A NOTE ON THE BIAS AND CONSISTENCY OF THE JACKKNIFE VARIANCE ESTIMATOR IN STRATIFIED SAMPLES

M. Bloznelis

Abstract. Using the ANOVA decomposition, we obtain an explicit formula for the bias of the jackknife variance estimator in stratified samples drawn without replacement. For a wide class of asymptotically linear statistics, we show the consistency of the jackknife variance estimator and establish the asymptotic normality of their Studentized versions.

1. Introduction

1.1. In this note, we provide exact finite sample theoretical results for the jackknife variance estimator of *nonlinear* finite population statistics based on samples *drawn without replacement*. For simplicity, we consider stratified samples where, from each stratum, a sample is drawn without replacement and independently across the strata (STSI samples).

Our analysis is based on an orthogonal decomposition of statistics which can be considered as a kind of linearization technique. The orthogonal decomposition (also called the ANOVA decomposition or the Hoeffding decomposition) of statistics was introduced by Hoeffding (1948). Efron and Stein (1981) applied the decomposition to study the jackknife variance estimator in the case of independent observations. Here we adapt the orthogonal decomposition technique to STSI samples and extend the results of Efron and Stein (1981) to samples drawn without replacement. In particular, we prove the Efron–Stein inequality for STSI samples.

It seems that the orthogonal decomposition for STSI samples was not known in the literature before. We believe that it can provide a convenient alternative to the commonly used Taylor linearization method (see, e.g., Rao and Wu (1988), Shao (1996), and Chen and Sitter (1999)) in the cases where this method is not applicable, e.g., for U- and L-statistics. Note that the

1991 *Mathematics Subject Classification.* Primary 62E20; secondary 60F05.

Key words and phrases. Hoeffding decomposition, jackknife variance estimation, stratified sample without replacement, Efron–Stein inequality, normal approximation.

influence functions used by Shao (1994) to study L-statistics are related to the linear part of the orthogonal decomposition.

Although the present paper focuses on the variance estimation, the orthogonal decomposition has other important applications, for example, second order approximations to the distributions of nonlinear statistics (see Helmers (1991), Bentkus, Götze, and van Zwet (1997), and Putter and van Zwet (1998), where i.i.d. samples were considered).

1.2. Let $\mathcal{X} = \{x_1, \dots, x_N\}$ be a population of N units which divided into h nonoverlapping subpopulations $\mathcal{X} = \mathcal{X}_1 \cup \dots \cup \mathcal{X}_h$, where $\mathcal{X}_k = \{x_{k\,1}, \dots, x_{k\,N_k}\}$. Therefore, $N_1 + \dots + N_h = N$. Given $n_k < N_k$, the simple random sample $\mathbb{X}_k = \{X_{k\,1}, \dots, X_{k\,n_k}\}$ is drawn without replacement from the stratum $\mathcal{X}_k$. We assume that the samples $\mathbb{X}_1, \dots, \mathbb{X}_h$ are independent. Let

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

be a real-valued statistic based on STSI sample $\mathbb{X} = (\mathbb{X}_1, \dots, \mathbb{X}_h)$, and let

$$\sigma_T^2 = \mathbf{Var}\, T$$

denote its variance. We shall write $T = t(\mathbb{X})$ for brevity. The only condition we impose on t is that the value of t is invariant under permutations of observations within every sample $\mathbb{X}_k$, $k = 1, \dots, h$. This is a very mild condition in view of the sampling model considered. We call the statistic T (kernel t) symmetric.

We consider the jackknife estimator $v^2(T)$ of σ_T^2 based on values of t on samples obtained by deleting observations from the enlarged STSI sample $\mathbb{X}' = (\mathbb{X}'_1, \dots, \mathbb{X}'_h)$. Here $\mathbb{X}'_k = \{X_{k\,1}, \dots, X_{k\,n_k+1}\}$ denotes the sample of size $n_k + 1$ drawn without replacement from $\mathcal{X}_k$ (the additional observation $X_{k\,n_k+1}$ is drawn from $\mathcal{X}_k \setminus \mathbb{X}_k$ and added to $\mathbb{X}_k$ to form $\mathbb{X}'_k$). For $j = 1, \dots, n_k$, let $T_{k|j} = t(\mathbb{X}_{k|j})$ denote the value of t on the sample $\mathbb{X}_{k|j}$ which is obtained from $\mathbb{X}$ by replacing the observation $X_{k\,j}$ by $X_{k\,n_k+1}$. Furthermore, put $\mathbb{X}_{k|n_k+1} = \mathbb{X}$ and write $T_{k|n_k+1} = t(\mathbb{X})$. The jackknife variance estimator is defined by

$$v^2(T) = \sum_{k=1}^{h} (1 - f_k) \sum_{j=1}^{n_k+1} \big(\overline{T}_k - T_{k|j}\big)^2, \qquad f_k = n_k/N_k, \tag{1.1}$$

where $\overline{T}_k = \dfrac{1}{n_k + 1} \sum_{j=1}^{n_k+1} T_{k|j}$.

Note that formula (1.1) involves values of T based on samples of the same size as $\mathbb{X}$. This allows us to consider a general class of kernels t but requires one additional observation from every stratum. For i.i.d. samples, jackknife estimators involving few (one or two) additional observations were considered by Beran (1984) and Putter and van Zwet (1998).

Let us outline the content of the paper. In Section 2, we introduce an orthogonal decomposition for STSI samples. In Section 3, we give an exact formula for the bias of $v^2(T)$ in terms of the decomposition components and derive the Efron–Stein inequality

$$\mathbf{E}\, v^2(T) \geq \sigma_T^2, \tag{1.2}$$

which shows that the jackknife variance estimator tends to be biased upwards. Note that the equality in (1.2) holds if and only if the statistic T is linear (= a statistic of the form $T = \sum_k \sum_i \psi_k(X_{k\,i})$). Section 4 contains large sample asymptotic results: the consistency of $v^2(T)$ and asymptotic normality of $(T - \mathbf{E}\, T)v^{-1}(T)$, where the use of the orthogonal decomposition allows us to treat a very wide class of statistics. The general results of the paper are illustrated by an example of U-statistic.

2. Orthogonal decomposition

The orthogonal decomposition of statistics based on stratified samples drawn *with replacement* was introduced by Lehman (1951, 1963) and Dwass (1956). Here we consider stratified samples drawn *without replacement*. In this case, calculations become much more complex. For convenience, we start with a simple example of a U-statistic.

2.1. *Example.* Assume that $H : \mathcal{X} \times \mathcal{X} \to \mathcal{R}$ is a symmetric function (i.e., $H(x_i, x_j) = H(x_j, x_i)$). In order to estimate the mean value

$$a = \frac{1}{\binom{N}{2}} \sum_{1 \leq i < j \leq N} H(x_i, x_j),$$

we use the STSI estimator

$$\begin{aligned} \hat{a} &= \sum_{1 \leq k \leq r \leq h} w_{k\,r} \hat{a}_{k\,r}, \\ \hat{a}_{k\,k} &= \frac{1}{\binom{n_k}{2}} \sum_{1 \leq i < j \leq n_k} H(X_{k\,i}, X_{k\,j}), \\ \hat{a}_{k\,r} &= \frac{1}{n_k n_r} \sum_{i=1}^{n_k} \sum_{j=1}^{n_r} H(X_{k\,i}, X_{r\,j}), \quad k < r, \end{aligned} \tag{2.1}$$

with weights $w_{k\,k} = \binom{N_k}{2}\binom{N}{2}^{-1}$ and $w_{k\,r} = N_k N_r \binom{N}{2}^{-1}$ for $k \neq r$.

