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

M. Bloznelis[1] and F. Götze

Vilnius University and Bielefeld University

Abstract. We study orthogonal decomposition of symmetric statistics based on samples drawn without replacement from finite populations. Under very mild smoothness conditions the first k terms of the decomposition provide stochastic expansion with remainder $O(N^{-k/2})$. Assuming that the linear part of the decomposition is nondegenerate we establish one term Edgeworth expansion of the distribution function of a general symmetric statistic. Several applications are discussed: second order asymptotics of the jackknife histogram, consistency of the jackknife estimator of variance, Efron–Stein inequality.

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) became 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 (Efron and Stein 1980, Karlin and Rinott 1982, Vitale 1992) and provides a natural framework for the first and the 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

---

1991 *Mathematics Subject Classification.* Primary 62E20; 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.

[1]Research supported by Alexander von Humboldt Foundation

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

random samples. We start with an overview of the orthogonal decomposition of general symmetric statistics based on simple random samples. In the following the decomposition is used to establish stochastic and asymptotic expansions of symmetric statistics. Important applications are: the second order approximation to the distribution of jackknife histogram (Wu 1990, Shao 1989, Politis and Romano 1994, Bickel, Götze and van Zwet 1997), Richardson extrapolation (Bickel and Yahav 1988, Booth and Hall 1993, Bertail 1997) and the bootstrap of finite population statistics (Gross 1980, Bickel and Freedman 1984, Chao and Lo 1985, Babu and Singh 1985, Booth, Butler and Hall 1994). Another applications include the Efron–Stein inequality and consistency of the jackknife variance estimator for 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 $\mathcal{X} = \{x_1, \dots, x_n\}$ 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

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

represents T by 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 $1 \le i_1 < \dots < i_k \le N$ and $1 \le j_1 < \dots < j_r \le N$ such that $r < k$. It is easy to verify that such decomposition is unique.

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

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

They can be evaluated by means of the recursive relation

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

We derive (2.3) in the Appendix. Here, the coefficients

$$d_{k,j} = \prod_{r=j}^{k-1} \frac{n-r}{n-r-j}, \qquad 1 \le j \le k-1, \quad k = 2, 3, \dots, N \tag{2.4}$$

and $d_{N,N} = 1$. The coefficients $M_{k,j}$, $1 \le j \le k$ satisfy the recursive relation

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

and $M_{k,k} = 1$. A simple calculation gives

$$g_1(x_1) = \frac{n-1}{n-N}\, h_1(x_1),$$
$$g_2(x_1,x_2) = \frac{n-2}{n-N}\,\frac{n-3}{n-N-1}\,\Big(h_2(x_1,x_2) - \frac{n-1}{n-2}\,\big(h_1(x_1)+h_1(x_2)\big)\Big).$$

Orthogonal decomposition of finite population U–statistics was constructed earlier by Zhao and Chen (1990), in their proof of the asymptotic normality of U–statistics.

For $1 \le j \le N$, denote

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

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

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

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

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

The proof of (2.5) and (2.6) is given in Lemmas 1 and 2 in the Appendix.

The identity (2.6) has an interesting consequence. 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, it entails a duality property, formulated in Proposition 1 below.

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

Proposition 1. *For $j \le N_*$ we have*

$$(2.8) \qquad U_j \equiv U'_j, \qquad where \qquad U'_j = (-1)^j \sum_{1 \le i_1 < \cdots < i_j \le N_*} g_j(X'_{i_1}, \dots, X'_{i_j}).$$

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

The proposition says that, in fact, T is a function of N_* random variables.

Proof of Proposition 1. For the linear statistic U_1 the identity (2.8) is a consequence of $\mathbf{E}U_1 = 0$. For $j = 2, \dots, N_*$, this identity follows from (2.2).

.

3. Stochastic and asymptotic expansions

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 a sufficiently precise approximation. To justify such approximation we introduce appropriate smoothness conditions.

Denote

$$D^j T = t(X_1, \dots, X_N) - t(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_N, X'_j), \quad X'_j = X_{N+j}.$$

A higher order difference operations are defined recursively

$$D^{j_1,j_2} T = D^{j_2}\big(D^{j_1} T\big), \qquad D^{j_1,j_2,j_3} T = D^{j_3}\big(D^{j_2}(D^{j_1} T)\big), \dots .$$

They are symmetric, i.e., $D^{j_1,j_2}T = D^{j_2,j_1}T$, etc. Given $k < N_*$ write

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

Theorem 1. *For $1 \le k < N_*$, we have*

$$(3.1) \qquad T = \mathbf{E}T + U_1 + \cdots + U_k + R_k, \qquad with \qquad \mathbf{E}R_k^2 \le N_*^{-(k-1)} \delta_{k+1}.$$

The proof of Theorem 1 is given in the Appendix.

