Probab. Th. Rel. Fields 81, 341–352 (1989)

Probability Theory and Related Fields

Asymptotic Normality for Two-Stage Sampling from a Finite Population

Esbjörn Ohlsson

Department of Mathematics, Royal Institute of Technology, S-10044 Stockholm, Sweden

Summary. In this paper we consider two-stage sampling from a finite population, and associated estimators of the population total, in a general setting which includes most two-stage procedures in the literature. The main result gives general conditions for asymptotic normality of the estimators. The proof is based on a martingale central limit theorem. It is indicated how the result can be extended to multi-stage procedures.

0. Introduction and Outline of the Paper

Two-stage sampling is a well established technique for sampling from a finite population; another term is "cluster sampling with subsampling". Most statisticians would conjecture that the corresponding estimator of the population total, being a sum of random variables, is approximately normally distributed when the first-stage sample size is large. When the variables forming the estimator are independent, this conjecture may be verified by application of the ordinary Liapunov central limit theorem, see Fuller (1975). However, in most cases the variables are dependent and hence the ordinary central limit theorem can not be directly applied. For an important special instance of the dependent case, viz. two-stage procedures with "succesive sampling with varying probabilities without replacement" at the first stage, asymptotic normality has been established by Sen (1980). To the best of our knowledge, no proof of asymptotic normality can be found in the literature when other procedures are used at the first stage.

For some single-stage procedures (which might be used at the first stage) asymptotic normality has been proved by different authors, for a list see Remark 2.2. However, it is not self-evident how these results can be generalized to two-stage sampling, where further randomness is introduced by sub-sampling. In the main theorem of the present paper general conditions are given, under which asymptotic normality for the first-stage procedure (regarded as a single-stage procedure) ensures asymptotic normality of the estimator in two-stage sampling. The proof of the theorem utilizes a martingale central limit theorem.

Here is an outline of the paper. In the next section we give a detailed description of two-stage sampling. Section 2 contains our main result, Theorem 2.1.

We also indicate how this result can be used to prove asymptotic normality for multi-stage procedures with more than two stages. Theorem 2.1 is proved in Sect. 3; the martingale central limit theorem used in the proof is stated in an appendix.

Theorem 2.1 is applicable to quite a wide class of two-stage sampling procedures. As an illustration, in Sect. 4 we discuss cases where simple random sampling without replacement is used at the second sampling stage. Throughout the paper, special attention is paid to the two-stage analogue of the Horvitz-Thompson estimator.

1. Two-Stage Sampling

We consider a finite population with N *primary stage units* (psu) labelled $1, 2, \ldots, N$. Each psu is a cluster of *second stage units* (ssu), the i^{th} psu consisting of M_i ssu's, $i=1, 2, \ldots, N$. To each ssu there is associated a variate value (i.e. a real number), denoted a_{ij} for the j^{th} ssu within the i^{th} psu, $i=1, 2, \ldots, N$, $j=1, 2, \ldots, M_i$. The corresponding *psu totals* and *psu means* are

$$a_{i\cdot}=\sum_{j=1}^{M_i} a_{ij}, \qquad \bar{a}_{i\cdot}=a_{i\cdot}/M_i, \quad i=1, 2, \ldots, N, \tag{1.1}$$

and the *population total* and *mean per psu* are

$$a_{\cdot\cdot}=\sum_{i=1}^{N} a_{i\cdot}, \qquad \bar{a}_{\cdot\cdot}=a_{\cdot\cdot}/N. \tag{1.2}$$

The a-values are supposed to be unknown; in order to estimate their total $a_{\cdot\cdot}$ a two-stage sampling procedure is employed. We assume that this procedure is of the following type (as most two-stage procedures in the literature in fact are).

In the first stage a sample I (i.e., a collection of integers from $\{1, 2, \ldots, N\}$) of psu's is generated by some arbitrary but specified sampling procedure Π_1. A psu may occur several times in I, in particular Π_1 may be a "with replacement" procedure.

