decomposition in their studies of the normal approximation and its refinements for various statistics $T = t(\mathbb{X}_1)$. However neither of these two papers provide a proof of the orthogonality property, see identity (9) below, which plays a crucial role for the decomposition. We give a combinatorial proof of this identity in a more general situation of several random subsets $\mathbb{X}_1, \ldots, \mathbb{X}_h$. In the case where $h = 1$ the random variable T defined by (1) reduces to that considered by Zhao and Chen [20] and Bloznelis and Götze [4]. In the case where $n_1 = \ldots = n_h = 1$ the underlying random variables are independent, and we are in the situation considered by Hoeffding [11].

The model (1) has numerous applications. In statistics, see, e.g., Cochran [6], Särndal, Swensson and Wretman [18], it is called the "stratified sampling without replacement" model. It is assumed there that a population $\mathcal{X}$ is broken up into non-overlapping subpopulations (strata) $\mathcal{X}_1, \ldots, \mathcal{X}_h$ and a statistic T, based on stratified sample drawn without replacement $(\mathbb{X}_1, \ldots, \mathbb{X}_h)$, is used to estimate some parameter of the population $\mathcal{X}$.

Another example is a subgraph count statistic, see, e.g., Barbour, Karoński and Ruciński [1], Janson [12]. Given an integer k let E denote the set of edges of the complete graph K_k based on k vertices and let $\mathbb{E}$ be a random subset of E of size $|\mathbb{E}| = n$ which is uniformly distributed among all n-subsets of E. We paint $\mathbb{E}$ edges blue thus obtaining the random graph $\mathcal{G}(k, n)$, see Bollobás [5]. The number of blue triangles $T = T(\mathbb{E})$ is a random variable of the form (1), where $h = 1$. Allowing several (independent) colors we obtain random variable (1) with $h > 1$.

Let us outline the content of the paper. In Section 2 we consider two examples. In Section 3 we introduce the orthogonal decomposition and discuss its main properties. Here we show the variance decomposition formula and construct bounds for the remainders of the approximation of T by the first few (say two or three) terms of the decomposition. Proofs are postponed into Section 4.

Acknowledgement. I thank anonymous referee for valuable comments and remarks.

2 Examples

For two subgraph count statistics we construct orthogonal decompositions and evaluate variances. Consider the complete graph based on $k \geq 3$ vertices. Let $\mathcal{X} = \{x_1, \dots, x_N\}$ denote the set of edges. Here $N = \binom{k}{2}$.

Example 1. Let $\mathbb{X} = \{X_1, \dots, X_n\} \subset \mathcal{X}$ be a random n-subset uniformly distributed over the class of n-subsets of $\mathcal{X}$. Here $n < N$. We paint edges $X_1, \dots, X_n$ blue. The number of blue triangles

$$\mathcal{N} = \sum_{1 \leq i < j < k \leq n} \Delta_{X_i X_j X_k}$$

is a U- statistic of degree three based on the random variables $X_1, \dots, X_n$ taking values in $\mathcal{X}$. Here $\Delta_{xyz} = 1$ if the edges x, y, z make up a triangle and $\Delta_{xyz} = 0$ otherwise.

By symmetry, the mean value

$$\mathbf{E}\mathcal{N} = \binom{n}{3} \mathbf{E}\Delta_{X_1 X_2 X_3}, \qquad \text{where} \qquad \mathbf{E}\Delta_{X_1 X_2 X_3} = \frac{k-2}{\binom{N-1}{2}}.$$

The orthogonal decomposition formula (3) gives $\mathcal{N} = \mathbf{E}\mathcal{N} + Q + K$, where

$$Q = \sum_{1 \leq i < j \leq n} g_2(X_i, X_j) \qquad \text{respectively} \qquad K = \sum_{1 \leq i < j < k \leq n} g_3(X_i, X_j, X_k)$$

denotes the quadratic, respectively, the cubic part. Note that the linear part L of the decomposition vanishes, cf (2). Here

$$\begin{aligned} g_2(x, y) &= \frac{n-2}{N-4}(l_{xy} - \mathbf{E}l_{X_1 X_2}), \qquad \mathbf{E}l_{X_1 X_2} = 2\frac{k-2}{N-1}, \\ g_3(x, y, z) &= \Delta_{xyz} - \mathbf{E}\Delta_{X_1 X_2 X_3} - \frac{1}{N-4}(l_{xy} + l_{yz} + l_{xz} - 3\mathbf{E}l_{X_1 X_2}). \end{aligned}$$

Here for $x, y \in \mathcal{X}$ we write $l_{xy} = 1$ if x is adjacent to y and $l_{xy} = 0$ otherwise. The random variables $g_2(X_{i_1}, X_{i_2})$ and $g_3(X_{j_1}, X_{j_2}, X_{j_3})$ are uncorrelated for arbitrary indices $i_1 < i_2$ and $j_1 < j_2 < j_3$.

Using (10) and (11), see below, we evaluate the variance

$$\begin{aligned} &\mathbf{Var}\mathcal{N} = \mathbf{Var}Q + \mathbf{Var}K, \\ &\mathbf{Var}Q = \frac{\binom{n}{2}\binom{N-n}{2}}{\binom{N-2}{2}}\sigma_2^2, \qquad \mathbf{Var}K = \frac{\binom{n}{3}\binom{N-n}{3}}{\binom{N-3}{3}}\sigma_3^2, \end{aligned}$$

where $\sigma_2^2 = \mathbf{E}g_2^2(X_1, X_2)$ and $\sigma_3^2 = \mathbf{E}g_3^2(X_1, X_2, X_3)$. A simple calculation shows

$$\begin{aligned} \sigma_2^2 &= (\frac{n-2}{N-4})^2 p_1(1-p_1), \\ \sigma_3^2 &= p_0(1-p_0) - \frac{6}{N-4}p_0(1-p_1) + \frac{1}{(N-4)^2}(3 - \frac{6}{N-2})p_1(1-p_1). \end{aligned}$$

Here we denote, for brevity, $p_1 = \mathbf{E} l_{X_1 X_2}$ and $p_0 = \mathbf{E}\Delta_{X_1 X_2 X_3}$.

Using the asymptotic relation $N \sim k^2/2$, as $k \to \infty$, we obtain

$$\mathbf{E}\Delta_{X_1 X_2 X_3} \sim 8/k^3, \qquad \sigma_2^2 \sim (\frac{n}{N})^2 \frac{4}{k}, \qquad \sigma_3^2 \sim \frac{8}{k^3}.$$

Here and below we write $a_k \sim b_k$ if $a_k/b_k \to 1$ as $k \to \infty$. Denoting $p = n/N$ and $q = 1 - p$, we have as $k \to \infty$

$$\begin{aligned} \mathbf{Var} Q &\sim 2^{-1} p^4 q^2 k^3, \qquad \mathbf{Var} K \sim 6^{-1} p^3 q^3 k^3, \\ \mathbf{E}\mathcal{N} &\sim 6^{-1} p^3 k^3, \qquad \mathbf{Var}\mathcal{N} \sim 6^{-1} p^3 q^3 k^3 (3\frac{p}{q} + 1). \end{aligned}$$

Example 2. Given integers $n_1, n_2, n_3 < N$, let $\mathbb{X}_i = \{X_{i,1}, \dots, X_{i,n_i}\}$ be random subsets of $\mathcal{X}$, $i = 1, 2, 3$. We assume that, for every i, $\mathbb{X}_i$ is uniformly distributed over the class of n_i-subsets of $\mathcal{X}$ and the random subsets $\mathbb{X}_1$, $\mathbb{X}_2$, $\mathbb{X}_3$ are independent. We paint edges $\mathbb{X}_1$ yellow, $\mathbb{X}_2$ green and $\mathbb{X}_3$ red. The number of triangles having all edges of different colors

$$\mathcal{N} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \Delta_{X_{1,i} X_{2,j} X_{3,k}}$$

is a U-statistic of degree three. Here $\Delta_{x\,y\,z} = 1$ if the edges x, y, z make up a triangle and $\Delta_{x\,y\,z} = 0$ otherwise. Note that in this model two vertices can be joined by at most three edges of different colors. Therefore, given three vertices there can be at most six differently colored triangles based on these vertices.

By symmetry, the mean value

$$\mathbf{E}\mathcal{N} = n_1 n_2 n_3 \mathbf{E}\Delta_{X_{1,1} X_{2,1} X_{3,1}}, \quad \text{where} \quad \mathbf{E}\Delta_{X_{1,1} X_{2,1} X_{3,1}} = 2(k-2)/N^2 =: \delta.$$

Before to write the orthogonal decomposition of $\mathcal{N}$ we introduce some more notation. Introduce the function $l_{xy} : \mathcal{X} \times \mathcal{X} \to \{0, 1\}$, where $l_{xy} = 1$ if $x \neq y$ and x and y are adjacent, otherwise put $l_{xy} = 0$. Denote $n_{12} := n_3$, $n_{13} := n_2$ and $n_{23} := n_1$. Write

$$v_i = n_i(N - n_i)/(N - 1), \quad p_i = n_i/N, \quad q_i = 1 - p_i, \quad p_{ij} = n_{ij}/N.$$

The orthogonal decomposition formula (3) shows $\mathcal{N} = \mathbf{E}\mathcal{N} + Q + K$. Here the quadratic part $Q = Q_{12} + Q_{13} + Q_{23}$, where

$$Q_{ij} = \sum_{r=1}^{n_i} \sum_{k=1}^{n_j} g_{ij}(X_{i,r}, X_{j,k}), \qquad g_{ij}(x, y) = n_{ij} N^{-1} (l_{xy} - \mathbf{E} l_{X_{1,1} X_{2,1}}),$$

where $\mathbf{E}l_{X_{1,1}X_{2,1}} = 2(k-2)/N$. The cubic part

$$K = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} h(X_{1,i}, X_{2,j}, X_{3,k}),$$

where $h(x,y,z) = \Delta_{xyz} - N^{-1}(l_{xy} + l_{xz} + l_{yz}) + 2\delta$.

