Functional central limit theorems in survey sampling

Hélène Boistard[1], Hendrik P. Lopuhaä[2], and Anne Ruiz-Gazen[3]

[1]Toulouse School of Economics
[2]Delft University of Technology
[3]Toulouse School of Economics

September 30, 2015

Abstract

For a joint model-based and design-based inference, we establish functional central limit theorems for the Horvitz-Thompson empirical process and the Hájek empirical process centered by their finite population mean as well as by their super-population mean in a survey sampling framework. The results apply to generic sampling designs and essentially only require conditions on higher order correlations. We apply our main results to a Hadamard differentiable statistical functional and illustrate its limit behavior by means of a computer simulation.

1 Introduction

Functional central limit theorems are well established in statistics. Much of the theory has been developed for empirical processes of independent summands. In combination with the functional delta-method they have become a very powerful tool for investigating the limit behavior for Hadamard or Fréchet differentiable statistical functionals (e.g., see [vdVW96] or [vdV98] for a rigorous treatment with several applications).

In survey sampling, results on functional central limit theorems are far from complete. At the same time there is a need for such results. For instance, in [Dd08] the limit distribution of several statistical functionals is investigated, under the assumption that such a limit theorem exists for a design-based empirical process, whereas in [BD09] the existence of a functional central limit theorem is assumed, to perform model-based inference on several Gini indices. Weak convergence of processes in combination with the delta method are treated in [Bha07], [Dav09], [BM11], but these results are tailor made for specific statistical functionals, and do not apply to the empirical processes that are typically considered in survey sampling.

Recently, functional central limit theorems for empirical processes in survey sampling have appeared in the literature. Most of them are concerned with empirical processes indexed by a class of functions, see [BW07], [SW13], and [BCC14]. However, the results in [BW07] and [SW13] are restricted to sampling schemes that have exchangeable inclusion indicators and constant inclusion probabilities, such as simple random sampling and Bernoulli sampling, whereas the approach in [BCC14] seems difficult to extend to sampling designs other than those that are closely related to Poisson sampling. [Wan12] considers empirical processes indexed by a real valued parameter. Unfortunately, this paper seems to miss a number of assumptions that cannot be avoided and, more importantly, it seems to contain a flaw in the proof. (see Section 7 for a more detailed discussion).

The main purpose of the present paper is to establish functional central limit theorems for the Horvitz-Thompson and the Hájek empirical distribution function that apply to general sampling designs. For design-based inference about finite population parameters, these empirical distribution functions will be centered around their population mean. On the other hand, in many situations involving survey data, one is interested in the corresponding model parameters (e.g., see [KG98] and [BR09]). Recently, Rubin-Bleuer and Schiopu Kratina [RBSK05] defined a mathematical framework for joint model-based and design-based inference through a probability product-space and introduced a general and unified methodology for studying the asymptotic properties of model parameter estimators. To incorporate both types of inferences, we consider the Horvitz-Thompson empirical process and the Hájek empirical process under the super-population model described in [RBSK05], both centered around their finite population mean as well as around their super-population mean. Our main results are functional central limit theorems for both empirical processes indexed by a real valued parameter and apply to generic sampling schemes. These results are established only requiring the usual standard assumptions that one encounters in asymptotic theory in survey sampling. Our approach was inspired by an unpublished manuscript from Philippe Fevrier and Nicolas Ragache, which was the outcome of an internship at INSEE in 2001.

The article is organized as follows. Notations and assumptions are discussed in Section 2. In particular we briefly discuss the joint model-based and design-based inference setting defined in [RBSK05]. In Sections 3 and 4, we list the assumptions and state our main results. Our assumptions essentially concern the inclusion probabilities of the sampling design up to the fourth order and a central limit theorem (CLT) for the Horvitz-Thompson estimator of a population total for i.i.d. bounded random variables. Our results allow random inclusion probabilities and are stated in terms of the design-based expected sample size, but we also formulate more detailed results in case these quantities are deterministic.

As an application of our results, in combination with the functional delta-method, we obtain the limit distribution of the poverty rate in Section 5. This example is further investigated in Section 6 by means of a simulation. Finally, in Section 7 we discuss in detail the differences of our results with the work by [BW07], [SW13], [Wan12], and [BCC14]. All proofs are deferred to Section 8 and some tedious technicalities can be found in [BLRG15].

2 Notations and assumptions

We adopt the super-population setup as described in [RBSK05]. Consider a sequence of finite populations $(\mathcal{U}^N)$, of sizes $N = 1, 2, \ldots$. With each population we associate a set of indices $U_N = \{1, 2, \ldots, N\}$. Furthermore, for each index $i \in U_N$, we have a tuple $(y_i, z_i) \in \mathbb{R} \times \mathbb{R}^q_+$. We denote $\mathbf{y}^N = (y_1, y_2, \ldots, y_N) \in \mathbb{R}^N$ and $\mathbf{z}^N \in \mathbb{R}^{q\times N}_+$ similarly. The vector $\mathbf{y}^N$ contains the values of the variable of interest and $\mathbf{z}^N$ contains information for the sampling design. We assume that the values in each finite population are realizations of random variables $(Y_i, Z_i) \in \mathbb{R} \times \mathbb{R}^q_+$, for $i = 1, 2, \ldots, N$, on a common probability space $(\Omega, \mathfrak{F}, \mathbb{P}_m)$. Similarly, we denote $\mathbf{Y}^N = (Y_1, Y_2, \ldots, Y_N) \in \mathbb{R}^N$ and $\mathbf{Z}^N \in \mathbb{R}^{q\times N}_+$. To incorporate the sampling design, a product space is defined as follows. For all $N = 1, 2, \ldots$, let $\mathcal{S}_N = \{s : s \subset U_N\}$ be the collection of subsets of U_N and let $\mathfrak{A}_N = \sigma(\mathcal{S}_N)$ be the σ-algebra generated by $\mathcal{S}_N$. A sampling design associated to some sampling scheme is a function $p : \mathfrak{A}_N \times \mathbb{R}^{q\times N}_+ \mapsto [0, 1]$, such that

(i) for all $s \in \mathcal{S}_N$, $\mathbf{z}^N \mapsto p(s, \mathbf{z}^N)$ is a Borel-measurable function on $\mathbb{R}^{q\times N}_+$.

(ii) for all $\mathbf{z}^N \in \mathbb{R}^{q\times N}_+$, $A \mapsto p(A, \mathbf{z}^N)$ is a probability measure on $\mathfrak{A}_N$.

Note that for each $\omega \in \Omega$, we can define a probability measure $A \mapsto \mathbb{P}_d(A, \omega) = \sum_{s\in A} p(s, \mathbf{Z}^N(\omega))$ on the design space $(\mathcal{S}_N, \mathfrak{A}_N)$. Corresponding expectations will be denoted by $\mathbb{E}_d(\cdot, \omega)$. Next, we define a product probability space that includes the super-population and the design space, under the premise that sample selection and the model characteristic are independent given the design variables. Let $(\mathcal{S}_N \times \Omega, \mathfrak{A}_N \times \mathfrak{F})$ be the product space with probability measure $\mathbb{P}_{d,m}$ defined on simple rectangles $\{s\} \times E \in \mathfrak{A}_N \times \mathfrak{F}$ by

$$\mathbb{P}_{d,m}(\{s\} \times E) = \int_E p(s, \mathbf{Z}^N(\omega))\, \mathrm{d}\mathbb{P}_m(\omega) = \int_E \mathbb{P}_d(\{s\}, \omega)\, \mathrm{d}\mathbb{P}_m(\omega).$$

When taking expectations or computing probabilities, we will emphasize whether this is with respect either to the measure $\mathbb{P}_{d,m}$ associated with the product space $(\mathcal{S}_N \times \Omega, \mathfrak{A}_N \times \mathfrak{F})$, or the measure $\mathbb{P}_d$ associated with the design space $(\mathcal{S}_N, \mathfrak{A}_N)$, or the measure $\mathbb{P}_m$ associated with the super-population space $(\Omega, \mathfrak{F})$.

If n_s denotes the size of sample s, then this may depend on the specific sampling design including the values of the design variables $Z_1(\omega), \dots, Z_N(\omega)$. Similarly, the inclusion probabilities may depend on the values of the design variables, $\pi_i(\omega) = \mathbb{E}_d(\xi_i, \omega) = \sum_{s \ni i} p\left(s, \mathbf{Z}^N(\omega)\right)$, where ξ_i is the indicator $\mathbb{1}_{\{s \ni i\}}$. Instead of n_s, we will consider $n = \mathbb{E}_d[n_s(\omega)] = \sum_{i=1}^N \mathbb{E}_d(\xi_i, \omega) = \sum_{i=1}^N \pi_i(\omega)$. This means that the inclusion probabilities and the design-based expected sample size may be random variables on $(\Omega, \mathfrak{F}, \mathbb{P}_m)$.