It is easy to see that $\hat{a}$ is an unbiased estimator of the parameter a. Clearly, $\hat{a}$ is a U-statistic. We decompose

$$\hat{a} = a + L + Q, \tag{2.2}$$

where the linear part

$$L = \sum_{k=1}^{h} \sum_{i=1}^{n_k} g_k(X_{k\,i})$$

and the quadratic part

$$Q = \sum_{1 \le k \le r \le h} Q_{k\,r},$$

$$Q_{k\,k} = \sum_{1 \le i < j \le n_k} g_{k\,k}(X_{k\,i}, X_{k\,j}), \quad Q_{k\,r} = \sum_{i=1}^{n_k} \sum_{j=1}^{n_r} g_{k\,r}(X_{k\,i}, X_{r\,j}), \quad k < r,$$

are uncorrelated. The kernels g are defined as follows. For $k \ne r$, $x, y \in \mathcal{X}_k$ and $z \in \mathcal{X}_r$, put

(2.3)

$$\begin{aligned}
g_k(x) &= \frac{w_{k\,k}}{\binom{n_k}{2}} \frac{N_k - 1}{N_k - 2} (n_k - 1) s_{k\,k}(x) + \sum_{r=1}^{h} \mathbb{I}_{\{r \ne k\}} \frac{w_{k\,r}}{n_k} s_{k\,r}(x), \\
g_{k\,k}(x, y) &= \frac{w_{k\,k}}{\binom{n_k}{2}} \Big(h_{k\,k}(x, y) - \frac{N_k - 1}{N_k - 2} \left(s_{k\,k}(x) + s_{k\,k}(y) \right) \Big), \\
g_{k\,r}(x, z) &= \frac{w_{k\,r}}{n_k n_r} \Big(h_{k\,r}(x, z) - s_{k\,r}(x) - s_{r\,k}(z) \Big).
\end{aligned}$$

Here we denote

$$\begin{aligned}
h_{kk}(x, y) &= H(x, y) - \mathbf{E}\, H(X_{k1}, X_{k2}), \\
h_{kr}(x, z) &= H(x, z) - \mathbf{E}\, H(X_{k1}, X_{r1}), \\
s_{kk}(x) &= \mathbf{E}\, \big(h_{kk}(X_{k1}, X_{k2}) \big| X_{k1} = x \big), \\
s_{kr}(x) &= \mathbf{E}\, \big(h_{kr}(X_{k1}, X_{r1}) \big| X_{k1} = x \big).
\end{aligned}$$

Note that the random variables

$$g_k(X_{k\,i}), \qquad g_{k\,k}(X_{k\,i}, X_{k\,j}), \qquad \text{and} \qquad g_{k\,r}(X_{k\,i}, X_{r\,m})$$

are centered. A straightforward calculation shows that identity (2.2) holds. Moreover, the random variables $g_k(X_{k\,i})$ from the linear part are *uncorrelated* with the random variables $g_{p\,q}(X_{p\,j}, X_{q\,m})$, $p \le q$, from the quadratic part for $k, p, q \in \{1, \dots, h\}$, $1 \le i \le n_k$, $1 \le j \le n_p$, and $1 \le m \le n_q$ (for $p = q$, we require $j \ne m$). In particular, the linear part L and the quadratic part Q in (2.2) are uncorrelated. We call (2.2) the orthogonal decomposition. One can show that L provides the best (in the sense of quadratic mean) approximation to $\hat{a}$ by a linear statistic. This implies the uniqueness of the orthogonal decomposition (2.2).

In order to introduce the orthogonal decomposition of a general symmetric statistic T, we need some additional notation.

2.2. *Notation.* Given k, let $(X_{k\,1}, \dots, X_{k\,N_k})$ be a random permutation of the ordered set of elements of $\mathcal{X}_k$. We assume that random permutations are independent across the strata so that, for every k, the first n_k elements of $(X_{k\,1}, \dots, X_{k\,N_k})$ represent a subsample $\mathbb{X}_k$.

By $\overline{a} = (a_1, \dots, a_h)$ we denote h-vectors with nonnegative integer coordinates and put $|\overline{a}| = a_1 + \dots + a_h$. We write $\overline{b} \le \overline{a}$ if $b_k \le a_k$ for every $k = 1, \dots, h$, and $\overline{b} < \overline{a}$ if, in addition, $\overline{b} \ne \overline{a}$. Introduce the vector

$$\overline{n}^* = (n_1^*, \dots, n_h^*), \qquad \text{where} \qquad n_k^* = \min\{n_k,\, N_k - n_k\}.$$

Furthermore, $\overline{0} = (0, \dots, 0)$, and $\overline{e}_k$ denotes the h-vector with all coordinates zero, except the kth coordinate 1.

Given an index vector $\overline{a} = (a_1, \dots, a_h)$ satisfying $\overline{a} \le \overline{n}^*$, consider a symmetric kernel $g_{\overline{a}}$ defined on h-tuples of subsets $(\mathcal{A}_1, \dots, \mathcal{A}_h)$, where

$$\mathcal{A}_1 = \{x_{1\,i_1}, \dots, x_{1\,i_{a_1}}\} \subset \mathcal{X}_1, \quad \dots, \quad \mathcal{A}_h = \{x_{h\,j_1}, \dots, x_{h\,j_{a_h}}\} \subset \mathcal{X}_h$$

are subsets of sizes $a_1, \dots, a_h$, respectively. Furthermore, for a h-tuple of subsets $(\mathbb{A}_1, \dots, \mathbb{A}_h)$ of elements of the random permutations, say

$$\mathbb{A}_1 = \{X_{1\,i_1}, \dots, X_{1\,i_{a_1}}\}, \quad \dots, \quad \mathbb{A}_h = \{X_{h\,j_1}, \dots, X_{h\,j_{a_h}}\},$$

the kernel $g_{\overline{a}}$ defines the random variable $g_{\overline{a}}(\mathbb{A}_1, \dots, \mathbb{A}_h)$. Finally, the kernel $g_{\overline{a}}$ defines the U-statistic

$$U_{\overline{a}} = \sum_{\mathbb{A}_1 \subset \mathbb{X}_1,\, |\mathbb{A}_1| = a_1} \cdots \sum_{\mathbb{A}_h \subset \mathbb{X}_h,\, |\mathbb{A}_h| = a_h} g_{\overline{a}}\big(\mathbb{A}_1, \dots, \mathbb{A}_h\big). \tag{2.4}$$

2.3. *Decomposition.* Orthogonal decomposition expands T into the sum of centered U-statistics of increasing order

$$\begin{aligned} T &= \mathbf{E}\,T + \sum_{\overline{a}:\, |\overline{a}|=1} U_{\overline{a}} + \sum_{\overline{a}:\, |\overline{a}|=2} U_{\overline{a}} + \dots \\ &= \mathbf{E}\,T + \sum_{\overline{0} < \overline{a} \le \overline{n}^*} U_{\overline{a}}\ . \end{aligned} \tag{2.5}$$

For a linear statistic T, we have $U_{\overline{a}} \equiv 0$ for $|\overline{a}| > 1$. Similarly, if T is a U-statistic of degree two, we have $U_{\overline{a}} \equiv 0$ for $|\overline{a}| > 2$ (cf. (2.2)).

The crucial property of decomposition (2.5) is that the contributing U statistics are mutually uncorrelated, that is, $\mathbf{E}\, U_{\overline{a}} U_{\overline{b}} = 0$ for $\overline{a} \neq \overline{b}$.

The kernels $g_{\overline{a}}$ defining $U_{\overline{a}}$ by means of (2.4) are constructed in Appendix. They are linear combinations of conditional expectations of T. Here we only mention that every random variable $g_{\overline{a}}(\mathbb{A}_1, \dots, \mathbb{A}_h)$ is centered and

$$\mathbf{E}\, g_{\overline{a}}(\mathbb{A}_1, \dots, \mathbb{A}_h) g_{\overline{b}}(\mathbb{B}_1, \dots, \mathbb{B}_h) = 0, \tag{2.6}$$

for every $\overline{b} \neq \overline{a}$ and every h-tuple $(\mathbb{B}_1, \dots, \mathbb{B}_h)$, where $\mathbb{B}_k \subset \{X_{k1}, \dots, X_{kN_k}\}$ is a subset of size b_k, $k = 1, \dots, h$. Note that (2.6) implies the orthogonality property $\mathbf{E}\, U_{\overline{a}} U_{\overline{b}} = 0$ for $\overline{a} \neq \overline{b}$. Identity (2.6) is a consequence of formula (5.3) in Appendix.

