Austrian Journal of Statistics
July 2021, Volume 50, 1–12.
http://www.ajs.or.at/
doi:10.17713/ajs.v50i3.631

On the Strong Law of Large Numbers for Nonnegative Random Variables. With an Application in Survey Sampling

Tobias Schoch
University of Applied Sciences Northwestern Switzerland
School of Business, Olten

Abstract

Strong laws of large numbers with arbitrary norming sequences for nonnegative not necessarily independent random variables are obtained. From these results we establish, among other things, stability results for weighted sums of nonnegative random variables. A survey sampling application is provided on strong consistency of the Horvitz–Thompson estimator and the ratio estimator.

Keywords: arbitrary normalization, survey sampling, Horvitz–Thompson estimator, ratio estimator.

1. Introduction

Consider the sequence $\{X_i, i \geq 1\}$ of square integrable random variables (r.v.) that take values in $[0, \infty)$. Put $S_n = \sum_{i \leq n} X_i$, and denote expectation and variance by, respectively, $\mathbb{E}X_i$ and $\mathbb{V}X_i$ for all i. We say that the sequence $\{X_i, i \geq 1\}$ satisfies the strong law of large numbers (SLLN) if $(S_n - \mathbb{E}S_n)/n \to 0$ almost surely (a.s.) as $n \to \infty$.

There exists a considerably large body of SLLN for nonnegative r.v.'s. Two strands of research are noteworthy: the results of (i) Etemadi (1983a,b) or Csörgő, Tandori, and Totik (1983) and (ii) the Petrov-type approach (which is inspired by Petrov (1969), hence the name); see e.g. Korchevsky (2015), Kuczmaszewska (2016) or Petrov (2009). Korchevsky (2015) provides a recent summary of the latter approach and Chen and Sung (2016) establish a theorem which unifies both strands. In what follows, we pursue the approach of Etemadi (1983a), which can be inferred from his theorem.

Theorem A (Etemadi, 1983a, Thm. 1)**.** *Let $\{X_i, i \geq 1\}$ be a sequence of nonnegative r.v.'s with finite second moment such that*

(i) $\sup_{i \geq 1} \mathbb{E}X_i < \infty$,

(ii) $\mathbb{E}[X_i X_j] \leq \mathbb{E}X_i \mathbb{E}X_j$ *for all $j > i$, and*

(iii) $\sum_{i=1}^{\infty} \mathbb{V}X_i/i^2 < \infty$.

Then as $n \to \infty$,

$$\frac{S_n - \mathbb{E}S_n}{n} \to 0 \text{ a.s.} \tag{1}$$

Remark. Walk (2005, Thm. 1) gave a generalization of Theorem A, having replaced hypothesis (ii) and (iii) by the weaker assumption $\sum_{i=1}^{\infty} \mathbb{V}(X_1 + \cdots + X_i)/i^3 < \infty$.

The two aforementioned strands differ primarily in terms of the assumptions imposed on the pth moments ($p \geq 1$) of the r.v.'s. The Petrov-type results replace hypotheses (ii) and (iii) of Theorem A and require instead that Petrov's condition holds (cf. Korchevsky 2015; Kuczmaszewska 2016)

$$\mathbb{E}|S_n - \mathbb{E}S_n|^2 = \mathcal{O}\left(\frac{n^2}{\psi(n)}\right) \qquad \text{for some } \psi \in \Psi_c. \tag{2}$$

The set of functions Ψ_c consists of all real-valued functions $\psi(x)$ that are (i) positive and nondecreasing in the domain of $x > x_0$ for some positive real x_0, and (ii) ψ is such that the series $\sum_n 1/(n\psi(n))$ is convergent. The value x_0 is not assumed to be the same for different ψ-functions. Examples of functions in Ψ_c are x^δ and $(\log x)^{1+\delta}$ for any $\delta > 0$.

Let $\{a_n, n \geq 1\}$ be an arbitrary monotone sequence of positive real numbers, $0 < a_n \uparrow \infty$ as $n \to \infty$. Korchevsky (2015) formulated the generalized Petrov condition,

$$\mathbb{E}|S_n - \mathbb{E}S_n|^p = \mathcal{O}\left(\frac{a_n^p}{\psi(a_n)}\right) \quad \text{for some } \psi \in \Psi_c \text{ and } p \geq 1. \tag{3}$$

Observe that in (3) the classical normalization n is replaced by the arbitrary norming sequence $\{a_n, n \geq 1\}$. Under the assumptions in (3) and the additional hypothesis $\mathbb{E}S_n = \mathcal{O}(a_n)$, Korchevsky (2015, Thm. 1) proved the following SLLN,

$$\frac{S_n - \mathbb{E}S_n}{a_n} \to 0 \text{ a.s.} \qquad \text{as } n \to \infty. \tag{4}$$

In the main section, we provide an extension of Theorem A to an arbitrary norming sequence in place of the classical one—similar to the Petrov-type result in (4). Our theorem provides two interesting corollaries: an SLLN for weighted sums and a result on the strong stability of sums of nonnegative r.v.'s. In addition, we give a generalization of Wu's (1981) Lemma 2 (see Lem. B) that holds for nonnegative but not necessarily independent r.v.'s.

Lemma B (Wu, 1981, Lem. 2)**.** *Let* $\{X_i, i \geq 1\}$ *be a sequence of independent r.v.'s with* $\mathbb{E}X_i = 0$ *and* $\mathbb{V}X_i = \sigma_i^2 < \infty$. *Suppose a sequence of positive real numbers* $\{A_n, n \geq 1\}$ *such that*

$$A_n \to \infty, \qquad \limsup_{n\to\infty} \frac{\left(\sum_{i\leq n} \sigma_i^2\right)^{1/2+\delta}}{A_n} < \infty \qquad \text{for some } \delta > 0. \tag{5}$$

Then as $n \to \infty$,

$$\frac{S_n}{A_n} \to 0 \text{ a.s.} \tag{6}$$

Under the hypothesis of independence, Lemma B proved to be a very popular and valuable tool in a large number of papers; see e.g., Fahrmeir and Kaufmann (1985) in the context of the generalized linear model or Xie and Yang (2003) who provide a generalization for double array sequences. In view of the wide applicability of Lemma B, our Corollary 5 can be useful in its own right.

