

Edgeworth expansions for two-stage sampling with applications to stratified and cluster sampling

Sherzod M. MIRAKHMEDOV¹, Sreenivasa R. JAMMALAMADAKA² and Magnus EKSTRÖM^{3*}

¹Department of Probability and Statistics, Institute of Mathematics, National University of Uzbekistan, Tashkent, Uzbekistan

²Department of Statistics and Applied Probability, University of California, Santa Barbara, U.S.A.

³Department of Statistics, Umeå University, Umeå, Sweden

Key words and phrases: Bootstrap; cluster sampling; confidence interval; Edgeworth expansion; simple random sampling; stratified sampling; two-stage sampling.

MSC 2010: Primary 62D05; secondary 62E20

Abstract: A two-term Edgeworth expansion for the standardized version of the sample total in a two-stage sampling design is derived. In particular, for the commonly used stratified and cluster sampling schemes, formal two-term asymptotic expansions are obtained for the *Studentized* versions of the sample total. These results are applied in conjunction with the bootstrap to construct more accurate confidence intervals for the unknown population total in such sampling schemes. *The Canadian Journal of Statistics* 43: 578–599; 2015 © 2015 The Authors. The Canadian Journal of Statistics published by Wiley Periodicals, Inc. on behalf of Statistical Society of Canada

Résumé: Les auteurs présentent un développement en deux termes d’une série d’Edgeworth pour l’estimateur du total basé sur un plan échantillonnal à deux phases. Ils obtiennent en particulier des développements asymptotiques formels pour la version studentisée du total échantillonnal basé sur un échantillonnage stratifié et sur un échantillonnage en grappes. Les auteurs utilisent ces résultats et des méthodes de rééchantillonnage afin de construire des intervalles de confiance plus précis pour le total de la population dans le contexte de ces plans d’échantillonnage. *La revue canadienne de statistique* 43: 578–599; 2015 © 2015 Les auteurs. La revue canadienne de statistique, publiée par Wiley Periodicals Inc. au nom de la Société statistique du Canada

1. INTRODUCTION

A common sampling strategy for surveying finite populations is to select the sampled units in several stages. Multistage sampling refers to sampling plans, where the selection of units is carried out in stages using smaller and smaller subunits at each stage. For instance, to do a national survey on unemployment or to conduct an opinion poll on a given topic, one may select certain states, then certain counties within those states, cities within these counties, etc. in order to draw the final sampling units. In particular, in a two-stage sampling design, the population is divided into several “primary” units from which a sample of primary units is selected, and then a sample of secondary units is selected within each selected primary unit. In each stage we use simple random

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

* Author to whom correspondence may be addressed.
E-mail: Magnus.Ekstrom@umu.se

sampling without replacement (SRSWOR), where a sample is taken without replacement and any sample of a given size has the same probability of being chosen. It should be noted that two of the most popular sampling designs—stratified sampling and cluster sampling—can be viewed as special cases of such two-stage sampling. A stratified random sample is a complete census of the primary units (the strata) followed by a sample of the secondary units within each primary unit. A cluster sample, on the other hand, is a sample of the primary units (the clusters) followed by a complete census of the secondary units within each selected primary unit. This allows considerable flexibility depending on the homogeneity of units at the primary or secondary level, as well as cost considerations.

Our aim in this paper is to study the asymptotic normality and develop Edgeworth expansions for the cumulative distribution function of the estimator of the population total (or the mean) in two-stage sampling, and then specialize these results to the case of stratified and cluster samplings. Such results, although very important and useful, become rather non-trivial because of the complex probability distribution on the selected subset of units induced by the sampling design. Large sample properties and statistical inferences for estimators in the context of finite populations are considerably more involved than in the independently and identically distributed (i.i.d) case because they depend not only on the characteristics of the finite population but also the sampling design employed. For instance, choice of units in stratified simple random sampling without replacement would correspond to a multivariate hypergeometric distribution, which can be viewed as independent binomials (with the same probability of selection) conditional on their sum. Such sampling schemes may be viewed as draws in “generalized urn models” and the resulting estimate of the population total as a sum of functions of resulting frequencies in such a context. This enables us to exploit some of the machinery developed in Mirakhmedov, Jammalamadaka, & Ibrahim (2014), as we do in Section 2.

The main objective is to make inferences on the parameters of the finite population using a sample selected from the finite population according to a specified probability sampling design. Even in simple situations, the exact distribution of the relevant estimators can be too complex to be determined analytically, and large sample theory and approximations provide a useful alternative for making such inferences. In this paper we shall consider two-stage designs, where it is assumed that population size as well as the sample sizes in each stage are sufficiently large.

One of the most commonly estimated finite population parameters is the “population total,” denoted by Y or the corresponding population mean. The estimator used for this purpose, say \hat{Y} (see Section 2), can be approximated by a Gaussian distribution under fairly general conditions, which also allows us to set confidence intervals for large samples as is usually done.

One of our primary goals in this paper is to obtain better approximations for this large sample distribution by studying the Edgeworth asymptotic expansion for this estimator. Not only are such analytical expressions of interest in their own right, but they also allow us to provide more accurate confidence intervals for the unknown population total (or mean), as we show. Under the single-stage SRSWOR design, asymptotic results for this estimator, which coincide with corresponding results on the sample mean from a finite population of a real numbers, are well studied in the literature: see, e.g., Erdős & Rényi (1959), Hájek (1960), Scott & Wu (1981), Robinson (1978), Bickel & van Zwet (1978), Sugden & Smith (1997), and Bloznelis (2000). For results on stratified and cluster sampling, see Rao (1973), Cochran (1977), Sen (1988) and Schenker & Welsh (1988), Krewski & Rao (1981), and Bickel & Freedman (1984), whereas Hájek (1964) and Prášková (1984) discuss results for unequal probability sampling. As we show in Section 2 the estimator under a two-stage scheme can be reduced to a weighted sample mean from a finite population of “random” variables. Asymptotic normality and Edgeworth asymptotic expansion for this sample mean have been considered by von Bahr (1972), Mirakhmedov (1983), Hu, Robinson, & Wang (2007), Mirakhmedov, Jammalamadaka, & Ibrahim (2014), and Ibrahim & Mirakhmedov (2013).

Our second aim is to focus specifically on the stratified and cluster sampling designs. We first apply the general Edgeworth asymptotic expansion result by writing down the first two terms in the expansion for standardized versions of the estimators. However their practical use is limited, as the standard deviation of these statistics is not known in practice. We hence derive formally a two-term asymptotic expansion for the Studentized version of these estimators; the result for stratified sampling design extends corresponding results by Babu & Singh (1985) and Sugden & Smith (1997) who considered a sample mean in the one-stage SRSWOR design. Finally, our third aim is to apply the theoretical results obtained here to construct skewness-adjusted confidence intervals for the unknown population total. Extensive Monte Carlo studies indicate that the additional terms in the asymptotic expansion do indeed provide better results than the usual normal approximation based confidence intervals, especially when combined with bootstrap methods.

A note about the notations. $\phi(x)$ and $\Phi(x)$ will denote the density and distribution functions of the standard normal distribution, respectively. c and C will denote some absolute finite constants and θ will denote a number whose absolute value does not exceed 1. It should be observed that c , C , and θ may be different when used in different equalities and inequalities or in different parts of the same equality or inequality. The main results are to be found in Theorems 1 and 2, as well as in Corollary 1, given in Sections 2 and 3. The long proofs of these two theorems are given in the Appendix. Construction of confidence intervals is discussed in Section 4, which contains two short Monte Carlo simulation studies as well.

2. ESTIMATORS IN TWO-STAGE SAMPLING

Suppose the population consists of N primary units, with the j th primary unit consisting of M_j secondary units. The value of the k th secondary unit within the j th primary unit is denoted by y_{jk} . Thus for the j th primary unit, the total and the mean are given by

$$Y_j = \sum_{k=1}^{M_j} y_{jk} \quad \text{and} \quad \bar{Y}_j = \frac{Y_j}{M_j}.$$

The quantity of interest that is to be estimated is the “population total,”

$$Y = \sum_{j=1}^N Y_j,$$

and we denote the mean per primary unit by $\bar{Y} = Y/N$. We draw a simple random sample without replacement at each stage: a simple random sample s of n primary units out of N , and a simple random sample s_j of m_j secondary units out of M_j within each primary unit $j \in s$. Let

$$f_1 = n/N \quad \text{and} \quad f_{2j} = m_j/M_j \tag{1}$$

represent the sampling fractions in these two stages, with the notation

$$g_1 = (1 - f_1) \quad \text{and} \quad g_{2j} = (1 - f_{2j}). \tag{2}$$

Denote the sample mean in the selected j th primary unit by \bar{y}_j , and let $y_j = m_j \bar{y}_j = \sum_{k \in s_j} y_{jk}$ be the corresponding sample total.

The unbiased estimator of Y_j is given by

$$\hat{Y}_j = M_j \bar{y}_j = \frac{y_j}{f_{2j}},$$

and the unbiased estimator of the population total Y is

$$\hat{Y}_{ts} = \frac{N}{n} \sum_{j \in s} \hat{Y}_j, \tag{3}$$

the subscript “ts” denoting two-stage, to distinguish it from other estimators coming later.