A straightforward consequence of decomposition (2.5) is the variance formula

$$\sigma_T^2 = \sum_{\overline{a} \le \overline{n}^*} \mathbf{Var}\, U_{\overline{a}} = \sum_{\overline{a} \le \overline{n}^*} c(\overline{a}) \sigma_{\overline{a}}^2, \qquad c(\overline{a}) = \prod_{k=1}^{h} \frac{\binom{n_k}{a_k}\binom{N_k - n_k}{a_k}}{\binom{N_k - a_k}{a_k}}, \tag{2.7}$$

where

$$\sigma_{\overline{a}}^2 = \mathbf{Var}\, g_{\overline{a}}\big(\{X_{1\,1}, \dots, X_{1\,a_1}\}, \dots, \{X_{h\,1}, \dots, X_{h\,a_h}\}\big),$$

and we write $\sigma_{\overline{0}}^2 = 0$. Identity (2.7) follows from (2.5) by orthogonality and from the following formula shown in Appendix:

$$\mathbf{Var}\, U_{\overline{a}} = c(\overline{a}) \sigma_{\overline{a}}^2. \tag{2.8}$$

Remark. The sum of the linear and quadratic terms

$$U = \mathbf{E}\, T + \sum_{\overline{a}\colon |\overline{a}|=1} U_{\overline{a}} + \sum_{\overline{a}\colon |\overline{a}|=2} U_{\overline{a}}$$

often provides a sufficiently precise approximation to T. Bentkus, Götze, and van Zwet (1997) used this fact to construct second order approximations for a general symmetric statistics based on i.i.d. samples. The case of simple random samples was considered by Bloznelis and Götze (2001). Furthermore, for i.i.d. samples, Putter and van Zwet (1998) used the approximation $T \approx U$ to construct an empirical Edgeworth expansion which provides a second order approximation to the distribution of T in probability. Their results were extended to simple random samples by Bloznelis (2001a). In view of the importance of STSI samples for practical inference (Cochran (1977) and

Särndal, Swensson, and Wretman (1997)), it is desirable to develop similar approximations in this case as well.

We complete this section by providing explicit formulas for the kernels $g_{\overline{a}}$, $|\overline{a}| \le 2$ (defining U), which are obtained from a general result (5.2) below,

(2.9)
$$\begin{aligned}
g_{\overline{e}_k}(x) &= \frac{N_k-1}{N_k-n_k}\,\varphi_{\overline{e}_k}(x),\\
g_{2\overline{e}_k}(x,y) &= \frac{N_k-2}{N_k-n_k}\,\frac{N_k-3}{N_k-n_k-1}\Big(\varphi_{2\overline{e}_k}(x,y)-\frac{N_k-1}{N_k-2}\,(\varphi_{\overline{e}_k}(x)+\varphi_{\overline{e}_k}(y))\Big),\\
g_{\overline{e}_k+\overline{e}_r}(x,z) &= \frac{N_k-1}{N_k-n_k}\,\frac{N_r-1}{N_r-n_r}\Big(\varphi_{\overline{e}_k+\overline{e}_r}(x,z)-\varphi_{\overline{e}_k}(x)-\varphi_{\overline{e}_r}(z)\Big),
\end{aligned}$$

where $x, y \in \mathcal{X}_k$, $z \in \mathcal{X}_r$, and

$$\begin{aligned}
\varphi_{\overline{e}_k}(x) &= \mathbf{E}\,\big(T-\mathbf{E}\,T\big|\,X_{k\,1}=x\big),\\
\varphi_{2\overline{e}_k}(x,y) &= \mathbf{E}\,\big(T-\mathbf{E}\,T\big|\,X_{k\,1}=x,\ X_{k,2}=y\big),\\
\varphi_{\overline{e}_k+\overline{e}_r}(x,z) &= \mathbf{E}\,\big(T-\mathbf{E}\,T\big|\,X_{k\,1}=x,\ X_{r\,1}=z\big).
\end{aligned}$$

Note that decomposition (2.5) applied to statistic (2.1) results in decomposition (2.2). Indeed, comparison of formulas (2.9) and (2.3) shows that $g_{\overline{e}_k} = g_k$ and $g_{\overline{e}_k+\overline{e}_r} = g_{kr}$ for $1 \le k \le r \le h$.

3. Bias

3.1. The main result of this section, Theorem 3.1, provides an exact formula for the bias of the jackknife variance estimator (1.1).

Theorem 3.1. *We have*

(3.1)
$$\mathbf{E}\,v^2(T) = \sigma_T^2 + \sum_{\overline{a}\le\overline{n}^*,\,|\overline{a}|\ge 2} (d(\overline{a})-1)\,c(\overline{a})\,\sigma_{\overline{a}}^2,$$

where $d(\overline{a}) = \sum_{k=1}^h d_k(\overline{a})$ *and* $d_k(\overline{a}) = a_k(N_k+1-a_k)/N_k$. *The coefficients* $c(\overline{a})$ *are defined by (2.7).*

Remark 3.1. It is easy to see that $d(\overline{a}) > 1$ for $|\overline{a}| \ge 2$. Hence, the sum $\sum(\dots)$ on the right side of (3.1) is always nonnegative. Therefore, (3.1) implies the Efron–Stein inequality (1.2). Furthermore, the sum on the right side in (3.1) equals zero if and only if $\sigma_{\overline{a}}^2 = 0$ for $|\overline{a}| \ge 2$. That is, $\mathbf{E}\,v^2(T) = \sigma_T^2$ if and only if T is a linear statistic.

Remark 3.2. Let us look at identity (3.1) in the case of simple random samples ($h = 1$). Let $\mathbb{X} = \{X_1, \dots, X_n\}$ be a simple random sample drawn without replacement from $\mathcal{X} = \{x_1, \dots, x_N\}$. Decomposition (2.5) reduces to

$$T = t(X_1, \dots, X_n) = \mathbf{E}\, T + \sum_{j=1}^{n^*} \sum_{\mathbb{A} \in \mathbb{X},\, |\mathbb{A}|=j} g_j(\mathbb{A}),$$

where $n^* = \min\{n,\, N-n\}$. Similarly, identity (3.1) reduces to

$$\mathbf{E}\, v^2(T) = \sigma_T^2 + \sum_{j=2}^{n^*} d(j)\, \frac{\binom{n}{j}\binom{N-n}{j}}{\binom{N-j}{j}}\, \sigma_j^2, \qquad d(j) = (j-1)\, \frac{N-j}{N}\,,$$

where we denote $\sigma_j^2 = \mathbf{E}\, g_j^2(X_1, \dots, X_j)$.

Example (continued). It follows from (3.1) that

$$\mathbf{E}\, v^2(\hat{a}) = \sigma_{\hat{a}}^2 + \sum_{k=1}^{h} \frac{N_k - 2}{N_k}\, \frac{\binom{n_k}{2}\binom{N_k-n_k}{2}}{\binom{N_k-2}{2}}\, \sigma_{kk}^2$$
$$+ \sum_{1 \le k < r \le h} \frac{n_k(N_k - n_k)}{N_k - 1}\, \frac{n_r(N_r - n_r)}{N_r - 1}\, \sigma_{kr}^2,$$

where $\sigma_{k\,r}^2 = \mathbf{E}\, g_{k\,r}^2(X_{k\,1}, X_{r\,2})$.