To each psu there is associated a drawing procedure, denoted Π_{2i} for the i^{th} psu. In the second stage of the two-stage procedure a subsample from the i^{th} psu may be generated by Π_{2i}. In many cases, e.g., those in Sect. 4, Π_{2i} is completely determined before the execution of Π_1. However, in our general setting we include the possibility that Π_{2i} depends on the outcome of Π_1, i.e., Π_{2i} is random (see Remark 1.1 below for an example).

For each i let T_i be an estimator of $a_{i\cdot}$ based on the a-values of the ssu's drawn by Π_{2i}. In practice Π_{2i} is executed and T_i is observed only for $i\in I$.

Let $\mathscr{G}$ be a σ-algebra such that I is $\mathscr{G}$-measurable. We shall think of $\mathscr{G}$ as containing all information on the first stage procedure. A natural choice is $\mathscr{G}=\sigma(I)$, i.e., the σ-algebra generated by I, but this choice is not always adequate (see Remark 1.3).

The T_i's are assumed to be conditionally independent given $\mathscr{G}$ and to be conditionally unbiased, i.e.,

$$E(T_i|\mathscr{G})=a_{i\cdot}, \qquad i=1, 2, \ldots, N. \tag{1.3}$$

We denote the (possibly random) conditional variance and fourth moment of T_i by

$$\sigma_i^2=V(T_i|\mathscr{G}), \qquad \mu_i^{(4)}=E[(T_i-a_{i\cdot})^4|\mathscr{G}], \quad i=1, 2, \ldots, N. \tag{1.4}$$

Remark 1.1. One example of a random Π_{2i} is simple random sampling with a sample size m_i which depends on the first-stage sample I; this is the case e.g., if m_i is to be chosen proportional to M_i, $i\in I$, and the total sample size $\sum_{i\in I} m_i$ is to be kept fixed. If in this case T_i is the "inflated" sample mean then (1.3) is fulfilled; note that σ_i^2 and $\mu_i^{(4)}$ are random here.

We consider estimators of the population total $a_{\cdot\cdot}$ of the following form, where $W_1, W_2, \ldots, W_N$ are $\mathscr{G}$-measurable random variables such that $W_i=0$ for $i\notin I$,

$$T=\sum_{i=1}^{N} W_i T_i=\sum_{i\in I} W_i T_i. \tag{1.5}$$

Thus W_i is a (random) weight for the i^{th} psu. The $\mathscr{G}$-measurability means that W_i only is allowed to depend on the outcome of the first stage sampling. As noted above, T_i is observed only for $i\in I$; hence the requirement that $W_i=0$ for $i\notin I$.

We will refer to a population and a two-stage procedure of the above type together as a *two-stage situation.*

Remark 1.2. As an illustration of the choice of W_i we mention the familiar two-stage analogue of the Horvitz-Thompson estimator. Suppose that the psu's are drawn without replacement with inclusion probabilities $\pi_i=P(i\in I)>0$. With $1\{\cdot\}$ denoting the indicator function, let $W_i=1\{i\in I\}/\pi_i$. For these W_i's, T takes the form

$$T=\sum_{i=1}^{N} 1\{i\in I\}\frac{T_i}{\pi_i}=\sum_{i\in I}\frac{T_i}{\pi_i}, \tag{1.6}$$

which we shall call just *the Horvitz-Thompson estimator.* For further discussion of this case see Remarks 2.4 and 4.2.

Remark 1.3. Suppose that the psu's are drawn by the RHC procedure (Rao, Hartley, and Cochran, 1962). Then the first-stage procedure contains a random stratification of the psu's which is performed before the actual drawing of the first-stage sample. Thereafter one psu is drawn from each stratum with probability proportional to some variable p_i, $i=1, 2, \ldots, N$. Let $\tilde{P}_i$ be the total of the p-values in the stratum to which the i^{th} psu belongs. Then the usual two-stage RHC estimator is given by (1.5) with $W_i=1\{i\in I\}\,\tilde{P}_i/p_i$. Here W_i depends on both the first-stage sample I and on the random stratification. If $\mathscr{G}$ is the σ-algebra generated by I and the stratification together, then W_i is $\mathscr{G}$-measurable but not $\sigma(I)$-measurable. Hence, in order that the above general description of a two-stage procedure and our results below should apply to this case, the setting with a general $\mathscr{G}$ is required. The setting usually found in the literature, e.g. in Raj (1966) and Rao (1975), allows the W_i's to depend only on I. Hence Raj's and Rao's applications of their results to the RHC case are not quite justified.

