AN EDGEWORTH EXPANSION FOR STUDENTIZED FINITE POPULATION STATISTICS

M. Bloznelis

Abstract. We show the validity of the one-term Edgeworth expansion for Studentized asymptotically linear statistics based on samples drawn without replacement from finite populations. Replacing the moments defining the expansion by their estimators we obtain an empirical Edgeworth expansion. We show the validity of the empirical Edgeworth expansion in probability.

1. Introduction and Results

1. Let $T = t(X_1, \dots, X_n)$ be a real valued statistic based on sample $X_1, \dots, X_n$ drawn without replacement from a finite population $\mathcal{X} = \{x_1, \dots, x_N\}$. Write

$$p = n/N, \qquad q = 1 - p, \qquad n_* = \min\{n,\, N - n\}.$$

Let σ_T^2 denote the variance of T and let

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

denote the jackknife estimator of variance based on sample $X_1, \dots, X_{n+1}$ (of size $n+1$) drawn without replacement form $\mathcal{X}$. Here $T_{(j)} = t(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_{n+1})$.

In the simplest case of a linear statistic $L = g(X_1) + \dots + g(X_n)$ (here $g : \mathcal{X} \to R$) the asymptotic normality as $n_* \to \infty$ of $(L - \mathbf{E}\,L)/S(L)$ follows from the central limit theorem combined with the law of large numbers. Edgeworth expansions were shown by Babu and Singh [1], see also [13].

Many important statistics are asymptotically linear as n increases. Consequently, their Studentized versions $(T - \mathbf{E}\,T)/S$ are asymptotically standard normal. In order to treat general asymptotically linear statistics we use linearization by means

---

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

Key words and phrases. Sampling without replacement, finite population, U-statistic, Studentized statistic, asymptotic expansion, Edgeworth expansion, jackknife.

of the orthogonal decomposition. This kind of decomposition of statistics, was first used by Hoeffding [9] in the case of *independent and identically distributed* observations. Orthogonal decomposition of symmetric statistics based on samples *drawn without replacement* was studied by Bloznelis and Götze [6], see also [14].

We shall assume in what follows that T is symmetric. That is, the kernel t is invariant under permutations of its arguments, i.e., $t(y_1, \dots, y_n) = t(y_{\pi(1)}, \dots, y_{\pi(n)})$, for any permutation π of indices $1, 2, \dots, n$. Note that the sample mean, the sample variance, sample quantiles, U-statistics, L-statistics, and many others are symmetric. Given a symmetric statistic T is decomposed into the sum of centered and *uncorrelated* $U-$statistics of increasing order,

$$T = \mathbf{E}\,T + \sum_{1 \le i \le n} g_1(X_i) + \sum_{1 \le i < j \le n} g_2(X_i, X_j) + \dots. \tag{1.1}$$

The first sum

$$L = \sum_{1 \le i \le n} g_1(X_i), \qquad g_1(X_i) = \frac{N-1}{N-n}\,\mathbf{E}\,(T - \mathbf{E}\,T | X_i), \tag{1.2}$$

is called the linear part of T. The second sum $Q = \sum_{i<j} g_2(X_i, X_j)$ is called the quadratic part. Here, for $i \neq j$,

$$g_2(X_i, X_j) = \frac{N-3}{N-n-1} \Big(\frac{N-2}{N-n}\,\mathbf{E}\,(T - \mathbf{E}\,T \big| X_i, X_j) - g_1(X_i) - g_1(X_j) \Big). \tag{1.3}$$

The random variables $g_1(X_k)$ and $g_2(X_i, X_j)$ are centered and uncorrelated for arbitrary $1 \le i, j, k \le n$, $i \neq j$. For a detailed description of the decomposition we refer to [6].

We shall assume that the linear part does not vanish, that is, $\sigma_1^2 > 0$, where $\sigma_1^2 = \mathbf{Var}\, g_1(X_1)$. Note that the variance of the linear part

$$\mathbf{Var}\, L = \tau^2 \sigma_1^2 N/(N-1), \qquad \text{where} \qquad \tau^2 = Npq.$$

Furthermore, $n_*/2 \le \tau^2 \le n_*$.