2. Main results

Unless otherwise stated, we adhere to the following notation in this section. Denote by

$\{X_i, i \geq 1\}$ a sequence of nonnegative r.v.'s with finite second moment, (7)

$\{w_i, i \geq 1\}$ a sequence of nonnegative real numbers, (8)

$\{a_i, i \geq 1\}$ a monotone sequence of positive real numbers, (9)

where $0 < a_i \uparrow \infty$ as $i \to \infty$. Inspired by the methods used in Etemadi (1983a) and Kuczmaszewska (2016), we prove the following theorem which replaces the classical normalization in Theorem A by an arbitrary normalizing sequence and thus generalizes the theorem.

Theorem 1. *Consider $\{X_i, i \geq 1\}$ and $\{a_i, i \geq 1\}$ in, respectively, (7) and (9). Suppose the hypotheses,*

(i) $\mathbb{E}S_n = \mathcal{O}(a_n)$,

(ii) $\mathbb{E}[X_iX_j] \leq \mathbb{E}X_i\mathbb{E}X_j$ *for all* $j > i$,

(iii) $\sum_{i=1}^{\infty} \mathbb{V}X_i/a_i^2 < \infty$,

then as $n \to \infty$,

$$\frac{S_n - \mathbb{E}S_n}{a_n} \to 0 \text{ a.s.} \tag{10}$$

Remark. The above theorem can be seen as a special case of the more general result in Chandra and Goswami (1992, Thm. 1). These authors require the existence of a double sequence $\{\rho_{ij}\}$ of real values such that $\mathbb{V}S_n \leq \sum_{i=1}^{n}\sum_{j=1}^{n} \rho_{ij}$ for all $n \geq 1$ with $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \rho_{ij}/a_{\max(i,j)}^2 < \infty$ instead of our hypotheses (ii) and (iii).

An immediate corollary of Theorem 1 is the following SLLN for weighted sums; cf. Etemadi (1983b). Consider $\{w_i, i \geq 1\}$ in (8), put

$$W_n = \sum_{i=1}^{n} w_i, \qquad T_n = \sum_{i=1}^{n} w_iX_i, \tag{11}$$

and let $\{w_i, i \geq 1\}$ be such that

$$\frac{w_n}{W_n} \to 0 \quad \text{and} \quad W_n \to \infty \quad \text{as } n \to \infty. \tag{12}$$

Corollary 2. *Let $\{X_i, i \geq 1\}$ be given in (7), and let $\{w_i, i \geq 1\}$ in (8) satisfy (12). Consider T_n and W_n in (11). If*

(i) $\mathbb{E}T_n = \mathcal{O}(W_n)$,

(ii) $\mathbb{E}[X_iX_j] \leq \mathbb{E}X_i\mathbb{E}X_j$ *for all* $j > i$,

(iii) $\sum_{i=1}^{\infty} w_i^2\mathbb{V}[X_i]/W_i^2 < \infty$,

then as $n \to \infty$,

$$\frac{T_n - \mathbb{E}T_n}{W_n} \to 0 \text{ a.s.} \tag{13}$$

The proof of Corollary 2 is straightforward. Put $a_i \equiv W_i$ and apply Theorem 1 using $\{w_iX_i, i \geq 1\}$ and $\{W_i, i \geq 1\}$ in place of $\{X_i, i \geq 1\}$ and $\{a_i, i \geq 1\}$, hence the assertion obtains.

Another interesting corollary of Theorem 1 refers to the (strong) stability of sums of r.v.'s. Consider $\{X_i, i \geq 1\}$ in (7) and let $S_n = \sum_{i \leq n} X_i$ be such that

$$\mathbb{E}S_n \to \infty \quad \text{and} \quad \frac{\mathbb{E}X_n}{\mathbb{E}S_n} \to 0 \quad \text{as } n \to \infty. \tag{14}$$

Corollary 3. *Let $\{X_i, i \geq 1\}$ in (7) satisfy (14). If*

(i) $\mathbb{E}[X_i X_j] \leq \mathbb{E}X_i \mathbb{E}X_j$ *for all* $j > i$,

(ii) $\sum_{n=1}^{\infty} \mathbb{V}[X_n]/(\mathbb{E}S_n)^2 < \infty$,

then as $n \to \infty$,

$$\frac{S_n}{\mathbb{E}S_n} \to 1 \ a.s. \tag{15}$$

The proof of Corollary 3 obtains from Theorem 1. Define $Y_i = X_i/\mathbb{E}X_i$ (where we assume without loss of generality that $\mathbb{E}X_i > 0$ for all i), and put $w_i = \mathbb{E}X_i$. Note that $T_n = \sum_{i\leq n} w_i Y_i = \sum_{i \leq n} X_i$ and $W_n = \sum_{i\leq n} w_i = \mathbb{E}S_n$, hence the assertion follows by application of Corollary 2.

The next theorem provides an SLLN under a suitable normalization when the sum of the variances of the elements in the partial sum S_n,

$$B_n = \sum_{i=1}^{n} \mathbb{V}X_i, \tag{16}$$

grows without bounds (i.e. $B_n \to \infty$ as $n \to \infty$). The result is comparable with Theorem 1 in Petrov (1969) except that it does not require the r.v.'s to be independently distributed.