We first consider the Horvitz-Thompson (HT) empirical processes, obtained from the HT empirical c.d.f.:

$$\mathbb{F}_N^{\mathrm{HT}}(t) = \frac{1}{N} \sum_{i=1}^N \frac{\xi_i \mathbb{1}_{\{Y_i \le t\}}}{\pi_i}, \quad t \in \mathbb{R}. \tag{2.1}$$

We will consider HT empirical process $\sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - \mathbb{F}_N)$, obtained by centering around the empirical c.d.f. $\mathbb{F}_N$ of $Y_1, \dots, Y_N$, as well as the HT empirical process $\sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - F)$, obtained by centering around the c.d.f. F of the Y_i's. A functional central limit theorem for both processes will be formulated in Section 3. In addition, we will consider the Hájek empirical c.d.f.:

$$\mathbb{F}_N^{\mathrm{HJ}}(t) = \frac{1}{\widehat{N}} \sum_{i=1}^N \frac{\xi_i \mathbb{1}_{\{Y_i \le t\}}}{\pi_i}, \quad t \in \mathbb{R}, \tag{2.2}$$

where $\widehat{N} = \sum_{i=1}^N \xi_i/\pi_i$ is the HT estimator for the population total N. Functional central limit theorems for $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - \mathbb{F}_N)$ and $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - F)$ will be provided in Section 4. The advantage of our results is that they allow general sampling schemes and that we primarily require bounds on the rate at which higher order correlations tend to zero ω-almost surely, under the design measure $\mathbb{P}_d$.

3 FCLT's for the Horvitz-Thompson empirical processes

A functional central limit theorem for $\sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - \mathbb{F}_N)$ and $\sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - F)$ is obtained by proving weak convergence of all finite dimensional distributions and tightness. In order to establish the latter for general sampling schemes, we impose a number of conditions that involve the sets

$$D_{\nu,N} = \Big\{(i_1, i_2, \dots, i_\nu) \in \{1, 2, \dots, N\}^\nu : i_1, i_2, \dots, i_\nu \text{ all different}\Big\}, \tag{3.1}$$

for the integers $1 \le \nu \le 4$. We assume the following conditions:

(C1) there exist constants K_1, K_2, such that for all $i = 1, 2, \dots, N$,

$$0 < K_1 \le \frac{N\pi_i}{n} \le K_2 < \infty, \quad \omega - \text{a.s.}$$

There exists a constant $K_3 > 0$, such that for all $N = 1, 2, \ldots$:

(C2) $\max_{(i,j)\in D_{2,N}} \left|\mathbb{E}_d(\xi_i - \pi_i)(\xi_j - \pi_j)\right| < K_3 n/N^2$,

(C3) $\max_{(i,j,k)\in D_{3,N}} \left|\mathbb{E}_d(\xi_i - \pi_i)(\xi_j - \pi_j)(\xi_k - \pi_k)\right| < K_3 n^2/N^3$,

(C4) $\max_{(i,j,k,l)\in D_{4,N}} \left|\mathbb{E}_d(\xi_i - \pi_i)(\xi_j - \pi_j)(\xi_k - \pi_k)(\xi_l - \pi_l)\right| < K_3 n^2/N^4$,

ω-almost surely. These conditions on higher order correlations are commonly used in the literature on survey sampling in order to derive asymptotic properties of estimators (e.g., see [BO00], and [CCGL10]). [BO00] proved that they hold for simple random sampling without replacement and stratified simple random sampling without replacement, whereas [BLRG12] proved that they hold also for rejective sampling. Lemma 2 from [BLRG12] allows us to reformulate the above conditions on higher order correlations into conditions on higher order inclusion probabilities.

To establish the convergence of finite dimensional distributions, for sequences of bounded i.i.d. random variables $V_1, V_2, \ldots$ on $(\Omega, \mathfrak{F}, \mathbb{P}_m)$, we will need a CLT for the HT estimator in the design space, conditionally on the V_i's. To this end, let S_N^2 be the (design-based) variance of the HT estimator of the population mean, i.e.,

$$S_N^2 = \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j} V_i V_j. \tag{3.2}$$

We assume that

(HT1) For N sufficiently large $S_N > 0$ and for any sequence of bounded i.i.d. random variables $V_1, V_2, \ldots$,

$$\frac{1}{S_N}\left(\frac{1}{N}\sum_{i=1}^N \frac{\xi_i V_i}{\pi_i} - \frac{1}{N}\sum_{i=1}^N V_i\right) \to N(0,1), \qquad \omega - \text{a.s.},$$

in distribution under $\mathbb{P}_d$.

Note that (HT1) holds for simple random sampling without replacement if $n(N-n)/N$ tends to infinity when N tends to infinity (see [Tho97]), as well as for Poisson sampling under some conditions on the first order inclusion probabilities (e.g., see [Ful09]). For rejective sampling, [Háj64] gives some sufficient conditions for (HT1) to hold.

We also need that nS_N^2 converges for the particular case where the V_i's are random vectors consisting of indicators $\mathbb{1}_{\{Y_j\leq t\}}$.

(HT2) For $k \in \{1, 2, \ldots\}$, $i = 1, 2, \ldots, k$ and $t_1, t_2, \ldots, t_k \in \mathbb{R}$, define $\mathbf{Y}_{ik}^t = \left(\mathbb{1}_{\{Y_i\leq t_1\}}, \ldots, \mathbb{1}_{\{Y_i\leq t_k\}}\right)$. There exists a deterministic matrix $\mathbf{\Sigma}_k^{\text{HT}}$, such that

$$\lim_{N\to\infty} \frac{n}{N^2}\sum_{i=1}^N\sum_{j=1}^N \frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}\mathbf{Y}_{ik}\mathbf{Y}_{jk}^t = \mathbf{\Sigma}_k^{\text{HT}}, \quad \omega - \text{a.s}. \tag{3.3}$$

This kind of assumption is quite standard in the literature on survey sampling and is usually imposed for general random vectors (see, for example [DS92], p.379, [FF91], condition 3 on page 457, or [KR81], condition C4 on page 1014). It suffices to require (3.3) for $\mathbf{Y}_{ik}^t = \left(\mathbb{1}_{\{Y_i\leq t_1\}}, \ldots, \mathbb{1}_{\{Y_i\leq t_k\}}\right)$. Moreover, if (C1)-(C2) hold, then the sequence in (3.3) is bounded, so that by dominated convergence it follows that

$$\mathbf{\Sigma}_k^{\text{HT}} = \lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \mathbb{E}_m\left[n\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}\mathbf{Y}_{ik}\mathbf{Y}_{jk}^t\right]. \tag{3.4}$$

This might help to get a more tractable expression for $\boldsymbol{\Sigma}_k^{\mathrm{HT}}$.

We are now able to formulate our first main result. Let $D(\mathbb{R})$ be the space of càdlàg functions on $\mathbb{R}$ equipped with the Skorohod topology.

Theorem 3.1. *Let $Y_1, \dots, Y_N$ be i.i.d. random variables with c.d.f. F and empirical c.d.f. $\mathbb{F}_N$ and let $\mathbb{F}_N^{\mathrm{HT}}$ be defined in (2.1). Suppose that conditions (C1)-(C4) and (HT1)-(HT2) hold. Then $\sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - \mathbb{F}_N)$ converges weakly in $D(\mathbb{R})$ to a mean zero Gaussian process $\mathbb{G}^{\mathrm{HT}}$ with covariance function*

$$\mathbb{E}_m \mathbb{G}^{\mathrm{HT}}(s)\mathbb{G}^{\mathrm{HT}}(t) = \lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}_m \left[n \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \mathbb{1}_{\{Y_i \le s\}} \mathbb{1}_{\{Y_j \le t\}} \right]$$

for $s, t \in \mathbb{R}$.

Note that Theorem 3.1 allows a random (design-based) expected sample size n and random inclusion probabilities. However, the expression of the covariance function of the limiting Gaussian process is somewhat unsatisfactory. When n and the inclusion probabilities are deterministic, we can obtain a functional CLT with a more precise expression for $\mathbb{E}_m \mathbb{G}^{\mathrm{HT}}(s)\mathbb{G}^{\mathrm{HT}}(t)$ under slightly weaker conditions. This is formulated in the proposition below. Note that with imposing conditions (i)-(ii) in Proposition 3.1 instead of (3.3), convergence of nS_N^2 is not necessarily guaranteed. However, this is established in Lemma 9.1 in [BLRG15] under (C1) and (C2).