Next we introduce a decomposition of T, which will be fundamental in our discussion of asymptotic normality. Set

$$Q=\sum_{i=1}^{N} W_i a_{i\cdot}, \tag{1.7}$$

$$M=\sum_{i=1}^{N} W_i(T_i-a_{i\cdot}). \tag{1.8}$$

The decomposition in question is

$$T=Q+M. \tag{1.9}$$

As will be shown in Sect. 3, M is a sum of martingale differences. Hence we have $E(M)=0$ and

$$E(T)=E(Q)=\sum_{i=1}^{N} E(W_i)\, a_{i\cdot}. \tag{1.10}$$

From (1.10) it is seen that T is unbiased for $a_{\cdot\cdot}$ if and only if $E(W_i)=1$, $i=1, 2, \ldots, N$. For the variance of T we have the following well-known formula,

$$V(T)=V(Q)+\sum_{i=1}^{N} E(W_i^2\, \sigma_i^2). \tag{1.11}$$

A version of the last result was derived by Raj (1966) and extended by Rao (1975). However, our version is slightly more general than these two, as seen from Remark 1.3. In Sect. 3 we will derive (1.11) as a by-product of other results.

Suppose that instead of subsampling from the psu's we could measure the $a_{i\cdot}$'s exactly, i.e., $T_i \equiv a_{i\cdot}$. Then by (1.7)–(1.9) we would have $T=Q$. Hence Q can be interpreted as an estimator of the population total based on single-stage sampling according to the procedure Π_1. Hereby (1.11) can be given the following interpretation. In order to get the variance of T for the two-stage procedure, first find the variance of the estimator based on the first stage procedure, i.e., $V(Q)$. Then find the second-stage variances, σ_i^2, and combine them with the first-stage quantities W_i^2, take expectations, and obtain $V(T)$ by adding everything. In the next section we will give a somewhat similar result on the asymptotic normality of T. There the recipe is: first prove asymptotic normality of Q, next verify two conditions concerning expectations of some functions of σ_i^2, $\mu_i^{(4)}$ and W_i; then the asymptotic normality of T will follow.

2. The Main Result

Let $\mathscr{L}(\cdot)$ denote probability distribution and $N(\mu, \sigma^2)$ the normal distribution with mean μ and variance σ^2. The aim of this paper is to justify the approximation

$$\mathscr{L}(T) \approx N(E(T), V(T)), \tag{2.1}$$

by asymptotic considerations. In order to give our results a stringent formulation we consider a sequence of two-stage situations, i.e., for each k ($k=1, 2, 3, \ldots$) we have a two-stage situation as described in Sect. 1. By attaching the sequence generating index k to any quantity we will indicate that it relates to situation k. To prepare for the presentation of our main result we present three conditions.

Let $C(\cdot, \cdot)$ denote covariance and $\xrightarrow{d}$ denote convergence in distribution.

(C1)
$$\mathscr{L}\left(\frac{Q_k - E(Q_k)}{\sqrt{V(Q_k)}}\right) \xrightarrow{d} N(0,1) \quad \text{as} \quad k \to \infty, \tag{2.2}$$

(C2)
$$\lim_{k\to\infty} \frac{\sum_i E(W_{ki}^4 \mu_{ki}^{(4)})}{V(T_k)^2} = 0, \tag{2.3}$$

(C3)
$$\overline{\lim_{k\to\infty}} \frac{\sum_{i\neq j} C(W_{ki}^2 \sigma_{ki}^2, W_{kj}^2 \sigma_{kj}^2)}{V(T_k)^2} \leqq 0. \tag{2.4}$$

The conditions are discussed after the next theorem, which is the main result of the paper.

Theorem 2.1. *For a sequence of two-stage situations which satisfies the conditions* (C1)–(C3) *we have*

$$\mathscr{L}\left(\frac{T_k - E(T_k)}{\sqrt{V(T_k)}}\right) \xrightarrow{d} N(0,1) \quad \text{as} \quad k \to \infty. \tag{2.5}$$

The proof of the theorem will be given in Sect. 3.

