VILNIUS UNIVERSITY

Andrius Čiginas

Approximations to distributions of linear combinations of order statistics in finite populations

Doctoral dissertation

Physical sciences, Mathematics (01P)

Vilnius, 2011

The scientific work was carried out during 2006-2011 at Vilnius University

Scientific supervisor:

Prof. Dr. Habil. Mindaugas Bloznelis (Vilnius University, Physical Sciences, Mathematics – 01P)

ISBN 978-9955-634-96-6

VILNIAUS UNIVERSITETAS

Andrius Čiginas

POZICINIŲ STATISTIKŲ TIESINIŲ KOMBINACIJŲ SKIRSTINIŲ APROKSIMACIJOS BAIGTINĖSE POPULIACIJOSE

Daktaro disertacija

Fiziniai mokslai, matematika (01P)

Vilnius, 2011

Disertacija rengta 2006-2011 metais Vilniaus universitete

Mokslinis vadovas:

Prof. habil. dr. Mindaugas Bloznelis (Vilniaus universitetas, fiziniai mokslai, matematika – 01P)

Contents

Notations

N denotes the population size.

$\mathcal{X}$ denotes the population of N numbers.

n denotes the sample size.

$\mathbb{X}$ denotes the sample of size n.

$\mathbb{X}_k$ denotes the sample of size $n+k$, $k = 1, 2, \ldots, N-n$.

n_* denotes $\min\{n, N-n\}$.

$\mathbb{N}$ denotes the set of natural numbers, $\mathbb{N} = \{1, 2, \ldots\}$.

$\mathbb{R}$ denotes the set of real numbers.

$\mathbb{C}$ denotes the set of complex numbers.

$[\cdot]$ denotes the greatest integer function.

$\mathbb{I}\{\cdot\}$ denotes the indicator function.

$\triangle_i$ denotes the difference $x_{i+1} - x_i$ of values from the population where $x_1 \leq \cdots \leq x_N$.

$\mathbf{\Delta}_{j:n}$ denotes the difference $X_{j+1:n} - X_{j:n}$ of the order statistics.

$\mathbf{P}\{A\}$ denotes the probability of an event A.

$\mathbf{E}\, X$ denotes the expectation of a random variable X.

$\mathbf{Var}\, X$ denotes the variance of a random variable X.

$\mathbf{Cov}\,(X, Y)$ denotes the covariance of random variables X and Y.

$\binom{m}{k}$ denotes the binomial coefficient $m!/[k!(m-k)!]$.

$\mathcal{H}_{N,n,i}(j)$ denotes the probability $\binom{i}{j}\binom{N-i}{n-j}\Big/\binom{N}{n}$ that a hypergeometric random variable with parameters N, n and i attains the value j.

$a_n = O(b_n)$ as $n \to \infty$ means that $|a_n| \,/\, |b_n| \leq C$, for some $C > 0$ and all n.

$a_n = o(b_n)$ as $n \to \infty$ means that $\lim_{n\to\infty}(a_n/b_n) = 0$.

$Y_n = O_P(b_n)$ as $n \to \infty$ means that the sequence $|Y_n| \, / \, |b_n|$ is bounded in probability.

$Y_n = o_P(b_n)$ as $n \to \infty$ means that the sequence $|Y_n| \, / \, |b_n|$ converges to zero in probability.

Introduction

An asymptotic theory for random variables (statistics), occurring in problems of mathematical statistics, plays an important role when we need good approximations to distributions of those statistics. The most fundamental asymptotic approximation is the normal approximation. Now it is well studied not only for sums of independent and identically distributed (i.i.d.) random variables, but also for much more complex statistics as well as for various sampling models. However, in many practical situations the accuracy of this classical approximation is not sufficient unless the sample size is (very) large. One of the known methods, which can improve the normal approximation, is Edgeworth expansions, i.e., the normal approximation plus one or more correction terms which reflect the specifics of an underlying statistic and a sampling model.

The main objects of this doctoral dissertation are one-term Edgeworth expansions for distributions of a general class of linear combinations of order statistics (L-statistics), where samples are *drawn without replacement* from a finite population. The work done also involves other related questions such as the same asymptotic normality, the analysis and estimation of variances of the statistics, an efficient estimation of parameters that define Edgeworth expansions, empirical Edgeworth expansions, bootstrap approximations, etc. We outline the following problems.

Aims and problems

- Construct a short Edgeworth expansion for L-statistics.
- Find explicit expressions of the main terms of the Hoeffding decomposition of L-statistics.
- Construct upper bounds for the variances of order statistics.
- Obtain simple sufficient conditions for the asymptotic normality and validity of the Edgeworth expansion.
- Construct estimators of variance and parameters that define the Edgeworth expansion of an L-statistic.

- Construct and analyze a one-term Edgeworth expansion for a Studentized L-statistic and empirical Edgeworth expansions.

Let $\mathcal{X} = \{x_1, \dots, x_N\}$ denote measurements of the study variable x of the population $\mathcal{U} = \{u_1, \dots, u_N\}$, i.e., a real function $f : \mathcal{U} \to \mathbb{R}$ assigns a fixed value for each element of the population $\mathcal{U}$. Let $\mathbb{X} = \{X_1, \dots, X_n\}$ be measurements of units of the simple random sample of size $n < N$ drawn without replacement from the population. The observations $X_1, \dots, X_n$ are identically distributed, but they are not independent. Let $X_{1:n} \leq \dots \leq X_{n:n}$ denote the order statistics of $\mathbb{X}$. Define the L-statistic

$$L_n = L_n(\mathbb{X}) = \frac{1}{n} \sum_{j=1}^{n} c_j X_{j:n}. \tag{1}$$

Here $c_1, \dots, c_n$ is a given sequence of real numbers called weights. L-statistics generalize the well-known estimators such as the sample mean (sum), trimmed means, empirical quantiles, and Gini's mean difference (each of them can be written in form (1)). Usually the weights $c_1, \dots, c_n$ are determined by the weight function $J : (0, 1) \to \mathbb{R}$ as follows

$$c_j = J\left(\frac{j}{n+1}\right), \quad 1 \leq j \leq n. \tag{2}$$

Denote $\sigma_L^2 = \mathbf{Var}\, L_n$. We present some L-statistics in more detail.

Example 1 The trimmed mean is defined as follows: for any fixed numbers $0 < t_1 < t_2 < 1$,

$$M_{t_1;t_2} = ([t_2 n] - [t_1 n])^{-1} \sum_{j=[t_1 n]+1}^{[t_2 n]} X_{j:n}, \tag{3}$$

where $[\cdot]$ represents the greatest integer function. Clearly, it is statistic (1), with the weight function $J(u) = (t_2 - t_1)^{-1} \mathbb{I}\{t_1 < u < t_2\}$. Here $\mathbb{I}\{\cdot\}$ is the indicator function. Note that the marginal case, where $t_1 = 0$ and $t_2 = 1$, represents the usual sample mean. In this case $J \equiv 1$. The trimmed means are applied in a robust estimation of a center of population $\mathcal{X}$.

Example 2 In the case of i.i.d. observations, the L-statistic, defined by the weight function $J(u) = 6u(1-u)$, is applied as an efficient estimator of the location parameter for the logistic distribution, see Chernoff et al. [23]. Therefore, if it is assumed that population $\mathcal{X}$ is obtained from the logistic distribution, the defined statistic may be useful in the estimation of a center of population $\mathcal{X}$.

Example 3 Gini's mean difference, known as a measure of dispersion,

$$U_G = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} |X_i - X_j|$$

is the U-statistic of degree 2 and it can be written in form (1) (see, e.g., Arnold et al. [4, pp. 229–230]), where $c_j = (n+1)J(j/(n+1))/(n-1)$, $1 \le j \le n$, with $J(u) = 2(2u-1)$. It is known that this statistic can be also applied to an efficient and robust against outliers estimation of the scale parameter (standard deviation) for the normal distribution, if we take $\sqrt{\pi}U_G/2$ instead of U_G. If it is assumed that population $\mathcal{X}$ is obtained from the normal distribution, this statistic may be useful in the estimation of the variance $\mathbf{Var}\, X_1$ of $\mathcal{X}$.

Note that in Examples 2 and 3, for the interpretation of L-statistics (-estimators), it was convenient to assume that a fixed finite population $\mathcal{X}$ is a random sample from an infinite population (also called a superpopulation) with a certain distribution function.

Further, when we talk about the asymptotics of L-statistics, we use centered statistics (1) with $n^{1/2}$ norming, i.e.,

$$S_n = n^{1/2}(L_n - \mathbf{E}\, L_n). \tag{4}$$

Denote $\tilde{\sigma}_n^2 = \mathbf{Var}\, S_n$. We are interested in approximations to the distribution function

$$F_n(x) = \mathbf{P}\left\{S_n \le x\tilde{\sigma}_n\right\}. \tag{5}$$

Denote

$$\tau^2 = Npq, \quad \text{where} \quad p = n/N, \quad q = 1-p, \tag{6}$$

and write

$$n_* = \min\{n, N-n\}. \tag{7}$$

The numbers τ^2 and n_* will be used in many further statements on the asymptotics of L-statistics instead of usual n. We note that τ^2 and n_* are approximately equivalent because of the inequalities $\tau^2 \le n_* \le 2\tau^2$. Clearly, if we fix the sample size n and let the population size $N \to \infty$ (the case of independent observations), then $\tau^2 \to n$.

Note that for correct formulations of asymptotic results for finite population statistics, we need to consider a sequence of populations $\mathcal{X}_r = \{x_{r,1}, \ldots, x_{r,N_r}\}$, with $N_r \to \infty$ as $r \to \infty$, and a sequence of statistics $L_{n_r}(\mathbb{X}_r)$, based on simple random samples $\mathbb{X}_r = \{X_{r,1}, \ldots, X_{r,n_r}\}$ drawn without replacement from $\mathcal{X}_r$. In order to keep the notation simple we shall skip the superscript r in what follows.

A separate case of the sample mean. First, we discuss the case of the sample mean, where $c_j \equiv 1$, $1 \le j \le n$. The following asymptotic results hold for samples drawn without replacement.

The most common approximation to (5) is the normal approximation. Write

$\sigma^2 = \mathbf{Var}\, X_1$. Let

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

denote the standard normal distribution function. We say that the random variable $\tilde{\sigma}_n^{-1} S_n$ or its distribution function $F_n(x)$ is asymptotically standard normal if, for every $x \in (-\infty, +\infty)$, we have $\lim_{n\to\infty} F_n(x) = \Phi(x)$.

Theorem 1 (Erdős and Rényi, [26]) *Assume that $N, n \to \infty$, such that $n < N$, and σ^2 remains bounded away from zero for all N. Say that, for every $\varepsilon > 0$,*

$$\mathbf{E}(X_1 - \mathbf{E}\, X_1)^2 \sigma^{-2} \mathbb{I}\{|X_1 - \mathbf{E}\, X_1| > \varepsilon\tau\sigma\} = o(1) \quad \textit{as} \quad N, n \to \infty. \tag{8}$$

Then $\tilde{\sigma}_n^{-1} S_n$ is asymptotically standard normal.

Here condition (8) is similar to the well-known Lindeberg condition, which ensures asymptotic normality in the traditional case of independent observations. It is called the Erdős–Rényi condition.

The closeness between $F_n(x)$ and $\Phi(x)$ was studied first by Bikelis [11], but the speed of convergence of (5) to the standard normal distribution is typically provided by the Berry–Esseen bound. Assuming that $\sigma^2 > 0$, we introduce the notation

$$\alpha_3 = \sigma^{-3}\, \mathbf{E}(X_1 - \mathbf{E}\, X_1)^3 \quad \text{and} \quad \beta_3 = \sigma^{-3}\, \mathbf{E}\, |X_1 - \mathbf{E}\, X_1|^3 .$$

The following theorem shows the accuracy of the normal approximation.

Theorem 2 (Höglund, [38]) *Assume that $\sigma^2 > 0$. There exists an absolute constant $C > 0$ such that, for every $1 \leq n < N$, we have*

$$\sup_{-\infty < x < +\infty} |F_n(x) - \Phi(x)| \leq \frac{C}{\tau} \beta_3.$$

It follows from Theorem 2 that, if $\mathbf{E}\, |X_1|^3 < \infty$ and also $\sigma^2 > 0$ for all N, then we have the approximation

$$\sup_{-\infty < x < +\infty} |F_n(x) - \Phi(x)| = O(\tau^{-1}) \quad \text{as} \quad \tau \to \infty.$$

In fact, Berry–Esseen bounds are of a purely theoretical significance.

One of possible ways to improve the normal approximation is an Edgeworth expansion. Typically only the first few terms of the Edgeworth expansion are taken, i.e., one- or two-term Edgeworth expansion can be convenient and sufficient for applications. Such an Edgeworth expansion was studied first by Robinson [61], but weaker conditions sufficient for the validity of Edgeworth expansions were obtained by Bloznelis [13]. Here, in the case of sample mean, and also further, we consider for illustration only one-term Edgeworth expansions. For the sample mean the one-term Edgeworth expansion is given by

$$G_n(x) = \Phi(x) - \frac{(q-p)\alpha_3}{6\tau}(x^2-1)\Phi'(x), \tag{9}$$

where $\Phi'(x)$ is the derivative of $\Phi(x)$. Depending on additional smoothness conditions, imposed on the distribution function $F_n(x)$, it is possible to obtain an improvement of the normal approximation,

$$\sup_{-\infty<x<+\infty} |F_n(x) - G_n(x)| = o(\tau^{-1}) \quad \text{as} \quad \tau \to \infty \tag{10}$$

or

$$\sup_{-\infty<x<+\infty} |F_n(x) - G_n(x)| = O(\tau^{-2}) \quad \text{as} \quad \tau \to \infty. \tag{11}$$

Given $g\colon \mathbb{R} \to \mathbb{C}$, write $\|g\|_{[a,b]} = \sup_{a<|t|<b} |g(t)|$. To obtain (10), we need the following nonlattice condition: for every $\varepsilon > 0$ and every $B > 0$, the function $\varphi(t) = \mathbf{E}\exp\{it\sigma^{-1}X_1\}$ satisfies

$$\liminf_{N,n\to\infty} \|\varphi\|_{[\varepsilon,B]} < 1. \tag{12}$$

To obtain (11), a more stringent Cramer-type condition should be used,

$$\liminf_{N,n\to\infty} \|\varphi\|_{[\varepsilon,\tau]} < 1. \tag{13}$$

Asymptotic conditions (12) and (13) are finite population analogues of the nonlattice and Cramer conditions familiar from the traditional case of i.i.d. observations.

Theorem 3 (Bloznelis, [13]) *Assume that $N, n \to \infty$, such that $n < N$, and σ^2 remains bounded away from zero for all N.*

(i) Assume that (12) *holds and* $\mathbf{E}|X_1|^3 < \infty$. *Let, for every $\varepsilon > 0$,*

$$\mathbf{E}|X_1 - \mathbf{E}X_1|^3 \sigma^{-3}\mathbb{I}\{|X_1 - \mathbf{E}X_1| > \varepsilon\tau\sigma\} = o(1) \quad \textit{as} \quad N, n \to \infty.$$

Then we get (10).

(ii) Assume that (13) *holds and* $\mathbf{E}|X_1|^4 < \infty$. *Then we obtain* (11).

Since L-statistics can be viewed as a certain generalization of the sample mean, one can expect that conditions, sufficient for the corresponding asymptotic statements, should be similar, but with some additional restrictions to the weights $c_1, \ldots, c_n$.

A class of symmetric statistics. L-statistics is a subclass of the more general class of symmetric statistics. The statistic $T = t(\mathbb{X})$ is called symmetric, if the function $t(\cdot)$ is invariant under permutations of its arguments $X_1, \ldots, X_n$. Symmetric statistics also include smooth functions of (multivariate) sample means, U-statistics and many others. A general asymptotic theory is well developed now not only for the case of i.i.d. observations, see Bentkus et al. [7], but also in the case of samples drawn without replacement, we refer to Bloznelis and Götze [20].

An asymptotic behavior of many important symmetric statistics (including L-statistics) differs not so much from that of the simplest linear statistic (the sample mean is an example), in the sense that usually it is possible to write

$$T - \mathbf{E}\, T = U_1 + R_1, \tag{14}$$

where U_1 is a linear statistic and R_1 is a stochastically smaller statistic. Then, under proper normalization, in (14) T is asymptotically standard normal if its linear part U_1 is asymptotically standard normal, and R_1 is a degenerate statistic as the sample size increases. Statistic (14), where $R_1 = o_P(1)$, is also called an asymptotically linear statistic. The method, used to decompose a symmetric statistic as in (14), usually depends on a form (properties) of the statistic T. A very common method is Taylor's expansion of the statistic, which is very suitable for many simple and popular statistics, e.g., for the members of a subclass of smooth functions of sample means. Unfortunately, this method cannot be applied, e.g., for many of L-statistics and U-statistics.

An alternative method is Hoeffding's decomposition. For U-statistics, based on i.i.d. observations, it was introduced by Hoeffding [37]. In the case of samples drawn without replacement, Hoeffding decompositions of U-statistics of the fixed degree m were studied in Zhao and Chen [78]. In the general case of symmetric statistics, based on the samples drawn without replacement, a decomposition of this type was studied by Bloznelis and Götze [20]. Hence, if it is aimed to prove the asymptotic normality of the symmetric statistic S_n (e.g., L-statistic defined by (4)), one can write, by Theorem 1 of [20], that

$$S_n = U_1 + R_1, \tag{15}$$

where the linear part

$$U_1 = \sum_{i=1}^{n} g_1(X_i)$$

and the remainder term R_1 are centered and uncorrelated. Here the variance of R_1 is bounded as follows: $\mathbf{E}\, R_1^2 \leq \delta_2$, where, in the case of the mentioned symmetric statistics, typically $\delta_2 = O(n_*^{-1})$ as $n_* \to \infty$. We will give later a more detailed description of the Hoeffding decomposition in Section 1.1. The main result on the asymptotic normality of the symmetric statistics is the following statement.

Proposition 4 (Bloznelis and Götze, [20]) *Assume that $n_* \to \infty$ and $\tilde{\sigma}_n$ remains bounded away from zero for all n_*. Let $\delta_2 = o(1)$ and for every $\varepsilon > 0$,*

$$n_* \, \mathbf{E}\, g_1^2(X_1) \mathbb{I}\{g_1^2(X_1) > \varepsilon\} = o(1) \quad \textit{as} \quad n_* \to \infty. \tag{16}$$

Then $\tilde{\sigma}_n^{-1} S_n$ is asymptotically standard normal.

An improvement of the normal approximation is provided by the one-term Edgeworth expansion. To write it, we need more terms of the Hoeffding decomposition. By Theorem 1 of Bloznelis and Götze [20],

$$S_n = U_1 + U_2 + R_2, \tag{17}$$

where the second term

$$U_2 = \sum_{1 \le i < j \le n} g_2(X_i, X_j)$$

is a U-statistic of degree 2, also called a quadratic part of the decomposition. Similarly as in the case of short expansion (15), U_1, U_2 and the remainder term R_2 are centered and mutually uncorrelated. Now $\mathbf{E}\, R_2^2 \le n_*^{-1} \delta_3$, where usually $\delta_3 = O(n_*^{-1})$ as $n_* \to \infty$. The first two terms of (17) are sufficient to write the one-term Edgeworth expansion for the distribution function $F_n(x)$, i.e., by Bloznelis and Götze [20],

$$G_n(x) = \Phi(x) - \frac{(q-p)\alpha + 3\kappa}{6\tau}(x^2 - 1)\Phi'(x), \tag{18}$$

where

$$\alpha = \sigma_1^{-3}\, \mathbf{E}\, g_1^3(X_1) \quad \text{and} \quad \kappa = \sigma_1^{-3} \tau^2\, \mathbf{E}\, g_2(X_1, X_2) g_1(X_1) g_1(X_2), \tag{19}$$

with $\sigma_1^2 = \mathbf{E}\, g_1^2(X_1)$. Note that the form of one-term Edgeworth expansion (18) for the general symmetric statistics differs from the corresponding expansion, in the case of the sample mean, see (9) above, only by the additional parameter κ, which reflects the influence of the quadratic part U_2 of the statistic (decomposition). The moment and smoothness conditions, sufficient for the approximation

$$\sup_{-\infty < x < +\infty} |F_n(x) - G_n(x)| = o(n_*^{-1/2}) \quad \text{as} \quad n_* \to \infty \tag{20}$$

or

$$\sup_{-\infty<x<+\infty} |F_n(x) - G_n(x)| = O(n_*^{-1}) \quad \text{as} \quad n_* \to \infty, \tag{21}$$

in the case of symmetric statistics, are also similar to those that are sufficient in the case of the sample mean. Now, by Bloznelis and Götze [20], we need to require for the validity of (20) that, for every $\varepsilon > 0$ and every $B > 0$, the function $\varphi(t) = \mathbf{E} \exp\{it\sigma_1^{-1} g_1(X_1)\}$ should satisfy

$$\liminf_{n_* \to \infty} \|\varphi\|_{[\varepsilon, B]} < 1, \tag{22}$$

and, respectively, for the validity of (21), we need

$$\liminf_{n_* \to \infty} \|\varphi\|_{[\varepsilon, \tau]} < 1. \tag{23}$$

We see that asymptotic conditions (22) and (23) are imposed on the linear part of the symmetric statistic only, as well as in the case of the usual sample mean. For the proof of (21), we also need the cubic part $U_3 = \sum_{1 \le i < j < k \le n} g_3(X_i, X_j, X_k)$ of the decomposition $S_n = U_1 + U_2 + U_3 + R_3$, where $\mathbf{E}\, R_3^2 \le n_*^{-2} \delta_4$ (see [20]). Let us introduce the moments

$$\begin{aligned} &\beta_s = \mathbf{E} \left| n_*^{1/2} g_1(X_1) \right|^s, \quad \gamma_s = \mathbf{E} \left| n_*^{3/2} g_2(X_1, X_2) \right|^s, \\ &\zeta_s = \mathbf{E} \left| n_*^{5/2} g_3(X_1, X_2, X_3) \right|^s. \end{aligned} \tag{24}$$

The following theorem is on the validity of approximations (20) and (21).

Theorem 5 (Bloznelis and Götze, [20]) *Assume that $n_* \to \infty$ and $\tilde{\sigma}_n$ remains bounded away from zero for all n_*.*

(i) Assume that (22) *holds, $\delta_3 = o(n_*^{-1/2})$ and, for some $\delta > 0$, the moments $\beta_{3+\delta}$ and $\gamma_{2+\delta}$ are bounded as $n_* \to \infty$. Then* (20) *holds.*

(ii) Assume that (23) *holds, $\delta_4 = O(n_*^{-1})$ and, the moments β_4, γ_4, ζ_2 are bounded as $n_* \to \infty$. Then* (21) *holds.*

The conditions, imposed on the quantities δ_k, $k = 2, 3, 4$ and moments (24) in Proposition 4 and Theorem 5, are quite general and it is difficult to verify them. Chapter 2 of this dissertation is devoted to a simplification of these general conditions, in the case of L-statistics. In fact, we will replace these conditions by the respective sufficient conditions expressed in terms of the weights $c_1, \ldots, c_n$ and moments of X_1. Moreover, in Chapter 1, we give explicit and quite convenient expressions of the functions $g_1(\cdot)$ and $g_2(\cdot, \cdot)$, i.e., explicit formulas of the parameters α and κ that define the Edgeworth correction term in (18). It opens new ways of a more efficient practical use of Edgeworth expansions, presented in Chapters 3 and 4.

Note that there are a few results on Edgeworth expansions, obtained before the work of Bloznelis and Götze [20], for some important subclasses of finite population symmetric statistics. One can mention the paper of Babu and Singh [5] on smooth functions of sample means, where the Taylor expansion was applied, and the work of Kokic and Weber [44] on U-statistics, where other methods than the finite population orthogonal decomposition were used as well. It is demonstrated in [20] that, in both cases, the orthogonal decomposition applies well, and gives practically simpler conditions for the validity of the one-term Edgeworth expansions.

L-statistics in the case of i.i.d. observations. In fact, in the case of samples drawn without replacement, there are no works on asymptotic approximations to distributions of L-statistics, except, e.g., the paper of Shao [64] on the asymptotic normality of L-statistics under more general sampling models, and the work of Chatterjee [22] on the asymptotic normality of the sample quantile. In the case of i.i.d. observations, L-statistics were studied by a number of authors. Strong laws of large numbers were obtained, e.g., by Wellner [76], van Zwet [73] and Norvaiša [52]. Laws of the iterated logarithms were established, e.g., by Wellner [77], Lea and Puri [45] and Norvaiša and Zitikis [53]. Asymptotic normality under various conditions was shown by Chernoff et al. [23], Shorack [66], Stigler [69] and Mason [50], among others. See also Serfling [63, Chapter 8]. Berry–Esseen bounds were obtained by Bjerve [12], Helmers [33], van Zwet [74], and others. Large deviations were considered by Vandemaele and Veraverbeke [75], Bentkus and Zitikis [8], Aleškevičienė [3] and Gao and Zhao [27]. Edgeworth expansions for L-statistics were established by Helmers [34], Putter and van Zwet [59] (see also Putter [58]), Gribkova and Helmers [28, 30], Alberink et al. [2], Maesono [47] and Maesono and Penev [48].

