Statistica Sinica **17**(2007), 1047-1064

ASYMPTOTIC NORMALITY UNDER TWO-PHASE SAMPLING DESIGNS

Jiahua Chen and J. N. K. Rao

University of British Columbia and Carleton University

Abstract: Large sample properties of statistical inferences in the context of finite populations are harder to determine than in the i.i.d. case due to their dependence jointly on the characteristics of the finite population and the sampling design employed. There have been many discussions on special inference procedures under special sampling designs in the literature. General and comprehensive results are still lacking. In this paper, we first present a surprising result on the weak law of large numbers under simple random sampling design: the sampling mean is not necessarily consistent for the population mean even if the population first absolute moment is bounded by a constant not depending on the evolving population size. Instead, a sufficient condition requires the boundedness of the $(1 + \delta)$th absolute population moment for some $\delta > 0$. Based on this result, we prove asymptotic normality of a class of estimators under two-phase sampling design. We show that these estimators can typically be decomposed as a sum of two random variables such that the first one is conditionally asymptotically normal and the second one is asymptotically normal. A theoretical result is derived to combine these two conclusions to prove the asymptotic normality of the estimators.

Key words and phrases: Asymptotic normality, non-response, PPS sampling, ratio estimator, regression estimator, simple random sampling, stratified sampling, weak law of large numbers.

1. Introduction

Finite population parameters are functionals of characteristics of interest associated with sampling units in the finite population under consideration. The fundamental problem in survey sampling is to make inferences on these parameters based on a sample selected according to a specified probability sampling design from the finite population.

In the design based approach, inferences are made according to the probability measure induced by the sampling design. The design specifies a probability distribution on a collection of subsets of the finite population. Even in simple situations, the induced exact distribution of the relevant estimators can be too complex to be determined analytically. Asymptotic theory provides a useful alternative for making inferences.

Some well known results on asymptotic normality of estimators include Erdös and Rényi (1959), Hájek (1960) and Scott and Wu (1981) for simple random sampling without replacement, Krewski and Rao (1981) and Bickel and Freedman (1984) for stratified random sampling, and Hájek (1964) and Prášková (1984) for unequal probability sampling without replacement For a review of asymptotic results in survey sampling, see Sen (1988). Surprisingly, some fundamental problems, such as the weak law of large numbers, are not discussed in depth in survey sampling. Asymptotic normality results for many commonly used estimators under completely general sampling designs are still not available.

Large sample properties in finite population problems are obtained under a special framework. We generally assume that both sample size and population size increase to infinity as some index increases to infinity. Under this framework and simple random sampling without replacement(SRS), one would expect the sampling mean to be consistent for the population mean if the population mean remains bounded by a constant. In Section 2, we show that this is not true. Instead, a sufficient condition for the consistency of the sample mean is that the $(1+\delta)$th population centralized absolute moment is bounded for some $\delta > 0$. In Section 3 we prove a central limit theorem. The result is particularly useful in studying asymptotic properties related to two-phase sampling discussed in subsequent sections.

In Section 4, we study the asymptotic normality of a class of estimators for the population mean or total under a two-phase design with SRS in both phases. We find that the estimators can be decomposed into a sum of two random variables such that one is conditionally asymptotic normal, and the other is asymptotically normal. This structure is then utilized to prove asymptotic normality. In the rest of the paper, we discuss asymptotic normality under variations of two-phase sampling, and under situations that can be viewed as two-phase sampling. In all cases, asymptotic normality of the estimators is established using the Section 3 result.

2. Weak Law of Large Numbers in SRS

Let $X_1, \ldots, X_n, \ldots$ be a sequence of independent and identically distributed (i.i.d.) random variables. It is well known that if $E|X_1|$ exists, $n^{-1}\sum_{i=1}^n X_i \to E(X_1)$ almost surely. Although similar results are often taken for granted in the context of finite populations, no such theorems, even week law of law of large numbers, seem to be available in the sampling literature. In the case of SRS, we prove a particularly useful weak law of large numbers under very general conditions.

Suppose that $\{Y_1, \ldots, Y_n\}$ is a simple random sample drawn without replacement from a finite population $\{y_1, \ldots, y_N\}$, and that $\{X_1, \ldots, X_n\}$ is a simple random sample drawn with replacement from the same population. Thus, X_i, $i = 1, \ldots, n$, are i.i.d. random variables. Further, X_1 and Y_1 have the same marginal distribution. If the finite population is fixed so that N is a constant, then we must have $n \leq N$. We consider the case when both n and N increase as an index ν attached to n and N increases to ∞; for simplicity we suppress the index ν. The corresponding asymptotic results are meaningful in the sense that they provide guidelines in situations when both n and N are large.

In Lemma 1 below, we prove a convergence result for $Y_1, \ldots, Y_n$ by linking them to $X_1, \ldots, X_n$. Let $\bar{Y}_N = N^{-1} \sum_{i=1}^N y_i$ be the finite population mean. Clearly, $\bar{Y}_N$ changes with ν in general. Without loss of generality we assume that $\bar{Y}_N = 0$.

Lemma 1. *Let $Y_1, \ldots, Y_n$ be a simple random sample without replacement from a finite population with population mean $\bar{Y}_N = 0$. Suppose that n and N increase to infinity as some index $\nu \to \infty$. Assume that $nN^{-1} \sum_{i=1}^N I(|y_i| > n) = o(1)$ as $\nu \to \infty$. Then $n^{-1} \sum_{i=1}^n [Y_i - E\{Y_i I(|Y_i| < n)\}] \to 0$ in probability as $\nu \to \infty$.*

Proof. Let $Y_i' = Y_i I(|Y_i| \leq n)$ be truncated versions of the random variables Y_i. Note that $Y_i' = Y_i$ unless $|Y_i| > n$. Hence,

$$P(\sum_{i=1}^n Y_i' \neq \sum_{i=1}^n Y_i) \leq \sum_{i=1}^n P(Y_i' \neq Y_i)$$
$$= nN^{-1} \sum_{i=1}^N I(|y_i| > n) = o(1).$$

Thus, we need only show $n^{-1} \sum_{i=1}^n \{Y_i' - E(Y_i')\} \to 0$ in probability as $\nu \to \infty$.