We are interested in the Edgeworth expansion for \hat{Y}_{ts} . In order to derive the expansion, we will use a somewhat different interpretation of the estimator \hat{Y}_{ts} . This will be done by reversing the two stages in the sampling plan. Now, in the first stage, a simple random sample s_j of m_j secondary units is drawn within *each* primary unit. As before we form the estimators \hat{Y}_j , but this time for $j = 1, \dots, N$ (rather than for $j \in s$). In the second stage, a simple random sample s' of size n is drawn from the “population” consisting of the “units” $\hat{Y}_1, \dots, \hat{Y}_N$, and thereafter we form the estimator

$$\hat{Y}'_{ts} = \frac{N}{n} \sum_{j \in s'} \hat{Y}_j.$$

It is clear that \hat{Y}_{ts} and \hat{Y}'_{ts} are equal in distribution, and that $f_1 \hat{Y}'_{ts}$ is a sample sum from a finite population of independent random variables.

Let

$$\alpha_{rl} = \frac{1}{N} \sum_{j=1}^N E \left[(\hat{Y}_j - \bar{Y})^r \right] (Y_j - \bar{Y})^l, \quad (4)$$

$$\sigma_{ts}^2 = \alpha_{20} - f_1 \alpha_{02}, \quad (5)$$

and

$$\Psi_{ts}(u) = \Phi(u) + \frac{(1-u^2)\phi(u)}{6n^{1/2}\sigma_{ts}^3}(\alpha_{30} - 3f_1\alpha_{21} + 2f_1^2\alpha_{03}), \quad (6)$$

and note that σ_{ts}^2 is the asymptotic variance of $n^{-1/2}\hat{Y}_{ts}$.

We wish to derive a suitable upper bound for

$$\sup_{u \in \mathbb{R}} \left| P \left\{ \frac{f_1(\hat{Y}'_{ts} - Y)}{n^{1/2}\sigma_{ts}} < u \right\} - \Psi_{ts}(u) \right|,$$

by utilizing Theorem 2 of Hu, Robinson, & Wang (2007). For doing this, we will need a technical condition, which ensures that the values y_{jk} in each secondary unit do not cluster around too few values (cf. Robinson (1978), Bickel & van Zwet (1978), and Mirakhhmedov (1983), where similar conditions are used). This technical condition, Condition C, is given in the Appendix.

Let

$$v_{rj} = M_j^{-1} \sum_{k=1}^{M_j} (y_{jk} - \bar{Y}_j)^r, \quad (7)$$

$$\rho = \frac{1}{N} \sum_{j=1}^N \frac{m_j}{f_{2j}^4} g_{2j}(v_{4j} + 3m_j g_{2j} v_{2j}^2) + \alpha_{04}, \quad (8)$$

and

$$\varrho = \left(n^{1/2} \log n \right) \exp \left[-n \left\{ 1 - 8^{1/2} \frac{1}{N} \sum_{j=1}^N (m_j g_{2j})^{1/2} \exp(-c(\epsilon, \delta) m_j g_{2j}) \right\} \right], \quad (9)$$

where ϵ and δ are positive constants defined in Condition C, and $c(\epsilon, \delta)$ is a positive constant depending only on ϵ and δ . We then have the following main result of this section whose proof is given in the Appendix.

Theorem 1. *If Condition C is fulfilled, then there exists a positive constant c such that*

$$\sup_{u \in \mathbb{R}} \left| P \left\{ \frac{f_1(\hat{Y}_{ts} - Y)}{n^{1/2}\sigma_{ts}} < u \right\} - \Psi_{ts}(u) \right| \leq \frac{c\rho}{n\sigma_{ts}^4} + 3\varrho,$$

where Y is the population total, and f_1 , \hat{Y}_{ts} , σ_{ts}^2 , $\Psi_{ts}(u)$, ρ , and ϱ are defined in (1), (3), (5), (6), (8), and (9), respectively.

Remarks 1. *The quantity $\varrho = \varrho(m_j, M_j, n, N)$ is exponentially decreasing when $m_j g_{2j}$ is increasing; the latter being a necessary condition for \hat{Y}_j to be asymptotically normal. In addition, $\varrho = \varrho(m_j, M_j, n, N)$ is exponentially small if $m_j g_{2j} < 1/8$.*

Remarks 2. *The quantities α_{20} , α_{21} , and α_{30} appearing in the definition of $\Psi_{ts}(u)$ can be computed through the following formulas:*

$$\alpha_{20} = \frac{1}{N} \sum_{j=1}^N \frac{m_j}{f_{2j}^2} g_{2j} v_{2j} \frac{M_j}{M_j - 1} + \alpha_{02}, \quad (10)$$

$$\alpha_{21} = \frac{1}{N} \sum_{j=1}^N \frac{m_j}{f_{2j}^2} g_{2j} v_{2j} \frac{M_j}{M_j - 1} (Y_j - \bar{Y}) + \alpha_{03}, \quad (11)$$

and

$$\alpha_{30} = \frac{1}{N} \sum_{j=1}^N \frac{m_j}{f_{2j}^3} (1 - 3f_{2j} + 2f_{2j}^2) v_{3j} \frac{M_j}{M_j - 1} \frac{M_j}{M_j - 2} + 3\alpha_{21} - 2\alpha_{03}, \quad (12)$$

where using (10), σ_{ts}^2 given in (5) can be rewritten as

$$\sigma_{ts}^2 = \frac{1}{N} \sum_{j=1}^N \frac{m_j}{f_{2j}^2} g_{2j} v_{2j} \frac{M_j}{M_j - 1} + g_1 \alpha_{02}.$$

Remarks 3. *To limit complexity, we have restricted ourselves to a two-term expansion in Theorem 1. However higher order terms can be obtained similar to the results in Ibrahim & Mirakhmedov (2013).*

3. STUDENTIZED ESTIMATORS IN STRATIFIED AND CLUSTER SAMPLINGS

We now consider two important special cases, namely: (i) stratified random sampling and (ii) cluster sampling. We continue to use the notations of Section 2.

- (i) Stratified random sampling is a special case of two-stage sampling where $n = N$. Hence for this case,

$$\sigma_{\text{ts}}^2 = \frac{1}{N} \sum_{j=1}^N \frac{m_j}{f_{2j}^2} g_{2j} v_{2j} \frac{M_j}{M_j - 1},$$

$$\Psi_{\text{ts}}(u) = \Phi(u) + \frac{(1-u^2)\phi(u)}{6n^{1/2}\sigma_{\text{ts}}^3} \frac{1}{N} \sum_{j=1}^N \frac{m_j}{f_{2j}^3} (1-f_{2j})(1-2f_{2j})v_{3j} \frac{M_j}{M_j-1} \frac{M_j}{M_j-2},$$

and ρ is defined as in (8).

- (ii) Cluster sampling is a special case of two-stage sampling with $m_j = M_j$, $j = 1, \dots, N$. Hence for this case,

$$\sigma_{\text{ts}}^2 = g_1 \alpha_{02},$$

$$\Psi_{\text{ts}}(u) = \Phi(u) + \frac{(1-u^2)\phi(u)(1-2f_1)\alpha_{03}}{6(n g_1)^{1/2} \alpha_{02}^{3/2}},$$

and $\rho = \alpha_{04}$.

Although the foregoing Edgeworth expansions obtained from Theorem 1 provide better approximations than the Central Limit Theorem for the standardized statistics, they are not useful in practice as the variances of the statistics under consideration are not known. Due to this fact we shall now obtain formally a two-term expansion for the Studentized versions of our statistics, both for the stratified and cluster sampling designs.

For the population total, Y , the estimator used in stratified random sampling is

$$\hat{Y}_{\text{str}} = \sum_{j=1}^N \hat{Y}_j. \quad (13)$$

Let $m = \sum_{j=1}^N m_j$. The variance of \hat{Y}_{str} is

$$m\sigma_{\text{str}}^2 = \sum_{j=1}^N \frac{m_j g_{2j} v_{2j}}{f_{2j}^2} \frac{M_j}{M_j - 1},$$

which is unbiasedly estimated by

$$mS_{\text{str}}^2 = \sum_{j=1}^N \frac{m_j g_{2j} S_j^2}{f_{2j}^2},$$

where

$$S_j^2 = \frac{1}{m_j - 1} \sum_{k \in s_j} (y_{jk} - \bar{y}_j)^2.$$

We now define the Studentized version of the estimator \hat{Y}_{str} as

$$T_{\text{str}} = \frac{\hat{Y}_{\text{str}} - Y}{m^{1/2} S_{\text{str}}}, \quad (14)$$

and seek a two-term Edgeworth expansion,

$$P\{T_{\text{str}} < u\} = \Phi(u) + \frac{\phi(u)}{m^{1/2}} \psi_m(u) + O\left(\frac{1}{m}\right). \quad (15)$$

The statistic \hat{Y}_{str} , defined in (13), can be viewed as a sum of independent but non-identically distributed random variables $\hat{Y}_j = y_j/f_{2j}$, $j = 1, \dots, N$, where y_j is the sample sum in stratum j . Hence the Studentized version T_{str} of \hat{Y}_{str} is a special case of Student's t -statistic based on independent but non-identically distributed random variables. Many authors have studied the problem of approximating the distribution of t -statistics by a normal distribution, see Hall & Wang (2004) and the references therein. In particular, Bentkus, Bloznelis, & Götze (1996) gave a Berry-Esseen bound for Student's t -statistic in the case of non-identically distributed data. From their Theorem 1.1, together with inequality (A4) in the Appendix of the current paper, we have

$$|P\{T_{\text{str}} < u\} - \Phi(u)| \leq \frac{c(\rho N)^{1/2}}{m\sigma_{\text{str}}^2}, \quad (16)$$

where ρ is defined as in (8). The right-hand side of (16) is of order $O(N^{-1/2})$ under some additional restrictions, e.g., under the following condition:

Condition A: f_{2j} and $mN^{-1}\sigma_{\text{str}}^2$ are bounded away from zero, whereas v_{4j} and $N^{-1} \sum_{j=1}^N m_j g_{2j}$ are bounded away from infinity.