Although the theorem applies to an arbitrary sample size N, we shall use it as an asymptotic result.

When speaking about the finite population asymptotics we shall assume a sequence of populations $\mathcal{X}_r = \{x_{r,1}, \dots, x_{r,n_r}\}$, 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 $\mathcal{X}_r$. 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

$$\delta_{k+1} = O(N_*^{-1}) \qquad \text{as} \qquad N_*, n \to \infty, \tag{3.2}$$

for some k. Clearly, (3.2) is a smoothness condition. It implies the validity of the stochastic expansion (3.1) with the remainder $R_k = O_P(N_*^{-k/2})$. The condition (3.2) is easy to handle. Below, we verify this condition for two typical examples: U–statistics and smooth functions of multivariate sample means.

In the remaining part of the 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),

$$s^2 = \tau^2 \sigma_1^2 n/(n-1), \qquad \tau^2 = npq, \qquad p = N/n, \quad q = 1 - p.$$

In Proposition 2 below we formulate sufficient conditions for the asymptotic normality.

Proposition 2. *Assume $\tilde{\sigma}$ remains bounded away from zero as $n \to \infty$. Assume that $\delta_2 = o_P(1)$ as $N_*, n \to \infty$. Then $\tilde{\sigma} - s = o(1)$. Suppose, in addition, that for every $\varepsilon > 0$,*

$$\mathbf{E}\, g_1^2(X_1) \sigma_1^{-2} \mathbb{I}_{|g_1(X_1)| > \varepsilon \tau \sigma_1} = o(1) \qquad \textit{as} \quad N_*, n \to \infty. \tag{3.3}$$

Then $\tilde{\sigma}^{-1}(T - \mathbf{E}T)$ and $s^{-1}(T - \mathbf{E}T)$ are asymptotically standard normal.

Proof of Proposition 2. In view of Theorem 1, the condition $\delta_2 = o_P(1)$ implies the validity of short stochastic expansion $T = \mathbf{E}T + U_1 + o_P(1)$. We also have $\tilde{\sigma}^2 - s^2 = o(1)$. Therefore, the linear part dominates the statistic and it suffices to prove the asymptotic normality of $s^{-1}U_1$. The asymptotic normality is ensured by a Lindeberg type Erdős–Rényi condition (3.3), see Erdős and Rényi (1959).

Assuming that (3.2) holds, with $k = 2$, 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 the order $O(N_*^{-1})$. Note that $U_1 + U_2$ is a U–statistic of degree two. Bloznelis and Götze (1999a) showed that one term asymptotic expansion

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

approximates the distribution function $\mathbf{P}\{U_1 + U_2 \leq x\,s\}$ with the reminder $O(N_*^{-1})$. 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 the 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 provide bounds $o(N_*^{-1/2})$ and $O(N_*^{-1})$ for

$$\Delta := \sup_x \bigl|F(x) - G(x)\bigr|, \qquad \text{where} \qquad F(x) = \mathbf{P}\{T \leq \mathbf{E}T + x\tilde{\sigma}\},$$
$$\Delta_1 := \sup_x \bigl|F_1(x) - G(x)\bigr|, \qquad \text{where} \qquad F_1(x) = \mathbf{P}\{T \leq \mathbf{E}T + x\sigma_1\tau\}.$$

In order to establish the validity of an Edgeworth expansion we need to impose an appropriate smoothness condition. It is the non-lattice condition in the case of the remainder $o(N_*^{-1/2})$ and it is a Cramér type condition if the remainder is $O(N_*^{-1})$. We shall impose these conditions on the linear part of the statistic.