Define $X_i' = X_i I(|X_i| \leq n)$ as the mirror version of Y_i' obtained from X_i. Since $\{X_1', \ldots, X_n'\}$ is a simple random sample with replacement, we have

$$\mathrm{Var}\,(\sum_{i=1}^n Y_i') = (1 - \frac{n}{N})\mathrm{Var}\,(\sum_{i=1}^n X_i') \leq \mathrm{Var}\,(\sum_{i=1}^n X_i').$$

Following Chow and Teicher (1997, p.128),

$$\mathrm{Var}\,(\sum_{i=1}^n X_i') = n\mathrm{Var}\,(X_1') \leq nE(X_1')^2$$
$$\leq n \sum_{j=1}^n j^2 \{P(|X_1| > j-1) - P(|X_1| > j)\}$$
$$= n[P(|X_1| > 0) - n^2 P(|X_1| > n) + \sum_{j=1}^{n-1} \{(j+1)^2 - j^2\} P(|X_1| > j)]$$

$$\leq 3n\{1 + \sum_{j=1}^{n-1} jP(|X_1| > j)\} = o(n^2),$$

noting that $\lim_{n\to\infty} n^{-1}\sum_{j=1}^{n-1} jP(|X_1| > j) = \lim_{n\to\infty} nP(|X_1| > n) = 0$. Hence, for all $\epsilon > 0$,

$$P\Big\{n^{-1}|\sum_{i=1}^{n}(Y_i' - EY_i')| \geq \epsilon\Big\} \leq \frac{\mathrm{Var}\,(\sum_{i=1}^{n} Y_i')}{n^2\epsilon^2} = o(1).$$

That is, $n^{-1}\sum_{i=1}^{n}(Y_i' - EY_i') \to 0$ in probability. This proves the result.

Surprisingly, for finite populations, the condition $N^{-1}\sum_{i=1}^{\infty}|y_i| < C < \infty$ for some constant C not depending on the index ν is not sufficient.

Example. Assume that the nth finite population consists of copies of two values $n^2/(1-n^2)$ and n^2 such that $P(Y_1 = n^2) = n^{-2}$ where Y_1 is a random sample from the population. It is seen that $P(Y_1 > n) = o(n^{-1})$ and $E(Y_1) = 0$. Further, $E|Y_1| = 2 < \infty$. However, $EY_1I(|Y_1| < n) = -1$ for $n \geq 2$. Hence, $n^{-1}\sum_{i=1}^{n} Y_i \to -1$ in probability instead of the population mean 0.

Clearly, this example is possible because the distribution of Y_1 depends on ν which is the nature of the triangle array: when $\nu = 10$ the observations are drawn from one finite population; when $\nu = 11$ the observations are drawn from another finite population. Thus, the former sample is not part of the latter sample.

Theorem 1. *Assume the conditions of Lemma* 1. *A sufficient condition for* $E\{Y_1I(|Y_1| > n)\} \to 0$ *as* $n \to \infty$ *in the finite population problem is that* $N^{-1}\sum_{i=1}^{N}|y_i|^{1+\delta} \leq C < \infty$ *for some constants* $\delta > 0$ *and* C, *and the sample mean is a consistent estimator of the population mean under this condition.*

Proof. By the Cauchy inequality,

$$E\{|Y_1|I(|Y_1| > n)\} \leq \{E|Y_1|^{1+\delta}\}^{\frac{1}{1+\delta}}\{P(|Y_1| > n)\}^{\frac{\delta}{1+\delta}}.$$

By the moment condition, we have $\{E|Y_1|^{1+\delta}\}^{1/(1+\delta)} \leq C^{1/(1+\delta)} < \infty$ and we have $P(|Y_1| > n) \to 0$ by the Markov inequality. Thus $E\{|Y_1|I(|Y_1| > n)\} \to 0$. This completes the proof.

It is a common belief that the sample variance based on a simple random sample without replacement is a consistent estimator of the population variance. Our result shows that this is not true in general unless the finite population satisfies additional conditions, such as having finite $(2+\delta)$th moment.

Without assuming $\bar{Y}_N = 0$, the condition in Theorem 1 should be replaced by $N^{-1}\sum_{i=1}^{N}|y_i - \bar{Y}_N|^{1+\delta} \leq C < \infty$. Hence, if a finite population sequence satisfies this condition, then any finite population sequence obtained by an arbitrary location shift still satisfies the condition.

3. Central Limit Theorem for Two-Phase Sampling

Estimators of the finite population means or totals in two-phase sampling can often be decomposed into a sum of two random variables such that one depends only on the first phase sample and finite population parameters, and the other on the second phase design and the outcome of the first phase sample. The following result is useful for studying their asymptotic properties.

Theorem 2. *Let U_n, V_n be two sequences of random variables and $\mathcal{B}_n$ be a σ-algebra. Assume that*

1. *there exists $\sigma_{1n} > 0$ such that $\sigma_{1n}^{-1} V_n \to N(0,1)$ in distribution as $n \to \infty$, and V_n is $\mathcal{B}_n$ measurable;*
2. *$E\{U_n|\mathcal{B}_n\} = 0$ and $\mathrm{Var}\,(U_n|\mathcal{B}_n) = \sigma_{2n}^2$ such that*
$$\sup_t |P(\sigma_{2n}^{-1} U_n \le t|\mathcal{B}_n) - \Phi(t)| = o_p(1), \tag{1}$$
where $\Phi(t)$ is the cumulative distribution function of the standard normal random variable;
3. *$\gamma_n^2 = \sigma_{1n}^2/\sigma_{2n}^2 \to \gamma^2$ in probability as $n \to \infty$.*

Then
$$\frac{U_n + V_n}{\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}} \to N(0,1) \tag{2}$$
in distribution as $n \to \infty$.

Remark. The σ-algebra allows us to study the asymptotic normality of U_n by regarding V_n as a non-random constant. It is often taken as the σ-algebra generated by a random variable or a random subset of the finite population in the case of survey sampling.

Since cumulative distribution functions are monotone and bounded, and the distribution of the standard normal random variable is continuous, Condition (1) is equivalent to the condition that for each t, $|P(\sigma_{2n}^{-1} U_n \le t|\mathcal{B}_n) - \Phi(t)| = o_p(1)$. This result is often referred as Polya's theorem (Schenker and Welsh (1988)).