Finally, we like to emphasize that if we would have imposed (HT2) for *any* sequence $\mathbf{Y}_1, \mathbf{Y}_2, \dots$ of bounded random vectors, then (HT2) would have implied conditions (i)-(ii) in the deterministic setup of Proposition 3.1.

Proposition 3.1. *Consider the setting of Theorem 3.1, where n and π_i, π_{ij}, for $i, j = 1, 2, \dots, N$, are deterministic. Suppose that (C1)-(C4) and (HT1) hold, but instead of (HT2) assume that there exist constants $\mu_{\pi 1}$, $\mu_{\pi 2} \in \mathbb{R}$ such that*

(i) $\displaystyle \lim_{N\to\infty} \frac{n}{N^2} \sum_{i=1}^N \left(\frac{1}{\pi_i} - 1 \right) = \mu_{\pi 1},$

(ii) $\displaystyle \lim_{N\to\infty} \frac{n}{N^2} \sum\sum_{i \ne j} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} = \mu_{\pi 2}.$

Then $\sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - \mathbb{F}_N)$ converges weakly in $D(\mathbb{R})$ to a mean zero Gaussian process $\mathbb{G}^{\mathrm{HT}}$ with covariance function $\mu_{\pi_1} F(s \wedge t) + \mu_{\pi 2} F(s)F(t)$, for $s, t \in \mathbb{R}$.

When $n/N \to \lambda \in [0, 1]$, then conditions (i)-(ii) hold with $\mu_{\pi 1} = 1 - \lambda$ and $\mu_{\pi 2} = \lambda - 1$ for simple random sampling without replacement. For Poisson sampling, (ii) holds trivially because the trials are independent. For rejective sampling, (i)-(ii) together with $n/N \to \lambda \in [0, 1]$, can be deduced from the associated Poisson sampling design. Indeed, suppose that (i) holds for Poisson sampling with first order inclusion probabilities $p_1, \dots, p_N$, such that $\sum_{i=1}^N p_i = n$. Then, from Theorem 1 in [BLRG12] it follows that if $d = \sum_{i=1}^N p_i(1 - p_i)$ tends to infinity, assumption (i) holds for rejective sampling. Furthermore, if $n/N \to \lambda \in [0, 1]$ and N/d has a finite limit, then also (ii) holds for rejective sampling.

Weak convergence of the process $\sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - F)$, where we center with F instead of $\mathbb{F}_N$, requires a CLT in the super-population space for

$$\sqrt{n} \left(\frac{1}{N} \sum_{i=1}^N \frac{\xi_i V_i}{\pi_i} - \mu_V \right), \qquad \text{where } \mu_V = \mathbb{E}_m(V_i), \tag{3.5}$$

for sequences of bounded i.i.d. random variables $V_1, V_2, \dots$ on $(\Omega, \mathfrak{F}, \mathbb{P}_m)$. Our approach to establish asymptotic normality of (3.5) is then to decompose as follows

$$\begin{aligned} &\sqrt{n}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i V_i}{\pi_i} - \mu_V\right) \\ &= \sqrt{n}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i V_i}{\pi_i} - \frac{1}{N}\sum_{i=1}^{N} V_i\right) + \frac{\sqrt{n}}{\sqrt{N}} \times \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} V_i - \mu_V\right). \end{aligned} \tag{3.6}$$

Since the V_i's are i.i.d. and bounded, for the second term on the right hand side, by the traditional CLT we immediately obtain

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} V_i - \mu_V\right) \to N(0, \sigma_V^2), \tag{3.7}$$

in distribution under $\mathbb{P}_m$, where σ_V^2 denotes the variance of the V_i's, whereas the first term on the right hand side can be handled with (HT1). [BW07] and [SW13] use a decomposition similar to the one in (3.6). Their approach assumes exchangeable ξ_i's and equal inclusion probabilities n/N, which allows the use of results on exchangeable weighted bootstrap to handle the first term on the right hand side of (3.6). Instead, we only require conditions (C2)-(C4) on higher order correlations for the ξ_i's and allow the π_i's to vary within certain bounds as described in (C1). To combine the two separate limits in (3.7) and (HT1), we will need

(HT3) $n/N \to \lambda \in [0, 1]$, ω-a.s.

We will then use Theorem 5.1(iii) from [RBSK05]. The finite dimensional projections of the processes involved turn out to be related to a particular HT estimator. In order to have the corresponding design-based variance converging to a strictly positive constant, we need the following condition.

(HT4) For all $k \in \{1, 2, \dots\}$ and $t_1, t_2, \dots, t_k \in \mathbb{R}$, the matrix $\mathbf{\Sigma}_k^{\mathrm{HT}}$ in (3.3) is positive definite.

We are now able to formulate our second main result.

Theorem 3.2. *Let $Y_1, \dots, Y_N$ be i.i.d. random variables met c.d.f. F and let $\mathbb{F}_N^{\mathrm{HT}}$ be defined in (2.1). Suppose that conditions (C1)-(C4) and (HT1)-(HT4) hold. Then $\sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - F)$ converges weakly in $D(\mathbb{R})$ to a mean zero Gaussian process $\mathbb{G}_F^{\mathrm{HT}}$ with covariance function $\mathbb{E}_{d,m}\mathbb{G}_F^{\mathrm{HT}}(s)\mathbb{G}_F^{\mathrm{HT}}(t)$ given by*

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}_m\left[n\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}\mathbb{1}_{\{Y_i \le s\}}\mathbb{1}_{\{Y_j \le t\}}\right] + \lambda\big\{F(s \wedge t) - F(s)F(t)\big\},$$

for $s, t \in \mathbb{R}$.

Theorem 3.2 allows random n and inclusion probabilities. As before, when the sample size n and inclusion probabilities are deterministic we can obtain a functional CLT under a simpler condition than (HT4) and with a more detailed description of the covariance function of the limiting process.

Proposition 3.2. *Consider the setting of Theorem 3.2, where n and π_i, π_{ij}, for $i, j = 1, 2, \dots, N$, are deterministic. Suppose that (C1)-(C4), (HT1)and (HT3) hold, but instead of (HT2) and (HT4) assume that there exist constants $\mu_{\pi 1}$, $\mu_{\pi 2} \in \mathbb{R}$ such that*

(i) $\displaystyle\lim_{N\to\infty}\frac{n}{N^2}\sum_{i=1}^{N}\left(\frac{1}{\pi_i} - 1\right) = \mu_{\pi 1} > 0,$

(ii) $\lim_{N\to\infty} \frac{n}{N^2} \sum\sum_{i\neq j} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j} = \mu_{\pi 2}$.

Then $\sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - F)$ *converges weakly in* $D(\mathbb{R})$ *to a mean zero Gaussian process* $\mathbb{G}^{\mathrm{HT}}$ *with covariance function* $(\mu_{\pi_1}+\lambda)F(s\wedge t) + (\mu_{\pi 2}-\lambda)F(s)F(t)$, *for* $s,t\in\mathbb{R}$.

Since $1/\pi_i \geq 1$, we will always have $\mu_{\pi 1} \geq 0$ in condition (i) in Proposition 3.2. This means that (i) is not very restrictive. For simple random sampling without replacement, condition (i) requires λ to be strictly smaller than one.

4 FCLT's for the Hájek empirical processes

To determine the behavior of the process $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - \mathbb{F}_N)$, it is useful to relate this process to the process

$$\mathbb{G}_N^{\pi}(t) = \frac{\sqrt{n}}{N}\sum_{i=1}^N \frac{\xi_i}{\pi_i}\Big(\mathbb{1}_{\{Y_i\leq t\}} - F(t)\Big). \tag{4.1}$$

We can then write

$$\sqrt{n}\left\{\mathbb{F}_N^{\mathrm{HJ}}(t) - \mathbb{F}_N(t)\right\} = \mathbb{Y}_N(t) + \left(\frac{N}{\widehat{N}} - 1\right)\mathbb{G}_N^{\pi}(t), \tag{4.2}$$

where

$$\mathbb{Y}_N(t) = \frac{\sqrt{n}}{N}\sum_{i=1}^N \left(\frac{\xi_i}{\pi_i} - 1\right)\left(\mathbb{1}_{\{Y_i\leq t\}} - F(t)\right). \tag{4.3}$$

As intermediate results we will first show that the process $\mathbb{G}_N^{\pi}$ converges weakly to a mean zero Gaussian process and that $\widehat{N}/N \to 1$ in probability. As a consequence, the limiting behavior of $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - \mathbb{F}_N)$ will be the same as that of $\mathbb{Y}_N$, which is an easier process to handle. Instead of (HT2) and (HT4) we now need

(HJ2) For $k\in\{1,2,\ldots\}$, $i=1,2,\ldots,k$ and $t_1,t_2,\ldots,t_k\in\mathbb{R}$, define

$$\widetilde{\mathbf{Y}}_{ik}^t = \left(\mathbb{1}_{\{Y_i\leq t_1\}} - F(t_1),\ldots,\mathbb{1}_{\{Y_i\leq t_k\}} - F(t_k)\right).$$

There exists a deterministic matrix $\mathbf{\Sigma}_k^{\mathrm{HJ}}$, such that

$$\lim_{N\to\infty}\frac{n}{N^2}\sum_{i=1}^N\sum_{j=1}^N \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\widetilde{\mathbf{Y}}_{ik}\widetilde{\mathbf{Y}}_{jk}^t = \mathbf{\Sigma}_k^{\mathrm{HJ}}, \quad \omega-\text{a.s.} \tag{4.4}$$

and

(HJ4) For all $k\in\{1,2,\ldots\}$ and $t_1,t_2,\ldots,t_k\in\mathbb{R}$, the matrix $\mathbf{\Sigma}_k^{\mathrm{HJ}}$ in (4.4) is positive definite.