We assume (Theorem 15 in Chapter 2) that the weight function $J\colon (0,1) \to \mathbb{R}$ is sufficiently smooth, and we impose very mild conditions on the finite population $\mathcal{X}$. These assumptions, sufficient for the asymptotic normality of L-statistics, are similar to that obtained by Stigler [69] in the i.i.d. case. The validity of Edgeworth expansion (Theorem 17) is ensured by similar but more stringent conditions for $J(\cdot)$ and $\mathcal{X}$. Our conditions are similar to that used in the i.i.d. situation, see Helmers [34] and Putter [58].

The structure of the thesis results

The thesis consists of four chapters and the bibliography. In most cases, the proofs of results are given at the end of each section.

- In Chapter 1, we obtain the form of the first three terms of the Hoeffding decomposition expressed explicitly via the weights $c_1, \dots, c_n$ and their differences, see Section 1.3. We similarly express the components of remainder terms of the decomposition.

- The main applications of the orthogonal decomposition are given in Chapter 2. Section 2.1 presents a new upper bound for the variance of the sample minimum and maximum. This bound is optimal in the form provided. Similar bounds are shown for the other order statistics. Sections 2.2 and 2.3 are on the asymptotic normality and the validity of Edgeworth expansion, respectively. In addition to the asymptotic normality of L-statistics of a more general form, we also consider the case of the trimmed mean.

- In Chapter 3, we consider the estimation of the variance and parameters α and κ that define the Edgeworth expansion of L-statistic. We examine two competitive methods: the classical jackknife and the finite population bootstrap of Booth et al. [21]. In the case of bootstrap, we give an exact formula of the bootstrap variance estimator, i.e., we reduce the computational burden and eliminate the approximation error, typically present in resampling approximations based on simulation. We also present similar efficient formulas for calculating the bootstrap estimators of α and κ directly from the sample.

- In Chapter 4, we consider several variants of the Edgeworth expansion, which are more close to practice: an Edgeworth expansion for the Studentized L-statistic, empirical Edgeworth expansions, and (related in a certain sense) non-parametric bootstrap approximations. We discuss their second-order correctness and compare their efficiencies for various L-statistics in the simulation study. In Section 4.2, we present a generalization of one-term Edgeworth expansions to the case of stratified simple random samples drawn without replacement, where the L-statistics are quantiles of the stratified sample. We give an explicit expression of the approximation to distribution of the quantile, and also its empirical version based on bootstrap.

Methods

The properties of statistics are explored using the Hoeffding decomposition. In the proofs of results, combinatorial and probabilistic methods are applied.

Novelty

New formulas of the parameters that define Edgeworth expansions are obtained, which are convenient for the construction of their exact bootstrap estimators.

Simple sufficient conditions are established, which ensure the improvement of the normal approximation to distribution of an L-statistic by the one-term Edgeworth expansion. An exact bootstrap variance estimator is obtained. The new optimal upper bound for variances of the sample minimum and maximum is constructed in the case of a sample drawn without replacement.

Maintaining statements

- A one-term Edgeworth expansion was constructed.
- The optimal upper bound for variances of sample extremes was obtained.
- Sufficient conditions for the asymptotic normality and the validity of the one-term Edgeworth expansion were expressed in terms of smoothness of the weight function that defines the statistics and boundedness of the moments of population. Special conditions sufficient for the asymptotic normality of the trimmed means were also presented.
- Exact formulas of bootstrap estimators of the variance and parameters that define the Edgeworth expansion of an L-statistic were obtained. Thus, the additional approximation errors, typically present in resampling approximations based on simulation, were eliminated.
- The simulation study has showed that the quality of Edgeworth approximations depends on the smoothness of the weight function of a statistic. It is also showed that, in the cases where the weight function is not smooth, empirical Edgeworth expansions with bootstrap estimates of the parameters are more efficient than the corresponding expansions with jackknife estimates.

Publications and presentations

The main results are published in the following articles:

1. A. Čiginas. Second-order approximations of finite population L-statistics. *Statistics*, 2011. (submitted)
2. A. Čiginas. An Edgeworth expansion for finite population L-statistics. *Lith. Math. J.*, 2011. (to appear); see also `arXiv:1103.4220v2 [math.ST]`.
3. A. Čiginas. An exact bootstrap for variance of finite-population L-statistic. *Lith. Math. J.*, 51:322–329, 2011.
4. A. Čiginas and T. Rudys. Approximations to distribution of median in stratified samples. *Liet. Mat. Rink. LMD darbai*, 52, 2011. (to appear)

5. A. Čiginas. Bootstrap, jackknife and Edgeworth approximations for finite population L-statistics. *Liet. Mat. Rink. LMD darbai*, 51:391–396, 2010.

6. A. Čiginas. Orthogonal decomposition of finite population L-statistics. *Liet. Mat. Rink. LMD darbai*, 50:287–292, 2009.

Several presentations at conferences were given on the topics of the thesis:

1. A. Čiginas. *On an optimal bound for the variance of sample maximum.* The Third Baltic-Nordic Conference on Survey Statistics, 13–17 June 2011, Norrfällsviken, Sweden.

2. A. Čiginas and T. Rudys. *Approximations to distribution of median in stratified samples.* LII Conference of the Lithuanian Mathematical Society, The General Jonas Žemaitis Military Academy of Lithuania, 16–17 June 2011, Vilnius, Lithuania.

3. A. Čiginas. *Bootstrap for variance of finite population L-statistic.* Workshop on Survey Sampling Theory and Methodology, Vilnius University, 23–27 August 2010, Vilnius, Lithuania.

4. A. Čiginas. *An Edgeworth expansion for finite population L-statistics.* 10th International Vilnius Conference on Probability and Mathematical Statistics, Vilnius University, 28 June – 02 July 2010, Vilnius, Lithuania.

5. A. Čiginas. *Bootstrap, jackknife and Edgeworth approximations for finite population L-statistics.* LI Conference of the Lithuanian Mathematical Society, Šiauliai University, 17–18 June 2010, Šiauliai, Lithuania.

6. A. Čiginas. *Orthogonal decomposition of finite population L-statistics.* The Baltic-Nordic-Ukrainian Summer School on Survey Statistics, 23–27 August 2009, Kyiv, Ukraine.

7. A. Čiginas. *Orthogonal decomposition of finite population L-statistics.* L Conference of the Lithuanian Mathematical Society, Vilnius University Institute of Mathematics and Informatics, 18–19 June 2009, Vilnius, Lithuania.

Chapter 1

Hoeffding decomposition

1.1 General formulas

Here we give the basic results of Bloznelis and Götze [20] on the Hoeffding decomposition of the symmetric statistics.

Since L-statistics are the symmetric statistics, our analysis of an asymptotic behavior of statistic (1) is based on the decomposition

$$L_n = \mathbf{E}\, L_n + U_1 + \cdots + U_n, \tag{1.1}$$

where

$$U_m = U_m(L_n) = \sum_{1 \le i_1 < \cdots < i_m \le n} g_m(X_{i_1}, \ldots, X_{i_m}), \qquad 1 \le m \le n.$$

Here symmetric and centered kernels g_m, $1 \le m \le n$ are certain linear combinations of conditional expectations

$$h_j(x_{k_1}, \ldots, x_{k_j}) = \mathbf{E}\left(L_n - \mathbf{E}\, L_n \,\Big|\, X_1 = x_{k_1}, \ldots, X_j = x_{k_j}\right), \qquad 1 \le j \le m,$$

such that U_m, U-statistics of order m, are mutually uncorrelated. The decomposition in (1.1) is also called an orthogonal decomposition of L_n. Bloznelis and Götze [20] provides expressions for the first three kernels of the decomposition as follows

$$g_1(x) = \frac{N-1}{N-n} h_1(x), \tag{1.2}$$

$$g_2(x,y) = \frac{N-2}{N-n}\frac{N-3}{N-n-1}\Big(h_2(x,y) - \frac{N-1}{N-2}\big(h_1(x) + h_1(y)\big)\Big), \tag{1.3}$$

$$
\begin{aligned}
g_3(x,y,z) = \frac{N-3}{N-n}\frac{N-4}{N-n-1}\frac{N-5}{N-n-2}\Big(& h_3(x,y,z) \\
& - \frac{N-2}{N-4}\big(h_2(x,y)+h_2(x,z)+h_2(y,z)\big) \\
& + \frac{N-1}{N-3}\frac{N-2}{N-4}\big(h_1(x)+h_1(y)+h_1(z)\big)\Big).
\end{aligned} \tag{1.4}
$$

See [20], on formula of the kernel of order m. Denote

$$
\sigma_m^2 = \mathbf{E}\, g_m^2(X_1,\dots,X_m), \quad 1 \le m \le n.
$$

Using the fact that the components of decomposition (1.1) are mutually uncorrelated, it is shown in [20] that the variance of (1) can be written as

$$
\sigma_L^2 = \sum_{m=1}^{n} \binom{n}{m}\binom{N-n}{m}\binom{N-m}{m}^{-1}\sigma_m^2. \tag{1.5}
$$

Decomposition (1.1) is a stochastic expansion of an L-statistic and the first few terms of the decomposition, defined by the kernels in (1.2)–(1.4) above, can be quite an excellent approximation to L_n, i.e., the first few terms of the sum in (1.5) can contain very large part of σ_L^2. In order to control the accuracy of approximation, one can use the smoothness conditions defined as follows. 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 population $\mathcal{X}$. For $j = 1,\dots,N-n$ denote $X_j' = X_{n+j}$. Define

$$
D^j L_n = L_n(X_1,\dots,X_n) - L_n(X_1,\dots,X_{j-1},X_{j+1},\dots,X_n,X_j').
$$

Higher order difference operations are defined recursively:

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

They are symmetric; that is, $D^{j_1,j_2}L_n = D^{j_2,j_1}L_n$, etc. Write

$$
\delta_k = \delta_k(L_n) = \mathbf{E}\left(n_*^{(k-1)}\mathbb{D}_k L_n\right)^2, \qquad \mathbb{D}_k L_n = D^{1,2,\dots,k}L_n, \quad 1 \le k < n_*.
$$

Then the following theorem holds.

Theorem 6 (Bloznelis and Götze, [20]) *For $1 \le k < n_*$, we have*

$$
L_n = \mathbf{E}\, L_n + U_1 + \dots + U_k + R_k, \quad \textit{with} \quad \mathbf{E}\, R_k^2 \le n_*^{-(k-1)}\delta_{k+1}.
$$

Now we have defined all general tools, which are necessary for further analysis of Proposition 4 and Theorem 5 on symmetric statistics, presented in Introduction.

1.2 Auxiliary lemmas

The binomial coefficients

$$\binom{m}{k} = \frac{m!}{k!(m-k)!}$$

will appear naturally in the further text, since, in the case of samples drawn without replacement, we usually need to count a number of ways to choose k elements from a set of m elements. For convenience, in this section and further, we use the conventions that, for integers $m \geq 0$ and $k \geq 1$,

$$\binom{m}{-k} = 0 \quad \text{and} \quad \binom{m}{m+k} = 0. \tag{1.6}$$

Next, we collect some well-known binomial identities. We give them in the following lemma without a proof.

Lemma 7 *For integers m, k, j, p and $m_1, \ldots, m_T, k_1, \ldots, k_T$ the following identities hold.*

(i) Let $m \geq 1$. Then

$$\binom{m}{k} = \binom{m-1}{k-1} + \binom{m-1}{k}. \tag{1.7}$$

(ii) Let $1 \leq k \leq m$. Then

$$\binom{m}{k} = \frac{m}{k}\binom{m-1}{k-1}. \tag{1.8}$$

(iii) Let $0 \leq k \leq m$. Then

$$\binom{m}{k} = \binom{m}{m-k}. \tag{1.9}$$

(iv) Let $0 \leq k \leq m$. Then

$$\sum_{j=k}^{m} \binom{j}{k} = \binom{m+1}{k+1}. \tag{1.10}$$

(v) Let $0 \leq j \leq k \leq m$. Then

$$\sum_{p=0}^{m} \binom{p}{j}\binom{m-p}{k-j} = \binom{m+1}{k+1}. \tag{1.11}$$

(vi) Let $0 \le k \le m$ and $0 \le p \le m$. Then

$$\sum_{j=0}^{k} \binom{p}{j} \binom{m-p}{k-j} = \binom{m}{k}. \tag{1.12}$$

(vii) Let $T \ge 3$, $0 \le k_t \le m_t$ for $t = 1, \dots, T$, and $m_1 + \cdots + m_T = m$. Then

$$\sum_{k_1+\cdots+k_T=k} \binom{m_1}{k_1} \cdots \binom{m_T}{k_T} = \binom{m}{k}. \tag{1.13}$$

Here (1.12) is called the Vandermonde identity, and (1.13) is its generalization.

The second simple lemma is useful for calculations of probability distributions of the order statistics in the case where the values $x_1, \dots, x_N$ of the population $\mathcal{X}$ are not necessarily all distinct.

Lemma 8 (Balakrishnan et al., [6]) *Assume that $\mathcal{Z} = \{1, \dots, N\}$ and consider a simple random sample $Z_1, \dots, Z_n$ of size $n < N$, drawn without replacement from $\mathcal{Z}$. Then, the ordered samples $X_{1:n} \le \cdots \le X_{n:n}$ from $\mathcal{X}$ and $Z_{1:n} < \cdots < Z_{n:n}$ from $\mathcal{Z}$ are related through*

$$(X_{1:n}, \dots, X_{n:n}) \stackrel{d}{=} (g(Z_{1:n}), \dots, g(Z_{n:n})),$$

where $g \colon \mathcal{Z} \to \mathcal{X}$ is given by $g(k) = x_k$, $k = 1, \dots, N$.

Assume that $x_1 \le \cdots \le x_N$. Introduce the number $x_0 := x_1$ and define $X_{0:n} := x_0$ so that, almost surely, $X_{0:n} \le X_{j:n}$ for each $1 \le j \le n$. Let $\mathbf{\Delta}_{j:n} = X_{j+1:n} - X_{j:n}$, $0 \le j \le n-1$ denote the sample spacings. We will need expressions of their moments $\mathbf{E}\,\mathbf{\Delta}_{u:n}$, $\mathbf{E}\,\mathbf{\Delta}_{u:n}^2$, $0 \le u \le n-1$, and $\mathbf{E}\,\mathbf{\Delta}_{u:n}\mathbf{\Delta}_{v:n}$, $0 \le u < v \le n-1$ in terms of the population differences $\triangle_i = x_{i+1} - x_i$, $i = 1, \dots, N-1$. Our expressions, see Lemma 9 below, are similar to those obtained by Jones and Balakrishnan [42], in the case of i.i.d. observations. Denote

$$h_i(u) = \binom{N}{n}^{-1} \binom{i}{u} \binom{N-i}{n-u}, \qquad 0 \le u \le n-1, \quad 1 \le i \le N-1, \tag{1.14}$$

$$h_{ij}(u) = \binom{N}{n}^{-1} \binom{i}{u} \binom{N-j}{n-u}, \qquad 0 \le u \le n-1, \quad 1 \le i < j \le N-1, \tag{1.15}$$

and

$$h_{ij}(u, v) = \binom{N}{n}^{-1} \binom{i}{u} \binom{j-i}{v-u} \binom{N-j}{n-v}, \qquad 0 \le u < v \le n-1, \quad 1 \le i < j \le N-1. \tag{1.16}$$

Lemma 9 *We have*

$$\mathbf{E}\,\boldsymbol{\Delta}_{u:n} = \sum_{i=1}^{N-1} h_i(u)\ \triangle_i, \qquad 0 \le u \le n-1, \tag{1.17}$$

$$\mathbf{E}\,\boldsymbol{\Delta}_{u:n}^2 = \sum_{i=1}^{N-1} h_i(u)\ \triangle_i^2 + 2 \sum_{1\le i<j\le N-1} h_{ij}(u)\ \triangle_i\triangle_j, \qquad 0 \le u \le n-1, \tag{1.18}$$

$$\mathbf{E}\,\boldsymbol{\Delta}_{u:n}\boldsymbol{\Delta}_{v:n} = \sum_{1\le i<j\le N-1} h_{ij}(u,v)\ \triangle_i\triangle_j, \qquad 0 \le u < v \le n-1. \tag{1.19}$$

Proof of Lemma 9

Here we need slightly modified Lemma 8. Now we take the ordered samples $X_{0:n} \le X_{1:n} \le \cdots \le X_{n:n}$ from $\mathcal{X}$ and $Z_{0:n} < Z_{1:n} < \cdots < Z_{n:n}$, where $Z_{0:n} := 0$, from $\mathcal{Z}$. They are related through

$$(X_{0:n}, X_{1:n}, \dots, X_{n:n}) \stackrel{d}{=} (\tilde{g}(Z_{0:n}), \tilde{g}(Z_{1:n}), \dots, \tilde{g}(Z_{n:n})),$$

where $\tilde{g}\colon \mathcal{Z} \cup \{0\} \to \mathcal{X} \cup \{x_0\}$ is given by $\tilde{g}(k) = x_k$, $k = 0, \dots, N$.

First, we will prove (1.17) for $1 \le u \le n-1$. Since

$$\boldsymbol{\Delta}_{u:n} \stackrel{d}{=} \tilde{g}(Z_{u+1:n}) - \tilde{g}(Z_{u:n})$$

and, for $1 \le k < l \le N$,

$$\mathbf{P}\{Z_{u:n} = k, Z_{u+1:n} = l\} = \binom{k-1}{u-1}\binom{l-k-1}{0}\binom{N-l}{n-u-1} \Big/ \binom{N}{n},$$

we have

$$\begin{aligned}
\mathbf{E}\,\boldsymbol{\Delta}_{u:n} &= \mathbf{E}\left(\tilde{g}(Z_{u+1:n}) - \tilde{g}(Z_{u:n})\right) \\
&= \binom{N}{n}^{-1} \sum_{1\le k<l\le N} \binom{k-1}{u-1}\binom{N-l}{n-u-1}(x_l - x_k) \\
&= \binom{N}{n}^{-1} \sum_{i=1}^{N-1} \left\{ \sum_{k=1}^{i} \binom{k-1}{u-1} \sum_{l=i+1}^{N} \binom{N-l}{n-u-1} \right\} \triangle_i,
\end{aligned}$$

where the last identity is obtained by writing

$$x_l - x_k = \sum_{i=k}^{l-1} \triangle_i$$

and collecting the terms with $\triangle_i$. Finally, application of identity (1.10) gives (1.17) for $1 \le u \le n-1$.

For $u = 0$, we observe that, for $1 \le l \le N$,

$$\mathbf{P}\{Z_{0:n} = 0, Z_{1:n} = l\} = \mathbf{P}\{Z_{1:n} = l\} = \binom{l-1}{0}\binom{N-l}{n-1} \Big/ \binom{N}{n}.$$

The remaining part of the proof is almost the same.

The proof of (1.18) uses the same arguments as that of (1.17). Here we also need to consider separately the cases $u = 0$ and $1 \le u \le n-1$. Here we apply the expansion

$$(x_l - x_k)^2 = \sum_{i=k}^{l-1} \triangle_i^2 + 2 \sum_{k \le i < j \le l-1} \triangle_i \triangle_j$$

and collect the terms with $\triangle_i^2$ and $\triangle_i \triangle_j$.

We will prove (1.19). Here we need to consider more separate cases. Let $1 \le u < v \le n-1$ and $u < v-1$. Since

$$\mathbf{\Delta}_{u:n}\mathbf{\Delta}_{v:n} \stackrel{d}{=} \left(\tilde{g}(Z_{u+1:n}) - \tilde{g}(Z_{u:n})\right)\left(\tilde{g}(Z_{v+1:n}) - \tilde{g}(Z_{v:n})\right)$$

and, for $1 \le k < l < s < t \le N$,

$$\begin{aligned}
&\mathbf{P}\{Z_{u:n} = k, Z_{u+1:n} = l, Z_{v:n} = s, Z_{v+1:n} = t\} \\
&\quad = \binom{k-1}{u-1}\binom{l-k-1}{0}\binom{s-l-1}{v-u-2}\binom{t-s-1}{0}\binom{N-t}{n-v-1} \Big/ \binom{N}{n},
\end{aligned}$$

we have

$$\begin{aligned}
&\mathbf{E}\,\mathbf{\Delta}_{u:n}\mathbf{\Delta}_{v:n} = \mathbf{E}\left(\tilde{g}(Z_{u+1:n}) - \tilde{g}(Z_{u:n})\right)\left(\tilde{g}(Z_{v+1:n}) - \tilde{g}(Z_{v:n})\right) \\
&\quad = \binom{N}{n}^{-1} \sum_{1 \le k < l < s < t \le N} \binom{k-1}{u-1}\binom{s-l-1}{v-u-2}\binom{N-t}{n-v-1}(x_l - x_k)(x_t - x_s) \\
&\quad = \binom{N}{n}^{-1} \sum_{1 \le i < j \le N-1} \left\{ \sum_{k=1}^{i} \binom{k-1}{u-1} \sum_{i+1 \le l < s \le j} \binom{s-l-1}{v-u-2} \right. \\
&\qquad\qquad\qquad \left. \times \sum_{t=j+1}^{N} \binom{N-t}{n-v-1} \right\} \triangle_i \triangle_j,
\end{aligned}$$

where the last identity is obtained by writing

$$(x_l - x_k)(x_t - x_s) = \sum_{i=k}^{l-1} \sum_{j=s}^{t-1} \triangle_i \triangle_j$$

and collecting the terms with $\triangle_i \triangle_j$. Finally, in the braces of the last term, we apply (1.10), and we do it twice for the middle sum.

Now let $1 \le u < v \le n-1$ and $u = v-1$. Here, for $1 \le k < l < s \le N$,

$$\begin{aligned}
&\mathbf{P}\left\{Z_{v-1:n} = k, Z_{v:n} = l, Z_{v+1:n} = s\right\} \\
&\quad = \binom{k-1}{v-2}\binom{l-k-1}{0}\binom{s-l-1}{0}\binom{N-s}{n-v-1}\Big/\binom{N}{n}.
\end{aligned}$$

The rest of the proof is very similar to that of the previous case.

For $u = 0$, we need to consider separately the cases $2 \le v \le n-1$ and $v = 1$. The proof of these special cases is very similar to that of the previous cases of (1.19); therefore, we omit it.

1.3 Explicit expressions

1.3.1 Kernels

Consider statistic (1). We assume that, without loss of generality, the values of the population $\mathcal{X}$ are arranged in non-decreasing order, i.e., $x_1 \le \cdots \le x_N$. Given $0 \le m \le n$ and $1 \le k_1 < \cdots < k_m \le N$, introduce the event

$$A_m = A_{x_{k_1}\cdots x_{k_m}} = \{X_1 = x_{k_1}, \ldots, X_m = x_{k_m}\}.$$

For convenience of notation, we define $k_0 := 0$ and $k_{m+1} := N+1$. Introduce numbers $x_0 := x_1$ and $x_{N+1} := x_N$ and define $X_{0:n} := x_0$ and $X_{n+1:n} := x_{N+1}$, so that, almost surely, $X_{0:n} \le X_{j:n} \le X_{n+1:n}$ for each $1 \le j \le n$. In the proof of Theorem 11 below we represent order statistics by sums of sample spacings

$$X_{j:n} = \sum_{r=0}^{j-1} \mathbf{\Delta}_{r:n} + x_0, \quad 1 \le j \le n. \tag{1.20}$$

Here $\mathbf{\Delta}_{r:n} = X_{r+1:n} - X_{r:n}$, $0 \le r \le n$ denote the sample spacings. Write $\triangle_i = x_{i+1} - x_i$, $0 \le i \le N$.

Lemma 10 *For any $m = 0, \ldots, n$ and $r = 0, \ldots, n$ we have*

$$\mathbf{E}\left(\mathbf{\Delta}_{r:n} \,|\, A_m\right) = \binom{N-m}{n-m}^{-1} \sum_{s=1}^{m+1} \sum_{i=k_{s-1}}^{k_s - 1} \binom{i-s+1}{r-s+1}\binom{N-i-m+s-1}{n-r-m+s-1} \triangle_i\,.$$

Define differences of the weights $c_1, \ldots, c_n$ recursively by

$$\Delta^0(c_j) = c_j, \quad \Delta^1(c_j) = c_j - c_{j-1}$$

and

$$\Delta^v(c_j) = \Delta^1(\Delta^{v-1}(c_j)), \quad \text{for} \quad v = 2, \ldots, n-1.$$

Denote by

$$\mathcal{H}_{N,n,i}(j) = \binom{i}{j}\binom{N-i}{n-j} \Big/ \binom{N}{n}$$

the probability that a hypergeometric random variable with parameters N, n and i attains the value j. Denote $[N]_j = N(N-1)\cdots(N-j+1)$. Next we give explicit and comparatively simple expressions of kernels (1.2)–(1.4).

Theorem 11

(i) For $1 \le k \le N$

$$g_1(x_k) = -n^{-1} \sum_{j=1}^{n} \Delta^0(c_j) \sum_{i=1}^{N-1} \varphi_k(i) \mathcal{H}_{N-2,n-1,i-1}(j-1) \, \triangle_i, \tag{1.21}$$

where

$$\varphi_k(i) = \begin{cases} -i/N & \text{if } 1 \le i < k, \\ 1 - i/N & \text{if } k \le i < N. \end{cases} \tag{1.22}$$

(ii) For $1 \le k < l \le N$

$$g_2(x_k, x_l) = -n^{-1} \sum_{j=2}^{n} \Delta^1(c_j) \sum_{i=1}^{N-1} \phi_{k,l}(i) \mathcal{H}_{N-4,n-2,i-2}(j-2) \, \triangle_i, \tag{1.23}$$

where

$$\phi_{k,l}(i) = \begin{cases} i(i-1)/B_2 & \text{if } 1 \le i < k, \\ -(i-1)(N-i-1)/B_2 & \text{if } k \le i < l, \\ (N-i-1)(N-i)/B_2 & \text{if } l \le i < N, \end{cases} \tag{1.24}$$

with $B_2 = [N-1]_2$.