Given $0 < a < b$ write $\rho(a, b) = \sup_{a<|t|<b} |\varphi(t)|$, where $\varphi(t)$ denotes the characteristic function of the random variable $\sigma_1^{-1} g_1(X_1)$. We shall say that the linear part is asymptotically non–lattice, if for every $\varepsilon > 0$ and every $B > 0$,

$$\liminf_{N_*, n\to\infty} \rho(\varepsilon, B) < 1. \tag{3.4}$$

A more stringent smoothness condition is a Cramér type condition

$$\liminf_{N_*, n\to\infty} \rho(\varepsilon, \tau) < 1. \tag{3.5}$$

Note that $\tau \to \infty$ as $N_*, n \to \infty$. Write

$$\beta_s = \mathbf{E}\sigma_1^{-s}|g_1(X_1)|^s, \quad \gamma_s = \mathbf{E}|N_*^{3/2}\gamma_2(X_1, X_2)|^s, \quad \zeta_s = \mathbf{E}|N_*^{5/2}\gamma_3(X_1, X_2, X_3)|^s.$$

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

(i) Assume that (3.4) 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

$$\Delta = o(N_*^{-1/2}) \qquad \textit{and} \qquad \Delta_1 = o(N_*^{-1/2}), \qquad \textit{as} \quad N_*, n \to \infty.$$

(ii) Assume that (3.5) holds, $\delta_4 = O(N_*^{-1})$ and, the moments β_4, γ_4, ζ_2 are bounded as $N_*, n \to \infty$. Then

$$\Delta = O(N_*^{-1}) \qquad \textit{and} \qquad \Delta_1 = O(N_*^{-1}) \qquad \textit{as} \quad N_*, n \to \infty.$$

Proof of Theorem 2. Note that either of the conditions (i) and (ii) implies

$$\tilde{\sigma}^2 = s^2 + O(N_*^{-1}) = \tau^2 \sigma_1^2 + O(N_*^{-1}). \tag{3.6}$$

Therefore, it suffics to construct bounds for Δ_1

Let us prove $\Delta_1 = o(N_*^{-1/2})$. In the case of U–statistics of degree two ($R_2 \equiv 0$), this bound is proved in Bloznelis and Götze (1999 b). A passage to the general case can be made by using a Slutzky type argument. Indeed, 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 (1999 c).

Note that (3.6) implies $\tau\sigma_1/\tilde{\sigma} = 1 + O(N_*^{-1})$. Therefore, we can replace G by G_0, where

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

Corollary. *Theorem 2 remains valid if we replace Δ by $\Delta_0 = \sup_x |F(x) - G_0(x)|$.*

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 approximation to symmetric statistics of i.i.d. observations.

In order to construct one term Edgeworth expansion G (respectively G_0) one needs to evaluate the parameters σ_1, α, κ (respectively $\tilde{\sigma}$, α_0, κ_0). For many statistics it reduces to routine calculations, see examples below. Another possible way is to substitute consistent estimators of these parameters. This question will be addressed elsewhere.

Earlier results on Edgeworth expansions for nonlinear asymptotically normal finite population statistics were obtained by Babu and Singh (1985), Babu and Bai (1996). They considered statistics which can be approximated by smooth functions of multivariate sample means and derived asymptotic expansions from corresponding expansions for multivariate sample means via delta method. This approach, though conceptually simpler, focuses on particular class of statistics (smooth functions of multivariate sample means). Furthermore, it often requires a bit restrictive Cramér type smoothness condition imposed on the underlying multivariate sample mean rather than on the linear part of the statistic itself, see also remark on a Cramér type condition in the Example 2 below.

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 $\mathcal{X} = \{x_1, \dots, x_n\}$. Define U–statistic

$$U = \sum_{1 \le i_1 < \cdots < i_m \le N} \varphi(X_{i_1}, \dots, X_{i_m}) \tag{3.7}$$

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

Write Hoeffding's decomposition for 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 (3.7) can be written in such form. Substituting these expressions in (3.7) we obtain Hoeffding's decomposition of U,

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

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

$$\alpha = \sigma_0^{-3} \mathbf{E}\tilde{g}_1^2(X_1), \qquad \kappa = \frac{(m-1)N}{N-1} \, \frac{q}{\sigma_0^3} \, \mathbf{E}\tilde{g}_2(X_1, X_2)\tilde{g}_1(X_1)\tilde{g}_1(X_2).$$

In order to estimate the moments (of differences) δ_k, we invoke variance formulas (2.6) and (5.19), see below. A straightforward calculation gives

$$\delta_k \le c(m) N_*^{k-2} \big(\mathbf{Var} U_k^2 + \cdots + \mathbf{Var} U_m^2\big), \tag{3.8}$$

where $c(m)$ denotes a constant which depends only on m. Variances $\mathbf{Var}U_j$, $k \le j \le m$, can obtained from (2.6), with $\sigma_j^2 = \binom{N-j}{m-j}^2 \mathbf{E}\tilde{g}_k^2(X_1, \dots, X_j)$. Note that (3.8) implies $N_*^{-(k-1)} \delta_{k+1} \le c(m) \mathbf{E} R_k^2$. That is, for U–statistics of degree m, the inequality (3.1) is optimal up to the constant $c(m)$.