Remark 2.2. By the discussion at the end of the preceding section, condition (C1) is seen to require asymptotic normality of an estimator of $a_{..}$ based on single-stage sampling. For some particular (single-stage) procedures conditions for asymptotic normality can be found in the literature. Below we list the (without replacement) cases that are known to us; we do not claim the list to be complete. For simple random sampling without replacement conditions are found in e.g., Hájek (1960) and Hájek (1961); for rejective sampling in Hájek (1964) and for "random replacement sampling" in Rosén (1967). Successive sampling with varying probabilities without replacement using different estimators is treated in several papers: with the sample total as estimator (this actually includes the Horvitz-Thompson estimator) in Rosén (1972), this result is generalized in Holst (1973) and also treated in Gordon (1983); with Des Raj's estimator in Rosén (1974) and with Murthy's estimator in Högfeldt (1980). The Rao-Hartley-Cochran procedure is considered in Ohlsson (1986a).

In order to facilitate applications of Theorem 2.1 we now give some sufficient conditions for (C2) and (C3) to hold.

Proposition 2.3. *Assume that* σ_{ki}^2 *and* $\mu_{ki}^{(4)}$ *are non-random, cf.* (1.4). (a) *If*

$$\lim_{k\to\infty} \frac{\sum_i \mu_{ki}^{(4)} E(W_{ki}^4)}{(\sum_i \sigma_{ki}^2 E(W_{ki}^2))^2} = 0, \tag{2.6}$$

then (C2) *holds true.*

(b) *If*

$$C(W_{ki}^2, W_{kj}^2) \leqq 0, \qquad i \neq j, \quad k = 1, 2, 3, \ldots, \tag{2.7}$$

then (C3) *is fulfilled.*

The result in (a) follows readily from (2.3) and (1.11). Part (b) is trivial but nevertheless useful: since W_{ki} is zero if $i \notin I$ one may expect (2.7) to be true in many cases. For example, when the first-stage procedure is simple random sampling without replacement, (2.7) is valid.

Remark 2.4. For a two-stage situation with the Horvitz-Thompson estimator (introduced in Remark 1.2) set $\pi_{ij} = P(i, j \in I)$. Here it is readily seen that

$$C(W_i^2, W_j^2) = (\pi_{ij} - \pi_i \pi_j)/(\pi_i^2 \pi_j^2).$$

Hence the covariances are non-positive, as required in (2.7), if and only if

$$\pi_{ij} \leqq \pi_i \pi_j, \qquad i \neq j. \tag{2.8}$$

The validity of (2.8) for different sampling procedures has been investigated by many authors in connection with discussions on the Yates-Grundy variance estimator (Cochran, 1977, p. 261). The reason for this is that (2.8) is sufficient also for the Yates-Grundy estimator to be non-negative, which is a desired property of variance estimators. For references and some results see Lanke (1974) where also an example showing that (2.8) does not always hold true is given.

For illustrations of the "Liapunov type" condition (C2) (or rather (2.6)) see Sect. 4.

Remark 2.5. Although Theorem 2.1 only deals with two-stage procedures it can be used to prove asymptotic normality for arbitrary multi-stage procedures as indicated below. The quality of the theorem is, as noted earlier, that given asymptotic normality of the first-stage procedure (i.e., given (C1)) we get asymptotic normality of the two-stage procedure by imposing the conditions (C2) and (C3). Now there is no restriction against the first-stage procedure actually being a multi-stage procedure itself. Hence Theorem 2.1 can be used inductively to prove asymptotic normality of procedures with any number of stages, each stage requiring its own versions of the conditions (C2) and (C3).