(iii) For $1 \le k < l < m \le N$

$$g_3(x_k, x_l, x_m) = -n^{-1} \sum_{j=3}^{n} \Delta^2(c_j) \sum_{i=1}^{N-1} \theta_{k,l,m}(i) \mathcal{H}_{N-6,n-3,i-3}(j-3) \, \triangle_i, \tag{1.25}$$

where

$$\theta_{k,l,m}(i) = \begin{cases} -i(i-1)(i-2)/B_3 & \text{if } 1 \le i < k, \\ (i-1)(i-2)(N-i-2)/B_3 & \text{if } k \le i < l, \\ -(i-2)(N-i-2)(N-i-1)/B_3 & \text{if } l \le i < m, \\ (N-i-2)(N-i-1)(N-i)/B_3 & \text{if } m \le i < N, \end{cases} \tag{1.26}$$

with $B_3 = [N-2]_3$.

Proof of Lemma 10

Note that the separate case $m = 0$, for $0 \le r \le n-1$, is already proved in Lemma 9. Here we give a different proof of this case.

Assume that $x_1 < \cdots < x_N$. Then, for any $m = 0, \ldots, n$ and $r = 0, \ldots, n+1$, straightforward combinatorial calculations give

$$\mathbf{E}\left(X_{r:n} \,|\, A_m\right) = \binom{N-m}{n-m}^{-1} \left[\sum_{s=1}^{m+1} \sum_{i=k_{s-1}+1}^{k_s-1} \binom{i-s}{r-s} \binom{N-i-m+s-1}{n-r-m+s-1} x_i \right.$$
$$\left. + \sum_{s=0}^{m+1} \binom{k_s - s}{r-s} \binom{N-k_s-m+s}{n-r-m+s} x_{k_s} \right].$$

The key idea is for $r = 0, \ldots, n$ to note that, by (1.7),

$$\mathbf{E}\left(X_{r+1:n} \,|\, A_m\right) = \binom{N-m}{n-m}^{-1} \left[\sum_{s=1}^{m+1} \sum_{i=k_{s-1}+1}^{k_s-1} \binom{i-s}{r-s+1} \delta'_{m,s,i}(r) x_i \right.$$
$$\left. + \sum_{s=0}^{m+1} \binom{k_s - s}{r-s+1} \binom{N-k_s-m+s}{n-r-m+s-1} x_{k_s} \right],$$

where

$$\delta'_{m,s,i}(r) = \binom{N-i-m+s}{n-r-m+s-1} - \binom{N-i-m+s-1}{n-r-m+s-1}$$

and

$$\mathbf{E}\left(X_{r:n} \,|\, A_m\right) = \binom{N-m}{n-m}^{-1} \left[\sum_{s=1}^{m+1} \sum_{i=k_{s-1}+1}^{k_s-1} \delta''_{m,s,i}(r) \binom{N-i-m+s-1}{n-r-m+s-1} x_i \right.$$
$$\left. + \sum_{s=0}^{m+1} \binom{k_s - s}{r-s} \binom{N-k_s-m+s}{n-r-m+s} x_{k_s} \right],$$

where

$$\delta''_{m,s,i}(r) = \binom{i-s+1}{r-s+1} - \binom{i-s}{r-s+1}.$$

Then, it is easy to verify that, for $r = 0, \ldots, n$, $\mathbf{E}\left(\boldsymbol{\Delta}_{r:n} \,|\, A_m\right)$ is the same as in the lemma's statement. For $x_1 \le \cdots \le x_N$ the result does not change. It follows from the argument similar to Lemma 8, i.e., it suffices to assume (without loss of generality) that coincident values of $\mathcal{X}$ are strictly ordered by their unique names, e.g., by sizes of their indexes.

Proof of Theorem 11

(i) First we write a kernel of orthogonal decomposition of the order statistic $X_{j:n}$, $1 \le j \le n$. For chosen $1 \le k \le N$, using representation (1.20), Lemma 10 for $m = 0, 1$ and applying binomial identities (1.8) and (1.9) we have

$$g_{1j}(x_k) = \binom{N-2}{n-1}^{-1} \sum_{r=0}^{j-1} \left\{ \sum_{i=1}^{k-1} \frac{i}{N} \theta_{21}(i,r) \bigtriangleup_i - \sum_{i=k}^{N-1} \left(1 - \frac{i}{N}\right) \theta_{22}(i,r) \bigtriangleup_i \right\},$$

where

$$\theta_{21}(i,r) = \frac{N}{i} \binom{i}{r} \left\{ \binom{N-i-1}{n-r-1} - \frac{n}{N} \binom{N-i}{n-r} \right\}$$

and

$$\theta_{22}(i,r) = -\frac{N}{N-i} \binom{N-i}{n-r} \left\{ \binom{i-1}{r-1} - \frac{n}{N} \binom{i}{r} \right\}.$$

It is easy to verify that $\theta_{21}(i,r) \equiv \theta_{22}(i,r)$. Next, using induction it is easy to show that for every $1 \le j \le n$

$$\sum_{r=0}^{j-1} \theta_{22}(i,r) = \binom{i-1}{j-1} \binom{N-i-1}{n-j}$$

and the proof of part *(i)* follows from a simple observation that

$$g_1(x_k) = n^{-1} \sum_{j=1}^{n} c_j g_{1j}(x_k), \quad 1 \le k \le N.$$

(ii) Similarly, for chosen $1 \le k < l \le N$, using representation (1.20), Lemma 10 for $m = 0, 1, 2$ and applying the simplest identities of Lemma 7, for a single order statistic we get

$$\begin{aligned} g_{2j}(x_k, x_l) = \binom{N-4}{n-2}^{-1} \sum_{r=0}^{j-1} \Bigg\{ & \sum_{i=1}^{k-1} \frac{i(i-1)}{[N-1]_2} \theta_{31}(i,r) \bigtriangleup_i \\ & - \sum_{i=k}^{l-1} \frac{(i-1)(N-i-1)}{[N-1]_2} \theta_{32}(i,r) \bigtriangleup_i \\ & + \sum_{i=l}^{N-1} \frac{(N-i)(N-i-1)}{[N-1]_2} \theta_{33}(i,r) \bigtriangleup_i \Bigg\}, \end{aligned}$$

where

$$\theta_{31}(i,r) = \frac{[N-1]_2}{i(i-1)} \binom{i}{r} \left\{ \binom{N-i-2}{n-r-2} - 2\frac{n-1}{N-2} \binom{N-i-1}{n-r-1} \right.$$
$$\left. + \frac{n(n-1)}{[N-1]_2} \binom{N-i}{n-r} \right\}$$

and

$$\theta_{32}(i,r) = -\frac{[N-1]_2}{(i-1)(N-i-1)} \left[\binom{i-1}{r-1} \left\{ \binom{N-i-1}{n-r-1} - \frac{n-1}{N-2} \binom{N-i}{n-r} \right\} \right.$$
$$\left. - \frac{n-1}{N-2} \binom{i}{r} \left\{ \binom{N-i-1}{n-r-1} - \frac{n}{N-1} \binom{N-i}{n-r} \right\} \right],$$

and

$$\theta_{33}(i,r) = \frac{[N-1]_2}{(N-i)(N-i-1)} \binom{N-i}{n-r} \left\{ \binom{i-2}{r-2} - 2\frac{n-1}{N-2} \binom{i-1}{r-1} \right.$$
$$\left. + \frac{n(n-1)}{[N-1]_2} \binom{i}{r} \right\}.$$

Similarly $\theta_{31}(i,r) \equiv \theta_{32}(i,r) \equiv \theta_{33}(i,r)$. Next, using induction one can show that, for every $1 \le j \le n$,

$$\sum_{r=0}^{j-1} \theta_{33}(i,r) = \binom{i-2}{j-1} \binom{N-i-2}{n-j-1} - \binom{i-2}{j-2} \binom{N-i-2}{n-j}.$$

To complete the proof of part *(ii)* we observe that

$$g_2(x_k, x_l) = n^{-1} \sum_{j=1}^{n} c_j g_{2j}(x_k, x_l), \quad 1 \le k < l \le N$$

and apply

$$\sum_{j=1}^{n} c_j (b_{j+1} - b_j) = c_n b_{n+1} - c_1 b_1 - \sum_{j=2}^{n} (c_j - c_{j-1}) b_j,$$

where

$$b_j = \binom{i-2}{j-2} \binom{N-i-2}{n-j}.$$

(iii) The proof of this part is very similar.

1.3.2 Remainder terms

Note that Theorem 11 represents the kernels of the Hoeffding decomposition in terms of differences of the weights $c_1, \ldots, c_n$. Here, for $k = 1, 2, 3, 4$, we write the differences $\mathbb{D}_k L_n$, defining quantities $\delta_k(L_n)$ in Theorem 6, in somewhat similar form.

For that purpose we need additional notation. Assume that $x_1 \leq \cdots \leq x_N$. For $k = 1,2,3,4$, denote by $X_{1:n+k} \leq \cdots \leq X_{n+k:n+k}$ order statistics which correspond to the sample $\mathbb{X}_k = \{X_1, \ldots, X_{n+k}\}$. Let $\mathbb{R}_k = \{R_1, \ldots, R_{n+k}\}$ be the ranks of the sample $\mathbb{X}_k$ assigned as follows. We decide that $R_i < R_j$ if $g^{-1}(X_i) < g^{-1}(X_j)$, where $g^{-1}(\cdot)$ is the inverse function of $g(\cdot)$ defined in Lemma 8. Thus ranks $\mathbb{R}_k$ are all distinct, i.e., $\mathbb{R}_k$ is a random permutation of the set $\{1, \ldots, n+k\}$. Further, denote $\mathbb{R}_k^* = \{R_1, \ldots, R_k, R_{n+1}, \ldots, R_{n+k}\}$ and let $\mathcal{R}_k$ be a set of all permutations of the set $\mathbb{R}_k^*$, where a particular permutation means the arrangement of elements of the set by size. Let $R_{1:2k} < \cdots < R_{k:2k} < R_{n+1:2k} < \cdots < R_{n+k:2k}$ denote order statistics which correspond to $\mathbb{R}_k^*$.

Lemma 12 *For each $k = 1,2,3,4$ there exists a random variable $d_k = d_k(\mathbb{R}_k^*)$ with values in $\{-1,0,1\}$ such that*

$$\mathbb{D}_k L_n = d_k n^{-1} \sum_{j=R_{k:2k}}^{R_{n+1:2k}-1} \Delta^{k-1}(c_j) \boldsymbol{\Delta}_{j:n+k}.$$

Proof of Lemma 12

The set $\mathcal{R}_1$ contains only 2 elements, therefore we elaborate the case $k = 1$. Let $R_1 < R_{n+1}$. Then

$$\begin{aligned} n\mathbb{D}_1 L_n &= n \left(L_n(\mathbb{X}_1 \backslash \{X_{n+1}\}) - L_n(\mathbb{X}_1 \backslash \{X_1\}) \right) \\ &= \left[\sum_{j=1}^{R_{n+1}-1} c_j X_{j:n+1} + \sum_{j=R_{n+1}}^{n} c_j X_{j+1:n+1} \right] \\ &\quad - \left[\sum_{j=1}^{R_1-1} c_j X_{j:n+1} + \sum_{j=R_1}^{n} c_j X_{j+1:n+1} \right] \\ &= - \sum_{j=R_1}^{R_{n+1}-1} c_j \boldsymbol{\Delta}_{j:n+1}. \end{aligned}$$

Let $R_1 > R_{n+1}$. Then

$$\begin{aligned} n\mathbb{D}_1 L_n &= n \left(L_n(\mathbb{X}_1 \backslash \{X_{n+1}\}) - L_n(\mathbb{X}_1 \backslash \{X_1\}) \right) \\ &= \left[\sum_{j=1}^{R_{n+1}-1} c_j X_{j:n+1} + \sum_{j=R_{n+1}}^{n} c_j X_{j+1:n+1} \right] \\ &\quad - \left[\sum_{j=1}^{R_1-1} c_j X_{j:n+1} + \sum_{j=R_1}^{n} c_j X_{j+1:n+1} \right] \\ &= \sum_{j=R_{n+1}}^{R_1-1} c_j \boldsymbol{\Delta}_{j:n+1}. \end{aligned}$$

Next we calculate $\mathbb{D}_k L_n$, $k = 2, 3, 4$ recursively. Write $\mathbb{D}_2 L_n = \mathbb{D}_1' L_n - \mathbb{D}_1'' L_n$, where

$$\begin{aligned}
\mathbb{D}_1' L_n &= L_n(\mathbb{X}_2 \backslash \{X_{n+1}, X_{n+2}\}) - L_n(\mathbb{X}_2 \backslash \{X_1, X_{n+2}\}), \\
\mathbb{D}_1'' L_n &= L_n(\mathbb{X}_2 \backslash \{X_2, X_{n+1}\}) - L_n(\mathbb{X}_2 \backslash \{X_1, X_2\}).
\end{aligned}$$

Note that both components of $\mathbb{D}_1' L_n$ are dependent on X_2 and are independent of X_{n+2}. For $\mathbb{D}_1'' L_n$ it is conversely. Now by constructing the set $\mathcal{R}_2$ from the set $\mathcal{R}_1$ we find the following cases. For $R_1 < R_{n+1}$,

$$n\mathbb{D}_1' L_n = \begin{cases} -\sum\limits_{j=R_1}^{R_{n+1}-1} c_{j-1} \boldsymbol{\Delta}_{j:n+2} & \text{for } R_{n+2} < R_1, \\ -\sum\limits_{j=R_1}^{R_{n+2}-1} c_j \boldsymbol{\Delta}_{j:n+2} - \sum\limits_{j=R_{n+2}}^{R_{n+1}-1} c_{j-1} \boldsymbol{\Delta}_{j:n+2} & \text{for } R_1 < R_{n+2} < R_{n+1}, \\ -\sum\limits_{j=R_1}^{R_{n+1}-1} c_j \boldsymbol{\Delta}_{j:n+2} & \text{for } R_{n+1} < R_{n+2}, \end{cases}$$

$$n\mathbb{D}_1'' L_n = \begin{cases} -\sum\limits_{j=R_1}^{R_{n+1}-1} c_{j-1} \boldsymbol{\Delta}_{j:n+2} & \text{for } R_2 < R_1, \\ -\sum\limits_{j=R_1}^{R_2-1} c_j \boldsymbol{\Delta}_{j:n+2} - \sum\limits_{j=R_2}^{R_{n+1}-1} c_{j-1} \boldsymbol{\Delta}_{j:n+2} & \text{for } R_1 < R_2 < R_{n+1}, \\ -\sum\limits_{j=R_1}^{R_{n+1}-1} c_j \boldsymbol{\Delta}_{j:n+2} & \text{for } R_{n+1} < R_2. \end{cases}$$

For $R_1 > R_{n+1}$,

$$n\mathbb{D}_1' L_n = \begin{cases} \sum\limits_{j=R_{n+1}}^{R_1-1} c_{j-1} \boldsymbol{\Delta}_{j:n+2} & \text{for } R_{n+2} < R_{n+1}, \\ \sum\limits_{j=R_{n+1}}^{R_{n+2}-1} c_j \boldsymbol{\Delta}_{j:n+2} + \sum\limits_{j=R_{n+2}}^{R_1-1} c_{j-1} \boldsymbol{\Delta}_{j:n+2} & \text{for } R_{n+1} < R_{n+2} < R_1, \\ \sum\limits_{j=R_{n+1}}^{R_1-1} c_j \boldsymbol{\Delta}_{j:n+2} & \text{for } R_1 < R_{n+2}, \end{cases}$$

$$n\mathbb{D}_1'' L_n = \begin{cases} \sum\limits_{j=R_{n+1}}^{R_1-1} c_{j-1} \boldsymbol{\Delta}_{j:n+2} & \text{for } R_2 < R_{n+1}, \\ \sum\limits_{j=R_{n+1}}^{R_2-1} c_j \boldsymbol{\Delta}_{j:n+2} + \sum\limits_{j=R_2}^{R_1-1} c_{j-1} \boldsymbol{\Delta}_{j:n+2} & \text{for } R_{n+1} < R_2 < R_1, \\ \sum\limits_{j=R_{n+1}}^{R_1-1} c_j \boldsymbol{\Delta}_{j:n+2} & \text{for } R_1 < R_2. \end{cases}$$

Note that, it suffices to consider only those elements of $\mathcal{R}_2$ for which we have $R_1 > R_2$.

Similarly, write $\mathbb{D}_3 L_n = \mathbb{D}_2' L_n - \mathbb{D}_2'' L_n$, where

$$
\begin{aligned}
\mathbb{D}_2' L_n &= L_n(\mathbb{X}_3 \backslash \{X_{n+1}, X_{n+2}, X_{n+3}\}) - L_n(\mathbb{X}_3 \backslash \{X_2, X_{n+1}, X_{n+3}\}) \\
&\qquad - L_n(\mathbb{X}_3 \backslash \{X_1, X_{n+2}, X_{n+3}\}) + L_n(\mathbb{X}_3 \backslash \{X_1, X_2, X_{n+3}\}), \\
\mathbb{D}_2'' L_n &= L_n(\mathbb{X}_3 \backslash \{X_3, X_{n+1}, X_{n+2}\}) - L_n(\mathbb{X}_3 \backslash \{X_2, X_3, X_{n+1}\}) \\
&\qquad - L_n(\mathbb{X}_3 \backslash \{X_1, X_3, X_{n+2}\}) + L_n(\mathbb{X}_3 \backslash \{X_1, X_2, X_3\}).
\end{aligned}
$$

Note that all the components of $\mathbb{D}_2' L_n$ are dependent on X_3 and are independent of X_{n+3}. For $\mathbb{D}_2'' L_n$ it is conversely. Now, by constructing the set $\mathcal{R}_3$ from the set $\mathcal{R}_2$, similarly as in the case $k = 2$, we can recursively find $\mathbb{D}_3 L_n$. Without loss of generality, we consider only those elements of $\mathcal{R}_3$ for which $R_1 > R_2 > R_3$. The calculations are routine, long and cumbersome; therefore we omit them.

The calculation of $\mathbb{D}_4 L_n$ is similar. Assuming that $R_1 > R_2 > R_3 > R_4$, it suffices to consider (only) $8!/4!$ elements of the set $\mathcal{R}_4$.

Chapter 2

Asymptotic approximations to distributions

2.1 Maximum variance of sample extremes

We assume further, without loss of generality, that $x_1 \leq \cdots \leq x_N$. To avoid trivialities, we assume in addition that $\mathbf{Var}\, X_1 > 0$ or, equivalently, $x_1 < x_N$.

Theorem 13 *For $m = 1$ and $m = n$ we have*

$$\mathbf{Var}\, X_{m:n} \leq n \frac{N-n}{N-1} \mathbf{Var}\, X_1. \tag{2.1}$$

Remark 14 *The bound* (2.1) *is optimal for $m = 1$ and $m = n$, i.e., there exist nontrivial populations where the equality is attained in* (2.1).

Variance bounds on order statistics were considered by Moriguti [51], Papadatos [54, 55], Rychlik [62] and Jasiński and Rychlik [41]. In particular, for sample extremes Papadatos [54] showed the bound $\mathbf{Var}\, X_{m:n} \leq n \,\mathbf{Var}\, X_1$, in the case of i.i.d. observations. Rychlik [62] extended this bound to arbitrarily dependent identically distributed observations. Theorem 13 improves these results by the finite population correction factor $(N-n)/(N-1) < 1$, in the case of samples drawn without replacement. Also, note that for i.i.d. samples, additionally assuming that X_1 has a symmetric distribution, for sample extremes Moriguti [51] obtained the bound $\mathbf{Var}\, X_{m:n} \leq (n/2)\, \mathbf{Var}\, X_1$.

With the references to [54, 55, 62] and [41] on the same sampling model as in [51], the best possible bounds are also obtained for the variances of order statistics $X_{m:n}$, $2 \leq m \leq n-1$. For samples drawn without replacement, the methods applied allow us to obtain bounds (2.8) on these order statistics, however, in general, these evaluations are not optimal. We give a disscusion on these bounds below the proof of Theorem 13.

Related problems are optimal bounds on the covariances and correlations of order statistics. For the results in the case of i.i.d. observations we refer to Papathanasiou [57], Chunsheng [24], Papadatos [56] and Terrell [72], Székely and Móri [70], respectively. There are a few analogous results for samples drawn without replacement. On optimal bounds on the covariances and correlations of order statistics were presented in Balakrishnan et al. [6] and López-Blázquez and Castaño-Martínez [46], respectively. In the case of samples drawn without replacement, the difficulties arising in similar problems were well discussed in Berred and Nevzorov [10]; see also Berred and Nevzorov [9]. Some properties of order statistics from finite populations and their connection to variance bounds were also discussed in Takahasi and Futatsuya [71] and Afendras et al. [1].

Proof of Theorem 13. We consider all $1 \leq m \leq n$. Clearly, $t^m(\mathbb{X}) = X_{m:n}$ is the symmetric statistic. The basic idea of the proof is an estimation of the error of approximation of $\mathbf{Var}\, t^m(\mathbb{X})$ by exactly zero terms of the Hoeffding decomposition. In particular, a slight and simple modification of Lemma 2 of Bloznelis and Götze [20], where for our purposes we take more strict inequalities (up to the constants), and a trivial extension of Theorem 6 (to the case $k = 0$) lead to the inequality

$$\mathbf{Var}\, X_{m:n} \leq \frac{1}{2} n \left(1 - \frac{n}{N}\right) \mathbf{E} \left(\mathbb{D}_1 X_{m:n}\right)^2, \tag{2.2}$$

where

$$\mathbb{D}_1 X_{m:n} = t^m(\mathbb{X}_1 \backslash \{X_{n+1}\}) - t^m(\mathbb{X}_1 \backslash \{X_1\}).$$

Now we evaluate $\mathbf{E} \left(\mathbb{D}_1 X_{m:n}\right)^2$. Introduce the set

$$\mathcal{I} = \{(i, j) : 1 \leq i < j \leq n + 1\}$$

and its subsets

$$\begin{aligned} \mathcal{I}_0^m &= \{(i, j) : 1 \leq i < j \leq m \text{ or } m < i < j \leq n + 1\}, \\ \mathcal{I}_1^m &= \{(i, j) : 1 \leq i \leq m \text{ and } m < j \leq n + 1\}. \end{aligned}$$

Clearly, $\mathcal{I} = \mathcal{I}_0^m \cup \mathcal{I}_1^m$ and $\mathcal{I}_0^m \cap \mathcal{I}_1^m = \emptyset$. By the proof of Lemma 12,

$$\mathbb{D}_1 X_{m:n} = \begin{cases} 0 & \text{if } (R_{1:2}, R_{n+1:2}) \in \mathcal{I}_0^m, \\ \pm \mathbf{\Delta}_{m:n+1} & \text{if } (R_{1:2}, R_{n+1:2}) \in \mathcal{I}_1^m. \end{cases}$$

The events $\mathfrak{R}_{1;ij} = \{R_{1:2} = i, R_{n+1:2} = j\}$ and $\{X_{m:n+1} = x_k, X_{m+1:n+1} = x_l\}$ for $1 \le k < l \le N$ are independent. Also, we find that for $1 \le i < j \le n+1$

$$p_{1;ij} := \mathbf{P}\{\mathfrak{R}_{1;ij}\} = \binom{n+1}{2}^{-1}. \tag{2.3}$$

Thus, the application of (1.18) of Lemma 9 to the sample of size $n+1$ and the use of simple identity (1.8) give

$$\begin{aligned} \mathbf{E}\left(\mathbb{D}_1 X_{m:n}\right)^2 &= \sum_{(i,j)\in\mathcal{I}_1^m} \mathbf{E}\left(\boldsymbol{\Delta}^2_{m:n+1} \,\middle|\, \mathfrak{R}_{1;ij}\right) p_{1;ij} = 2\frac{m(n+1-m)}{(n+1)n}\,\mathbf{E}\,\boldsymbol{\Delta}^2_{m:n+1} \\ &= 2\frac{N}{N-1}\Bigg[\sum_{i=1}^{N-1}\frac{i}{N}\left(1-\frac{i}{N}\right)p_i(m)\ \triangle_i^2 \\ &\qquad + 2\sum_{1\le i<j\le N-1}\frac{i}{N}\left(1-\frac{j}{N}\right)p_{ij}(m)\ \triangle_i\triangle_j\Bigg], \end{aligned} \tag{2.4}$$

where we denote

$$p_i(m) = \binom{i-1}{m-1}\binom{N-i-1}{n-m}\bigg/\binom{N-2}{n-1}, \qquad 1 \le i \le N-1,$$

and

$$p_{ij}(m) = \binom{i-1}{m-1}\binom{N-j-1}{n-m}\bigg/\binom{N-2}{n-1}, \qquad 1 \le i < j \le N-1.$$

It is easy to verify that, applying

$$(x_l - x_k)^2 = \sum_{i=k}^{l-1}\triangle_i^2 + 2\sum_{k\le i<j\le l-1}\triangle_i\triangle_j$$

and collecting the terms with $\triangle_i^2$ and $\triangle_i\triangle_j$ we get

$$\begin{aligned} \mathbf{Var}\,X_1 &= \frac{1}{N^2}\sum_{1\le k<l\le N}(x_l-x_k)^2 \\ &= \sum_{i=1}^{N-1}\frac{i}{N}\left(1-\frac{i}{N}\right)\triangle_i^2 + 2\sum_{1\le i<j\le N-1}\frac{i}{N}\left(1-\frac{j}{N}\right)\triangle_i\triangle_j\ . \end{aligned} \tag{2.5}$$

In addition, the inequality

$$\max_{1\le i<j\le N-1} p_{ij}(m) \le \max_{1\le i\le N-1} p_i(m) \quad \text{for all} \quad 1 \le m \le n,$$

holds. Thus, it follows from (2.4) that

$$\mathbf{E}\,(\mathbb{D}_1 X_{m:n})^2 \le 2\frac{N}{N-1}\max_{1\le i\le N-1} p_i(m)\,\mathbf{Var}\,X_1. \tag{2.6}$$

For $m=1$ and $m=n$, this inequality together with (2.2) yields (2.1).