Proof. Note that
$$P\{(\sigma_{1n}^2 + \sigma_{2n}^2)^{-\frac{1}{2}}(U_n + V_n) \le t\} = P\{U_n \le t\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2} - V_n\}$$
$$= E\left\{P(\sigma_{2n}^{-1} U_n \le t\sqrt{1 + \frac{\sigma_{1n}^2}{\sigma_{2n}^2}} - \sigma_{2n}^{-1} V_n|\mathcal{B}_n)\right\}$$

$$= E\left\{\Phi(t\sqrt{1+\gamma_n^2} - \sigma_{2n}^{-1}V_n)\right\} + E[\Delta_n(t)], \tag{3}$$

where $\Delta_n(t) = P(\sigma_{2n}^{-1}U_n \le t\sqrt{1+\sigma_{1n}^2/\sigma_{2n}^2} - \sigma_{2n}^{-1}V_n|\mathcal{B}_n) - \Phi(t\sqrt{1+\gamma_n^2} - \sigma_{2n}^{-1}V_n)$. By Condition 2, $\Delta_n(t) = o_p(1)$ uniformly in t. Further, $\Delta_n(t)$ is bounded, implying $E\Delta_n(t) = o(1)$. Let $Z_n = \sigma_{1n}^{-1}V_n$ so by Condition 1, $Z_n \to Z$ in distribution for some standard normal random variable Z. Since $\Phi(\cdot)$ is a bounded continuous function, we have

$$E\{\Phi(t\sqrt{1+\gamma_n^2} - \gamma_n Z_n)\} \to E\{\Phi(t\sqrt{1+\gamma^2} - \gamma Z)\},$$

For mathematical simplicity, let U be a standard normal random variable independent of Z. We have

$$\begin{aligned} E\{\Phi(t\sqrt{1+\gamma^2} - \gamma Z)\} &= E\{P(U \le t\sqrt{1+\gamma^2} - \gamma Z|Z\} \\ &= P(U + \gamma Z \le t\sqrt{1+\gamma^2}) = \Phi(t), \end{aligned}$$

noting that $U+\gamma Z$ is a normal random variable with mean 0 and variance $1+\gamma^2$. This completes the proof.

Our result is different from Lemma 1 of Schenker and Welsh (1988) or Nielsen (2003). They require almost sure convergence in (1). In finite population problems, each sample is part of a triangular array. As we have seen in our Theorem 1, even a weak law of large numbers is difficult to establish in finite population problems. Establishing almost sure convergence may require conditions on higher order moments. In comparison, Condition (1) in Theorem 2 is more convenient to verify in applications.

In most applications, the parameters in σ_{1n}^2 and σ_{2n}^2 are replaced by their consistent estimators, resulting in studentized statistics. For simple random sampling, Theorem 1 shows that consistent estimators are readily obtained by the method of moments, under mild conditions.

4. Two-Phase Sampling: SRS in Both Phases

Assume there is a finite population consisting of N sampling units with measurements (x_i, y_i) for $i = 1, \ldots, N$. A two-phase sampling design with SRS in both phases is as follows.

Phase 1. A simple random sample S_1 of size n_1 without replacement is drawn from $\{1, 2, \ldots, N\}$ and all the x_i, $i \in S_1$ are obtained.

Phase 2. A simple random sample S of size n without replacement from S_1 is drawn and all the y_i, $i \in S$ are obtained.

A two-phase design is often used when the measurement of the characteristic of interest, y, is more expensive than the measurement of an auxiliary variable,

x, related to y. In this design, covariate x is measured on a large sample of the population in the first phase. In the second phase, the characteristic of interest y is measured on a random subset of the sample in the first phase (see Cochran (1977, Chap. 12)). We discuss the asymptotic normality of a number of two-phase estimators of the population mean $\bar{Y}$ in this section, under simple random sample without replacement in both phases. First, we specify some assumptions.

4.1. Finite population assumptions

In order for the asymptotic results to be applicable, the finite population and the corresponding design must satisfy certain conditions. Due to marked differences in sampling designs, we only attempt to specify some general conditions that are required for a two-phase design with SRS in both phases.

We assume that each unit in the population has a pair of characteristics (x_i, y_i), $i = 1, \ldots, N$. Denote the finite population means, variances and the covariance of x and y as

$$\bar{X} = N^{-1} \sum_{i=1}^{N} x_i, \quad \bar{Y} = N^{-1} \sum_{i=1}^{N} y_i,$$

$$\sigma_X^2 = (N-1)^{-1} \sum_{i=1}^{N} (x_i - \bar{X})^2, \quad \sigma_Y^2 = (N-1)^{-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2,$$

$$\sigma_{XY} = (N-1)^{-1} \sum_{i=1}^{N} (x_i - \bar{X})(y_i - \bar{Y}).$$

Let $\rho_{XY} = \sigma_{XY}/(\sigma_X \sigma_Y)$ be the correlation coefficient. The population totals are denoted as $X = N\bar{X}$ and $Y = N\bar{Y}$.

As before, we assume that there exists an index ν such that when ν increases, the finite population evolves, but some conditions remain satisfied. Let us denote $\mathcal{P}_\nu = \{(x_i, y_i) : i = 1, 2, \ldots, N_\nu\}$. Suppressing index ν we list some commonly assumed conditions on $\mathcal{P}_\nu$.

A1. $N \to \infty$ as $\nu \to \infty$.

A2. There exist some generic constants $M_1, M_2, \delta > 0$ and ρ_0, such that for all ν,

$$0 < M_1 \le \sigma_X^2,\ \sigma_Y^2 \le M_2 < \infty,$$

$$\nu_x = N^{-1} \sum_{i=1}^{N} |x_i|^{2+\delta} \le M_2 < \infty,$$

$$\nu_y = N^{-1} \sum_{i=1}^{N} |y_i|^{2+\delta} \le M_2 < \infty,$$

$$|\rho_{XY}| \le \rho_0 < 1.$$

A3. As $\nu \to \infty$, σ_Y^2/σ_X^2 and σ_{XY}/σ_X^2 converge to some constants.

The above conditions require that the finite population consists of units whose characteristics are not severely skewed, and that X and Y are not perfectly correlated. The former makes the normal approximation hold with reasonable precision, while the latter makes the problem under consideration non-trivial. According to Bickel and Freedman (1984), under simple random sampling without replacement (SRS), sample means are asymptotically normal under Conditions A1−A3 if n and $N-n$ both tend to infinity as $\nu \to \infty$.