As in the case of (3.4), if (C1)-(C2) hold, then (HJ2) implies

$$\mathbf{\Sigma}_k^{\mathrm{HJ}} = \lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \mathbb{E}_m\left[n\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\widetilde{\mathbf{Y}}_{ik}\widetilde{\mathbf{Y}}_{jk}^t\right]. \tag{4.5}$$

Theorem 4.1. *Let* $\mathbb{G}_N^{\pi}$ *be defined in* (4.1) *and let* $\widehat{N} = \sum_{i=1}^N \xi_i/\pi_i$. *Suppose* $n\to\infty$, ω*-a.s., and that there exists* $\sigma_\pi^2 \geq 0$, *such that*

$$\frac{n}{N^2}\sum_{i=1}^N\sum_{i=1}^N \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j} \to \sigma_\pi^2, \quad \omega - a.s. \tag{4.6}$$

If in addition,

(i) (HT1) hold, then $\widehat{N}/N \to 1$ in $\mathbb{P}_{d,m}$ probability.

(ii) (C1)-C(4), (HT1), (HT3), (HJ2) and (HJ4) hold, then $\mathbb{G}_N^\pi$ converges weakly in $D(\mathbb{R})$ to a mean zero Gaussian process $\mathbb{G}^\pi$ with covariance function $\mathbb{E}_{d,m}\mathbb{G}^\pi(s)\mathbb{G}^\pi(t)$ given by

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}_m \left[n \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \left(\mathbb{1}_{\{Y_i \le s\}} - F(s) \right) \left(\mathbb{1}_{\{Y_i \le t\}} - F(t) \right) \right] + \lambda \left(F(s \wedge t) - F(s)F(t) \right), \quad s, t \in \mathbb{R}.$$

Note that in view of condition (HT3), the condition $n \to \infty$ is immediate, if $\lambda > 0$. We proceed by establishing weak convergence of $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - \mathbb{F}_N)$.

Theorem 4.2. *Let $Y_1, \ldots, Y_N$ be i.i.d. random variables with c.d.f. F and empirical c.d.f. $\mathbb{F}_N$ and let $\mathbb{F}_N^{\mathrm{HJ}}$ be defined in (2.2). Suppose $n \to \infty$, ω-a.s., and that (C1)-C(4), (HT1), (HT3), and (HJ2) hold, as well as condition (4.6). Then $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - \mathbb{F}_N)$ converges weakly in $D(\mathbb{R})$ to a mean zero Gaussian process $\mathbb{G}^{\mathrm{HJ}}$ with covariance function $\mathbb{E}_{d,m}\mathbb{G}^{\mathrm{HJ}}(s)\mathbb{G}^{\mathrm{HJ}}(t)$ given by*

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}_m \left[n \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \left(\mathbb{1}_{\{Y_i \le s\}} - F(s) \right) \left(\mathbb{1}_{\{Y_i \le t\}} - F(t) \right) \right],$$

for $s, t \in \mathbb{R}$.

Note that we do not need condition (HJ4) in Theorem 4.2. This condition is only needed in Theorem 4.1 to establish the limit distribution of the finite dimensional projections of the process $\mathbb{G}_N^\pi$. For Theorem 4.2 we only need that $\mathbb{G}_N^\pi$ is tight.

As before, below we obtain a functional CLT for $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - \mathbb{F}_N)$ in the case that n and the inclusion probabilities are deterministic. Similar to the remark we made after Theorem 3.1, note that if we would have imposed (HJ2) for any sequence of bounded random vectors, then this would imply conditions (i)-(ii) of Proposition 3.1, which can then be left out in Theorem 4.1.

Proposition 4.1. *Consider the setting of Theorem 4.2, where n and π_i, π_{ij}, for $i, j = 1, 2, \ldots, N$, are deterministic. Suppose $n \to \infty$ and that (C1)-(C4), (HT1) and (HT3) hold, as well as conditions (i)-(ii) from Proposition 3.1. Then $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - \mathbb{F}_N)$ converges weakly in $D(\mathbb{R})$ to a mean zero Gaussian process $\mathbb{G}^{\mathrm{HT}}$ with covariance function $\mu_{\pi 1}\left(F(s \wedge t) - F(s)F(t)\right)$, for $s, t \in \mathbb{R}$.*

Finally, we consider $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - F)$. Again, we relate this process to (4.1) and write

$$\sqrt{n}\left(\mathbb{F}_N^{\mathrm{HJ}}(t) - F(t)\right) = \frac{N}{\widehat{N}} \mathbb{G}_N^\pi(t). \tag{4.7}$$

Since $\widehat{N}/N \to 1$ in probability, this implies that $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - F)$ has the same limiting behavior as $\mathbb{G}_N^\pi$.

Theorem 4.3. *Let $Y_1, \ldots, Y_N$ be i.i.d. random variables with c.d.f. F and let $\mathbb{F}_N^{\mathrm{HJ}}$ be defined in (2.2). Suppose $n \to \infty$, ω-a.s., and that (C1)-C(4), (HT1), (HT3), (HJ2) and (HJ4) hold, as well as condition (4.6). Then $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - F)$ converges weakly in $D(\mathbb{R})$ to a mean zero Gaussian process $\mathbb{G}_F^{\mathrm{HJ}}$ with covariance function $\mathbb{E}_{d,m}\mathbb{G}^\pi(s)\mathbb{G}^\pi(t)$ given by*

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}_m \left[n \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \left(\mathbb{1}_{\{Y_i \le s\}} - F(s) \right) \left(\mathbb{1}_{\{Y_i \le t\}} - F(t) \right) \right] + \lambda \left(F(s \wedge t) - F(s)F(t) \right), \quad s, t \in \mathbb{R}.$$

With Theorem 4.3 we recover Theorem 1 in [Wan12]. Our assumptions are comparable to those in [Wan12], although this paper seems to miss a condition on the convergence of the variance, such as our condition (HJ2).

We conclude this section by establishing a functional CLT for $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - F)$ in the case of deterministic n and inclusion probabilities.

Proposition 4.2. *Consider the setting of Theorem 4.3, where n and π_i, π_{ij}, for $i, j = 1, 2, \ldots, N$, are deterministic. Suppose $n \to \infty$ and that (C1)-(C4), (HT1) and (HT3) hold, as well as conditions (i)-(ii) from Proposition 3.2. Then $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - F)$ converges weakly in $D(\mathbb{R})$ to a mean zero Gaussian process $\mathbb{G}^{\mathrm{HJ}}$ with covariance function $(\mu_{\pi_1} + \lambda)\left(F(s \wedge t) - F(s)F(t)\right)$, for $s, t \in \mathbb{R}$.*

5 Hadamard-differentiable functionals

Theorem 4.3 provides an elegant means to study the limit behavior of estimators that can be described as $\phi(\mathbb{F}_N^{\mathrm{HJ}})$, where ϕ is a Hadamard-differentiable functional. Given such a ϕ, the functional delta-method, e.g., see Theorems 3.9.4 and 3.9.5 in [vdVW96] or Theorem 20.8 in [vdV98], enables one to establish the limit distribution of $\phi(\mathbb{F}_N^{\mathrm{HJ}})$. Similarly, this holds for Theorems 3.1, 3.2, and 4.2, or Propositions 3.1, 3.2, 4.1, and 4.2 in the special case of deterministic n and inclusion probabilities.