To the best of our knowledge, existing results on Edgeworth expansions for Student's t -statistic consider the independent and identically distributed case only; we refer to Hall (1987) and Hall & Wang (2004), who have established Edgeworth expansion results in this case under different refinements. From their result, it follows that, if Cramer's continuity condition $\limsup_{|t| \rightarrow \infty} |E[e^{itX}]| < 1$ holds where X is the observed random variable, then, under some moment conditions, the remainder term of the two-term Edgeworth expansion is $O(N^{-1})$. Extensions of such results to non-identically distributed data is a difficult probabilistic problem, which we do not consider here. Instead we obtain the second term of the Edgeworth expansion in (15), see Theorem 2, using a procedure somewhat similar to that outlined in Sugden & Smith (1997, Section 2). The details are given in the Appendix.

Theorem 2. *In the second term of the Edgeworth expansion in (15), we have*

$$\frac{1}{m^{1/2}} \psi_m(u) = \frac{\sum_{j=1}^N m_j f_{2j}^{-3} g_{2j} v_{3j} ((2 - f_{2j})(u^2 - 1) + 3g_{2j})}{6 \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} \right)^{3/2}}, \quad (17)$$

where f_{2j} , g_{2j} , and v_{rj} are defined in (1), (2), and (7), respectively.

The above theorem is sufficient to establish (15) formally. In addition, we conjecture that Equation (15), with respect to the remainder term, holds true under Conditions A and C, because in this case Cramer's continuity condition, $\limsup_{|t| \rightarrow \infty} |E[e^{it\hat{Y}_j}]| < 1$, follows from inequality (A9) in the Appendix.

The case when $N = 1$ with just one primary unit corresponds to the simple random sampling design considered by Sugden & Smith (1997), and their two-term expansion result follows from our theorem above. For this case, we have

$$T_{\text{str}} = \frac{f_{21}(\hat{Y}_1 - Y_1)}{(m_1 g_{21})^{1/2} S_1}$$

which is the Studentized version of the estimator \hat{Y}_1 of the population total Y_1 in simple random sampling from a population of size M_1 , and the two-term Edgeworth expansion for $P\{f_{21}(\hat{Y}_1 - Y_1) < u(m_1 g_{21})^{1/2} S_1\}$ equals

$$\begin{aligned} \Phi(u) + \frac{\phi(u)}{6m_1^{1/2}} \frac{v_{31}}{v_{21}^{3/2}} \left(\frac{2 - f_{21}}{g_{21}^{1/2}} (u^2 - 1) + 3g_{21}^{1/2} \right) \\ = \Phi(u) + \frac{\phi(u)}{6(m_1 g_{21})^{1/2}} \frac{v_{31}}{v_{21}^{3/2}} \left(3g_{21} u^2 + (1 - u^2)(1 - 2f_{21}) \right). \end{aligned} \quad (18)$$

As yet another special case of two-stage sampling, we now consider the case of cluster sampling which corresponds to taking $m_j = M_j$ for $j = 1, \dots, N$. For the population total, Y , the estimator used in cluster sampling is

$$\hat{Y}_{\text{clu}} = \frac{N}{n} \sum_{j \in s} Y_j.$$

For this case, we define the Studentized version as

$$T_{\text{clu}} = \frac{f_1(\hat{Y}_{\text{clu}} - Y)}{(ng_1)^{1/2} S_{\text{clu}}},$$

where

$$S_{\text{clu}}^2 = \frac{1}{n-1} \sum_{j \in s} \left(Y_j - \frac{1}{n} \sum_{j' \in s} Y_{j'} \right)^2,$$

and seek a two-term Edgeworth expansion,

$$P\{T_{\text{clu}} < u\} = \Phi(u) + \frac{\phi(u)}{n^{1/2}} \psi_n(u) + (\text{remainder term}). \quad (19)$$

By noting that cluster sampling can be viewed as simple random sampling from a “population” consisting of the “units,” i.e., cluster totals, Y_1, \dots, Y_N , and by (18), we obtain the following result.

Corollary 1. *In the second term of the Edgeworth expansion in (19), we have*

$$\frac{1}{n^{1/2}} \psi_n(u) = \frac{1}{6(ng_1)^{1/2}} \frac{\alpha_{03}}{\alpha_{02}^{3/2}} \left(3g_1 u^2 + (1 - u^2)(1 - 2f_1) \right),$$

where f_1 , g_1 , and α_{0l} are defined in (1), (2), and (4), respectively.

Remarks 4. *By using results from Babu & Singh (1985, p. 265) it follows, under some assumptions that include a limiting non-latticeness condition, that the remainder term in the two-term Edgeworth expansion (19) is at least $o(n^{-1/2})$.*

Remarks 5. The quantities v_{rj} and α_{0r} , $r = 2, 3$, in Theorem 2 and Corollary 1 are unbiasedly estimated by

$$\hat{v}_{rj} = \frac{f_{2j}^{r-1}}{m_j} \sum_{k \in s_j} \left(y_{jk} - \frac{1}{m_j} \sum_{k' \in s_j} y_{jk'} \right)^r \prod_{k=1}^{r-1} \frac{M_j - k}{m_j - k} \quad (20)$$

and

$$\hat{\alpha}_{0r} = \frac{f_1^{r-1}}{n} \sum_{j \in s} \left(\hat{Y}_j - \frac{1}{n} \sum_{j' \in s} \hat{Y}_{j'} \right)^r \prod_{i=1}^{r-1} \frac{N - i}{n - i}.$$

The estimators $\hat{\alpha}_{0r}$ and \hat{v}_{rj} are \sqrt{n} - and $\sqrt{m_j}$ -consistent, respectively, i.e., for every $\epsilon > 0$ there exists a $c > 0$ such that

$$P\{\sqrt{n}|\hat{\alpha}_{0r} - \alpha_{0r}| > c\} \leq \epsilon, \quad (21)$$

$$P\{\sqrt{m_j}|\hat{v}_{rj} - v_{rj}| > c\} \leq \epsilon. \quad (22)$$

Let $\mu_k = N^{-1} \sum_{j=1}^N |\hat{Y}_j|^k$. That (21) holds true is a direct consequence of Chebyshev's inequality together with the inequality $\text{var}(\hat{\alpha}_{0r}) \leq c\mu_{2r}/n$, $r = 2, 3$. For $r = 2$, the latter inequality follows from an application of Theorem 1 in Cho, Cho, & Eltinge (2005) and the fact that $\mu_l \leq \mu_k^{l/k}$, $1 \leq l \leq k$, and for $r = 3$, from routine computations like those in the proof of Theorem 1 in Cho, Cho, & Eltinge (2005). The same reasoning shows that (22) holds true when $r = 2, 3$.

4. CONFIDENCE INTERVALS AND MONTE CARLO SIMULATIONS

Fletcher & Webster (1996) note that clumping in the spatial distribution of animal and plant species can lead to a high degree of skewness in the population, which can then carry over to the distribution of a stratified sample mean. For this reason, they studied several ways of calculating skewness-adjusted confidence intervals for population means from stratified random samples under the assumption that the sampling fraction in each stratum is negligible (i.e., that $m_j \ll M_j$ for all j). The set-up of our first Monte Carlo study will be similar to that in Fletcher & Webster (1996). By using the result of Theorem 2, we will show that Fletcher & Webster's conclusions can be extended to the case where the sampling fractions are not negligible. Our second Monte Carlo study uses data from the U.S. 1992 Census of Agriculture.