If, for large n, the linear part dominates the statistic, we call T asymptotically linear. In this case Bloznelis and Götze [6] showed that the one-term Edgeworth expansion

$$G(x) = \Phi(x) - \frac{(q-p)\alpha + 3\kappa}{6\tau}\,\Phi'(x)(x^2 - 1)$$

approximates the distribution function $F(x) = \mathbf{P}\{T - \mathbf{E}\,T \le \sigma_T x\}$ up to the error $o(n_*^{-1/2})$. The moments

$$\alpha = \sigma_1^{-3} \mathbf{E}\, g_1^3(X_1), \qquad \kappa = \sigma_1^{-3} \tau^2 \mathbf{E}\, g_2(X_1, X_2) g_1(X_1) g_1(X_2) \tag{1.4}$$

refer to the linear and the quadratic part of the decomposition.

We shall show in Theorem 2 below that the one-term Edgeworth expansion

$$H(x) = \Phi(x) + \frac{(q - p + (q+1)x^2)\alpha + 3(x^2+1)\kappa}{6\tau}\,\Phi'(x) \tag{1.5}$$

approximates the distribution function of the Studentized statistic

$$F_S(x) = \mathbf{P}\{T - \mathbf{E}\,T \le S(T)x\}$$

up to the error $o(n_*^{-1/2})$.

Note that in order to write the expansion (1.5) one does not need to evaluate all terms of the decomposition (1.1), but the moments (1.4) of the linear and the quadratic part only. Furthermore, these moments can be estimated.

Let us define the jackknife estimators. In what follows $\{X_1, \ldots, X_m\}$, for $m = n, n+1, n+2$, denote simple random samples drawn without replacement from $\mathcal{X}$. It is convenient to represent the sample $\{X_1, \ldots, X_m\}$ by the set of the first m variables of the random permutation $(X_1, \ldots, X_N)$ of the ordered set $(x_1, \ldots, x_N)$. For $1 \le k \le n+1$, $1 \le i, j, r \le n+2$, $i \ne j$, denote

$$V_k = \overline{T} - T_{(k)}, \qquad \tilde{V}_r = \tilde{T} - \overline{T}_{(r)}, \qquad W_{ij} = \tilde{T} - \overline{T}_{(i)} - \overline{T}_{(j)} + T_{(i,j)},$$

where

$$\overline{T}_{(r)} = \frac{1}{n+1} \sum_{1 \le j \le n+2,\, j \ne r} T_{(r,j)}, \qquad \tilde{T} = \frac{1}{\binom{n+2}{2}} \sum_{1 \le i < j \le n+2} T_{(i,j)}.$$

Here $T_{(i,j)}$ denotes the value of t at the sample $\{X_1, \ldots, X_{n+2}\} \setminus \{X_i, X_j\}$. Write

$$\hat{\alpha}_J = \frac{\sqrt{n}}{\hat{\sigma}_J^3} \sum_{k=1}^{n+1} V_k^3, \qquad \hat{\kappa} = q\,\frac{2\sqrt{n}}{\hat{\sigma}_J^3} \sum_{1 \le i < j \le n+2} W_{ij} \tilde{V}_i \tilde{V}_j, \tag{1.6}$$

where $\hat{\sigma}_J^2 = \sum_{k=1}^{n+1} V_k^2$.

Bloznelis [4] showed that $\hat{\alpha}$, $\hat{\kappa}$ and $S^2(T)\,(= q\hat{\sigma}_J^2)$ are consistent estimators of α, κ and σ_T^2 as $n \to \infty$. Using this fact we show in Theorem 3 below that the empirical Edgeworth expansion

$$\hat{H}(x) = \Phi(x) + \frac{(q - p + (q+1)x^2)\hat{\alpha} + 3(x^2+1)\hat{\kappa}}{6\tau}\,\Phi'(x) \tag{1.7}$$

approximates $F_S(x)$ up to the error $o(n_*^{-1/2})$ in probability.

One-term Edgeworth expansions for Studentized U-statistics based on *independent and identically distributed* observations were constructed by Helmers [8]. Expansions for Studentized versions of general symmetric statistics were shown by Putter and van Zwet [12], see also [2], [11]. Empirical Edgeworth expansions that use jackknife estimators were studied by Beran [3]. Putter and van Zwet [12] constructed such expansions for general symmetric statistics and their Studentized versions.