Theorem 4. *Let $\{X_i, i \geq 1\}$ be defined as in (7). Suppose the hypotheses*

(i) $\mathbb{E}S_n = \mathcal{O}(a_n)$;

(ii) $\mathbb{E}[X_i X_j] \leq \mathbb{E}X_i \mathbb{E}X_j$ *for all* $j > i$.

Let $B_n \to \infty$ *as* $n \to \infty$, *then*

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{B_n \psi(B_n)}} \to 0 \ a.s. \qquad \text{for any } \psi \in \Psi_c. \tag{17}$$

As a corollary of Theorem 4 we obtain the next result, which can be regarded as a generalization of Lemma 2 in Wu (1981) – see Lemma B – under the hypothesis of nonnegative but not necessarily independent r.v.'s.

Corollary 5. *Suppose that the hypotheses of Theorem 4 hold. Let $\{w_n, n \geq 1\}$ be a sequence of positive real numbers such that*

$$w_n \to \infty, \qquad \text{and} \qquad \limsup_{n\to\infty} \frac{B_n^{1/2+\delta}}{w_n} < \infty \qquad \text{for some } \delta > 0. \tag{18}$$

Then as $n \to \infty$,

$$\frac{S_n - \mathbb{E}S_n}{w_n} \to 0 \ a.s. \tag{19}$$

3. Application

We provide an application in the context of survey sampling, where the restriction to nonnegative r.v.'s is quite natural. Our focus is on (strong) consistency of the Horvitz–Thompson (HT) estimator and the ratio estimator for the population mean.

Suppose a finite population U_t of units labeled $i = 1, \ldots, N_t$ (the index t is unimportant for the moment). Associated with the ith element in U_t is the (nonnegative) study variable y_i, which is unknown for all $i \in U_t$. The goal is to estimate the population y-mean, $\bar{y}_{U_t} = \sum_{i\in U_t} y_i/N_t$.

Sampling

Suppose a without-replacement sample s_t of fixed (i.e. non-random) size n_t, $0 < n_t < N_t$, is drawn from U_t, and error-free measurements y_i, $i \in s_t$, are recorded. Denote by $\mathscr{S}_t = \{i \in U_t : \mathbb{1}_{i \in s_t} = 1\}$ a random set, where $\mathbb{1}_{i \in s_t}$ is the sample-inclusion indicator which equals one (zero) in the presence (absence) of element i in the sample. The logical status of $\mathbb{1}_{i \in s_t}$ is that of a r.v.; in the context of sampling, y_i is regarded as a fixed quantity. Expectation with respect to (w.r.t.) the randomization distribution (abbreviated by p-distr.) is denoted by $\mathbb{E}_p$; the operators $\mathbb{V}_p$ and Cov_p shall be defined analogously. Each sample s_t is regarded as a realization of $\mathscr{S}_t$, the probability distribution of which is called sampling design. The first- and second-order sample-inclusion probabilities are denoted by, respectively, $\pi_{i,t} = \mathbb{E}_p(\mathbb{1}_{i \in s_t})$ and $\pi_{ij,t} = \mathbb{E}_p(\mathbb{1}_{i \in s_t} \mathbb{1}_{j \in s_t})$, $i, j \in U_t$; the covariance of the sample-inclusion indicators will be denoted by

$$\Delta_{ij,t} = \mathrm{Cov}_p(\mathbb{1}_{i \in s_t}, \mathbb{1}_{j \in s_t}) = \pi_{ij,t} - \pi_{i,t}\pi_{j,t}. \tag{20}$$

By our hypothesis of fixed-size sampling designs, the identities (cf. Särndal, Swensson, and Wretman 1992, Result 2.6.2),

$$\sum_{i=1}^{N_t} \pi_{i,t} = n_t, \quad \text{and} \quad \sum_{j=1,\, j \neq i}^{N_t} \pi_{ij,t} = (n_t - 1)\pi_{i,t} \quad \text{for all } i \in U_t, \tag{21}$$

obtain, which imply

$$\sum_{j=1,\, j \neq i}^{N_t} \Delta_{ij,t} = \pi_{i,t}(\pi_{i,t} - 1) \quad \text{for all } i \in U_t. \tag{22}$$

Asymptotic framework

Our asymptotic framework is for the most part that of Robinson and Särndal (1983); see also Isaki and Fuller (1982). Let $\{U_t, t \geq 1\}$ denote the nested sequence of populations U_t of size N_t; the samples s_t form an analogous but not necessarily nested sequence. All limiting processes will be taken as $t \to \infty$. In what follows, we will impose some assumptions on the behavior of the sampling design.

Assumption A. *(i) The r.v.'s $\mathbb{1}_{i \in s_t}$ are independent of the study variable y_i for all $i \in U_t$.*

(ii) There exists a constant λ such that $\min_{i \in U_t}(\pi_{i,t}) \geq \lambda > 0$ for all N_t.

Assumption B. *For all elements $i, j \in U_t$ such that $i \neq j$, $\Delta_{ij,t} \leq 0$.*

Designs that satisfy hypothesis (i) of Assumption A are called non-informative designs (Särndal *et al.* 1992, Remark 2.4.4). Assumption B refers to the covariance of the sample-inclusion indicators. We say that two observations $i, j \in U_t$, $i \neq j$, are "tied" if $\Delta_{ij,t} > 0$. As Robinson (1982, 237) shows (for the case of the Horvitz–Thompson estimator), tying can lessen the rate of convergence. For extreme types of designs (e.g., when an ideal coin is tossed to decide whether the sample should consist entirely of even natural numbers, or of odd ones), the Horvitz–Thompson estimator will not be consistent. Clearly, our Assumption B does exclude such situations. In fact, the vast majority of single-stage without-replacement sampling designs obeys Assumption B; e.g., simple random sampling (SRS), stratified SRS, designs with (unequal) probabilities proportional to certain size measures (πps-designs), etc. However, Assumption B usually fails for one-stage cluster sampling and multistage designs (due to positive correlation within clusters).