We illustrate this by discussing the poverty rate. This indicator has recently been revisited by [GT14] and [OAB15]. This example has also been discussed by [Dd08], but under the assumption of weak convergence of $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - \mathbb{F}_N)$ to some centered continuous Gaussian process. Note that this assumption is now covered by our Theorem 4.2 and Proposition 4.1. Let $\mathbb{D}_\phi \subset D(\mathbb{R})$ consist of $F \in D(\mathbb{R})$ that are non-decreasing. Then for $F \in \mathbb{D}_\phi$, the poverty rate is defined as

$$\phi(F) = F\left(\beta F^{-1}(\alpha)\right) \tag{5.1}$$

for fixed $0 < \alpha, \beta < 1$, where $F^{-1}(\alpha) = \inf\{t : F(t) \geq \alpha\}$. Typical choices are $\alpha = 0.5$ and $\beta = 0.5$ (INSEE) or $\beta = 0.6$ (EUROSTAT). Its Hadamard derivative is given by

$$\phi_F'(h) = -\beta \frac{f(\beta F^{-1}(\alpha))}{f(F^{-1}(\alpha))} h(F^{-1}(\alpha)) + h(\beta F^{-1}(\alpha)). \tag{5.2}$$

See [BLRG15] for details.

We then have the following corollaries for the Horvitz-Thompson estimator $\phi(\mathbb{F}_N^{\mathrm{HT}})$ and the Hájek estimator $\phi(\mathbb{F}_N^{\mathrm{HJ}})$ for the poverty rate $\phi(F)$.

Corollary 5.1. *Let ϕ be defined by (5.1) and suppose that the conditions of Proposition 3.2 hold. Then, if F is differentiable at $F^{-1}(\alpha)$, the random variable $\sqrt{n}(\phi(\mathbb{F}_N^{\mathrm{HT}}) - \phi(F))$ converges in distribution to a mean zero normal random variable with variance*

$$\begin{aligned}\sigma^2_{\mathrm{HT},\alpha,\beta} &= \beta^2 \frac{f(\beta F^{-1}(\alpha))^2}{f(F^{-1}(\alpha))^2}\left(\gamma_{\pi 1}\alpha + \gamma_{\pi 2}\alpha^2\right) \\ &\quad + \gamma_{\pi 1}\phi(F) + \gamma_{\pi 2}\phi(F)^2 - 2\beta \frac{f(\beta F^{-1}(\alpha))}{f(F^{-1}(\alpha))}\phi(F)\big(\gamma_{\pi 1} + \gamma_{\pi 2}\alpha\big),\end{aligned} \tag{5.3}$$

where $\gamma_{\pi 1} = \mu_{\pi 1} + \lambda$ and $\gamma_{\pi 2} = \mu_{\pi 2} - \lambda$. If in addition $n/N \to 0$, then $\sqrt{n}(\phi(\mathbb{F}_N^{\mathrm{HT}}) - \phi(\mathbb{F}_N))$ converges in distribution to a mean zero normal random variable with variance $\sigma^2_{\mathrm{HT},\alpha,\beta}$.

Corollary 5.2. *Let ϕ be defined by (5.1). and suppose that the conditions of Proposition 4.2 hold. Then, if F is differentiable at $F^{-1}(\alpha)$, the random variable $\sqrt{n}(\phi(\mathbb{F}_N^{\mathrm{HJ}}) - \phi(F))$ converges in* *distribution to a mean zero normal random variable with variance*

$$\sigma^2_{\mathrm{HJ},\alpha,\beta} = \beta^2 \frac{f(\beta F^{-1}(\alpha))^2}{f(F^{-1}(\alpha))^2}\gamma_{\pi 1}\alpha(1-\alpha) + \gamma_{\pi 1}\phi(F)\big(1-\phi(F))\big) - 2\beta\frac{f(\beta F^{-1}(\alpha))}{f(F^{-1}(\alpha))}\phi(F)\gamma_{\pi 1}(1-\alpha), \tag{5.4}$$

where $\gamma_{\pi 1} = \mu_{\pi 1} + \lambda$*. If in addition* $n/N \to 0$*, then* $\sqrt{n}(\phi(\mathbb{F}^{\mathrm{HJ}}_N) - \phi(\mathbb{F}_N))$ *converges in distribution to a mean zero normal random variable with variance* $\sigma^2_{\mathrm{HJ},\alpha,\beta}$*.*

6 Simulation study

The objective of this simulation study is to investigate the performance of the Horvitz-Thompson (HT) and the Hájek (HJ) estimators for the poverty rate, as defined in (5.1), at the finite population level and at the super-population level. The asymptotic results from Corollary 5.1 and 5.2 are used to obtain variance estimators whose performance is also assessed in this small study.

Six simulation schemes are implemented with different population sizes and (design-based) expected sample sizes, namely $N = 10\,000$ and 1000 and $n = 500$, 100, and 50. The samples are drawn according to three different sampling designs. The first one is simple random sampling without replacement (SI) with size n. The second design is Bernoulli sampling (BE) with parameter n/N. The third one is Poisson sampling (PO) with first order inclusion probabilities equal to $0.4n/N$ for the first half of the population and equal to $1.6n/N$ for the other half of the population, where the population is randomly ordered. The first order inclusion probabilities are deterministic for the three designs and the sample size n_s is fixed for the SI design, while it is random with respect to the design for the BE and PO designs. Moreover, the SI and BE designs are equal probability designs, while PO is an unequal probability design. The results are obtained by replicating $N_R = 1000$ populations. For each population, $n_R = 1000$ samples are drawn according to the different designs. The variable of interest Y is generated for each population according to an exponential distribution with rate parameter equal to one. For this distribution and given α and β, the poverty rate has an explicit expression $\phi(F) = 1 - \exp(\beta \ln(1-\alpha))$. In what follows, $\alpha = 0.5$ and $\beta = 0.6$ and $\phi(F) \simeq 0.34$. These are the same values for α and β as considered in [Dd08].

The Horvitz-Thompson estimator and Hájek estimator for $\phi(F)$ or $\phi(\mathbb{F}_N)$ are denoted by $\widehat{\phi}_{\mathrm{HT}}$ and $\widehat{\phi}_{\mathrm{HJ}}$, respectively. They are obtained by plugging in the empirical c.d.f.'s $\mathbb{F}^{\mathrm{HT}}_N$ and $\mathbb{F}^{\mathrm{HJ}}_N$, respectively, for F in expression (5.1). The empirical quantiles are calculated by using the function `wtd.quantile` from the R package `Hmisc` for the Hájek estimator and by adapting the function for the Horvitz-Thompson estimator. For the SI sampling design, the two estimators are the same. The performance of the estimators for the parameters $\phi(F)$ and $\phi(\mathbb{F}_N)$ is evaluated using some Monte-Carlo relative bias (RB). This is reported in Table 1. When estimating the super-population parameter $\phi(F)$, if $\widehat{\phi}_{ij}$ denotes the estimate (either $\widehat{\phi}_{\mathrm{HT}}$ or $\widehat{\phi}_{\mathrm{HJ}}$) for the ith generated population and the jth drawn sample, the Monte Carlo relative bias of $\widehat{\phi}$ in percentages has the following expression

$$\mathrm{RB}_F(\widehat{\phi}) = \frac{100}{N_R\, n_R}\sum_{i=1}^{N_R}\sum_{j=1}^{n_R}\frac{\widehat{\phi}_{ij} - \phi(F)}{\phi(F)}.$$

When estimating the finite population parameter $\phi(\mathbb{F}_N)$, the parameter depends on the generated population N_i, for each $i = 1, \ldots, N_R$, and will be denoted by $\phi(\mathbb{F}_{N_i})$. The Monte Carlo relative bias of $\widehat{\phi}$ is then computed by replacing F by $\mathbb{F}_{N_i}$ in the above expression. Concerning the relative biases reported in Table 1, the values are small and never exceed 3%. As expected, these values increase when n decreases. When the centering is relative to $\phi(\mathbb{F}_N)$, the relative bias is in general

Table 1: RB (in %) of the HT and the HJ estimators for the finite population $\phi(\mathbb{F}_N)$ and the super-population $\phi(F)$ poverty rate parameter

			$N = 10\,000$			$N = 1000$		
			$n = 500$	$n = 100$	$n = 50$	$n = 500$	$n = 100$	$n = 50$
SI	HT-HJ	$\phi(\mathbb{F}_N)$	−0.17	−0.89	−1.82	−0.05	−0.84	−1.62
		$\phi(F)$	−0.20	−0.91	−1.86	−0.18	−0.72	−1.85
BE	HT	$\phi(\mathbb{F}_N)$	−0.12	−0.66	−1.29	0.01	−0.65	−1.12
		$\phi(F)$	−0.15	−0.68	−1.34	−0.12	−0.54	−1.36
	HJ	$\phi(\mathbb{F}_N)$	−0.17	−0.92	−1.87	−0.04	−0.88	−1.68
		$\phi(F)$	−0.20	−0.93	−1.92	−0.17	−0.76	−1.91
PO	HT	$\phi(\mathbb{F}_N)$	−0.05	−1.05	−2.06	−0.06	−0.30	−0.37
		$\phi(F)$	−0.08	−1.07	−2.11	−0.19	−0.19	−0.63
	HJ	$\phi(\mathbb{F}_N)$	−0.20	−1.27	−2.95	−0.04	−1.08	−1.99
		$\phi(F)$	−0.23	−1.28	−3.00	−0.17	−0.97	−2.23