3. Proof of Theorem 2.1

First we derive some auxiliary results. For this purpose we return to the case with just one two-stage situation so that we may drop the k's. We introduce some further notation. With $\mathscr{B}_1 \vee \mathscr{B}_2$ denoting the minimal σ-algebra containing both $\mathscr{B}_1$ and $\mathscr{B}_2$, set

$$\mathscr{F}_0 = \mathscr{G}, \qquad \mathscr{F}_i = \mathscr{G} \vee \sigma(T_1, T_2, \ldots, T_i), \quad i = 1, 2, \ldots, N. \tag{3.1}$$

Note that $\{\mathscr{F}_i;\ i=0, 1, \ldots, N\}$ is a filtration i.e., $\mathscr{F}_{i-1} \subset \mathscr{F}_i$, $i=1, 2, \ldots, n$. Also let

$$X_i = W_i(T_i - a_{i.}). \tag{3.2}$$

Note that by (1.8)

$$M = \sum_i X_i. \tag{3.3}$$

Lemma 3.1. $\{X_i;\ i=1, 2, \ldots, N\}$ *is a martingale difference sequence relative to the filtration* $\{\mathscr{F}_i;\ i=0, 1, \ldots, N\}$, *i.e., each* X_i *is* $\mathscr{F}_i$*-measurable and satisfies*

$$E(X_i \mid \mathscr{F}_{i-1}) = 0. \tag{3.4}$$

Proof. The $\mathscr{F}_i$-measurability follows from the $\mathscr{G}$-measurability of W_i. Since the T_i's are independent given $\mathscr{G}$ we have $E(T_i \mid \mathscr{F}_{i-1}) = E(T_i \mid \mathscr{G})$, which by (1.3) equals $a_{i.}$. This readily yields (3.4).

From Lemma 3.1 and (3.3) we see that $E(M)=0$ as claimed in Sect. 1.

Lemma 3.2.

(a) $$V(T) = V(Q) + V(M). \tag{3.5}$$

(b) $$V(M) = \sum_{i=1}^{N} E(X_i^2). \tag{3.6}$$

Proof. By (1.7), the $\mathscr{G}$-measurability of the W_i's imply that Q is $\mathscr{G}$-measurable. From this fact, (3.3) and (3.4) it is readily seen that M and Q are uncorrelated so that we get (3.5) from (1.9). Since the X_i's are martingale differences they have zero means and are uncorrelated. This and (3.3) yields (3.6).

Note that by (3.2) and (1.4)

$$E(X_i^2 \mid \mathscr{G}) = W_i^2\, E((T_i - a_{i.})^2 \mid \mathscr{G}) = W_i^2\, \sigma_i^2. \tag{3.7}$$

By taking expectations in (3.7) and using (3.5) and (3.6) we get (1.11); this constitutes the proof of (1.11) announced in Sect. 1.

Lemma 3.3.

$$E(X_i^4)=E(W_i^4\,\mu_i^{(4)}),\qquad i=1,2,\ldots,N. \tag{3.8}$$

This result follows directly by using (3.2), conditioning on $\mathscr{G}$ and recalling (1.4).

Lemma 3.4.

$$E\{\sum_i E(X_i^2\,|\,\mathscr{F}_{i-1})-V(M)\}^2\leqq\sum_i E(W_i^4\,\mu_i^{(4)})+\sum_{i\neq j}C(W_i^2\,\sigma_i^2,W_j^2\,\sigma_j^2). \tag{3.9}$$

Proof. Since the T_i's are independent given $\mathscr{G}$ and since W_i is $\mathscr{G}$-measurable, we can replace $\mathscr{F}_{i-1}$ by $\mathscr{G}$ in the left-hand side of (3.9). This fact together with (3.6) yields that the left-hand side of (3.9) equals

$$V\{\sum_i E(X_i^2\,|\,\mathscr{G})\}=\sum_i V\{E(X_i^2\,|\,\mathscr{G})\}+\sum_{i\neq j}C\{E(X_i^2\,|\,\mathscr{G}),E(X_j^2\,|\,\mathscr{G})\}. \tag{3.10}$$

By the elementary formula $V(Y)=E[V(Y\,|\,\mathscr{G})]+V[E(Y\,|\,\mathscr{G})]$ we have

$$V(E(X_i^2\,|\,\mathscr{G}))\leqq V(X_i^2)\leqq E(X_i^4). \tag{3.11}$$

From (3.10), (3.11), (3.8) and (3.7) we now get (3.9).