The random variables $g_{12}(X_{1,i_1}, X_{2,j_1})$, $g_{13}(X_{1,i_2}, X_{3,k_1})$, $g_{23}(X_{2,j_2}, X_{3,k_2})$ and $h(X_{1,i_3}, X_{2,j_3}, X_{3,k_3})$ are uncorrelated for every $1 \le i_1, i_2, i_3 \le n_1$, $1 \le j_1, j_2, j_3 \le n_2$ and $1 \le k_1, k_2, k_3 \le n_3$. Furthermore, a calculation shows

$$\begin{aligned}\sigma_{ij}^2 &:= \mathbf{E}g_{ij}^2(X_{i,1}, X_{j,1}) = p_{ij}^2 N\delta(1-N\delta),\\ \sigma_{123}^2 &:= \mathbf{E}h^2(X_{1,1}, X_{2,1}, X_{3,1}) = (1-3/N)\delta + 2\delta^2.\end{aligned}$$

To show the variance we write (using (10), see below)

$\mathbf{Var}\mathcal{N} = \sum_{i<j} \mathbf{Var}Q_{ij} + \mathbf{Var}K$. Denoting $\overline{p} = p_1p_2p_3$ and $\overline{q} = q_1q_2q_3$ we obtain from (11), see below, that

$$\begin{aligned}\mathbf{Var}Q &= \sum_{1\le i<j\le 3} \mathbf{Var}Q_{ij} = \sum_{1\le i<j\le 3} v_iv_j\sigma_{ij}^2\\ &= \frac{N^4}{(N-1)^2}\overline{p}\,\overline{q}(\frac{p_1}{q_1} + \frac{p_2}{q_2} + \frac{p_3}{q_3})N\delta(1-N\delta),\\ \mathbf{Var}K &= v_1v_2v_3\sigma_{123}^2 = \frac{N^6}{(N-1)^3}\overline{p}\,\overline{q}((1-3/N)\delta + 2\delta^2).\end{aligned}$$

Using the asymptotic relation $N \sim k^2/2$ as $k \to \infty$, we obtain

$$\begin{aligned}\mathbf{Var}Q &\sim k^3\overline{p}\,\overline{q}(\frac{p_1}{q_1} + \frac{p_2}{q_2} + \frac{p_3}{q_3}), \qquad \mathbf{Var}K \sim k^3\overline{p}\,\overline{q},\\ \mathbf{E}\mathcal{N} &\sim \overline{p}\,k^3, \qquad \mathbf{Var}\mathcal{N} \sim k^3\overline{p}\,\overline{q}(\frac{p_1}{q_1} + \frac{p_2}{q_2} + \frac{p_3}{q_3} + 1).\end{aligned}$$

3 Orthogonal decomposition

We can assume without loss of generality that $\mathbf{E}T = 0$.

3.1 Notation

Given $k = 1, \dots, h$, let $\mathcal{X}_k^* = (X_{k,1}, \dots, X_{k,N_k})$ be a random permutation of the ordered set $(x_{k,1}, \dots, x_{k,N_k})$. We assume that random permutations $\mathcal{X}_1^*, \dots, \mathcal{X}_h^*$ are independent. Note that the group $\{X_{k,1}, \dots, X_{k,n_k}\}$ of the first n_k values of the permutation $\mathcal{X}_k^*$ represents the random subset $\mathbb{X}_k$. In what follows we use the representation $\mathbb{X}_k = \{X_{k,1}, \ldots, X_{k,n_k}\}$, for $1 \le k \le h$, and write

$$T = t(\{X_{1,1}, \ldots, X_{1,n_1}\}, \ldots, \{X_{h,1}, \ldots, X_{h,n_h}\}).$$

Here we assume that t is invariant under permutations within every group $\{X_{k,1}, \ldots, X_{k,n_k}\}$ of its arguments (the invariance property agrees with the formula (1) where t is considered as a function defined on subsets.

For $r = 1, 2, \ldots,$ denote $\Omega_r = \{1, \ldots, r\}$. Write $n_k^* = \min\{n_k, N_k - n_k\}$ and $n^* = n_1^* + \ldots + n_h^*$. By $\overline{a} = (a_1, \ldots, a_h)$ and $\overline{b} = (b_1, \ldots, b_h)$ we denote h-dimensional vectors with non-negative integer coordinates and write $\overline{b} \le \overline{a}$ if $b_k \le a_k$ for every $k = 1, \ldots, h$. Furthermore, write $\overline{b} < \overline{a}$ if $\overline{b} \le \overline{a}$ and $\overline{b} \ne \overline{a}$. Clearly, $\overline{n}^* \le \overline{n}$, where $\overline{n} = (n_1, \ldots, n_h)$ and $\overline{n}^* = (n_1^*, \ldots, n_h^*)$. By $e_k = (0, \ldots, 0, 1, 0, \ldots, 0)$ we denote the k-th coordinate vector and write $\overline{0} = (0, \ldots, 0)$. Furthermore, write $|\overline{a}| = a_1 + \ldots + a_h$.

In what follows $\overline{a}$ will be used to mark the sizes of sets of a h-tuple $(\mathcal{A}_1, \ldots, \mathcal{A}_h)$, where $\mathcal{A}_k \subset \mathcal{X}_k$, for $1 \le k \le h$, so that $|\mathcal{A}_k| = a_k$, for every k. Similarly, $(\mathbb{A}_1, \ldots, \mathbb{A}_h)$ will denote a h-tuple of sets $\mathbb{A}_k = \{X_{k,i_1}, \ldots, X_{k,i_{a_k}}\}$ of random variables, $1 \le k \le h$. Note that $\mathbb{A}_k$ represents a random subset of $\mathcal{X}_k$ of size a_k, which is uniformly distributed over the class of a_k-subsets of $\mathcal{X}_k$.

Given a real random variable G, we denote by $\mathbf{E}(G|\mathbb{A}_1, \ldots, \mathbb{A}_h)$ the conditional expectation of G given the random variables $\{X_{k,i} : X_{k,i} \in \mathbb{A}_k, k = 1, \ldots, h\}$. Furthermore, given $\overline{a}$ and a h-tuple $(\mathcal{A}_1, \ldots, \mathcal{A}_h)$ of subsets $\mathcal{A}_k = \{x_{k,j_1}, \ldots, x_{k,j_{a_k}}\} \subset \mathcal{X}_k$, $1 \le k \le h$, denote

$$\varphi_{\overline{a}}(\mathcal{A}_1, \ldots, \mathcal{A}_h) := \mathbf{E}(T \,|\, X_{k,1} = x_{k,j_1}, \ \ldots, \ X_{k,a_k} = x_{k,j_{a_k}}, \ 1 \le k \le h).$$

Note that,

$$\varphi_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}_h) = \mathbf{E}(T|\mathbb{A}_1, \ldots, \mathbb{A}_h).$$

Finally, by $\mathbf{E}_k$ we shall denote the conditional expectation given all the random variables but $\mathcal{X}_k^*$.

3.2 Decomposition

The orthogonal decomposition

$$T = \sum_{\overline{a} \le \overline{n}} U(\overline{a}) \tag{3}$$

expands T into the sum of mutually uncorrelated U statistics

$$U(\overline{a}) = \sum_{|\mathbb{A}_1| = a_1, \, \mathbb{A}_1 \subset \mathbb{X}_1} \cdots \sum_{|\mathbb{A}_h| = a_h, \, \mathbb{A}_h \subset \mathbb{X}_h} g_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}_h).$$

Here $\sum_{\overline{a} \le \overline{n}}$ denotes the sum over all vectors $\overline{a} = (a_1, \ldots, a_h)$ such that $0 \le a_k \le n_k$, for $k = 1, \ldots, h$.

Given $\overline{a}$, the real function $g_{\overline{a}}$ is defined on h-tuples $(\mathcal{A}_1, \ldots, \mathcal{A}_h)$ of subsets $\mathcal{A}_k \subset \mathcal{X}_k$ of sizes $|\mathcal{A}_k| = a_k$, $k = 1, \ldots, h$. We define functions $g_{\overline{a}}$, $\overline{a} \leq \overline{n}$, using induction on increasing values of $a_k = 0, 1, \ldots, n_k$, $1 \leq k \leq h$. To this aim we introduce auxiliary functions $\psi_{\overline{a}}$, $\overline{a} \leq \overline{n}$, which differ from $g_{\overline{a}}$ by multiplicative constants,

$$g_{\overline{n}} = \psi_{\overline{n}} \qquad \text{and} \qquad g_{\overline{b}} = C(\overline{n}, \overline{b})\psi_{\overline{b}}, \qquad \text{for} \qquad \overline{b} < \overline{n}. \tag{4}$$

The constants $C(\overline{n}, \overline{b})$ are specified in (7) below.

Define $\psi_{\overline{0}} \equiv 0$ and, for $1 \leq k \leq h$, put

$$\psi_{\overline{e}_k}(\{x\}) = \varphi_{\overline{e}_k}(\{x\}) = \mathbf{E}(T \,|X_{k,1} = x), \quad x \in \mathcal{X}_k.$$

Given $\overline{a} \leq \overline{n}$ we assume that the functions $\psi_{\overline{b}}$, $\overline{b} < \overline{a}$ are already defined and put

$$\begin{aligned}\psi_{\overline{a}}(\mathcal{A}_1, \ldots, \mathcal{A}_h) &= \varphi_{\overline{a}}(\mathcal{A}_1, \ldots, \mathcal{A}_h) \\ &\quad - \sum_{\overline{b} < \overline{a}} C(\overline{a}, \overline{b}) \sum_{|\mathcal{B}_1| = b_1,\, \mathcal{B}_1 \subset \mathcal{A}_1} \cdots \sum_{|\mathcal{B}_h| = b_h,\, \mathcal{B}_h \subset \mathcal{A}_h} \psi_{\overline{b}}(\mathcal{B}_1, \ldots, \mathcal{B}_h).\end{aligned} \tag{5}$$

We choose the numbers $C(\overline{a}, \overline{b})$ so that almost surely

$$\mathbf{E}(\psi_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}_h) \,|\mathbb{B}_1, \ldots, \mathbb{B}_h) = 0, \tag{6}$$

for every h-tuple $(\mathbb{B}_1, \ldots, \mathbb{B}_h)$ satisfying $|\mathbb{B}_k| < a_k$, for some k. Here $\mathbb{B}_j = \{X_{j,k_1}, \ldots, X_{j,k_{b_j}}\}$ denotes a collection of random variables of the permutation $\mathcal{X}_j^*$, $1 \leq j \leq h$.