3.2. *Proof of Theorem 3.1* We can assume without loss of generality that $\mathbf{E}\, T = 0$. Denote, for brevity, $v_k^2(T) = \sum_{j=1}^{n_k+1} (\overline{T}_k - T_{k|j})^2$.

In order to prove (3.1) it suffices to show that, for every k,

$$(1 - f_k)\mathbf{E}\, v_k^2(T) = \sum_{\overline{a} \le \overline{n}^*} d_k(\overline{a}) c(\overline{a}) \sigma_{\overline{a}}^2. \tag{3.2}$$

Indeed, since $\sigma_{\overline{0}}^2 = 0$ and $d(\overline{a}) = 0$ for $|\overline{a}| = 1$, identity (3.2) and (2.7) imply (3.1).

Let us show (3.2). Write $v_k^2(T)$ in the form

$$\sum_{j=1}^{n_k+1} T_{k|j}^2 - (n_k + 1)^{-1} H_k^2, \qquad H_k = T_{k|1} + \cdots + T_{k|n_k+1}.$$

By symmetry we have

$$\mathbf{E}\, v_k^2(T) = (n_k + 1)\mathbf{E}\, T_{k|n_k+1}^2 - (n_k + 1)^{-1}\mathbf{E}\, H_k^2. \tag{3.3}$$

In order to evaluate $\mathbf{E}\, H_k^2$, we expand $T_{k|j}$ by means of (2.5) for every $j = 1, \dots, n_k + 1$ and obtain

$$H_k = \sum_{\overline{a} \le \overline{n}^*} (n_k + 1 - a_k) U_{\overline{a}|k},$$

where $U_{\overline{a}|k}$ is defined in the same way as $U_{\overline{a}}$ but with the kth sum $\sum_{\mathbb{A}_k\subset\mathbb{X}_k,\,|\mathbb{A}_k|=a_k}$ replaced by $\sum_{\mathbb{A}_k\subset\mathbb{X}'_k,\,|\mathbb{A}_k|=a_k}$ in (2.4). Since, by (2.6), the random variables $U_{\overline{a}|k}$, $\overline{a}\le\overline{n}^*$, are uncorrelated, we have

$$\mathbf{E}\,H_k^2=\sum_{\overline{a}\le\overline{n}^*}(n_k+1-a_k)^2\mathbf{E}\,U_{\overline{a}|k}^2.$$

Furthermore, by (2.8), $\mathbf{E}\,U_{\overline{a}|k}^2=c_k(\overline{a})\sigma_{\overline{a}}^2$, where $c_k(\overline{a})$ is defined in the same way as $c(\overline{a})$ but with n_k replaced by n_k+1. Finally, invoking, in (3.3), the formula for $\mathbf{E}\,H_k^2$ and the formula (2.7) for $\mathbf{E}\,T_{k|n_k+1}^2=\mathbf{Var}\,T$, we obtain (3.2).

4. Consistency

4.1. Consider a sequence of stratified populations $\mathcal{X}(\nu)=\mathcal{X}_1(\nu)\cup\cdots\cup\mathcal{X}_h(\nu)$, where $\nu=1,2,\ldots$. We assume that the total sample size $n=n_1+\cdots+n_h$ tends to ∞ in any way, e.g., many small samples, or a few large samples, or some combination thereof. No condition is imposed on the number of strata h (it can be bounded or increase) and on the strata sizes N_k. More precisely, we suppose that h, N_k, n_k, and the kernel t all depend on an index ν such that $n(\nu)=n_1(\nu)+\cdots+n_h(\nu)\to\infty$ as $\nu\to\infty$. The index ν will be suppressed in the sequel. We say that the estimator $v^2(T)$ is consistent if $v^2(T)/\sigma_T^2\to1$ in probability. Here and in what follows, limits are taken as $\nu\to\infty$.

Write $T=\mathbf{E}\,T+L+R$, where

$$L=\sum_{\overline{a}\colon|\overline{a}|=1}U_{\overline{a}}=\sum_{k=1}^{h}\sum_{i=1}^{n_h}g_{\overline{e}_k}(X_{ki})$$

denotes the linear part of decomposition (2.5), and $R=\sum_{\overline{a}\colon|\overline{a}|\ge2}U_{\overline{a}}$ denotes the remainder. We consider statistics with dominant linear part as $\nu\to\infty$. More precisely, we assume that

$$\sigma_L^2/\sigma_T^2\to1,\qquad\text{where}\qquad\sigma_L^2=\mathbf{Var}\,L. \tag{4.1}$$

Note that, in the case where (4.1) is violated, the estimator $v^2(T)$ can be inconsistent. This can be seen, e.g., from formula (3.1). Indeed, the jackknife variance estimator such as (1.1) targets the linear part L and estimates its variance σ_L^2 consistently under the very mild Lindeberg-type condition (4.3) (see Bickel and Freedman (1984)). However, such estimator becomes inconsistent when applied to the nonlinear parts of decomposition (2.5). Therefore, we need $\sigma_R^2/\sigma_T^2\to0$ in order to make the influence of the nonlinear part R negligible. But this is equivalent to (4.1), since

$$\sigma_L^2+\sigma_R^2=\sigma_T^2 \tag{4.2}$$

by the fact that L and R are uncorrelated.

Theorem 4.1. *Assume that*

$$\forall \varepsilon > 0, \qquad \sigma_L^{-2} \sum_{k=1}^{h} n_k^* \mathbf{E}\, g_{\overline{e}_k}^2(X_{k\,1}) \mathbb{I}_{\{g_{\overline{e}_k}^2(X_{k\,1}) \geq \varepsilon \sigma_L^2\}} \to 0. \tag{4.3}$$

Here $n_k^* = \min\{N_k,\, N_k - n_k\}$. *Suppose that*

$$\Delta^2/\sigma_L^2 \to 0, \qquad \text{where} \qquad \Delta^2 = \sum_{\overline{a} \leq \overline{n}^*,\, |\overline{a}| \geq 2} d(\overline{a}) c(\overline{a}) \sigma_{\overline{a}}^2, \tag{4.4}$$

where the numbers $d(\overline{a})$ *are the same as in Theorem 3.1. Then*

i) $\quad v^2(T)/\sigma_T^2 \to 1 \quad$ *in probability*

and

ii) $\quad (T - \mathbf{E}\, T) v^{-1}(T) \to N(0,1) \quad$ *in law.*

Theorem 4.1 provides an extension of a result of Bickel and Freedman (1984) to nonlinear symmetric statistics. Here, in addition to the Lindeberg-type condition (4.3), we impose condition (4.4) which controls the nonlinear part R and is a bit stronger than (4.1). In order to inspect the optimality of (4.4), we compare (3.1) and (4.4). Firstly, note that (4.4) is equivalent to $\Delta^2/\sigma_T^2 \to 0$.

A simple calculation shows that $d(\overline{a}) - 1 \geq d(\overline{a})/4$ for $|\overline{a}| \geq 2$. Therefore, $d(\overline{a})/4 \leq d(\overline{a}) - 1 \leq d(\overline{a})$. This and the identity (see (2.7)) $\sigma_R^2 = \sum_{\overline{a} \leq \overline{n}^*,\, |\overline{a}| \geq 2} c(\overline{a}) \sigma_{\overline{a}}^2$ imply that

$$\Delta^2/4 \leq \Delta^2 - \sigma_R^2 \leq \Delta^2 \qquad \text{and} \qquad \sigma_R^2 \leq 3\Delta^2/4. \tag{4.5}$$

It follows from (4.2) and (4.5) that

$$\Delta^2/\sigma_L^2 \geq (\Delta^2 - \sigma_R^2)/(\sigma_L^2 + \sigma_R^2) \geq 4^{-1}\Delta^2/(\sigma_L^2 + 3\Delta^2/4).$$

Finally, identity (3.1) written in the form $\mathbf{E}\, v^2(T) - \sigma_T^2 = \Delta^2 - \sigma_R^2$ implies that

$$\frac{\Delta^2}{\sigma_L^2} \geq \mathbf{E}\Big(\frac{v^2(T)}{\sigma_T^2} - 1\Big) \geq \frac{1}{4}\, \frac{1}{\sigma_L^2/\Delta^2 + 3/4}\,. \tag{4.6}$$

We obtain the following corollary.

