The Annals of Statistics
2001, Vol. 29, No. 3, 899–917

ORTHOGONAL DECOMPOSITION OF FINITE POPULATION STATISTICS AND ITS APPLICATIONS TO DISTRIBUTIONAL ASYMPTOTICS

BY M. BLOZNELIS[1] AND F. GÖTZE

Vilnius University and Institute of Mathematics and Informatics, Bielefeld Universität

We study orthogonal decomposition of symmetric statistics based on samples drawn without replacement from finite populations. Several applications to finite population statistics are given: we establish one-term Edgeworth expansions for general asymptotically normal symmetric statistics, prove an Efron–Stein inequality and the consistency of the jackknife estimator of variance. Our expansions provide second order a.s. approximations to Wu's jackknife histogram.

1. Introduction. Orthogonal decomposition of statistics were introduced by Hoeffding (1948) in his proof of the asymptotic normality of U-statistics. Since then the orthogonal decomposition (called also ANOVA decomposition or Hoeffding's decomposition) has become an indispensable tool of analysis of distributional properties of statistics based on independent observations. In particular, it plays a crucial role in the analysis of variance components [Efron and Stein (1981), Karlin and Rinott (1982), Vitale (1992)] and provides a natural framework for first- and second-order asymptotics of statistics [Hajek (1968), Rubin and Vitale (1980), van Zwet (1984), Bentkus, Götze and van Zwet (1997)].

We study orthogonal decomposition of statistics based on *samples drawn without replacement* from finite populations. For simplicity we consider the case of simple random samples. We start with an overview of the orthogonal decomposition of general symmetric statistics based on simple random samples; see Section 2 below. Here we also provide bounds for the remainders of the approximation of statistics by a fixed number, say two or three, of terms of the Hoeffding decomposition. Orthogonal decompositions of finite population U-statistics of fixed degree k were used first in Zhao and Chen (1990) without providing uniform estimates for the remainders as k increases together with the number of observations.

In Section 3 some brief applications are given. Here we prove the consistency of the jackknife variance estimator for symmetric statistics based on samples drawn without replacement and the finite population Efron–Stein inequality. We discuss second-order approximations to the distribution of

Received February 2000; revised December 2000.

[1]Supported by Alexander von Humboldt Foundation.

AMS 2000 *subject classifications*. Primary 62F20; secondary 60F05.

Key words and phrases. ANOVA, Hoeffding decomposition, sampling without replacement, finite population, asymptotic expansion, Edgeworth expansion, stochastic expansion, jackknife estimator of variance, Efron–Stein inequality, jackknife histogram.

jackknife histograms [Shao (1989), Wu (1990), Booth and Hall (1993)] and subsampling [Politis and Romano (1994), Bickel, Götze and van Zwet (1997), Bertail (1997)]. In Section 4 the Hoeffding decomposition is used to establish asymptotic expansions for distribution functions of general symmetric finite population statistics.

2. Hoeffding's decomposition. Let $T = t(X_1, \dots, X_n)$ denote a statistic based on simple random sample $X_1, \dots, X_n$ drawn without replacement from a finite population $\mathscr{X} = \{x_1, \dots, x_N\}$ consisting of N units. Clearly, $n < N$. We shall assume that the function t is invariant under permutations of its arguments. Therefore, T is a symmetric statistic.

The Hoeffding decomposition,

$$T = \mathbf{E}T + \sum_{1\le i\le n} g_1(X_i) + \sum_{1\le i<j\le n} g_2(X_i, X_j) + \cdots \tag{2.1}$$

represents T by the sum of n mutually uncorrelated U-statistics of increasing order. Here g_k, $k = 1, 2, \dots, n$, denote symmetric kernels, which satisfy

$$\mathbf{E}(g_k(X_{i_1}, \dots, X_{i_k}) | X_{j_1}, \dots, X_{j_r}) = 0, \tag{2.2}$$

for every $1 \le i_1 < \cdots < i_k \le n$ and $1 \le j_1 < \cdots < j_r \le n$ such that $r < k$. It is easy to verify that such a decomposition is unique.

The functions g_k, $k = 1, \dots, n$, are linear combinations of conditional expectations

$$h_j(x_{i_1}, \dots, x_{i_j}) = \mathbf{E}(T - \mathbf{E}T | X_1 = x_{i_1}, \dots, X_j = x_{i_j}).$$

We show in the Appendix that

$$g_k(x_1, \dots, x_k) = d_{n,k} \sum_{j=1}^{k} M_{k,j} \sum_{1\le i_1<\cdots<i_j\le k} h_j(x_{i_1}, \dots, x_{i_j}). \tag{2.3}$$

Here, for $k = 2, 3, \dots, n$,

$$d_{k,j} = \prod_{r=j}^{k-1} \frac{N-r}{N-r-j}, \qquad 1 \le j \le \min\{k-1, N-k\}. \tag{2.4}$$

In the case where $2k > N+1$ we put $d_{k,j} = 0$ for $N-k < j \le k-1$. Finally, we write $d_{n,n} = 1$ for $2n \le N+1$ and $d_{n,n} = 0$ for $2n > N+1$. Furthermore, the coefficients $M_{k,j}$, for k satisfying the inequality $2k \le N+1$, are given by the recursive relation,

$$M_{k,j} = -\sum_{i=j}^{k-1} d_{k,i} M_{i,j} \binom{k-j}{i-j}, \qquad 1 \le j \le k-1,$$

and we put $M_{k,k} = 1$. For $2k > N+1$ we write $M_{k,j} = 0$.

A simple calculation gives

$$g_1(X_1) = \frac{N-1}{N-n} h_1(X_1),$$
$$g_2(X_1, Y_2) = \frac{N-2}{N-n} \frac{N-3}{N-n-1} \Big(h_2(X_1, Y_2) - \frac{N-1}{N-2} (h_1(X_1) + h_1(Y_2)) \Big).$$

Let U_j, $1 \le j \le n$ denote the jth sum in (2.1),

$$U_j = U_j(T) = \sum_{1 \le i_1 < \cdots < i_j \le n} g_j(X_{i_1}, \ldots, X_{i_j}).$$

Clearly, (2.2) implies $\mathbf{E} U_k U_r = 0$, for $k \neq r$. That is, the U-statistics of the decomposition (2.1) are mutually uncorrelated. Note that, contrary to the i.i.d. case, the random variables $g_j(X_{i_1}, \ldots, X_{i_j})$ and $g_j(X_{k_1}, \ldots, X_{k_j})$ are not uncorrelated. Indeed, for m denoting the number of elements of the intersection $\{i_1, \ldots, i_j\} \cap \{k_1, \ldots, k_j\}$ we have

$$s_{j,m} := \mathbf{E} g_j(X_{i_1}, \ldots, X_{i_j}) g_j(X_{k_1}, \ldots, X_{k_j}) = \frac{(-1)^{j-m}}{\binom{N-j}{j-m}} \sigma_j^2. \tag{2.5}$$

Here we donote $\sigma_j^2 = \mathbf{E} g_j^2(X_{i_1}, \ldots, X_{i_j})$. Invoking a simple combinatorial argument we evaluate the variances

$$\mathbf{Var}\, U_j = \frac{\binom{n}{j}\binom{N-n}{j}}{\binom{N-j}{j}} \sigma_j^2 \quad \text{and} \quad \mathbf{Var}\, T = \sum_{j=1}^{n} \frac{\binom{n}{j}\binom{N-n}{j}}{\binom{N-j}{j}} \sigma_j^2. \tag{2.6}$$

The formulas (2.5) and (2.6) have been used in Zhao and Chen (1990) for U-statistics of fixed degree k. For convenience, we include the proof of (2.5) and (2.6); see Lemmas 1 and 2 in the Appendix.

Here we shall develop several consequences of (2.3) and (2.6) which are new and have important applications. It follows from (2.3) and (2.6) that for $j > N-n$ we have $U_j \equiv 0$. That is, the decomposition (2.1) reduces to

$$T = \mathbf{E}T + U_1 + \cdots + U_{n_*}, \qquad n_* = \min\{n, N-n\}. \tag{2.7}$$

Moreover, (2.2) entails the duality property, formulated in Proposition 1 below.

Let $(X_1, \ldots, X_N)$ denote a random permutation of the ordered set $(x_1, \ldots, x_N)$ which is uniformly distributed over the class of permutations. Then the first n observations $X_1, \ldots, X_n$ represent a simple random sample from $\mathscr{X}$. For $j = 1, \ldots, N-n$ denote $X'_j = X_{n+j}$.