The fact that its is possible to choose such numbers $C(\overline{a}, \overline{b})$ is not obvious. We show in Lemmas 4.2 and 4.3 below that (6) holds with the (unique choice of) constants

$$C(\overline{a}, \overline{b}) = \prod_{k=1}^{h} V_k(a_k, b_k). \tag{7}$$

Here, for $a_k + b_k \leq N_k$, we put

$$V_k(a_k, b_k) = \frac{\binom{N_k - b_k}{b_k}}{\binom{N_k - a_k}{b_k}}.$$

For $a_k + b_k > N_k$ we put $V_k(a_k, b_k) = 0$ with one exception in the case where N_k is odd (write $N_k = 2r_k + 1$) and $a_k = b_k = r_k + 1$. In this case we put $V_k(a_k, a_k) = 1$.

The identity (3.3) applied to $\overline{a} = \overline{n}$ gives (3). Indeed, for arbitrary h-tuple $(\mathcal{A}_1, \ldots, \mathcal{A}_h)$ of subsets $\mathcal{A}_k \subset \mathcal{X}_k$ with $|\mathcal{A}_k| = n_k$ we obtain from (3.3)

$$\begin{aligned} t(\mathcal{A}_1, \ldots, \mathcal{A}_h) &= \varphi_{\overline{n}}(\mathcal{A}_1, \ldots, \mathcal{A}_h) \\ &= \sum_{\overline{b} \leq \overline{n}} \sum_{|\mathcal{B}_1| = b_1,\, \mathcal{B}_1 \subset \mathcal{A}_1} \cdots \sum_{|\mathcal{B}_h| = b_h,\, \mathcal{B}_h \subset \mathcal{A}_h} g_{\overline{b}}(\mathcal{B}_1, \ldots, \mathcal{B}_h),\end{aligned}$$

3.3 Properties of kernels $g_{\overline{a}}$

Remark. *For those $\overline{b} \le \overline{n}$, which fail to satisfy $\overline{b} \le \overline{n}^*$, we have $g_{\overline{b}} \equiv 0$.*

To prove the remark fix such $\overline{b}$. We have $n_k^* < b_k \le n_k$, for some k. Therefore, $N_k - n_k < b_k$. In the case where $b_k = n_k$ we have $2b_k > N_k$ and, by Lemma 4.3, we obtain $\psi_{\overline{b}} \equiv 0$. In view of (4) this implies $g_{\overline{b}} \equiv 0$. In the case where $b_k < n_k$ we have $\overline{b} < \overline{n}$. The inequality $N_k - n_k < b_k < n_k$ imply $C(\overline{n}, \overline{b}) = 0$. In view of (4) we obtain $g_{\overline{b}} \equiv 0$, thus completing the proof of the remark.

It follows from the remark that $U(\overline{a}) \equiv 0$ for those $\overline{a}$ which fail to satisfy $\overline{a} \le \overline{n}^*$. Therefore, the sum (3) reduces to the sum

$$T = \sum_{\overline{a} \le \overline{n}^*} U(\overline{a}).$$

Furthermore, one can represent T by the sum of uncorrelated U-statistics U_s of increasing order $s = 1, 2, \ldots, n^*$,

$$T = U_1 + U_2 + \ldots + U_{n^*}, \qquad U_s = \sum_{|\overline{a}|=s,\, \overline{a} \le \overline{n}^*} U(\overline{a}). \tag{8}$$

Here U_1 is called the linear part, U_2 is called the quadratic part etc. (in (2) we denote $L = U_1$, $Q = U_2$).

The fact that U_s and U_t are uncorrelated for $s \ne t$ and $U(\overline{a})$ and $U(\overline{b})$ are uncorrelated for $\overline{a} \ne \overline{b}$ follows from the identity

$$\mathbf{E} g_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}_h) g_{\overline{b}}(\mathbb{B}_1, \ldots, \mathbb{B}_h) = 0 \qquad \text{for} \qquad \overline{a} \ne \overline{b}.$$

Here $\mathbb{A}_k$, $\mathbb{B}_k$ are arbitrary collections of random variables of the random permutation $\mathcal{X}_k^*$ such that $|\mathbb{A}_k| = a_k$ and $|\mathbb{B}_k| = b_k$, $1 \le k \le h$. This identity is a consequence of the orthogonality property:

$$\mathbf{E}\big(g_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}_h) \,|\, \mathbb{B}_1, \ldots, \mathbb{B}_h\big) = 0 \qquad \text{a.s.}, \tag{9}$$

whenever $|\mathbb{A}_k| > |\mathbb{B}_k|$ for some $1 \le k \le h$. Note that (9) follows from (6). Choosing $\mathbb{B}_1 = \emptyset$, ..., $\mathbb{B}_h = \emptyset$ we obtain from (9)

$$\mathbf{E} g_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}_h) = 0, \qquad \text{for} \qquad \overline{a} \ne \overline{0}.$$

3.4 Dual representation

An interesting consequence of (9) is the duality property of the U-statistic $U(\overline{a})$. Introduce the sets $\mathbb{X}'_k = \mathcal{X}_k \setminus \mathbb{X}_k$. For $\overline{a} \le \overline{n}^*$ the identity (9) implies that almost surely $U(\overline{a}) = U'(\overline{a})$, where

$$U'(\overline{a}) = \sum_{|\mathbb{A}_1|=a_1,\, \mathbb{A}_1 \subset \mathbb{X}'_1} \cdots \sum_{|\mathbb{A}_h|=a_h,\, \mathbb{A}_h \subset \mathbb{X}'_h} g'_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}_h), \qquad g'_{\overline{a}} \equiv (-1)^{|\overline{a}|} g_{\overline{a}}.$$

Recall that $|\overline{a}| = a_1 + \ldots + a_h$. Furthermore, denoting

$$U'_s = \sum_{\overline{a} \leq \overline{n}^* : \, |\overline{a}| = s} U'(\overline{a}),$$

we obtain almost surely $U_s = U'_s$. From (8), we have

$$T = U_1 + \ldots + U_{n^*} = U'_1 + \ldots + U'_{n^*}.$$

Therefore, T can be considered as a statistic of the "dual sample" $\mathbb{X}'_1, \ldots, \mathbb{X}'_h$.

3.5 Variance decomposition

As a consequence of (3), (8) and the fact that the contributing U-statistics are uncorrelated we obtain the identity

$$\mathbf{Var} T = \sum_{s=1}^{n^*} \mathbf{Var} U_s = \sum_{\overline{a} \leq \overline{n}^*} \mathbf{Var} U(\overline{a}). \qquad (10)$$

A calculation shows

$$\mathbf{Var} U(\overline{a}) = \sigma^2_{\overline{a}} C(\overline{a}), \quad C(\overline{a}) = \prod_{k=1}^{h} \frac{\binom{n_k}{a_k}\binom{N_k - n_k}{a_k}}{\binom{N_k - a_k}{a_k}}. \qquad (11)$$

Here we denote $\sigma^2_{\overline{0}} = 0$ and write for $\overline{a} > \overline{0}$

$$\sigma^2_{\overline{a}} = \mathbf{E} g^2_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}_h).$$

Note that, by symmetry, the expectation in the right hand side is the same for arbitrary $(\mathbb{A}_1, \ldots, \mathbb{A}_h)$, satisfying $|\mathbb{A}_k| = a_k$, $1 \leq k \leq h$. The proof of (11) is given in the Appendix below. Combining (10) and (11) we obtain the variance decomposition

$$\mathbf{Var} T = \sum_{a_1=1}^{n_1^*} \ldots \sum_{a_h=1}^{n_h^*} C(\overline{a}) \, \sigma^2_{\overline{a}} = \sum_{\overline{a} \leq \overline{n}^*} C(\overline{a}) \sigma^2_{\overline{a}}.$$

3.6 Bounds for remainders

The partial sums

$$U_{s,*} = U_1 + \ldots + U_s, \qquad s < n^*, \qquad n^* = n_1^* + \ldots + n_h^*$$

often provide satisfactory approximations to T. In order to control the remainder $R_s = T - U_{s,*}$ we construct an upper bound for

$$\mathbf{E} R_s^2 = \sum_{\overline{a} \leq \overline{n}^*, \, |\overline{a}| > s} C(\overline{a}) \sigma^2_{\overline{a}}.$$

A motivation for such approximations comes from mathematical statistics. A number of important statistics T are asymptotically linear, that is, for large n the linear part dominates the statistic T. This implies the asymptotic normality of T, see Hoeffding [11], Hajek [10], Lehmann [16], Koroljuk and Borovskikh [15] for results in the case of *independent observations.* Furthermore, the approximation by the linear and the quadratic part is used to obtain a higher order asymptotic results (one-term Edgeworth expansion), see Bentkus, Götze and van Zwet [2].