Corollary 4.2. *Condition (4.4) is necessary and sufficient for the asymptotic unbiasedness of* $v^2(T)/\sigma_T^2$. *Similarly, the condition* $\Delta^2 = o(1)$ *is necessary and sufficient for* $\mathbf{E}\, v^2(T) - \sigma_T^2 = o(1)$.

The corollary follows from (4.5) and (4.6).

Remark 4.1. For simple random samples, using the notation of Remark 3.2, we have $\sigma_L^2 = \sigma_1^2 n(N-n)/(N-1)$ and

$$\Delta^2 = n \sum_{j=2}^{n^*} c_j \sigma_j^2, \qquad c_j = \frac{N+1-j}{N} \frac{\binom{n-1}{j-1}\binom{N-n}{j}}{\binom{N-j}{j}}.$$

Therefore, (4.4) reduces to $\sigma_1^{-2} \dfrac{N-1}{N-n} \sum_{j=2}^{n^*} c_j \sigma_j^2 = o(1)$. Furthermore, letting $N \to \infty$ and keeping the sample size n fixed, we approach the i.i.d. situation. In this case, we have $c_j = \binom{n-1}{j-1}$, and (4.4) reduces to $\sigma_1^{-2} \sum_{j=2}^{n} \binom{n-1}{j-1} \sigma_j^2 = o(1)$ as $n \to \infty$. This condition appears in van Zwet (1990), who considered the i.i.d. case. The consistency of the jackknife variance estimator of U-statistics, based on simple random samples, was studied by Krewski (1978).

Example (continued). For the U-statistic $\hat{a}$ from (2.1) we have $\sigma_R^2 \le \Delta^2 \le 2\sigma_R^2$ since $d(\overline{a}) \le 2$ for $|\overline{a}| \le 2$. Therefore, (4.4) is equivalent to $\sigma_R^2/\sigma_L^2 \to 0$, and the latter is equivalent to (4.1), which is a very mild condition.

Although condition (4.4) is precise and suitable for U-statistics, it can be hardly verifiable for other classes of statistics. Van Zwet (1984) and Bloznelis and Götze (2001) introduce a kind of smoothness conditions which allow one to control the sums like (4.4). In Corollary 4.3 below, we replace (4.4) by a more restrictive but simpler smoothness condition (4.7).

Before formulating the corollary, we introduce some notation. Given $k = 1, \dots, h$ and $j = 1, 2$, define the 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 $\mathbb{X}_k^j = (\mathbb{X}_k \setminus \{X_{k,j}\}) \cup \{X_{k,n_k+j}\}$. Thus, in order to define $\delta_{k|j}$, we need $n_k + j$ observations drawn without replacement from $\mathcal{X}_k$. Application of the difference operation $\delta_{k|2}$ to the statistic $\delta_{k|1}T$ results in random variable $\delta_{k|2}\delta_{k|1}T$. For $i = 1, 2$, write $\mathbb{D}_{k|i} = \delta_{k|i}\delta_{k|i-1} \dots \delta_{k|0}$, where $\delta_{k|0}T = T$. Finally, given $\overline{a} \le \overline{n}^*$, denote $\mathbb{D}_{\overline{a}} = \mathbb{D}_{h|a_h} \dots \mathbb{D}_{1|a_1}$. Write $n_{\overline{a}} = (n_1^*)^{a_1} \dots (n_h^*)^{a_h}$ and denote

$$\tilde{\Delta}^2 = \sum_{\overline{a}\colon |\overline{a}|=2} n_{\overline{a}} \mathbf{E}\, (\mathbb{D}_{\overline{a}} T)^2.$$

Corollary 4.3. *We have $\Delta^2 \le \tilde{\Delta}^2$. Theorem 4.1 remains valid with (4.4) replaced by*

$$\tilde{\Delta}^2/\sigma_L^2 \to 0. \tag{4.7}$$

Condition (4.7) applies easily to smooth functions of linear statistics and smooth functions of U-statistics. In particular, for $\mathcal{X} \subset R^d$ and $T = g(X_{1\,\cdot}, \dots, X_{h\,\cdot})$ (here, the dot is the averaging operator), the second-order continuous differentiability of g in a neighborhood of the point $(\mathbf{E}\, X_{1\,1}, \dots, \mathbf{E}\, X_{h\,1})$ is sufficient to verify (4.7).

4.2. Here we prove Theorem 4.1 and Corollary 4.3. Without loss of generality, we assume that $\mathbf{E}\, T = 0$. For $r = 1, 2, \dots,$ write $\Omega_r = \{1, \dots, r\}$.

Proof of Theorem 4.1. Let us prove i). Given $k \in \Omega_h$ and $j \in \Omega_{n_k+1}$, write $\overline{T}_k - T_{k|j}$ in the form

$$\overline{T}_k - T_{k|j} = \sum_{i=1}^{n_k^*} V_{k|j}(i), \qquad V_{k|j}(i) = \sum_{\mathbb{A}\subset\mathbb{X}_k',\, |\mathbb{A}|=i} \big(\mathbb{I}_{X_{k\,j}\in\mathbb{A}} - \frac{i}{n_k+1}\big) T_k(\mathbb{A}),$$

where, given $\mathcal{A} \subset \mathcal{X}_k$ with $|\mathcal{A}| \le n_k^*$, we denote

$$T_k(\mathcal{A}) = \sum_{\overline{a}\le\overline{n}^*,\, a_k=|\mathcal{A}|} \ \sum_{\overline{a}|k} g_{\overline{a}}(\mathbb{A}_1, \dots, \mathbb{A}_{k-1}, \mathcal{A}, \mathbb{A}_{k+1}, \dots, \mathbb{A}_h). \tag{4.8}$$

Here the first sum runs over all $\overline{a} \le \overline{n}^*$ satisfying $a_k = |\mathcal{A}|$. Furthermore, by $\sum_{\overline{a}|k}$ we denote the multiple sum as in (2.4) but with the kth sum omitted. Split $V_{k|j}(1) = s_{k|j} + t_{k|j}$, where

$$s_{k|j} = \sum_{x\in\mathbb{X}_k'} \big(\mathbb{I}_{X_{k\,j}=x} - \frac{1}{n_k+1}\big) g_{\overline{e}_k}(x),$$

and where $t_{k|j} = V_{k|j}(1) - s_{k|j}$ denotes the remainder. Write

$$\overline{T}_k - T_{k|j} = s_{k|j} + r_{k|j}, \qquad r_{k|j} = t_{k|j} + \sum_{i=2}^{n_k^*} V_{k|j}(i).$$

Finally, from (1.1) we obtain that $v^2(T) = S^2 + Q^2 + 2Z$, where

$$S^2 = \sum_{k=1}^{h} (1-f_k) S_k^2, \qquad S_k^2 = \sum_{j=1}^{n_k+1} s_{k|j}^2,$$

$$Q^2 = \sum_{k=1}^{h} (1-f_k) Q_k^2, \qquad Q_k^2 = \sum_{j=1}^{n_k+1} r_{k|j}^2,$$

$$Z = \sum_{k=1}^{h} (1-f_k) \sum_{j=1}^{n_k+1} s_{k|j} r_{k|j}.$$

Note that $S^2 = v^2(L)$. In view of Theorem 3 of Bickel and Freedman (1984), (4.3) implies that $S^2/\sigma_L^2 \to 1$ in probability. Furthermore, since (4.4) implies (4.1), we obtain $S^2/\sigma_T^2 \to 1$ in probability. It remains to show that Q^2 and Z can be neglected. We have $|Z| \le S\,Q$ by Cauchy–Schwartz. Therefore, it suffices to show that $\mathbf{E}\,Q^2/\sigma_T^2 \to 0$. For this purpose, we show that $\mathbf{E}\,Q^2 = \Delta^2$ and apply (4.4).