One-term Edgeworth expansion for U-statistics of degree two based on samples *drawn without replacement* was constructed by Kokic and Weber [10]. Bloznelis and Götze [6] established the validity of the one-term Edgeworth expansion for general symmetric finite population statistics. Corresponding empirical Edgeworth expansions were constructed by Bloznelis [4]. Since often the variance of the underlying statistic (estimator) is unknown, it is important, for practical purposes, to have such approximations for Studentized versions of statistics too. This question is addressed in the present paper. The one-term Edgeworth expansion for Studentized finite population statistics (1.5) and the corresponding empirical Edgeworth expansion (1.7) seem to be new and not known in the literature before.

2. Results. In order to formulate asymptotic results for finite population statistics we introduce a sequence of populations $\mathcal{X}_\nu = \{x_{\nu\,1}\ldots, x_{\nu\,N_\nu}\}$, $\nu = 1, 2, \ldots$, and a sequence of symmetric statistics $T_\nu = t_\nu(X_{\nu\,1}, \ldots, X_{\nu\,n_\nu})$. Here $\{X_{\nu\,1}, \ldots, X_{\nu\,n_\nu}\}$ denotes a sample drawn without replacement from $\mathcal{X}_\nu$. Orthogonal decomposition expands T_ν into the sum of uncorellated centered U-statistics

$$(1.8)\quad T_\nu = \mathbf{E}\,T_\nu + U_{\nu\,1} + \cdots + U_{\nu\,n_\nu}, \qquad U_{\nu\,k} = \sum_{1\le i_1<\cdots<i_k\le n_\nu} g_{\nu\,k}(X_{\nu\,i_1}, \ldots X_{\nu\,i_k}).$$

Let $L_\nu = U_{\nu\,1}$ and $Q_\nu = U_{\nu\,2}$ denote the linear and the quadratic part respectively. Furthermore, denote $n_{\nu\,*} = \min\{N_\nu - n_\nu,\, n_\nu\}$ and $\sigma^2_{\nu\,1} = \mathbf{Var}\, g^2_{\nu\,1}(X_{\nu\,1})$,

$$\beta_{\nu\,s} = \sigma^{-s}_{\nu\,1}\mathbf{E}\,|g_{\nu\,1}(X_{\nu\,1})|^s, \qquad \gamma_{\nu\,s} = \sigma^{-s}_{\nu\,1}\tau^{2s}_\nu\mathbf{E}\,|g_{\nu\,2}(X_{\nu\,1}, X_{\nu\,2})|^s,$$

where $s > 0$ and where $\tau_\nu = N_\nu p_\nu q_\nu$, $p_\nu = n_\nu/N_\nu$, $q_\nu = 1 - p_\nu$. Let α_ν, κ_ν and $\hat\alpha_\nu$, $\hat\kappa_\nu$ denote the moments defined by (1.4) and their jackknife estimators (1.6) respectively. Furthermore, let $S^2_\nu = S^2_\nu(T_\nu)$ denote the jackknife estimator of the variance $\sigma^2_{\nu\,T}$ of T_ν. Write $\Psi_\nu(t) = \mathbf{E}\,\exp\{it\sigma^{-1}_{\nu\,1}g_{\nu\,1}(X_{\nu\,1})\}$.

We shall assume that $n_{\nu\,*}$ tends to infinity as $\nu \to \infty$ and construct bounds for

$$\Delta_\nu = \sup_x |F_{\nu\,S}(x) - H_\nu(x)| \qquad \text{and} \qquad \hat\Delta_\nu = \sup_x |F_{\nu\,S}(x) - \hat H_\nu(x)|.$$

Here $F_{\nu\,S}(x) = \mathbf{P}\{(T_\nu - \mathbf{E}\,T_\nu)/S_\nu(T_\nu) \le x\}$. The functions $H_\nu(x)$ and $\hat H_\nu(x)$ are defined by (1.5) and (1.7), but using α_ν, κ_ν and $\hat\alpha_\nu$, $\hat\kappa_\nu$.