Let the sample means in the second phase be $\bar{x} = n^{-1}\sum_{i\in S} x_i$, $\bar{y} = n^{-1}\sum_{i\in S} y_i$, and in the first phase be $\bar{x}_1 = n_1^{-1}\sum_{i\in S_1} x_i$, $\bar{y}_1 = n_1^{-1}\sum_{i\in S_1} y_i$. Also, denote the sample variance and covariances as

$$s_{1x}^2 = (n_1-1)^{-1}\sum_{i\in S_1}(x_i-\bar{x}_1)^2; \qquad s_{1y}^2 = (n_1-1)^{-1}\sum_{i\in S_1}(y_i-\bar{y}_1)^2,$$
$$s_x^2 = (n-1)^{-1}\sum_{i\in S}(x_i-\bar{x})^2; \qquad s_y^2 = (n-1)^{-1}\sum_{i\in S}(y_i-\bar{y})^2,$$
$$s_{1xy} = (n-1)^{-1}\sum_{i\in S_1}(x_i-\bar{x}_1)(y_i-\bar{y}_1); \quad s_{xy} = (n-1)^{-1}\sum_{i\in S}(x_i-\bar{x})(y_i-\bar{y}).$$

4.2. Regression estimator

The finite population mean $\bar{Y}$ can be estimated by a difference estimator given by

$$\hat{\bar{Y}}_b = \bar{y} + b(\bar{x}_1 - \bar{x}). \tag{4}$$

Note that for any given constant b, we have $E[\hat{\bar{Y}}_b] = \bar{Y}$. An optimal choice b minimizes the variance. Recalling that S is a sample from S_1, we have

$$\begin{aligned} E\{\mathrm{Var}\,(\hat{\bar{Y}}_b|S_1)\} &= (n^{-1}-n_1^{-1})E(s_{1y}^2 + b^2 s_{1x}^2 - 2bs_{1xy}) \\ &= (n^{-1}-n_1^{-1})(\sigma_Y^2 + b^2\sigma_X^2 - 2b\sigma_{XY}). \end{aligned} \tag{5}$$

In addition $\mathrm{Var}\,\{E(\hat{\bar{Y}}_b|S_1)\} = \mathrm{Var}\,(\bar{y}_1) = (n_1^{-1} - N^{-1})\sigma_Y^2$, so $\mathrm{Var}\,(\hat{\bar{Y}}_b) = (n^{-1} - N^{-1})\sigma_Y^2 + (n^{-1}-n_1^{-1})(b^2\sigma_X^2 - 2b\sigma_{XY})$, see Cochran (1977, p.239). The optimal choice of b is given by $b_{opt} = \sigma_{XY}/\sigma_X^2$. In applications, b is chosen as $\hat{b} = s_{xy}/s_x^2$, the least squares regression coefficient of y_i on x_i computed from the second phase sample. When b is estimated, the variance formula is an approximation.

Under Conditions A1−A3, $\hat{b} - b_{opt} = s_{xy}/s_x^2 - b_{opt} = o_p(1)$ by Theorem 1. Hence, the asymptotic normality of $\hat{\bar{Y}}_b$ to be developed remains valid when b is replaced by $\hat{b}$. We thus assume that b is a pre-chosen non-random constant.

We use the following decomposition of $\hat{\bar{Y}}_b - \bar{Y}$:

$$\hat{\bar{Y}}_b - \bar{Y} = \{(\bar{y} - \bar{y}_1) + b(\bar{x}_1 - \bar{x})\} + (\bar{y}_1 - \bar{Y}) =: U_n + V_n, \tag{6}$$

where $U_n = (\bar{y} - \bar{y}_1) + b(\bar{x}_1 - \bar{x})$ and $V_n = \bar{y}_1 - \bar{Y}$. Both U_n and V_n are asymptotically normal after rescaling. The asymptotic normality of their sum can be thus established by verifying conditions in Theorem 2.

Lemma 2. *Let U_n be defined as in* (6) *and denote* $\sigma_{2n}^2 = \mathrm{Var}\,(U_n|S_1) = (n^{-1} - n_1^{-1})(s_{1y}^2 - 2bs_{1xy} + b^2 s_{1x}^2)$. *Under Conditions* A1–A3, *we have* $\sup_t |P(\sigma_{2n}^{-1}U_n \le t|S_1) - \Phi(t)| = o_p(1)$ *as n and $n_1 - n$ tend to infinity.*

Proof. For each ν, a subset $S_1 = S_{1,\nu}$ is obtained in the first phase sample. Given S_1, $\{(x_i, y_i), i \in S\}_\nu$ is a simple random sample without replacement from the finite population $\{(x_i, y_i), i \in S_1\}$. Thus we are considering a sequence of finite populations $\mathcal{P}_{1,\nu} = \{(x_i, y_i), i \in S_1\}$ for the purpose of asymptotics. Given any sequence of finite populations $\mathcal{P}_{1,\nu}, \nu = 1, 2, \ldots$, satisfying Conditions A1–A3 for some M_1, M_2 and δ, and under the condition that n and $n_1 - n$ both tend to infinity, we have (Bickel and Freedman (1984)) $\sigma_{2n}^{-1}U_n \to N(0, 1)$. Applying Polya's theorem, this result implies that

$$\sup_t |P(\sigma_{2n}^{-1}U_n \le t|S_1) - \Phi(t)| \to 0 \tag{7}$$

for this specified realization of the first phase sample.

Since $\mathcal{P}_{1,\nu}$ is random there is a small chance that the first phase sample forms a population sequence which does not satisfy Conditions A1–A3. However, this chance tends to zero as $\nu \to \infty$ as seen in following. Note that the original population sequence $\mathcal{P}_\nu, \nu = 1, 2, \ldots$, from which the first phase sample is drawn, satisfies Conditions A1–A3. By the weak law of large numbers for SRS proved in Theorem 1 and the condition that ν_x and ν_y are finite for some $\delta > 0$, we have that for each $0 < \delta' < \delta$,

$$n_1^{-1} \sum_{i \in S_1} |x_i|^{2+\delta'} - N^{-1} \sum_{i=1}^{N} |x_i|^{2+\delta'} \to 0, \quad n_1^{-1} \sum_{i \in S_1} |y_i|^{2+\delta'} - N^{-1} \sum_{i=1}^{N} |x_i|^{2+\delta'} \to 0,$$

in probability. Thus, with probability tending to one, $n_1^{-1} \sum_{i \in S_1} |x_i|^{2+\delta'}$ and $n_1^{-1} \sum_{i \in S_1} |y_i|^{2+\delta'}$ are asymptotically bounded. The random population sequence $\mathcal{P}_{1,\nu}$ hence satisfies Conditions A1–A3 with the generic constant δ'. That is, (7) holds in probability which implies that as a random variable, $\sup_t |P(\sigma_{2n}^{-1}U_n \le t|S_1) - \Phi(t)| \to 0$ in probability.