PROPOSITION 1. *For $j \le n_*$ we have*

$$U_j \equiv U'_j \quad \textit{where } U'_j = (-1)^j \sum_{1 \le i_1 < \cdots < i_j \le N-n} g_j(X'_{i_1}, \ldots, X'_{i_j}). \tag{2.8}$$

Therefore, $T \equiv T'$, *where* $T' = \mathbf{E}T + U'_1 + \cdots + U'_{n_*}$.

The proposition says that, in a sense, T is a function of n_* random variables. In particular, if $n > N/2$, one may replace the statistic by a U-statistic based on $n_* < N/2$ observations.

PROOF. For the linear statistic U_1 the identity (2.8) is a consequence of $\mathbf{E}U_1 = 0$. For $j = 2, \ldots, n_*$, this identity follows from (2.2). $\square$

One may view the decomposition (2.1) as a stochastic expansion of the statistic T. Indeed, for a number of statistics the first few terms of the decomposition provide sufficiently precise approximations. To bound the errors of such approximations we introduce appropriate smoothness conditions.

Denote

$$\begin{aligned} D^j T = {} & t(X_1, \ldots, X_n) \\ & - t(X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n, X'_j), \qquad X'_j = X_{n+j}. \end{aligned}$$

Higher order difference operations are defined recursively,

$$D^{j_1, j_2} T = D^{j_2}(D^{j_1} T), \quad D^{j_1, j_2, j_3} T = D^{j_3}(D^{j_2}(D^{j_1} T)), \ldots .$$

They are symmetric; that is, $D^{j_1, j_2} T = D^{j_2, j_1} T$, etc. Given $k < n_*$, write

$$\delta_j = \delta_j(T) = \mathbf{E}\Big(n_*^{(j-1)} \mathbb{D}_j T\Big)^2, \qquad \mathbb{D}_j T = D^{1,2,\ldots,j} T,\ 1 \le j \le k.$$

In Examples 1 and 2 below we estimate the moments δ_j for U-statistics and smooth functions of sample means.

THEOREM 1. *For $1 \le k \le n_*$, we have*

$$T = \mathbf{E}T + U_1 + \cdots + U_k + R_k \quad \text{with } \mathbf{E}R_k^2 \le n_*^{-(k-1)} \delta_{k+1}. \tag{2.9}$$

The proof of Theorem 1 is given in the Appendix.

3. Applications.

Jackknife estimator of variance. The Quenouille–Tukey jackknife estimator of variance is a symmetric statistic of observations $X_1, \ldots, X_{n+1}$,

$$\sigma_J^2 = \sigma_J^2(T) = \sum_{j=1}^{n+1} (T_{(j)} - \overline{T})^2, \qquad \overline{T} = \frac{1}{n+1} \sum_{j=1}^{n+1} T_{(j)},$$

where we write $T_{(j)} = t(X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n, X_{n+1})$.

In the case of *independent and identically distributed* observations the jackknife estimator of variance is asymptotically consistent if the underlying statistic is sufficiently smooth; see, for example, Miller (1974), Parr (1985) and Shao and Wu (1989), where in the later paper several smoothness conditions are discussed.

Here we consider statistics based on *samples drawn without replacement*. Let $X_1, \ldots, X_n$ be a simple random sample drawn without replacement from the population $\mathscr{X} = \{x_1, \dots, x_N\}$. The jackknife variance estimator for the statistic $T = t(X_1, \dots, X_n)$ is defined by

$$\sigma^2_{FJ} = \sigma^2_{FJ}(T) = q\sigma^2_J(T) \quad \text{where} \quad q = (N-n)/N.$$

Note that $\sigma^2_{FJ}(T)$ is a symmetric statistic of the sample $X_1, \dots, X_{n+1}$ drawn without replacement from the population $\mathscr{X}$. For a linear statistic $T = \mathbf{E}T + \sum_{i=1}^n g_1(X_i)$ it is easy to show that $\mathbf{E}\sigma^2_{FJ}(T) = \mathbf{Var}\ T$.

Our first application of the orthogonal decomposition (2.1) is the finite population Efron–Stein inequality: for an arbitrary symmetric finite population statistic $T = t(X_1, \dots, X_n)$ we have

$$\mathbf{E}\sigma^2_{FJ}(T) \geq \mathbf{Var}\ T. \tag{3.1}$$

That is, the jackknife variance estimator tends to be biased upward. In the i.i.d. case the Efron–Stein inequality was proved by Efron and Stein (1981). The proof of (3.1) is given in the Appendix.

Another application of (2.1) is a general consistency result for the estimator σ^2_{FJ}. Assuming that n and $N \to \infty$ we prove the consistency of σ^2_{FJ} for asymptotically linear symmetric finite population statistics. In order to formulate the consistency result we consider a sequence of statistics $T_n = t_n(X_1, \dots, X_n)$. That is, we show that for every $\varepsilon > 0$,

$$\mathbf{P}\{|\sigma^2_{FJ}(T_n) - \mathbf{Var}\ T_n| > \varepsilon\} = o(1) \quad \text{as } n,\ N \to \infty. \tag{3.2}$$

Let $T_{i,n}$ denote the summand $g_1(X_i)$ of the linear part of decomposition (2.1) for the statistic T_n.

PROPOSITION 2. *Assume that N and $n_* = \min\{n, N-n\} \to \infty$. Assume that*:

(i) *For some $0 < c_1 < c_2 < \infty$, we have $c_1 \leq \mathbf{Var}\ T_n \leq c_2$ and $\delta_2(T_n) = o(1)$.*
(ii) *For every $\varepsilon > 0$,*

$$n_*\mathbf{E}T^2_{1,n}\mathbb{I}_{T^2_{1,n} > \varepsilon} = o(1). \tag{3.3}$$

Then (3.2) *holds.*

The proof of Proposition 2 is given in the Appendix.

Recall that $\delta_2(T_n) = n_*^2\mathbf{E}(\mathbb{D}_2T_n)^2$, where

$$\begin{aligned}\mathbb{D}_2T_n = t_n(X_1, \dots, X_n) &- t_n(X_2, \dots, X_n, X_{n+1}) \\ &- t_n(X_1, X_3, \dots, X_n, X_{n+2}) + t_n(X_3, \dots, X_{n+2}).\end{aligned}$$

Note that condition (i) implies that T_n is asymptotically linear as $n_* \to \infty$. That is,

$$T_n = \mathbf{E}T_n + \sum_{i=1}^n T_{i,n} + op(1) \quad \text{with } \mathbf{Var}\left(\sum_{i=1}^n T_{i,n}\right) \geq c_1 - o(1). \tag{3.4}$$

Here $\mathbf{E}T_n + \sum_{i=1}^n T_{i,n}$ denotes the linear part of the decomposition (2.1) of T_n. Indeed, by (2.1), we have $T_n = \mathbf{E}T_n + \sum_{i=1}^n T_{i,n} + r_n$, where the remainder

r_n and the linear part are uncorrelated. The condition $\delta_2(T_n) = o(1)$ implies the bound $\mathbf{E}r_n^2 = o(1)$; see Theorem 1. Therefore, (3.4) follows. Note that the uniform integrability condition (3.3) can be replaced by a more restrictive moment condition $\limsup_n \mathbf{E}(T_{1,n}^2 n_*)^{1+\delta} < \infty$, for some $\delta > 0$.

Subsampling. Let $Y_1, \ldots, Y_n$ be independent observations from a probability distribution P. Let $\theta_n = \theta_n(Y_1, \ldots, Y_n)$ be an estimator of a real-valued parameter $\theta = \theta(P)$. In order to make inferences about θ one estimates the distribution of $\theta_n - \theta$. Assuming that the distribution K_n of $\tau_n(\theta_n - \theta)$ converges weakly to a limit law, Politis and Romano (1994) showed that the conditional distribution function,

$$\widehat{K}_m(x) = \mathbf{P}\{\tau_m(\theta_m(X_1, \ldots, X_m) - \theta_n) \le x | Y_1, \ldots, Y_n\}$$

estimates the true distribution function $K_n(x) = \mathbf{P}\{\tau_n(\theta_n - \theta) \le x\}$ consistently as $n, m \to \infty$ so that $m/n \to 0$ and $\tau_m/\tau_n \to 0$. Here τ_n denotes a nonrandom sequence of normalizing constants and $X_1, \ldots, X_m$ denotes a random sample drawn without replacement from $\{Y_1, \ldots, Y_n\}$. Assuming in addition that $K_n(x)$ admits an Edgeworth expansion, Bertail (1997) showed that $\widehat{K}_m(x)$ admits a corresponding stochastic expansion. The proofs of Politis and Romano (1994) and Bertail (1997) exploit the U-statistic structure of the conditional distribution function $\widehat{K}_m(x)$ and rely on the law of large numbers for U-statistics.