Proof of Theorem 2.1. In this proof we shall use a version of the martingale central limit theorem which is presented as Theorem A.1 in the appendix. We shall verify that for a sequence of two-stage situations satisfying (C1)–(C3), the conditions of Theorem A.1 are fulfilled with

$$\xi_{ki}=X_{ki}/\sqrt{V(T_k)},\qquad \beta_k^2=V(M_k)/V(T_k),$$
$$Y_k=(Q_k-E(Q_k))/\sqrt{V(T_k)},\qquad \mathscr{L}_0=N(0,1), \tag{3.12}$$

and $\mathscr{F}_{ki}$ as defined in (3.1). By Lemma 3.1, $\{\xi_{ki}\}$ is a martingale difference sequence relative to $\{\mathscr{F}_{ki}\}$. As noted earlier, Q_k is $\mathscr{G}_k$-measurable; hence by (3.1) Y_k is $\mathscr{F}_{k0}$-measurable.

Next we show that (i) and (ii) in Theorem A.1 are implied by (C2) and (C3). By (3.8) condition (C2) is equivalent to (i) in the present case. By (3.5) $\sup_k \beta_k^2\leqq 1<\infty$. From (3.9) it is readily seen that (C2) and (C3) together imply (ii).

By (3.12) and (3.5), (C1) implies (iii) with $\mathscr{L}_0=N(0,1)$.

By (3.12) and (3.3), S_k in (A.1) here takes the form $S_k=M_k/\sqrt{V(T_k)}$. By this, the definition of Y_k in (3.12), (1.9) and (1.10) we have $Y_k+S_k=(T_k-E(T_k))/\sqrt{V(T_k)}$. Since all the conditions of Theorem A.1 are fulfilled we now get (2.5) from (A.5).

4. Applications to a Special Class of Two-Stage Procedures

In this section we shall illustrate the usage of Theorem 2.1, and in particular condition (C2), by considering two-stage procedures where simple random sampling without replacement (henceforth *srs*) is used at the second stage. Such two-stage procedures are often discussed in the literature. First we treat this case without making any additional assumptions on the first-stage procedure, beside those made in Sect. 1. Then we specialize to cases with the Horvitz-Thompson estimator.

We first specify what we mean by assuming that srs is used at the second stage. Let the *second-stage sample sizes* $m_1, m_2, \ldots, m_N$ be fixed integers such that $1 \leqq m_i \leqq M_i$, $i=1, 2, \ldots, N$; in particular the m_i's must not depend on the outcome of the first-stage procedure Π_1. For each i, the outcome of the second-stage procedure Π_{2i} is a size m_i simple random sample J_i of ssu's drawn without replacement from the i^{th} psu. The J_i's are assumed to be mutually independent and independent of $\mathscr{G}$. An unbiased estimator of the psu total $a_{i.}$ is given by

$$T_i = \frac{M_i}{m_i} \sum_{j \in J_i} a_{ij}. \tag{4.1}$$

The variance within psu's is denoted by

$$S_{2i}^2 = \frac{1}{M_i} \sum_{j=1}^{M_i} (a_{ij} - \bar{a}_{i.})^2. \tag{4.2}$$

The division by M_i instead of $M_i - 1$ is made for convenience in the formulas below. The variance σ_i^2 of T_i is well known, see e.g. Cochran (1977, formula (11.13)).

Now let there be given a sequence of two-stage situations with second stage procedures as described above. We are interested in the validity of the conditions (C1)–(C3) of Theorem 2.1. The condition (C1) only concerns the first stage procedure; the same is true for (2.7) which implies (C3). Hence we concentrate on (C2) here and we shall give sufficient conditions for it to hold. We will not try to give the best (weakest) conditions, but rather to give comparatively simple ones.

Proposition 4.1. *Suppose that srs is used at the second stage. If the following two conditions are fulfilled then so is* (C2):

$$0 < \varliminf_{k \to \infty} \min_i \frac{m_{ki}}{M_{ki}} \leqq \varlimsup_{k \to \infty} \max_i \frac{m_{ki}}{M_{ki}} < 1, \tag{4.3}$$

$$\lim_{k \to \infty} \frac{\sum_i \frac{M_{ki}^4}{m_{ki}^2} S_{k2i}^4 E(W_{ki}^4)}{\left(\sum_i \frac{M_{ki}^2}{m_{ki}} S_{k2i}^2 E(W_{ki}^2) \right)^2} = 0. \tag{4.4}$$

This proposition follows from Proposition 2.3(a), see Ohlsson (1986b) for a detailed proof. In that paper we also give a weaker version of (4.3).