Let us prove $\mathbf{E}\,Q^2 = \Delta^2$. By symmetry we have

$$\mathbf{E}\,Q^2 = \sum_k (1-f_k)(n_k+1)\mathbf{E}\,r_{k|n_k+1}^2. \tag{4.9}$$

Since $t_{k|n_k+1}$ and $V_{k|n_k+1}(i)$, $i = 2, \dots, n_k^*$, are uncorrelated, we have

$$\mathbf{E}\,r_{k|n_k+1}^2 = \mathbf{E}\,t_{k|n_k+1}^2 + \sum_{i=2}^{n_k^*} \mathbf{E}\,V_{k|n_k+1}^2(i). \tag{4.10}$$

Finally, combining (4.10) and the formulas

$$\mathbf{E}\,V_{k|n_k+1}^2(i) = d_k'(i)\varkappa_k^2(i), \quad i \ge 2, \tag{4.11}$$

$$\mathbf{E}\,t_{k|n_k+1}^2 = d_k'(1)\varkappa_k^2(1), \tag{4.12}$$

where

$$\varkappa_k^2(i) := \sum_{\overline{a}\le\overline{n}^*,\, a_k=i,\, |\overline{a}|\ge 2} c(\overline{a})\sigma_{\overline{a}}^2, \qquad d_k'(i) = \frac{i}{n_k+1}\;\frac{N_k-i+1}{N_k-n_k}\,,$$

from (4.9) we obtain the identity $\mathbf{E}\,Q^2 = \Delta^2$, thus, completing the proof of part *i*). The proof of (4.11) and (4.12) is technical and is given in Appendix.

Let us prove ii). The asymptotic normality of L/σ_L is shown by Bickel and Freedman (1984). In view of (4.1), we can replace σ_L and L by σ_T and T, thus, obtaining the asymptotic normality of T/σ_T. Indeed, $\mathbf{E}\,(T-L)^2/\sigma_T^2 = \sigma_R^2/\sigma_T^2 \to 0$. The statement i) completes the proof.

Proof of Corollary 4.3. Write $[x]_i = x(x-1)\dots(x-i+1)$ and denote

$$v_k(i,j) = 2^i\,\frac{[j]_i[N_k-j+1]_i}{[n_k]_i[N_k-n_k]_i}\,.$$

It follows from the identities

$$\mathbf{E}\,(\mathbb{D}_{\overline{e}_k+\overline{e}_r}T)^2 = \sum_{\overline{a}\le\overline{n}^*} v_k(1,a_k)v_r(1,a_r)c(\overline{a})\sigma_{\overline{a}}^2, \quad k \ne r, \tag{4.13}$$

and

$$\mathbf{E}\,(\mathbb{D}_{2\overline{e}_k}T)^2 = \sum_{\overline{a}\le\overline{n}^*} v_k(2,a_k)c(\overline{a})\sigma^2_{\overline{a}} \tag{4.14}$$

that

$$\sum_{1\le i\le j\le h} n_i^* n_j^* \mathbf{E}\,(\mathbb{D}_{\overline{e}_i+\overline{e}_j}T)^2 = \sum_{\overline{a}\le\overline{n}^*,|\overline{a}|\ge 2} (u(\overline{a})+v(\overline{a}))c(\overline{a})\sigma^2_{\overline{a}}. \tag{4.15}$$

Here we denote

$$v(\overline{a}) = \sum_{i=1}^h (n_i^*)^2 v_i(2,a_i), \qquad u(\overline{a}) = \sum_{i=1}^h u_i(\overline{a}),$$
$$u_i(\overline{a}) = \sum_{1\le j\le h,\, j\ne i} n_i^* n_j^* v_i(1,a_i)v_j(1,a_j).$$

The corollary follows from (4.15) and from the inequality

$$v(\overline{a}) + u(\overline{a}) \ge d(\overline{a}), \qquad \text{for} \quad |\overline{a}|\ge 2.$$

To prove this inequality note that, given $\overline{a}$, we have $d_k(\overline{a}) \le (n_k^*)^2 v_k(2,a_k)$ for $a_k\ge 2$. Finally, for $a_k = 1$, we have $d_k(\overline{a}) = 1 \le u_k(\overline{a})$, where the last inequality follows from $|\overline{a}|\ge 2$. Auxiliary identities (4.13) and (4.14) are shown in Appendix.

5. Appendix

Without loss of generality, we can assume that $\mathbf{E}\,T = 0$. Otherwise, the argument below applies to the statistic $T - \mathbf{E}\,T$.

5.1. Here we sketch the construction of decomposition (2.5). The details involving tedious combinatorial calculation are given in the accompanying paper of Bloznelis (2001b).

By $\varphi_{\overline{a}}$ and $\psi_{\overline{a}}$ we denote the real functions defined on h-tuples $(\mathcal{A}_1,\dots,\mathcal{A}_h)$ of subsets $\mathcal{A}_k\subset\mathcal{X}_k$ such that $|\mathcal{A}_k| = a_k$. Given $\overline{b}$ and an h-tuple $(\mathcal{B}_1,\dots,\mathcal{B}_h)$, where, for every k, $\mathcal{B}_k = \{x_{k\ i_k(1)},\dots,x_{k\ i_k(b_k)}\}\subset\mathcal{X}_k$ is a subset of size b_k, write

$$\varphi_{\overline{b}}(\mathcal{B}_1,\dots,\mathcal{B}_h) = \mathbf{E}\,\big(T\big|\, X_{k\,r} = x_{k\ i_k(r)},\ \ 1\le r\le b_k,\ \ 1\le k\le h\big).$$

Let $\psi_{\overline{a}}$ be the linear combination of conditional expectations $\varphi_{\overline{b}}$, $\overline{b}\le\overline{a}$, defined as follows. Put $\psi_{\overline{0}}\equiv 0$ and $\psi_{\overline{e}_k} = \varphi_{\overline{e}_k}$ for every k. The functions $\psi_{\overline{a}}$,

$|\overline{a}| > 1$, are defined by induction over increasing values of a_k, $k = 1, \ldots, h$. We put

$$(5.1)\quad \psi_{\overline{a}}(\mathcal{A}_1, \ldots, \mathcal{A}_h) = \varphi_{\overline{a}}(\mathcal{A}_1, \ldots, \mathcal{A}_h) - \sum_{\overline{b} < \overline{a}} \prod_{k=1}^{h} V_k(a_k, b_k) \sum_{|\mathcal{B}_1| = b_1} \cdots \sum_{|\mathcal{B}_h| = b_h} \psi_{\overline{b}}(\mathcal{B}_1, \ldots, \mathcal{B}_h),$$

where the sum $\sum_{|\mathcal{B}_k| = b_k}$ runs over all subsets $\mathcal{B}_k \subset \mathcal{A}_k$ of size b_k. Here

$$V_k(a_k, b_k) = \frac{\binom{N_k - b_k}{b_k}}{\binom{N_k - a_k}{b_k}} \qquad \text{for} \qquad a_k + b_k \le N_k.$$

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 (that is, $N_k = 2r_k + 1$ for some integer r_k), we put $V_k(r_k + 1, r_k + 1) = 1$.