Another way to construct higher order approximations to $\widehat{K}_m(x)$ is based on conditional asymptotic expansions given $\{Y_1, \ldots, Y_n\}$. Let $v_m = v_m(Y_1, \ldots, Y_n)$ [respectively, $e_m = e_m(Y_1, \ldots, Y_n)$] denote the conditional variance (respectively, the mean value) of $\tau_m(\theta_m(X_1, \ldots, X_m) - \theta_n)$, given $Y_1, \ldots, Y_n$. Theorem 2 below provides the conditional asymptotic expansion,

$$\widehat{K}_m(xv_m + e_m) = \Phi(x) - \hat{q}_m(x)\Phi^{(3)}(x)(m(1 - m/n))^{1/2} + O(m_*^{-1}) \tag{3.5}$$

almost surely as $m_* = \min\{m, n - m\} \to \infty$ and $n \to \infty$. An explicit formula for the first term of the expansion $\hat{q}_m(x)$ is provided in Section 4 below.

Wu (1990) used a one-term asymptotic expansion of finite population Studentized mean due to Babu and Singh (1985) to construct a second-order approximation like (3.5) to the jackknife histogram of a Studentized mean Clearly, (3.5) provides such approximations with remainder $O(m_*^{-1})$ for a broad class of asymptotically linear statistics; see also Bickel, Götze and van Zwet (1997) for other possible applications of (3.5).

Let us mention that for some classes of statistics the order of the approximation of $\widehat{K}_m(x)$ can be further improved by using Richardson extrapolation; see Bickel and Yahav (1988), Booth and Hall (1993), Bertail (1997).

Finally, we discuss applications to resampling of finite population statistics. Using orthogonal decomposition of Section 2 and Edgeworth expansions of Section 4, one can extend i.i.d. results of Putter and van Zwet (1998) on empirical Edgeworth expansions to samples drawn without replacement. This question is addressed in Bloznelis (2000). Furthermore, the orthogonal decomposition of Section 2 and the expansions of Section 4 below could be extended to stratified sampling without replacement models and applied to resampling schemes like finite the population bootstrap [Gross (1980), Bickel and Freedman (1984), Chao and Lo (1985), Babu and Singh (1985), Chen and Sitter (1993), Booth, Butler and Hall (1994), Helmers and Wegkamp (1998)] and its modifications.

4. Stochastic and asymptotic expansions. We shall apply (2.9) to study the asymptotics of the distribution of T.

When speaking about the finite population asymptotics we assume that we have a sequence of populations $\mathscr{X}_r = \{x_{r,1}, \dots, x_{r,N_r}\}$, with $N_r \to \infty$ as $r \to \infty$, and a sequence of symmetric statistics $T_r = t_r(X_{r,1}, \dots, X_{r,n_r})$, based on samples $X_{r,1}, \dots, X_{r,n_r}$ drawn without replacement from $\mathscr{X}_r$. We shall assume that the variances $\tilde{\sigma}_r^2 = \mathbf{Var}\ T_r$ remain bounded away from zero as $r \to \infty$. In order to keep the notation simple we drop the subscript r in what follows.

In typical situations (U-statistics, smooth functions of sample means, Student's t and many others) we have $U_j = O_P(n_*^{(1-j)/2})$, for $j = 1, \dots, k$, and

$$\delta_{k+1} = O(n_*^{-1}) \quad \text{as } n_*, N \to \infty \tag{4.1}$$

for some k. Clearly, (4.1) is a smoothness condition. It implies the validity of the stochastic expansion (2.9) with the remainder $R_k = O_P(n_*^{-k/2})$. The condition (4.1) is easy to handle. Below, we verify this condition for two classes of statistics: smooth functions of multivariate sample means and U-statistics.

In the remaining part of this section we study first- and second-order approximations of asymptotically linear statistics. We shall assume that the linear part U_1 is nondegenerate; that is, $s^2 := \mathbf{Var}\ U_1 > 0$. Note that, by (2.6),

$$s^2 = \tau^2 \sigma_1^2 N/(N-1) \quad \text{where } \tau^2 = Npq,\ \ p = n/N,\ \ q = 1-p.$$

Clearly, $n_*/2 \le \tau^2 \le n_*$. In Proposition 3 below we formulate sufficient conditions for asymptotic normality.

PROPOSITION 3. *Assume that $\tilde{\sigma}$ remains bounded away from zero and $\delta_2 = o(1)$ as $n_*, N \to \infty$. Then $\tilde{\sigma} - s = o(1)$. Suppose, in addition, that* (3.3) *holds. Then $\tilde{\sigma}^{-1}(T - \mathbf{E}T)$, $s^{-1}(T - \mathbf{E}T)$ and $(T - \mathbf{E}T)/\sigma_{FJ}$ are asymptotically standard normal.*

PROOF. In view of Theorem 1, the condition $\delta_2 = o(1)$ implies the validity of the short stochastic expansion $T = \mathbf{E}T + U_1 + o_P(1)$. We also have $\tilde{\sigma}^2 - s^2 = o(1)$ and, by Proposition 2, $\sigma_{FJ}^2/\tilde{\sigma}^2 = \sigma_{FJ}^2(T/\tilde{\sigma}) = 1 + o_P(1)$. Therefore, the linear part dominates the statistic T and it suffices to prove the asymptotic normality of $s^{-1}U_1$. The asymptotic normality is ensured by (3.3) which (under the conditions of the proposition) implies a Lindeberg-type Erdős–Rényi condition,

$$\mathbf{E}g_1^2(X_1)\sigma_1^{-2}\mathbb{I}_{|g_1(X_1)|>\varepsilon\tau\sigma_1} = o(1) \quad \text{as } n_*, N \to \infty$$

for every $\varepsilon > 0$; see Erdős and Rényi (1959). Note that (3.3) is equivalent to the Erdős–Rényi condition if, in addition, $\tilde{\sigma}$ is bounded as $n_*, N \to \infty$. □

Assuming that (4.1) holds, for $k = 2$, we obtain from Theorem 1 the stochastic expansion $T = \mathbf{E}T + U_1 + U_2 + O_P(n_*^{-1})$. It suggests that Edgeworth expansions of $T - \mathbf{E}T$ and $U_1 + U_2$ should coincide up to $O(n_*^{-1})$. Note that $U_1 + U_2$ is a U-statistic of degree two. Bloznelis and Götze (2000) showed that a one-term asymptotic expansion,

$$G(x) = \Phi(x) - \frac{(q-p)\alpha + 3\kappa}{6\tau}\Phi^{(3)}(x),$$

approximates the distribution function $\mathbf{P}\{U_1 + U_2 \le xs\}$ with the remainder $O(n_*^{-1})$. Here $\Phi^{(3)}(x)$ denotes the third derivative of the standard normal distribution function $\Phi(x)$,

$$\alpha = \sigma_1^{-3}\mathbf{E}g_1^3(x_1), \qquad \kappa = \sigma_1^{-3}\tau^2\mathbf{E}g_2(X_1, X_2)g_1(X_1)g_1(X_2).$$

Using Theorem 1, one may extend this result to arbitrary symmetric statistics. In particular, in order to construct a one-term Edgeworth expansion of T we do not need to evaluate all the summands of (2.1), but (moments of) the first few terms only. A general result formulated in Theorem 2 below provides the bounds $o(n_*^{-1/2})$ and $O(n_*^{-1})$ for the error of the expansion,

$$\Delta := \sup_x |F(x) - G(x)| \quad \text{where } F(x) = \mathbf{P}\{T \le \mathbf{E}T + x\tilde{\sigma}\}.$$

Similar bounds hold for

$$\Delta_1 := \sup_x |F_1(x) - G(x)| \quad \text{where } F_1(x) = \mathbf{P}\{T \le \mathbf{E}T + x\sigma_1\tau\}.$$

In order to establish the validity of an Edgeworth expansion we need to impose an appropriate smoothness condition. It is the nonlattice condition in the case of the remainder $o(n_*^{-1/2})$ and it is a Cramér type condition in the case of the remainder $O(n_*^{-1})$. Either of these conditions will be imposed on the linear part of the statistic.