As in Fletcher & Webster (1996) we will consider four different methods for computing confidence intervals; in our case, for the population total, Y . The first one is based on the usual normal approximation of the distribution of $T_{\text{str}} = m^{-1/2} S_{\text{str}}^{-1} (\hat{Y}_{\text{str}} - Y)$, the Studentized version of the estimated population total, \hat{Y}_{str} , in stratified random sampling. That is, this $100(1 - \alpha)\%$ interval is given by

$$(\hat{Y}_{\text{str}} - z_{\alpha} m^{1/2} S_{\text{str}}, \hat{Y}_{\text{str}} + z_{\alpha} m^{1/2} S_{\text{str}}), \quad (23)$$

where z_{α} is the upper $\alpha/2$ percentile of the standard normal distribution.

In order to make use of the Edgeworth expansion in Theorem 2, we will use Hall's (1992, p. 123) idea to find an invertible cubic transformation, $F_{\text{str}} = f(T_{\text{str}})$, whose coefficients depend on the form of (17), such that the distribution of F_{str} is closer to the standard normal distribution than is the distribution of T_{str} (cf. Fletcher & Webster, 1996). Direct application of Hall's idea

gives the following cubic transformation:

$$F_{\text{str}} = f(T_{\text{str}}) = T_{\text{str}} + m^{-1/2}\hat{\gamma}_1 T_{\text{str}}^2 + \frac{1}{3}m^{-1}\hat{\gamma}_1^2 T_{\text{str}}^3 + m^{-1/2}\hat{\gamma}_2,$$

where

$$\hat{\gamma}_1 = \frac{m^{-1} \sum_{j=1}^N m_j f_{2j}^{-3} g_{2j} \hat{v}_{3j} (2 - f_{2j})}{6 \left(m^{-1} \sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} \hat{v}_{2j} \right)^{3/2}} \quad \text{and} \quad \hat{\gamma}_2 = \frac{m^{-1} \sum_{j=1}^N m_j f_{2j}^{-3} g_{2j} \hat{v}_{3j} (3g_{2j} - (2 - f_{2j}))}{6 \left(m^{-1} \sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} \hat{v}_{2j} \right)^{3/2}},$$

and \hat{v}_{rj} , $r = 2, 3$ is defined in (20). The inverse of the aforementioned transformation, needed for computing the confidence interval, is

$$f^{-1}(x) = \frac{m^{1/2}}{\hat{\gamma}_1} \left(\left(1 + \frac{3\hat{\gamma}_1}{m^{1/2}} \left(x - \frac{\hat{\gamma}_2}{m^{1/2}} \right) \right)^{1/3} - 1 \right),$$

and the corresponding $100(1 - \alpha)\%$ interval is then given by

$$(\hat{Y}_{\text{str}} - f^{-1}(z_\alpha) m^{1/2} S_{\text{str}}, \hat{Y}_{\text{str}} - f^{-1}(-z_\alpha) m^{1/2} S_{\text{str}}). \quad (24)$$

Remarks 6. By Theorem 2 we have $P\{T_{\text{str}} < u\} = \Phi(u) + m^{-1/2}(\gamma_1 u^2 + \gamma_2)\phi(u) + r_m$, where r_m is the remainder in (15), and γ_1 and γ_2 are defined as $\hat{\gamma}_1$ and $\hat{\gamma}_2$, respectively, but with \hat{v}_{2j} and \hat{v}_{3j} replaced by v_{2j} and v_{3j} . Let $\tilde{m} = \min_{1 \leq j \leq N} m_j$. It is easy to see that $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are $\tilde{m}^{1/2}$ -consistent estimators of γ_1 and γ_2 , respectively (cf. Remark 5). Following Hall (1992, pp. 122–123), define $g(t) = t + m^{-1/2}\hat{\gamma}_1 t^2 + m^{-1/2}\hat{\gamma}_2$. If $\hat{\gamma}_1 > 0$ and u is fixed, then $g(t) < u$ is equivalent to $-m^{-1/2}\hat{\gamma}_1^{-1} + O_p(1) < t < u - m^{-1/2}(\hat{\gamma}_1 u^2 + \hat{\gamma}_2) + O_p(m^{-1})$. From this relation and a similar one for the case $\hat{\gamma}_1 < 0$, we deduce as in Hall (1992, p. 123) that if $\hat{\gamma}_1 \neq 0$, then

$$P\{g(T_{\text{str}}) < u\} = \Phi(u) + O((\tilde{m}m)^{-1/2}) + r_m. \quad (25)$$

The disadvantage of the quadratic transformation g is that it is generally not one-to-one, and this is why we use the aforementioned cubic transformation f instead. As the added term in the cubic transformation f is of order m^{-1} , formula (25) will not be affected if we replace transformation g with f . Thus, $f(T_{\text{str}})$ admits an Edgeworth expansion in which the second term is $O((\tilde{m}m)^{-1/2}) + r_m$ rather than $O(m^{-1/2})$.

The next and final two confidence intervals use the bootstrap to estimate the distributions of T_{str} and F_{str} , respectively. Ordinary bootstrap (applied independently within each stratum) is not recommended, as it involves *with-replacement* samples of size m_j in stratum j and so does not mimic how the original data were sampled. One way to deal with this is to use the population bootstrap, as described in, e.g., Davison & Hinkley (1997, Section 3.7). Then, in the j th stratum and if $l_j = M_j/m_j$ is an integer, a fake stratum of size M_j is constructed by replicating each sample observation y_{jk} , $k \in s_j$, l_j times, and a bootstrap replicate of $\{y_{jk}; k \in s_j\}$ is generated by taking a sample of size m_j without replacement from the constructed fake stratum. More generally, let l_j be the integer part of M_j/m_j and $d_j = M_j - l_j m_j$. Then a fake stratum is obtained by taking l_j copies of each y_{jk} , $k \in s_j$, and adding to them a sample of size d_j taken without replacement from y_{jk} , $k \in s_j$. By applying the population bootstrap independently within each stratum, a bootstrap replicate of $\{y_{jk}; k \in s_j, j = 1, \dots, N\}$ is obtained. From bootstrap replicate b , $b = 1, \dots, B$, $\hat{Y}_{\text{str},b}$ and $S_{\text{str},b}$ are constructed in the same way as \hat{Y}_{str} and S_{str} are constructed from the original stratified sample. The bootstrap version of the statistic T_{str} is then defined as

TABLE 1: Summary statistics for the finite population in the Monte Carlo study; δ is the proportion of zeros, and for the nonzero values μ and σ^2 are the mean and variance on the log scale.

Density	δ	μ	σ^2
Low	0.46	4.13	2.26
High	0.32	5.96	1.33

$T_{\text{str},b} = m^{-1/2} S_{\text{str},b}^{-1} (\hat{Y}_{\text{str},b} - \hat{Y}_{\text{str}})$, and a $100(1 - \alpha)\%$ confidence interval for the population total is given by

$$(\hat{Y}_{\text{str}} - t_U m^{1/2} S_{\text{str}}, \hat{Y}_{\text{str}} - t_L m^{1/2} S_{\text{str}}), \quad (26)$$

where t_L and t_U are the lower and upper $\alpha/2$ percentiles of the empirical distribution of $\{T_{\text{str},b}\}_{b=1}^B$. Likewise, the bootstrap version of F_{str} is defined as

$$F_{\text{str},b} = m^{-1/2} \hat{\gamma}_{1,b} + T_{\text{str},b} + m^{-1/2} \hat{\gamma}_{2,b} T_{\text{str},b}^2 + \frac{1}{3} m^{-1} \hat{\gamma}_{2,b}^2 T_{\text{str},b}^3,$$

where $\hat{\gamma}_{1,b}$ and $\hat{\gamma}_{2,b}$ are constructed in the same way as $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are constructed from the original stratified sample. A $100(1 - \alpha)\%$ confidence interval for the population total is then obtained as

$$(\hat{Y}_{\text{str}} - f^{-1}(f_U) m^{1/2} S_{\text{str}}, \hat{Y}_{\text{str}} - f^{-1}(f_L) m^{1/2} S_{\text{str}}), \quad (27)$$

where f_L and f_U are, respectively, the lower and upper $\alpha/2$ percentiles of the empirical distribution of $\{F_{\text{str},b}\}_{b=1}^B$.

We henceforth will refer to the confidence intervals (23)–(27) as the NT, NF, BT, and BF interval, respectively. In the Monte Carlo study, three basic types of survey are considered for each total sample size m . Type I consists of as many strata as possible (for computing the \hat{v}_{3j} , at least three units are needed in each stratum), type III has only two strata, and type II is chosen as an intermediate between type I and type III. For every type of survey and sample size considered, half the strata are “high density,” consisting in total of 120 units, and half “low density,” consisting in total of 1,080 units. The values of the population units were generated from the corresponding infinite population in Fletcher & Webster (1996), and summary statistics of our finite population are given in Table 1. Survey types and total sample sizes considered are presented in Table 2. It should be noted that the use of equal sample sizes in the strata corresponds approximately to Neyman allocation.

TABLE 2: Survey types and total sample sizes in the Monte Carlo study.

Total sample size (m)	I		II		III	
	N	m_j	N	m_j	N	m_j
30	10	3	6	5	2	15
48	16	3	8	6	2	24
90	30	3	10	9	2	45
144	48	3	12	12	2	72

TABLE 3: Lower error rates (%).