Firstly, we consider a special case of U-statistics. We write $T_\nu = U_\nu$, where

$$U_\nu = \sum_{1\le i<j\le n_\nu} h_\nu(X_{\nu\, i}, X_{\nu\, j}). \tag{1.9}$$

In this case the decomposition (1.8) reduces to $U_\nu = \mathbf{E}\, U_\nu + L_\nu + Q_\nu$. The kernels defining the linear and the quadratic part are obtained from (1.2) and (1.3)

$$g_{\nu\,1}(x) = (n_\nu - 1)\,\frac{N_\nu - 1}{N_\nu - 2}\,\mathbf{E}\,\Big(h_\nu(X_{\nu\,1}, X_{\nu\,2}) - \mathbf{E}\, h_\nu(X_{\nu\,1}, X_{\nu\,2})\Big| X_{\nu\,1} = x\Big),$$
$$g_{\nu\,2}(x,y) = h_\nu(x,y) - \mathbf{E}\, h_\nu(X_{\nu\,1}, X_{\nu\,2}) - (n_\nu - 1)^{-1}(g_{\nu\,1}(x) + g_{\nu\,1}(y)).$$

Theorem 1. *Let T_ν be a U-statistic of the form (1.9). Assume that $n_{\nu\,*} \to \infty$ as $\nu \to \infty$. Assume that there exist absolute constants $s > 6$, $C_1 > 0$, a positive continuous function ϕ on $(0, +\infty)$, and sequences $\{\xi_\nu\} \uparrow \infty$ and $\{\eta_\nu\} \uparrow \infty$ such that for $\nu = 1, 2, \dots,$*

$$\beta_{\nu\, s} \le C_1, \qquad \gamma_{\nu\, s} \le C_1, \tag{1.10}$$
$$|\Psi(t)| \le 1 - \phi(|t|), \qquad \textit{for} \quad 0 < |t| \le \eta_\nu, \tag{1.11}$$
$$n_\nu \le N_\nu - \xi_\nu N_\nu^{2/3}. \tag{1.12}$$

Then there exists a sequence $\{\psi_\nu\} \downarrow 0$ depending only on C_1, ϕ, $\{\xi_\nu\}$, and $\{\eta_\nu\}$ such that, for every $\nu = 1, 2 \dots,$

$$\Delta_\nu \le \psi_\nu \tau_\nu^{-1}, \tag{1.13}$$
$$\mathbf{P}\big\{\hat{\Delta}_\nu > \psi_\nu \tau_\nu^{-1}\big\} \le \psi_\nu. \tag{1.14}$$

Remark 1. Under the moment condition (1.10), the non-lattice condition (1.11) and (1.12) the results (1.13) respectively (1.14) establish the bounds as $n_{\nu *} \to \infty$

$$\Delta_\nu = o(n_{\nu *}^{-1/2}) \qquad \text{respectively} \qquad \hat{\Delta}_\nu = o_P(n_{\nu *}^{-1/2}) \tag{1.15}$$

(in probability). Here $n_{\nu *}$ plays the same role as the sample size does in the i.i.d. situation, see [7].

Remark 2. Let us note that the non-lattice condition (1.11) is the weakest possible smoothness condition. The condition (1.12) is very mild. The moment condition (1.10) is far from the optimal one. Here one would expect the uniform integrability of $\beta_{\nu\,3}$ and $\gamma_{\nu\,5/3}$, for $\nu = 1, 2, \dots,$ instead of (1.10). In the proof no effort was made to obtain the result (1.15) under the optimal moment conditions. A modification of our proof involving truncation would probably reduce the moment condition (1.10) up to $\beta_{\nu\, s} < C_1$ and $\gamma_{\nu\, t} < C_1$, for $s > 3$ and $t > 2$, cf. [5].

Let us consider a general symmetric statistic T_ν. Using (1.8) we write

$$T_\nu = U_\nu + R_\nu,$$

where $U_\nu = \mathbf{E}\, T_\nu + L_\nu + Q_\nu$ is a finite population U-statistic of degree two and where the remainder $R_\nu = \sum_{j\ge 3} U_{\nu\, j}$. For a typical standardized asymptotically linear statistic $\sigma_{\nu T}^{-1} T_\nu$ admitting one-term expansion we have as $n_{\nu *} \to \infty$

$$\sigma_{\nu T}^{-1} R_\nu = o_P(n_{\nu *}^{-1/2}) \qquad \text{and} \qquad S_\nu(T_\nu)/S_\nu(U_\nu) = 1 + o_P(n_{\nu *}^{-1/2}) \tag{1.16}$$

in probability. Consequently, one-term expansions for $(T_\nu - \mathbf{E}\, T_\nu)/S_\nu(T_\nu)$ and $(U_\nu - \mathbf{E}\, U_\nu)/S_\nu(U_\nu)$ are, in fact, the same.