Table 2: RB (in %) for the variance estimator of the HT and the HJ estimators for the poverty rate parameter

		$N = 10\,000$			$N = 1000$		
		$n = 500$	$n = 100$	$n = 50$	$n = 500$	$n = 100$	$n = 50$
SI	HT-HJ	−2.21	−3.08	−2.97	−2.25	−3.26	−3.00
BE	HT	−4.15	−5.11	−4.21	−3.31	−5.11	−4.19
	HJ	−2.22	−3.06	−3.03	−2.26	−3.24	−3.03
PO	HT	−4.43	−4.96	−3.45	−3.74	−5.72	−4.59
	HJ	−2.36	−3.43	−3.36	−2.44	−3.75	−4.13

somewhat smaller than when centering with $\phi(F)$. This behavior is most prominent when $N = 1000$ and $n = 500$, which suggests that the estimates are typically closer to the population poverty rate $\phi(\mathbb{F}_N)$ than to the model parameter $\phi(F)$. The Hájek estimator has a larger relative bias than the Horvitz-Thompson estimator in all situations but in particular for the Poisson sampling design when the size of the population is 1000. Note that all values in Table 1 are negative, which illustrates the fact that the estimators typically underestimate the population and model poverty rates.

In Table 2, the estimators of the variance of $\widehat{\phi}_{\mathrm{HT}}$ and $\widehat{\phi}_{\mathrm{HJ}}$ are obtained by plugging in the empirical c.d.f.'s $\mathbb{F}_N^{\mathrm{HT}}$ and $\mathbb{F}_N^{\mathrm{HJ}}$, respectively, for F in the expressions (5.3) and (5.4). To estimate f in the variance of $\widehat{\phi}_{\mathrm{HJ}}$, we follow [BS03], who propose a Hájek type kernel estimator with a Gaussian kernel function. For the variance of $\widehat{\phi}_{\mathrm{HT}}$, we use a corresponding Horvitz-Thompson estimator by replacing $\widehat{N}$ by N. Based on [Sil86], pages 45-47, we choose $b = 0.79Rn_s^{-1/5}$, where R denotes the interquartile range. This differs from [BS03], who propose a similar bandwidth of the order $N^{-1/5}$. However, this severely underestimates the optimal bandwidth, leading to large variances of the kernel estimator. Usual bias variance trade-off computations show that the optimal bandwidth is of the order $n_s^{-1/5}$.

For the SI sampling design, (5.3) and (5.4) are identical and can be calculated in an explicit way using the fact that $\mu_{\pi 1} + \lambda = 1$ and $\mu_{\pi 2} - \lambda = -1$. For the BE design, $\mu_{\pi 1} + \lambda = 1$, whereas for Poisson sampling, the value $(n/N^2)\sum_{i=1}^N 1/\pi_i$ is taken for $\mu_{\pi 1} + \lambda$. For these designs, $\mu_{\pi 2} - \lambda = -\lambda$, where we take n/N as the value of λ.

In order to compute the relative bias of the variance estimates, the asymptotic variance is taken as reference. This asymptotic variance $\mathrm{AV}(\widehat{\phi})$ of the estimator $\widehat{\phi}$ (either $\widehat{\phi}_{\mathrm{HT}}$ or $\widehat{\phi}_{\mathrm{HJ}}$) is computed from (5.3) and (5.4). The expressions $f(\beta F^{-1}(\alpha))$ and $f(F^{-1}(\alpha))$ are explicit in the case of an

Table 3: Coverage probabilities (in %) for 95% confidence intervals of the HT and the HJ estimators for the finite population $\phi(\mathbb{F}_N)$ and the super-population $\phi(F)$ poverty rate parameter

			$N = 10\,000$			$N = 1000$		
			$n = 500$	$n = 100$	$n = 50$	$n = 500$	$n = 100$	$n = 50$
SI	HT-HJ	$\phi(\mathbb{F}_N)$	95.2	94.4	93.5	98.8	95.1	94.6
		$\phi(F)$	94.6	93.2	92.2	94.7	93.2	92.0
BE	HT	$\phi(\mathbb{F}_N)$	94.9	94.3	94.6	98.4	94.8	94.6
		$\phi(F)$	94.4	93.7	94.9	94.6	93.6	94.7
	HJ	$\phi(\mathbb{F}_N)$	95.1	94.3	93.9	98.7	94.9	94.2
		$\phi(F)$	94.7	94.2	93.9	94.7	94.2	93.9
PO	HT	$\phi(\mathbb{F}_N)$	94.5	94.2	94.3	96.8	94.0	93.6
		$\phi(F)$	94.5	94.0	94.3	94.6	93.6	93.5
	HJ	$\phi(\mathbb{F}_N)$	94.8	93.9	93.6	97.2	94.2	93.3
		$\phi(F)$	94.6	93.9	93.6	94.6	93.9	93.2

exponential distribution. Furthermore, for $\mu_{\pi 1} + \lambda$ and $\mu_{\pi 2} - \lambda$ we use the same expressions as mentioned above. The Monte Carlo relative bias of the variance estimator $\widehat{\text{AV}}(\widehat{\phi})$ in percentages, is defined by

$$\text{RB}(\widehat{\text{AV}}(\widehat{\phi})) = \frac{100}{N_R\, n_R} \sum_{i=1}^{N_R} \sum_{j=1}^{n_R} \frac{\widehat{\text{AV}}(\widehat{\phi}_{ij}) - \text{AV}(\widehat{\phi})}{\text{AV}(\widehat{\phi})},$$

where $\widehat{\text{AV}}(\widehat{\phi}_{ij})$ denotes the variance estimate for the ith generated population and the jth drawn sample.

Table 3 gives the Monte-Carlo coverage probabilities for a nominal coverage probability of 95% for the two parameters $\phi(\mathbb{F}_N)$ and $\phi(F)$, the Horvitz-Thompson and the Hájek estimators and the different simulation schemes. In general the coverage probabilities are somewhat smaller than 95%, which is due to the underestimation of the asymptotic variance, as can be seen from Table 2. The case $N = 1000$ and $n = 500$ for $\widehat{\phi}_{\text{HJ}}$ forms an exception, which is probably due to the fact that in this case $\lambda = n/N$ is far from zero, so that the limit distribution of $\sqrt{n}(\phi(\mathbb{F}_N^{\text{HT}}) - \phi(\mathbb{F}_N))$ and $\sqrt{n}(\phi(\mathbb{F}_N^{\text{HJ}}) - \phi(\mathbb{F}_N))$ has a larger variance than the ones reported in Corollaries 5.1 and 5.2. When looking at Table 2, the relative biases are smaller than 5% when n is 500. The biases are larger for the Horvitz-Thompson estimator than for the Hájek estimator. Again all relative biases are negative, which illustrates the fact that the asymptotic variance is typically underestimated.

7 Discussion

[Wan12] formulates a functional central limit theorem (see his Theorem 1) for the Hájek empirical c.d.f. from (2.2) centered around F. It is also claimed that a similar result holds for the Horvitz-Thompson process in (2.1), but details are not provided. The paper seems to miss a number of assumptions that cannot be avoided. For instance, the proof of his Theorem 1 requires convergence in probability of the covariance matrix of the vector $\sqrt{n^*}(F_{n\pi}(t) - F_N(t), F_{n\pi}(s) - F_N(s))$. This assumption is comparable with our condition (HJ2), but is missing in [Wan12]. More severely, the argument establishing Billingsley's tightness condition seems to contain a serious mistake, which cannot be repaired easily (see the inequality on line 6 page 678 in [Wan12]; the inequality can be shown not to hold for instance for sampling designs with independent inclusion indicators). As a consequence, assumption 5 in [Wan12] differs somewhat from our conditions (C2)-(C4). The remaining assumptions in [Wan12] are comparable to the conditions needed for our Theorem 4.3.

Note that, in addition to the latter theorem, we also establish Theorems 3.1, 3.2, and 4.2 for other empirical processes of interest.