We now use this result to obtain the asymptotic normality of the regression estimator $\hat{\bar{Y}}_b$.

Theorem 3. *Assume n, $n_1 - n$ and $N - n_1$ tend to infinity as ν goes to infinity, and Conditions* A1−A3 *hold for the finite population. We have* $\sigma_n^{-1}(\hat{\bar{Y}}_b - \bar{Y}) \to N(0,1)$ *in distribution where* $\sigma_n^2 = (n^{-1} - n_1^{-1})(s_{1y}^2 + b^2 s_{1x}^2 + 2bs_{1xy}) + (n_1^{-1} - N^{-1})\sigma_Y^2$.

Proof. Let $U_n = \{(\bar{y} - \bar{y}_1) + b(\bar{x} - \bar{x}_1)\}$, $V_n = \bar{y}_1 - \bar{Y}$. Also, let $\sigma_{2n}^2 = (n^{-1} - n_1^{-1})(s_{1y}^2 + b^2 s_{1x}^2 + 2bs_{1xy})$ and $\sigma_{1n}^2 = (n_1^{-1} - N^{-1})\sigma_Y^2$. By the weak law of large numbers and Condition A3, $\sigma_{2n}^2/\sigma_{1n}^2$ converges to a constant. Thus U_n, V_n satisfy the conditions of Theorem 2, and the asymptotic normality of $\hat{\bar{Y}}_b$ is proved.

The condition that $n_1 - n$ tends to infinity is not restrictive. If $n_1 - n$ remains finite, then U_n converges to zero faster than V_n. Thus, the limiting distribution of $\hat{\bar{Y}}_b$ is determined by that of V_n. Further, one motivation of the two-phase sampling plan is to save the cost through a large sample size difference. Hence large $n_1 - n$ is also a practical requirement.

In applications, σ_{1n}^2 depends on unknown population parameters, and σ_{2n}^2 is a function of the first phase sample which includes the unobserved y-values. Thus, both of them have to be estimated before making statistical inference. The asymptotic result, is not affected when $\sigma_{1n}^2 + \sigma_{2n}^2$ is replaced by a consistent estimator, in view of Slutsky's theorem.

4.3. Ratio estimator

When a proportional relationship is suspected between the response variable y and the covariate x, a ratio estimator might be used for estimating the population mean, $\bar{Y}$. In this case, the ratio estimator of $\bar{Y}$ is given by $\hat{\bar{Y}}_R = (\bar{y}/\bar{x})\bar{x}_1$. Let $B = \bar{Y}/\bar{X}$. In this section, we assume that the limits of $\bar{Y}$ and $\bar{X}$ both exist as $\nu \to \infty$ in addition to A1−A3.

We have

$$\hat{\bar{Y}}_R - \bar{Y} = \frac{\bar{x}_1(\bar{y} - \bar{Y}) + \bar{Y}(\bar{x}_1 - \bar{x})}{\bar{x}}$$
$$= \frac{\bar{X}(\bar{y}_1 - \bar{Y}) + \bar{X}(\bar{y} - \bar{y}_1) + \bar{Y}(\bar{x}_1 - \bar{x})}{\bar{X}} + o_p(n^{-\frac{1}{2}}). \tag{8}$$

Let $V_n = \bar{y}_1 - \bar{Y}$ and $U_n = (\bar{y} - \bar{y}_1) - B(\bar{x} - \bar{x}_1)$. Similar to case of the regression estimator, V_n is asymptotically normal since $\bar{y}_1$ is the sample mean of a simple random sample without replacement. Its asymptotic variance is given by $\sigma_{1n}^2 = (n_1^{-1} - N^{-1})\sigma_Y^2$.

Further, given the sample from the first phase, $\bar{X}\bar{y} - \bar{Y}\bar{x}$ can be regarded the sample mean of a simple random sample without replacement from a population consist of $\bar{X}y_i - \bar{Y}x_i$, $i \in S_1$. Similar to Section 4.1, the finite population formed by the sample units in S_1 satisfies Conditions A1−A3 in probability. Thus, with probability converging to 1, U_n is conditionally asymptotically normal with conditional asymptotic variance $\sigma_{2n}^2 = (n^{-1} - n_1^{-1})\{s_y^2 - 2Bs_{xy} + B^2 s_x^2\}$. Thus, V_n and U_n defined above satisfy the conditions of Theorem 2. Consequently, $(\hat{\bar{Y}}_R - \bar{Y})/\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}$ is asymptotically N(0, 1).

In this section, we assumed that the sampling schemes in Phases 1 and 2 are both simple random sampling without replacement. Our technique, however, is applicable to situations in which U_n is conditionally asymptotically normal and V_n is asymptotically normal. Some of those cases will be discussed further in Sections 5 and 6.

5. Two-Phase Sampling: PPS Sampling in the First Phase

Consider the case where each population unit has three characteristics (x, y, z), with "sizes" $z_1, \ldots, z_N$ assumed to be known. A sample S_1 of size n_1 is drawn in the first phase with probability proportional to size (PPS) z_i, with replacement, and the x-values of sampled units are observed. A simple random sample S of size n from S_1 is then drawn without replacement in the second phase and y-values are observed.

Unbiased estimators of the population means $\bar{X}$ and $\bar{Y}$ based on the first phase sample are the usual PPS estimators

$$\hat{\bar{X}}_1 = n_1^{-1} \sum_{i \in S_1} \frac{x_i}{Np_i} \quad \text{and} \quad \hat{\bar{Y}}_1 = n_1^{-1} \sum_{i \in S_1} \frac{y_i}{Np_i},$$

where $p_i = z_i/Z$ and Z is the known population total of z. As usual, $\hat{\bar{Y}}_1$ is unknown to us. Let

$$\hat{\hat{X}} = \frac{1}{n} \sum_{i \in S} \frac{x_i}{Np_i}, \quad \text{and} \quad \hat{\hat{Y}} = \frac{1}{n} \sum_{i \in S} \frac{y_i}{Np_i}$$

be the estimators of $\hat{\bar{X}}_1$ and $\hat{\bar{Y}}_1$ based on the second phase sample. The difference estimator of the population mean $\bar{Y}$ is given by $\hat{\bar{Y}}_b = \hat{\hat{Y}} + b(\hat{\bar{X}}_1 - \hat{\hat{X}})$ for some constant b. The optimal choice of b can be estimated from the sample. As before, we regard b as a constant for the purpose of asymptotics.