We show that bound (2.1) is optimal. In order to obtain the equality in (2.6) for any $1\le m\le n$, we need to take a population with the values

$$x_1=\cdots=x_{i_0}<x_{i_0+1}=\cdots=x_N, \tag{2.7}$$

where

$$i_0=i_0(m)=\arg\max_{1\le i\le N-1} p_i(m).$$

For this population Lemma 8 implies that $X_{m:n}\overset{d}{=}g(Z_{m:n})$ has the distribution with two values $x_{i_0(m)}$ and $x_{i_0(m)+1}$, with

$$r_m=\mathbf{P}\left\{X_{m:n}=x_{i_0(m)}\right\}=\mathbf{P}\left\{Z_{m:n}\le i_0(m)\right\}=\binom{N}{n}^{-1}\sum_{k=1}^{i_0(m)}\binom{k-1}{m-1}\binom{N-k}{n-m}.$$

Therefore $\mathbf{Var}\,X_{m:n}=r_m(1-r_m)\,\triangle^2_{i_0(m)}$. Then, it is easy to verify that, as $m=1$ and $m=n$, for which $i_0(1)=1$ and $i_0(n)=N-1$, the choice of (2.7) also gives the equality in (2.1). The theorem is proven. $\square$

The proof of Theorem 13 also sets the bounds on variances of the order statistics $X_{m:n}$, $2\le m\le n-1$. We obtain from (2.6) and (2.2) that

$$\mathbf{Var}\,X_{m:n}\le n\frac{N-n}{N-1}\max_{1\le i\le N-1}\mathcal{H}_{N-2,n-1,i-1}(m-1)\,\mathbf{Var}\,X_1. \tag{2.8}$$

In general, for $2\le m\le n-1$, these bounds are not optimal in the sense that inequality (2.8) is always strict for $n_*=\min\{n,N-n\}\ge 3$. For populations other than (2.7), this easily follows from (2.4) and (2.6). For a population of the form (2.7), the strict inequality in (2.8) follows from the strict inequality in (2.2). Indeed, one can show that, for a population of type (2.7), we have, by using, e.g., Theorem 11,

$$\max\Bigg\{\sum_{1\le i<j\le N} g_2^2(x_i,x_j),\sum_{1\le i<j<k\le N} g_3^2(x_i,x_j,x_k)\Bigg\}>0,$$

which is equivalent to $\max\{\mathbf{Var}\,U_2(\mathbb{D}_1X_{m:n}),\mathbf{Var}\,U_3(\mathbb{D}_1X_{m:n})\}>0$, see (A.21) in [20]. Here $U_j(\mathbb{D}_1X_{m:n})$, $j=2,3$ are the components of Hoeffding decomposition (1.1) for the statistic $\mathbb{D}_1X_{m:n}$. The latter inequality implies that inequality (2.2) is strict.

Let us mention that for simple random samples our bounds (2.8) outperform the corresponding bounds of Rychlik [62] in cases where the finite population correction factor $(N-n)/(N-1)$ is sufficiently small. Clearly, in the case of sample extremes, if we fix the sample size n and let the population size $N \to \infty$, then the sampling without replacement approximates the case of i.i.d. observations and bound (2.1) becomes the same as in Papadatos [54].

2.2 Asymptotic normality

Consider the normalized L-statistic S_n defined by (4). Clearly, for the statistic S_n, the results on the Hoeffding decomposition in Theorem 11 and Lemma 12 must be multiplied by $n^{1/2}$ only. Recall the notions $\tilde{\sigma}_n^2 = \mathbf{Var}\, S_n$ and $\sigma_1^2 = \mathbf{E}\, g_1^2(X_1)$, and the numbers τ^2 and n_* defined in (6) and (7). Let the weights $c_1, \ldots, c_n$ be determined by the weight function $J\colon (0,1) \to \mathbb{R}$ as in (2).

First, we give a version of Proposition 4 where the general smoothness conditions (for symmetric statistics) are replaced by that imposed on the weight function $J(\cdot)$ and the moments of X_1. Recall a Lindeberg-type Erdős–Rényi condition familiar from the case of sample mean (see Theorem 1): for every $\varepsilon > 0$,

$$\mathbf{E}\, g_1^2(X_1)\sigma_1^{-2}\mathbb{I}\{|g_1(X_1)| > \varepsilon\tau\sigma_1\} = o(1) \quad \text{as} \quad n_* \to \infty. \tag{2.9}$$

We say that the function $J(\cdot)$ satisfies the Hölder condition of order δ on $(0,1)$ if there are nonnegative real constants B, δ, such that $|J(u) - J(v)| \leq B\,|u-v|^\delta$ for all $u, v \in (0,1)$.

Theorem 15 *Assume that $n_* \to \infty$ and $\tilde{\sigma}_n$ remains bounded away from zero for all n_*. Suppose that $\mathbf{E}\, X_1^2 < \infty$ and that $J(\cdot)$ is bounded and satisfies the Hölder condition of order $\delta > 1/2$ on $(0,1)$. Let (2.9) hold. Then $\tilde{\sigma}_n^{-1} S_n$ is asymptotically standard normal.*

Second, Theorem 16 below gives sufficient conditions under which the trimmed means, defined in (3), are asymptotically standard normal. Note that in this case, the weight function $J(u)$ (recall Example 1 presented in Introduction) is not sufficiently smooth, i.e., $J(u)$ is bounded, but it does not satisfy the Hölder condition. Introduce an additional smoothness condition for the population $\mathcal{X}$. Assume that, without loss of generality, $x_1 \leq \cdots \leq x_N$. Suppose that, for some constants $C > 0$ and $1/2 < \delta \leq 1$,

$$|x_m - x_l| \leq C N^{-\delta}\,|m - l| \tag{2.10}$$

is satisfied for all $1 \leq l < m \leq N$.

Theorem 16 *Assume that $n_* \to \infty$ and $\tilde{\sigma}_n$ remains bounded away from zero for all n_*. Say that $\mathbf{E}\, X_1^2 < \infty$. Assume that* (2.10) *is satisfied for some $1/2 < \delta \le 1$, and $(1 - n/N)^{-2}\, nN^{2(\delta-1)} \to \infty$. Let* (2.9) *hold. Then, in the case of a trimmed mean, $\tilde{\sigma}_n^{-1} S_n$ is asymptotically standard normal.*

In the case of i.i.d. observations, it was shown by Stigler [68] that in order for the trimmed mean to be asymptotically normal, it is necessary and sufficient that the sample is trimmed at sample quantiles for which the corresponding population quantiles are uniquely defined. Thus, the finite population smoothness conditions of Theorem 16 seem too strong. On the other hand, for samples drawn without replacement, condition (2.10) has a specific interpretation. Let us take $l = 1$ and $m = N$. If the population $\mathcal{X}$ is bounded, then the condition is satisfied for $\delta = 1$. In the marginal case of $\delta = 1/2$, (2.10) is satisfied for any finite population by Nair–Thomson inequality $x_N - x_1 \le \sigma\sqrt{2N}$, where $\sigma^2 = \mathbf{Var}\, X_1$ (see, e.g., [6]). Thus, condition (2.10) seems very mild for small $\epsilon > 0$ in $\delta = 1/2 + \epsilon$, i.e., it holds for most of possible populations. Obviously, if we are interested in the asymptotic normality of the trimmed means, then, by the conditions of Theorem 16, for small ϵ we should have $n \to \infty$ quite quickly as $N \to \infty$, while in the case $\delta = 1$ it suffices that $n \to \infty$ arbitrarily slowly with respect to the grow of the population size N.

In the proofs of Theorems 15 and 16 below, we assume that, without loss of generality, $x_1 \le \cdots \le x_N$.

Proof of Theorem 15

First we show that $\tilde{\sigma}_n$ is bounded as $n_* \to \infty$. Then (2.9) is equivalent to (16). Arguing as in the proof of Theorem 13, we have

$$\tilde{\sigma}_n^2 \le \frac{1}{2} n \left(1 - \frac{n}{N}\right) \mathbf{E}\,(\mathbb{D}_1 S_n)^2 . \tag{2.11}$$

Since $J(\cdot)$ is bounded, there exists an absolute constant a that

$$\max_{1 \le p \le n} |c_p| \le a \tag{2.12}$$

for all n. Introduce the events $\mathfrak{R}_{1;ij} = \{R_{1:2} = i, R_{n+1:2} = j\}$, $1 \le i < j \le n+1$, as in the proof of Theorem 13, and recall their probabilities $p_{1;ij}$, given in (2.3). By Lemma 12 and (2.12) we obtain

$$\begin{aligned} \mathbf{E}\,(\mathbb{D}_1 S_n)^2 &= \sum_{1 \le i < j \le n+1} \mathbf{E}\left[(\mathbb{D}_1 S_n)^2 \,\middle|\, \mathfrak{R}_{1;ij} \right] p_{1;ij} \\ &\le a^2 n^{-1} \sum_{1 \le i < j \le n+1} \mathbf{E}\left[(X_{j:n+1} - X_{i:n+1})^2 \,\middle|\, \mathfrak{R}_{1;ij} \right] p_{1;ij}. \end{aligned} \tag{2.13}$$

Since the events $\mathfrak{R}_{1;ij}$ and $\mathfrak{B}_{1;ijlm} = \{X_{i:n+1} = x_l, X_{j:n+1} = x_m\}$, $1 \le l < m \le N$ are independent, for $x_1 < \cdots < x_N$ we get

$$p_{1;ijlm} := \mathbf{P}\left\{\mathfrak{B}_{1;ijlm} \,|\, \mathfrak{R}_{1;ij}\right\} = \binom{l-1}{i-1}\binom{m-l-1}{j-i-1}\binom{N-m}{n+1-j} \Big/ \binom{N}{n+1}.$$

For $x_1 \le \cdots \le x_N$ these probabilities are the same. This fact follows from Lemma 8. We also have that, by the generalized Vandermonde identity, see the case $T = 3$ in (1.13),

$$\begin{aligned}
\sum_{1 \le i < j \le n+1} p_{1;ijlm} &= \binom{N}{n+1}^{-1} \sum_{s=0}^{n-1} \sum_{t=0}^{n-1-s} \binom{l-1}{s}\binom{m-l-1}{t}\binom{N-m}{n-1-s-t} \\
&= \binom{N}{n+1}^{-1}\binom{N-2}{n-1}.
\end{aligned}$$

Then we recall the first expression of $\mathbf{Var}\, X_1$ in (2.5) and continue (2.13),

$$\begin{aligned}
\mathbf{E}\left(\mathbb{D}_1 S_n\right)^2 &\le a^2 n^{-1} \sum_{1 \le i < j \le n+1} \left\{ \sum_{1 \le l < m \le N} (x_m - x_l)^2 p_{1;ijlm} \right\} p_{1;ij} \\
&= a^2 n^{-1} \binom{n+1}{2}^{-1} \binom{N}{n+1}^{-1} \binom{N-2}{n-1} \sum_{1 \le l < m \le N} (x_m - x_l)^2 \\
&= 2a^2 n^{-1} \frac{N}{N-1} \mathbf{Var}\, X_1 \le 4a^2 n^{-1} \mathbf{Var}\, X_1.
\end{aligned}$$

Finally, from (2.11) we get

$$\tilde{\sigma}_n^2 \le 2a^2 \left(1 - \frac{n}{N}\right) \mathbf{Var}\, X_1 = O(1) \quad \text{as} \quad n_* \to \infty.$$

Next we show that, under the conditions of the theorem, $\delta_2(S_n) = o(1)$ is satisfied. Then the theorem will follow from Proposition 4. Since $J(\cdot)$ satisfies the Hölder condition of order $\delta > 1/2$ on $(0,1)$, we find that

$$|c_p - c_{p-1}| = \left| J\left(\frac{p}{n+1}\right) - J\left(\frac{p-1}{n+1}\right) \right| \le B(n+1)^{-\delta}$$

or

$$\max_{2 \le p \le n} |c_p - c_{p-1}| \le B(n+1)^{-\delta}, \quad \text{for some} \quad \delta > 1/2. \tag{2.14}$$

Similarly, introduce the events $\mathfrak{R}_{2;ij} = \{R_{2:4} = i, R_{n+1:4} = j\}$, $1 \le i < j \le n+2$. Now

$$p_{2;ij} := \mathbf{P}\left\{\mathfrak{R}_{2;ij}\right\} = \binom{i-1}{1}\binom{n+2-j}{1} \Big/ \binom{n+2}{4}. \tag{2.15}$$

We also have

$$p_{2;ijlm} := \mathbf{P}\left\{\mathfrak{B}_{2;ijlm} \,|\, \mathfrak{R}_{2;ij}\right\} = \binom{l-1}{i-1}\binom{m-l-1}{j-i-1}\binom{N-m}{n+2-j}\Big/\binom{N}{n+2},$$

where the events $\mathfrak{R}_{2;ij}$ and $\mathfrak{B}_{2;ijlm} = \{X_{i:n+2} = x_l, X_{j:n+2} = x_m\}$, $1 \le l < m \le N$ are independent. By Lemma 12 and (2.14), we obtain

$$\begin{aligned}
\delta_2(S_n) &= \mathbf{E}\left(n_*\mathbb{D}_2 S_n\right)^2 = n_*^2 \sum_{1\le i<j\le n+2} \mathbf{E}\left[(\mathbb{D}_2 S_n)^2 \,\Big|\, \mathfrak{R}_{2;ij}\right] p_{2;ij} \\
&\le B^2 \frac{n_*^2 n^{-1}}{(n+1)^{2\delta}} \sum_{1\le i<j\le n+2} \mathbf{E}\left[(X_{j:n+2} - X_{i:n+2})^2 \,\Big|\, \mathfrak{R}_{2;ij}\right] p_{2;ij} \qquad (2.16)\\
&= B^2 \frac{n_*^2 n^{-1}}{(n+1)^{2\delta}} \sum_{1\le l<m\le N} \lambda_{2;lm}(x_m - x_l)^2,
\end{aligned}$$

where

$$\lambda_{2;lm} = \sum_{1\le i<j\le n+2} p_{2;ij} p_{2;ijlm}.$$

Taking $j = i+1$ and applying $\max_{0\le u\le 1} u(1-u) \le 1/4$, for all $1 \le i < j \le n+2$, we get the inequalities

$$p_{2;ij} \le n^2 \binom{n+2}{4}^{-1} \frac{i-1}{n}\left(1 - \frac{i-1}{n}\right) \le \frac{1}{4} n^2 \binom{n+2}{4}^{-1}.$$

Then, noting that, by identity (1.13) as $T = 3$,

$$\sum_{1\le i<j\le n+2} p_{2;ijlm} = \binom{N}{n+2}^{-1}\binom{N-2}{n},$$

we obtain, for all $1 \le l < m \le N$,

$$\lambda_{2;lm} \le \frac{1}{4} n^2 \binom{n+2}{4}^{-1}\binom{N}{n+2}^{-1}\binom{N-2}{n} \le 24N^{-2}.$$

Finally, it follows from this bound and (2.16) that

$$\delta_2(S_n) \le 24B^2 \frac{n_*^2 n^{-1}}{(n+1)^{2\delta}} \,\mathbf{Var}\, X_1 = o(1) \quad \text{as} \quad n_* \to \infty.$$

It means that the theorem is proven.

Proof of Theorem 16

By the first part of the proof of Theorem 15, condition (2.9) is equivalent to (16). Thus we need to verify the condition $\delta_2(S_n) = o(1)$ of Proposition 4 only.

Write, for short, $s = [t_1 n] + 1$ and $t = [t_2 n]$. Similarly, as in the proof of Theorem 15, by Lemma 12 we obtain

$$\delta_2(S_n) \leq \frac{n_*^2 n}{(t-s+1)^2} \sum_{1 \leq i < j \leq n+2} p_{2;ij} \, \mathbf{E} \, A_{ij}^2(s,t), \tag{2.17}$$

where $p_{2;ij}$ is defined in (2.15) and

$$A_{ij}(s,t) = \sum_{p=i}^{j-1} (\tilde{c}_p - \tilde{c}_{p-1}) \boldsymbol{\Delta}_{p:n+2}, \quad \text{with} \quad \tilde{c}_p = \mathbb{I}\{s \leq p \leq t\}.$$

Assuming, without loss of generality, that $n > (t_2 - t_1)^{-1}$, we have $s < t$. It also follows, from the same assumption, that the inequality $[t_2 n] - [t_1 n] \geq t_2 n - 1 - t_1 n$ implies that, for some constant $C_1 > 0$,

$$\frac{n^2}{(t-s+1)^2} \leq \left(t_2 - t_1 - \frac{1}{n}\right)^{-2} \leq C_1. \tag{2.18}$$

Let us decompose $\mathcal{I} = \{(i,j) : 2 \leq i < j \leq n+1\}$, for fixed $s < t$, into mutually disjoint subsets

$$\begin{aligned}
\mathcal{I}_1 &= \{(i,j) : t+2 \leq i < j \leq n+1\}, \\
\mathcal{I}_2 &= \{(i,j) : 2 \leq i < j \leq s\}, \\
\mathcal{I}_3 &= \{(i,j) : s+1 \leq i < j \leq t+1\}, \\
\mathcal{I}_4 &= \{(i,j) : s+1 \leq i \leq t+1, \, t+2 \leq j \leq n+1\}, \\
\mathcal{I}_5 &= \{(i,j) : 2 \leq i \leq s, \, s+1 \leq j \leq t+1\}, \\
\mathcal{I}_6 &= \{(i,j) : 2 \leq i \leq s, \, t+2 \leq j \leq n+1\},
\end{aligned}$$

such that $\mathcal{I} = \mathcal{I}_1 \cup \cdots \cup \mathcal{I}_6$. Then we get

$$A_{ij}(s,t) = \begin{cases} 0 & \text{if } (i,j) \in \mathcal{I}_1 \cup \mathcal{I}_2 \cup \mathcal{I}_3, \\ -\tilde{c}_t \boldsymbol{\Delta}_{t+1:n+2} & \text{if } (i,j) \in \mathcal{I}_4, \\ \tilde{c}_s \boldsymbol{\Delta}_{s:n+2} & \text{if } (i,j) \in \mathcal{I}_5, \\ \tilde{c}_s \boldsymbol{\Delta}_{s:n+2} - \tilde{c}_t \boldsymbol{\Delta}_{t+1:n+2} & \text{if } (i,j) \in \mathcal{I}_6. \end{cases}$$

Now, by collecting the terms of the sum $\sum_{i<j}$ with the same value of $\mathbf{E} \, A_{ij}^2(s,t)$ in (2.17), applying $\mathbf{E}(\boldsymbol{\Delta}_{t+1:n+2} - \boldsymbol{\Delta}_{s:n+2})^2 \leq \mathbf{E} \, \boldsymbol{\Delta}_{t+1:n+2}^2 + \mathbf{E} \, \boldsymbol{\Delta}_{s:n+2}^2$, and then collecting terms with $\mathbf{E} \, \boldsymbol{\Delta}_{t+1:n+2}^2$ and $\mathbf{E} \, \boldsymbol{\Delta}_{s:n+2}^2$, and also invoking inequality (2.18),

we obtain

$$\begin{aligned}\delta_2(S_n) \le C_1 n_*^2 n^{-1} \binom{n+2}{4}^{-1} \Bigg[& \binom{t+1}{2}\binom{n-t+1}{2} \mathbf{E}\,\mathbf{\Delta}^2_{t+1:n+2} \\ & + \binom{s}{2}\left\{(n-t+1)^2 - \binom{n-s+1}{2}\right\} \mathbf{E}\,\mathbf{\Delta}^2_{s:n+2}\Bigg].\end{aligned} \tag{2.19}$$

By applying the simple inequality $\binom{u}{v} \le u^v/v!$, we derive

$$\binom{t+1}{2}\binom{n-t+1}{2} \le \frac{(n+2)^4}{4}\left[\frac{t+1}{n+2}\left(1-\frac{t+1}{n+2}\right)\right]^2 \le \frac{(n+2)^4}{64}. \tag{2.20}$$

Taking $s = t$, very similarly we get

$$\binom{s}{2}(n-t+1)^2 \le \frac{(n+1)^4}{2}\left[\frac{t}{n+1}\left(1-\frac{t}{n+1}\right)\right]^2 \le \frac{(n+1)^4}{32}. \tag{2.21}$$

Next, by Lemma 8, for $1 \le p \le n+1$,

$$\mathbf{E}\,\mathbf{\Delta}^2_{p:n+2} = \binom{N}{n+2}^{-1} \sum_{1\le l<m\le N} \binom{l-1}{p-1}\binom{m-l-1}{0}\binom{N-m}{n+1-p}(x_m - x_l)^2.$$

Then, by (2.10),

$$\begin{aligned}\mathbf{E}\,\mathbf{\Delta}^2_{p:n+2} &\le \frac{C^2}{N^{2\delta}}\binom{N}{n+2}^{-1} \sum_{1\le l<m\le N} (m-l)^2 \binom{l-1}{p-1}\binom{N-m}{n+1-p} \\ &= \frac{C^2}{N^{2\delta}} \frac{(N+1)(2N-n)}{(n+3)(n+4)}.\end{aligned} \tag{2.22}$$

Here the last equality is easily obtained by using simple binomial identities (1.10), (1.8) and (1.11). Indeed, for instance,

$$\begin{aligned}\sum_{1\le l<m\le N} m^2 \binom{l-1}{p-1}\binom{N-m}{n+1-p} &= \sum_{m=2}^{N}\left\{\sum_{l=1}^{m-1}\binom{l-1}{p-1}\right\} m^2 \binom{N-m}{n+1-p} \\ &= \sum_{m=2}^{N} m^2 \binom{m-1}{p}\binom{N-m}{n+1-p} = (p+1)\sum_{m=2}^{N} m \binom{m}{p+1}\binom{N-m}{n+1-p} \\ &= (p+1)(p+2)\sum_{m=2}^{N}\binom{m+1}{p+2}\binom{N-m}{n+1-p} \\ &\qquad - (p+1)\sum_{m=2}^{N}\binom{m}{p+1}\binom{N-m}{n+1-p} \\ &= (p+1)(p+2)\binom{N+2}{n+4} - (p+1)\binom{N+1}{n+3},\end{aligned}$$

and so on. Finally, applying (2.20), (2.21) and (2.22), and $n_* \le 2n(1-n/N)$, we continue (2.19),

$$\begin{aligned}\delta_2(S_n) &\le C_1 n_*^2 n^{-1} \binom{n+2}{4}^{-1} \frac{C^2}{N^{2\delta}} \frac{(N+1)(2N-n)}{(n+3)(n+4)} \left[\frac{(n+2)^4}{64} + \frac{(n+1)^4}{32}\right] \\ &\le C_2 \left(1-\frac{n}{N}\right)^2 \frac{N^{2(1-\delta)}}{n},\end{aligned}$$

for some constant $C_2 > 0$. The theorem is proven.

2.3 Edgeworth expansion

Consider the L-statistic S_n defined in (4). We assume that the weight function $J\colon (0,1) \to \mathbb{R}$ generates the weights $c_1, \ldots, c_n$ as in (2). Theorem 5 implies the following result on the validity of (20) and (21), where the distribution function of L-statistic $\tilde{\sigma}_n^{-1} S_n$ is approximated by its one-term Edgeworth expansion (18). We consider the case where the weight function $J(\cdot)$ is sufficiently smooth.