[BW07] and [SW13] obtain weak convergence of the empirical process (in our notation)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{\xi_i}{\pi_i}f(Y_i)-\mathbb{E}_m f(Y_i)\right\},\qquad f\in\mathcal{F}. \tag{7.1}$$

Weak convergence is established under finite population two-phase stratified sampling. This process is comparable to our Horvitz-Thompson empirical process in Theorem 3.2. Although their functional CLT allows general function classes, it only covers sampling designs with equal inclusion probabilities within strata that assume exchangeability of the inclusion indicators $\xi_1,\dots,\xi_N$, such as simple random sampling and Bernoulli sampling. Their approach views two-phase stratified sampling as a form of bootstrap and uses results on exchangeable weighted bootstrap for empirical processes from [PW93], as incorporated in [vdVW96]. This approach, in particular the application of Theorem 3.6.13 in [vdVW96], seems difficult to extend to more complex sampling designs that go beyond exchangeable inclusion indicators. Although our results only correspond to the class of indicators $f_t(y)=\mathbb{1}_{(-\infty,t]}(y)$, for $t\in\mathbb{R}$, the advantage of our results is that they are applicable to general sampling designs. Moreover, our results also include empirical processes centered with the population mean.

[BCC14] establish a functional CLT, for the Poisson-like empirical process

$$\widetilde{\mathbb{G}}^{\mathbf{p}}_{T_N}(f)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\xi_i-p_i)\left\{\frac{f(Y_i)}{p_i}-\theta_{N,\mathbf{p}}(f)\right\},\qquad f\in\mathcal{F}, \tag{7.2}$$

where $\mathbf{p}=(p_1,\dots,p_N)$ is the vector of inclusion probabilities corresponding to a Poisson sampling design and

$$\theta_{N,\mathbf{p}}(f)=\frac{1}{d_N}\sum_{i=1}^{N}(1-p_i)f(Y_i),\quad d_N=\sum_{i=1}^{N}p_i(1-p_i).$$

However, the functional CLT is obtained conditionally on the $Y_1,Y_2\dots$. In this case, the terms in the summation in (7.2) are independent, which allows the use of Theorem 2.11.1 from [vdVW96]. From their result a functional CLT under rejective sampling can then be established for the design-based Horvitz-Thompson process

$$\mathbb{G}_{N,\pi}(f)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{\xi_i}{\pi_i}f(Y_i)-f(Y_i)\right\},\qquad f\in\mathcal{F}, \tag{7.3}$$

ω-almost surely. This is due to the close connection between Poisson sampling and rejective sampling. For this reason, the approach used in [BCC14] seems difficult to extend to other sampling designs. For the class of indicators $f_t(y)=\mathbb{1}_{(-\infty,t]}(y)$, for $t\in\mathbb{R}$, the process in (7.3) is similar to the one in our Theorem 3.1, but this theorem allows general sampling designs. Moreover, our results also include empirical processes centered with the superpopulation mean.

8 Proofs

We will use Theorem 13.5 from [Bil99], which requires convergence of finite dimensional distributions and a tightness condition (see (13.14) in [Bil99]. We will first establish the tightness condition, as stated in the following lemma.

Lemma 8.1. *Let $Y_1, \dots, Y_N$ be i.i.d. random variables with c.d.f. F and empirical c.d.f. $\mathbb{F}_N$ and let $\mathbb{F}_N^{\mathrm{HT}}$ be defined according to (2.1). Let $\mathbb{X}_N = \sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - \mathbb{F}_N)$ and suppose that (C1)-(C4) hold. Then there exists a constant $K > 0$ independent of N, such that for any t_1, t_2 and $-\infty < t_1 \le t \le t_2 < \infty$,*

$$\mathbb{E}_{d,m}\left[(\mathbb{X}_N(t) - \mathbb{X}_N(t_1))^2 (\mathbb{X}_N(t_2) - \mathbb{X}_N(t))^2\right] \le K\Big(F(t_2) - F(t_1)\Big)^2.$$

Proof. First note that

$$\mathbb{X}_N(t) = \frac{\sqrt{n}}{N}\sum_{i=1}^N \left(\frac{\xi_i}{\pi_i} - 1\right)\mathbb{1}_{\{Y_i \le t\}}.$$

For the sake of brevity, for $-\infty < t_1 \le t \le t_2 < \infty$, and $i = 1, 2, \dots, N$, we define $p_1 = F(t) - F(t_1)$, $p_2 = F(t_2) - F(t)$, $A_i = \mathbb{1}_{\{t_1 < Y_i \le t\}}$, and $B_i = \mathbb{1}_{\{t < Y_i \le t_2\}}$. Furthermore, let $\alpha_i = (\xi_i - \pi_i)A_i/\pi_i$ and $\beta_i = (\xi_i - \pi_i)B_i/\pi_i$. Then, according to the fact that $p_1 p_2 \le (F(t_2) - F(t_1))^2$, due to the monotonicity of F, it suffices to show

$$\frac{1}{N^4}\mathbb{E}_{d,m}\left[n^2\left(\sum_{i=1}^N \alpha_i\right)^2\left(\sum_{j=1}^N \beta_j\right)^2\right] \le K p_1 p_2. \tag{8.1}$$

The expectation on the left hand side can be decomposed as follows

$$\begin{aligned}&\sum_{i=1}^N\sum_{k=1}^N \mathbb{E}_{d,m}\left[n^2\alpha_i^2\beta_k^2\right] + \sum_{i=1}^N\sum_{j\neq i}\sum_{k=1}^N \mathbb{E}_{d,m}\left[n^2\alpha_i\alpha_j\beta_k^2\right]\\ &+\sum_{k=1}^N\sum_{l\neq k}\sum_{i=1}^N \mathbb{E}_{d,m}\left[n^2\alpha_i^2\beta_k\beta_l\right] + \sum_{i=1}^N\sum_{j\neq i}\sum_{k=1}^N\sum_{l\neq k} \mathbb{E}_{d,m}\left[n^2\alpha_i\alpha_j\beta_k\beta_l\right].\end{aligned} \tag{8.2}$$

Note that by symmetry, sums two and three on the right hand side can be handled similarly, so that essentially we have to deal with three summations. We consider them one by one.

First note that, since $\mathbb{1}_{\{t_1 < Y_i \le t\}}\mathbb{1}_{\{t < Y_i \le t_2\}} = 0$, we will only have non-zero expectations when $\{i, j\}$ and $\{k, l\}$ are disjoint. With (C1), we find

$$\begin{aligned}\frac{1}{N^4}\sum_{i=1}^N\sum_{k=1}^N \mathbb{E}_{d,m}\left[n^2\alpha_i^2\beta_k^2\right] &= \frac{1}{N^4}\sum\sum_{(i,k)\in D_{2,N}} \mathbb{E}_{d,m}\left[n^2\alpha_i^2\beta_k^2\right]\\ &= \frac{1}{N^4}\sum\sum_{(i,k)\in D_{2,N}} \mathbb{E}_m\left[n^2\frac{A_iB_k}{\pi_i^2\pi_k^2}\mathbb{E}_d(\xi_i - \pi_i)^2(\xi_k - \pi_k)^2\right]\\ &\le \frac{1}{K_1^4}\sum\sum_{(i,k)\in D_{2,N}} \mathbb{E}_m\left[\frac{A_iB_k}{n^2}\mathbb{E}_d(\xi_i - \pi_i)^2(\xi_k - \pi_k)^2\right]\end{aligned} \tag{8.3}$$

Straightforward computation shows that $\mathbb{E}_d(\xi_i - \pi_i)^2(\xi_k - \pi_k)^2$ equals

$$(\pi_{ik} - \pi_i\pi_k)(1 - 2\pi_i)(1 - 2\pi_k) + \pi_i\pi_k(1 - \pi_i)(1 - \pi_k).$$

Hence, with (C1)-(C2) we find that

$$\mathbb{E}_d(\xi_i - \pi_i)^2(\xi_k - \pi_k)^2 \le |\mathbb{E}_d(\xi_i - \pi_i)(\xi_k - \pi_k)| + K_2^2\frac{n^2}{N^2} = O\left(\frac{n^2}{N^2}\right),$$

ω-almost surely. It follows that

$$\frac{1}{N^4}\sum_{i=1}^{N}\sum_{k=1}^{N}\mathbb{E}_{d,m}\left[n^2\alpha_i^2\beta_k^2\right] \le O\left(\frac{1}{N^2}\right)\sum\sum_{(i,k)\in D_{2,N}}\mathbb{E}_m\left[A_iB_k\right].$$

Since $D_{2,N}$ has $N(N-1)$ elements and $\mathbb{E}_m[A_iB_j] = p_1p_2$ for $(i,j)\in D_{2,N}$, it follows that

$$\frac{1}{N^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}_{d,m}\left[n^2\alpha_i^2\beta_j^2\right] \le Kp_1p_2. \tag{8.4}$$