In the model (1) of the present paper the sample sizes $n_1, \dots, n_h$ are bounded ($n_k < N_k$, $1 \le k \le h$). In order to speak of the asymptotic distribution of T we introduce a sequence of collections of sets $\{\mathcal{X}_1^{(r)}, \dots, \mathcal{X}_{h_r}^{(r)}\}$, $r = 1, 2, \dots$, and a sequence of collections of random subsets $\{\mathbb{X}_1^{(r)}, \dots, \mathbb{X}_{h_r}^{(r)}\}$, $r = 1, 2, \dots$. Denote $|\mathbb{X}_k^{(r)}| = n_k^{(r)}$, $1 \le k \le h_r$, $r = 1, 2, \dots$. We assume that $\mathbb{X}_k^{(r)}$ is a random subset of $\mathcal{X}_k^{(r)}$ of size $n_k^{(r)}$ and suppose that given r, the random subsets $\mathbb{X}_1^{(r)}, \dots, \mathbb{X}_{h_r}^{(r)}$ are independent. Furthermore, we assume that $n^{(r)} = n_1^{(r)} + \dots + n_{h_r}^{(r)} \to \infty$ and h_r is bounded as $r \to \infty$. We are interested in the asymptotic distribution of $T^{(r)} = t^{(r)}(\mathbb{X}_1^{(r)}, \dots, \mathbb{X}_{h_r}^{(r)})$. In what follows we skip the superscript (r).

In the simplest case of a linear statistic $T = U_{1,*}$, the asymptotic normality was proved by Erdős and Rény [9] and Bickel and Freedman [3] under very mild Lindeberg type condition. For asymptotically linear statistic ($T \approx U_{1,*}$), by the central limit theorem, for large n, the distribution of T can be approximated by the normal distribution. Furthermore, using the approximation $T \approx U_{s,*}$ one can construct asymptotic expansions to the distribution of T. Bloznelis and Götze [4] showed the validity of one-term asymptotic expansion in the case where $h = 1$.

Let us construct an upper bound for the remainder R_s of the approximation $T = U_{s,*} + R_s$. For this purpose we use moments of finite differences of T. Given $k = 1, \dots, h$ and $j = 1, \dots, n_k^*$ define the first order difference

$$\delta_{k|j} T = t(\mathbb{X}_1, \dots, \mathbb{X}_h) - t(\mathbb{X}_1, \dots, \mathbb{X}_{k-1}, \mathbb{X}_k^j, \mathbb{X}_{k+1}, \dots, \mathbb{X}_h),$$

where we denote $\mathbb{X}_k^j = (\mathbb{X}_k \setminus \{X_{k,j}\}) \cup \{X_{k,n_k+j}\}$. The difference operation $\delta_{k|j}$ can be applied to every function of random variables such that $X_{k,j}$ is among its arguments and X_{k,n_k+j} is not. In particular, given $i \in \Omega_{n_k^*} \setminus \{j\}$ an application of the difference $\delta_{k|i}$ to the statistic $\delta_{k|j} T$ results in random variable $\delta_{k|i} \delta_{k|j} T$ called the second order difference. For $i \in \Omega_{n_k^*}$ write $\Delta_{k|i} = \delta_{k|i} \delta_{k|i-1} \dots \delta_{k|1}$ and given $\overline{a} \le \overline{n}^*$ denote $\Delta_{\overline{a}} = \Delta_{h|a_h} \dots \Delta_{1|a_1}$.

Given $\overline{a} < \overline{n}^*$ introduce the random variable $T_u(\overline{a}) = \sum_{\overline{a} \le \overline{b} \le \overline{n}^*} U(\overline{b})$.

Theorem 3.1 *For $\overline{a} \le \overline{n}^*$ we have*

$$\mathbf{E} T_u^2(\overline{a}) \le n_{\overline{a}} 2^{-|\overline{a}|} \mathbf{E}(\Delta_{\overline{a}} T)^2, \qquad (12)$$

where $n_{\overline{a}} = (n_1^*)^{a_1} \dots (n_h^*)^{a_h}$. *For* $s = 1, \dots, n^* - 1$ *we have*

$$\mathbf{E}R_s^2 \le \sum_{\overline{a}:\, |\overline{a}| = s+1} n_{\overline{a}} 2^{-|\overline{a}|} \mathbf{E}(\Delta_{\overline{a}} T)^2. \tag{13}$$

Informally one can consider $U_{s,*}$ as s-th order polynomial in variables $X_{k,i}$. Thus, it seems natural to formulate results about the error of the approximation $T \approx U_{s,*}$ in terms of finite differences, like $\Delta_{\overline{a}} T$, where $|\overline{a}| = s + 1$. Similar differences were introduced and used by van Zwet [19], Bentkus, Götze and van Zwet [2] in the case of independent observations. Often it is much easier to estimate moments $\mathbf{E}(\Delta_{\overline{a}} T)^2$ than to construct a bound for $\mathbf{E}R_s^2$ directly, cf. Bloznelis and Götze [4], where the case $h = 1$ is considered.

Proof of Theorem 3.1. The inequality (13) follows from (12) and the inequality

$$\mathbf{E}R_s^2 = \sum_{\overline{a} \le \overline{n}^*:\, |\overline{a}| \ge s+1} \mathbf{E}U^2(\overline{a}) \le \sum_{\overline{a} \le \overline{n}^*:\, |\overline{a}| = s+1} \mathbf{E}T_u^2(\overline{a}).$$

Let us prove (12). The simplest case, $h = 1$, is considered in Bloznelis and Götze [4]. For convenience we recall some argument of the proof given ibidem. Write for brevity $\overline{a} = a$, $n_1 = n$, $n_1^* = n^*$ and $X_{1,j} = X_j$, for $1 \le j \le n$. We have $R_s = T_u(s+1)$,

$$U_{s,*} = \sum_{a=0}^{s} U(a), \quad R_s = \sum_{a=s+1}^{n^*} U(a), \quad U(a) = \sum_{1 \le i_1 < \dots < i_a \le n} g_a(X_{i_1}, \dots, X_{i_a}).$$

By (11), $\mathbf{E}R_s^2 = \sum_{a=s+1}^{n^*} \sigma_a^2 C(a)$.

For $s = 0$, we have $\mathbf{E}R_0^2 = \sum_{a=1}^{n^*} \sigma_a^2 C(a) = \mathbf{Var}T$ and
$\Delta_1 T = \sum_{a=1}^{n^*} \tilde{U}(a)$, where

$$\tilde{U}(1) = g_1(X_1) - g_1(X_{n+1}), \qquad \tilde{U}(2) = \sum_{j=2}^{n} (g_2(X_1, X_j) - g_2(X_{n+1}, X_j)), \dots .$$

Clearly, $\mathbf{E}(\Delta_1 T)^2 = \sum_{a=1}^{n^*} \sigma_a^2 \tilde{C}_1(a)$, for some constants $\tilde{C}_1(a) > 0$. In order to prove $\mathbf{E}R_0^2 \le (n^*/2)\mathbf{E}(\Delta_1 T)^2$ we show $C(a) \le (n^*/2)\tilde{C}_1(a)$, for $a = 1, \dots, n^*$.

Similarly, in order to prove

$$\mathbf{E}R_s^2 \le (n^*/2)^{s+1} \mathbf{E}(\Delta_{s+1} T)^2 \tag{14}$$

we evaluate the constants $\tilde{C}_{s+1}(a)$ of the expression

$$\mathbf{E}(\Delta_{s+1} T)^2 = \sum_{a=s+1}^{n^*} \sigma_a^2 \tilde{C}_{s+1}(a)$$

and show the inequalities $C(a) \leq (n^*/2)^{s+1}\tilde{C}_{s+1}(a)$. Detailed calculation is given in Bloznelis and Götze [4].

Let us prove (12) for $h > 1$. Introduce random variables

$$V_k = \Delta_{k|a_k} \dots \Delta_{1|a_1} T_u(\overline{a}), \quad k = 1, \dots, h,$$

and put $V_0 = T_u(\overline{a})$. Since (12) is valid for $h = 1$ we can apply this inequality to the statistic V_{k-1} conditionally given all the random variables but $\mathcal{X}_k^*$. Recall that $\mathbf{E}_k$ denotes the conditional expectation given all the random variables, but $\mathcal{X}_k^*$. We obtain from (14)

$$\mathbf{E}_k V_{k-1}^2 \leq (n_k^*/2)^{a_k} \mathbf{E}_k (\Delta_{k|a_k} V_{k-1})^2 = (n_k^*/2)^{a_k} \mathbf{E}_k V_k^2.$$

Taking expected value we replace conditional expectations by the unconditional ones. Thus, we have

$$\mathbf{E} V_{k-1}^2 \leq (n_k^*/2)^{a_k} \mathbf{E} V_k^2.$$

Choosing $k = 1, \dots, h$ we obtain a chain of inequalities which implies

$$\mathbf{E} T_u(\overline{a})^2 \leq (n_1^*/2)^{a_1} \dots (n_h^*/2)^{a_h} \mathbf{E} V_h^2.$$

Finally, since $V_h = \Delta_{\overline{a}} T_u(\overline{a})$ and the random variables $\Delta_{\overline{a}} T_u(\overline{a})$ and $\Delta_{\overline{a}} T$ coincide we obtain the inequality (12).

4 Appendix

We can assume without loss of generality that $\mathbf{E}T = 0$. Otherwise the argument below applies to the statistic $T - \mathbf{E}T$. Furthermore, with a set $\mathbb{A}_k = \{X_{k,j_1}, \dots, \mathbb{X}_{k,j_{a_k}}\}$ of random elements of the permutation $\mathcal{X}_k^*$ we associate the corresponding index set $A_k = \{j_1, \dots, j_{a_k}\} \subset \Omega_{N_k}$. The conditional expectation $\mathbf{E}(\dots|\mathbb{A}_1, \dots, \mathbb{A}_h)$ will be denoted by $\mathbf{E}(\dots|A_1, \dots, A_h)$.

4.1 Proof of (6).

In the proof we use the following identity, see, e.g., Zhao and Chen [20],

$$\sum_{v=0}^{\min\{s,k\}} (-1)^v \binom{s}{v} \binom{k}{v} \binom{u}{v}^{-1} = \binom{u-s}{k} \binom{u}{k}^{-1}, \tag{15}$$

where the integers $s, t, u \geq 0$ and $u \geq \max\{s; k\}$.