Bloznelis and Götze [6] introduced simple conditions that ensure the validity of approximations like (1.16). These conditions are formulated in terms of moments of differences. Recall that $(X_{\nu\, 1}, \dots, X_{\nu\, N_\nu})$ denotes a random permutation of the ordered population $(x_{\nu\, 1}, \dots, x_{\nu\, N_\nu})$. For $j < n_{\nu\, *}$ define

$$D^j T_\nu = t_\nu(X_{\nu 1}, \dots, X_{\nu n}) - t_\nu(X_{\nu 1}, \dots, X_{\nu j-1},\, X_{\nu j+1}, \dots, X_{\nu n},\, X_{\nu n+j}).$$

Higher order differences are defined recursively: $D^{ij} T_\nu = D^j\big(D^i T_\nu\big)$, for $i \ne j$; $D^{ijk} T_\nu = D^k\big(D^j(D^i T_\nu)\big)$, for $i \ne j \ne k$; $\dots$. Write

$$\delta_{\nu\, j} = \delta_{\nu\, j}(T_\nu) = \mathbf{E}\left(n_{\nu *}^{(j-1)} \mathbb{D}_j T_\nu\right)^2, \qquad \mathbb{D}_j T_\nu := D^{12\dots j} T_\nu.$$

Theorem 2. *The statement (1.13) holds true for a sequence of general symmetric statistic $\{T_\nu\}$ if in addition to (1.10), (1.11) and (1.12) we assume that*

$$\delta_{\nu\, 3} \le \varepsilon_\nu n_{\nu\, *}^{-2/3} \sigma_{\nu\, T}^2,$$

for some decreasing sequence $\{\varepsilon_\nu\} \downarrow 0$ as $\nu \to \infty$.

Theorem 3. *The statement (1.14) holds true for a sequence of general symmetric statistic $\{T_\nu\}$ if in addition to (1.10), (1.11) and (1.12) we assume that*

$$\delta_{\nu\, 2} \le \varepsilon_\nu n_\nu^{-1/3} \sigma_{\nu\, T}^2, \qquad \delta_{\nu\, 3} \le \varepsilon_\nu n_{\nu\, *}^{-2/3} \sigma_{\nu\, T}^2,$$

for some decreasing sequence $\{\varepsilon_\nu\} \downarrow 0$ as $\nu \to \infty$.

Theorems 2 and 3 establish the bounds (1.15) for general symmetric statistic.

2. Proofs

We shall prove only Theorem 1. The proof of Theorems 2 and 3 is almost the same. The extension of the argument of the proof of Theorem 1 to general symmetric statistics needs only a minor modification, see [4] and [12]. In the proof we shall use the variance decomposition, see formula (2.6) of [6],

$$\mathbf{Var}\, T_\nu = \sum_{k=1}^{n_\nu} \mathbf{Var}\, U_{\nu\, k}, \qquad \mathbf{Var}\, U_{\nu k} = \frac{\binom{n_\nu}{k}\binom{N_\nu - n_\nu}{k}}{\binom{N_\nu - k}{k}}\, \sigma^2_{\nu\, k}. \tag{2.1}$$

Here $\sigma^2_{\nu\, k} = \mathbf{Var}\, g_{\nu k}(X_{\nu\, 1}, \dots, X_{\nu\, k})$, and $\mathbf{Var}\, U_{\nu k} = 0$, for $k > n_{\nu *}$.

Proof of Theorem 1. We assume without loss of generality that $\mathbf{E}\, U_\nu = 0$ and $\mathbf{Var}\, U_\nu = 1$, for $\nu = 1, 2, \dots$.

Denote, for brevity, $\sigma^2_{\nu U} = \mathbf{Var}\, U_\nu$, $\;S^2_\nu = S^2_\nu(U_\nu)$, $\;Y_{\nu i} = g_{\nu 1}(X_{\nu i})$, and $Y_{\nu ij} = g_{\nu 2}(X_{\nu i}, X_{\nu j})$. In order to simplify the notation we shall write $o(n^s_{\nu *})$ to denote the sequence $\tilde\psi_\nu n^s_{\nu *}$, where $\{\tilde\psi_\nu\} \downarrow 0$ is a sequence depending only on the constants C_1, the function ϕ and the sequences $\{\xi_\nu\}$ and $\{\eta_\nu\}$. Furthermore, we shall drop the subscript ν whenever this does not cause an ambiguity. By c we denote a constant which depends only on C_1.