Method	<i>m</i> = 30			<i>m</i> = 48			<i>m</i> = 90			<i>m</i> = 144		
	I	II	III	I	II	III	I	II	III	I	II	III
NT	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.2	0.2	0.2
NF	1.0	1.2	1.3	1.3	1.5	1.6	1.7	1.8	1.5	2.2	2.5	2.3
BT	0.0	0.1	0.6	0.0	0.5	1.0	0.1	1.0	1.8	0.1	1.6	2.1
BF	0.5	4.8	2.2	0.8	3.8	2.5	1.1	3.2	2.9	1.6	3.6	3.4

For the simulation results presented in Tables 3 and 4, the nominal error rate is set to $100\alpha = 5\%$, and $B = 1,000$. For each survey type and total sample size, a lower (upper) error rate is computed as the percentage of simulation replicates for which the lower (upper) limit of the confidence interval is above (below) the population total, Y . Ideally, these error rates should both be 2.5%; however, due to skewness of the actual sampling distribution, the actual rates may be far from the desired values. In Tables 3 and 4, each pair of lower and upper rates is based on 10,000 repeated stratified samples from the defined population.

For NT, NF, and BT, the lower rates are too low and the upper rates are too high, and the upper rates are worse than the lower ones. This holds true also for the BF method, except that it produces too high lower rates in some cases. The NT method is worse than the other methods, except that the BT method produces even worse lower rates when the stratum sizes are small. The best lower rates in the case of small and intermediate stratum sizes are given by the NF method, whereas the BF method appears to give better lower rates for large stratum sizes. With respect to the upper error rates, the BF method is uniformly better than all the other methods, and BT and NF perform about equally well, although the latter is not as good when the total sample size is large. The best overall error rates are provided by BT and BF, where the latter is to be preferred if the stratum sizes are small.

The results in Table 5 show that BT, NF, and BF produce wider intervals than NT. For example, for survey type I and total sample size $m = 30$, NF produces intervals 2.0 times wider than NT, on average. The NF method results in narrower intervals than the BF method. For larger strata and total sample sizes, BT produces narrower intervals than NF, but the opposite holds true when stratum or total sample sizes are small.

We conclude this section by considering sampling from two real-world populations. The first real-world example uses data from the U.S. 1992 Census of Agriculture. As our variable of study, we use the number of farms with 1,000 acres or more. For this example, we use the four census

TABLE 4: Upper error rates (%).

Method	<i>m</i> = 30			<i>m</i> = 48			<i>m</i> = 90			<i>m</i> = 144		
	I	II	III	I	II	III	I	II	III	I	II	III
NT	22.6	22.1	21.4	18.6	17.2	18.2	15.0	14.7	14.7	13.7	13.4	13.4
NF	11.1	10.6	9.7	9.4	8.2	9.0	7.3	6.7	6.1	8.1	7.5	7.2
BT	11.9	11.2	10.4	10.1	8.8	9.2	7.4	6.7	6.0	6.9	5.5	5.6
BF	7.9	9.2	9.4	6.9	7.4	8.3	5.1	5.0	4.8	5.0	4.8	5.1

TABLE 5: Average length of confidence intervals BT, NF, and BF, relative the NT interval.

Method	<i>m</i> = 30			<i>m</i> = 48			<i>m</i> = 90			<i>m</i> = 144		
	I	II	III	I	II	III	I	II	III	I	II	III
NF	2.0	1.9	2.0	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9
BT	4.2	2.7	2.7	3.1	2.2	2.1	2.5	1.8	1.8	2.1	1.7	1.7
BF	3.9	2.6	2.5	3.2	2.4	2.3	2.9	2.3	2.2	2.8	2.3	2.3

regions of the United States – Northeast, North Central, South, and West – as strata, consisting of $M_1 = 220$, $M_2 = 1054$, $M_3 = 1382$, and $M_4 = 422$ counties, respectively. The population is illustrated in Figure 1, and we see that it has a high degree of skewness (but not quite as extreme as in our previous example). We repeatedly take stratified samples from the population, with $m_1 = 21$, $m_2 = 103$, $m_3 = 135$, and $m_4 = 41$, so each sampling fraction, f_{2j} , is in the interval $(0.090, 0.098)$. Again, we use $\alpha = 0.05$ and $B = 1,000$. The results, based on 10,000 repeated stratified samples from the population, are given in Table 6. The NT method is worse than the other methods. The overall winner is the NF method, with an overall error rate equal to $2.50 + 2.25 = 4.75$, and with the BT and BF methods close behind. The NF, BT, and BF intervals are slightly wider than the NT intervals (about 1.4%, 2.6%, and 2.1% wider, on average).

The second real-world population consists of the 284 municipalities of Sweden and is called the MU284 population; it can be found in Särndal, Swensson, & Wretman (1992, pp. 652–659). We use the variable CS82, which is the number of Conservative seats in a municipal council,

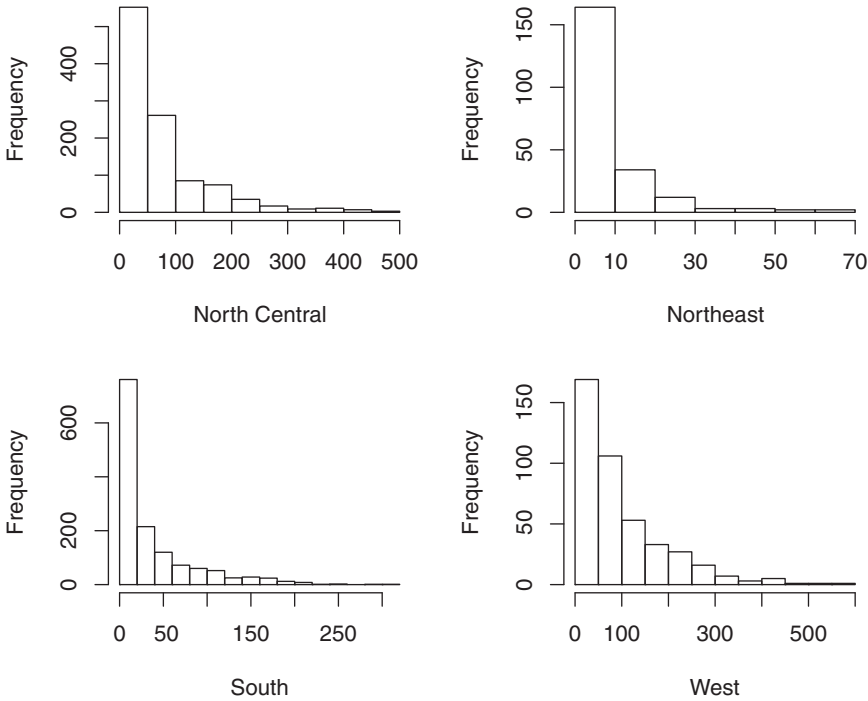


FIGURE 1: Histograms of the number of farms with 1,000 acres or more in the four census regions of the United States.

TABLE 6: Lower (L) and upper (U) error rates (%).

Method	L	U
NT	1.69	3.66
NF	2.50	2.25
BT	2.48	2.18
BF	2.63	2.18

and the $N = 50$ clusters as defined in Särndal, Swensson, & Wretman (1992). As before, we use $\alpha = 0.05$ and $B = 1,000$, and NT, NF, BT, and BF intervals are defined analogously to those in the case of stratified sampling. The results, based on 10,000 repeated cluster samples of size $n = 16$ from the MU284 population, are given in Table 7. We see that the lower rates are too low and the upper rates are too high, except for the NF lower rate which is slightly larger than the nominal lower rate. The NF has the best lower rate, whereas the BT method has the best upper and overall rates (but the worst lower rate). The NF, BT, and BF intervals are wider than the NT intervals (about 2.0%, 12.0%, and 15.2% wider, on average), but the NF interval is only slightly wider.

5. CONCLUDING REMARKS

In this paper we have derived a two-term Edgeworth expansion for the standardized sample total in two-stage sampling. For two very important special cases, namely stratified random sampling and cluster sampling, formal two-term Edgeworth expansions have been obtained for Studentized sample totals. We have illustrated that such results can be very useful for calculating skewness-adjusted confidence intervals for the population total. By itself, the second-order terms in the expansion for the Studentized stratified sample total can improve coverage error to an extent comparable to what is achieved by using the bootstrap, and appears to be even better than the bootstrap on the lower limit.

Further improvements, at least on the upper limit, can be achieved by using these higher order terms together with the bootstrap (adapted to the finite population setting).

APPENDIX

First we specify the Condition C, needed for ensuring the result of Theorem 1.

Condition C: For each $j = 1, \dots, N$, there exists $\varepsilon > 0$ and $\delta > 0$ not depending on m_j and M_j such that

$$\#\{k : \left| u f_{2j}^{-1} y_{jk} - x - 2\pi m \right| > \varepsilon\} \geq \delta M_j,$$

TABLE 7: Lower (L) and upper (U) error rates (%).