The next assumption is concerned with the behavior of the sequence $\{y_i, i \geq 1\}$ of real numbers.

Assumption C. *Let $\{y_i, i \geq 1\}$ be such that*

$$\lim_{t \to \infty} \frac{1}{N_t} \sum_{i=1}^{N_t} y_i < \infty \qquad \textit{and} \qquad \lim_{t \to \infty} \frac{1}{N_t} \sum_{i=1}^{N_t} y_i^2 < \infty.$$

Horvitz–Thompson estimator

We shall derive sufficient conditions for the Horvitz–Thompson type estimator of the mean to be strongly consistent for the population mean. Denote the HT estimator of the population y-total $T_{y,t} = \sum_{i\in U_t} y_i$ by

$$\hat{T}_{y,t} = \sum_{i\in U_t} \frac{y_i}{\pi_{i,t}} \mathbb{1}_{i\in s_t}. \tag{23}$$

The HT estimator of the x-total, $T_{x,t}$, shall be defined in the same way. An obvious estimator of the population y-mean, $\bar{y}_{U_t}$, is the HT-type estimator given by

$$\hat{\bar{y}}_{HT,t} = \frac{\hat{T}_{y,t}}{N_t}. \tag{24}$$

Robinson (1982) studied mean square (weak) consistency of this estimator. For our purposes, it is instructive to present Robinson's main arguments (see his Thm. 1 and 2) in condensed form. To this end, define the sets, for all $i = 1, \ldots, N_t$,

$$D_i^+ = \{j \in U_t :\ \Delta_{ij,t} > 0\}, \quad \text{and} \quad D_i^- = \{j \in U_t :\ \Delta_{ij,t} \le 0\}, \tag{25}$$

having suppressed the index t for the sake of simplicity. Note that these definitions enable us to "separate" tied form untied observations. Let $D^+ = \cup_{i=1}^{N_t} D_i^+$ and $D^- = \cup_{i=1}^{N_t} D_i^-$. Under the hypothesis of fixed-size designs, the variance of $\hat{\bar{y}}_{HT,t}$ is given by (cf. Särndal *et al.* 1992, Result 2.8.2)

$$N_t^2 \mathbb{V}_p\big(\hat{\bar{y}}_{HT,t}\big) = -\frac{1}{2}\sum\sum_{i\neq j} \Delta_{ij,t}\left(\frac{y_i}{\pi_{i,t}} - \frac{y_j}{\pi_{j,t}}\right)^2 \le -\frac{1}{2}\sum\sum_{i\in D^-} \Delta_{ij,t}\left(\frac{y_i}{\pi_{i,t}} - \frac{y_j}{\pi_{j,t}}\right)^2$$

(having exploited that the double sum over the index set D^- is non-positive), moreover,

$$\le -\sum\sum_{i\in D^-} \Delta_{ij,t}\frac{y_i^2}{\pi_{i,t}^2} = -\sum_{i=1}^{N_t} \frac{y_i^2}{\pi_{i,t}^2}\sum_{D_i^-}\Delta_{ij,t} = -\sum_{i=1}^{N_t} \frac{y_i^2}{\pi_{i,t}^2}\left(\sum_{j=1,\, j\neq i}^{N_t} \Delta_{ij,t} - \sum_{D_i^+}\Delta_{ij,t}\right),$$

which then, together with (21) and (22), implies

$$= \sum_{i=1}^{N_t} \frac{y_i^2}{\pi_{i,t}^2}\left(\pi_{i,t}(1-\pi_{i,t}) + \sum_{D_i^+}\Delta_{ij,t}\right) \le \sum_{i=1}^{N_t} \frac{y_i^2}{\pi_{i,t}^2}\left(\pi_{i,t} + \sum_{D_i^+}\Delta_{ij,t}\right). \tag{26}$$

Under our Assumption C, Robinson (1982, Thm. 2) establishes from (26) that

$$N_t^2 \mathbb{V}_p\big(\hat{\bar{y}}_{HT,t}\big) = \mathcal{O}(N_t)\big(\min_{i\in U_t}\pi_{i,t}\big)^{-1}\left(1 + \big(\min_{i\in U_t}\pi_{i,t}\big)^{-1}\max_{i\in U_t}\sum_{D_i^+}\Delta_{ij,t}\right) \tag{27}$$

which implies, for large N_t,

$$\hat{\bar{y}}_{HT,t} - \bar{y}_{U_t} = \mathcal{O}(N_t^{-1/2}\delta^{-1/2}) + \mathcal{O}(N_t^{-1/2}\delta^{-1}\zeta), \tag{28}$$

where

$$\delta = \min_{i\in U_t}\pi_{i,t}, \qquad \text{and} \qquad \zeta = \max_{i\in U_t}\sum_{D_i^+}\Delta_{ij,t}. \tag{29}$$

Now, sufficient conditions of (weak) consistency of $\hat{\bar{y}}_{HT,t}$ for the population mean are i) $\delta N_t \to \infty$ and ii) $\zeta = o(\delta {N_t}^{1/2})$ as $t \to \infty$. Note that condition ii) means that the observations are not strongly tied. Moreover, as Robinson (1982, 237) points out, it is desirable if $\Delta_{ij,t} \le 0$ (cf. our Assumption B), for then the sets D_i^+ are empty sets for all $i \in U_t$, hence $\zeta = 0$. If in addition the terms $\pi_{i,t}N_t/n_t$ are bounded away from zero it follows that $\hat{\bar{y}}_{HT,t} - \bar{y}_{U_t}$ is $\mathcal{O}({n_t}^{-1/2})$, implying weak consistency as $n_t \to \infty$ $(t \to \infty)$.