Note that (1.14) follows from (1.13) and the consistency results established in Lemmas 2.1-3 in [4]: as $\nu \to \infty$

$$|S^2_\nu - \sigma^2_{\nu U}| = o_P(1), \quad |\hat\alpha_\nu - \alpha_\nu| = o_P(1), \quad |\hat\kappa_\nu - \kappa_\nu| = o_P(1).$$

Let us mention that Lemma 2.2 in [4], which establishes the consistency result for $\hat\alpha_\nu$, assumes, in addition, that $\delta_2(U_\nu) = o(n_\nu^{-1/3})$. For U statistics of degree two (and such that $\mathbf{Var}\, U = 1$) this condition can be replaced by the condition formulated in the second inequality of (1.10).

Let us prove (1.13). It follows from (2.1) that $\sigma^2_U = \mathbf{Var}\, L + \mathbf{Var}\, Q$,

$$\mathbf{Var}\, L = \frac{\binom{n}{1}\binom{N-n}{1}}{\binom{N-1}{1}}\, \sigma^2_1, \qquad \mathbf{Var}\, Q = \frac{\binom{n}{2}\binom{N-n}{2}}{\binom{N-2}{2}}\, \sigma^2_2. \tag{2.2}$$

By the assumption $\sigma^2_U = 1$, we have $\mathbf{Var}\, L \le 1$. Therefore, $\sigma^2_1 \le \tau^{-2}$. Invoking Hölder's inequality $\gamma_2 \le \gamma_6^{1/3}$ we obtain from (1.10)

$$\sigma^2_2 = \gamma_2 \sigma^2_1 \tau^{-4} \le C_1^{1/3} \sigma^2_1 \tau^{-4} \le c\tau^{-6}. \tag{2.3}$$

The identities $\sigma^2_U = 1$ and (2.2) combined with (2.3) show

$$1 - c\tau^{-2} \le \mathbf{Var}\, L \le 1 \qquad \text{and} \qquad \tau^{-2} \ge \sigma^2_1 \ge \tau^{-2} - c\tau^{-4}. \tag{2.4}$$

The remaining part of the proof splits in two steps.

Step 1. In this step we show that that there exists a sequence $\{\tilde\psi_\nu\} \downarrow 0$ (which depends only on C_1) such that for every $\nu = 1, 2, \dots$

$$S_\nu^2 = \sigma_{\nu U}^2 + L_S + M_\nu, \qquad L_S = \sum_{j=1}^{n_\nu+1} f(X_{\nu i}), \tag{2.5}$$

$$f(X_{\nu i}) = q_\nu\Big(1 - \frac{1}{n_\nu+1}\Big)(Y_{\nu i}^2 - \sigma_{\nu 1}^2) + 2q_\nu n_\nu \frac{N_\nu - 1}{N_\nu - 2}\, \mathbf{E}\,(Y_{\nu j} Y_{\nu ij} | X_{\nu i}).$$

Here the sequence of random variables $\{M_\nu\}$ satisfies

$$\mathbf{P}\{|M_\nu| \ge \tilde\psi_\nu \tau_\nu^{-1}\} \le \tilde\psi_\nu \tau_\nu^{-1}. \tag{2.6}$$

To show (2.5) and (2.6) fix k and split $V_k = \overline{U} - U_{(k)} = Z_k + W_k$, where

$$Z_k = \sum_{j=1}^{n+1} Y_j \big(\mathbb{I}_{\{j=k\}} - \frac{1}{n+1}\big), \qquad W_k = \sum_{1\le i<j\le n+1} Y_{ij}\big(\mathbb{I}_{\{k\in\{i,j\}\}} - \frac{2}{n+1}\big),$$

and write

$$S^2 = q\sum_k V_k^2 = q\sum_k Z_k^2 + 2q\sum_k Z_k W_k + q\sum_k W_k^2. \tag{2.7}$$