Method	L	U
NT	2.13	4.73
NF	2.62	3.86
BT	1.86	2.98
BF	2.17	3.34

for any fixed x , all $M_j > 1$, $1 \leq m_j \leq M_j$, all integers m , and all u satisfying

$$\frac{8\sigma_{ts}^2}{5 \cdot 10^5 f_1 \rho^{3/4}} \leq u \leq \frac{16n^{1/2}\sigma_{ts}}{f_1}.$$

Proof of Theorem 1. Recall that $f_1 \hat{Y}'_{ts}$ can be viewed as a sample sum from the finite population $(\hat{Y}_1, \dots, \hat{Y}_N)$. Therefore, using Theorem 2 of Hu, Robinson, & Wang (2007), with $X_j = \hat{Y}_j - \bar{Y}$, $\delta_0 = 8 \cdot 10^5$ (which satisfy their condition for δ_0), and $L_0 = N^{-1} \sum_{j=1}^N E[|\hat{Y}_j - \bar{Y}|^3]$, we see that

$$\sup_{u \in \mathbb{R}} \left| P \left\{ \frac{f_1(\hat{Y}'_{ts} - Y)}{n^{1/2}\sigma_{ts}} < u \right\} - \Psi_{ts}(u) \right| \leq \frac{C\alpha_{40}}{n\sigma_{ts}^4} + 3\chi(\delta_1, \delta_2), \quad (A1)$$

where

$$\chi(d_1, d_2) = n^{1/2} \log n \exp \left(-n \left(1 - \sup_{d_1 \leq |u| \leq d_2} \frac{1}{N} \sum_{j=1}^N |E[\exp(iu\hat{Y}_j)]| \right) \right),$$

$$\delta_1 = \frac{8\sigma_{ts}^2}{10^5 L_0} \quad \text{and} \quad \delta_2 = 16n^{1/2}\sigma_{ts}.$$

Note that $y_j = f_{2j}\hat{Y}_j$ is a sample sum obtained by simple random sampling from the j th primary unit. Therefore, by Finucan, Galbraith, & Stone (1974, pp. 152–153),

$$\begin{aligned} E[(\hat{Y}_j - Y_j)^2] &= \frac{m_j}{f_{2j}^2} g_{2j} v_{2j} \frac{M_j}{M_j - 1}, \\ E[(\hat{Y}_j - Y_j)^3] &= \frac{m_j}{f_{2j}^3} g_{2j} (1 - 2f_{2j}) v_{3j} \frac{M_j}{M_j - 1} \frac{M_j}{M_j - 2}, \\ E[(\hat{Y}_j - Y_j)^4] &\leq \frac{m_j}{f_{2j}^4} g_{2j} (v_{4j} + 3m_j g_{2j} v_{2j}^2) \left(1 + O\left(\frac{1}{M_j}\right) \right). \end{aligned} \quad (A2)$$

Thus,

$$\alpha_{0l} = \frac{1}{N} \sum_{j=1}^N (Y_j - \bar{Y})^l, \quad (A3)$$

and

$$\begin{aligned} \alpha_{20} &= \frac{1}{N} \sum_{j=1}^N E[(\hat{Y}_j - \bar{Y})^2] \\ &= \frac{1}{N} \sum_{j=1}^N E[(\hat{Y}_j - Y_j) + (Y_j - \bar{Y})^2] \\ &= \frac{1}{N} \sum_{j=1}^N \frac{m_j}{f_{2j}^2} g_{2j} v_{2j} \frac{M_j}{M_j - 1} + \alpha_{02}, \end{aligned}$$

and the formulas (11)–(12) for α_{21} and α_{30} can be derived in a similar manner.

Next, consider the right-hand side of the inequality (A1). We want to find suitable upper bounds for L_0 and $\chi(d_1, d_2)$. We have

$$\alpha_{40} = \frac{1}{N} \sum_{j=1}^N E[(\hat{Y}_j - Y_j) + (Y_j - \bar{Y})^4] \leq 8 \left(\frac{1}{N} \sum_{j=1}^N E[(\hat{Y}_j - Y_j)^4] + \alpha_{04} \right) \leq 8\rho, \quad (\text{A4})$$

where the last inequality follows from (A2) and ρ is defined as in (8). By Hölder's inequality and (A4), $L_0 \leq \alpha_{40}^{3/4} < 5\rho^{3/4}$. Hence

$$\delta_1 = \frac{8\sigma_{\text{ts}}^2}{10^5 L_0} \geq \delta'_1 = \frac{8\sigma_{\text{ts}}^2}{5 \times 10^5 \rho^{3/4}},$$

implying

$$\chi(\delta_1, \delta_2) \leq \chi(\delta'_1, \delta_2). \quad (\text{A5})$$

By the Erdős & Rényi (1959) formula (see also formula (2.5) in Mirakhmedov, Jammalamadaka, & Ibrahim (2014)), we have

$$E \left[\exp \left(iu \hat{Y}_j \right) \right] = \left(2\pi \binom{M_j}{m_j} f_{2j}^{m_j} g_{2j}^{M_j - m_j} \right)^{-1} \int_{-\pi}^{\pi} \prod_{j=1}^{M_j} \psi_{jk}(u, \tau) d\tau, \quad (\text{A6})$$

where

$$\psi_{jk}(u, \tau) = E \left[\exp \left(iu f_{2j}^{-1} y_{jk} \xi_j + i\tau (\xi_j - f_{2j}) \right) \right] = e^{i\tau g_{2j}} \left(1 + f_{2j} \left(e^{i(\tau + u f_{2j}^{-1} y_{jk})} - 1 \right) \right)$$

and ξ_1, \dots, ξ_N are independent Bernoulli random variables with probability of success equal to f_1 . We have $|E[e^{iu f_{2j}^{-1} y_{jk} \xi_j + i\tau \xi_j}]|^2 = 1 - 2f_{2j}g_{2j}(1 - \cos(u f_{2j}^{-1} y_{jk} + \tau))$, and under Condition C we get

$$\begin{aligned} \prod_{k=1}^{M_j} |\psi_{jk}(u, \tau)| &\leq \exp \left(-f_{2j}g_{2j} \sum_{k=1}^{M_j} (1 - \cos(u f_{2j}^{-1} y_{jk} + \tau)) \right) \\ &\leq \exp(-c(\epsilon, \delta) M_j f_{2j} g_{2j}) \\ &= \exp(-c(\epsilon, \delta) m_j g_{2j}). \end{aligned} \quad (\text{A7})$$

By Lemma 1 in Höglund (1978) we have

$$\frac{\pi^{1/2}}{2} \leq (2\pi m_j g_{2j})^{1/2} \left(\binom{M_j}{m_j} f_{2j}^{m_j} g_{2j}^{M_j - m_j} \right) < 1. \quad (\text{A8})$$

By combining (A6), (A7), and (A8), we see that

$$\left| E \left[\exp \left(iu \hat{Y}_j \right) \right] \right| \leq (8m_j g_{2j})^{1/2} \exp(-c(\epsilon, \delta) m_j g_{2j}) \quad (\text{A9})$$

for $f_1^{-1} \delta'_1 \leq u \leq f_1^{-1} \delta_2$, implying that

$$\chi(\delta'_1, \delta_2) \leq \varrho. \quad (\text{A10})$$

The proof of the theorem is concluded by combining (A1), (A4), (A5), and (A10). ■

Proof of Theorem 2. Denote the statistic in (14) by $T_{\text{str},m}$, to reflect its dependence on $m = \sum_{j=1}^N m_j$. We will derive formally a two-term asymptotic expansion for $P\{T_{\text{str},m} < u\}$ of the form

$$\Phi(u) + \frac{\phi(u)}{m^{1/2}} \psi_m(u). \quad (\text{A11})$$

Assume for the moment that $E[T_{\text{str},m}] \rightarrow 0$ and $\text{Var}[T_{\text{str},m}] \rightarrow 1$ as $m \rightarrow \infty$ (we will verify that this assumption holds true at the end of the proof). By Hall (1992, Section 2.4), see also Sugden & Smith (1997, Section 2), it follows that if the cumulants $\kappa_{j,m}$ of $T_{\text{str},m}$ are of order $m^{-(j-2)/2}$, $j = 1, 2, 3$, and can be expressed as $\kappa_{j,m} = m^{-(j-2)/2} (a_{j,1} + m^{-1}a_{j,2} + \dots)$, where $a_{1,1} = 0$ and $a_{2,1} = 1$ by assumption, then only $a_{1,2}$ and $a_{3,1}$ contribute to the $O(m^{-1/2})$ term in (A11), and

$$\psi_m(u) = -\left(a_{1,2} + \frac{a_{3,1}}{6}(u^2 - 1)\right). \quad (\text{A12})$$

We derive asymptotic expansions for the cumulants of $T_{\text{str},m}$ by using a procedure somewhat similar to that outlined in Sugden & Smith (1997, Section 2).