Theorem 17 *Assume that $n_* \to \infty$ and $\tilde{\sigma}_n$ remains bounded away from zero for all n_*.*

(i) Assume that (22) *holds, $J(\cdot)$ has a bounded second derivative $J''(\cdot)$ on* $(0,1)$ *and, for some $\delta > 0$, $\mathbf{E}\,|X_1|^{3+\delta} < \infty$. Then* (20) *holds.*

(ii) Assume that (23) *holds, $J(\cdot)$ has a bounded third derivative $J'''(\cdot)$ on* $(0,1)$ *and, $\mathbf{E}\,|X_1|^4 < \infty$. Then* (21) *holds.*

The conditions of Theorem 17 on the boundedness of the derivatives of $J(\cdot)$ are quite restrictive. Thus the theorem cannot be applied to such important statistics as, e.g., trimmed means. Then an interesting question arises whether it is possible to modify Theorem 17, as a consequence of Theorem 5, so that to impose less requirements on the weights $c_1, \ldots, c_n$. In the case of i.i.d. observations, a certain answer is given in Gribkova and Helmers [28], where the trimmed means were considered. In particular, it is shown in Lemma A.2 of [28] that there are essential difficulties in proving that a one-term Edgeworth expansion approximates the distribution function of the trimmed mean with an error of the order $o(n^{-1/2})$, if we try to infer directly from the general result for the symmetric statistics of Putter and van Zwet [59]. Thus, it seems that, in the case of samples drawn without replacement, the trimmed means and other similar statistics also need different methods for the proof of the validity of the one-term Edgeworth expansion. The next example, in the case of i.i.d. observations, is Alberink et al. [2], where the general result for the symmetric statistics of Bentkus et al. [7] was applied. In particular, in [2] only a little weaker conditions for $J(\cdot)$ were obtained. However, here, as well as in the case of trimmed means (see [28]), additional smoothness conditions are imposed on the distribution function of X_1.

Before the proof of Theorem 17, we give a simulation example, which shows how the one-term Edgeworth expansion G_n, defined in (18), improves the usual normal approximation.

Simulation 1 A population $\mathcal{X}$ of size $N = 100$ was simulated from the logistic distribution $\mathcal{L}(0, 1)$. Our chosen population $\mathcal{X}$ has the mean 0.004 and variance 3.270. Consider the L-statistic given in Example 2. Note that the smoothness of its weight function is suitable to apply Theorem 17.

For samples of sizes $n = 5, 15, 30$, we present several q-quantiles of the functions $\tilde{F}_n$, G_n and Φ in Table 2.1 below. Here $\tilde{F}_n$ is the Monte–Carlo approximation to F_n, see Appendix A.1, where $C = 10^7$. Evidently, the functions $g_1(\cdot)$ and $g_2(\cdot,\cdot)$, given in Theorem 11, were used for the calculation of G_n.

Table 2.1: Approximations to $\tilde{F}_n$, $n = 5, 15, 30$.

$q =$	0.01	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.99
$\tilde{F}_5^{-1}(q)$	-2.12	-1.53	-1.22	-0.69	-0.06	0.63	1.32	1.75	2.58
$G_5^{-1}(q)$	-2.08	-1.56	-1.25	-0.70	-0.06	0.64	1.32	1.75	2.56
$\tilde{F}_{15}^{-1}(q)$	-2.16	-1.58	-1.25	-0.69	-0.04	0.65	1.31	1.71	2.49
$G_{15}^{-1}(q)$	-2.16	-1.59	-1.26	-0.70	-0.04	0.65	1.31	1.71	2.49
$\tilde{F}_{30}^{-1}(q)$	-2.20	-1.60	-1.27	-0.69	-0.03	0.66	1.30	1.69	2.43
$G_{30}^{-1}(q)$	-2.22	-1.61	-1.27	-0.69	-0.03	0.66	1.30	1.69	2.43
$\Phi^{-1}(q)$	-2.33	-1.64	-1.28	-0.67	0.00	0.67	1.28	1.64	2.33

Table 2.1 shows that, even for a small sample of size $n = 5$, the Edgeworth expansion remains much more efficient than the normal approximation.

Proof of Theorem 17

Firstly, we note that it follows from $|J''(y)| < \infty$ for all $y \in (0, 1)$ that for some constants a, b and c we have

$$\max_{1\le j\le n} \left|\Delta^0(c_j)\right| \le a, \quad n \max_{2\le j\le n} \left|\Delta^1(c_j)\right| \le b, \quad n^2 \max_{3\le j\le n} \left|\Delta^2(c_j)\right| \le c, \tag{2.23}$$

for all n. Similarly, $|J'''(y)| < \infty$ for all $y \in (0, 1)$ implies that for some constant d we have

$$n^3 \max_{4\le j\le n} \left|\Delta^3(c_j)\right| \le d \tag{2.24}$$

for all n.

Following van Zwet [74] we introduce the functions

$$\begin{aligned} G(x) &= \int_{-\infty}^{x} F(y)\,dy, \quad H(x) = \int_{x}^{+\infty} (1 - F(y))\,dy, \\ M(x) &= \int_{-\infty}^{x} F(y)(1 - F(y))\,dy, \end{aligned} \tag{2.25}$$

where

$$F(y) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\{x_i \le y\} \tag{2.26}$$

is the distribution function of the random variable X_1.

Assume further that, without loss of generality, $x_1 \le \cdots \le x_N$. A simple integration of (2.26) and some work with sums yield that at the points $x = x_k$, $1 \le k \le N$ we have

$$\begin{aligned} G(x_k) &= \sum_{i=1}^{k-1} \frac{i}{N}\, \triangle_i, \quad H(x_k) = \sum_{i=k}^{N-1} \left(1 - \frac{i}{N}\right) \triangle_i, \\ M(x_k) &= \sum_{i=1}^{k-1} \frac{i}{N} \left(1 - \frac{i}{N}\right) \triangle_i\,. \end{aligned}$$

Functions (2.25) are finite and monotone, and, similarly as in Putter [58], we obtain for $1 \le k \le N$

$$G(x_k) + H(x_k) \le \mathbf{E}\,|X_1| + |x_k| \tag{2.27}$$

and

$$M(x_N) \le G(x_k) + H(x_k) \le \mathbf{E}\,|X_1| + |x_k|\,. \tag{2.28}$$

One can show (we omit a detailed proof) that the quantities $\phi_{k,l}(i)$ and $\theta_{k,l,m}(i)$, defined by (1.24) and (1.26), satisfy

$$|\phi_{k,l}(i)| \le \left(\frac{N}{N-1}\right)^2 \left|\phi'_{k,l}(i)\right| \le 4\left|\phi'_{k,l}(i)\right| \tag{2.29}$$

and

$$|\theta_{k,l,m}(i)| \le \left(\frac{N}{N-2}\right)^3 \left|\theta'_{k,l,m}(i)\right| \le 27\left|\theta'_{k,l,m}(i)\right|, \tag{2.30}$$

where we write

$$\phi'_{k,l}(i) = \left(\mathbb{I}\{i \ge k\} - i/N\right)\left(\mathbb{I}\{i \ge l\} - i/N\right)$$

and

$$\theta'_{k,l,m}(i) = \left(\mathbb{I}\{i \ge k\} - i/N\right)\left(\mathbb{I}\{i \ge l\} - i/N\right)\left(\mathbb{I}\{i \ge m\} - i/N\right).$$

Then, by (2.23), (2.27) and because $\sum_{j=1}^{n} \mathcal{H}_{N-2,n-1,i-1}(j-1) = 1$ (by (1.12)), for $1 \leq k \leq N$,

$$\begin{aligned}|g_1(x_k)| &\leq an^{-1/2} \sum_{i=1}^{N-1} |\varphi_k(i)| \, \triangle_i = an^{-1/2} \left[G(x_k) + H(x_k) \right] \\ &\leq an^{-1/2} \left[\mathbf{E} |X_1| + |x_k| \right],\end{aligned}$$

and thus for $s \geq 1$,

$$\begin{aligned}\mathbf{E} |g_1(X_1)|^s &= \frac{1}{N} \sum_{k=1}^{N} |g_1(x_k)|^s \leq a^s n^{-s/2} \frac{1}{N} \sum_{k=1}^{N} \left[\mathbf{E} |X_1| + |x_k| \right]^s \\ &\leq a^s n^{-s/2} \frac{1}{N} \sum_{k=1}^{N} 2^{s-1} \left[(\mathbf{E} |X_1|)^s + |x_k|^s \right] \leq 2^s a^s n^{-s/2} \, \mathbf{E} |X_1|^s .\end{aligned} \tag{2.31}$$

Also, by (2.23), (2.29), (2.27), (2.28), because of monotonicity of G, H, M and $\sum_{j=2}^{n} \mathcal{H}_{N-4,n-2,i-2}(j-2) = 1$, for $1 \leq k < l \leq N$,

$$\begin{aligned}|g_2(x_k, x_l)| &\leq bn^{-3/2} \sum_{i=1}^{N-1} |\phi_{k,l}(i)| \, \triangle_i \leq 4bn^{-3/2} \sum_{i=1}^{N-1} \left| \phi'_{k,l}(i) \right| \triangle_i \\ &\leq 4bn^{-3/2} \left[G(x_l) + M(x_N) + H(x_l) \right] \leq 8bn^{-3/2} \left[\mathbf{E} |X_1| + |x_l| \right],\end{aligned}$$

and then for $s \geq 1$,

$$\begin{aligned}\mathbf{E} |g_2(X_1, X_2)|^s &= \binom{N}{2}^{-1} \sum_{1 \leq k < l \leq N} |g_2(x_k, x_l)|^s \\ &\leq 2^{3s} b^s n^{-3s/2} \binom{N}{2}^{-1} \sum_{1 \leq k < l \leq N} \left[\mathbf{E} |X_1| + |x_l| \right]^s \\ &\leq 2^{3s} b^s n^{-3s/2} \binom{N}{2}^{-1} \sum_{1 \leq k < l \leq N} 2^{s-1} \left[(\mathbf{E} |X_1|)^s + |x_l|^s \right] \\ &\leq 2^{4s+1} b^s n^{-3s/2} \, \mathbf{E} |X_1|^s .\end{aligned} \tag{2.32}$$

Similarly, by (2.23), (2.30), (2.27), (2.28), because of monotonicity of G, H, M and $\sum_{j=3}^{n} \mathcal{H}_{N-6,n-3,i-3}(j-3) = 1$, for $1 \leq k < l < m \leq N$,

$$|g_3(x_k, x_l, x_m)| \leq 3^4 cn^{-5/2} \left[\mathbf{E} |X_1| + |x_m| \right],$$

and then for $s \geq 1$,

$$\mathbf{E} |g_3(X_1, X_2, X_3)|^s \leq 2^{s+2} 3^{4s} c^s n^{-5s/2} \, \mathbf{E} |X_1|^s . \tag{2.33}$$

Note that, it follows from inequalities (2.31), (2.32) and (2.33) that, for both cases *(i)* and *(ii)*, the moments β_s, γ_s and ζ_s in Theorem 5 are bounded if the corresponding moments $\mathbf{E}\,|X_1|^s$ are finite.

Next we evaluate the quantities $\delta_3(S_n)$ and $\delta_4(S_n)$. Let us introduce the events $\mathfrak{R}_{k;ij} = \{R_{k:2k} = i, R_{n+1:2k} = j\}$, $1 \le i < j \le n+k$, $k = 3, 4$. Similarly, as in the cases $k = 1, 2$ (recall (2.3) and (2.15), respectively), we find that

$$p_{k;ij} := \mathbf{P}\left\{\mathfrak{R}_{k;ij}\right\} = \binom{i-1}{k-1}\binom{n+k-j}{k-1} \Big/ \binom{n+k}{2k}.$$

Next, we proceed very similarly as in the proof of Theorem 15. We write $\mathfrak{B}_{k;ijlm} = \{X_{i:n+k} = x_l, X_{j:n+k} = x_m\}$, $1 \le l < m \le N$, $k = 3, 4$, and, by invoking Lemma 8, calculate

$$p_{k;ijlm} := \mathbf{P}\left\{\mathfrak{B}_{k;ijlm}\right\} = \binom{l-1}{i-1}\binom{m-l-1}{j-i-1}\binom{N-m}{n+k-j} \Big/ \binom{N}{n+k}.$$

Note that the events $\mathfrak{R}_{k;ij}$ and $\mathfrak{B}_{k;ijlm}$, $1 \le l < m \le N$ are independent. Denote $C_1 = \max\left\{c^2, d^2\right\}$. Then application of Lemma 12 and the use of the corresponding conditions (2.23) and (2.24) for $k = 3, 4$ yield

$$\begin{aligned}
\delta_k(S_n) = \mathbf{E}\left(n_*^{(k-1)}\mathbb{D}_k S_n\right)^2 &= n_*^{2(k-1)} \sum_{1\le i<j\le n+k} \mathbf{E}\left[(\mathbb{D}_k S_n)^2 \,\Big|\, \mathfrak{R}_{k;ij}\right] p_{k;ij} \\
&\le C_1 n^{-1} \sum_{1\le i<j\le n+k} \mathbf{E}\left[(X_{j:n+k} - X_{i:n+k})^2 \,\Big|\, \mathfrak{R}_{k;ij}\right] p_{k;ij} \\
&= C_1 n^{-1} \sum_{1\le l<m\le N} \lambda_{k;lm}(x_m - x_l)^2,
\end{aligned}$$

where

$$\lambda_{k;lm} = \sum_{1\le i<j\le n+k} p_{k;ij}\, p_{k;ijlm}.$$

Recall the first expression of $\mathbf{Var}\, X_1$ in (2.5). Clearly, to prove the bounds $\delta_k(S_n) = O(n_*^{-1})$, $k = 3, 4$, it will now suffice to show that $\lambda_{k;lm} = O(N^{-2})$ for all $1 \le l < m \le N$.

Using $\binom{u}{v} \le u^v/v!$, taking $j = i+1$ and applying $\max_{0\le u\le 1} u(1-u) \le 1/4$, for all $1 \le i < j \le n+k$ we obtain

$$\begin{aligned}
\binom{n+k}{2k} p_{k;ij} &\le \frac{(i-1)^{k-1}}{(k-1)!}\frac{(n+k-j)^{k-1}}{(k-1)!} \le \frac{[(i-1)(n+k-1-i)]^{k-1}}{[(k-1)!]^2} \\
&= \frac{(n+k-2)^{2(k-1)}}{[(k-1)!]^2}\left[\frac{i-1}{n+k-2}\left(1 - \frac{i-1}{n+k-2}\right)\right]^{k-1} \\
&\le \left(\frac{1}{4}\right)^{k-1}\frac{(n+k-2)^{2(k-1)}}{[(k-1)!]^2}.
\end{aligned}$$

Also, applying the case $T = 3$ of (1.13), for all $1 \le l < m \le N$

$$\begin{aligned}
&\sum_{1 \le i < j \le n+k} p_{k;ijlm} \\
&\quad = \binom{N}{n+k}^{-1} \sum_{s=0}^{n+k-2} \sum_{t=0}^{n+k-2-s} \binom{l-1}{s} \binom{m-l-1}{t} \binom{N-m}{n+k-2-s-t} \\
&\quad = \binom{N}{n+k}^{-1} \binom{N-2}{n+k-2}.
\end{aligned}$$

Finally, it follows from the last two evaluations that, for all $1 \le l < m \le N$, $\lambda_{k;lm} \le C_k N^{-2}$, where

$$C_k = 2\left(\frac{1}{4}\right)^{k-1} \frac{(2k)!}{[(k-1)!]^2} \frac{(2k-2)^{2k-3}}{(2k-3)!}, \qquad k = 3, 4.$$

Application of Theorem 5 completes the proof of both cases *(i)* and *(ii)* of the theorem.

Chapter 3

Estimation of parameters

3.1 Variance

3.1.1 Jackknife estimator

A very common method, used for estimating variance of an L-statistic, is the jackknife. In the case of i.i.d. observations, see, e.g., Efron and Stein [25] and Karlin and Rinott [43], regarding the classical Quenouille–Tukey jackknife estimator of variance. For samples drawn without replacement, the jackknife estimator of $\sigma_L^2 = \mathbf{Var}\, L_n$ is defined as follows. Let the weights $c_1, \dots, c_n$ be determined as in (2). Then, given the sample $\mathbb{X}$ drawn without replacement from $\mathcal{X}$, the jackknife estimator is

$$S^2(L_n) = \Big(1 - \frac{n}{N}\Big) \frac{n-1}{n} \sum_{k=1}^{n} \Big(L_{(k)} - \overline{L}\Big)^2, \qquad \overline{L} = \frac{1}{n} \sum_{k=1}^{n} L_{(k)}, \tag{3.1}$$

where $L_{(k)} = L_{n-1}(\mathbb{X} \backslash \{X_k\})$, $1 \le k \le n$ are L-statistics (1) with the weights $c'_j = J(j/n)$, $1 \le j \le n-1$.

Note that, other than in the case of i.i.d. observations, jackknife estimator (3.1) includes the finite population correction factor. The properties of this estimator (such as bias and consistency) were studied by Bloznelis [15], see also Bloznelis and Götze [20], in the case of finite population symmetric statistics. It is known that, in the case of i.i.d. observations, the quality of the jackknife variance estimator depends on the smoothness of the underlying statistic, see, e.g., Shao and Wu [65]. In the case of L-statistics, we understand it as the smoothness of the weight function $J\colon (0,1) \to \mathbb{R}$.

Next, we give a different expression of (3.1), which will be useful later:

$$S^2(L_n) = \Big(1 - \frac{n}{N}\Big) \frac{1}{n(n-1)} \sum_{k=1}^{n} \Big(Y_k - \overline{Y}\Big)^2, \qquad \overline{Y} = \frac{1}{n} \sum_{k=1}^{n} Y_k, \tag{3.2}$$

where $Y_k = Y_{k-1} + c'_{k-1}\mathbf{\Delta}_{k-1:n}$ for $2 \le k \le n$, and $Y_1 := 0$. We omit the proof of this identity, which is quite straightforward.

3.1.2 Bootstrap estimator

For samples drawn without replacement, there are a few bootstrap procedures considered in literature, see, e.g., Rao and Wu [60] and Sitter [67]. We consider here the finite population bootstrap of Booth et al. [21]. Write $N = mn+t$, where $0 \le t < n$. The empirical (bootstrap) population $\tilde{\mathcal{X}}$ is defined by taking m copies $\mathcal{X}_j = \{X_{j1}, \dots, X_{jn}\}$, $1 \le j \le m$ of $\mathbb{X}$ and, if $t > 0$, drawing the simple random sample $\mathcal{Y} = \{Y_1, \dots, Y_t\}$ of size t without replacement from $\mathbb{X}$. If $t = 0$, then put $\mathcal{Y} = \emptyset$. Then

$$\tilde{\mathcal{X}} = \left(\bigcup_{j=1}^{m} \mathcal{X}_j \right) \cup \mathcal{Y}. \tag{3.3}$$

For any population parameter (characteristic) $\theta = \theta(\mathcal{X})$, the bootstrap estimator is then defined as the conditional expectation

$$\hat{\theta}_B = \mathbf{E}\left(\theta(\tilde{\mathcal{X}}) \,\middle|\, \mathbb{X} \right), \tag{3.4}$$

i.e., expectation over all empirical populations conditional on $\mathbb{X}$. Note that the case $t = 0$ was first considered by Gross [31].

Note that, for L-statistics, the properties of bootstrap estimator (3.4) are still not well known. As to discussions about possible bias of estimator (3.4) in the case of variance of statistics, see [21] and references therein and also Bloznelis [17].

Next, we give an exact expression of (3.4) for the parameter σ_L^2. Obviously, in practice the approximation to bootstrap estimate (3.4) can be obtained by using the Monte–Carlo method (see, e.g., [21]), but it also implies an additional error. Thus, we eliminate this error. In the case of i.i.d. observations, a similar problem in the case of the naive bootstrap for variance of an L-statistic was solved by Hutson and Ernst [40], see also Huang [39].

Let us write

$$\sigma_L^2 = n^{-2}\left[\sum_{p=1}^{n} c_p^2 \,\mathbf{Var}\, X_{p:n} + 2 \sum_{1 \le p < r \le n} c_p c_r \,\mathbf{Cov}\,(X_{p:n}, X_{r:n}) \right]. \tag{3.5}$$

First, we express (3.5) in terms of the moments $\mathbf{E}\,\mathbf{\Delta}_{u:n}$, $\mathbf{E}\,\mathbf{\Delta}_{u:n}^2$, $0 \le u \le n-1$ and $\mathbf{E}\,\mathbf{\Delta}_{u:n}\mathbf{\Delta}_{v:n}$, $0 \le u < v \le n-1$, which are calculated in Lemma 9 above. Actually, using representation (1.20), it is easy to get

$$\mathbf{Cov}\,(X_{p:n}, X_{r:n}) = \mathbf{Var}\, X_{p:n} + \sum_{u=0}^{p-1} \sum_{v=p}^{r-1} \mathbf{Cov}\,(\mathbf{\Delta}_{u:n}, \mathbf{\Delta}_{v:n}), \tag{3.6}$$

for $1 \le p < r \le n$, with

$$\mathbf{Var}\, X_{p:n} = \sum_{u=0}^{p-1} \mathbf{Var}\, \mathbf{\Delta}_{u:n} + 2 \sum_{0 \le u < v \le p-1} \mathbf{Cov}\left(\mathbf{\Delta}_{u:n}, \mathbf{\Delta}_{v:n}\right), \tag{3.7}$$

for $1 \le p \le n$. These expressions of $\mathbf{Var}\, X_{p:n}$ and $\mathbf{Cov}\,(X_{p:n}, X_{r:n})$ are based on the moments of spacings only. Second, we find bootstrap estimator (3.4) for any of the population parameters $\theta_u = \mathbf{E}\, \mathbf{\Delta}_{u:n}$, $\theta_{uu} = \mathbf{E}\, \mathbf{\Delta}^2_{u:n}$, $0 \le u \le n-1$ and $\theta_{uv} = \mathbf{E}\, \mathbf{\Delta}_{u:n}\mathbf{\Delta}_{v:n}$, $0 \le u < v \le n-1$. Write, for short, $b_{ik} = \mathcal{H}_{n,t,i}(k)$, $0 \le k \le t$, $1 \le i \le n-1$. Denote

$$b_{ijkl} = \binom{n}{t}^{-1} \binom{i}{k} \binom{j-i}{l-k} \binom{n-j}{t-l}, \qquad 0 \le k \le l \le t, \quad 1 \le i < j \le n-1.$$

Recall the numbers $h_i(u)$, $h_{ij}(u)$ and $h_{ij}(u,v)$ defined in (1.14), (1.15) and (1.16).

Theorem 18 *The bootstrap estimators of θ_u, θ_{uu}, $0 \le u \le n-1$ and θ_{uv}, $0 \le u < v \le n-1$ are*

$$\hat{\theta}_{uB} = \sum_{i=1}^{n-1} \sum_{k=0}^{t} h_{mi+k}(u) b_{ik} \mathbf{\Delta}_{i:n}, \tag{3.8}$$

$$\begin{aligned} \hat{\theta}_{uuB} &= \sum_{i=1}^{n-1} \sum_{k=0}^{t} h_{mi+k}(u) b_{ik} \mathbf{\Delta}^2_{i:n} \\ &\quad + 2 \sum_{1 \le i < j \le n-1} \sum_{0 \le k \le l \le t} h_{mi+k;mj+l}(u) b_{ijkl} \mathbf{\Delta}_{i:n} \mathbf{\Delta}_{j:n}, \end{aligned} \tag{3.9}$$

$$\hat{\theta}_{uvB} = \sum_{1 \le i < j \le n-1} \sum_{0 \le k \le l \le t} h_{mi+k;mj+l}(u,v) b_{ijkl} \mathbf{\Delta}_{i:n} \mathbf{\Delta}_{j:n}, \tag{3.10}$$

respectively.

Finally, replacing the moments of spacings in (3.7) and (3.6) by their bootstrap estimators given in Theorem 18, we obtain from (3.5) the bootstrap estimator

$$\begin{aligned} \hat{\sigma}^2_B = n^{-2} \Bigg[\sum_{p=1}^{n} c_p^2 \widehat{\mathbf{Var}} X_{p:n} + 2 \sum_{1 \le p < r \le n} c_p c_r \Bigg\{ & \widehat{\mathbf{Var}} X_{p:n} \\ & + \sum_{u=0}^{p-1} \sum_{v=p}^{r-1} (\hat{\theta}_{uvB} - \hat{\theta}_{uB} \hat{\theta}_{vB}) \Bigg\} \Bigg], \end{aligned} \tag{3.11}$$

where

$$\widehat{\mathbf{Var}} X_{p:n} = \sum_{u=0}^{p-1} (\hat{\theta}_{uuB} - \hat{\theta}^2_{uB}) + 2 \sum_{0 \le u < v \le p-1} (\hat{\theta}_{uvB} - \hat{\theta}_{uB} \hat{\theta}_{vB}), \qquad 1 \le p \le n.$$

Proof of Theorem 18

Consider the empirical population $\tilde{\mathcal{X}} = \{X_{1:n}, \ldots, X_{1:n}, \ldots, X_{n:n}, \ldots, X_{n:n}\}$.

We prove (3.8). For the population parameter $\theta_u = \theta_u(\mathcal{X})$, given in (1.17), we have, for any $0 \leq u \leq n-1$, that

$$\theta_u(\tilde{\mathcal{X}}) = \sum_{i=1}^{n-1} h_{p_i}(u) \mathbf{\Delta}_{i:n},$$

where p_i, $1 \leq i \leq n-1$ is a random number from the set $\{1, \ldots, N-1\}$ with $\mathbf{P}\{p_i = mi + k\} = b_{ik}$, $0 \leq k \leq t$. Thus, for $1 \leq i \leq n-1$, we get

$$\mathbf{E}\left(h_{p_i}(u) \mathbf{\Delta}_{i:n} \,|\, \mathbb{X}\right) = \mathbf{\Delta}_{i:n} \sum_{k=0}^{t} h_{mi+k}(u) b_{ik}.$$

Formula (3.8) is proven.