A calculation shows

$$\sum_k Z_k^2 = \big(1 - \frac{1}{n+1}\big)\sum_{k=1}^{n+1} Y_k^2 - \frac{2}{n+1}\sum_{1\le i<j\le n+1} Y_i Y_j,$$

$$\sum_k Z_k W_k = \sum_{1\le i<j\le n+1} (Y_i + Y_j) Y_{ij} - \frac{2}{n+1} H_1 H_2,$$

$$\sum_k W_k^2 = 2\sum_{1\le i<j\le n+1} Y_{ij}^2 + 2H_3 - \frac{8}{n+1} H_2^2 + \frac{4}{(n+1)^2} H_2^2.$$

Here $H_1 = \sum_{1\le k\le n+1} Y_k$ and $H_2 = \sum_{1\le i<j\le n+1} Y_{ij}$, and

$$H_3 = \sum_{1\le i\le n+1}\ \sum_{1\le j<k\le n+1} Y_{ij} Y_{ik} \mathbb{I}_{\{i\notin\{j,k\}\}}.$$

Finally, from (2.7) we obtain (2.5) with $M_\nu = m_1 + q(m_2 + \dots + m_8)$. Here

$$m_1 = nq\sigma_1^2 - \sigma_U^2, \qquad m_2 = -\frac{2}{n+1}\sum_{1\le i<j\le n+1} Y_i Y_j,$$

$$m_3 = -\frac{4}{n+1} H_1 H_2, \qquad m_4 = 2(H - L_H),$$

where

$$H = \sum_{1\le i<j\le n+1} (Y_i + Y_j)Y_{ij} \qquad \text{and} \qquad L_H = \sum_{1\le i\le n+1} n\,\frac{N-1}{N-2}\,\mathbf{E}\,(Y_jY_{ij}|X_i),$$

and where

$$m_5 = 2\sum_{1\le i<j\le n+1} Y_{ij}^2, \quad m_6 = 2H_3, \quad m_7 = -\,\frac{8}{n+1}\,H_2^2, \quad m_8 = \frac{4}{(n+1)^2}\,H_2^2.$$

In order to prove (2.6) we show that

$$\mathbf{E}\,H_1^2 < 2, \qquad \mathbf{E}\,H_2^2 \le c\tau^{-2}, \qquad |\mathbf{E}\,H_3| \le cq^{-2}\tau^{-2},$$

$$|m_1| \le c\tau^{-2}, \qquad |\mathbf{E}\,m_2| \le \tau^{-2} \qquad \mathbf{E}\,m_5 \le cq^{-2}\tau^{-2}, \tag{2.8}$$

$$\mathbf{Var}\,m_2 \le cn^{-2}, \qquad \mathbf{Var}\,m_5 \le cq^{-2}\tau^{-6},$$

$$\mathbf{Var}\,(H - L_H) \le c\tau^{-4}, \qquad \mathbf{Var}\,H_3 \le cq^{-2}\tau^{-4}. \tag{2.9}$$

It is easy to show that the bounds (2.8) and (2.9) imply (2.6), provided that $q_\nu\tau_\nu \to \infty$ as $\nu \to \infty$. The latter condition is equivalent to (1.12).

The bounds for expectations (2.8) are simple consequences of (2.1), (2.2), (2.3) and (2.4). In order to bound the variances (2.9) we decompose the random variables m_2, m_5, $H - L_H$ (which are U-statistics of degree two) and H_3 (which is a U-statistic of degree three) by means of (1.8) and use the identity (2.1).

Step 2. Denote $\tilde L_S = \sum_{i=1}^n f(X_i)$ and write $\tilde S^2 = 1 + \tilde L_S$, for $\tilde L_S > -1$, and put $\tilde S^2 = 2$, for $\tilde L_S \le -1$. Denote $\tilde\Delta = \sup_x \tilde\Delta(x)$, where $\tilde\Delta(x) = |\mathbf{P}\{U/\tilde S \le x\} - H(x)|$. Using Chebyshev's inequality we obtain

$$\mathbf{P}\{|\tilde S - 1| > 1/2\} \le \mathbf{P}\{|\tilde L_S| > 1/2\} \le 4\mathbf{E}\,\tilde L_S^2 \le c\,\tau^{-2}. \tag{2.10}$$

Using (2.5) and $\sigma_U^2 = 1$ we write

$$S/\tilde S = \sqrt{1 + (M + f(X_{n+1}))/\tilde S^2} = 1 + \tilde M.$$

By Lagrange mean value theorem, we obtain form (2.6) and (2.10) that the remainder $\tilde M$ satisfies (2.6). Since $xH'(x)$ is bounded, this implies $\Delta \le \tilde\Delta + o(n_*^{-1/2})$. In what follows we construct the bound $\tilde\Delta = o(n_*^{-1/2})$.