Given $g: \mathbb{R} \to \mathbb{C}$ write $\|g\|_{[a,b]} = \sup_{a<|t|<b}|g(t)|$. We shall say that the linear part is asymptotically nonlattice, if for every $\varepsilon > 0$ and every $B > 0$, the characteristic function $\varphi(t) = \mathbf{E}\exp\{it\sigma_1^{-1}g_1(X_1)\}$ of the random variable $\sigma_1^{-1}g_1(X_1)$ satisfies

$$\liminf_{n_*, N\to\infty} \|\varphi\|_{[\varepsilon, B]} < 1. \tag{4.2}$$

A more stringent smoothness condition is a Cramér-type condition,

$$\liminf_{n_*, N\to\infty} \|\varphi\|_{[\varepsilon, \tau]} < 1. \tag{4.3}$$

Note that $\tau \to \infty$ as $n_*, N \to \infty$. Write

$$\beta_s = \mathbf{E}|n_*^{1/2} g_1(X_1)|^s, \quad \gamma_s = \mathbf{E}|n_*^{3/2} g_2(X_1, X_2)|^s, \quad \zeta_s = \mathbf{E}|n_*^{5/2} g_3(X_1, X_2, X_3)|^s.$$

THEOREM 2. *Assume that $\tilde{\sigma}$ remains bounded away from zero as $N \to \infty$.*

(i) *Assume that* (4.2) *holds, $\delta_3 = o(n_*^{-1/2})$ and, for some $\delta > 0$, the moments $\beta_{3+\delta}$ and $\gamma_{2+\delta}$ are bounded as $n_*, N \to \infty$. Then*

$$\Delta = o(n_*^{-1/2}) \quad \text{and} \quad \Delta_1 = o(n_*^{-1/2}) \quad \text{as } n_*, N \to \infty.$$

(ii) *Assume that* (4.3) *holds, $\delta_4 = O(n_*^{-1})$ and, the moments $\beta_4, \gamma_4, \zeta_2$ are bounded as $n_*, N \to \infty$. Then*

$$\Delta = O(n_*^{-1}) \quad \text{and} \quad \Delta_1 = O(n_*^{-1}) \quad \text{as } n_*, N \to \infty.$$

PROOF. Note that either of the conditions (i) and (ii) implies

$$\tilde{\sigma}^2 = s^2 + O(n_*^{-1}) = \tau^2 \sigma_1^2 + O(n_*^{-1}). \tag{4.4}$$

Therefore, it suffices to construct bounds for Δ_1.

Let us prove $\Delta_1 = o(n_*^{-1/2})$. In the case of U-statistics of degree two [the remainder in (2.9) $R_2 \equiv 0$], this bound is proved in Bloznelis and Götze (1999a). A passage to the general case can be made by using a Slutzky-type argument. Indeed, by (2.9), under condition (i), we have $\mathbf{P}\{|R_2| > \varepsilon n_*^{-1/2}\} \le \varepsilon^{-2}\delta_3 = o(n_*^{-1/2})$, for every $\varepsilon > 0$. Note that $\sup_x |G^{(1)}(x)|$ remains bounded as $n_*, N \to \infty$.

The proof of the bound $\Delta_1 = O(n_*^{-1})$ is rather complex and laborious. It is given in a more technical paper [Bloznelis and Götze (1999b)]. □

Note that if $\tilde{\sigma}$ remains bounded away from zero as $n_*, N \to \infty$, then (4.4) implies $\tau\sigma_1/\tilde{\sigma} = 1 + O(n_*^{-1})$. Therefore, we can replace G by G_0, where

$$G_0(x) = \Phi(x) - \frac{\tau^2}{\tilde{\sigma}^2} \frac{(q-p)\alpha_0 + 3\tau^2\kappa_0}{6\tilde{\sigma}} \Phi^{(3)}(x),$$
$$\alpha_0 = \mathbf{E} g_1^3(X_1), \qquad \kappa_0 = \mathbf{E} g_2(X_1, X_2) g_1(X_1) g_1(X_2).$$

COROLLARY. *Theorem* 2 *remains valid if we replace G by G_0 in the definition of Δ and Δ_1.*

If $n, N \to \infty$ and $n^2 = o(N)$, the simple random sample model approaches the i.i.d. situation. In this case Theorem 2 and the Corollary agree with the corresponding results of Bentkus, Götze and van Zwet (1997), who constructed second-order approximations to symmetric statistics of i.i.d. observations.

In order to construct a one-term Edgeworth expansion G (respectively, G_0), one needs to evaluate the parameters σ_1, α, κ (respectively, $\tilde{\sigma}, \alpha_0, \kappa_0$). For some classes of statistics it reduces to routine calculations; see examples below. Another possible way is to substitute consistent estimators of these parameters. The consistency of the corresponding jackknife estimators is established in Bloznelis (2000).

Earlier results on Edgeworth expansions for nonlinear asymptotically normal finite population statistics by Babu and Singh (1985), Babu and Bai (1996) apply to statistics which can be approximated by smooth functions of multivariate sample means. Their approach combines linearization and expansions for multivariate sample means. This approach, though conceptually simpler, focuses on a particular class of statistics (smooth functions of multivariate sample means). Furthermore, it often requires a somewhat restrictive Cramér-type smoothness condition imposed on the underlying multivariate sample mean rather than on the linear part of the statistic itself. Another approach was used by Kokic and Weber (1990) to prove the validity of one-term Edgeworth expansions for finite population U-statistics. However, they did not use a finite population orthogonal decomposition. In contrast to Theorem 2 all of the above-mentioned results establish the validity of the expansions under some additional conditions on the sample fraction $p = n/N$.

In what follows we consider two examples: U-statistics and smooth functions of sample means.

EXAMPLE 1 (U-statistics). Given an integer m let φ denote a real symmetric function defined on m-subsets of the population $\mathscr{X} = \{x_1, \dots, x_N\}$. Define a U-statistic

$$U = \sum_{1 \le i_1 < \cdots < i_m \le n} \varphi(X_{i_1}, \dots, X_{i_m}) \tag{4.5}$$

based on the simple random sample $X_1, \dots, X_n$ $(n > m)$ drawn without replacement from $\mathscr{X}$. We shall assume that $\mathbf{E}U = 0$ and construct a one-term Edgeworth expansion $G(x)$. To this end we evaluate the parameters σ_1, α and κ.

Write Hoeffding's decomposition for a symmetric statistic $\varphi(X_1, \dots, X_m)$,

$$\varphi(X_1, \dots, X_m) = \sum_{1 \le k \le m} \sum_{1 \le i_1 < \cdots < i_k \le m} \tilde{g}_k(X_{i_1}, \dots, X_{i_k}),$$

where symmetric kernels $\tilde{g}_k$ are defined by (2.3). Clearly, every summand $\varphi(X_{i_1}, \dots, X_{i_m})$ of (4.5) can be written in such form. Substitution of these expressions in (4.5) yields Hoeffding's decomposition of U,

$$U = \sum_{1 \le k \le m} \sum_{1 \le i_1 < \cdots < i_k \le n} g_k(X_{i_1}, \dots, X_{i_k}) \quad \text{where } g_k = \binom{n-k}{m-k} \tilde{g}_k.$$

Denoting $\sigma_0^2 = \mathbf{E}\tilde{g}_1{}^2(X_1)$ we have $\sigma_1^2 = \binom{n-1}{m-1}^2 \sigma_0^2$,

$$\alpha = \sigma_0^{-3} \mathbf{E}\tilde{g}_1^3(X_1), \qquad \kappa = \frac{(m-1)n}{n-1} \frac{q}{\sigma_0^3} \mathbf{E}\tilde{g}_2(X_1, X_2)\tilde{g}_1(X_1)\tilde{g}_1(X_2).$$

In order to estimate the moments (of differences) δ_k, we invoke variance formulas (2.6) and (A.21); see below. A straightforward calculation gives

$$\delta_k \le c(m) n_*^{k-2} (\mathbf{Var}\, U_k^2 + \cdots + \mathbf{Var}\, U_m^2), \tag{4.6}$$

where $c(m)$ denotes a constant which depends only on m. Note that (4.6) implies $n_*^{-(k-1)} \delta_{k+1} \le c(m) \mathbf{E} R_k^2$, where R_k denotes the remainder of (2.9). That is, for U-statistics of degree m, the inequality (2.9) is precise [up to the constant $c(m)$].