Example 2. **Smooth functions of multivariate sample means.** Assume that $\mathcal{X} \subset \mathbb{R}^k$ 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 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

$$\sigma_h^2 = \mathbf{E}(h^{(1)}(0)X_1)^2, \qquad \alpha_h = \sigma_h^{-3}\mathbf{E}(h^{(1)}(0)X_1)^3,$$
$$\kappa_h = \sigma_h^{-3}\mathbf{E}(h^{(1)}(0)X_1)(h^{(1)}(0)X_2)(h^{(2)}(0)X_1X_2)$$

and write

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

Here $h^{(s)}(y)$, $s = 1, 2$, denote the s–th derivative of h at the point $y \in \mathbb{R}^k$. We write $h^{(s)}(y)z_1 \ldots z_s$ 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 shows that α_h and $q\kappa_h$ provides sufficiently precise approximations to α and $\tau^2\kappa$ so that G can be replaced by G_h with impunity. Furthermore, it follows from Theorem 2 that $G_h(x)$ provides one term asymptotic expansion for the distribution function $\mathbf{P}\{T - \mathbf{E}T \le x\sigma_h\}$. The verification of the conditions of the theorem reduces to routine calculations. We skip most technical details and focus on the smoothness conditions only.

Let $\|h^{(s)}\|$ 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 an element $z_i = (z_{i,1}, \ldots, z_{i,k}) \in \mathbb{R}^k$. We say that the s–th derivative is bounded if $\|h^{(s)}\|_\infty := \sup_y \|h^{(s)}(y)\|$ is finite.

Assuming that $\mathbf{E}|X_1|^4$ and $\|h^{(s)}\|_\infty$ remain bounded as $N_*, n \to \infty$ one may show that $g_1(X_1) = \tau^{-1}h^{(1)}(0)X_1 + O_P(N_*^{-3/2})$. An important consequence of this representation is that it suffices to verify the conditions (3.4) (or (3.5)) for the distribution of $\sigma_h^{-1}h^{(1)}(0)X_1$.

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 $j = 1, 2, \ldots,$

$$\delta_j(T) \le 2^j \,\frac{n^j}{[n]_j}\, \|h^{(j)}\|_\infty (\mathbf{E}|X_1|^2)^j \tau^{-2}, \tag{3.9}$$

where $[n]_j = n(n-1)\ldots(n-j+1)$. Let $\tau, \tau_1, \tau_2, \ldots$ 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 $f(x+y) - f(x) = \mathbf{E}_\tau f^{(1)}(x+\tau y)y$. Here $\mathbf{E}_\tau$ denotes the conditional expectation given all the random variables but τ. Split

$\overline{X} = \overline{X}_{(j)} + N^{-1}(X_1 + \cdots + X_j)$ and write $u_i = N^{-1}(X_i - X_i')$. By the mean value formula

$$\mathbb{D}_j h(\overline{X}) = \mathbf{E}_{\tau_1} \ldots \mathbf{E}_{\tau_j} h^{(1)}(\overline{X}_{(j)} + \tau_1 u_1 + \cdots + \tau_j u_j) u_1 \cdots u_j, \tag{3.10}$$

Furthermore, invoking the simple bound $\mathbf{E}|X_{i_1}|^2 \ldots |X_{i_j}|^2 \le n^j/[n]_j(\mathbf{E}|X_1|^2)^j$, for $i_1 < \cdots < i_j$, we obtain $\mathbf{E}|u_1 \ldots u_j|^2 \le 2^j n^j/[n]_j N^{-2j}(\mathbf{E}|X_1|^2)^j$. The last inequality in combination with (3.10) implies (3.9).

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

4. Applications

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

$$\sigma_J^2 = \sigma_J^2(T) = \sum_{j=0}^{N} (T_{(j)} - \overline{T})^2, \qquad \overline{T} = \frac{1}{N+1} \sum_{j=0}^{N} T_{(j)},$$

where we write $T_{(j)} = t(X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N, X_{N+1})$, for $1 \le j \le N$, and put $T_{(0)} = t(X_1, \ldots, X_N)$.