The identity (5.1) applied to $\overline{a} = \overline{n}$ gives (2.5). Indeed, denoting

$$(5.2)\qquad g_{\overline{n}} = \psi_{\overline{n}} \qquad \text{and} \qquad g_{\overline{b}} = \psi_{\overline{b}} \prod_{k=1}^{h} V_k(n_k, b_k) \qquad \text{for} \quad \overline{b} < \overline{n},$$

from (5.1) we obtain that

$$t(\mathcal{A}_1, \ldots, \mathcal{A}_h) = \varphi_{\overline{n}}(\mathcal{A}_1, \ldots, \mathcal{A}_h) = \sum_{\overline{b} \le \overline{n}} \sum_{|\mathcal{B}_1| = b_1} \cdots \sum_{|\mathcal{B}_h| = b_h} g_{\overline{b}}(\mathcal{B}_1, \ldots, \mathcal{B}_h)$$

for an arbitrary h-tuple $(\mathcal{A}_1, \ldots, \mathcal{A}_h)$ of subsets $\mathcal{A}_k \subset \mathcal{X}_k$ such that $|\mathcal{A}_k| = n_k$. One can show that $g_{\overline{a}} \equiv 0$ for $\overline{a} \le \overline{n}$ that fail to satisfy the inequality $\overline{a} \le \overline{n}^*$. Therefore, the sum in (2.5) effectively extends over $\overline{a}$ satisfying $\overline{a} \le \overline{n}^*$ only. The most important property of the kernels $g_{\overline{a}}$ is their orthogonality: for all $\mathbb{A}_k$, $\mathbb{B}_k \subset \mathbb{X}_k$, where $|\mathbb{A}_k| = a_k$, $k \in \Omega_h$, we have

$$(5.3)\qquad \mathbf{E}\left(g_{\overline{a}}(\mathbb{A}_1, \ldots, \mathbb{A}_h) \,\big|\, \mathbb{B}_1, \ldots, \mathbb{B}_h\right) = 0 \qquad \text{almost surely}$$

provided that $|\mathbb{B}_j| < a_j$ for some j. Clearly, (5.3) implies (2.6).

5.2. Here we prove (2.8), (4.11), (4.12), (4.13), and (4.14).

For integers $s, t, u \ge 0$ such that $u \ge \max\{s; t\}$, the identity

$$(5.4)\qquad \sum_{\nu} \frac{\binom{s}{\nu}\binom{t}{\nu+i}}{\binom{u}{\nu+i}} (-1)^{\nu} = \frac{\binom{u-i-s}{t-i}}{\binom{u}{t}} \qquad i = 0, 1$$

is obtained from the formula (12.15) of Chapter 2 of Feller (1970) by replacing $\binom{t}{\nu+i}\binom{u}{\nu+i}^{-1}$ by $\binom{u-\nu-i}{u-t}\binom{u}{t}^{-1}$ on the left side. For $i = 0$, this identity has been used by Zhao and Chen (1990).

By $\mathbf{E}_k$ we denote the conditional expectation given $\{\mathbb{X}_1, \ldots, \mathbb{X}_{k-1}, \mathbb{X}_{k+1}, \ldots, \mathbb{X}_h\}$. Recall the notation $\Omega_r = \{1, \ldots, r\}$ for $r = 1, 2, \ldots$.

Proof of (2.8). Given $\overline{a} \le \overline{n}$, write $\mathbb{A}_k^0 = \{X_{k\,1}, \ldots, X_{k\,a_k}\}$ for $k \in \Omega_h$. Introduce the random variables

$$U_{[k]}(\overline{a}) = \sum_{\overline{a}||k} g_{\overline{a}}(\mathbb{A}_1^0, \ldots, \mathbb{A}_{k-1}^0, \mathbb{A}_k, \ldots, \mathbb{A}_h)$$

and write $U_{[h+1]}(\overline{a}) = g_{\overline{a}}(\mathbb{A}_1^0, \ldots, \mathbb{A}_h^0)$. Here by $\sum_{\overline{a}||k}$ we denote the multiple sum

$$\sum_{\mathbb{A}_k \subset \mathbb{X}_k,\, |\mathbb{A}_k| = a_k} \quad \sum_{\mathbb{A}_{k+1} \subset \mathbb{X}_{k+1},\, |\mathbb{A}_{k+1}| = a_{k+1}} \cdots \sum_{\mathbb{A}_h \subset \mathbb{X}_h,\, |\mathbb{A}_h| = a_h} .$$

In order to prove (2.8) it suffices to show that, for every $k \in \Omega_h$,

$$\mathbf{E}_k U_{[k]}^2(\overline{a}) = c_{[k]}(\overline{a}) \mathbf{E}_k U_{[k+1]}^2(\overline{a}), \qquad c_{[k]}(\overline{a}) = \frac{{n_k \atop a_k}\ {N_k - n_k \atop a_k}}{{N_k - a_k \atop a_k}} . \tag{5.5}$$

Indeed, (5.5) implies $\mathbf{E}\, U_{[k]}^2(\overline{a}) = c_{[k]}(\overline{a}) \mathbf{E}\, U_{[k+1]}^2(\overline{a})$, and, therefore,

$$\mathbf{E}\, U_{[1]}^2(\overline{a}) = c_{[1]}(\overline{a}) \mathbf{E}\, U_{[2]}^2(\overline{a}) = \cdots = c_{[1]}(\overline{a}) \ldots c_{[h]}(\overline{a}) \mathbf{E}\, U_{[h+1]}^2(\overline{a}).$$

Since $U_{\overline{a}} = U_{[1]}(\overline{a})$ and $\mathbf{E}\, U_{[h+1]}^2(\overline{a}) = \sigma_{\overline{a}}^2$, we obtain (2.8).

Let us prove (5.5). Write $U_{[k]}(\overline{a})$ in the form

$$U_{[k]}(\overline{a}) = \sum_{\mathbb{A}_k \subset \mathbb{X}_k,\, |\mathbb{A}_k| = a_k} G(\mathbb{A}_k), \tag{5.6}$$

where, for every $\mathcal{A}_k \subset \mathcal{X}_k$ such that $|\mathcal{A}_k| = a_k$, we write

$$G(\mathcal{A}_k) = \sum_{\overline{a}||k+1} g_{\overline{a}}(\mathbb{A}_1^0, \ldots, \mathbb{A}_{k-1}^0, \mathcal{A}_k, \mathbb{A}_{k+1} \ldots, \mathbb{A}_h).$$

From (5.3) it follows that, for $\mathbb{A}_k, \mathbb{B}_k \subset \mathbb{X}_k$ such that $|\mathbb{B}_k| < a_k$, we have

$$\mathbf{E}_k (G(\mathbb{A}_k) \,|\, \mathbb{B}_k) = 0 \qquad \text{almost surely.} \tag{5.7}$$

Using this identity it is easy to show (cf. Lemma 1 in Bloznelis and Götze (2001)) that for any $\mathbb{A}_k, \mathbb{A}_k' \subset \mathbb{X}_k$ such that $|\mathbb{A}_k| = |\mathbb{A}_k'| = a_k$ and $|\mathbb{A}_k \cap \mathbb{A}_k'| = j$, we have

$$\mathbf{E}_k G(\mathbb{A}_k) G(\mathbb{A}_k') = \frac{(-1)^{a_k - j}}{\binom{N_k - a_k}{a_k - j}} \mathbf{E}_k G^2(\mathbb{A}_k). \tag{5.8}$$

This identity with (5.6) and (5.4) gives

$$\mathbf{E}\, G(\mathbb{A}_k^0) U_{[k]}(\overline{a}) = \frac{\binom{N_k - n_k}{a_k}}{\binom{N_k - a_k}{a_k}} \mathbf{E}_k G^2(\mathbb{A}_k^0). \tag{5.9}$$

Note that $G(\mathbb{A}_k^0) = U_{[k+1]}(\overline{a})$. Finally, by symmetry from (5.9) it follows that

$$\mathbf{E}_k U_{[k]}^2(\overline{a}) = \binom{n_k}{a_k} \mathbf{E}\, G(\mathbb{A}_k^0) U_{[k]}(\overline{a}) = c_{[k]}(\overline{a}) \mathbf{E}_k U_{[k+1]}^2(\overline{a}).$$