Now we prove (3.10). For the population parameter $\theta_{uv} = \theta_{uv}(\mathcal{X})$, given in (1.19), we have, for all $0 \leq u < v \leq n-1$, that

$$\theta_{uv}(\tilde{\mathcal{X}}) = \sum_{1 \leq i < j \leq n-1} h_{p_i r_j}(u, v) \mathbf{\Delta}_{i:n} \mathbf{\Delta}_{j:n},$$

where (p_i, r_j), $1 \leq i < j \leq n-1$ is a pair of random numbers from the set $\{1, \ldots, N-1\}$ with $\mathbf{P}\{p_i = mi + k, r_j = mj + l\} = b_{ijkl}$, $0 \leq k \leq l \leq t$. Thus, for $1 \leq i < j \leq n-1$, we obtain

$$\mathbf{E}\left(h_{p_i r_j}(u, v) \mathbf{\Delta}_{i:n} \mathbf{\Delta}_{j:n} \,\middle|\, \mathbb{X}\right) = \mathbf{\Delta}_{i:n} \mathbf{\Delta}_{j:n} \sum_{0 \leq k \leq l \leq t} h_{mi+k;mj+l}(u, v) b_{ijkl}.$$

Formula (3.10) is proven.

The proof of (3.9) is, in fact, the same as that of (3.8) and (3.10).

3.1.3 Numerical comparisons

We present some more examples of L-statistics. For more details on the following examples see Chernoff et al. [23].

Example 4 The L-statistic, defined by the weight function $J(u) = \Phi^{-1}(u)$, is applied as an efficient estimator of the scale parameter for the normal distribution. Denote such an L-statistic by N_σ. In the finite population context, the motivation of application of N_σ is the same as in Example 3.

Example 5 The L-statistic with weights, generated by the weight function $J(u) = \sin 4\pi\,(u - 1/2) \,/\, \tan \pi\,(u - 1/2)$, is applied as an efficient estimator of the location parameter for the Cauchy distribution. Denote such an L-statistic by C_μ. If we assume that the population $\mathcal{X}$ is obtained from the superpopulation with the Cauchy distribution, C_μ will estimate a center of $\mathcal{X}$.

Example 6 The L-statistic with weights, generated by the weight function $J(u) = 8\tan\pi\,(u-1/2)\,/\sec^4\pi\,(u-1/2)$, is applied as an efficient estimator of the scale parameter for the Cauchy distribution. Denote such an L-statistic by C_σ. This statistic can be useful in the estimation of the interquartile range of $\mathcal{X}$.

The quality of any variance estimator is always important, when, e.g., we construct confidence intervals for an L-statistic (-estimator) or, we need to choose between two or more competing L-statistics. Next, for some of the discussed L-statistics, we compare the efficiencies of variance estimators $S^2 = S^2(L_n)$ and $\hat{\sigma}^2_B$, given by (3.1) and (3.11), recpectively.

Simulation 2 Let us consider two different populations of sizes $N = 50$. The first population $\mathcal{X}_\mathcal{N}$ was simulated from the normal distribution $\mathcal{N}(2,4)$. Our chosen population $\mathcal{X}_\mathcal{N}$ has the mean 2.002 and variance 3.995. The second population $\mathcal{X}_\mathcal{C}$ was simulated from the Cauchy distribution $\mathcal{C}(2,1)$. Our chosen population $\mathcal{X}_\mathcal{C}$ has the median 1.996 and interquartile range 2.013. We choose $n = 20$ for both populations.

Table 3.1 presents numerical results for the statistics $M_{0;1}$, $M_{0.2;0.8}$ and C_μ, see Examples 1 and 5. Table 3.2 shows simulation results for the statistics U_G, N_σ and C_σ, see Examples 3, 4 and 6, recpectively. In particular, for each of the L-statistics, we give the value of its variance σ^2_L and estimated values of the mean square errors (MSEs) and biases (BIASes) of its estimators S^2 and $\hat{\sigma}^2_B$. In order to estimate MSE and BIAS, we draw independently $R = 200$ samples from the population of interest, and, e.g., for realizations $\hat{\sigma}^2_{B;r}$, $1 \le r \le R$, of the bootstrap estimator $\hat{\sigma}^2_B$, we take

$$\widehat{\mathrm{MSE}}(\hat{\sigma}^2_B) = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\sigma}^2_{B;r} - \sigma^2_L\right)^2$$

and

$$\widehat{\mathrm{BIAS}}(\hat{\sigma}^2_B) = \frac{1}{R}\sum_{r=1}^{R}\hat{\sigma}^2_{B;r} - \sigma^2_L. \tag{3.12}$$

It is seen that for the L-statistics, which are used as estimators of location, see Table 3.1, the jackknife variance estimator S^2 is a little more efficient (see MSE) as compared to the bootstrap estimator $\hat{\sigma}^2_B$. On the other hand, in the case of estimators of scale (Table 3.2), $\hat{\sigma}^2_B$ is more efficient compared to S^2. Note that for both the populations $\mathcal{X}_\mathcal{N}$ and $\mathcal{X}_\mathcal{C}$, the estimator $\hat{\sigma}^2_B$ outperforms the estimator S^2

Table 3.1: Variances of the estimators of location for $\mathcal{X}_\mathcal{N}$ and $\mathcal{X}_\mathcal{C}$.

	$\mathcal{X}_\mathcal{N}$			$\mathcal{X}_\mathcal{C}$		
	$M_{0;1}$	$M_{0.2;0.8}$	C_μ	$M_{0;1}$	$M_{0.2;0.8}$	C_μ
σ_L^2	0.118	0.105	0.092	0.370	0.103	0.074
$10^3\,\widehat{\mathrm{MSE}}(S^2)$	1.13	3.30	5.69	31.98	8.36	1.63
$10^3\,\widehat{\mathrm{MSE}}(\hat\sigma_B^2)$	1.06	4.06	10.65	30.47	13.77	5.89
$10^3\,\widehat{\mathrm{BIAS}}(S^2)$	7.73	0.30	-2.83	33.37	5.45	-6.96
$10^3\,\widehat{\mathrm{BIAS}}(\hat\sigma_B^2)$	5.52	26.92	45.71	26.23	57.27	37.55

Table 3.2: Variances of the estimators of scale for $\mathcal{X}_\mathcal{N}$ and $\mathcal{X}_\mathcal{C}$.

	$\mathcal{X}_\mathcal{N}$			$\mathcal{X}_\mathcal{C}$		
	U_G	N_σ	C_σ	U_G	N_σ	C_σ
σ_L^2	0.095	0.046	0.079	0.670	0.374	0.118
$10^3\,\widehat{\mathrm{MSE}}(S^2)$	0.70	0.18	2.16	112.10	32.60	15.94
$10^3\,\widehat{\mathrm{MSE}}(\hat\sigma_B^2)$	0.49	0.13	1.67	82.06	23.57	32.09
$10^3\,\widehat{\mathrm{BIAS}}(S^2)$	7.65	3.43	15.62	62.11	27.47	30.58
$10^3\,\widehat{\mathrm{BIAS}}(\hat\sigma_B^2)$	-3.70	-2.55	18.40	-36.75	-40.88	94.07

for the statistics $M_{0;1}$, U_G and N_σ, whereas S^2 is better for the statistics $M_{0.2;0.8}$ and C_μ.

3.2 Parameters defining the Edgeworth expansion

3.2.1 Jackknife estimators

Consider the parameters α and κ given in (19). We define jackknife estimators $\hat\alpha_J$ and $\hat\kappa_J$ of these parameters similarly as in the case of symmetric statistics, see Bloznelis [14]. Our estimators are based on the sample of size n, whereas, in the case of symmetric statistics, it is feasible for the sample of size $n+2$ only. Recall the notation used for the definition of $S^2(L_n)$ in (3.1). For $1 \le k \le n$, $1 \le i, j, r \le n$, $i \ne j$, denote

$$V_k = \overline{L} - L_{(k)}, \qquad \tilde V_r = \overline{\overline{L}} - \overline{L}_{(r)}, \qquad W_{ij} = \overline{\overline{L}} - \overline{L}_{(i)} - \overline{L}_{(j)} + L_{(i,j)},$$

where

$$\overline{L}_{(r)} = \frac{1}{n-1} \sum_{1 \le j \le n,\ j \ne r} L_{(r,j)}, \qquad \overline{\overline{L}} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} L_{(i,j)}.$$

Here $L_{(i,j)} = L_{n-2}(\mathbb{X}\backslash\{X_i, X_j\})$ is the statistic of form (1), with the weights, say, $c_l'' = J(l/(n-1))$, $1 \le l \le n-2$. Then,

$$
\begin{aligned}
\hat{\alpha}_J &= \hat{\sigma}_J^{-3} n^{1/2} \sum_{k=1}^{n} V_k^3, \\
\hat{\kappa}_J &= 2\hat{\sigma}_J^{-3} \left(1 - \frac{n}{N}\right) n^{1/2} \sum_{1 \le i < j \le n} W_{ij} \tilde{V}_i \tilde{V}_j,
\end{aligned}
\tag{3.13}
$$

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

The consistency of very similar jackknife estimators of α and κ was proven by Bloznelis [14] in the case of general symmetric finite population statistics.

3.2.2 Bootstrap estimators

We present exact expressions of Booth et al. [21] bootstrap estimator (3.4) for the parameters α and κ. First, similarly as in the case of variance (see Section 3.1.2), we find the exact bootstrap estimators $\hat{g}_{1B}(k)$ and $\hat{g}_{2B}(k,l)$ for any of the population characteristics $g_1(k) := g_1(x_k)$, $1 \le k \le N$ and $g_2(k,l) := g_2(x_k, x_l)$, $1 \le k < l \le N$ defined in (1.21) and (1.23), respectively. Write

$$
g_1(x_k) = \sum_{i=1}^{N-1} u_i(k)\, \triangle_i, \quad 1 \le k \le N
$$

and

$$
g_2(x_k, x_l) = \sum_{i=1}^{N-1} v_i(k,l)\, \triangle_i, \quad 1 \le k < l \le N,
$$

where, for $1 \le i \le N-1$, we denote

$$
u_i(k) = -n^{-1} \left(\mathbb{I}\{i \ge k\} - \frac{i}{N}\right) \sum_{p=1}^{n} c_p \mathcal{H}_{N-2,n-1,i-1}(p-1), \qquad 1 \le k \le N
$$

and

$$
v_i(k,l) = -n^{-1} \phi_{k,l}(i) \sum_{p=2}^{n} (c_p - c_{p-1}) \mathcal{H}_{N-4,n-2,i-2}(p-2), \qquad 1 \le k < l \le N,
$$

with $\phi_{k,l}(i)$ given by (1.24).

Theorem 19

(i) For $1 \le k \le N$

$$
\hat{g}_{1B}(k) = \sum_{j=1}^{n-1} \sum_{i=mj}^{mj+t} u_i(k) \mathcal{H}_{n,t,j}(i - mj) \boldsymbol{\Delta}_{j:n}.
\tag{3.14}
$$

(ii) For $1 \leq k < l \leq N$

$$\hat{g}_{2B}(k,l) = \sum_{j=1}^{n-1} \sum_{i=mj}^{mj+t} v_i(k,l)\mathcal{H}_{n,t,j}(i-mj)\boldsymbol{\Delta}_{j:n}. \tag{3.15}$$

Next, we substitute estimators (3.14) and (3.15) into (19) thus obtaining the bootstrap estimators

$$\begin{aligned} \hat{\alpha}_B &= \hat{\sigma}_{1B}^{-3} \frac{1}{N} \sum_{k=1}^{N} \hat{g}_{1B}^3(k), \\ \hat{\kappa}_B &= \hat{\sigma}_{1B}^{-3} n \left(1 - \frac{n}{N}\right) \binom{N}{2}^{-1} \sum_{1 \leq k < l \leq N} \hat{g}_{2B}(k,l)\hat{g}_{1B}(k)\hat{g}_{1B}(l), \end{aligned} \tag{3.16}$$

where $\hat{\sigma}_{1B}^2 = N^{-1} \sum_{k=1}^{N} \hat{g}_{1B}^2(k)$.

Proof of Theorem 19

The proofs of (3.14) and (3.15) are, in fact, the same as the proof of (3.8) in Theorem 18. Therefore we omit them.

Chapter 4

Applications

4.1 Approximations to distributions

4.1.1 Edgeworth expansion for a Studentized statistic

In many practical situations, the variance σ_L^2 of L-statistic (1) is unknown. Therefore, if, e.g., we aim to evaluate the reliability of an estimate L_n by constructing a confidence interval, we need approximations to the distribution function of a Studentized L-statistic

$$F_{nS}(x) = \mathbf{P}\left\{L_n - \mathbf{E}\, L_n \leq xS(L_n)\right\}, \tag{4.1}$$

where $S^2(L_n)$ is the estimate of σ_L^2 based on the sample $\mathbb{X}$. Here we consider the frequently used jackknife variance estimator $S^2(L_n)$ given by (3.1). Next, in this section, we give some theoretical insights on the asymptotic normality and the one-term Edgeworth expansion, see (4.2) below, of (4.1).

Asymptotic properties of distribution (4.1) are similar to that of $F_n(x)$ given by (5). By Proposition 3 of Bloznelis and Götze [20], where, in the case of symmetric statistics, $S^2(L_n)$ is defined slightly different (it is based on the extended sample $\mathbb{X}_1 = \{X_1, \ldots, X_{n+1}\}$), the conditions, sufficient for the asymptotic normality of the Studentized L-statistic $(L_n - \mathbf{E}\, L_n)/S(L_n)$, should be the same as in Theorems 15 and 16, since a difference between both similar jackknife estimators of variance is (asymptotically) negligible.

By Bloznelis [16], where the jackknife variance estimator for symmetric statistics is also based on the sample $\mathbb{X}_1$, the one-term Edgeworth expansion of (4.1) is

$$G_{nS}(x) = \Phi(x) + \frac{(q - p + (q+1)x^2)\alpha + 3(x^2+1)\kappa}{6\tau}\Phi'(x). \tag{4.2}$$

Here the numbers τ, p, q are defined in (6), and α and κ are the same characteristics of the Hoeffding decomposition of (1), see (19). Theorem 2 in [16] provides sufficient conditions, which ensure that, in the case of symmetric statistics, the one-term Edgeworth expansion approximates the distribution function of the Studentized statistic up to the error $o(n_*^{-1/2})$. Since jackknife variance estimator (3.1) is very close to that defined in [16], we can formulate a statement on the validity of the one-term Edgeworth expansion (4.2), which is similar to *(i)* of Theorem 17. Indeed, the condition $\mathbf{E}\,|X_1|^{3+\delta} < \infty$, for some $\delta > 0$, of Theorem 17 should be replaced by $\mathbf{E}\,|X_1|^{6+\delta} < \infty$ and the additional condition $q\tau \to \infty$.

4.1.2 Empirical Edgeworth expansions

One-term Edgeworth expansions (18) and (4.2) of distributions (5) and (4.1), respectively, cannot be applied directly if the population parameters α and κ, that define them, are unknown. A reasonable alternative to the true expansions (18) and (4.2) are empirical Edgeworth expansions, where unknown parameters are replaced by their estimators. Here we use the jackknife estimators, given by (3.13), and the bootstrap estimators, see (3.16).

Replacing the true parameters α and κ in (18) and (4.2) by their jackknife estimators $\hat{\alpha}_J$ and $\hat{\kappa}_J$, we obtain the corresponding empirical Edgeworth expansions

$$G_{nJ}(x) = \Phi(x) - \frac{(q-p)\hat{\alpha}_J + 3\hat{\kappa}_J}{6\tau}(x^2 - 1)\Phi'(x) \tag{4.3}$$

and

$$G_{nSJ}(x) = \Phi(x) + \frac{(q-p+(q+1)x^2)\hat{\alpha}_J + 3(x^2+1)\hat{\kappa}_J}{6\tau}\Phi'(x). \tag{4.4}$$

Replacing the true parameters α and κ in (18) and (4.2) by their bootstrap estimators $\hat{\alpha}_B$ and $\hat{\kappa}_B$, we obtain the corresponding empirical Edgeworth expansions

$$G_{nB}(x) = \Phi(x) - \frac{(q-p)\hat{\alpha}_B + 3\hat{\kappa}_B}{6\tau}(x^2 - 1)\Phi'(x) \tag{4.5}$$

and

$$G_{nSB}(x) = \Phi(x) + \frac{(q-p+(q+1)x^2)\hat{\alpha}_B + 3(x^2+1)\hat{\kappa}_B}{6\tau}\Phi'(x). \tag{4.6}$$

Since the defined empirical Edgeworth expansions depend on the random sample $\mathbb{X}$, the validity of these expansions is understood as the validity in probability. To prove the validity of an empirical Edgeworth expansion, if it has already been proven in the case of a true expansion, it suffices to show the consistency of the parameters estimators, which define the empirical expansion.

In the case of symmetric finite population statistics, empirical Edgeworth expansions with jackknife estimators of α and κ were considered by Bloznelis [14, 16], where the conditions sufficient for the validity of true expansions (18) and (4.2), are also sufficient for the consistency of estimators. Note that our estimators $\hat{\alpha}_J$ and $\hat{\kappa}_J$ are almost the same as in [14].

In the case of L-statistics, the properties of the bootstrap estimators, such as $\hat{\alpha}_B$ and $\hat{\kappa}_B$, are still not quite understood. We examine them by numerical simulations in Section 4.1.4, as well as the jackknife estimators.

4.1.3 Bootstrap approximations

Non-parametric bootstrap approximations are, in a sense, close to the one-term Edgeworth approximations, thus the latter are often used in the evaluations of accuracy of bootstrap approximations. Typically, the accuracy of bootstrap approximation is of the same order as in the case of the Edgeworth expansion. We consider here bootstrap approximations to distributions (5) and (4.1) of the L-statistics.

Let $\tilde{\mathcal{X}}$ be the empirical population defined by (3.3). We draw a simple random sample without replacement $\tilde{\mathbb{X}} = \{\tilde{X}_1, \ldots, \tilde{X}_n\}$ from $\tilde{\mathcal{X}}$. Then the bootstrap estimator of statistic (1) is $\tilde{L}_n = L_n(\tilde{\mathbb{X}})$. Denote $\tilde{\sigma}_L^2 = \mathbf{Var}\,(\tilde{L}_n \mid \mathbb{X}, \mathcal{Y})$ and let $\tilde{S}^2(L_n) = S^2(\tilde{L}_n)$ be jackknife estimate (3.1) of $\tilde{\sigma}_L^2$ based on the sample $\tilde{\mathbb{X}}$. Then bootstrap estimators for the distribution functions $F_n(x)$ and $F_{nS}(x)$ are

$$F_n^*(x) = \mathbf{P}\,\{\tilde{L}_n - \mathbf{E}\,(\tilde{L}_n \mid \mathbb{X}, \mathcal{Y}) \leq x\tilde{\sigma}_L \mid \mathbb{X}\} \tag{4.7}$$

and

$$F_{nS}^*(x) = \mathbf{P}\,\{\tilde{L}_n - \mathbf{E}\,(\tilde{L}_n \mid \mathbb{X}, \mathcal{Y}) \leq x\tilde{S}(L_n) \mid \mathbb{X}\}, \tag{4.8}$$

respectively. The theoretical analysis shows that this kind of bootstrap approximation is second-order correct for statistics which are smooth functions of multivariate sample means (see Booth et al. [21]), and U-statistics (see Bloznelis [17]). However, the case of L-statistics is still not well explored for samples drawn without replacement. Similarly, in the case of i.i.d. observations, there are only several results in the special case of trimmed means (see, e.g., Hall and Padmanabhan [32], and Gribkova and Helmers [29]) for the classical Efron bootstrap or the m out of n bootstrap. A hint on the extension of bootstrap results for U-statistics to the case of more general L-statistics appeared in Helmers [35], but it is omitted in the final version of the same paper, see [36]. In the case of samples drawn without replacement, we similarly expect that the case of U-statistics in [17] can be extended to the case of L-statistics. We note in addition that, in the special case of Gini's mean difference (Example 3), the results of [17] are applicable, and the general conditions can be very similarly simplified as in the case of Edgeworth expansion for Studentized L-statistics (recall the discussion at the end of Section

4.1.1), since, in the proofs of [17], the basic tool is Theorem 1 of Bloznelis [16] on the validity of Edgeworth expansion for Studentized U-statistics.

In Section 4.1.4, we examine the accuracy of bootstrap approximations for L-statistics via computer simulation. Note that usually bootstrap estimators (4.7) and (4.8) cannot be applied directly, since for many statistics it is difficult, if not impossible, to find their exact expressions. Therefore, in numerical examples, we apply Monte–Carlo approximations to $F_n^*(x)$ and $F_{nS}^*(x)$ proposed by Booth et al. [21]. We conclude this section with an example of L-statistic, where the bootstrap distribution can be computed analytically.

Example 7 A simple estimator of the finite population q-quantile $F^{-1}(q) = \inf\{x : F(x) \geq q\}$, $0 < q < 1$, where $F(x)$ is the distribution function of $\mathcal{X}$, see (2.26), is a single order statistic (empirical quantile) $X_{[qn]+1:n}$.

Proposition 20 *For a single order statistic $X_{r:n}$, where $1 \leq r \leq n$, we have for $1 \leq j \leq n$*

$$\begin{aligned} \mathbf{P}\left\{\tilde{X}_{r:n} = X_{j:n} \mid \mathbb{X}\right\} = &\sum_{s=1}^{t+1} \sum_{i=m(j-1)+s}^{mj+s-1} p_r(i) \binom{j-1}{s-1}\binom{n-j}{t+1-s} \Big/ \binom{n}{t} \\ &+ \sum_{s=1}^{t} \sum_{i=m(j-1)+s}^{mj+s} p_r(i) \binom{j-1}{s-1}\binom{n-j}{t-s} \Big/ \binom{n}{t}, \end{aligned} \tag{4.9}$$

where

$$p_r(i) = \mathbf{P}\left\{X_{r:n} = x_i\right\} = \binom{i-1}{r-1}\binom{N-i}{n-r} \Big/ \binom{N}{n}, \qquad 1 \leq i \leq N.$$

Proof of Proposition 20

For $1 \leq j \leq n$, consider a pair of random variables $(u_j, v_j) \in \{(k,l) : 1 \leq k \leq l \leq N\}$, where u_j and v_j are the lowest and the highest positions of $X_{j:n}$ in the ordered empirical population $\tilde{\mathcal{X}} = \{X_{1:n}, \ldots, X_{1:n}, \ldots, X_{n:n}, \ldots, X_{n:n}\}$. Consider the events

$$\mathcal{A}_s = \{(u_j, v_j) = (m(j-1)+s, mj+s-1)\}, \qquad 1 \leq s \leq t+1$$

and

$$\mathcal{B}_s = \{(u_j, v_j) = (m(j-1)+s, mj+s)\}, \qquad 1 \leq s \leq t.$$

These events are mutually exclusive and appear with the probabilities

$$\mathbf{P}\left\{\mathcal{A}_s\right\} = \binom{j-1}{s-1}\binom{1}{0}\binom{n-j}{t+1-s} \Big/ \binom{n}{t}, \qquad 1 \leq s \leq t+1$$

and

$$\mathbf{P}\{\mathcal{B}_s\} = \binom{j-1}{s-1}\binom{1}{1}\binom{n-j}{t-s} \Big/ \binom{n}{t}, \qquad 1 \le s \le t,$$

which sum up to 1. The law of the total probability yields (4.9).

4.1.4 Simulation study

Note that, in the simulation examples below, we use Monte-Carlo approximations $\tilde{F}_n$ and $\tilde{F}_{nS}$ to the exact distributions F_n and F_{nS}, respectively, see Appendix A.1, where we choose $C = 10^6$. Similarly, we approximate bootstrap distributions F_n^* and F_{nS}^* by $\tilde{F}_n^*$ and $\tilde{F}_{nS}^*$, respectively, see Appendix A.2, where $B = 10^2$ and $R = 10^4$.

In the tables below, we compare the distributions $\tilde{F}_n$, Φ, G_n, G_{nJ}, G_{nB}, $\tilde{F}_n^*$ and also $\tilde{F}_{nS}$, Φ, G_{nS}, G_{nSJ}, G_{nSB}, $\tilde{F}_{nS}^*$ by taking their q-quantiles, $q = 0.01, 0.05, 0.10, 0.90, 0.95, 0.99$. Specifically, for the empirical functions G_{nJ}, G_{nB}, $\tilde{F}_n^*$ and G_{nSJ}, G_{nSB}, $\tilde{F}_{nS}^*$ we give two characterstics of the empirical quantile: estimated values of its expectation and the standard error (SE) based on $R = 200$ samples drawn independently and without replacement from $\mathcal{X}$. That is, e.g., for realizations $G_{nJ;r}^{-1}(q)$, $1 \le r \le R$, of the empirical quantile $G_{nJ}^{-1}(q)$, the expectation $\mathbf{E}\, G_{nJ}^{-1}(q)$ is estimated by the formula

$$\widehat{\mathbf{E}}\, G_{nJ}^{-1}(q) = \frac{1}{R}\sum_{r=1}^{R} G_{nJ;r}^{-1}(q), \tag{4.10}$$

and the standard error $\mathbf{S}\, G_{nJ}^{-1}(q)$ is estimated by

$$\widehat{\mathbf{S}}\, G_{nJ}^{-1}(q) = \left(\frac{1}{R}\sum_{r=1}^{R}\left(G_{nJ;r}^{-1}(q) - \widehat{\mathbf{E}}\, G_{nJ}^{-1}(q)\right)^2\right)^{1/2}. \tag{4.11}$$

We also give values of the parameters α and κ, and estimated values of the biases (BIASes) and SEs of their estimators $\hat{\alpha}_J$, $\hat{\alpha}_B$ and $\hat{\kappa}_J$, $\hat{\kappa}_B$. Biases are estimated just like in (3.12).