Given $\overline{a} \leq \overline{n}$ let $f_{\overline{a}}$ denote a real function defined on h-tuples of sets $(\mathcal{A}_1, \dots, \mathcal{A}_h)$ such that $\mathcal{A}_k \subset \mathcal{X}_k$ and $|\mathcal{A}_k| = a_k$, $k = 1, \dots, h$. Given $1 \leq i \leq h$ and $B_i \subset \Omega_{N_i}$ let $\mathbf{E}_i(f_{\overline{a}}(\mathbb{A}_1, \dots, \mathbb{A}_h)\,|B_i)$ denote the conditional expectation

$$\mathbf{E}(f_{\overline{a}}(\mathbb{A}_1, \dots, \mathbb{A}_h)\,|\Omega_{N_1}, \dots, \Omega_{N_{i-1}}, B_i, \Omega_{N_{i+1}}, \dots, \Omega_{N_h}).$$

Lemma 4.1. *Let $j \in \{1, \dots, h\}$.*

i) Given a number $b_j \le a_j$ assume that

$$\mathbf{E}_j(f_{\overline{a}}(\mathbb{A}_1, \dots, \mathbb{A}_h)\,|B_j) = 0 \tag{16}$$

for every $B_j \subset A_j$ such that $|B_j| = b_j$. Then (16) holds for every $B_j \subset \Omega_{N_j}$ satisfying $|B_j| = b_j$.

ii) Assume that (16) holds for every $B_j \subset A_j$ such that $|B_j| < a_j$. Then (16) holds for every $B_j \subset \Omega_{N_j}$ satisfying $|B_j| < a_j$. Furthermore, for any $D_j \subset \Omega_{N_j}$ and $w \in A_j \setminus D_j$ we have

$$\begin{aligned}&\mathbf{E}_j(f_{\overline{a}}(\mathbb{A}_1, \dots, \mathbb{A}_h)\,|D_j)\\ &\quad= \frac{-1}{N_j - |A_j \cup D_j| + 1} \sum_{y \in \mathbb{D}_j \setminus \mathbb{A}_j} \mathbf{E}_j(f_{\overline{a}}(\mathbb{A}_1, \dots, \mathbb{A}_j^*(y), \dots \mathbb{A}_h)\,|D_j).\end{aligned} \tag{17}$$

Here we assume that $A_j \setminus D_j$ is nonempty and denote $\mathbb{A}_j^*(y) = \mathbb{A}_j^* \cup \{y\}$, where $\mathbb{A}_j^* = \{X_{j,i} : i \in A_j^*\}$ and $A_j^* = A_j \setminus \{w\}$.

Proof of Lemma 4.1. Let us prove (i) in the case where $j = 1$. Given a set $B_1 \subset \Omega_{N_1}$ with $|B_1| = b_1$ let W_t denote the class of subsets $D \subset \Omega_{N_1}$ of size a_1 such that $|B_1 \setminus D| = t$. In particular, W_0 consists of sets D of size a_1 such that $B_1 \subset D$. By W_t^* we denote the class of subsets D of size a_1 such that $|B_1 \setminus D| \le t$. That is, $W_t^* = W_0 \cup \ldots \cup W_t$. We show that if W_t is nonempty for some $t > 0$ then

(A): for every $D \in W_t$ the conditional expectation

$$\mathbf{E}_1(f_{\overline{a}}(\mathbb{D}, \mathbb{A}_2, \dots, \mathbb{A}_h)\,|\,B_1)$$

is a linear combination of conditional expectations

$$\mathbf{E}_1(f_{\overline{a}}(\mathbb{D}_j, \mathbb{A}_2, \dots, \mathbb{A}_h)\,|\,B_1)$$

where $D_j \in W_{t-1}^*$.

Here $\mathbb{D} = \{X_{1,i},\ i \in D\}$ and $\mathbb{D}_j = \{X_{1,i},\ i \in D_j\}$.

Note that if the statement (A) is true then the validity of the identity (16) (with $j = 1$) for every $A_1 \in W_{t-1}^*$ implies the validity of (16) for arbitrary $A_1 \in W_t$. Using the fact that (16) is valid for every $A_1 \in W_0$ (this, in fact, is the condition of the lemma) we derive the identity (16) for arbitrary $A_1 \in W_t$ using induction over increasing values of $t = 1, 2, \ldots$. Hence, we obtain (i).

In order to prove the statement (A) fix $D \in W_t$ and subset $B^* \subset D$ of size b_1 such that $B_1 \cap D = B_1 \cap B^*$. Let $\mathcal{K}_1$ denote the class of subsets $V \subset \Omega_{N_1} \setminus B^*$ of size $a_1 - b_1$; $\mathcal{K}_2$ denote the class of subsets $V \subset \Omega_{N_1} \setminus (B^* \cup B_1)$ of size $a_1 - b_1$. Clearly, $\mathcal{K}_2 \subset \mathcal{K}_1$. Denote $\mathcal{K}_3 = \mathcal{K}_1 \setminus \mathcal{K}_2$. That is, $\mathcal{K}_3$ consists of those $V \subset \Omega_{N_1} \setminus B^*$ of size $a_1 - b_1$ which satisfy $V \cap (B_1 \setminus B^*) \neq \emptyset$. In particular every union $B^* \cup V$, $V \in \mathcal{K}_3$, is an element of W^*_{t-1}. Denote

$$S_i = \sum_{V \in \mathcal{K}_i} f_{\overline{a}}(\mathbb{A}_{V \cup B^*}, \mathbb{A}_2, \ldots, \mathbb{A}_h), \qquad i = 1, 2, 3.$$

Here $\mathbb{A}_{V \cup B^*}$ denotes the set of random variables $\{X_{1,j},\, j \in V \cup B^*\} \subset \mathcal{X}^*_1$. Clearly, $S_1 - S_2 = S_3$. We have

$$\frac{S_1}{\binom{N_1 - |B^*|}{a_1 - b_1}} = \mathbf{E}_1(f_{\overline{a}}(\mathbb{D}, \mathbb{A}_2, \ldots, \mathbb{A}_h)|\, B^*),$$
$$\frac{S_2}{\binom{N_1 - |B_1 \cup B^*|}{a_1 - b_1}} = \mathbf{E}_1(f_{\overline{a}}(\mathbb{D}, \mathbb{A}_2, \ldots, \mathbb{A}_h)|\, B_1 \cup B^*).$$

Denote the latter (conditional) expectation by $\mathcal{E}$. Since, by our assumption (16), $S_1 = 0$, we obtain $S_2 = -S_3$. Now the identity

$$\mathbf{E}_1(f_{\overline{a}}(\mathbb{D}, \mathbb{A}_2, \ldots, \mathbb{A}_h)|\, B_1) = \mathbf{E}_1(\mathcal{E}|B_1) = \frac{-1}{\binom{N_1 - |B_1 \cup B^*|}{a_1 - b_1}} \mathbf{E}_1(S_3|B_1)$$

(in the last step we replaced S_2 by $-S_3$) completes the proof of the statement (A). Indeed, S_3 is a linear combination of $f_{\overline{a}}(\mathbb{D}_j, \mathbb{A}_2, \ldots, \mathbb{A}_h)$, where $D_j \in W^*_{t-1}$. Hence, (i) is proved.

Let us prove (ii). In order to prove (4.3), fix D_j and $w \in A_j \setminus D_j$. We have

$$\mathbf{E}_j(f_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}_h)\,|D_j) = \mathbf{E}_j(\mathcal{E}|D_j), \qquad \mathcal{E} := \mathbf{E}_j(f_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}_h)\,|D_j \cup A^*_j). \tag{18}$$

Clearly,

$$\mathcal{E} = \frac{1}{N_j - |D_j \cup A^*_j|} \sum_{y \in \mathcal{X}^*_j \setminus (\mathbb{D}_j \cup \mathbb{A}^*_j)} f_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}^*_j(y), \ldots, \mathbb{A}_h). \tag{19}$$

Write $\mathcal{E}_1 = \mathbf{E}_j(f_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}_h)\,|\, A^*_j)$. By (16), $\mathcal{E}_1 = 0$. Therefore,

$$(N_j - |A^*_j|)\mathcal{E}_1 = \sum_{y \in \mathcal{X}^*_j \setminus \mathbb{A}^*_j} f_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}^*_j(y), \ldots, \mathbb{A}_h) = 0. \tag{20}$$

Combining (19) and (20) we obtain

$$\mathcal{E} = \frac{-1}{N_j - |D_j \cup A^*_j|} \sum_{y \in \mathbb{D}_j \setminus \mathbb{A}_j} f_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}^*_j(y), \ldots, \mathbb{A}_h).$$

Substitution of this expression of $\mathcal{E}$ in (18) yields (4.3).

Let us prove (16) for B_j such that $|B_j| < a_j$ and $t = |B_j \setminus A_j|$ is a positive number. For $t = 1$ the result follows from (4.3) applied to $D_j = B_j$. For $t > 1$ an application of the identity (4.3) reduces the problem to the case of $t-1$. The desired result follows after t iterations of application of (4.3).

Lemma 4.2. *Given $\overline{a}$ satisfying*

$$2a_k \leq N_k + 1, \qquad 1 \leq k \leq h, \tag{21}$$

assume that (6) holds for every $\psi_{\overline{b}}$ with $\overline{b} \leq \overline{a}$. Then the coefficients $C(\overline{a}, \overline{b})$ of (3.3) satisfy (7).

Proof of Lemma 4.2. We prove the lemma in the case where $h = 2$. The proof for $h = 3, 4, \ldots$ is much the same.