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 e.g., Miller (1974), Parr (1985) and Shao and Wu (1989) where in the later paper several "smoothness" conditions are discussed.

In the case where sample $X_1, \ldots, X_N$ is drawn without replacement from the population $\mathcal{X} = \{x_1, \ldots, x_n\}$ we need a finite population correction. Write

$$\sigma_{FJ}^2 = q\sigma_J^2, \qquad \text{where} \qquad q = (n - N)/n.$$

For linear statistic T we have $\mathbf{E}\sigma_{FJ}^2 = \mathbf{Var}\, T$. For a general symmetric statistic T the Efron–Stein inequality holds

$$\mathbf{E}\sigma_{FJ}^2 \ge \mathbf{Var}\, T. \tag{4.1}$$

Moreover under very mild smoothness condition on T we prove consistency of σ_{FP}^2.

Proposition 3. *Assume that*

(i) for some $0 < c_1 < c_2 < \infty$, we have $c_1 \le \tilde{\sigma} \le c_2$ and $\delta_2(T) = o(1)$ as $N_*, n \to \infty$;

(ii) condition (3.3) holds.
Then, for every $\delta > 0$, we have

$$\mathbf{P}\{|\sigma_{FJ}^2 - \tilde{\sigma}^2| > \delta\} = o(1) \qquad as \qquad N_*, n \to \infty. \tag{4.2}$$

Combining Propositions 2 and 3 we conclude that the conditions of Proposition 3 are sufficient for the asymptotic normality of $(T - \mathbf{E}T)/\sigma_{FJ}$.

Proofs of (4.1) and Proposition 3 are given in the Appendix.

Resampling. Let $Y_1, Y_2, \dots$ be a sequence of independent identically distributed observations. Given $1 < N < n$, let $T_N = t_N(X_1, \dots, X_N)$ be a statistic based on simple random sample $X_1, \dots, X_N$ drawn without replacement from the population $\mathcal{Y} := \{Y_1, \dots, Y_n\}$. The jackknife histogram (Wu, 1990) and $\binom{n}{m}$ bootstrap (Politis and Romano, 1994; Bickel, Götze and van Zwet, 1997) studies asymptotic properties of the conditional distribution of T_N given $Y_1, \dots Y_n$. Let $F(x|\mathcal{Y})$ denotes the conditional distribution function of T_N given $Y_1, \dots, Y_N$. Theorem 2 provides an second order approximation to $F(x|\mathcal{Y})$ as $N_*, n \to \infty$.

5. Appendix

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

Denote $\Omega_k = \{1, \dots, k\}$, for $k = 1, 2, \dots$, and put $\Omega = \Omega_n$. Given a statistic $V = V(X_1, \dots, X_n)$ write

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

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

Proof of (2.3)**.** Introduce random variables Q_A, for $A \subset \Omega_N$, with $|A| \geq 1$. For $|A| = 1$, we put $Q_A = \mathbf{E}(T\big|A)$. For $|A| = 2, 3, \dots, N$, we define Q_A recursively as follows. Given $A \subset \Omega_N$, with $|A| = k$, write

$$Q_A = \mathbf{E}(T\big|A) - d_{k,k-1} \sum_{B \subset A,\, |B| = k-1} Q_B - \dots - d_{k,1} \sum_{B \subset A,\, |B| = 1} Q_B, \tag{5.1}$$

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

$$\mathbf{E}(Q_A|B) = 0, \qquad |B| < |A|. \tag{5.2}$$

A straightforward calculation gives (2.4). For $k = N$, the relation (5.1) yields

$$T = \mathbf{E}(T\big|\Omega_N) = \sum_{B \subset \Omega_N,\, |B| \geq 1} T_B, \qquad T_B = d_{N,|B|} Q_B, \tag{5.3}$$

where, for every $B \subset \Omega_N$, the random variable T_B satisfies

$$\mathbf{E}(T_B|C) = 0, \qquad \text{for} \quad |C| < |B|, \tag{5.4}$$

for every $C \subset B$, see (5.2). In Lemma 1 we extend the identity (5.4) to arbitrary $B, C \subset \Omega_N$ satisfying $|C| < |B|$. Denoting

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

we obtain (2.1) from (5.3) and (2.2) from (5.4). Finally, invoking a simple combinatoric calculation we derive (2.3) from (5.1) and (5.3).