The next theorem establishes strong consistency of the Horvitz–Thompson type estimator $\hat{\bar{y}}_{HT,t}$ under hypotheses similar to the ones used by Robinson (1982).

Theorem 6. *Let the sequences of populations $\{U_t, t \geq 1\}$ and samples $\{s_t, t \geq 1\}$ be as described. Suppose Assumptions A and B. Let $\{y_i, i \geq 1\}$ be a sequence of nonnegative real numbers that satisfies Assumption C. Then, as $t \to \infty$,*

$$\hat{\bar{y}}_{HT,t} - \bar{y}_{U_t} \to 0 \quad a.s.. \tag{30}$$

Remarks. (i) In contrast to Robinson (1982, Thm. 2), our Theorem 6 requires $\{y_i, i \geq 1\}$ to be a sequence of *nonnegative* real numbers. With this additional assumption, however, we obtain strong instead of weak consistency. The restriction to non-negativity does not limit the applicability of the result in any noticeable manner as population totals are only meaningful population characteristics in conjunction with nonnegative data.

(ii) The hypotheses of Theorem 6 can be slightly weakened. The assertion is still true when the y_i^2's are allowed to grow (although not too much); thus, instead of Assumption C, we require that $\sum_{i \leq N_t} y_i^2 = \mathcal{O}(N_t^{2-\delta})$ with $\delta > 0$ and $\sum_{i \leq N_t} y_i = \mathcal{O}(N_t)$ as $N_t \to \infty$.

(iii) The results of Robinson (1982) establish error estimates (e.g., $\hat{\bar{y}}_{HT,t} - \bar{y}_{U_t}$ is of order $\mathcal{O}({n_t}^{-1/2})$ in probability), while our results do not. However, we can easily obtain error estimates by specifying assumptions in terms of the asymptotic behavior of $\pi_{i,t} N_t / n_t$.

Ratio estimator

Let $\{x_i, i \geq 1\}$ denote a sequence of nonnegative real numbers that are known for all $i \in U_t$; hence, the population x-mean, $\bar{x}_{U_t}$, is a known quantity. We shall assume that the population-level relationship between y_i and x_i can be *approximated* by the heteroscedastic model ξ: $y_i = x_i \beta + E_i$, $i \in U_t$, where $\mathbb{E}_\xi(E_i \mid x_i) = 0$ and

$$\mathbb{E}_\xi(E_i E_j \mid x_i, x_j) = \begin{cases} x_i \sigma^2 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{31}$$

The parameters $\beta \in \mathbb{R}^+$ and $\sigma^2 \in \mathbb{R}^+$ are not known. Note that model ξ is merely motivated as an *assisting model*. In this context, the ratio estimator of the population y-mean,

$$\hat{\bar{y}}_{\text{rat},t} = \bar{x}_{U_t} \hat{\beta}_t, \quad \text{where} \quad \hat{\beta}_t = \frac{\hat{\bar{y}}_{HT,t}}{\hat{\bar{x}}_{HT,t}}, \tag{32}$$

is among survey statisticians' preferred estimators because of its simplicity. Under model ξ, estimator $\hat{\bar{y}}_{\text{rat},t}$ proves to be more efficient as an estimator of $\bar{y}_{U_t}$ than the Horvitz–Thompson type estimator of the mean, $\hat{\bar{y}}_{HT,t}$.

The following result establishes that the ratio estimator $\hat{\bar{y}}_{\text{rat},t}$ is a strongly consistent estimator of $\bar{y}_{U_t}$ in the above asymptotic framework.

Theorem 7. *Let the sequences of populations $\{U_t, t \geq 1\}$ and samples $\{s_t, t \geq 1\}$ be as described, and suppose that Assumptions A and B hold. Let $\{y_i, i \geq 1\}$ and $\{x_i, i \geq 1\}$ denote, respectively, a nonnegative and a positive sequence of real numbers; each sequence is assumed to satisfy Assumption C. Then, as $t \to \infty$,*

$$\hat{\bar{y}}_{rat,t} - \bar{y}_{U_t} \to 0 \quad a.s.\ (w.r.t.\ p\text{-}distr.)$$

Note that $\hat{\bar{y}}_{rat,t}$ is a strongly consistent estimator of $\bar{y}_{U_t}$, whether model ξ holds or not. We can do without the model because we have specified the behavior of the sequence $\{y_i, i \geq 1\}$ instead.

The next result makes explicit usage of model ξ. It is important to note that the model ξ and the sampling design p induce separate stochastic behavior. It is thus natural to study strong consistency of the estimator $\hat{\beta}_t$ for β (more precisely, the sequence $\{\hat{\beta}_t, t \geq 1\}$) under the *compound* design–model distribution (abbreviated by ξp-distr.).

Theorem 8. *Let the sequences of populations $\{U_t, t \geq 1\}$ and samples $\{s_t, t \geq 1\}$ be as described, and suppose Assumptions A and B. Let $\{x_i, i \geq 1\}$ denote a sequence of positive real numbers that satisfies Assumption C. Suppose model ξ and consider $\hat{\beta}_t$ defined in (32). Then, as $t \to \infty$,*

$$\hat{\hat{y}}_{rat,t} \to \bar{y}_{U_t} \quad \textit{a.s. (w.r.t. } \xi p\textit{-distr.).}$$

Theorems 7 and 8 extend some of the results in Robinson and Särndal (1983) under the hypothesis of nonnegative r.v.'s.