EXAMPLE 2 (Smooth functions of multivariate sample means). Assume that $\mathscr{X} \subset \mathbb{R}^k$ and consider the statistic $T = \sqrt{n/q}(h(\overline{X}) - h(a))$, where h: $\mathbb{R}^k \to \mathbb{R}$. Here $\overline{X} = n^{-1}(X_1 + \cdots + X_n)$ and $a = \mathbf{E} X_1$. Assuming that h is three times differentiable and derivatives are bounded, we construct a one-term Edgeworth expansion and bound δ_3. In order to bound δ_4 we need one more derivative.

We may assume without loss of generality that $\mathbf{E} X_1 = 0$.

Denote $Y_i = h^{(1)}(0) X_i$ and $Y_{i,j} = h^{(2)}(0) X_i X_j$ and write

$$\sigma_h^2 = \mathbf{E} Y_1^2, \qquad \alpha_h = \sigma_h^{-3} \mathbf{E} Y_1^3, \qquad \kappa_h = \sigma_h^{-3} \mathbf{E} Y_1 Y_2 Y_{1,2}.$$

Here $h^{(s)}(y)$ denotes the sth derivative of h at the point $y \in \mathbb{R}^k$. We write $h^{(s)}(y) z_1 \cdots z_s$ to denote the value of the s-linear form $h^{(s)}(y)$ with arguments $z_1, \ldots, z_s \in \mathbb{R}^k$.

Straightforward, but tedious, calculations show that α_h and $q\kappa_h$ provide sufficiently precise approximations to α and κ. Therefore, by Theorem 2,

$$\Phi(x) - \frac{(q-p)\alpha_h + 3q\kappa_h}{6\tau} \Phi^{(3)}(x)$$

can be used as a one-term asymptotic expansion of the distribution function $\mathbf{P}\{T - \mathbf{E}T \le x\sigma_h\}$. The verification of the conditions of Theorem 2 reduces to routine, but cumbersome calculations. We skip most of technical details and focus on the smoothness condition (4.1) only.

Let $\|h^{(s)}(y)\|$ denote the smallest $c > 0$ such that $|h^{(s)}(y) z_1 \cdots z_s| \le c|z_1| \cdots |z_s|$. Here $|z_i|^2 = z_{i,1}^2 + \cdots + z_{i,k}^2$ denotes the Euclidean norm of a vector $z_i = (z_{i,1}, \ldots, z_{i,k}) \in \mathbb{R}^k$. We say that the sth derivative is bounded if $\|h^{(s)}\|_\infty := \sup_y \|h^{(s)}(y)\|$ is finite.

Assuming that $\|h^{(j)}\|_\infty$ and $\mathbf{E}|X_1|^2$ remain bounded as $n_*, N \to \infty$ we prove that $\delta_j(T) = O(n_*^{-1})$. More precisely, we show that, for every fixed $j = 1, 2, \ldots, n_*$,

$$\delta_j(T) \le 2^j \frac{N^j}{[N]_j} \|h^{(j)}\|_\infty^2 (\mathbf{E}|X_1|^2)^j \frac{n_*^{2j-2}}{qn^{2j-1}}, \tag{4.7}$$

where $[N]_j = N(N-1)\cdots(N-j+1)$. Let $\varkappa_1, \varkappa_2, \ldots$ be a sequence of independent random variables uniformly distributed in [0,1]. We assume that the sequence and the random permutation $(X_1, \ldots, X_N)$ are independent. Given a differentiable function f we use the mean value formula $f(x+y) - f(x) =$

$\mathbf{E}_{\varkappa_1} f^{(1)}(x + \varkappa_1 y)y$. Here $\mathbf{E}_{\varkappa_1}$ denotes the conditional expectation given all the random variables but $\varkappa_1$. Write $u_i = n^{-1}(X_i - X_i')$. By the mean value formula,

$$\mathbb{D}_j h(\overline{X}) = \mathbf{E}_{\varkappa_1} \cdots \mathbf{E}_{\varkappa_j} h^{(j)}(\overline{X} - (\varkappa_1 u_1 + \cdots + \varkappa_j u_j))u_1 \cdots u_j. \tag{4.8}$$

Furthermore, invoking the simple bound $\mathbf{E}|X_{i_1}|^2 \cdots |X_{i_j}|^2 \leq N^j/[N]_j \times (\mathbf{E}|X_1|^2)^j$, for $i_1 < \cdots < i_j$, we obtain $\mathbf{E}|u_1 \cdots u_j|^2 \leq 2^j N^j/[N]_j n^{-2j}(\mathbf{E}|X_1|^2)^j$. The last inequality in combination with (4.8) implies (4.7).

The smoothness condition on h can be relaxed. By the law of large numbers, $\overline{X}$ concentrates around $a = \mathbf{E}X_1$ with high probability. Therefore, it suffices to impose smoothness conditions on h in a neighbourhood of a only.

APPENDIX

We may assume without loss of generality that $\mathbf{E}T = 0$. Recall that $(X_1, \ldots, X_N)$ denotes random permutation of the ordered set $(x_1, \ldots, x_N)$.

Denote $\Omega_k = \{1, \ldots, k\}$, for $k = 1, 2, \ldots$, and $\Omega = \Omega_N$. Given a statistic $V = V(X_1, \ldots, X_N)$, write

$$\mathbf{E}(V|A) = \mathbf{E}(V|X_i, i \in A), \qquad A \subset \Omega,$$

and denote $\mathbf{E}(V|\varnothing) = \mathbf{E}V$.

PROOF OF (2.3). Introduce random variables Q_A, for $A \subset \Omega_n$, with $|A| \geq 1$. For $|A| = 1$, we put $Q_A = \mathbf{E}(T|A)$. Let n_0 be the largest integer such that $2n_0 - 1 \leq N$.

For $|A| = 2, 3, \ldots, \min\{n_0, n\}$, we define Q_A recursively as follows. Given $A \subset \Omega_n$, with $|A| = k$, write

$$Q_A = \mathbf{E}(T|A) - d_{k,k-1} \sum_{B \subset A, |B|=k-1} Q_B - \cdots - d_{k,1} \sum_{B \subset A, |B|=1} Q_B, \tag{A.1}$$

where the numbers $d_{k,j}$ are chosen so that for each $B \subset A$,

$$\mathbf{E}(Q_A|B) = 0, \qquad |B| < |A|. \tag{A.2}$$

A straightforward calculation gives (2.4). In Lemma 1 we extend the identity (A.2) to arbitrary $B \subset \Omega_N$ satisfying $|B| < |A|$. Now (A.2) implies that the sums $\sum_{|B|=i} Q_B$ and $\sum_{|B|=j} Q_B$ in (A.1) are uncorrelated for $1 \leq i < j \leq k$. Therefore, (A.1) provides an orthogonal decomposition for the statistic $\mathbf{E}(T|A)$,

$$\mathbf{E}(T|A) = \sum_{j=1}^{k} d_{k,j} \sum_{B \subset A, |B|=j} Q_B, \tag{A.3}$$

where we put $d_{k,k} = 1$.

For $n \leq n_0$ this identity yields the decomposition for T,

$$T = \mathbf{E}(T|\Omega_n) = \sum_{B \subset \Omega_n, |B| \geq 1} T_B, \qquad T_B = d_{n,|B|} Q_B, \tag{A.4}$$

where for every $B \subset \Omega_n$ and $C \subset \Omega_N$ we have almost surely,

$$\mathbf{E}(T_B|C) = 0 \quad \text{for } |C| < |B|. \tag{A.5}$$

Denoting

$$g_k(x_1, \dots, x_k) = \mathbf{E}(T_{\Omega_k}|X_1 = x_1, \dots, X_k = x_k), \qquad k = 1, 2, \dots, n,$$

we obtain (2.1) from (A.4) and (2.2) from (A.5).