Remark 4.2. In practice it appears to be common to draw psu's with varying probabilities (pps) and then draw ssu's with srs as described above. Suppose that the Horvitz-Thompson estimator is used (cf. Remarks 1.2 and 2.4). Then it is readily seen that condition (4.4) is implied by

$$\lim_{k\to\infty} \frac{\max_i \dfrac{M_{ki}^2 S_{k2i}^2}{m_{ki}\pi_{ki}^2}}{\sum_i \dfrac{M_{ki}^2 S_{k2i}^2}{m_{ki}\pi_{ki}}} = 0. \tag{4.5}$$

Here the π_{ki}'s depend on the particular pps procedure which is used. In order to simplify our conditions even further we consider the important special case when the first-stage sample size n_k is prescribed and π_{ki} is proportional to M_{ki}, i.e. $\pi_{ki}=n_k M_{ki}/\sum_i M_{ki}$. In this case (4.5) reduces to

$$\lim_{k\to\infty} \frac{1}{n_k} \frac{(\sum_i M_{ki}) \max_i \dfrac{S_{k2i}^2}{m_{ki}}}{\sum_i M_{ki} \dfrac{S_{k2i}^2}{m_{ki}}} = 0, \tag{4.6}$$

which is fulfilled e.g., when the second-stage sample sizes m_{ki} are equal, the variances S_{k2i}^2 within psu's are of the same magnitude and the first-stage sample size $n_k \to \infty$.

Appendix

Here we state a version of the martingale central limit theorem which is used in the proof of Theorem 2.1.

Theorem A.1. *For* $k=1, 2, 3, \ldots$ *let* $\{\xi_{ki};\ i=1, 2, \ldots, N_k\}$ *be a martingale difference sequence relative to the filtration* $\{\mathscr{F}_{ki};\ i=0, 1, \ldots, N_k\}$ *and let* Y_k *be an* $\mathscr{F}_{k0}$*-measurable random variable. Set*

$$S_k = \sum_{i=1}^{N_k} \xi_{ki}. \tag{A.1}$$

Suppose that the following three conditions are fulfilled.

(i)
$$\lim_{k\to\infty} \sum_{i=1}^{N_k} E(\xi_{ki}^4) = 0. \tag{A.2}$$

(ii) *For some sequence of non-negative real numbers* $\{\beta_k;\ k=1, 2, 3, \ldots\}$, *with* $\sup_k \beta_k < \infty$ *we have*

$$\lim_{k\to\infty} E\left\{\sum_{i=1}^{N_k} E(\xi_{ki}^2 \mid \mathscr{F}_{k,i-1}) - \beta_k^2\right\}^2 = 0. \tag{A.3}$$

(iii) *For some probability distribution* $\mathscr{L}_0$, * *denoting convolution,*

$$\mathscr{L}(Y_k) * N(0, \beta_k^2) \xrightarrow{d} \mathscr{L}_0 \quad as \quad k \to \infty. \tag{A.4}$$

Then we have

$$\mathscr{L}(Y_k + S_k) \xrightarrow{d} \mathscr{L}_0 \quad as \quad k \to \infty. \tag{A.5}$$

We shall not give a complete proof of this result, but merely indicate how it can be derived from the ordinary martingale central limit theorem (see e.g., Hall and Heyde, 1980, Corollary 3.1). By the latter we have under (i) and (ii), in terms of characteristic functions,

$$E\, e^{itS_k} - e^{-\beta_k^2 t^2/2} \to 0 \quad \text{as} \quad k \to \infty. \tag{A.6}$$

Assume that we also know that

$$E\,|E(e^{itS_k} | \mathscr{F}_{k0}) - e^{-\beta_k^2 t^2/2}| \to 0 \quad \text{as} \quad k \to \infty. \tag{A.7}$$