To establish the asymptotic normality of $\hat{\bar{Y}}_b$ under the current design, the finite population must satisfy an additional condition. Let Λ and Γ be two random variables such that

$$P\left(\Lambda = \frac{y_i}{Np_i}, \Gamma = \frac{x_i}{Np_i}\right) = p_i$$

for $i = 1, \ldots, N$. Hence $E(\Lambda) = \bar{Y}$, $E(\Gamma) = \bar{X}$, and

$$\sigma_\lambda^2 = \mathrm{Var}\,(\Lambda) = \sum_{i=1}^N p_i (\frac{y_i}{Np_i} - \bar{Y})^2, \quad \sigma_\gamma^2 = \mathrm{Var}\,(\Gamma) = \sum_{i=1}^N p_i (\frac{x_i}{Np_i} - \bar{X})^2,$$

$$\sigma_{\lambda,\gamma} = \mathrm{Cov}(\Lambda, \Gamma) = \sum_{i=1}^{N} p_i\Big(\frac{x_i}{Np_i} - \bar{X}\Big)\Big(\frac{y_i}{Np_i} - \bar{Y}\Big).$$

We now state conditions regarding (Λ, Γ).

A4. There exist generic constants M_1, M_2, $\delta > 0$ and ρ_0 such that, for all ν,

$$0 < M_1 \le \sigma^2_\lambda, \sigma^2_\gamma < M_2 < \infty; \quad |\sigma_{\lambda,\gamma}|/\sigma_\lambda\sigma_\gamma \le \rho_0 < 1,$$

$$E|\Lambda|^{2+\delta} < M_2, \quad E|\Gamma|^{2+\delta} < M_2.$$

A5. As $\nu \to \infty$, $\sigma^2_\lambda/\sigma^2_\gamma$ and $\sigma^2_{\lambda,\gamma}/\sigma^2_\gamma$ converge to some constants.

Assume A1, A4 and A5. Decompose $\hat{\bar{Y}}_b - \bar{Y}$ as $\hat{\bar{Y}}_b - \bar{Y} = \{\hat{\bar{Y}} - \hat{\bar{Y}}_1 + b(\hat{\bar{X}}_1 - \hat{\bar{X}})\} + (\hat{\bar{Y}}_1 - \bar{Y})$. Let $U_n = \hat{\bar{Y}} - \hat{\bar{Y}}_1 + b(\hat{\bar{X}}_1 - \hat{\bar{X}})$ and $V_n = \hat{\bar{Y}}_1 - \bar{Y}$. Due to PPS with replacement, we can write $\hat{\bar{Y}}_1 = n_1^{-1}\sum_{i=1}^{n_1}\Lambda_i$ with Λ_i being i.i.d. copies of Λ. Applying central limit theory for triangular arrays of i.i.d. random variables (Serfling (1980, p.32)), $V_n = \hat{\bar{Y}}_1 - \bar{Y}$ is asymptotically normal after standardization. The asymptotic variance of V_n is given by $\sigma^2_{1n} = n_1^{-1}\sigma^2_\lambda$.

We now examine the asymptotic distribution of U_n. Define $\gamma_i = (Np_i)^{-1}(y_i - bx_i)$ for $i \in S_1$, $\mu_\gamma = n_1^{-1}\sum_{i\in S_1}\gamma_i$ and $\sigma^2_\gamma = (n_1-1)^{-1}\sum_{i\in S_1}(\gamma_i - \mu_\gamma)^2$. Given S_1, U_n is the mean of a simple random sample without replacement of size n from the population $\{\gamma_i : i \in S_1\}$. Due to Condition A4, $\{\gamma_i : i \in S_1\}$ satisfies Conditions A1−A3 with probability approaching 1. Hence, U_n is conditionally asymptotically normal with asymptotic variance $\sigma^2_{2n} = (n^{-1} - n_1^{-1})\sigma^2_\gamma$.

In conclusion, our decomposition of U_n and V_n meets the conditions of Theorem 2. Thus, $(\hat{\bar{Y}}_b - \bar{Y})/\sqrt{\sigma^2_{1n} + \sigma^2_{2n}}$ asymptotically N(0, 1).

6. Two-Phase Sampling: PPS in the Second Phase

We now consider PPS sampling in the second phase. Here the first phase is a simple random sample without replacement of size n_1 and x-values are measured. The second phase sample is drawn with probability proportional to x_i and with replacement from the first phase sample.

Let $p_{1i} = x_i/\sum_{j\in S_1} x_j = x_i/(n_1\bar{x}_1)$ where $\bar{x}_1$ is the first-phase sample mean of x. Then the population mean $\bar{Y}$ is estimated by

$$\hat{\bar{Y}} = \frac{1}{n_1}\Big\{\frac{1}{n}\sum_{i\in S}\frac{y_i}{p_{1i}}\Big\} = \frac{\bar{x}_1}{n}\sum_{i\in S}\frac{y_i}{x_i}.$$

We now seek a decomposition. It is readily seen that

$$E\Big[\sum_{i\in S}\frac{y_i}{x_i}|S_1\Big] = \frac{n\bar{y}_1}{\bar{x}_1},$$

where $\bar{y}_1$ is the mean of the first phase sample. Hence, we decompose $\hat{\hat{Y}} - \bar{Y}$ as

$$\hat{\hat{Y}} - \bar{Y} = \bar{x}_1 \left(\frac{1}{n} \sum_{i \in S} \frac{y_i}{x_i} - \frac{\bar{y}_1}{\bar{x}_1} \right) + \bar{y}_1 - \bar{Y}.$$

Let

$$U_n = \bar{x}_1 \left(\frac{1}{n} \sum_{i \in S} \frac{y_i}{x_i} - \frac{\bar{y}_1}{\bar{x}_1} \right) = \frac{1}{n} \sum_{i \in S} \left(\frac{y_i}{x_i} \bar{x}_1 - \bar{y}_1 \right)$$

and $V_n = \bar{y}_1 - \bar{Y}$. Clearly, V_n is asymptotically normal with asymptotic variance $\sigma_{1n}^2 = (n_1^{-1} - N^{-1}) \sigma_Y^2$.

Given S_1, U_n is a mean of independent and identically distributed random variables. Thus, it is asymptotically normal under mild conditions similar to A4, which we omit for the sake of space. Its conditional asymptotic variance is given by

$$\sigma_{2n}^2 = \frac{1}{n} \sum_{i \in S_1} p_{1i} \left(\frac{y_i}{x_i} \bar{x}_1 - \bar{y}_1 \right)^2.$$

Consequently, $(\hat{\hat{Y}} - \bar{Y}) / \sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}$ is asymptotically N(0, 1).