Simulation 3 Consider Gini's mean difference U_G, see Example 3. A population $\mathcal{X}$ of size $N = 150$ was simulated from the normal distribution $\mathcal{N}(2, 4)$. Our chosen population $\mathcal{X}$ has the mean 2.01 and variance 4.03. The sample size is $n = 45$.

Table 4.1 shows that both G_n and Φ approximate $\tilde{F}_n$ similarly and quite well, i.e., the normal approximation seems sufficient in this simulation example. Next, G_{nB} (as the estimate of G_n) has a larger bias compared to G_{nJ}. It means that the jackknife estimate of κ is more successful compared to the bootstrap estimate, see

Table 4.1: Simulation 3. Approximations to $\tilde{F}_n$.

$q =$	0.01	0.05	0.10	0.90	0.95	0.99
$\tilde{F}_n^{-1}(q)$	-2.29	-1.64	-1.29	1.29	1.65	2.31
$\Phi^{-1}(q)$	-2.33	-1.64	-1.28	1.28	1.64	2.33
$G_n^{-1}(q)$	-2.32	-1.64	-1.28	1.28	1.65	2.33
$\widehat{\mathbf{E}}\, G_{nJ}^{-1}(q)$	-2.30	-1.64	-1.28	1.28	1.65	2.35
$\widehat{\mathbf{E}}\, G_{nB}^{-1}(q)$	-2.36	-1.66	-1.29	1.28	1.63	2.29
$\widehat{\mathbf{E}}\, \tilde{F}_n^{*-1}(q)$	-2.33	-1.66	-1.29	1.28	1.63	2.28
$\widehat{\mathbf{S}}\, G_{nJ}^{-1}(q)$	0.06	0.02	0.01	0.01	0.02	0.06
$\widehat{\mathbf{S}}\, G_{nB}^{-1}(q)$	0.05	0.02	0.01	0.01	0.02	0.05
$\widehat{\mathbf{S}}\, \tilde{F}_n^{*-1}(q)$	0.06	0.02	0.01	0.01	0.02	0.05

Table 4.2: Simulation 3. Approximations to $\tilde{F}_{nS}$.

$q =$	0.01	0.05	0.10	0.90	0.95	0.99
$\tilde{F}_{nS}^{-1}(q)$	-2.92	-1.94	-1.47	1.18	1.49	2.05
$\Phi^{-1}(q)$	-2.33	-1.64	-1.28	1.28	1.64	2.33
$G_{nS}^{-1}(q)$	-2.68	-1.88	-1.43	1.16	1.45	1.89
$\widehat{\mathbf{E}}\, G_{nSJ}^{-1}(q)$	-2.65	-1.86	-1.42	1.17	1.46	1.93
$\widehat{\mathbf{E}}\, G_{nSB}^{-1}(q)$	-2.60	-1.81	-1.39	1.20	1.50	2.00
$\widehat{\mathbf{E}}\, \tilde{F}_{nS}^{*-1}(q)$	-2.94	-1.92	-1.44	1.21	1.53	2.12
$\widehat{\mathbf{S}}\, G_{nSJ}^{-1}(q)$	0.13	0.10	0.08	0.06	0.09	0.18
$\widehat{\mathbf{S}}\, G_{nSB}^{-1}(q)$	0.13	0.10	0.07	0.05	0.08	0.17
$\widehat{\mathbf{S}}\, \tilde{F}_{nS}^{*-1}(q)$	0.30	0.16	0.11	0.04	0.06	0.12

Table 4.3: Simulation 3. Parameters α and κ, and their estimates.

α	κ		$\hat{\alpha}_J$	$\hat{\alpha}_B$	$\hat{\kappa}_J$	$\hat{\kappa}_B$
2.04	-0.25	BIAS	-0.25	-0.25	0.07	-0.07
		SE	0.51	0.50	0.09	0.08

Table 4.3, although the quality of both estimates of κ is almost the same. Note that the quality of $\tilde{F}_n^*$ is similar to the quality of G_{nB}.

Table 4.2 shows that G_{nS} significantly improves Φ. Here G_{nSJ} and G_{nSB} estimate G_{nS} analogously as in the previous case, but now their biases and variabilities are larger. Nonetheless, both G_{nSJ} and G_{nSB} seem more efficient than Φ. The last approximation $\tilde{F}_{nS}^*$ is the most unbiased on the left tail of $\tilde{F}_{nS}$, but here its SE is the highest one compared to the other empirical approximations. On the right tail the situation is converse.

Simulation 4 Consider the trimmed mean $M_{0.2;0.8}$, see Example 1. A population $\mathcal{X}$ of size $N = 150$ was simulated from the exponential distribution $\mathcal{E}(0.5)$. Our chosen population $\mathcal{X}$ has the mean 1.99 and variance 3.9. The sample size is

$n = 45$.

Table 4.4: Simulation 4. Approximations to $\tilde{F}_n$.

$q =$	0.01	0.05	0.10	0.90	0.95	0.99
$\tilde{F}_n^{-1}(q)$	-2.09	-1.55	-1.24	1.31	1.73	2.55
$\Phi^{-1}(q)$	-2.33	-1.64	-1.28	1.28	1.64	2.33
$G_n^{-1}(q)$	-2.10	-1.57	-1.25	1.32	1.74	2.54
$\widehat{\mathbf{E}}\, G_{nJ}^{-1}(q)$	-2.15	-1.58	-1.26	1.33	1.74	2.50
$\widehat{\mathbf{E}}\, G_{nB}^{-1}(q)$	-2.10	-1.57	-1.25	1.32	1.74	2.54
$\widehat{\mathbf{E}}\, \tilde{F}_n^{*-1}(q)$	-2.09	-1.55	-1.24	1.31	1.73	2.56
$\widehat{\mathbf{S}}\, G_{nJ}^{-1}(q)$	0.19	0.07	0.03	0.13	0.15	0.19
$\widehat{\mathbf{S}}\, G_{nB}^{-1}(q)$	0.09	0.03	0.01	0.02	0.05	0.08
$\widehat{\mathbf{S}}\, \tilde{F}_n^{*-1}(q)$	0.09	0.04	0.02	0.01	0.04	0.11

Table 4.5: Simulation 4. Approximations to $\tilde{F}_{nS}$.

$q =$	0.01	0.05	0.10	0.90	0.95	0.99
$\tilde{F}_{nS}^{-1}(q)$	-2.88	-1.98	-1.50	1.17	1.50	2.13
$\Phi^{-1}(q)$	-2.33	-1.64	-1.28	1.28	1.64	2.33
$G_{nS}^{-1}(q)$	-2.65	-1.87	-1.45	1.14	1.43	1.90
$\widehat{\mathbf{E}}\, G_{nSJ}^{-1}(q)$	-2.59	-1.84	-1.43	1.15	1.46	1.98
$\widehat{\mathbf{E}}\, G_{nSB}^{-1}(q)$	-2.65	-1.87	-1.45	1.13	1.43	1.90
$\widehat{\mathbf{E}}\, \tilde{F}_{nS}^{*-1}(q)$	-2.96	-1.98	-1.50	1.16	1.49	2.12
$\widehat{\mathbf{S}}\, G_{nSJ}^{-1}(q)$	0.20	0.19	0.18	0.15	0.20	0.32
$\widehat{\mathbf{S}}\, G_{nSB}^{-1}(q)$	0.11	0.09	0.07	0.06	0.08	0.17
$\widehat{\mathbf{S}}\, \tilde{F}_{nS}^{*-1}(q)$	0.32	0.16	0.09	0.05	0.07	0.12

Table 4.6: Simulation 4. Parameters α and κ, and their estimates.

α	κ		$\hat{\alpha}_J$	$\hat{\alpha}_B$	$\hat{\kappa}_J$	$\hat{\kappa}_B$
0.34	0.53	BIAS	-0.04	0.03	0.01	0.01
		SE	0.17	0.16	1.04	0.24

Table 4.4 shows that G_n not only improves Φ, it also approximates $\tilde{F}_n$ quite accurate. G_{nJ} is a much more biased estimate of G_n for 0.01 and 0.99 quantiles compared to G_{nB}, and its SE is larger. It can be explained by large SE of $\hat{\kappa}_J$, see Table 4.6. Thus, taking the variability into account, G_{nB} is more efficient than Φ and G_{nJ}. Next, $\tilde{F}_n^*$ approximates $\tilde{F}_n$ similarly as G_{nB}.

Table 4.5 shows that G_{nS} outperforms Φ. Now G_{nSJ} is a little more biased estimate of G_{nS} as compared to the previous case, and also has large SE. Thus, the approximation G_{nSB} is reasonable again. $\tilde{F}_{nS}^*$ is the mostly unbiased estimate of $\tilde{F}_{nS}$. Its drawback is large SE on the left tail.

Simulation 5 Let us modify Simulation 4 by taking the finite combination of sample quantiles (Example 7) as follows

$$n^{-1}(X_{[0.3n]+1:n} + 2X_{[0.4n]+1:n} + 3X_{[0.5n]+1:n} + 2X_{[0.6n]+1:n} + X_{[0.7n]+1:n})/9. \quad (4.12)$$

Table 4.7: Simulation 5. Approximations to $\tilde{F}_n$.

$q =$	0.01	0.05	0.10	0.90	0.95	0.99
$\tilde{F}_n^{-1}(q)$	-2.13	-1.55	-1.25	1.30	1.73	2.51
$\Phi^{-1}(q)$	-2.33	-1.64	-1.28	1.28	1.64	2.33
$G_n^{-1}(q)$	-2.14	-1.58	-1.26	1.31	1.72	2.51
$\widehat{\mathbf{E}}\, G_{nJ}^{-1}(q)$	-1.94	-1.45	-1.13	1.42	1.72	2.28
$\widehat{\mathbf{E}}\, G_{nB}^{-1}(q)$	-2.13	-1.57	-1.26	1.32	1.74	2.52
$\widehat{\mathbf{E}}\, \tilde{F}_n^{*-1}(q)$	-2.09	-1.55	-1.24	1.30	1.72	2.53
$\widehat{\mathbf{S}}\, G_{nJ}^{-1}(q)$	1.57	1.33	1.21	0.88	0.96	1.12
$\widehat{\mathbf{S}}\, G_{nB}^{-1}(q)$	0.12	0.04	0.02	0.03	0.07	0.12
$\widehat{\mathbf{S}}\, \tilde{F}_n^{*-1}(q)$	0.12	0.07	0.05	0.05	0.07	0.18

Table 4.8: Simulation 5. Approximations to $\tilde{F}_{nS}$.

$q =$	0.01	0.05	0.10	0.90	0.95	0.99
$\tilde{F}_{nS}^{-1}(q)$	-4.46	-2.47	-1.76	1.40	2.02	3.59
$\Phi^{-1}(q)$	-2.33	-1.64	-1.28	1.28	1.64	2.33
$G_{nS}^{-1}(q)$	-2.58	-1.82	-1.41	1.17	1.48	2.01
$\widehat{\mathbf{E}}\, G_{nSJ}^{-1}(q)$	-2.24	-1.80	-1.53	0.89	1.16	1.59
$\widehat{\mathbf{E}}\, G_{nSB}^{-1}(q)$	-2.59	-1.83	-1.43	1.16	1.46	1.97
$\widehat{\mathbf{S}}\, G_{nSJ}^{-1}(q)$	1.39	1.22	1.17	1.56	1.68	1.93
$\widehat{\mathbf{S}}\, G_{nSB}^{-1}(q)$	0.15	0.12	0.10	0.08	0.11	0.22

Table 4.9: Simulation 5. Parameters α and κ, and their estimates.

α	κ		$\hat{\alpha}_J$	$\hat{\alpha}_B$	$\hat{\kappa}_J$	$\hat{\kappa}_B$
0.12	0.46	BIAS	-0.02	0.03	38.54	0.05
		SE	0.19	0.10	167.36	0.35

Table 4.7 shows a similar efficiency of approximations G_n, G_{nB} and $\tilde{F}_n^*$ as in Table 4.4, i.e., now only the quality of empirical approximations G_{nB} and $\tilde{F}_n^*$ is slightly lower. Here the empirical approximation G_{nJ} evidently fails, since the estimate $\hat{\kappa}_J$ of κ is very unstable, see Table 4.9.

The distribution $\tilde{F}_{nS}$ has comparatively heavy tails, see Table 4.8. Now all its Edgeworth approximations seem to be untrustworthy, i.e., they do not mimic $\tilde{F}_{nS}$

as in the previous simulation examples. Moreover, the bootstrap approximation $\tilde{F}^*_{nS}$ is omitted in Table 4.8, since it is undefined, because for a part of resamples drawn from empirical population (3.3), the jackknife estimate of variance of (4.12) is exactly 0. It is evident from (3.2).

4.1.5 Discussions

It is evident from the simulation examples of Section 4.1.4 that a positive effect of the second-order approximations is most observable for Studentized L-statistics, and in the case, where the distribution of the population is asymmetric. Clearly, here the sample size is also important. Thus, the chosen moderate sample size $n = 45$ ensures a closeness of distribution of L-statistic to the normal distribution in the first case of Simulation 3 only.

The L-statistics, considered in Simulations 3–5, have different smoothness properties, in the sense of smoothness of the weight function $J\colon (0,1) \to \mathbb{R}$. Thus, for a smooth L-statistic of Simulation 3, we conclude that Edgeworth expansions and empirical Edgeworth expansions with jackknife estimates of the parameters are efficient. In Simulations 4 and 5, the L-statistics are not smooth. In these cases (except the case of a Studentized statistic of Simulation 5), Edgeworth expansions are also much more efficient than the normal approximation. However, it is not necessarily the case for empirical Edgeworth expansions. In particular, for the trimmed mean of Simulation 4, the jackknife estimate of κ is of a poorer quality and its quality is very low for a much more not smooth L-statistic of Simulation 5. For empirical Edgeworth approximations with bootstrap estimates of the parameters, we have noticed only one drawback, i.e., they are biased in Simulation 3. But they are quite unbiased and stable, and thus efficient in Simulation 4 and in the first case of Simulation 5. The last non-parametric bootstrap approximations behave similarly in some cases, and they can be very unbiased for Studentized L-statistics, but here their variability is higher.

Less successful results of empirical Edgeworth approximations with jackknife estimates of the parameters, in the case of not smooth statistics, and inaccuracy of all Edgeworth approximations, in the case of a Studentized L-statistic of Simulation 5, can be clarified. There are at least two possible theoretical reasons, which are familiar from the case of i.i.d. observations, but still not well explored for samples drawn without replacement. The first one is that, in the cases where L-statistics are not smooth (or less smooth), typically, the validity of Edgeworth expansions is ensured by imposing additional smoothness conditions on the distribution function of the underlying population, see, e.g., Gribkova and Helmers [28, 30] and Alberink et al. [2]. The finite population, even assuming that it is obtained from a smooth superpopulation, is not necessarily sufficiently smooth if its size N is relatively small. The second reason, why the approximations are so complicated, particularly in Simulation 5, is the well-known phenomenon of inconsistency of the classical jackknife variance estimator applied to the sample quantile, see, e.g., Martin [49], where it is also shown that the asymptotic distribution of the Studentized quantile is nonnormal (heavy-tailed). It seems that the jackknife variance estimator also fails in the case of the finite combination of the sample quantiles of Simulation 5. We also refer to Shao and Wu [65], where an alternative delete-d jackknife is proposed, with the number d (of deleted observations in the jackknife) depending on smoothness of a statistic.

4.2 Approximations to distributions of quantiles in stratified samples

4.2.1 Hoeffding decomposition and approximations

Consider a population $\mathcal{X} = \{x_1, \dots, x_N\}$ of size N. We assume, without loss of generality, that $x_1 \le \cdots \le x_N$. Let $\mathcal{X}$ be divided into $h \ge 1$ nonoverlapping strata $\mathcal{X} = \mathcal{X}_1' \cup \cdots \cup \mathcal{X}_h'$, where $\mathcal{X}_k' = \{x_{k,1}, \dots, x_{k,N_k}\}$, $1 \le k \le h$. Evidently, $N = N_1 + \cdots + N_h$. For convenience, we also assume here that $x_{k,1} \le \cdots \le x_{k,N_k}$. Let $\mathbb{X}_k' = \{X_{k,1}, \dots, X_{k,n_k}\}$ be a simple random sample of size $n_k < N_k$ drawn without replacement from the stratum $\mathcal{X}_k'$. We assume that the samples $\mathbb{X}_1', \dots, \mathbb{X}_h'$ are independent. Write $\mathbb{X}' = \mathbb{X}_1' \cup \cdots \cup \mathbb{X}_h'$ and denote $n = n_1 + \cdots + n_h$. Denote the distribution function of the stratum k and its empirical analogue as follows:

$$F_{N,k}(x) = \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbb{I}\{x_{k,i} \le x\} \tag{4.13}$$

and

$$\widehat{F}_{n,k}(x) = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbb{I}\{X_{k,i} \le x\}, \tag{4.14}$$

respectively. Then the distribution function of the population $\mathcal{X}$ and its estimator are

$$F(x) = \sum_{k=1}^{h} \frac{N_k}{N} F_{N,k}(x)$$

(it is the same as (2.26)) and

$$\widehat{F}_n(x) = \sum_{k=1}^{h} \frac{N_k}{N} \widehat{F}_{n,k}(x),$$

respectively. Consider the population β-quantile, $0 < \beta < 1$, defined as follows: $F^{-1}(\beta) = \inf\{x : F(x) \geq \beta\}$. Define its estimator

$$X_\beta = \widehat{F}_n^{-1}(\beta) = \inf\{x : \widehat{F}_n(x) \geq \beta\}.$$

Denote $\sigma_\beta^2 = \mathbf{Var}\, X_\beta$. We are interested in approximations to the distribution function

$$F_\beta(x) = \mathbf{P}\{X_\beta - \mathbf{E}\, X_\beta \leq x\sigma_\beta\}.$$

The asymptotic normality of the quantile X_β, under a stratified simple random sampling without replacement, was considered by Shao [64], see also Gross [31]. Here we present an Edgeworth type approximation to $F_\beta(x)$, see (4.15) below, and its empirical analogue based on the bootstrap of Booth et al. [21]. Our main tool is again Hoeffding's decomposition

$$X_\beta = \mathbf{E}\, X_\beta + L + Q + R,$$

constructed by Bloznelis [15] for general symmetric statistics in the case of stratified simple random samples drawn without replacement. Here, like in the one-stratum case, L and Q are called linear and quadratic parts of the decomposition, and R is a remainder term. In the case of U-statistics, where $R \equiv 0$, Edgeworth expansions were constructed and their second-order correctness was shown by Bloznelis [18]. Thus, we expect that, if R is negligible, those Edgeworth expansions will also approximate $F_\beta(x)$ well. In particular, we suggest to approximate $F_\beta(x)$ by

$$H_\beta(x) = \Phi(x) - \frac{\alpha_\beta + 3\kappa_\beta}{6\sigma_\beta^3}\Phi'(x)(x^2-1), \tag{4.15}$$

obtained in [18]. Here

$$\alpha_\beta = \sum_{k=1}^{h}(1 - 2n_k/N_k)\tau_k^2\alpha_k \quad \text{and} \quad \kappa_\beta = \sum_{k=1}^{h}\tau_k^4\kappa_{kk} + 2\sum_{1\leq k<u\leq h}\tau_k^2\tau_u^2\kappa_{ku},$$

with $\tau_k^2 = n_k(1 - n_k/N_k)$. Here the moments

$$\begin{aligned}
\alpha_k &= \frac{1}{N_k}\sum_{s=1}^{N_k} g_k^3(x_{k,s}),\\
\kappa_{kk} &= \binom{N_k}{2}^{-1}\sum_{1\leq s<r\leq N_k}\psi_k(x_{k,s}, x_{k,r})g_k(x_{k,s})g_k(x_{k,r}),\\
\kappa_{ku} &= \frac{1}{N_kN_u}\sum_{1\leq s\leq N_k,\,1\leq r\leq N_u}\psi_{ku}(x_{k,s}, x_{u,r})g_k(x_{k,s})g_u(x_{u,r}),
\end{aligned}$$

established in [18], are based on the functions

$$g_k(x_{k,s}) = \frac{N_k - 1}{N_k - n_k} \sum_{i=1}^{N-1} \Big(p_i(x_{k,s}) - p_i \Big) \triangle_i, \tag{4.16}$$

$$\begin{aligned} \psi_k(x_{k,s}, x_{k,r}) = \frac{N_k - 2}{N_k - n_k} \frac{N_k - 3}{N_k - n_k - 1} \sum_{i=1}^{N-1} \Big(& p_i(x_{k,s}, x_{k,r}) \\ & - \frac{N_k - 1}{N_k - 2} \Big(p_i(x_{k,s}) + p_i(x_{k,r}) \Big) + \frac{N_k}{N_k - 2} p_i \Big) \triangle_i, \end{aligned} \tag{4.17}$$

$$\begin{aligned} \psi_{ku}(x_{k,s}, x_{u,r}) = \frac{N_k - 1}{N_k - n_k} \frac{N_u - 1}{N_u - n_u} \sum_{i=1}^{N-1} \Big(& p_i(x_{k,s}, x_{u,r}) - p_i(x_{k,s}) \\ & - p_i(x_{u,r}) + p_i \Big) \triangle_i, \end{aligned} \tag{4.18}$$

where, for $1 \le i \le N - 1$, we denote the probabilities

$$p_i = \mathbf{P}\{X_\beta > x_i\},$$

$$p_i(x_{k,s}) = \mathbf{P}\left\{ X_\beta > x_i \,|\, X_{k,1} = x_{k,s} \right\},$$

$$p_i(x_{k,s}, x_{k,r}) = \mathbf{P}\left\{ X_\beta > x_i \,|\, X_{k,1} = x_{k,s}, X_{k,2} = x_{k,r} \right\},$$

$$p_i(x_{k,s}, x_{u,r}) = \mathbf{P}\left\{ X_\beta > x_i \,|\, X_{k,1} = x_{k,s}, X_{u,1} = x_{u,r} \right\}.$$

We give these probabilities in (4.19) and in Proposition 21 below. Note that expressions (4.16)–(4.18) are obtained directly from (11) in [15], using the definitions of expectation and conditional expectations, and applying the summation by parts formula $\sum_{i=1}^{N} (p_{i-1} - p_i) x_i = -p_N x_N + p_0 x_1 + \sum_{i=1}^{N-1} p_i \triangle_i$ (in the case of expectation) and noting that, by definition, $p_N = 0$ and $p_0 = 1$, and so forth.

Let $\mathcal{T}$ be the set of h-tuples $(t_1, \dots, t_h) \in \{0, \dots, n_1\} \times \cdots \times \{0, \dots, n_h\}$, which satisfy the condition $\sum_{j=1}^{h} w_j t_j < \beta$. Here $w_j = N_j/(N n_j)$. Denote $d_{ij} := N_j F_{N,j}(x_i)$. It is shown in Gross [31] that, for $0 \le i \le N$,

$$p_i = \sum_{\mathcal{T}} \prod_{1 \le j \le h} \mathcal{H}_{N_j, n_j, d_{ij}}(t_j), \tag{4.19}$$

and then the variance of X_β in (4.15) is

$$\sigma_\beta^2 = \sum_{i=1}^{N} (p_{i-1} - p_i) x_i^2 - \left(\sum_{i=1}^{N} (p_{i-1} - p_i) x_i \right)^2. \tag{4.20}$$

Next, we give explicit expressions of the conditional probabilities.