Proof of (4.11) and (4.12). Let us prove (4.11). Introduce the random variables

$$H = \sum_{\mathbb{A} \subset \mathbb{X}_k,\, |\mathbb{A}| = i} T_k(\mathbb{A}), \qquad K = \sum_{\mathbb{B} \subset \mathbb{X}_k,\, |\mathbb{B}| = i-1} T_k(\mathbb{B}'),$$

where T_k is defined by (4.8) and $\mathbb{B}' = \mathbb{B} \cup \{X_{k,\, n_k+1}\}$. From the simple identity

$$V_{k|n_k+1}(i) = u\, K - v\, H, \qquad u = 1 - v, \quad v = \frac{i}{n_k + 1}, \tag{5.10}$$

it follows that $\mathbf{E}\, V_{k|n_k+1}^2(i) = u^2 \mathbf{E}\, K^2 - 2uv \mathbf{E}\, K\, H + v^2 \mathbf{E}\, H^2$. Finally, invoking the identities

$$\mathbf{E}\, K^2 = \frac{v}{u} \, \frac{N_k - 2i + 1}{N_k - n_k} \mathbf{E}\, H^2, \qquad \mathbf{E}\, K\, H = -\frac{i}{N_k - n_k} \mathbf{E}\, H^2, \tag{5.11}$$

and $\mathbf{E}\, H^2 = \varkappa_k^2(i)$, we obtain (4.11). The identity $\mathbf{E}\, H^2 = \varkappa_k^2(i)$ follows from (2.8) and the fact that random variables $U_{\overline{a}}$, $\overline{a} \le \overline{n}^*$, are uncorrelated.

Let us prove (5.11). For this purpose, we show (5.11) in the case where $\mathbf{E}$ is replaced by $\mathbf{E}_k$ (i.e., we show (5.11) for conditional expectations). Note that (5.3) implies (5.7) for the kernel T_k as well. Therefore, (5.8) remains valid with G replaced by T_k. That is,

$$\mathbf{E}_k T_k(\mathbb{A}_k) T_k(\mathbb{A}_k') = \frac{(-1)^{a_k - j}}{\binom{N_k - a_k}{a_k - j}} \mathbf{E}_k T^2(\mathbb{A}_k) \tag{5.12}$$

for any $\mathbb{A}_k, \mathbb{A}'_k \subset \mathbb{X}'_k$ such that $|\mathbb{A}_k| = |\mathbb{A}'_k| = a_k$ and $|\mathbb{A}_k \cap \mathbb{A}'_k| = j$.

In order to show the first identity of (5.11), we write, by symmetry, $\mathbf{E}_k K^2 = \binom{n_k}{i-1} \mathbf{E}_k T_k(\mathbb{A}^0_k) K$. In the right side we obtain a sum of conditional expectations like (5.12). Therefore, combining (5.12) and identity (5.4), we obtain the first identity of (5.11) for the conditional expectation $\mathbf{E}_k$. The proof of the second identity of (5.11) is similar.

Let us prove (4.12). Given $\mathcal{A} \subset \mathcal{X}_k$, define $T'_k(\mathcal{A})$ by formula (4.8), where the first sum runs over $\overline{a}$ that satisfy $\overline{a} \le \overline{n}^*$, $a_k = |\mathcal{A}|$ and, in addition, $|\overline{a}| \ge 2$. Since $t_{k|n_k+1}$ is obtained from $V_{k|n_k+1}(1)$ by excluding the sum $s_{k|n_k+1}$, we can write a representation of $t_{k|n_k+1}$ similar to that of $V_{k|n_k+1}(1)$ in (5.10) above. More precisely, we replace the sums K and H in (5.10) by similar sums with T_k is replaced by T'_k. The remaining part of the proof of (4.12) is similar to that of (4.11).

Proof of (4.13) and (4.14). We shall prove (4.14) only. Proof of (4.13) is similar. We show (4.14) for $k = 1$. For this purpose, we write (2.5) in the form

$$T = \sum_{j=0}^{n_1} \tilde{U}_j, \qquad \tilde{U}_j = \sum_{\mathbb{A} \subset \mathbb{X}_1,\, |\mathbb{A}| = j} T_1(\mathbb{A}),$$

where T_1 is given by (4.8). Since, for $j = 0, 1$, we have $\mathbb{D}_{2\overline{e}_1} \tilde{U}_j = 0$, from the identity $\mathbb{D}_{\overline{a}} T = \sum_j \mathbb{D}_{\overline{a}} \tilde{U}_j$ it follows, by orthogonality, that $\mathbf{E}_1 (\mathbb{D}_{2\overline{e}_1} T)^2 = \sum_{j=2}^{n_1} \mathbf{E}_1 (\mathbb{D}_{2\overline{e}_1} \tilde{U}_j)^2$. Finally, invoking the identity (see Lemma 2 of Bloznelis and Götze (2001)) $\mathbf{E}_1 (\mathbb{D}_{2\overline{e}_1} \tilde{U}_j)^2 = v_1(2, j) \mathbf{E}_1 \tilde{U}_j^2$, we obtain

$$\mathbf{E}\, (\mathbb{D}_{2\overline{e}_1} T)^2 = \sum_{j=2}^{n_1} v_1(2, j) \mathbf{E}\, \tilde{U}_j^2 = \sum_{\overline{a} \le \overline{n}^*,\, a_1 \ge 2} v_1(2, a_1) c(\overline{a}) \sigma^2_{\overline{a}}.$$

References

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

R. Beran (1984), *Jackknife approximations to bootstrap estimates*, Ann. Statist. **12**, 101–118.

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

M. Bloznelis (2001a), *Empirical Edgeworth expansion for finite population statistics I, II*, Lithuanian Math. Journal. **41**.

M. Bloznelis (2001b), *Orthogonal decomposition of symmetric functions defined on random permutations*, Preprint 2001–05 Faculty of Mathematics and Informatics, Vilnius University.

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

Chen Jiahua and R. R. Sitter (1999), *A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys*, Statistica Sinica **9**, 385–406.

W. G. Cochran (1977), *Sampling techniques*, 3rd ed. Wiley New York.

M. Dwass (1956), *The large–sample power of rank tests in the two–samples problem*, Ann. Math. Statist. **27**, 352–374.

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

R. Helmers (1991), *On the Edgeworth expansion and the bootstrap approximation for a Studentized U-statistic*, Ann. Statist. **19**, 470–484.

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

D. Krewski (1978), *Jackknifing U-statistics in finite populations*, Commun. Statist.-Theor. Meth. A. **7**, 1–12.

E. L. Lehmann (1951), *Consistency and unbiasedness of certain nonparametric test*, Ann. Math. Statist. **22**, 165–179.

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

H. Putter and W. R. van Zwet (1998), *Empirical Edgeworth expansions for symmetric statistics*, Ann. Statist. **26**, 1540–1569.

J. N. K. Rao and C. F. J. Wu (1988), *Resampling inference with complex survey data*, Journal of the American Statistical Association **83**, 231–241.

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

J. Shao (1994), *L-statistics in complex surveys*, Ann. Statist. **22**, 946–967.

J. Shao (1996), *Resampling methods in sample surveys*, Statistics **27**, 203–254.

Zhao Lin Cheng and Chen Xi Ru (1990), *Normal approximation for finite-population U-statistics*, Acta mathematicae applicatae Sinica **6**, 263–272.

van W. R. van Zwet (1984), *A Berry–Esseen bound for symmetric statistics*, Z. Wahrsch. Verw. Gebiete **66**, 425–440.

W. R. van Zwet (1990), *Asymptotic statistics*, Unpublished lecture notes, Chapel Hill.