7. Two-Phase Sampling: Stratification in the Second Phase

In some applications, a simple random sample without replacement of size n_1 is obtained in the first phase. The sample is then stratified into L strata according to values of auxiliary variable x collected in the first phase. Let n_{1h} be the number of first-phase sample units in each stratum h; $h = 1, \ldots, L$. We assume that when $n_1 \to \infty$, $P(\min_{1 \le h \le L} n_{1h} \ge 2) \to 1$. This will be the case when L is fixed and the proportion in each stratum remains non-zero. We also assume that each stratum, when regarded as a finite population, satisfies Conditions A1−A3 with common constants M_1, M_2 and δ.

Let $n_h = n_{1h} \nu_h$ for some fixed $\nu_h \in (0, 1)$. In the second phase of sampling, a simple random sample without replacement of size n_h is drawn from the hth stratum (Rao (1973)). Note that if n_h is not an integer, we can round it off without affecting the asymptotic properties as $n_{1h} \to \infty$. Let $w_h = n_{1h}/n_1$ be the first-phase strata weights. The two-phase sampling estimator is then given by the stratified mean $\hat{\hat{Y}}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h$ where $\bar{y}_h$ are the second phase stratum sample means. The estimator itself does not depend on the true stratum weights which are likely unknown in this application.

We assume that the strata are predetermined at the population level according to x and the same stratification rule is used for each first-phase sample. For stratum h, let N_h be the population stratum size, $W_h = N_h/N$ be the population stratum weight and $\bar{y}_{1h}$ be the first-phase stratum sample mean. Letting the first phase sample mean $\bar{y}_1 = \sum_{h=1}^{L} w_h \bar{y}_{1h}$, we can write $\hat{\bar{Y}}_{st} - \bar{Y} = \sum_{h=1}^{L} w_h(\bar{y}_h - \bar{y}_{1h}) + (\bar{y}_1 - \bar{Y})$.

Let $V_n = \bar{y}_1 - \bar{Y}$. Its asymptotic normality is readily verified under Conditions A1−A3. The corresponding asymptotic variance is given by $\sigma_{1n}^2 = (n_1^{-1} - N^{-1})\sigma_Y^2$. Similarly, let $U_n = \sum_{h=1}^{L} w_h(\bar{y}_h - \bar{y}_{1h})$. Given the first-phase sample, which can be regarded as the corresponding finite population, U_n is the stratified (and centered) sample mean.

To establish asymptotic normality, we assume that L remains a constant as the finite population evolves. Further, we assume that each stratum, when regarded as a finite population itself, satisfies Conditions A1−A3. Consequently, the first phase sample forms a stratified finite population with every stratum satisfying A1−A3 in probability. Thus, using the asymptotic normality result in Bickel and Freedman (1984), U_n is conditionally asymptotic normal with conditional asymptotic variance $\sigma_{2n}^2 = \sum_{h=1}^{L}(n_h^{-1} - n_{1h}^{-1})w_h^2 s_{1h}^2$, with s_{1h}^2 being the first phase sample variance of y-values in stratum h.

Since not all first phase sample y-values are observed, σ_{2n}^2 has to be replaced in practice by a consistent estimator based on the second-phase sample. One such choice is to replace s_{1h}^2 in σ_{2n}^2 by the second phase sample variance s_{2h}^2 which is consistent by the weak law of large numbers according to Theorem 1. Regardless of the choice, using the conclusion of Theorem 2, we conclude that $(\bar{y}_{st} - \bar{Y})/\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}$ is asymptotically N(0, 1).

8. Two-Phase Sampling in Other Contexts

In some situations, a two-phase sampling design is not used, but the analysis resembles that of two-phase sampling. We discuss two cases here.

8.1. SRS with uniform non-response

Non-response occurs in most survey applications. If the probability of response is uniform over the finite population, then the sample mean of respondents is a good estimator of the population mean. Otherwise, more sophisticated techniques are needed to avoid severe bias of the estimator and to obtain approximately unbiased variance estimators.

In simple situations, we can view a sample containing non-response as a sample from a two-phase sampling design. The first phase sample contains all the sampling units according to simple random sampling. A Bernoulli experiment is then performed so that only a subset of the sample have their response variable measured.

Let y_i, $i \in s$ be the response values of the units in the sample. Let z_i be the response indicator taking the value 1 or 0 according as the unit is a respondent or not. The sample mean of the respondents is given by

$$\bar{y}_r = \frac{\sum_{i \in s} z_i y_i}{\sum_{i \in s} z_i}.$$

We assume that z_i, $i \in s$ are independent of each other and that the response mechanism is uniform, i.e., $P(z_i = 1) = p$ with $0 < p < 1$.

Notice that under Conditions A1−A3,

$$\bar{y}_r - \bar{Y} = \frac{1}{np} \sum_{i \in s} z_i (y_i - \bar{y}) + (\bar{y} - \bar{Y}) + o_p(n^{-\frac{1}{2}}),$$

where $\bar{y}$ is the sample mean for the full sample s. Clearly, $\bar{y} - \bar{Y}$ is asymptotically normal with asymptotic variance $n^{-1}(1-f)\sigma_Y^2$ where σ_Y^2 is the population variance and f is the sample fraction. We now establish the asymptotic normality of $\sqrt{n}(\bar{y}_r - \bar{Y})$ through that of $n^{-1/2} \sum_{i \in s} z_i (y_i - \bar{y})$. Given the sample s, $n^{-1/2} \sum_{i \in s} z_i s_y^{-1} (y_i - \bar{y})$ is a sum of independent random variables with mean 0, where s_y^2 is the full sample variance. It satisfies the Lindberg condition when $\sum_{i \in s} |y_i - \bar{Y}|^{2+\delta} / (n s_y^{2+\delta}) \to 0$ for some $\delta > 0$. By the law of large numbers proved in Theorem 2 and Conditions A1−A3, this condition is satisfied in probability. Thus, we have

$$\left| P[\{np(1-p)\}^{-1/2} \sum_{i \in s} z_i s_y^{-1} (y_i - \bar{Y}) \le x | s] - \Phi(x) \right| \to 0$$

in probability.