Put $W_{jk} = (y_{jk} - \bar{Y}_j) v_{2j}^{-1/2}$ and $V_{jk} = W_{jk}^2 - 1$, and define the sample means $\bar{w}_j = m_j^{-1} \sum_{k \in s_j} W_{jk}$ and $\bar{v}_j = m_j^{-1} \sum_{k \in s_j} V_{jk}$. Then $(m_j - 1)S_j^2 = m_j v_{2j} (1 + \bar{v}_j - \bar{w}_j^2)$ and $E[\bar{w}_j] = E[\bar{v}_j] = 0$, and both \bar{w}_j and \bar{v}_j are $O(m_j^{-1/2})$. Hence, using that $\hat{Y}_{\text{str}} - Y = \sum_{j=1}^N (\hat{Y}_j - Y_j) = \sum_{j=1}^N M_j(\bar{y}_j - \bar{Y}_j) = \sum_{j=1}^N f_{2j}^{-1} \sum_{k \in s_j} (y_{jk} - \bar{Y}_j) = \sum_{j=1}^N m_j f_{2j}^{-1} v_{2j}^{1/2} \bar{w}_j$, we obtain the following stochastic expansion of $T_{\text{str},m}$ (and its powers to an appropriate order):

$$\begin{aligned} T_{\text{str},m} &= \frac{\hat{Y}_{\text{str}} - Y}{\left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} (1 + \bar{v}_j - \bar{w}_j^2)\right)^{1/2}} \\ &= \frac{\sum_{j=1}^N m_j f_{2j}^{-1} v_{2j}^{1/2} \bar{w}_j}{\left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}\right)^{1/2}} \left(1 - \frac{\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} (\bar{v}_j - \bar{w}_j^2)}{2 \sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}}\right) \\ &\quad + \frac{3 \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} (\bar{v}_j - \bar{w}_j^2)\right)^2}{8 \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}\right)^2} + \dots \\ &= \frac{\sum_{j=1}^N m_j f_{2j}^{-1} v_{2j}^{1/2} \bar{w}_j}{\left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}\right)^{1/2}} - \frac{\sum_{j=1}^N m_j f_{2j}^{-1} v_{2j}^{1/2} \bar{w}_j \sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} \bar{v}_j}{2 \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}\right)^{3/2}} \\ &\quad + \frac{3 \sum_{j=1}^N m_j f_{2j}^{-1} v_{2j}^{1/2} \bar{w}_j \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} \bar{v}_j\right)^2}{8 \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}\right)^{5/2}} \\ &\quad + \frac{\sum_{j=1}^N m_j f_{2j}^{-1} v_{2j}^{1/2} \bar{w}_j \sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} \bar{w}_j^2}{2 \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}\right)^{3/2}} + \dots \end{aligned}$$

$$= \frac{\sum_{j=1}^N m_j f_{2j}^{-1} v_{2j}^{1/2} \bar{w}_j}{\left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}\right)^{1/2}} - \frac{\sum_{j=1}^N m_j f_{2j}^{-1} v_{2j}^{1/2} \bar{w}_j \sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} \bar{v}_j}{2 \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}\right)^{3/2}} + O_P\left(m^{-3/2}\right).$$

Note, only the terms which will contribute to $a_{1,2}$ and $a_{3,1}$ are given explicitly on the right-hand side above. Also, note that $\bar{v}_1, \dots, \bar{v}_N$ and $\bar{w}_1, \dots, \bar{w}_N$ are two sequences of independent random variables, and that \bar{v}_j and $\bar{w}_{j'}, j \neq j'$, are independent random variables.

Below we will focus on the leading terms in the asymptotic expansions, and the notation $l.t.(x)$ will be used to refer to the leading term of a quantity x . Recalling that $E[\bar{w}_j] = 0$, we have

$$l.t.(E[T_{\text{str},m}]) = -\frac{\sum_{j=1}^N m_j^2 f_{2j}^{-3} g_{2j} v_{2j}^{3/2} E[\bar{w}_j \bar{v}_j]}{2 \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}\right)^{3/2}}, \quad (\text{A13})$$

$$l.t.(E[T_{\text{str},m}^2]) = \frac{E\left[\left(\sum_{j=1}^N m_j f_{2j}^{-1} v_{2j}^{1/2} \bar{w}_j\right)^2\right]}{\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}} = \frac{\sum_{j=1}^N m_j^2 f_{2j}^{-2} v_{2j} E[\bar{w}_j^2]}{\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}}, \quad (\text{A14})$$

and

$$l.t.(E[T_{\text{str},m}^3]) = \frac{E\left[\left(\sum_{j=1}^N m_j f_{2j}^{-1} v_{2j}^{1/2} \bar{w}_j\right)^3\right]}{\left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}\right)^{3/2}} - \frac{3E\left[\left(\sum_{j=1}^N m_j f_{2j}^{-1} v_{2j}^{1/2} \bar{w}_j\right)^3 \sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} \bar{v}_j\right]}{2 \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j}\right)^{5/2}}. \quad (\text{A15})$$

Let $\eta_{jk} = 1$ if y_{jk} appears in the sample s_j , and 0 otherwise. Note that $P(\eta_{jk} = 1) = f_{2j}$, and that the sample means \bar{w}_j and \bar{v}_j may be written as

$$\bar{w}_j = \frac{1}{m_j} \sum_{k=1}^{M_j} W_{jk} \eta_{jk} \quad \text{and} \quad \bar{v}_j = \frac{1}{m_j} \sum_{k=1}^{M_j} V_{jk} \eta_{jk},$$

which is useful when deriving the expected values of $\bar{w}_j^a \bar{v}_j^b$, where a and b are non-negative integers. In particular, we obtain

$$\begin{aligned} E[\bar{w}_j \bar{v}_j] &= \frac{1}{m_j^2} \sum_{k'=1}^{M_j} \sum_{k=1}^{M_j} W_{jk'} (W_{jk}^2 - 1) E[\eta_{jk'} \eta_{jk}] \\ &= \frac{m_j - 1}{m_j M_j (M_j - 1)} \sum_{k'=1, k' \neq k}^{M_j} \sum_{k=1}^{M_j} W_{jk'} (W_{jk}^2 - 1) + \frac{1}{m_j M_j} \sum_{k=1}^{M_j} W_{jk} (W_{jk}^2 - 1) \\ &= -\frac{m_j - 1}{m_j M_j (M_j - 1)} \sum_{k=1}^{M_j} W_{jk} (W_{jk}^2 - 1) + \frac{1}{m_j M_j} \sum_{k=1}^{M_j} W_{jk} (W_{jk}^2 - 1), \end{aligned}$$

where we have used that $\bar{W}_j = M_j^{-1} \sum_{k=1}^{M_j} W_{jk} = 0$ and $\bar{V}_j = M_j^{-1} \sum_{k=1}^{M_j} V_{jk} = 0$, and thus

$$l.t.(E[\bar{w}_j \bar{v}_j]) = \frac{g_{2j}}{m_j} \frac{v_{3j}}{v_{2j}^{3/2}}. \quad (A16)$$

In the same vein, we can find the leading terms of $E[\bar{w}_j^2]$, $E[\bar{w}_j^3]$, and $E[\bar{w}_j^3 \bar{v}_j]$. A simpler option is to use results of Finucan, Galbraith, & Stone (1974, p. 152), from which it follows that

$$l.t.(E[\bar{w}_j^2]) = \frac{g_{2j}}{m_j} \quad \text{and} \quad l.t.(E[\bar{w}_j^3]) = \frac{g_{2j}(1-2f_{2j})}{m_j^2} \frac{v_{3j}}{v_{2j}^{3/2}}, \quad (A17)$$

and Theorem 2 of Nath (1968), from which we obtain

$$l.t.(E[\bar{w}_j^3 \bar{v}_j]) = \frac{3g_{2j}^2}{m_j^2} \frac{v_{3j}}{v_{2j}^{3/2}}. \quad (A18)$$

Thus, by applying (A16) and (A17) in (A13) and (A14), respectively, we see that

$$l.t.(E[T_{\text{str},m}]) = -\frac{\sum_{j=1}^N m_j f_{2j}^{-3} g_{2j}^2 v_{3j}}{2 \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} \right)^{3/2}} = \frac{a_{1,2}}{m^{1/2}}, \quad (A19)$$

where

$$a_{1,2} = -\frac{m^{-1} \sum_{j=1}^N m_j f_{2j}^{-3} g_{2j}^2 v_{3j}}{2 \left(m^{-1} \sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} \right)^{3/2}}, \quad (A20)$$

and

$$E[T_{\text{str},m}^2] = 1 + o(1). \quad (A21)$$

Consider the numerator in the first term on the right-hand side of (A15). By (A17), and by noting that \bar{w}_j , $j = 1, \dots, N$, are independent random variables, we get

$$E \left[\left(\sum_{j=1}^N \frac{m_j v_{2j}^{1/2}}{f_{2j}} \bar{w}_j \right)^3 \right] = \sum_{j=1}^N \frac{m_j g_{2j} (1-2f_{2j}) v_{3j}}{f_{2j}^3}. \quad (A22)$$