Fix $\overline{b} = (b_1, b_2)$ such that $\overline{b} < \overline{a}$. Let $B_1, D_1 \subset \Omega_{N_1}$ and $B_2, D_2 \subset \Omega_{N_2}$ be such that $|B_1| = |D_1| = b_1$ and $|B_2| = |D_2| = b_2$. Denote $r = |B_1 \setminus D_1|$ and $s = |B_2 \setminus D_2|$. Write $\psi = \psi_{\overline{b}}$ for short. We start with an auxiliary identity

$$\mathbf{E}(\psi(\mathbb{D}_1, \mathbb{D}_2) \,|\, B_1, B_2) = K_{b_1,r} K_{b_2,s} \psi(\mathbb{B}_1, \mathbb{B}_2), \qquad K_{u,t} = \frac{(-1)^t}{\binom{N_1 - u}{t}} \tag{22}$$

Write $\mathcal{E}_1 = \mathbf{E}_1(\psi(\mathbb{D}_1, \mathbb{D}_2)|B_1)$ and $\mathcal{E}_2 = \mathbf{E}_2(\psi(\mathbb{B}_1, \mathbb{D}_2)|B_2)$. If we had shown that

$$\mathcal{E}_1 = K_{b_1,r} \psi(\mathbb{B}_1, \mathbb{D}_2), \qquad \mathcal{E}_2 = K_{b_2,s} \psi(\mathbb{B}_1, \mathbb{B}_2) \tag{23}$$

then (22) would follow from the identities

$$\begin{aligned} \mathbf{E}(\psi(\mathbb{D}_1, \mathbb{D}_2)|B_1, B_2) &= \mathbf{E}(\mathcal{E}_1|B_1, B_2) = K_{b_1,r}\, \mathbf{E}(\psi(\mathbb{B}_1, \mathbb{D}_2)|B_1, B_2) \\ &= K_{b_1,r} \mathbf{E}(\mathcal{E}_2|B_1, B_2) = K_{b_1,r} K_{b_2,s}\, \psi(\mathbb{B}_1, \mathbb{B}_2). \end{aligned}$$

Therefore, in order to prove (22) it suffices to show (23). We shall prove the first identity of (23) only. Note that for $r = 0$ this identity is obvious. In what follows we consider the case where $r > 0$.

A subset of Ω_{N_1} of size b_1 is said to belong to the class W_t if it has exactly $b_1 - t$ common elements with the set B_1. We claim that if H belongs to W_t, $t \geq 1$, then

$$\mathbf{E}_1(\psi(\mathbb{H}, \mathbb{D}_2) \,|\, B_1) = \frac{-1}{N_1 - b_1 - t + 1} \sum_{i=1}^{t} \mathbf{E}_1(\psi(\mathbb{V}^{(i)}, \mathbb{D}_2) \,|\, B_1), \tag{24}$$

where $V^{(1)}, \ldots, V^{(t)}$ are distinct elements of W_{t-1}. Here we denote $\mathbb{H} = \{X_{1,j},\ j \in H\}$ and $\mathbb{V}^{(i)} = \{X_{1,j},\ j \in V^{(i)}\}$. Indeed, (24) follows from (4.3).

Starting with $H = D_1$ we iterate (24) until obtain a sum of conditional expectations $\mathbf{E}_1(\psi(\mathbb{B}_1, \mathbb{D}_2)|B_1) = \psi(\mathbb{B}_1, \mathbb{D}_2)$ in the right hand side. Therefore, after r iteration steps we have, for some number K, $\mathbf{E}_1(\psi(\mathbb{D}_1, \mathbb{D}_2)|B_1) = K\, \psi(\mathbb{B}_1, \mathbb{D}_2)$. A straightforward calculation shows that

$$K = (-1)^r \prod_{i=0}^{r-1} \frac{r - i}{N_1 - b_1 - r + 1 + i} = \frac{(-1)^r}{\binom{N_1 - b_1}{r}} = K_{b_1,r}.$$

Now we derive (7). Fix (A_1, A_2) and (B_1, B_2) such that $B_k \subset A_k \subset \Omega_{N_k}$ for $k = 1, 2$ and the (size) vectors $\overline{a}$ and $\overline{b}$ satisfy $\overline{b} < \overline{a}$. By (3.3), the conditional expectation $\mathbf{E}(\psi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2) \,|\, B_1, B_2)$ equals

$$\varphi_{\overline{b}}(\mathbb{B}_1, \mathbb{B}_2) - \sum_{\overline{d}<\overline{a}} \sum_{|D_1|=d_1} \sum_{|D_2|=d_2} C(\overline{a}, \overline{d}) \mathbf{E}(\psi_{\overline{d}}(\mathbb{D}_1, \mathbb{D}_2) \,|\, B_1, B_2). \qquad (25)$$

Since $\mathbf{E}(\psi_{\overline{d}}(\mathbb{D}_1, \mathbb{D}_2) \,|\, B_1, B_2) = 0$ for any $\overline{d} < \overline{a}$ which fails to satisfy $\overline{d} \le \overline{b}$ we can write (25) as follows

$$\varphi_{\overline{b}}(\mathbb{B}_1, \mathbb{B}_2) - S, \quad S = \sum_{\overline{d}\le\overline{b}} \sum_{|D_1|=d_1} \sum_{|D_2|=d_2} C(\overline{a}, \overline{d}) \mathbf{E}(\psi_{\overline{d}}(\mathbb{D}_1, \mathbb{D}_2) \,|\, B_1, B_2).$$

Here and above $\sum_{|D_k|=d_k}$ denotes the sum over all subsets $D_k \subset A_k$ of size d_k. Let us collect the terms $\varphi_{\overline{b}}(\mathbb{B}_1, \mathbb{B}_2)$ in (25). Clearly, for $\overline{d} < \overline{b}$ the conditional expectation $\mathbf{E}(\psi_{\overline{d}}(\mathbb{D}_1, \mathbb{D}_2) \,|\, B_1, B_2)$ has no such a term, cf. (3.3). Split $S = S_1 + S_2$, where the sum S_1 includes the summands of S with indices $\overline{d}$ satisfying $\overline{d} < \overline{b}$. Furthermore,

$$S_2 = C(\overline{a}, \overline{b}) \sum_{|D_1|=b_1} \sum_{|D_2|=b_2} \mathbf{E}(\psi_{\overline{d}}(\mathbb{D}_1, \mathbb{D}_2) \,|\, B_1, B_2).$$

By (22), $S_2 = C(\overline{a}, \overline{b})\, V\, \psi_{\overline{b}}(\mathbb{B}_1, \mathbb{B}_2)$, where

$$V = V_1^* V_2^*, \qquad V_k^* = \sum_{|D_k|=b_k, D_k\subset A_k} K_{b_k, |B_k \setminus D_k|}, \quad k = 1, 2. \qquad (26)$$

In particular, $V_k^* = 1$ for $b_k = a_k$. For $b_k < a_k$ an application of the identity (15) gives $V_k^* = V_k^{-1}(a_k, b_k)$. Hence, the coefficient of $\varphi_{\overline{b}}(\mathbb{B}_1, \mathbb{B}_2)$ in (25) is $1 - C(\overline{a}, \overline{b})V$ with $V^{-1} = V_1(a_1, b_1)V_2(a_2, b_2)$. In order to ensure (6) we make this coefficient zero, that is, choose $C(\overline{a}, \overline{b}) = V^{-1}$. We arrive to (7).

Lemma 4.3. *Let $\overline{a} \le \overline{n}$. The function $\psi_{\overline{a}}$ defined by (3.3) satisfies (6). If $2a_k > N_k$ for some k then $\psi_{\overline{a}} \equiv 0$ almost surely.*

Proof of Lemma 4.3. We shall prove the lemma in the case where $h = 2$. The proof for $h = 3, 4, \ldots$ is almost the same. We split the proof in two steps.

Step 1. Here we prove (6) for $\overline{a}$ satisfying (21). The proof uses induction over increasing values of a_1, a_2. For $\overline{a} = (0, 1)$ and $\overline{a} = (1, 0)$, the identity (6) follows from $\mathbf{E}T = 0$. Given $\overline{a}$ with $a_1 + a_2 > 1$ we assume that (6) holds for every $\psi_{\overline{b}}$ with $\overline{b} < \overline{a}$ (induction hypothesis) and derive (6) for $\psi_{\overline{a}}$.

It suffices to show that for every numbers b_1, b_2 such that $b_1 < a_1$ and $b_2 < a_2$ and arbitrary sets $B_k, A_k \subset \Omega_{N_k}$ with $|A_k| = a_k$, $|B_k| = b_k$, $k = 1, 2$, we have almost surely

$$\mathbf{E}_1(\psi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2) \,|\, B_1) = 0, \qquad (27)$$

$$\mathbf{E}_2(\psi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2) \,|\, B_2) = 0.$$

We prove (27) only.

Given $\overline{h} = (h_1, h_2)$ and $H_k \subset \Omega_{N_k}$ with $|H_k| = h_k$, $k = 1, 2$ we write, for $\overline{d} \le \overline{h}$,

$$U_{H_1,H_2}(\overline{d}) = C(\overline{h}, \overline{d}) \sum_{|D_1|=d_1} \sum_{|D_2|=d_2} \mathbf{E}_1(\psi_{\overline{d}}(\mathbb{D}_1, \mathbb{D}_2) \,|\, B_1), \qquad (28)$$

where the sum $\sum_{|D_k|=d_k}$ is taken over all subsets $D_k \subset H_k$ of size $|D_k| = d_k$.