Now assume that $n > n_0$. For $k = n_0 + 1, \dots n$ we show that (A.3) remains valid if we choose $d_{k,j} = 0$, for $j = N - k + 1, \dots, k$. Let Q_A with ($|A| = k$) be defined by (A.1). A calculation shows that if, for $j = 1, \dots, N - k$ the numbers $d_{k,j}$ are given by (2.4) and $d_{k,j} = 0$ for $j > N - k$ then (A.2) holds for every $B \subset A$ with $|B| \le N - k$. Proceeding as in the proof of (A.8) below, one can show that (A.2) extends to arbitrary $B \subset \Omega_N$ such that $|B| \le N - k$. In particular, for $A = \Omega_k$ and $B = \Omega_N \backslash \Omega_k$ we have $\mathbf{E}(Q_A|B) = 0$ almost surely. Since $\mathbf{E}(Q_A|B) = Q_A$, we obtain $Q_A = 0$ almost surely, thus proving (A.3). In the case where $A = \Omega_n$, the identity (A.3) provides the orthogonal decomposition for T,

$$T = \mathbf{E}(T|\Omega_n) = \sum_{j=1}^{N-n} \sum_{B \subset \Omega_n, |B| = j} T_B, \qquad T_B = d_{n,|B|} Q_B, \tag{A.6}$$

where Q_B is given by (A.1) and satisfies (A.2) and (A.5). Finally, invoking a simple combinatoric calculation we derive (2.3) from (A.1) and (A.4), (A.6).

Before formulating Lemmas 1 and 2 we introduce some notation. Define the random variable T_A for arbitrary $A = \{i_1, \dots, i_r\} \subset \Omega$, with cardinality $r \le n$, by putting $T_A = g_r(X_{i_1}, \dots, X_{i_r})$. Let us write also $T_\varnothing = 0$. Note that T_A is a centered symmetric statistic of observations $X_i, i \in A$. Two random variables T_A and T_B are identically distributed if $|A| = |B|$. The difference operation D_i can be applied to T_A provided that $i' = n + i \notin A$. We write $D^i T_A = T_A - T_{A(\{i\})}$ if $i \in A$ and put $D^i T_A = 0$ otherwise. Here $A(\{i\}) = (A \backslash \{i\}) \cup \{i'\}$. Higher order differences $\mathbb{D}_i = D^i \cdots D^1$ are defined recursively: $\mathbb{D}_2 T_A = D^2 T_A - D^2 T_{A(\{1\})}$, etc. We shall apply these differences to T_A, with $A \in \Omega_n$. Note that $\mathbb{D}_i T_A = 0$, whenever $\Omega_i \not\subseteq A$. If $\Omega_i \subseteq A$, we can write $A = \Omega_i \cup B$, for some $B \subset \Omega_n \backslash \Omega_i$. In this case we have

$$\mathbb{D}_i T_{\Omega_i \cup B} = \sum_{C \subset \Omega_i} (-1)^{|C|} T_{\Omega_i(C) \cup B}, \tag{A.7}$$

$$\Omega_i(C) = (\Omega_i \backslash C) \cup C', \qquad C' = \{l' \colon l \in C\}, \; l' = l + n.$$

Here we write also $\Omega_i(\varnothing) = \Omega_i$.

Given $A, B \in \Omega$, with $|A \cap B| = k$ and $|A| = |B| = j$, $j \le n$, denote

$$\sigma_j^2 = \mathbf{E} T_A^2, \qquad s_{j,k} = \mathbf{E} T_A T_B.$$

If, in addition, $A, B \subset \Omega_n \backslash \Omega_i$, we write

$$\sigma_{i,j+i}^2 = \mathbf{E}(\mathbb{D}_i T_{\Omega_i \cup A})^2, \qquad s_{i,j,k} = \mathbf{E} \mathbb{D}_i T_{\Omega_i \cup A} \mathbb{D}_i T_{\Omega_i \cup B}.$$

Put $\sigma_{0,j}^2 = \sigma_j^2$ and $s_{0,j,k} = s_{j,k}$.

LEMMA 1. *The following identities hold*:

$$\mathbf{E}(T_G|H) = 0 \quad \textit{for every } G, H \subset \Omega \textit{ with } |H| < |G|, \tag{A.8}$$

$$s_{j,k} = \frac{(-1)^{j-k}}{\binom{N-j}{j-k}}\sigma_j^2, \qquad 0 \le k \le j \le n_0, \tag{A.9}$$

$$s_{i,j,k} = \frac{(-1)^{j-k}}{\binom{N-j-2i}{j-k}}\sigma_{i,i+j}^2, \qquad 0 \le k \le j \le n_0 - i, \tag{A.10}$$

$$\sigma_{i,j}^2 = \frac{N-j+1}{N-j-i+1}2^i\sigma_j^2, \qquad i \le j \le n_0. \tag{A.11}$$

PROOF. We start with an auxiliary identity (A.12). Fix $C, D \subset \Omega$ such that $1 \le |C| \le |D|$ and $|C\backslash D| = 1$. Denote $C_1 = C \cap D$. We have

$$\mathbf{E}(T_C|D) = \frac{1}{N-|D|}\sum_{i\in\Omega\backslash D} T_{\{i\}\cup C_1} = \frac{-1}{N-|D|}\sum_{i\in D\backslash C} T_{\{i\}\cup C_1}, \tag{A.12}$$

since $(\Omega\backslash D) \cup (D\backslash C) = \Omega\backslash C_1$ and, by (A.5),

$$\sum_{i\in\Omega\backslash C_1} T_{\{i\}\cup C_1} = (N - |C_1|)\mathbf{E}(T_C|C_1) = 0.$$

Let us prove (A.8). For $H \subset G \subset \Omega_n$, (A.8) follows from (A.5). By symmetry, it still holds for $H \subset G \subset \Omega$. For $|H\backslash G| = k$, we prove (A.8) by induction. Assume that (A.8) holds for every $k \le r$. Given $G, H \subset \Omega$, with $|H\backslash G| = r+1$, fix $a \in G\backslash H$ and denote $G_a = G\backslash\{a\}$. Write $\mathbf{E}(T_G|H) = \mathbf{E}(V_a|H)$, where $V_a = \mathbf{E}(T_G|G_a \cup H)$. An application of (A.12) to V_a gives $\mathbf{E}(V_a|H) = 0$, by induction hypothesis, with $k = r$. Hence, $\mathbf{E}(T_G|H) = 0$ and we obtain (A.8), with $k = r+1$.

Let us prove (A.9). Given $A, B \subset \Omega$, with $|A| = |B| = j \ge 1$ and $|A \cap B| = k < j$, fix $i_a \in A\backslash B$ and denote $A_1 = A\backslash\{i_a\}$. An application of (A.12) gives

$$\begin{aligned} s_{j,k} = \mathbf{E}T_B\mathbf{E}(T_A|A_1 \cup B) &= \frac{-1}{N-(2j-k-1)}\sum_{i\in B\backslash A}\mathbf{E}T_BT_{\{i\}\cup A_1} \\ &= \frac{(-1)(j-k)}{N-(2j-k-1)}s_{j,k+1}, \end{aligned} \tag{A.13}$$

where the last identity follows by symmetry. Applying (A.13) several times, for increasing k, we obtain (A.9).

Let us prove (A.10). Let A, B be as above and assume in addition that $A, B \subset \Omega_n\backslash\Omega_i$. Fix $i_a \in A\backslash B$ and $i_b \in B\backslash A$. Denote $B_1 = B\backslash\{i_b\}$ and $B_2 = B_1 \cup \{i_a\}$. It follows from (A.7) that

$$s_{i,j,k} = \sum_{C\subset\Omega_i}(-1)^{|C|}\mathbf{E}\mathbb{D}_iT_{\Omega_i\cup A}h_C, \qquad h_C = T_{\Omega_i(C)\cup B}. \tag{A.14}$$

Write $H = A \cup \Omega_i \cup \Omega_i' \cup B_1$, where $\Omega_i' = \{1', \dots, i'\}$. By (A.12),

$$\mathbf{E}(h_C|H) = \frac{-1}{N - |H|} M, \qquad M = \sum_{r \in H \setminus (\Omega_i(C) \cup B_1)} T_{\Omega_i(C) \cup B_1 \cup \{r\}}.$$

Split $M = K_C + L_C$,

$$K_C = \sum_{r \in A \setminus B} T_{\Omega_i(C) \cup B_1 \cup \{r\}}, \qquad L_C = \sum_{r \in (\Omega_i \cup \Omega_i') \setminus \Omega_i(C)} T_{\Omega_i(C) \cup B_1 \cup \{r\}}$$

and substitute the expression $\mathbf{E}(h_C|H) = -(N - |H|)^{-1}(K_C + L_C)$, in (A.14) to get

$$\text{(A.15)} \qquad \begin{aligned} s_{i,\,j,\,k} &= \sum_{C \subset \Omega_i} (-1)^{|C|} \mathbf{E}\mathbb{D}_i T_{\Omega_i \cup A} \mathbf{E}(h_C|H) = \frac{-1}{N - |H|}(S_K + S_L), \\ S_K &= \sum_{C \subset \Omega_i} (-1)^{|C|} \mathbf{E}\mathbb{D}_i T_{\Omega_i \cup A} K_C, \quad S_L = \sum_{C \subset \Omega_i} (-1)^{|C|} \mathbf{E}\mathbb{D}_i T_{\Omega_i \cup A} L_C. \end{aligned}$$