Next, consider the numerator in the second term on the right-hand side of (A15). We have

$$\begin{aligned} & E \left[\left(\sum_{j=1}^N \frac{m_j v_{2j}^{1/2}}{f_{2j}} \bar{w}_j \right)^3 \sum_{j=1}^N \frac{m_j g_{2j} v_{2j}}{f_{2j}^2} \bar{v}_j \right] \\ &= \sum_{j=1}^N \frac{m_j^4 g_{2j} v_{2j}^{5/2}}{f_{2j}^5} E[\bar{w}_j^3 \bar{v}_j] + 3E \left[\sum_{j'=1, j' \neq j}^N \sum_{j=1}^N \frac{m_{j'}^2 v_{2j'}}{f_{2j'}^2} \frac{m_j v_{2j}^{1/2}}{f_{2j}} \bar{w}_{j'}^2 \bar{w}_j \sum_{l=1}^N \frac{m_l g_{2l} v_{2l}}{f_{2l}^2} \bar{v}_l \right] \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^N \frac{m_j^4 g_{2j} v_{2j}^{5/2}}{f_{2j}^5} E[\bar{w}_j^3 \bar{v}_j] + 3 \sum_{j'=1, j' \neq j}^N \sum_{j=1}^N \frac{m_{j'}^2 v_{2j'}}{f_{2j'}^2} \frac{m_j^2 g_{2j} v_{2j}^{3/2}}{f_{2j}^3} E[\bar{w}_{j'}^2] E[\bar{v}_j \bar{w}_j] \\
&= \sum_{j=1}^N \frac{m_j^4 v_{2j}^{5/2}}{f_{2j}^5} g_{2j} E[\bar{w}_j^3 \bar{v}_j] \\
&\quad + 3 \left(\sum_{j=1}^N \frac{m_j^2 v_{2j}}{f_{2j}^2} E[\bar{w}_j^2] \sum_{j'=1}^N \frac{m_{j'}^2 g_{2j'} v_{2j'}^{3/2}}{f_{2j'}^3} E[\bar{v}_{j'} \bar{w}_{j'}] - \sum_{j=1}^N \frac{m_j^4 g_{2j} v_{2j}^{5/2}}{f_{2j}^5} E[\bar{w}_j^2] E[\bar{v}_j \bar{w}_j] \right),
\end{aligned}$$

and from this, together with (A16), (A17), and (A18), we obtain the leading term of the second term on the right-hand side of (A15),

$$\frac{9 \sum_{j=1}^N m_j f_{2j}^{-3} g_{2j}^2 v_{3j}}{2 \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} \right)^{3/2}}.$$

This together with (A15) and (A22) yield

$$l.t.(E[T_{\text{str},m}^3]) = \frac{\sum_{j=1}^N m_j f_{2j}^{-3} (1 - 2f_{2j}) g_{2j} v_{3j}}{\left(\sum_{j=1}^N m_j g_{2j} f_{2j}^{-2} v_{2j} \right)^{3/2}} - \frac{9 \sum_{j=1}^N m_j f_{2j}^{-3} g_{2j}^2 v_{3j}}{2 \left(\sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} \right)^{3/2}}. \quad (\text{A23})$$

By (A19) and (A21), $E[T_{\text{str},m}] \rightarrow 0$ and $\text{Var}[T_{\text{str},m}] \rightarrow 1$ as $m \rightarrow \infty$, and we have $\kappa_{1,m} = E[T_{\text{str},m}]$ and $\kappa_{2,m} = E[T_{\text{str},m}^2] - (E[T_{\text{str},m}])^2$, where $a_{1,1} = 0$ and $a_{2,1} = 1$. We have $\kappa_{3,m} = E[T_{\text{str},m}^3] - 3E[T_{\text{str},m}^2]E[T_{\text{str},m}] + 2(E[T_{\text{str},m}])^3$, and by (A19) and (A21), $l.t.(\kappa_{3,m}) = E[T_{\text{str},m}^3] - 3E[T_{\text{str},m}]$. By (A19) and (A23) we obtain

$$\begin{aligned}
a_{3,1} &= \frac{m^{-1} \left(\sum_{j=1}^N m_j f_{2j}^{-3} g_{2j} (1 - 2f_{2j}) v_{3j} - 3 \sum_{j=1}^N m_j f_{2j}^{-3} g_{2j}^2 v_{3j} \right)}{\left(m^{-1} \sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} \right)^{3/2}} \\
&= - \frac{m^{-1} \sum_{j=1}^N m_j f_{2j}^{-3} g_{2j} (2 - f_{2j}) v_{3j}}{\left(m^{-1} \sum_{j=1}^N m_j f_{2j}^{-2} g_{2j} v_{2j} \right)^{3/2}}. \quad (\text{A24})
\end{aligned}$$

Thus, (A11), (A12), (A20), and (A24) imply the desired result. ■

ACKNOWLEDGEMENTS

The authors would like to thank the Associate Editor and the referees for their many helpful comments and suggestions. The research of M. Ekström was supported by a grant from the Swedish Research Council.

BIBLIOGRAPHY

- Babu, G. J. & Singh, K. (1985). Edgeworth expansions for sampling without replacement from finite populations. *Journal of Multivariate Analysis*, 17, 261–278.
- Bentkus, V., Bloznelis, M., & Götze, F. (1996). The Berry-Esseen bound for Student's statistic in the non-i.i.d. case. *Journal of Theoretical Probability*, 9, 765–796.

- Bickel, P. J. & Freedman, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Annals of Statistics*, 12, 470–482.
- Bickel, P. J. & van Zwet, W. R. (1978). Asymptotic expansions for the power of distribution-free tests in the two-sample problem. *Annals of Statistics*, 6, 937–1004.
- Blznelis, M. (2000). One- and two-term Edgeworth expansions for a finite population sample mean. Exact results. I. *Lithuanian Mathematical Journal*, 40, 213–227.
- Cho, E. C., Cho, M. J., & Eltinge, J. L. (2005). The variance of sample variance from a finite population. *The International Journal of Pure and Applied Mathematics*, 21, 387–394.
- Cochran, W. G. (1977). *Sampling Techniques*, 3rd ed., John Wiley, New York.
- Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge.
- Erdős, P. & Rényi, A. (1959). On the central limit theorem for samples from a finite population. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 4, 49–61.
- Finucan, H. M., Galbraith, R. F., & Stone, M. (1974). Moments without tears in simple random sampling from a finite population. *Biometrika*, 61, 151–154.
- Fletcher, D. & Webster, R. (1996). Skewness-adjusted confidence intervals in stratified biological surveys. *Journal of Agricultural Biological and Environmental Statistics*, 1, 120–130.
- Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, 361–374.
- Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite populations. *Annals of Mathematical Statistics*, 35, 1491–1523.
- Hall, P. (1987). Edgeworth expansion for Student's t statistic under minimal moment conditions. *Annals of Probability*, 15, 920–931.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.
- Hall, P. & Wang, Q. (2004). Exact convergence rate and leading term in central limit theorem for Student's t statistic. *Annals of Probability*, 32, 1419–1437.
- Hu, Z., Robinson, J., & Wang, Q. (2007). Edgeworth expansions for a sample sum from a finite set of independent random variables. *Electronic Journal of Probability*, 12, 1402–1417.
- Ibrahim, B. M. & Mirakhmedov, S. M. (2013). Approximation by normal distribution for a sample sum in sampling without replacement from a finite population. arXiv:1312.3256
- Krewski, D. & Rao, J. N. K. (1981). Inference from stratified samples: Properties of linearization, jackknife and balanced repeated replication methods. *Annals of Statistics*, 9, 1010–1019.
- Mirakhmedov, S. A. (1983). An asymptotic expansion for a sample sum from a finite sample. *Theory of Probability and Its Applications*, 28, 492–502.
- Mirakhmedov, S. M., Jammalamadaka, S. R., & Ibrahim, B. M. (2014). On Edgeworth expansions in generalized urn models. *Journal of Theoretical Probability*, 27, 725–753.
- Nath, S. N. (1968). On product moments from a finite universe. *Journal of the American Statistical Association*, 63, 535–541.
- Prášková, Z. (1984). On the rate of convergence in Samford-Durbin sampling from a finite population. *Statistics & Decisions*, 2, 339–350.
- Rao, J. N. K. (1973). Double sampling for stratification and analytical surveys. *Biometrika*, 60, 125–133.
- Robinson, J. (1978). An asymptotic expansion for samples from a finite population. *Annals of Statistics*, 6, 1004–1011.
- Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer, New York.
- Schenker, N. & Welsh, A. H. (1988). Asymptotic results for multiple imputation. *Annals of Statistics*, 16, 1550–1566.
- Scott, A. J. & Wu, C. F. J. (1981). On the asymptotic distribution of ratio and regression estimators. *Journal of the American Statistical Association*, 76, 98–102.

- Sen, P. K. (1988). Asymptotics in finite population sampling. *Handbook of Statistics*, 6, 291–331.
- Sugden, R. A. & Smith, T. M. F. (1997). Edgeworth approximations to the distribution of the sample mean under simple random sampling. *Statistics & Probability Letters*, 34, 293–299.
- von Bahr, B. (1972). On sampling from a finite set of independent random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 24, 279–286.
-

Received 20 July 2014

Accepted 15 August 2015