In view of Lemma 4.1 it suffices to prove (27) for B_1 satisfying $B_1 \subset A_1$. Denote $\overline{u} = (b_1, a_2)$. Since, by induction hypothesis, $\mathbf{E}_1(\psi_{\overline{d}}(\mathbb{D}_1, \mathbb{D}_2) \,|\, B_1) = 0$ for $\overline{d} < \overline{a}$ with $d_1 > b_1$, we obtain from (3.3) that

$$\mathbf{E}_1(\psi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2) \,|\, B_1) = \varphi_{\overline{u}}(\mathbb{B}_1, \mathbb{A}_2) - S, \quad S = \sum_{\overline{d} \le \overline{u}} U_{A_1,A_2}(\overline{d}), \qquad (29)$$

Split $S = S_1 + S_2$, where $S_1 = U_{A_1,A_2}(\overline{u})$ and $S_2 = \sum_{\overline{d}<\overline{u}} U_{A_1,A_2}(\overline{d})$. The identity (23) combined with (15) gives $S_1 = V_1^{-1}(a_1, b_1) C(\overline{a}, \overline{u}) \psi_{\overline{u}}(\mathbb{B}_1, \mathbb{A}_2)$, cf (26). Note that, by induction hypothesis, (6) holds for every $\overline{d} < \overline{a}$ and, therefore, it holds for every $\overline{d} \le \overline{u}$ (since $\overline{u} < \overline{a}$). Hence, conditions of Lemma 4.2 are satisfied for $\overline{d} \le \overline{u}$ and we are allowed to use (23) here.) Since $V_1^{-1}(a_1, b_1) C(\overline{a}, \overline{u}) = 1$, we obtain $S_1 = \psi_{\overline{u}}(\mathbb{B}_1, \mathbb{A}_2)$. Note that in order to prove (27) it suffices to show that

$$S_2 = \varphi_{\overline{u}}(\mathbb{B}_1, \mathbb{A}_2) - \psi_{\overline{u}}(\mathbb{B}_1, \mathbb{A}_2). \qquad (30)$$

Let us prove (30). It follows from (3.3) (applied to $\psi_{\overline{u}}$) that (30) is equivalent to the identity $S_2 = \sum_{\overline{d}<\overline{u}} U_{B_1,A_2}(\overline{d})$, since

$$U_{B_1,A_2}(\overline{d}) = C(\overline{u}, \overline{d}) \sum_{|D_1|=d_1, D_1 \subset B_1} \sum_{|D_2|=d_2, D_2 \subset A_2} \psi_{\overline{d}}(\mathbb{D}_1, \mathbb{D}_2).$$

Therefore, in order to prove (30) it suffices to show that for every $\overline{d} < \overline{u}$

$$U_{A_1,A_2}(\overline{d}) = U_{B_1,A_2}(\overline{d}). \qquad (31)$$

Firstly we prove (31) for $\overline{d}$ such that $d_1 = b_1$ and $d_2 < a_2$. The identity (23) combined with (15) (cf. (26)) gives

$$U_{A_1,A_2}(\overline{d}) = C(\overline{a}, \overline{d}) \sum_{D_2 \subset A_2, |D_2|=d_2} V_1^{-1}(a_1, b_1) \psi_{\overline{u}}(\mathbb{B}_1, \mathbb{D}_2) = U_{B_1,A_2}(\overline{d}).$$

In the last step we applied the identity $V_1^{-1}(a_1, b_1) C(\overline{a}, \overline{d}) = C(\overline{u}, \overline{d})$.

Let us prove (31) for $\overline{d}$ such that $d_1 < b_1$. Given $D_2 \subset A_2$ denote

$$\mathcal{U} = \sum_{|D_1|=d_1, D_1 \subset A_1} \mathbf{E}_1(\psi_{\overline{d}}(\mathbb{D}_1, \mathbb{D}_2)|B_1), \qquad \mathcal{U}^* = \sum_{|D_1|=d_1, D_1 \subset B_1} \psi_{\overline{d}}(\mathbb{D}_1, \mathbb{D}_2).$$

In order to prove (31) we show that

$$C(\overline{a},\overline{d})\,\mathcal{U}=C(\overline{u},\overline{d})\,\mathcal{U}^*. \tag{32}$$

Given an integer $t\ge 0$ let Q_t denote the class of subsets $D_1\subset A_1$ such that $|D_1|=d_1$ and $|D_1\setminus B_1|=t$. Clearly, $0\le t\le t_0=\min\{d_1,a_1-b_1\}$. Split

$$\mathcal{U}=U_0+\ldots+U_{t_0},\qquad U_t=\sum_{D_1\in Q_t}\mathbf{E}_1(\psi_{\overline{d}}(\mathbb{D}_1,\mathbb{D}_2)\,|\,B_1). \tag{33}$$

We shall show below that, for $t=0,1,\ldots,t_0$,

$$U_t=\frac{\binom{d_1}{d_1-t}\binom{a_1-b_1}{t}(-1)^t}{\binom{N_1-b_1}{t}}\mathcal{U}^*. \tag{34}$$

Collecting these expressions in (33) and then using (15) we obtain (32).

It remains to prove (34). For $t=0$ (34) is obvious. Let $t>0$. Given $D_1\subset Q_t$ write $D_1=K\cup L$, where $L=D_1\cap B_1$ and $K=D_1\setminus B_1$. Write for convenience $K=\{k_1,\ldots,k_t\}$. Furthermore, given a subset $\{w_1,\ldots,w_s\}\subset B_1\setminus D_1$ denote

$$\begin{aligned}D_1(w_1,\ldots,w_s)&=L\cup\{w_1,\ldots,w_s\}\cup\{k_{s+1},\ldots,k_t\},\\ \mathbb{D}_1(w_1,\ldots,w_s)&=\{X_{1,j}\,:\,j\in D_1(w_1,\ldots,w_s)\}.\end{aligned}$$

An application of (4.3) (with $w=k_1$) gives

$$\mathbf{E}_1(\psi_{\overline{d}}(\mathbb{D}_1,\mathbb{D}_2)\,|\,B_1)=\frac{-1}{m_0}\sum_{w_1\in B_1\setminus D_1}\mathbf{E}_1(\psi_{\overline{d}}(\mathbb{D}_1(w_1),\mathbb{D}_2)\,|\,B_1),$$

where $m_0=N_1-|B_1\cup D_1|+1$. Iterated applications of (4.3) (with $w=k_1,k_2,\ldots,k_t$ respectively) yields

$$\mathbf{E}_1(\psi_{\overline{d}}(\mathbb{D}_1,\mathbb{D}_2)\,|\,B_1)=\prod_{s=1}^{t}\frac{-1}{m_{s-1}}\sum\nolimits^{*}\mathbf{E}_1(\psi_{\overline{d}}(\mathbb{D}_1(w_1,\ldots,w_t),\mathbb{D}_2)\,|\,B_1), \tag{35}$$

where $m_s=N_1-|B_1\cup D_1(w_1,\ldots,w_s)|+1$ and where $\sum^*$ denotes the sum

$$\sum_{w_1\in B_1\setminus D_1}\quad\sum_{w_2\in B_1\setminus D_1(w_1)}\cdots\sum_{w_t\in B_1\setminus D_1(w_1,\ldots,w_{t-1})}.$$

Clearly, $D_1(w_1,\ldots,w_t)\subset B_1$. Therefore,

$$\mathbf{E}_1(\psi_{\overline{d}}(\mathbb{D}_1(w_1,\ldots,w_t),\mathbb{D}_2)\,|\,B_1)=\psi_{\overline{d}}(\mathbb{D}_1(w_1,\ldots,w_t),\mathbb{D}_2).$$

A simple calculation shows that

$$\sum\nolimits^{*}\psi_{\overline{d}}(\mathbb{D}_1(w_1,\ldots,w_t),\mathbb{D}_2)=t!\,U_t^*(L),$$

$$U_t^*(L) = \sum_{M\subset B_1\setminus L,\,|M|=t} \psi_{\overline{d}}(\mathbb{D}_{M\cup L}, \mathbb{D}_2),$$

where $\mathbb{D}_{M\cup L} = \{X_{1,j} : j \in M \cup L\}$. Therefore, we obtain from (35)

$$\mathbf{E}_1(\psi_{\overline{d}}(\mathbb{D}_1, \mathbb{D}_2)\,|\,B_1) = \Big(\prod_{s=1}^{t} \frac{-1}{m_{s-1}}\Big)\, t!\, U_t^*(L) = \frac{(-1)^t}{\binom{N_1-b_1}{t}} U_t^*(L).$$

Finally, using this formula we derive (34):

$$\begin{aligned} U_t &= \sum_{L\subset B_1,\,|L|=d_1-t} \ \sum_{K\subset A_1\setminus B_1,\,|K|=t} \mathbf{E}_1(\psi_{\overline{d}}(\mathbb{D}_{L\cup K}, \mathbb{D}_2)|B_1) \\ &= \sum_{L\subset B_1,\,|L|=d_1-t} (-1)^t \frac{\binom{a_1-b_1}{t}}{\binom{N_1-b_1}{t}} U_t^*(L) \\ &= (-1)^t \frac{\binom{d_1}{d_1-t}\binom{a_1-b_1}{t}}{\binom{N_1-b_1}{t}} \mathcal{U}^*. \end{aligned}$$

In the last step we use the fact that given $D_1 \subset B_1$ of size d_1 there are $\binom{d_1}{d_1-t}$ different ways to write $D_1 = L \cup M$, where $|L| = d_1 - t$ and $|M| = t$. We arrive to (34) thus completing the proof of (32).

Step 2. Here we prove (6) for $\psi_{\overline{a}}$ where $\overline{a}$ fails to satisfy (21). Given such a vector $\overline{a}$ denote

$$\overline{A}_{\overline{a}} = \{\overline{d} : \overline{d} < \overline{a} \ \text{ and } \ C(\overline{a}, \overline{d}) \neq 0\}.$$

By (3.3), we have

$$\psi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2) = \varphi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2) - \sum_{\overline{d}\in\overline{A}_{\overline{a}}} C(\overline{a}, \overline{d}) \sum_{|D_1|=d_1} \sum_{|D_2|=d_2} \psi_{\overline{d}}(\mathbb{D}_1, \mathbb{D}_2), \qquad (36)$$

where $\sum_{|D_k|=d_k}$ denotes the sum over all subsets $D_k \subset A_k$ of size $|D_k| = d_k$. We shall show that almost surely

$$\psi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2) = 0. \qquad (37)$$

Clearly, (37) implies (6).