We shall show below that $S_L = 0$. Now consider S_K. By symmetry,

$$\mathbf{E}\mathbb{D}_i T_{\Omega_i \cup A} K_C = |A \setminus B| \mathbf{E}\mathbb{D}_i T_{\Omega_i \cup A} T_{\Omega_i(C) \cup B_2}.$$

Therefore, $S_K = (j - k)S$, where, by (A.7),

$$S = \mathbf{E}\mathbb{D}_i T_{\Omega_i \cup A} \sum_{C \subset \Omega_i} (-1)^{|C|} T_{\Omega_i(C) \cup B_2} = \mathbf{E}\mathbb{D}_i T_{\Omega_i \cup A} \mathbb{D}_i T_{\Omega_i \cup B_2}.$$

We obtain $S_K = (j - k)s_{i,\,j,\,k+1}$. Finally, by (A.15) and the identity $S_L = 0$,

$$\text{(A.16)} \qquad s_{i,\,j,\,k} = \frac{(-1)}{N - |H|} S_K = \frac{(-1)(j - k)}{N - (2j + 2i - k - 1)} s_{i,\,j,\,k+1}.$$

Applying this identity several times, for increasing k, we obtain (A.10).

It remains to prove $S_L = 0$. To this end we shall show that, almost surely,

$$\text{(A.17)} \qquad \sum_{C \subset \Omega_i} (-1)^{|C|} L_C = \sum_{C \subset \Omega_i} (-1)^{|C|} \sum_{r \in \Omega_i^*(C)} T_{\Omega_i(C) \cup B_1 \cup \{r\}} = 0.$$

Here $\Omega_i^*(C) = (\Omega_i \cup \Omega_i') \setminus \Omega_i(C)$. Note that for any fixed $C \subset \Omega_i$ and $j_1 \in \Omega_i^*(C)$ there exists a unique pair D, j_2, with $D \neq C$, where $D \subset \Omega_i$ and $j_2 \in \Omega_i^*(D)$ such that

$$\text{(A.18)} \qquad \Omega_i(C) \cup \{j_1\} = \Omega_i(D) \cup \{j_2\}.$$

Namely, if $j_1 \in \Omega_i'$ then $j_1 = j'$ for some $j \in \Omega_i$ and in this case $D = C \cup \{j\}$. If $j_1 \in \Omega_i$ then necessarily $j_1 \in C$ and in this case $D = C \setminus \{j_1\}$. In both cases we have $||C| - |D|| = 1$ and therefore, $(-1)^{|C|} + (-1)^{|D|} = 0$. Hence, for every random variable $T_{\Omega_i(C) \cup B_1 \cup \{j_1\}}$ of the sum (A.17) there exists a unique counterpart $T_{\Omega_i(D) \cup B_1 \cup \{j_2\}}$ (in the same sum) satisfying (A.18). Clearly, we have

$$(-1)^{|C|} T_{\Omega_i(C) \cup B_1 \cup \{j_1\}} + (-1)^{|D|} T_{\Omega_i(D) \cup B_1 \cup \{j_2\}} = 0$$

and thus (A.17) follows.

Let us prove (5.11). Fix $A \subset \Omega_n \backslash \Omega_i$ with $|A| = j - i$. We have

$$\begin{aligned}
\sigma_{i,j}^2 &= \mathbf{E}(\mathbb{D}_i T_{\Omega_i \cup A})^2 = \mathbf{E}(\mathbb{D}_{i-1} T_{\Omega_i \cup A} - \mathbb{D}_{i-1} T_{\Omega_i(\{i\}) \cup A})^2 \\
&= \mathbf{E}(\mathbb{D}_{i-1} T_{\Omega_i \cup A})^2 + \mathbf{E}(\mathbb{D}_{i-1} T_{\Omega_i(\{i\}) \cup A})^2 - 2\mathbf{E}\mathbb{D}_{i-1} T_{\Omega_i \cup A} \mathbb{D}_{i-1} T_{\Omega_i(\{i\}) \cup A} \\
&= 2\sigma_{i-1,j}^2 - 2s_{i-1,\,j-i+1,\,j-i} = 2\left(1 + \frac{1}{N - (j+i-1)}\right)\sigma_{i-1,j}^2.
\end{aligned}$$

In the last step we used (A.16). Applying this identity several times, for decreasing i, we obtain (A.11), thus completing the proof of the lemma. □

We shall consider statistics of the form $V = \sum_{B \subset \mathscr{H}} T_B$, where $\mathscr{H}$ denotes some class of subsets B of Ω with $|B| \le n$. Denote $U_j(V) = \sum_{B \subset \mathscr{H},\,|B|=j} T_B$ and write $e_j(V) = \sigma_j^{-2} \mathbf{Var}\, U_j(V)$, for $\sigma_j^2 > 0$. Otherwise put $e_j(V) = 0$. By (A.8), random variables $U_r(V)$ and $U_k(V)$ are uncorrelated for $r \neq k$. Therefore,

$$\begin{aligned}
\mathbf{Var}\, V &= \mathbf{Var}\, U_1(V) + \cdots + \mathbf{Var}\, U_n(V) \\
&= e_1(V)\sigma_1^2 + \cdots + e_n(V)\sigma_n^2.
\end{aligned} \tag{A.19}$$

In what follows we use the formula.

$$\sum_{v=0}^{\min\{s,k\}} (-1)^v \binom{s}{v}\binom{k}{v}\binom{u}{v}^{-1} = \binom{u-s}{k}\binom{u}{k}^{-1}, \tag{A.20}$$

where the integers $s, t, u \ge 0$ [see, e.g., Zhao and Chen (1990)].

Write $r_{i,j} = \binom{N-n-i}{j-i}\binom{N-i-j}{j-i}^{-1}$.

LEMMA 2. *The formulas* (2.6) *hold true. For every* $1 \le i \le j \le n_*$, *we have*

$$\mathbf{Var}\, U_j(\mathbb{D}_i T) = \binom{n-i}{j-i} r_{i,j}\sigma_{i,j}^2 = \binom{n-i}{j-i} r_{i,j} \frac{N-j+1}{N-i-j+1} 2^i \sigma_j^2, \tag{A.21}$$

$$\mathbf{Var}\, U_j(T) \le (n_*/2)^i \mathbf{Var}\, U_j(\mathbb{D}_i T). \tag{A.22}$$

PROOF. Let us prove the first part of (2.6). By symmetry,

$$\mathbf{Var}\, U_j(T) = \binom{n}{j}\mathbf{E} T_{\Omega_j} U_j(T) \quad \text{and} \quad \mathbf{E} T_{\Omega_j} U_j(T) = \sum_{v=0}^{n-j} m_v s_{j,\,j-v}, \tag{A.23}$$

where m_v denotes the number of subsets $B \subset \Omega_n$, with $|B| = j$, satisfying $|B \cap (\Omega_n \backslash \Omega_j)| = v$. Clearly, $m_v = \binom{n-j}{v}\binom{j}{j-v}$. Therefore,

$$\mathbf{E} T_{\Omega_j} U_j(T) = \sum_{v=0}^{v_0} \binom{n-j}{v}\binom{j}{v} s_{j,\,j-v}, \qquad v_0 := \min\{n-j,\, j\}.$$

Invoking (A.9) and then using (A.20) we obtain $\mathbf{E} T_{\Omega_j} U_j(T) = r_{0,j}\sigma_j^2$. This identity in combination with (A.23) gives the first part of (2.6). The second part is trivial [cf. (A.19)].

Let us prove (A.21). We have

$$U_j(\mathbb{D}_i T) = \sum_{A\subset\Omega_n\setminus\Omega_i,\,|A|=j-i} \mathbb{D}_i T_{\Omega_i\cup A}, \qquad i \le j \le n.$$

By symmetry,

$$\mathbf{Var}\, U_j(\mathbb{D}_i T) = \binom{n-i}{j-i}\mathbf{E}\mathbb{D}_i T_{\Omega_j} U_j(\mathbb{D}_i T),$$

$$\mathbf{E}\mathbb{D}_i T_{\Omega_j} U_j(\mathbb{D}_i T) = \sum_{u=0}^{\min\{n-j,\,j-i\}} \binom{n-j}{u}\binom{j-i}{j-i-u} s_{i,\,j-i,\,j-i-u}.$$

Invoking (A.10) and then using (A.20) we obtain $\mathbf{E}\mathbb{D}_i T_{\Omega_j} U_j(\mathbb{D}_i T) = r_{i,\,j}\sigma^2_{i,\,j}$, thus proving the first identity of (A.21). The second one follows from (A.11).