Before to formulate Lemmas 1 and 2 we introduce some notation. Recall that $(X_1, \dots, X_n)$ denotes a random permutation of $(x_1, \dots, x_n)$. 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 $T_\emptyset = 0$. 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 $i' = N + i \notin A$. In this case we write $D^i T_A = T_A - T_{A(\{i\})}$ if $i \in A$ and put $D^i T_A = 0$ otherwise. Here $A(\{i\}) = (A \setminus \{i\}) \cup \{i'\}$. A higher order differences $\mathbb{D}_i = D^i \dots D^1$ 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 $\Omega_i \nsubseteq A$. If $\Omega_i \subseteq A$, we can write $A = \Omega_i \cup B$, for some $B \subset \Omega_N \setminus \Omega_i$. In this case we have

$$\mathbb{D}_i T_{\Omega_i \cup B} = \sum_{C \subset \Omega_i} (-1)^{|C|} T_{\Omega_i(C) \cup B}, \tag{5.5}$$

$$\Omega_i(C) = (\Omega_i \setminus C) \cup C', \qquad C' = \{l' : l \in C\}, \quad l' = l + N.$$

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

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, $A, B \subset \Omega_N \setminus \Omega_i$ we write

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

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

Lemma 1. *The following identities hold*

$$\mathbf{E}(T_G|H) = 0, \quad \textit{for every} \quad G, H \subset \Omega \qquad \textit{with} \quad |H| < |G|, \tag{5.6}$$

$$s_{j,k} = \frac{(-1)^{j-k}}{\binom{n-j}{j-k}} \sigma_j^2, \qquad 0 \le k \le j \le N, \tag{5.7}$$

$$s_{i,j,k} = \frac{(-1)^{j-k}}{\binom{n-j-2i}{j-k}} \sigma_{i,i+j}^2, \quad 0 \le k \le j \le N - i, \tag{5.8}$$

$$\sigma_{i,j}^2 = \frac{n - j + 1}{n - j - i + 1} 2^i \sigma_j^2, \quad i \le j \le N. \tag{5.9}$$

Proof of Lemma 1. We start with an auxiliary identity (5.10). Fix $C, D \subset \Omega$ such that $1 \le |C| \le |D|$ and $|C \setminus D| = 1$. Denote $C_1 = C \cap D$. We have

$$\mathbf{E}(T_C \,\big|D) = \frac{1}{n-|D|} \sum_{i \in \Omega \setminus D} T_{\{i\} \cup C_1} = \frac{-1}{n-|D|} \sum_{i \in D \setminus C} T_{\{i\} \cup C_1}, \tag{5.10}$$

since, $(\Omega \setminus D) \cup (D \setminus C) = \Omega \setminus C_1$ and, by (5.4),

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

Let us prove (5.6). For $H \subset G \subset \Omega_N$, (5.6) follows from (5.4). By symmetry, it still holds for $H \subset G \subset \Omega$. For $|H \setminus G| = k$, we prove (5.6) by induction. Assume that (5.6) holds for every $k \le r$. Given $G, H \subset \Omega$, with $|H \setminus G| = r+1$, fix $a \in G \setminus H$ and denote $G_a = G \setminus \{a\}$. 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 (5.10) 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 (5.6), with $k = r + 1$.

Let us prove (5.7). Given $A, B \subset \Omega$, with $|A| = |B| = j \ge 1$ and $|A \cap B| = k < j$, fix $i_a \in A \setminus B$ and denote $A_1 = A \setminus \{i_a\}$. An application of (5.10) gives

$$\begin{aligned} s_{j,k} = \mathbf{E}T_B\mathbf{E}(T_A|A_1 \cup B) &= \frac{-1}{n-(2j-k-1)} \sum_{i \in B \setminus A} \mathbf{E}T_B T_{\{i\} \cup A_1} \\ &= \frac{(-1)(j-k)}{n-(2j-k-1)}\, s_{j,k+1}, \end{aligned} \tag{5.11}$$

where the last identity follows by symmetry. Applying (5.11) several times, for increasing k, we obtain (5.7).