Since $\overline{a}$ fails to satisfy (21) there exists a nonempty subset $Z_{\overline{a}} \subset \{1, 2\}$ defined by $k \in Z_{\overline{a}} \Leftrightarrow 2a_k > N_k + 1$. In order to prove (37) we show that given $k \in Z_{\overline{a}}$ and arbitrary $A_j \in \Omega_{N_j}$ with $|A_j| = a_j$, $j = 1, 2$, we have almost surely

$$\mathbf{E}_k(\psi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2)\,|\,B_k) = 0, \qquad (38)$$

for every set $B_k \subset \Omega_{N_k}$ of size $|B_k| = N_k - a_k$. Indeed, write $B_k' = \Omega_{N_k} \setminus A_k$. Since $\mathbf{E}_k(\psi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2)|B_k') = \psi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2)$ we obtain from (38)

$$0 = \mathbf{E}_k(\psi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2)\,|\,B_k') = \psi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2). \qquad (39)$$

Note that in view of Lemma 4.1 (i) it suffices to prove (38) for $B_k \subset A_k$ (observe that the inequality $2a_k > N_k + 1$ implies $a_k > |B_k| = N_k - a_k$).

Firstly we prove (38) in the case where $Z_{\overline{a}}$ has only one element, say $Z_{\overline{a}} = \{1\}$. Since $2 \notin Z_{\overline{a}}$ we have $2a_2 \leq N_2 + 1$. Therefore, given a subset $B_1 \subset A_1$ of size $b_1 = N_1 - a_1$ the vector $\overline{u} = (b_1, a_2)$ satisfies (21). In particular, the argument used to prove (27) in *Step 1* can be used to derive (38) for $k = 1$ in the case where $B_1 \subset A_1$. Hence, (38) follows. We conclude that (37) holds for $\overline{a}$ satisfying $|Z_{\overline{a}}| = 1$.

Now assume that $Z_{\overline{a}} = \{1, 2\}$. We prove (38) for $k = 1$ for arbitrary $B_1 \subset A_1$ of size $b_1 = N_1 - a_1$. Write $\overline{u} = (b_1, a_2)$ and note that $\overline{A}_{\overline{a}} = \overline{A}_{\overline{u}}$. From (36) we obtain

$$\mathbf{E}_1(\psi_{\overline{a}}(\mathbb{A}_1, \mathbb{A}_2) \,|\, B_1) = \varphi_{\overline{u}}(\mathbb{B}_1, \mathbb{A}_2) - \sum_{\overline{d} \in \overline{A}_{\overline{u}}} U_{A_1, A_2}(\overline{d}). \tag{40}$$

Here we use the notation (28). Using (31) we can replace $U_{A_1,A_2}(\overline{d})$ by $U_{B_1,A_2}(\overline{d})$. After this replacement the right hand side of (40) coincides with the expression (36) applied to $\psi_{\overline{u}}(\mathbb{B}_1, \mathbb{A}_2)$. But $\psi_{\overline{u}}(\mathbb{B}_1, \mathbb{A}_2) = 0$ almost surely, by (37). Therefore, the right-hand side of (40) is zero almost surely. We obtain (38). Note that the use of (37) is legitimate since $|Z_{\overline{u}}| = 1$ and we have already proved the validity of (37) for this case.

Finally we complete the proof of the lemma by showing that almost surely $\psi_{\overline{a}} \equiv 0$, provided $2a_k > N_k$ for some k. We have $N_k - a_k < a_k$. By (6), (38) holds for every B_k of size $N_k - a_k$. Given A_k choose $B_k = \Omega_{N_k} \setminus A_k$. Now (39) provides desired result. The lemma is proved.

4.2 Proof of the variance decomposition formula (11).

Lemma 4.4. *For every $\overline{a} \leq \overline{n}^*$ the identity (11) holds.*

Proof of Lemma 4.4. For $\overline{a} \leq \overline{n}^*$ denote $M_j = \prod_{i=1}^{j} V_i^{-1}(n_i, a_i)$. Fix a h-tuple $(D_1, \ldots, D_h)$ where $D_k \subset \Omega_{n_k}$ with $|D_k| = a_k$, $1 \leq k \leq h$, and denote $g_* = g_{\overline{a}}(\mathbb{D}_1, \ldots, \mathbb{D}_h)$.

We have, by symmetry,

$$\mathbf{E}U^2(\overline{a}) = L\, \mathbf{E}U(\overline{a})\, g_*.$$

Here $L = \prod_{k=1}^{h} \binom{n_k}{a_k}$ denotes the number of different summands in $U(\overline{a})$. Since $L\, M_h = C(\overline{a})$, we obtain (11) form the identity

$$\mathbf{E}U(\overline{a})\, g_* = \sigma_{\overline{a}}^2\, M_h. \tag{41}$$

Let us prove (41). Write $G_0 = U(\overline{a})\, g_*$ and $G_h = M_h \sigma_{\overline{a}}^2$. For $j = 1, \ldots, h-1$ introduce the random variables

$$G_j = M_j \sum_{|A_{j+1}| = a_{j+1}} \cdots \sum_{|A_h| = a_h} \mathbf{E}_1 \ldots \mathbf{E}_j g_{\overline{a}}(\mathbb{D}_1, \ldots, \mathbb{D}_j, \mathbb{A}_{j+1}, \ldots, \mathbb{A}_h)\, g_*.$$

Here the sum $\sum_{|A_k|=a_k}$ is taken over all subsets $A_k \subset \Omega_{n_k}$, $j+1 \le k \le h$. We shall show that

$$\mathbf{E}_j G_{j-1} = G_j, \qquad \text{for} \qquad j = 1, \dots, h. \tag{42}$$

These identities imply (41). Indeed, we have

$$\mathbf{E} G_0 = \mathbf{E}_h \dots \mathbf{E}_1 G_0 = \mathbf{E}_h \dots \mathbf{E}_2 G_1 = \dots = \mathbf{E}_h G_{h-1} = G_h.$$

It remains to prove (42). For this purpose it suffices to show that given $A_{j+1}, \dots, A_h$,

$$\sum_{|A_j|=a_j} \mathbf{E}_j g_{\overline{a}}(\mathbb{D}_1, \dots, \mathbb{D}_{j-1}, \mathbb{A}_j, \dots, \mathbb{A}_h)\, g_* \tag{43}$$

$$= V_j^{-1}(n_j, a_j) \mathbf{E}_j g_{\overline{a}}(\mathbb{D}_1, \dots, \mathbb{D}_j, \mathbb{A}_{j+1}, \dots, \mathbb{A}_h)\, g_*.$$

Proceeding as in proof of (23) we obtain, for every $A_j \subset \Omega_{n_j}$ with $|A_j| = a_j$

$$\mathbf{E}_j (g_{\overline{a}}(\mathbb{D}_1, \dots, \mathbb{D}_{j-1}, \mathbb{A}_j, \dots, \mathbb{A}_h) |\, D_j)$$

$$= K_{a_j, |D_j \setminus A_j|} g_{\overline{a}}(\mathbb{D}_1, \dots, \mathbb{D}_j, \mathbb{A}_{j+1}, \dots, \mathbb{A}_h).$$

Therefore, the left-hand side of (43) equals

$$S\, \mathbf{E}_j g_{\overline{a}}(\mathbb{D}_1, \dots, \mathbb{D}_j, \mathbb{A}_{j+1}, \dots, \mathbb{A}_h)\, g_*, \qquad \text{where} \qquad S = \sum_{|A_j|=a_j} K_{a_j, |D_j \setminus A_j|}.$$

Since there are $\binom{n_j - a_j}{v}\binom{a_j}{a_j - v}$ different possibilities to choose $A_j \subset \Omega_{n_j}$ of size a_j such that $|D_j \setminus A_j| = v$, we have

$$S = \sum_v \binom{n_j - a_j}{v} \binom{a_j}{a_j - v} K_{a_j, v} = V_j^{-1}(n_j, a_j).$$

In the last step we applied (15). We arrive to (43) thus completing the proof of (42).

References

[1] Barbour, A. D., Karoński, M. and Ruciński, A. (1989) A central limit theorem for decomposable random variables with applications to random graphs. *J. Comb. Theory B.* **47** 125–145.

[2] Bentkus, V., Götze, F. and van Zwet, W. R. (1997) An Edgeworth expansion for symmetric statistics. *Ann. Statist.* **25** 851–896.

[3] Bickel, P. J. and Freedman, D. A. (1984) Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist.* **12** 470–482.

[4] Bloznelis, M. and Götze, F. (2001) Orthogonal decomposition of finite population statistics and its applications to distributional asymptotics. *Ann. Statist.* **29** 899–917.

[5] Bollobás, B. (1985) *Random graphs.* Academic Press.

[6] Cochran, W. G.(1977) *Sampling techniques.* Wiley.

[7] De Jong, P. (1996) A central limit theorem with applications to random hypergraphs. *Random Struct. Algorithms* **8** 105–120.

[8] Efron, B. and Stein, C. (1981) The jackknife estimate of variance. *Ann. Statist.* **9** 586–596.

[9] Erdős, P. and Rényi, A. (1959) On the central limit theorem for samples from a finite population. *Publ. Math. Inst. Hungar. Acad. Sci.* **4** 49–61.

[10] Hajek, J. (1968) Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.

[11] Hoeffding, W. (1948) A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

[12] Janson, S. (1990) A functional limit theorem for random graphs with applications to subgraph count statistics. *Random. Struct. Algorithms* **1** 15–37.

[13] Janson, S. (1994) Orthogonal decompositions and Functional Limit Theorems for Random Graph Statistics. *Memoirs Amer. Math. Soc.* **534**.

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

[15] Koroljuk, V. S., Borovskikh, Yu. V. (1994) *Theory of U-statistics* Kluwer.

[16] Lehmann, E. L. (1963) Robust estimation in analysis of variance. *Ann. Math. Statist.* **34** 957–966.

[17] Rubin, H. and Vitale, R. A. (1980) Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8** 165–170.

[18] Särndal, C.-E., Swensson, B. and Wretman, J. (1997) *Model assisted survey sampling.* Springer.

[19] Van Zwet W. (1984) A Berry–Esseen bound for symmetric statistics. *Z. Wahrsch. Verw. Gebiete* **66** 425–440.

[20] Zhao L. C. and Chen X. R. (1990) Normal approximation for finite-population U-statistics *Acta mathematicae applicatae Sinica* **6** 263–272.