The inequality (A.22) is a simple consequence of the identity

$$\frac{\mathbf{Var}\, U_j(\mathbb{D}_i T)}{\mathbf{Var}\, U_j(T)} = 2^i \frac{[j]_i[N-j+1]_i}{[n]_i[N-n]_i},$$

which follows from (2.6) and (A.21). $\square$

PROOF OF THEOREM 1. Combining (2.7) and (A.22) we obtain

$$\begin{aligned}\mathbf{E}R_k^2 &= \mathbf{Var}\, U_{k+1}(T) + \cdots + \mathbf{Var}\, U_{n_*}(T)\\ &\le (n_*/2)^{k+1}(\mathbf{Var}\, U_{k+1}(\mathbb{D}_{k+1}T) + \cdots + \mathbf{Var}\, U_{n_*}(\mathbb{D}_{k+1}T))\\ &= n_*^{1-k}2^{-1-k}\delta_{k+1}. \qquad \square\end{aligned}$$

PROOF OF (3.1). Using the identity $\sigma_J^2 = \sum_{i=1}^{n+1} T_i^2 - (n+1)\overline{T}^2$ it is easy to show that (3.1) is equivalent to the inequality $(n+1-q^{-1})\mathbf{E}T^2 \ge (n+1)\mathbf{E}\overline{T}^2$. In order to prove this inequality, it suffices to show that for every $j = 1, \dots, n_*$,

$$(n+1-q^{-1})\mathbf{Var}\, U_j(T) \ge (n+1)^{-1}\,\mathbf{Var}\, U_j(H), \qquad H = (n+1)\overline{T}. \tag{A.24}$$

Let us evaluate $\mathbf{Var}\, U_j(H)$. An application of (2.1) to $T_{(1)}, \dots, T_{(n+1)}$ gives

$$H = \sum_{j=1}^{n_*}(n+1-j)W_j, \qquad W_j = \sum_{B\subset\Omega_{n+1},\,|B|=j} T_B.$$

Proceeding as in the proof of (A.23), we obtain

$$\mathbf{E}W_j^2 = \binom{n+1}{j}\binom{N-n-1}{j}\binom{N-j}{j}^{-1}\sigma_j^2.$$

Therefore, we have an explicit formula for $\mathbf{Var}\, U_j(H) = (n+1-j)^2\mathbf{E}W_j^2$. Invoking (2.6) we obtain an explicit formula for the left-hand side of (A.24) as well. Now simple arithmetic proves (A.24). $\square$

PROOF OF PROPOSITION 2. Under condition (i) we have $s^2 - \tilde{\sigma}^2 = o(1)$ as $n_*, N \to \infty$. In particular, $s^2 = O(1)$. Let $V^2 = \sigma_J^2(U_1(T))$ denote the jackknife variance estimator of $U_1(T)$, the linear part of T. In order to prove (3.2) it suffices to show that as $n_*, N \to \infty$,

$$q(\sigma_J^2 - V^2) = o_P(1) \quad \text{and} \quad qV^2 - s^2 = o_P(1).$$

The first relation is implied by the smoothness condition $\delta_2 = o(1)$. The second relation follows by the (weak) law of large numbers [use (3.3) and the fact that $s^2 = O(1)$]. □

REFERENCES

BABU, G. J. and BAI, Z. D. (1996). Mixtures of global and local Edgeworth expansions and their applications. *J. Multivariate Anal.* **59** 282–307.

BABU, G. J. and SINGH, K. (1985). Edgeworth expansions for sampling without replacement from finite populations. *J. Multivariate Anal.* **17** 261–278.

BENTKUS, V., GÖTZE, F. and VAN ZWET, W. R. (1997). An Edgeworth expansion for symmetric statistics. *Ann. Statist.* **25** 851–896.

BERTAIL, P. (1997). Second-order properties of an extrapolated bootstrap without replacement under weak assumptions. *Bernoulli* **3** 149–179.

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

BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1997). Resampling fewer than n observations: gains, losses, and remedies for losses. *Statist. Sinica* **7** 1–31.

BICKEL, P. J. and YAHAV, J. A. (1988). Richardson extrapolation and the bootstrap. *J. Amer. Statist. Assoc.* **83** 387–393.

BLOZNELIS, M. (2000). Empirical Edgeworth expansion for finite population statistics. Preprint 2000-14 Faculty of Mathematics and Informatics, Vilnius Univ.

BLOZNELIS, M. and GÖTZE, F. (1999a). One term Edgeworth expansion for finite population U-statistics of degree two. *Acta Applicandae Math.* **58** 75–90.

BLOZNELIS, M. and GÖTZE, F. (1999b). An Edgeworth expansion for finite population symmetric statistics. Preprint 00-042 SFB 343 Bielefeld Univ. Available at www.mathematik.uni-bielefeld.de/sfb343/Welcome.html.

BLOZNELIS, M. and GÖTZE, F. (2000). An Edgeworth expansion for finite population U-statistics. *Bernoulli* **6** 729–760.

BOOTH, J. G., BUTLER, RONALD W. and HALL, P. (1994). Bootstrap methods for finite populations. *J. Amer. Statist. Assoc.* **89** 1282–1289.

BOOTH, J. G. and HALL, P. (1993). An improvement of the jackknife distribution function estimator. *Ann. Statist.* **21** 1476–1485.

CHAO, M. T. and LO, S. H. (1985). A bootstrap method for finite population. *Sankhyā Ser. A* **47** 399–405.

CHEN, J. and SITTER, R. R. (1993). Edgeworth expansion and the bootstrap for stratified sampling without replacement from finite population. *Canad. J. Statist.* **21** 347–357.

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

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.

GROSS, S. (1980). Median estimation in sample surveys (1980). *Proceedings of the Section on Survey Research Methods.* 181–184. Amer. Statist. Assoc.

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

HELMERS, R. and WEGKAMP, M. (1998). Wild bootstrapping in finite populations with auxiliary information. *Scand. J. Statist.* **25** 383–399.

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

KARLIN, S. and RINOTT, Y. (1982). Applications of ANOVA type decompositions for comparisons of conditional variance statistics including jackknife estimates. *Ann. Statist.* **10** 485–501.

KOKIC, P. N. and WEBER, N. C. (1990). An Edgeworth expansion for U-statistics based on samples from finite populations. *Ann. Probab.* **18** 390–404.

MILLER, R. G. (1974). The jackknife—a review. *Biometrica* **61** 1–15.

PARR, W. C. (1985). Jackknifing differentiable statistical functionals. *J. Roy. Statist. Soc. Ser. B* **47** 56–66.

POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050.

PUTTER, H. and VAN ZWET, W. R. (1998). Empirical Edgeworth expansions for symmetric statistics. *Ann. Statist.* **26** 1540–1569.

RUBIN, H. and VITALE, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8** 165–170.

SHAO, J. (1989). The efficiency and consistency of approximations to the jackknife variance estimators. *J. Amer. Statist. Assoc.* **84** 114–119.

SHAO, J. and WU, C.-F. J. (1989). A general theory for jackknife variance estimation. *Ann. Statist.* **17** 1176–1197.

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

VITALE, R. A. (1992). Covariances of symmetric statistics. *J. Multivariate Anal.* **41** 14–26.

WU, C.-F. J. (1990). On the asymptotic properties of the jackknife histogram. *Ann. Statist.* **18** 1438–1452.

ZHAO, L. C. and CHEN, X. R. (1990). Normal approximation for finite-population U-statistics. *Acta Math. Appl. Sinica* **6** 263–272.

DEPARTMENT OF MATHEMATICS
VILNIUS UNIVERSITY
NAUGARDUKO 24
VILNIUS 2006
LITHUANIA
E-MAIL: Mindaugas.Bloznelis@maf.vu.lt

FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄT BIELEFELD
POSTFACH 100131
33501 BIELEFELD 1
GERMANY
E-MAIL: goetze@mathematik.uni-bielefeld.de