4. Proofs

Proof of Thm. 1. Let $\eta > 1$ be a real number and $m, n \in \mathbb{N}$. For all $m \geq 1$, define

$$n_m = \inf\big\{n : a_n \geq \eta^m\big\}. \tag{33}$$

Note that $\{n_m, m \geq 1\}$ is a monotone sequence of positive integers, $0 < n_1 \leq n_2 \leq \cdots \leq n_m \uparrow \infty$ as $m \to \infty$ since $\{a_n, n \geq 1\}$ is monotone and $0 < a_n \uparrow \infty$ as $n \to \infty$. By Chebyshev's inequality and hypothesis (ii), for any $\epsilon > 0$,

$$\begin{aligned}\epsilon^2 \sum_{m=1}^{\infty} \mathbb{P}\big\{\big|S_{n_m} - \mathbb{E}S_{n_m}\big| > \epsilon \cdot a_{n_m}\big\} &\leq \sum_{m=1}^{\infty} \frac{\mathbb{V}S_{n_m}}{a_{n_m}^2} \\ &\leq \sum_{m=1}^{\infty} \frac{1}{a_{n_m}^2} \sum_{i=1}^{n_m} \mathbb{V}X_i = \sum_{i=1}^{\infty} \mathbb{V}[X_i] t_i,\end{aligned} \tag{34}$$

where

$$t_i = \sum_{m \in M} \frac{1}{a_{n_m}^2} \qquad \text{and} \qquad M_i = \{m : n_m \geq i\}. \tag{35}$$

Next, we use an argument similar to the one in Etemadi (1983b). Since $a_n \uparrow \infty$ as $n \to \infty$, it follows that $a_{n_m} \sim \eta^m$ for all large m. Thus for some constant $C_1 > 0$ and every $i = 1, 2, 3, \ldots$,

$$M_i = \big\{m : n_m \geq i\big\} \subset \big\{m : a_{n_m} \geq a_i\big\} \subset \big\{m : C_1 \eta^m \geq a_i\big\} =: M_i^*, \text{ say,} \tag{36}$$

since $n_m \geq i$ and monotonicity of $\{a_n, n \geq 1\}$ imply $a_{n_m} \geq a_i$. By this and (33), the geometric series on the far left in (35) obtains for all $i \geq 1$

$$t_i = \sum_{m \in M_i} a_{n_m}^{-2} \leq \sum_{m \in M_i} \eta^{-2m} \leq \sum_{m \in M_i^*} \eta^{-2m} = \frac{C_\eta}{\eta^{2m_i}}, \tag{37}$$

where $m_i = \inf M_i^*$ and $C_\eta = (1 - 1/\eta^2)^{-1}$ is a constant. From (36) and the fact that $m_i \in M_i^*$, it is easy to see that $C_1^2/a_i^2 \geq \eta^{-2m_i}$. This together with hypothesis (iii) and (34) implies

$$\epsilon^2 \sum_{m=1}^{\infty} \mathbb{P}\big\{\big|S_{n_m} - \mathbb{E}S_{n_m}\big| > \epsilon \cdot a_{n_m}\big\} \leq C_1^2 C_\eta \sum_{i=1}^{\infty} \frac{\mathbb{V}X_i}{a_i^2} < \infty. \tag{38}$$

Since $\epsilon > 0$ is arbitrary, (38) and the Borel–Cantelli lemma imply $\mathbb{P}\{|S_{n_m} - \mathbb{E}S_{n_m}| > \epsilon \cdot a_{n_m} \text{ i.o.}\} = 0$ (where i.o. stands for infinitely often), therefore as $m \to \infty$

$$\frac{S_{n_m} - \mathbb{E}S_{n_m}}{a_{n_m}} \to 0 \text{ a.s.} \tag{39}$$

Thus we showed the desired result for the subsequence. This result can be extended to the whole sequence. Let $m \in [n_m, n_{m+1})$. By monotonicity of S_m, we have

$$\frac{|S_m - \mathbb{E}S_m|}{a_m} \leq \frac{|S_{n_{m+1}} - \mathbb{E}S_{n_m}|}{a_{n_m}} \leq \frac{a_{n_{m+1}}}{a_{n_m}} \frac{|S_{n_{m+1}} - \mathbb{E}S_{n_{m+1}}|}{a_{n_{m+1}}} + \frac{|\mathbb{E}S_{n_{m+1}} - \mathbb{E}S_{n_m}|}{a_{n_m}}. \tag{40}$$

By (39) and since $a_{n_{m+1}}/a_{n_m} \to \eta$ (as $m \to \infty$), the first summand on the far right of (40) converges a.s. to zero as $m \to \infty$. By hypothesis (i), and for m large enough, there is a constant C_2 such that $\mathbb{E}S_m \le C_2 a_m$. Hence, for large m, we have obtain for the second summand in (40) that

$$\frac{|\mathbb{E}S_{n_{m+1}} - \mathbb{E}S_{n_m}|}{a_{n_m}} \le C_2 \frac{|a_{n_{m+1}} - a_{n_m}|}{a_{n_m}} \to C_2(\eta - 1) \qquad \text{(as } m \to \infty\text{)}, \tag{41}$$

thus,

$$\limsup_{m\to\infty} \left| \frac{S_m - \mathbb{E}S_m}{a_m} \right| \le C_2(\eta - 1). \tag{42}$$

Likewise we obtain a lower bound of the left-hand side in (40), which then provides that $\liminf_{m\to\infty} |(S_m - \mathbb{E}S_m)/a_m| \ge C_2(1 - 1/\eta)$. Therefore,

$$\left(1 - \frac{1}{\eta}\right) C_2 \le \liminf_{m\to\infty} \left| \frac{S_m - \mathbb{E}S_m}{a_m} \right| \le \limsup_{m\to\infty} \left| \frac{S_m - \mathbb{E}S_m}{a_m} \right| \le (\eta - 1)C_2,$$

and as $\eta > 1$ is arbitrary, $|(S_m - \mathbb{E}S_m)/a_m| \xrightarrow{\eta\downarrow 1} 0$ a.s. This completes the proof. □

Before we give the proof of Theorem 4, we need the following lemma.