The bound $\sup\{\tilde\Delta(x) : |x| \ge \log n_*\} = o(n_*^{-1/2})$ follows from the result of Theorem 2 of [6], see also Theorem 2.1 in [4],

$$\sup_x |P\{U/\sigma_U \le x\} - G(x)| = o(n_*^{-1/2}), \tag{2.11}$$

and the fact that $|G(x)|$ and $|H(x)|$ decrease exponentially as $|x| \to \infty$. Here we also use $\sigma_U^2 = 1$ and (2.10).

Let $|x| < \log n_*$. For $|\tilde{L}_S| \le 4/5$ we use the inequalities $A \le \tilde{S} \le B$, where $A = 1 + \tilde{L}_S/2 - \tilde{L}_S^2/4$ and $B = 1 + \tilde{L}_S/2$. Since, by (2.10), $\mathbf{P}\{|\tilde{L}_S| > 4/5\} = o(n_*^{-1/2})$, we obtain $\tilde{\Delta}(x) \le \max\{\Delta_A, \Delta_B\} + o(n_*^{-1/2})$, where

$$\Delta_A = \sup_{|x| \le \log n_*} |\mathbf{P}\{U \le Ax\} - H(x)|, \qquad \Delta_B = \sup_{|x| \le \log n_*} |\mathbf{P}\{U \le Bx\} - H(x)|.$$

It is easy to show (for instance, by using Chebyshev's inequality) that $\mathbf{P}\{|A - B| > n_*^{-8/15}\} = o(n_*^{-1/2})$. Since $xH'(x)$ is bounded this implies $\Delta_A \le \Delta_B + o(n_*^{-1/2})$. Finally, an application of (2.11) gives $\Delta_B = o(n_*^{-1/2})$ thus completing the proof of the bound $\tilde{\Delta} = o(n_*^{-1/2})$.

References

1. Babu, G. J. and Singh, K. (1985), *Edgeworth expansions for sampling without replacement from finite populations*, J. Multivar. Anal. **17**, 261–278.
2. Bentkus, V., Götze, F. and van Zwet, W. R. (1997), *An Edgeworth expansion for symmetric statistics*, Ann. Statist. **25**, 851–896.
3. Beran, R. (1984), *Jackknife approximations to bootstrap estimates*, Ann. Statist. **12**, 101–118.
4. Bloznelis, M. (2001), *Empirical Edgeworth expansion for finite population statistics.I*, Liet. Matem. Rink. **41**, 154–171.
5. Bloznelis, M. and Götze, F. (1999), *One term Edgeworth expansion for finite population U–statistics of degree two*, Acta Applicandae Mathematicae **58**, 75–90.
6. Bloznelis, M. and Götze, F. (2001), *Orthogonal decomposition of finite population statistics and its applications to distributional asymptotics*, Ann. Statist. **29**, 899–917.
7. Bloznelis, M. and Götze, F. (2002), *An Edgeworth expansion for finite population symmetric statistics*, Ann. Probab. **30**.
8. Helmers R. (1991), *On the Edgeworth expansion and the bootstrap approximation for a Studentized U-statistic*, Ann. Statist. **19**, 470–484.
9. Hoeffding, W. (1948), *A class of statistics with asymptotically normal distribution*, Ann. Math. Statist. **19**, 293–325.
10. 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.
11. Putter, H. (1994), *Consistency of Resampling Methods.*, PhD Thesis, Leiden University.
12. Putter, H. and van Zwet, W.R. (1998), *Empirical Edgeworth expansions for symmetric statistics*, Ann. Statist. **26**, 1540–1569.
13. Sugden, R. A., Smith, T. M. F. and Jones, R. P. (2000), *Cochran's rule for simple random sampling*, J.R. Statist. Soc. B **62**, 787–793.
14. Zhao Lin Cheng and Chen Xi Ru (1990), *Normal approximation for finite-population U-statistics*, Acta mathematicae applicatae Sinica **6**, 263–272.