Using Theorem 1, we conclude that $\bar{y}_r - \bar{Y}$ is asymptotically normal with mean 0 and asymptotic variance $n^{-1}(p^{-1} - f)\sigma_Y^2$, noting that s_y^2 converges in probability to σ_Y^2. A consistent estimator of the asymptotic variance is given by $(r^{-1} - N^{-1}) s_{2y}^2$, where r is the number of respondents in the sample and s_{2y}^2 is the sample variance of the respondents.

8.2. Sampling on two occasions

In practice, the same population is often sampled on two or more occasions and the same study variable is measured on each occasion. In this section, we confine attention to sampling on two occasions and denote the study variable as x and y for occasions 1 and 2, respectively. A simple random sample without replacement of size n_1 is drawn from the finite population on the first occasion and x-values are obtained. On the second occasion, we take a simple random sample without replacement of size n_m from the first occasion sample, where $n_m = n_1 - n_u$. An additional sample of size n_u is obtained without replacement from the rest of the finite population. The samples are then combined to estimate the population mean.

Let $\bar{y}_m$ and $\bar{y}_u$ be sample means from the second occasion sample corresponding to the matched and un-matched sample units. Let $\bar{x}_1$ be the sample mean of x from the first occasion sample, and $\bar{x}_m$ be the sample mean of x based on the matched sample.

The population mean $\bar{Y}$ on the second occasion is then estimated by a linear combination of the regression estimator $\bar{y}_m + b(\bar{x}_1 - \bar{x})$ and the unmatched mean $\bar{y}_u$:

$$\hat{\bar{Y}} = \alpha\{\bar{y}_m + b(\bar{x}_1 - \bar{x}_m)\} + (1-\alpha)\bar{y}_u,$$

for some constants α and b. The optimal choice of b is the regression coefficient of y on x. The optimal choice of α minimizes the variance of the linear combination. When either or both of them are replaced by consistent estimators, the limiting distribution of $\hat{\bar{Y}}$ will not be affected. Thus, for simplicity we assume that α and b are both non-random constants.

Under the above sampling design, the finite population is divided into two strata: The first stratum consists of all units in the first occasion sample, and the remaining units in the population form the second stratum. Hence, stratification here is random.

Among the terms in $\hat{\bar{Y}}$, $\bar{x}_1$ is completely defined by the first occasion sample. A decomposition of $\hat{\bar{Y}} - \bar{Y}$ is given by

$$\hat{\bar{Y}} - \bar{Y} = \alpha(\bar{y}_1 - \bar{Y}) + [\alpha\{(\bar{y}_m - \bar{y}_1) + b(\bar{x}_1 - \bar{x}_m)\} + (1-\alpha)(\bar{y}_u - \bar{Y})],$$

where $\bar{y}_1$ is the unobserved mean of y for the first occasion sample. Let $V_n = \alpha(\bar{y}_1 - \bar{Y})$ and $U_n = \alpha\{(\bar{y}_m - \bar{y}_1) + b(\bar{x}_1 - \bar{x}_m)\} + (1-\alpha)(\bar{y}_u - \bar{Y})$. Since $\bar{y}_1$ is the sample mean of a simple random sample drawn without replacement, it follows that V_n is asymptotically normal with mean 0 and variance $\sigma_{1n}^2 = (n_1^{-1} - N^{-1})\alpha^2\sigma_Y^2$.

We next consider the asymptotic normality of U_n given the first occasion sample. Note that U_n is a linear combination of two conditionally independent terms given the first occasion sample. The first term is the usual difference estimator based on the matched sample. Thus, it is conditionally asymptotically normal. The second term is equivalent to a sample mean from a simple random sample without replacement and hence it is also conditionally asymptotically normal. With the conditional independence, we arrive at the conclusion that U_n is conditionally asymptotically normal with asymptotic conditional variance

$$\sigma_{2n}^2 = \alpha^2(n_m^{-1} - n_1^{-1})\{s_{1y}^2 - 2bs_{1xy} + b^2 s_{1x}^2\} + (1-\alpha)^2\{n_u^{-1} - (N-n_1)^{-1}\}s_y^2,$$

where s_{1y}^2, s_{1xy} and s_{1x}^2 are sample variances and covariance based on the first occasion sample, and s_y^2 is population variance of y when the first phase sample is excluded. By Theorem 2 again, we claim that $(\hat{\hat{Y}} - \bar{Y})/\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}$ is asymptotically N(0, 1).

9. Conclusion

In this paper, the asymptotic normality of estimators of totals and means under several two-phase sampling designs is studied. The case of simple random sampling with uniform non-response is also considered. Extensions to imputation for missing item values under a two-phase sampling approach will be reported in a separate paper.

Acknowledgement

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada. We thank an anonymous referee whose careful review lead to substantial improvement of the paper.

References

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

Cochran, W. G. (1977). *Sampling Techniques.* 3rd Edition. Wiley, New York.

Chow, Y. S. and Teicher, H. (1997). *Probability Theory, Independence, Interchangeability, Martingales.* Springer Texts in Statistics. Third Edition. Springer, New York.

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

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Pub. Math. Inst. Hungar. Acad. Sci.* **5**, 361-374.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite populations. *Ann. Math. Statist.* **35**, 1491-1523.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13-30.

Krewski, D. and Rao, J. N. K. (1981). Inference from stratified samples: properties of linearization, jackknife and balanced repeated replication methods. *Ann. Statist* **9**, 1010-1019.

Nielsen, S. F. (2003). Proper and improper multiple imputation (with discussion). *Internat. Statist. Rev.* **71**, 593-627.

Prášková, Z. (1984). On the rate of convergence in Samford-Durbin sampling from a finite population. *Statist. Decisions* **2**, 339-350.

Rao, J. N. K. (1973). Double sampling for stratification and analytical surveys. *Biometrika* **60**, 125-133.

Schenker, N. and Welsh, A. H. (1988). Asymptotic results for multiple imputation. *Ann. Statist.* **16**, 1550-1566.

Scott, A. J. and Wu, C. F. J. (1981). On the asymptotic distribution of ratio and regression estimators. *J. Amer. Statist. Assoc.* **76**, 98-102.

Sen, P. K. (1988). Asymptotics in finite population sampling. *Handbook of Statistics* **6**, 291-331.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Department of Statistics, 333-6356 Agriculture Road, University of British Columbia, Vancouver BC V6T1Z2, Canada.

E-mail: jhchen@stat.ubc.ca

School of Mathematics and Statistics, Carleton University, 1125 Colonel by Drive, Ottawa, On K1S5B6, Canada.

E-mail: jrao@math.carelton.ca

(Received August 2004; accepted November 2005)