Lemma 9 (generalized Abel–Dini Thm., cf. Petrov, 1969, Lem. 1)**.** *Let $\{b_i, i \ge 1\}$ denote a sequence of nonnegative real numbers, put $B_n = \sum_{i\le n} b_i$, and assume that $B_n \to \infty$ as $n \to \infty$. Then the series $\sum_n^\infty b_n/(B_n \psi(B_n))$ converges for any $\psi \in \Psi_c$.*

With Lemma 9 we are in the position to give the proof of Thm. 4.

Proof of Thm. 4. Put $b_n = \sqrt{B_n \psi(B_n)}$, where $B_n = \sum_{i\le n} \mathbb{V}X_i$, and note that $b_n \uparrow \infty$ as $n \to \infty$. Application of Lemma 9 implies hypothesis (iii) of Theorem 1, hence the assertion follows by Kronecker's lemma. □

Proof of Thm. 6. To facilitate the proof, we introduce a lemma on convergence of a series of ratios of real numbers. The class of functions Ψ_c appearing in the lemma is defined in the text following Equation (2); see Section 1.

Lemma 10 (Korchevsky, 2015, Thm. 5)**.** *(i) Let $\{a_i, i \ge 1\}$ be a non-decreasing, unbounded sequence of positive real numbers such that*

$$\frac{a_{2n}}{a_n} \le q \qquad \textit{(for sufficiently large } n\textit{)},$$

where q is a constant.

(ii) Let $\{b_i, i \ge 1\}$ be a sequence of non-negative real numbers and suppose that

$$\sum_{i=1}^{n} b_i = \mathcal{O}\left(\frac{a_n^2}{\psi(n)}\right) \qquad \textit{for some } \psi \in \Psi_c.$$

Under the hypotheses i) and ii), we have

$$\sum_{i=1}^{\infty} \frac{b_i}{a_i^2} < \infty.$$

The proof of Theorem 6 now follows by application of Theorem 1 and Lemma 10. Define the r.v. $Y_i = y_i \mathbb{1}_{i\in s_t}/\pi_{i,t}$, for all $i \in U_t$, and put $\hat{T}_{N_t} = \sum_{i\le N_t} Y_i$. Hypotheses i) and ii) of Theorem 1 are easily verified under our assumptions. Let us consider hypothesis iii). By Assumption A, there exist $\lambda > 0$ such that

$$\sum_{i=1}^{\infty} \frac{\mathbb{V}_p Y_i}{i^2} = \sum_{i=1}^{\infty} \frac{y_i^2}{i^2}\left(\frac{1-\pi_{i,t}}{\pi_{i,t}}\right) \le \sum_{i=1}^{\infty} \frac{y_i^2}{i^2 \pi_{i,t}} \le \left(\min_{1\le i\le N_t}(\pi_{i,t})\right)^{-1} \sum_{i=1}^{\infty} \frac{y_i^2}{i^2} \le \frac{1}{\lambda}\sum_{i=1}^{\infty}\frac{y_i^2}{i^2}. \tag{43}$$

Now, by application of Lemma 10 with $a_i = i$, $b_i = y_i^2$, and $\psi(n)$ defined as $n \mapsto n^\delta$ with $\delta > 0$, we conclude that

$$\sum_{i=1}^{\infty} \frac{y_i^2}{i^2} < \infty$$

if

$$\sum_{i=1}^{N_t} y_i^2 = \mathcal{O}(N_t^{2-\delta}). \tag{44}$$

Note that Assumption C can be written as $\sum_{i\le N_t} y_i^2 = \mathcal{O}(N_t)$. From this we see that (44) is (a fortiori) verified. Thus, the lemma and (43) imply hypothesis iii) $\lim_{t\to\infty}\sum_{i\le N_t}\mathbb{V}_p(Y_i)/i^2 < \infty$ of Theorem 1. □

Proof of Thm. 7. Under the Assumptions A, B, and C, Theorem 6 implies $(\hat{\bar{y}}_{HT,t}, \hat{\bar{x}}_{HT,t}) \to (\bar{y}_{U_t}, \bar{x}_{U_t})$ a.s. (as $t \to \infty$) w.r.t. p-distr.; thus, we conclude that

$$\hat{\bar{y}}_{rat,t} - \bar{y}_{U_t} = \bar{x}_{U_t}\left(\frac{\hat{\bar{y}}_{HT,t}}{\hat{\bar{x}}_{HT,t}} - \frac{\bar{y}_{U_t}}{\bar{x}_{U_t}}\right) \to 0 \text{ a.s. (as } t \to \infty)$$

by application of the continuous mapping theorem (see e.g., Van der Vaart 1998, Thm. 2.3). □

Proof of Thm. 8. For ease of simplicity, we write $\bar{x}_U$ for $\bar{x}_{U_t}$ and $\hat{\bar{x}}$ instead of $\hat{\bar{x}}_{HT,t}$ (the same applies for the y-means). Write

$$\hat{\bar{y}}_{rat} - \bar{x}_U\beta = \bar{x}_U\big(\hat{\beta} - \beta\big), \qquad \text{where} \quad \hat{\beta} = \frac{\hat{\bar{y}}}{\hat{\bar{x}}}. \tag{45}$$