Let us prove (5.8). Let A, B be as above and assume in addition that $A, B \subset \Omega_N \setminus \Omega_i$. Fix $i_a \in A \setminus B$ and $i_b \in B \setminus A$. Denote $B_1 = B \setminus \{i_b\}$ and $B_2 = B_1 \cup \{i_a\}$. It follows from (5.5) that

$$s_{i,j,k} = \sum_{C \subset \Omega_i} (-1)^{|C|} \mathbf{E}\mathbb{D}_i T_{\Omega_i \cup A} h_C, \qquad h_C = T_{\Omega_i(C) \cup B}. \tag{5.12}$$

Write $H = A \cup \Omega_i \cup \Omega_i' \cup B_1$, where $\Omega_i' = \{1', \ldots, i'\}$. By (5.10),

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

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 $\mathbf{E}(h_C \mid H) = -(n - |H|)^{-1}(K_C + L_C)$, in (5.12) to get

$$s_{i,j,k} = \sum_{C \subset \Omega_i} (-1)^{|C|} \mathbf{E}\mathbb{D}_i T_{\Omega_i \cup A} \mathbf{E}(h_C \mid H) = \frac{-1}{n - |H|}(S_K + S_L), \tag{5.13}$$

$$S_K = \sum_{C \subset \Omega_i} (-1)^{|C|} \mathbf{E}\mathbb{D}_i T_{\Omega_i \cup A} K_C, \qquad S_L = \sum_{C \subset \Omega_i} (-1)^{|C|} \mathbf{E}\mathbb{D}_i T_{\Omega_i \cup A} L_C.$$

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 (5.5),

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

We obtain $S_K = (j - k)s_{i,j,k+1}$. Finally, by (5.13) and the identity $S_L = 0$,

$$s_{i,j,k} = \frac{(-1)}{n - |H|} S_K = \frac{(-1)(j - k)}{n - (2j + 2i - k - 1)} s_{i,j,k+1}. \tag{5.14}$$

Applying this identity several times, for increasing k, we obtain (5.8).

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

$$\sum_{C \subset \Omega_i} (-1)^{|C|} L_C = \sum_{C \subset \Omega_i} (-1)^{|C|} \sum_{r \in \Omega_i^*(C)} T_{\Omega_i(C) \cup B_1 \cup \{r\}} = 0. \tag{5.15}$$

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

$$\Omega_i(C) \cup \{j_1\} = \Omega_i(D) \cup \{j_2\}. \tag{5.16}$$

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 $\big||C| - |D|\big| = 1$ 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 (5.15) there exists a unique counterpart $T_{\Omega_i(D) \cup B_1 \cup \{j_2\}}$ (in the same sum) satisfying (5.16). Clearly, we have

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

and thus, (5.15) follows.

Let us prove (5.9). Fix $A \subset \Omega_N \setminus \Omega_i$ with $|A| = j - i$. We have

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

In the last step we used (5.14). Applying this identity several times, for decreasing i, we obtain (5.9) thus completing the proof of the lemma.

We shall consider statistics of the form: $V = \sum_{B \subset \mathcal{H}} T_B$, where $\mathcal{H}$ denotes some class of subsets B of Ω with $|B| \le N$. Denote $U_j(V) = \sum_{B \subset \mathcal{H},\, |B| = j} T_B$ 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 (5.6), random variables $U_r(V)$ and $U_k(V)$ are uncorrelated, for $r \ne k$. Therefore,

$$\mathbf{Var}\, V = \mathbf{Var}\, U_1(V) + \cdots + \mathbf{Var}\, U_N(V) = e_1(V)\sigma_1^2 + \cdots + e_N(V)\sigma_N^2 \tag{5.17}$$

In what follows we use the formula

$$\sum_{v=0}^{\min\{s,k\}} (-1)^v \binom{s}{v}\binom{k}{v}\binom{u}{v}^{-1} = \binom{u-s}{k}\binom{u}{k}^{-1}, \tag{5.18}$$

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

For $0 \le i \le j \le N$, write $r_{i,j} = \binom{n-N-i}{j-i}\binom{n-i-j}{j-i}^{-1}$.