Then (A.5) follows from (A.4) and the $\mathscr{F}_{k0}$-measurability of Y_k since

$$\begin{aligned} &|E\, e^{it(Y_k+S_k)} - E(e^{itY_k})\, e^{-\beta_k^2 t^2/2}| \\ &= |E(e^{itY_k}\{E(e^{itS_k} | \mathscr{F}_{k0}) - e^{-\beta_k^2 t^2/2}\})| \\ &\leqq E\,|E(e^{itS_k} | \mathscr{F}_{k0}) - e^{-\beta_k^2 t^2/2}|. \end{aligned} \tag{A.8}$$

That (A.7) is in fact true under (i) and (ii) can be brought back on the ordinary martingale central limit theorem. A closely related result is given in Eagleson (1975, Corollary 2, p. 560). For a detailed proof of Theorem A.1 we refer the reader to Ohlsson (1985). Since (i) and (ii) imply the conditions (L) and (Q*) there, the present Theorem A.1 follows from Corollary 4.1 in that paper.

Acknowledgement. I am grateful to Professor Bengt Rosén for suggesting the problem and giving valuable comments on the manuscript.

References

1. Cochran, W.G.: Sampling techniques, 3rd ed. New York: Wiley 1977
2. Eagleson, G.K.: Martingale convergence to mixtures of infinitely divisible laws. Ann. Probab. **3**, 557–562 (1975)
3. Fuller, W.A.: Regression analysis for sample survey. Sankhyā C**37**, 117–132 (1975)
4. Gordon, L.: Succesive sampling in large finite populations. Ann. Stat. **11**, 702–706 (1983)
5. Hájek, J.: Limiting distributions in simple random sampling from a finite population. Publ. Math. Inst. Hungar. Acad. Sci. **5**, 361–374 (1960)
6. Hájek, J.: Some extensions of the Wald-Wolfowitz-Noether theorem. Ann. Math. Stat. **32**, 506–523 (1961)
7. Hájek, J.: Asymptotic theory of rejective sampling with varying probabilities from a finite population. Ann. Math. Stat. **35**, 1491–1523 (1964)

8. Hall, P., Heyde, C.C.: Martingale limit theory and its applications. New York: Academic Press 1980
9. Holst, L.: Some limit theorems with applications in sampling theory. Ann. Stat. **1**, 644–658 (1973)
10. Högfeldt, P.: Asymptotic theory for Murthy's estimator. Tech. report TRITA-MAT-1980-13, Royal Institute of Technology, Stockholm (1980)
11. Lanke, J.: On non-negative variance estimators in survey sampling. Sankhyā C**36**, 33–42 (1974)
12. Ohlsson, E.: Some remarks on the martingale central limit theorem. Tech. report TRITA-MAT-1985-19, Royal Institute of Technology, Stockholm (1985)
13. Ohlsson, E.: Asymptotic normality of the Rao-Hartley-Cochran estimator: An application of the martingale CLT. Scand. J. Stat. **13**, 17–28 (1986a)
14. Ohlsson, E.: Asymptotic normality for two-stage sampling from a finite population. Tech. report TRITA-MAT-1986-26, Royal Institute of Technology, Stockholm (1986b)
15. Raj, Des: Some remarks on a simple procedure of sampling without replacement. J. Am. Stat. Assoc. **61**, 391–396 (1966)
16. Rao, J.N.K., Hartley, H.O., Cochran, W.G.: On a simple procedure of unequal probability sampling without replacement. J. Roy. Stat. Soc. Ser. B**24**, 482–491 (1962)
17. Rao, J.N.K.: Unbiased variance estimation for multistage designs. Sankhyā C**37**, 133–139 (1975)
18. Rosén, B.: On the central limit theorem for a class of sampling procedures. Z. Wahrscheinlichkeitstheor. Verw. Geb. **7**, 103–115 (1967)
19. Rosén, B.: Asymptotic theory for successive sampling with varying probabilities without replacement, I and II. Ann. Math. Stat. **43**, 373–397 and 748–776 (1972)
20. Rosén, B.: Asymptotic theory for Des Raj's estimator, I and II. Scand. J. Stat. **1**, 71–83 and 135–144 (1974)
21. Sen, P.K.: Limit theorems for an extended coupon collector's problem and for successive subsampling with varying probabilities. Calcutta Stat. Assoc. Bull. **29**, 113–132 (1980)

Received March 5, 1988