Under model ξ, we have $y_i = x_i\beta + E_i$. Substituting $x_i\beta + E_i$ for y_i in $\hat{\beta} = \hat{\bar{y}}/\hat{\bar{x}}$, we obtain

$$\hat{\beta} = \beta + \frac{\hat{\bar{e}}}{\hat{\bar{x}}}, \tag{46}$$

where

$$\hat{\bar{e}} = \frac{1}{N_t}\sum_{i=1}^{N_t} e_i \qquad \text{with} \qquad e_i = \frac{E_i \mathbb{1}_{i\in s_t}}{\pi_i}.$$

We now show that $\hat{\bar{e}} \to 0$ a.s. (w.r.t. the ξp-distr.) by application of Theorem 1. Note that the compound expectation is $\mathbb{E}_{\xi p}(e_i) = 0$ (for all $i = 1, 2, \ldots$). Also, with the help of the identity $\mathbb{V}_{\xi p}(e_i) = \mathbb{V}_\xi\{\mathbb{E}_p(e_i)\} + \mathbb{E}_\xi\{\mathbb{V}_p(e_i)\}$, we have (for all $i = 1, 2, \ldots$)

$$\mathbb{V}_{\xi p}(e_i) = \mathbb{V}_\xi(E_i) + \mathbb{E}_\xi\left(E_i^2\frac{1-\pi_i}{\pi_i}\right) = x_i\sigma^2 + \frac{1-\pi_i}{\pi_i}x_i\sigma^2 = \frac{x_i\sigma^2}{\pi_i}.$$

With this and Assumption A, there exists λ such that

$$\sum_{i=1}^{\infty}\frac{\mathbb{V}_{\xi p}(e_i)}{i^2} \le \frac{\sigma^2}{\lambda}\sum_{i=1}^{\infty}\frac{x_i}{i^2} < \infty,$$

where the last inequality follows from Assumption C and application of Lemma 10 with $a_i = i$, $b_i = x_i$, and $\psi(u) = u^2$. This implies hypothesis iii) of Theorem 1. Hypothesis i) is implied by Assumption C. Lastly, hypothesis ii) holds because $\text{cov}_{\xi p}(e_i e_j) = 0$ for all e_i and e_j, $i \neq j$. Hence, we conclude that $\hat{\bar{e}} \to 0$ a.s. (w.r.t. the ξp-distr.) for $t \to \infty$. Moreover, by application of Theorem 6, we have $\hat{\bar{x}} \to \bar{x}_U$ a.s. with respect to the p-distr. as $t \to \infty$. Thus, by application of the continuous mapping theorem (see e.g., Van der Vaart 1998, Thm. 2.3), we conclude that

$$\frac{\hat{\bar{e}}}{\hat{\bar{x}}} \to 0 \quad \text{a.s.}$$

which implies in (46) that $\hat{\beta} - \beta \to 0$ a.s. for $t \to \infty$. □

Acknowledgements and funding

I would like to thank the anonymous reviewers for numerous helpful suggestions which led to an improvement of the paper. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Chandra T, Goswami A (1992). "Cesàro Uniform Integrability and the Strong Laws of Large Numbers." *Sankhyā, Series A*, **54**, 215–231.

Chen P, Sung S (2016). "A Strong Law of Large Numbers for Nonnegative Random Variables and Applications." *Statistics & Probability Letters*, **118**, 80–86.

Csörgő S, Tandori K, Totik V (1983). "On the Strong Law of Large Numbers for Pairwise Independent Random Variables." *Acta Mathematica Hungarica*, **48**, 319–330.

Etemadi N (1983a). "On the Laws of Large Number for Nonnegative Random Variables." *Journal of Multivariate Analysis*, **13**, 187–193.

Etemadi N (1983b). "Stability of Sums of Weighted Nonnegative Random Variables." *Journal of Multivariate Analysis*, **13**, 360–365.

Fahrmeir L, Kaufmann H (1985). "Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models." *Annals of Statistics*, **13**, 342–368.

Isaki C, Fuller W (1982). "Survey Design under the Regression Superpopulation Model." *Journal of the American Statistical Association*, **77**, 89–96.

Korchevsky V (2015). "A Generalization of the Petrov Strong Law of Large Numbers." *Statistics & Probability Letters*, **104**, 102–108.

Kuczmaszewska A (2016). "Convergence Rate in the Petrov SLLN for Dependent Random Variables." *Acta Mathematica Hungarica*, **148**, 56–72.

Petrov V (1969). "On the Strengthened Law of Large Numbers." *Theory of Probability and its Applications*, **14**, 193–202.

Petrov V (2009). "On the Strong Law of Large Numbers for Nonnegative Random Variables." *Theory of Probability and its Applications*, **53**, 346–349.

Robinson P (1982). "On the Convergence of the Horvitz–Thompson Estimator." *Australian Journal of Statistics*, **24**, 234–238.

Robinson P, Särndal C (1983). "Asymptotic Properties of the Generalized Regression Estimator in Probability Sampling." *Sankhyā. Series B*, **45**, 240–248.

Särndal C, Swensson B, Wretman J (1992). *Model Assisted Survey Sampling*. Springer, New York.

Van der Vaart AW (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Walk H (2005). "Strong Laws of Large Numbers by Elementary Tauberian Arguments." *Monatshefte für Mathematik*, **144**, 329–346.

Wu CF (1981). "Asymptotic Theory of Nonlinear Least Squares Estimation." *Annals of Statistics*, **9**, 501–513.

Xie M, Yang Y (2003). "Asymptotics for Generalized Estimating Equations with Large Cluster Sizes." *Annals of Statistics*, **31**, 310–347.

Affiliation:

Tobias Schoch
University of Applied Sciences Northwestern Switzerland
School of Business, Institute ICC
Riggenbachstrasse 16
CH-4600 Olten, Switzerland
E-mail: tobias.schoch@fhnw.ch

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 50 *Submitted:* 2016-12-19
July 2021 *Accepted:* 2020-01-27