Lemma 2. *The formulas (2.6) holds true. For every $1 \le i \le j \le N_*$, we have*

$$\mathbf{Var} U_j(\mathbb{D}_i T) = \binom{N-i}{j-i} r_{i,j}\sigma_{i,j}^2 = \binom{N-i}{j-i} r_{i,j}\, \frac{n-j+1}{n-i-j+1}\, 2^i \sigma_j^2, \tag{5.19}$$

$$\mathbf{Var} U_j(T) \le (N_*/2)^i \mathbf{Var} U_j(\mathbb{D}_i T). \tag{5.20}$$

Proof of Lemma 2. Let us prove the first part of (2.6). By symmetry,

$$\mathbf{Var} U_j(T) = \binom{N}{j} \mathbf{E} T_{\Omega_j} U_j(T) \quad \text{and} \quad \mathbf{E} T_{\Omega_j} U_j(T) = \sum_{v=0}^{N-j} m_v s_{j,j-v}, \tag{5.21}$$

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

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

Invoking (5.7) and then using (5.18) we obtain $\mathbf{E}T_{\Omega_j}U_j(T) = r_{0,j}\sigma_j^2$. This identity in combination with (5.21) gives the first part of (2.6). The second part is trivial, cf (5.17).

Let us prove (5.19). We have

$$U_j(\mathbb{D}_iT) = \sum_{A\subset\Omega_N\setminus\Omega_i,\,|A|=j-i} \mathbb{D}_iT_{\Omega_i\cup A}, \qquad i\le j\le N.$$

By symmetry,

$$\begin{aligned}\mathbf{Var}U_j(\mathbb{D}_iT) &= \binom{N-i}{j-i}\mathbf{E}\mathbb{D}_iT_{\Omega_j}U_j(\mathbb{D}_iT),\\ \mathbf{E}\mathbb{D}_iT_{\Omega_j}U_j(\mathbb{D}_iT) &= \sum_{u=0}^{\min\{N-j,j-i\}}\binom{N-j}{u}\binom{j-i}{j-i-u}s_{i,j-i,j-i-u}.\end{aligned}$$

Invoking (5.8) and then using (5.18) we obtain $\mathbf{E}\mathbb{D}_iT_{\Omega_j}U_j(\mathbb{D}_iT) = r_{i,j}\sigma_{i,j}^2$, thus, proving the first identity of (5.19). The second one follows from (5.9).

The inequality (5.20) is a simple consequence of the identity

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

which follows from (2.6) and (5.19).

Proof of Theorem 1. Combining (2.7) and (5.20) we obtain

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

Proof of (4.1). 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 (4.1) 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 $j=1,\dots,N_*$,

$$(5.22)\qquad (N+1-q^{-1})\mathbf{Var}\,U_j(T)\ge(N+1)^{-1}\mathbf{Var}U_j(H), \qquad H=(N+1)\overline{T}.$$

Let us evaluate $\mathbf{Var}U_j(H)$. An application of (2.1) to $T_{(0)},\dots,T_{(N)}$ 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 proof of (5.21) we obtain

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

This yields 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 (5.22) as well. Now a simple arithmetics proves (5.22).

Proof of Proposition 3. 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 (4.2) it suffices to show that as $N_*, n \to \infty$

$$q(\sigma_J^2 - V^2) = o_P(1) \qquad \text{and} \qquad qV^2 - s^2 = o_P(1) \qquad \text{as} \qquad N_*, n \to \infty.$$

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

References

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

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

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

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

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

Bickel, P. J., Götze, F. and van Zwet, W. R., *Resampling fewer than n observations: gains, losses, and remedies for losses*, Statist. Sinica **7** (1997), 1–31.

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

Bloznelis, M. and Götze, F., *An Edgeworth expansion for finite population U–statistics*, To appear in Bernoulli (1999 a).

Bloznelis, M. and Götze, F., *One term Edgeworth expansion for finite population U–statistics of degree two*, Acta Applicanda Mathematicae **58** (1999 b), 75–90.

Bloznelis, M. and Götze, F., *An Edgeworth expansion for finite population symmetric statistics* (1999 c), Manuscript.

Booth, J. G., Butler, Ronald W. and Hall, P., *Bootstrap methods for finite populations*, J. Amer. Statist. Assoc. **89** (1994), 1282–1289.

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

Chao, M. T. and Lo, S. H., *A bootstrap method for finite population*, Sankhya Ser. A **47** (1985), 399–405.

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

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

Gross, S., *Median estimation in sample surveys* (1980), Proceedings of the Section on Survey Research Methods,, American Statistical Association, 181–184.

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

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

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

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

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

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

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

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

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

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

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

Wu, C.F.J., *On the asymptotic properties of the jackknife histogram*, Ann. Statist. **18**, 1438–1452.

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

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