Proposition 21 *Let $1 \leq i \leq N-1$.*

(i) For $1 \leq k \leq h$ and $1 \leq s \leq N_k$, we have

$$p_i(x_{k,s}) = \sum_{\mathcal{T}} \varphi_{k,s}(i) \prod_{1\leq j\leq h,\, j\neq k} \mathcal{H}_{N_j,n_j,d_{ij}}(t_j),$$

where

$$\varphi_{k,s}(i) = \begin{cases} \mathcal{H}_{N_k-1,n_k-1,d_{ik}}(t_k) & \text{if } i \in \mathcal{I}_{21}, \\ \mathcal{H}_{N_k-1,n_k-1,d_{ik}-1}(t_k-1) & \text{if } i \in \mathcal{I}_{22}, \end{cases}$$

with

$$\mathcal{I}_{21} = \{i : x_i < x_{k,s}\}, \quad \mathcal{I}_{22} = \{i : x_i \geq x_{k,s}\}.$$

(ii) For $1 \leq k \leq h$ and $1 \leq s < r \leq N_k$, we have

$$p_i(x_{k,s}, x_{k,r}) = \sum_{\mathcal{T}} \phi_{k,s;k,r}(i) \prod_{1\leq j\leq h,\, j\neq k} \mathcal{H}_{N_j,n_j,d_{ij}}(t_j),$$

where

$$\phi_{k,s;k,r}(i) = \begin{cases} \mathcal{H}_{N_k-2,n_k-2,d_{ik}}(t_k) & \text{if } i \in \mathcal{I}_{31}, \\ \mathcal{H}_{N_k-2,n_k-2,d_{ik}-1}(t_k-1) & \text{if } i \in \mathcal{I}_{32}, \\ \mathcal{H}_{N_k-2,n_k-2,d_{ik}-2}(t_k-2) & \text{if } i \in \mathcal{I}_{33}, \end{cases}$$

with

$$\mathcal{I}_{31} = \{i : x_i < x_{k,s} \leq x_{k,r}\}, \quad \mathcal{I}_{32} = \{i : x_{k,s} \leq x_i < x_{k,r}\},$$
$$\mathcal{I}_{33} = \{i : x_{k,s} \leq x_{k,r} \leq x_i\}.$$

(iii) For $1 \leq k < u \leq h$ and $1 \leq s \leq N_k$, $1 \leq r \leq N_u$, we have

$$p_i(x_{k,s}, x_{u,r}) = \sum_{\mathcal{T}} \theta_{k,s;u,r}(i) \prod_{1\leq j\leq h,\, j\neq k,u} \mathcal{H}_{N_j,n_j,d_{ij}}(t_j),$$

where

$$\theta_{k,s;u,r}(i) = \begin{cases} \mathcal{H}_{N_k-1,n_k-1,d_{ik}}(t_k)\, \mathcal{H}_{N_u-1,n_u-1,d_{iu}}(t_u) & \text{if } i \in \mathcal{I}_{41}, \\ \mathcal{H}_{N_k-1,n_k-1,d_{ik}-1}(t_k-1)\, \mathcal{H}_{N_u-1,n_u-1,d_{iu}}(t_u) & \text{if } i \in \mathcal{I}_{42}, \\ \mathcal{H}_{N_k-1,n_k-1,d_{ik}}(t_k)\, \mathcal{H}_{N_u-1,n_u-1,d_{iu}-1}(t_u-1) & \text{if } i \in \mathcal{I}_{43}, \\ \mathcal{H}_{N_k-1,n_k-1,d_{ik}-1}(t_k-1)\, \mathcal{H}_{N_u-1,n_u-1,d_{iu}-1}(t_u-1) & \text{if } i \in \mathcal{I}_{44}, \end{cases}$$

with

$$\mathcal{I}_{41} = \{i : x_i < x_{k,s},\, x_i < x_{u,r}\}, \quad \mathcal{I}_{42} = \{i : x_i \geq x_{k,s},\, x_i < x_{u,r}\},$$
$$\mathcal{I}_{43} = \{i : x_i < x_{k,s},\, x_i \geq x_{u,r}\}, \quad \mathcal{I}_{44} = \{i : x_i \geq x_{k,s},\, x_i \geq x_{u,r}\}.$$

Proof. Calculations of all the conditional probabilities are based on the same arguments as the derivation of (4.19) in Gross [31]. Here, for each case from *(i)*–*(iii)*, we need to consider, under fixed conditions of the conditional probabilities, a few different positions of x_i only. Note that the set $\mathcal{T}$ is the same for all probabilities, since we use conventions (1.6). $\square$

Empirical approximation. The parameters $\alpha_\beta = \alpha_\beta(\mathcal{X})$, $\kappa_\beta = \kappa_\beta(\mathcal{X})$ and $\sigma_\beta^2 = \sigma_\beta^2(\mathcal{X})$ that define approximation (4.15) are usually unknown characteristics of the population $\mathcal{X}$. Thus, they should be estimated in practice. In Gross [31], for the estimation of the parameter σ_β^2, a convenient plug-in rule was proposed, where strata distribution functions (4.13) were replaced by their corresponding empirical versions (4.14). However, it is not convenient for the estimation of α_β and κ_β. Another way is to replace the population parameters by their jackknife estimators, see Bloznelis [18]. But it is well known that, in the case of sample quantiles, jackknife estimators (of variance) often fail. Recall also the discussion at the end of Section 4.1.5.

We consider here the bootstrap estimators of the parameters. For $1 \leq k \leq h$ write $N_k = m_k n_k + l_k$, where $0 \leq l_k < n_k$. For each $1 \leq k \leq h$ we construct an empirical stratum $\tilde{\mathcal{X}}_k'$, as in the one-stratum case, see (3.3). Then $\tilde{\mathcal{X}}' = \tilde{\mathcal{X}}_1' \cup \cdots \cup \tilde{\mathcal{X}}_h'$ is an empirical population, and the bootstrap estimator of the population parameter $\theta = \theta(\mathcal{X})$ is

$$\hat{\theta}_B = \mathbf{E}\left(\theta(\tilde{\mathcal{X}}') \,\middle|\, \mathbb{X}'\right), \tag{4.21}$$

similarly as in (3.4). Thus, we have the bootstrap estimators $\hat{\alpha}_{\beta B}$, $\hat{\kappa}_{\beta B}$ and $\hat{\sigma}_{\beta B}^2$ of α_β, κ_β and σ_β^2. However, it is difficult to obtain their explicit expresions. Therefore, we apply Monte–Carlo (M–C) approximations to the parameters we are interested in. In particular, let $\tilde{\mathcal{X}}_{(1)}', \ldots, \tilde{\mathcal{X}}_{(B)}'$ be B empirical populations constructed independently as described above, i.e., we randomly and with replacement select B empirical populations from all possible $\prod_{k=1}^h \binom{n_k}{l_k}$. Then M–C approximation to (4.21) is

$$\tilde{\theta}_B = \frac{1}{B}\sum_{b=1}^{B} \theta(\tilde{\mathcal{X}}_{(b)}'). \tag{4.22}$$

Finally, replacing the true parameters α_β, κ_β and σ_β^2 in (4.15) by their estimates

$\tilde{\alpha}_{\beta B}$, $\tilde{\kappa}_{\beta B}$ and $\tilde{\sigma}^2_{\beta B}$, we obtain the empirical approximation

$$\tilde{H}_\beta(x) = \Phi(x) - \frac{\tilde{\alpha}_{\beta B} + 3\tilde{\kappa}_{\beta B}}{6\tilde{\sigma}^3_{\beta B}} \Phi'(x)(x^2 - 1) \tag{4.23}$$

to $F_\beta(x)$.

4.2.2 Numerical simulations

In the simulation example below, we approximate the exact distribution F_β similarly as in the case of one-stratum, see Appendix A.1, i.e., we obtain its approximation $\tilde{F}_\beta$ by the M–C simulations, by drawing independently 10^5 stratified samples from $\mathcal{X}$. Now the variance of X_β is given by (4.20) and the expectation is $\mu_\beta = \sum_{i=1}^{N} (p_{i-1} - p_i) x_i$.

In the tables below, we present q-quantiles, $q = 0.01, 0.05, 0.10, 0.90, 0.95, 0.99$, of $\tilde{F}_\beta$, Φ, H_β, and $\tilde{H}_\beta$. For the approximation $\tilde{H}_\beta$ we present two characteristics for each of the empirical q-quantiles: estimated value $\widehat{\mathbf{E}}\,\tilde{H}_\beta^{-1}(q)$ of its expectation $\mathbf{E}\,\tilde{H}_\beta^{-1}(q)$ and estimated value $\widehat{\mathbf{S}}\,\tilde{H}_\beta^{-1}(q)$ of its standard error $\mathbf{S}\,\tilde{H}_\beta^{-1}(q)$, based on $R = 100$ stratified samples drawn independently from $\mathcal{X}$, see formulas (4.10) and (4.11). To estimate the parameters α_β, κ_β and σ^2_β by (4.22), we take $B = 30$.

Simulation 6 We consider the case of a sample median, i.e., we take $\beta = 0.5$. From the real finite population, which consists of Lithuanian service enterprises, we take three completely sampled strata, which belong to the economic activity classified as 'combined facilities support activities'. The strata sizes are $N_1 = 25$, $N_2 = 7$ and $N_3 = 13$. Using the measurements of turnover and the number of persons employed, we form two different populations: $\mathcal{X}_{(1)} = \mathcal{X}'_{(1)1} \cup \mathcal{X}'_{(1)2} \cup \mathcal{X}'_{(1)3}$ and $\mathcal{X}_{(2)} = \mathcal{X}'_{(2)1} \cup \mathcal{X}'_{(2)2} \cup \mathcal{X}'_{(2)3}$. Here we use the first-quarter data of 2011. The simulation results for these populations are presented in Tables 4.10 and 4.11, respectively. We choose sample sizes $n_1 = 10$, $n_2 = 3$ and $n_3 = 5$.

Table 4.10: The population $\mathcal{X}_{(1)}$. Approximations to $\tilde{F}_{0.5}$.

$q =$	0.01	0.05	0.10	0.90	0.95	0.99
$\tilde{F}_{0.5}^{-1}(q)$	-2.85	-2.01	-1.94	0.96	0.96	0.96
$\Phi^{-1}(q)$	-2.33	-1.64	-1.28	1.28	1.64	2.33
$H_{0.5}^{-1}(q)$	-2.82	-1.95	-1.41	1.21	1.46	1.81
$\widehat{\mathbf{E}}\,\tilde{H}_{0.5}^{-1}(q)$	-2.21	-1.64	-1.28	1.32	1.74	2.42
$\widehat{\mathbf{S}}\,\tilde{H}_{0.5}^{-1}(q)$	0.42	0.20	0.08	0.09	0.21	0.42

Table 4.10 shows that $H_{0.5}$ significantly improves Φ. However, it is not the case for its empirical version $\tilde{H}_{0.5}$, since, for a large part of the samples, this approximation to $\tilde{F}_{0.5}$ is less accurate than Φ. Table 4.11 shows that $H_{0.5}$ evidently

Table 4.11: The population $\mathcal{X}_{(2)}$. Approximations to $\tilde{F}_{0.5}$.

$q =$	0.01	0.05	0.10	0.90	0.95	0.99
$\tilde{F}_{0.5}^{-1}(q)$	-1.60	-1.26	-0.82	1.90	1.99	3.85
$\Phi^{-1}(q)$	-2.33	-1.64	-1.28	1.28	1.64	2.33
$H_{0.5}^{-1}(q)$	-1.84	-1.47	-1.22	1.40	1.92	2.79
$\widehat{\mathbf{E}}\,\tilde{H}_{0.5}^{-1}(q)$	-2.00	-1.53	-1.24	1.35	1.81	2.63
$\widehat{\mathbf{S}}\,\tilde{H}_{0.5}^{-1}(q)$	0.17	0.06	0.02	0.04	0.10	0.16

outperforms Φ. Here the estimated variability $\widehat{\mathbf{S}}\,\tilde{H}_{0.5}^{-1}(q)$ is comparatively small, therefore $\tilde{H}_{0.5}$ is also more efficient than Φ.

We stress that the proposed approximations may be very efficient in real surveys, where we need to measure the accuracy of a sample quantile in small domains of a population (for some collections of strata) and where populations are highly skewed.

Appendix A

Monte–Carlo approximations

A.1 Approximations to distributions

Here we present Monte–Carlo approximations to distribution functions $F_n(x)$ and $F_{nS}(x)$ given by (5) and (4.1), respectively. We draw independently C samples $\mathbb{X}^{(c)} = \{X_1^{(c)}, \ldots, X_n^{(c)}\}$, $1 \le c \le C$ of size n without replacement from $\mathcal{X}$, and take

$$\tilde{F}_n(x) = \frac{1}{C} \sum_{c=1}^{C} \mathbb{I}\Big\{L_n(\mathbb{X}^{(c)}) - \mu_L(\mathcal{X}) \le x\sigma_L(\mathcal{X})\Big\}$$

and

$$\tilde{F}_{nS}(x) = \frac{1}{C} \sum_{c=1}^{C} \mathbb{I}\Big\{L_n(\mathbb{X}^{(c)}) - \mu_L(\mathcal{X}) \le xS(L_n(\mathbb{X}^{(c)}))\Big\}.$$

Here the population $\mathcal{X}$ characteristics $\mu_L(\mathcal{X}) = \mathbf{E}\, L_n$ and $\sigma_L^2(\mathcal{X}) = \mathbf{Var}\, L_n$ are expressed by

$$\mu_L(\mathcal{X}) = \frac{1}{n} \sum_{p=1}^{n} c_p\, \mathbf{E}\, X_{p:n}$$

and (3.5), respectively, where the moments of the order statistics (inside of these expressions) are

$$\mathbf{Var}\, X_{p:n} = \binom{N}{n}^{-1} \sum_{i=1}^{N} \binom{i-1}{p-1} \binom{N-i}{n-p} x_i^2 - \left(\mathbf{E}\, X_{p:n}\right)^2, \qquad 1 \le p \le n \tag{A.1}$$

and

$$\begin{aligned} \mathbf{Cov}(X_{p:n}, X_{r:n}) = &\binom{N}{n}^{-1} \sum_{1 \le i < j \le N} \binom{i-1}{p-1} \binom{j-i-1}{r-p-1} \binom{N-j}{n-r} x_i x_j \\ &- \mathbf{E}\, X_{p:n}\, \mathbf{E}\, X_{r:n}, \qquad 1 \le p < r \le n, \end{aligned} \tag{A.2}$$

with

$$\mathbf{E}\, X_{p:n} = \binom{N}{n}^{-1} \sum_{i=1}^{N} \binom{i-1}{p-1}\binom{N-i}{n-p} x_i, \qquad 1 \le p \le n. \tag{A.3}$$

If it is assumed that $x_1 < \cdots < x_N$, the proof of (A.1)–(A.3) is simple. To obtain these formulas in the case of $x_1 \le \cdots \le x_N$, apply Lemma 8.

A.2 Approximations to bootstrap distributions

Here we present Monte–Carlo approximations to distribution functions $F_n^*(x)$ and $F_{nS}^*(x)$ given by (4.7) and (4.8), respectively. Given the sample $\mathbb{X}$ of size n, we construct independently B empirical populations $\tilde{\mathcal{X}}^{(b)}$, $1 \le b \le B$. Next, for every $1 \le b \le B$, we draw independently R resamples $\tilde{\mathbb{X}}^{(b,r)} = \{\tilde{X}_1^{(b,r)}, \ldots, \tilde{X}_n^{(b,r)}\}$, $1 \le r \le R$ of size n without replacement from $\tilde{\mathcal{X}}^{(b)}$, and take

$$\tilde{F}_n^*(x) = \frac{1}{BR}\sum_{b=1}^{B}\sum_{r=1}^{R} \mathbb{I}\Big\{L_n(\tilde{\mathbb{X}}^{(b,r)}) - \mu_L(\tilde{\mathcal{X}}^{(b)}) \le x\sigma_L(\tilde{\mathcal{X}}^{(b)})\Big\}$$

and

$$\tilde{F}_{nS}^*(x) = \frac{1}{BR}\sum_{b=1}^{B}\sum_{r=1}^{R} \mathbb{I}\Big\{L_n(\tilde{\mathbb{X}}^{(b,r)}) - \mu_L(\tilde{\mathcal{X}}^{(b)}) \le xS(L_n(\tilde{\mathbb{X}}^{(b,r)}))\Big\}.$$

Here the expressions of $\mu_L(\cdot)$ and $\sigma_L^2(\cdot)$ are given in Appendix A.1.

Bibliography

[1] G. Afendras, N. Papadatos, and V. Papathanasiou. The discrete Mohr and Noll inequality with applications to variance bounds. *Sankhyā*, 69:162–189, 2007.

[2] I. Alberink, G. Pap, and M. van Zuijlen. Edgeworth expansions for L-statistics. *Prob. Math. Statist.*, 21:277–302, 2001.

[3] A. Aleškevičienė. Large and moderate deviations for L-statistics. *Lith. Math. J.*, 31:145–156, 1991.

[4] B. Arnold, N. Balakrishnan, and H. Nagaraja. *A First Course in Order Statistics.* SIAM, Philadelphia, 2008.

[5] G. Babu and K. Singh. Edgeworth expansions for sampling without replacement from finite populations. *J. Multivariate Anal.*, 17:261–278, 1985.

[6] N. Balakrishnan, C. Charalambides, and N. Papadatos. Bounds on expectation of order statistics from a finite population. *J. Statist. Plann. Inference*, 113:569–588, 2003.

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

[8] V. Bentkus and R. Zitikis. Probabilities of large deviations for L-statistics. *Lith. Math. J.*, 30:215–222, 1990.

[9] A. Berred and V. Nevzorov. Samples without replacement: discrete analogs of some continuous distributions. *J. Math. Sci.*, 133:1224–1230, 2006.

[10] A. Berred and V. Nevzorov. Characterizations based on order statistics under sampling without replacement. *J. Statist. Plann. Inference*, 139:547–557, 2009.

[11] A. Bikelis. On the estimation of the remainder term in the central limit theorem for samples from finite populations. *Studia Sci. Math. Hungar.*, 4:345–354, 1969.

[12] S. Bjerve. Error bounds for linear combinations of order statistics. *Ann. Statist.*, 5:357–369, 1977.

[13] M. Bloznelis. One- and two-term Edgeworth expansions for a finite population sample mean. Exact results. I,II. *Lith. Math. J.*, 40:213–227, 329–340, 2000.

[14] M. Bloznelis. Empirical Edgeworth expansion for finite population statistics I,II. *Lith. Math. J.*, 41:120–134, 207–218, 2001.

[15] M. Bloznelis. A note on the bias and consistency of the jackknife variance estimator in stratified samples. *Statistics*, 37:489–504, 2003.

[16] M. Bloznelis. Edgeworth expansions for Studentized versions of symmetric finite population statistics. *Lith. Math. J.*, 43:221–240, 2003.

[17] M. Bloznelis. Bootstrap approximation to distributions of finite population U-statistics. *Acta Appl. Math.*, 96:71–86, 2007.

[18] M. Bloznelis. Second-order and resampling approximation of finite population U-statistics based on stratified samples. *Statistics*, 41:321–332, 2007.

[19] M. Bloznelis. The unfinished manuscript of a book, 2008.

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

[21] J. Booth, R. Butler, and P. Hall. Bootstrap methods for finite populations. *J. Amer. Statist. Assoc.*, 89:1282–1289, 1994.

[22] A. Chatterjee. Asymptotic properties of sample quantiles from a finite population. *Ann. Inst. Statist. Math.*, 63:157–179, 2011.

[23] H. Chernoff, J. Gastwirth, and M. J. Jr. Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Ann. Math. Statist.*, 38:52–72, 1967.

[24] M. Chunsheng. Variance bound of function of order statistics. *Stat. Probab. Lett.*, 13:25–27, 1992.

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

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

[27] F. Gao and X. Zhao. Delta method in large deviations and moderate deviations for estimators. *Ann. Statist.*, 39:1211–1240, 2011.

[28] N. Gribkova and R. Helmers. The empirical Edgeworth expansion for a Studentized trimmed mean. *Math. Methods Statist.*, 15:61–87, 2006.

[29] N. Gribkova and R. Helmers. On the Edgeworth expansion and the M out of N bootstrap accuracy for a Studentized trimmed mean. *Math. Methods Statist.*, 16:142–176, 2007.

[30] N. Gribkova and R. Helmers. Second order approximations for slightly trimmed sums. *Theory Probab. Appl.*, 2011. (to appear); see also `arXiv:1104.3347v1 [math.PR]`.

[31] S. Gross. Median estimation in sample surveys. In *Proceedings of the Section on Survey Research Methods*, pages 181–184. American Statistical Association, 1980.

[32] P. Hall and A. Padmanabhan. On the bootstrap and the trimmed mean. *J. Multivariate Anal.*, 41:132–153, 1992.

[33] R. Helmers. The order of the normal approximation for linear combinations of order statistics with smooth weight functions. *Ann. Probab.*, 5:940–953, 1977.

[34] R. Helmers. Edgeworth expansions for linear combinations of order statistics with smooth weight functions. *Ann. Statist.*, 8:1361–1374, 1980.

[35] R. Helmers. On the Edgeworth expansion and the bootstrap approximation for a Studentized U-statistic. Technical Report Tech. Rep. MS-R8708, CWI, Amsterdam, The Netherlands, 1987.

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

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

[38] T. Höglund. Sampling from a finite population. A remainder term estimate. *Scand. J. Statist.*, 5:69–71, 1978.

[39] J. Huang. Efficient computation of the performance of bootstrap and jackknife estimators of the variance of L-statistics. *J. Statist. Comput. Simul.*, 38:45–56, 1991.

[40] A. Hutson and M. Ernst. The exact bootstrap mean and variance of an L-estimator. *J. Roy. Statist. Soc. B*, 62:89–94, 2000.

[41] K. Jasiński and T. Rychlik. Maximum variance of order statistics from symmetric populations revisited. Submitted, 2010.

[42] M. Jones and N. Balakrishnan. How are moments and moments of spacings related to distribution functions? *J. Statist. Plann. Inference*, 103:377–390, 2002.

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

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

[45] C.-D. Lea and M. Puri. Asymptotic properties of linear functions of order statistics. *J. Statist. Plann. Inference*, 18:203–223, 1988.

[46] F. López-Blázquez and A. Castaño-Martínez. Upper and lower bounds for the correlation ratio of order statistics from a sample without replacement. *J. Statist. Plann. Inference*, 136:43–52, 2006.

[47] Y. Maesono. An Edgeworth expansion and a normalizing transformation for L-statistics. *Bull. Inform. Cybernet.*, 39:25–43, 2007.

[48] Y. Maesono and S. Penev. Edgeworth expansion for the kernel quantile estimator. *Ann. Inst. Statist. Math.*, 63:617–644, 2011.

[49] M. Martin. On using the jackknife to estimate quantile variance. *Canad. J. Statist.*, 18:149–153, 1990.

[50] D. Mason. Asymptotic normality of linear combinations of order statistics with a smooth score function. *Ann. Statist.*, 9:899–908, 1981.

[51] S. Moriguti. Extremal properties of extreme value distributions. *Ann. Math. Statist.*, 22:523–536, 1951.

[52] R. Norvaiša. Laws of large numbers for L-statistics. *J. Appl. Math. Stoch. Anal.*, 7:125–143, 1994.

[53] R. Norvaiša and R. Zitikis. Asymptotic behaviour of linear combinations of functions of order statistics. *J. Statist. Plann. Inference*, 28:305–317, 1991.

[54] N. Papadatos. Maximum variance of order statistics. *Ann. Inst. Statist. Math.*, 47:185–193, 1995.

[55] N. Papadatos. A note on maximum variance of order statistics from symmetric populations. *Ann. Inst. Statist. Math.*, 49:117–121, 1997.

[56] N. Papadatos. Upper bound for the covariance of extreme order statistics from a sample of size three. *Sankhyā, Ser. A*, 61:229–240, 1999.

[57] V. Papathanasiou. Some characterizations of distributions based on order statistics. *Stat. Probab. Lett.*, 9:145–147, 1990.

[58] H. Putter. *Consistency of Resampling Methods.* PhD thesis, Leiden University, 1994.

[59] H. Putter and W. van Zwet. Empirical Edgeworth expansions for symmetric statistics. *Ann. Statist.*, 26:1540–1569, 1998.

[60] J. Rao and C. Wu. Resampling inference with complex survey data. *J. Amer. Statist. Assoc.*, 83:231–241, 1988.

[61] J. Robinson. An asymptotic expansion for samples from a finite population. *Ann. Statist.*, 6:1005–1011, 1978.

[62] T. Rychlik. Extreme variances of order statistics in dependent samples. *Stat. Probab. Lett.*, 78:1577–1582, 2008.

[63] R. Serfling. *Approximation Theorems of Mathematical Statistics.* Wiley, New York, 1980.

[64] J. Shao. L-statistics in complex survey problems. *Ann. Statist.*, 22:946–967, 1994.

[65] J. Shao and C. Wu. A general theory for jackknife variance estimation. *Ann. Statist.*, 17:1176–1197, 1989.

[66] G. Shorack. Functions of order statistics. *Ann. Math. Statist.*, 43:412–427, 1972.

[67] R. Sitter. A resampling procedure for complex survey data. *J. Amer. Statist. Assoc.*, 87:755–765, 1992.

[68] S. Stigler. The asymptotic distribution of the trimmed mean. *Ann. Statist.*, 1:472–477, 1973.

[69] S. Stigler. Linear functions of order statistics with smooth weight functions. *Ann. Statist.*, 2:676–693, 1974.

[70] G. Székely and T. Móri. An extremal property of rectangular distributions. *Stat. Probab. Lett.*, 3:107–109, 1985.

[71] K. Takahasi and M. Futatsuya. Dependence between order statistics in samples from finite population and its application to ranked set sampling. *Ann. Inst. Statist. Math.*, 50:49–70, 1998.

[72] G. Terrell. A characterization of rectangular distributions. *Ann. Probab.*, 11:823–826, 1983.

[73] W. van Zwet. A strong law for linear functions of order statistics. *Ann. Probab.*, 8:986–990, 1980.

[74] W. van Zwet. A Berry-Esseen bound for symmetric statistics. *Z. Wahrsch. Verw. Gebiete*, 66:425–440, 1984.

[75] M. Vandemaele and N. Veraverbeke. Cramér type large deviations for linear combinations of order statistics. *Ann. Probab.*, 10:423–434, 1982.

[76] J. Wellner. A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. *Ann. Statist.*, 5:473–480, 1977.

[77] J. Wellner. A law of the iterated logarithm for functions of order statistics. *Ann. Statist.*, 5:481–494, 1977.

[78] L. Zhao and X. Chen. Normal approximation for finite-population U-statistics. *Acta Math. Appl. Sin.*, 6:263–272, 1990.