Consider the second (and third) summation on the right hand side of (8.2). Similarly to (8.3), we can then write

$$\begin{aligned}
&\frac{1}{N^4}\left|\sum_{i=1}^{N}\sum_{j\neq i}\sum_{k=1}^{N}\mathbb{E}_{d,m}\left[n^2\alpha_i\alpha_j\beta_k^2\right]\right| = \frac{1}{N^4}\left|\sum\sum\sum_{(i,j,k)\in D_{3,N}}\mathbb{E}_{d,m}\left[n^2\alpha_i\alpha_j\beta_k^2\right]\right| \\
&\le \frac{1}{N^4}\sum\sum\sum_{(i,j,k)\in D_{3,N}}\left|\mathbb{E}_{d,m}\left[n^2\frac{A_iA_jB_k}{\pi_i\pi_j\pi_k^2}(\xi_i-\pi_i)(\xi_j-\pi_j)(\xi_k-\pi_k)^2\right]\right| \\
&\le \frac{1}{N^4}\sum\sum\sum_{(i,j,k)\in D_{3,N}}\mathbb{E}_m\left[n^2\frac{A_iA_jB_k}{\pi_i\pi_j\pi_k^2}\left|\mathbb{E}_d(\xi_i-\pi_i)(\xi_j-\pi_j)(\xi_k-\pi_k)^2\right|\right] \\
&\le \frac{1}{K_1^4}\sum\sum\sum_{(i,j,k)\in D_{3,N}}\mathbb{E}_m\left[\frac{A_iA_jB_k}{n^2}\left|\mathbb{E}_d(\xi_i-\pi_i)(\xi_j-\pi_j)(\xi_k-\pi_k)^2\right|\right].
\end{aligned}$$

We find that $\mathbb{E}_d(\xi_i-\pi_i)(\xi_j-\pi_j)(\xi_k-\pi_k)^2$ equals

$$(1-2\pi_k)\mathbb{E}_d(\xi_i-\pi_i)(\xi_j-\pi_j)(\xi_k-\pi_k) + \pi_k(1-\pi_k)\mathbb{E}_d(\xi_i-\pi_i)(\xi_j-\pi_j)$$

With (C1)-(C3), this means $|\mathbb{E}_d(\xi_i-\pi_i)(\xi_j-\pi_j)(\xi_k-\pi_k)^2| = O(n^2/N^3)$, ω-almost surely. It follows that

$$\frac{1}{N^4}\left|\sum_{i=1}^{N}\sum_{j\neq i}\sum_{k=1}^{N}\mathbb{E}_{d,m}\left[n^2\alpha_i\alpha_j\beta_k^2\right]\right| = O\left(\frac{1}{N^3}\right)\sum\sum\sum_{(i,j,k)\in D_{3,N}}\mathbb{E}_m\left[A_iA_jB_k\right].$$

Since $D_{3,N}$ has $N(N-1)(N-2)$ elements and $\mathbb{E}_{d,m}[A_iA_jB_k] = p_1^2p_2$, for $(i,j,k)\in D_{3,N}$, we find

$$\frac{1}{N^4}\left|\sum_{i=1}^{N}\sum_{j\neq i}\sum_{k=1}^{N}\mathbb{E}_{d,m}\left[n^2\alpha_i\alpha_j\beta_k^2\right]\right| \le Kp_1p_2. \tag{8.5}$$

The computations for the third summation in (8.2) are completely similar. Finally, consider the last summation in (8.2). As before, this summation can be bounded by

$$\frac{1}{K_1^4}\sum_{(i,j,k,l)\in D_{4,N}}\mathbb{E}_m\left[\frac{A_iA_jB_kB_l}{n^2}\left|\mathbb{E}_d(\xi_i-\pi_i)(\xi_j-\pi_j)(\xi_k-\pi_k)(\xi_l-\pi_l)\right|\right].$$

Since $D_{4,N}$ has $N(N-1)(N-2)(N-3)$ elements and $\mathbb{E}_m[A_iA_jB_kB_l] = p_1^2p_2^2$, for $(i,j,k,l)\in D_{4,N}$, with (C4) we conclude that

$$\frac{1}{N^4}\left|\sum_{i=1}^{N}\sum_{j\neq i}\sum_{k=1}^{N}\sum_{l\neq k}\mathbb{E}_{d,m}\left[n^2\alpha_i\alpha_j\beta_k\beta_l\right]\right| \le Kp_1p_2. \tag{8.6}$$

Together with (8.4), (8.5) and decomposition (8.2), this proves (8.1). □

Lemma 8.2. *Let $\mathbb{X}_N = \sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - \mathbb{F}_N)$ and suppose that (C1)-(C2),(HT1)-(HT2) hold. For any $k \in \{1,2,\dots\}$, and $t_1,\dots,t_k \in \mathbb{R}$, $\big(\mathbb{X}_N(t_1),\dots,\mathbb{X}_N(t_k)\big)$ converges in distribution under $\mathbb{P}_{d,m}$ to a k-variate mean zero normal random vector with covariance matrix $\mathbf{\Sigma}_k^{\mathrm{HT}}$ given in (3.4).*

Proof. We will use the Cramér-Wold device. Note that any linear combination

$$a_1\sqrt{n}\left\{\mathbb{F}_N^{\mathrm{HT}}(t_1) - \mathbb{F}_N(t_1)\right\} + \cdots + a_k\sqrt{n}\left\{\mathbb{F}_N^{\mathrm{HT}}(t_k) - \mathbb{F}_N(t_k)\right\} \tag{8.7}$$

can be written as

$$\sqrt{n}\left\{\frac{1}{N}\sum_{i=1}^N \frac{\xi_i}{\pi_i}V_{ik} - \frac{1}{N}\sum_{i=1}^N V_{ik}\right\}, \tag{8.8}$$

where

$$V_{ik} = a_1\mathbb{1}_{\{Y_i\le t_1\}} + \cdots + a_k\mathbb{1}_{\{Y_i\le t_k\}} = \mathbf{a}_k^t\mathbf{Y}_{ik} \tag{8.9}$$

with $\mathbf{Y}_{ik}^t = (\mathbb{1}_{\{Y_i\le t_1\}},\dots,\mathbb{1}_{\{Y_i\le t_k\}})$ and $\mathbf{a}_k^t = (a_1,\dots,a_k)$. For the corresponding design-based variance, we have

$$\begin{aligned} nS_N^2 &= \frac{n}{N^2}\sum_{i=1}^N\sum_{j=1}^N \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}V_{ik}V_{jk} \\ &= \mathbf{a}_k^t\left(\frac{n}{N^2}\sum_{i=1}^N\sum_{j=1}^N \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\mathbf{Y}_{ik}\mathbf{Y}_{jk}^t\right)\mathbf{a}_k \to \mathbf{a}_k^t\mathbf{\Sigma}_k^{\mathrm{HT}}\mathbf{a}_k, \end{aligned} \tag{8.10}$$

ω-almost surely, according to (HT2), where $\mathbf{\Sigma}_k^{\mathrm{HT}}$ can obtained from (3.4). Together with (HT1), it follows that (8.7) converges in distribution to a mean zero normal random variable with variance $\mathbf{a}_k^t\mathbf{\Sigma}_k^{\mathrm{HT}}\mathbf{a}_k$. We conclude that (8.7) converges in distribution to $a_1N_1+\cdots+a_kN_k$, where $(N_1,\dots,N_k)$ has a k-variate mean zero normal distribution with covariance matrix $\mathbf{\Sigma}_k^{\mathrm{HT}}$. According to the Cramér-Wold device this proves the lemma. □

Proof of Theorem 3.1 We first consider $\mathbb{X}_N = \sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - \mathbb{F}_N)$ for the case that the Y_i's follow a uniform distribution on $[0,1]$. We apply Theorem 13.5 from [Bil99]. Lemma 8.2 provides the limiting distribution of the finite dimensional projections $(\mathbb{X}_N(t_1),\dots,\mathbb{X}_N(t_k))$, which is the same as that of the vector $(\mathbb{G}^{\mathrm{HT}}(t_1),\dots,\mathbb{G}^{\mathrm{HT}}(t_k))$, where $\mathbb{G}^{\mathrm{HT}}$ is a mean zero Gaussian process with covariance function

$$\mathbb{E}_m\mathbb{G}^{\mathrm{HT}}(s)\mathbb{G}^{\mathrm{HT}}(t) = \lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \mathbb{E}_m\left[n\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\mathbb{1}_{\{Y_i\le s\}}\mathbb{1}_{\{Y_j\le t\}}\right],$$

for all $s,t\in\mathbb{R}$. Tightness condition (13.14) in [Bil99] is provided by Lemma 8.1. Since $\mathbb{G}^{\mathrm{HT}}$ is continuous at 1, the theorem now follows from Theorem 13.5 in [Bil99] for the case that the Y_i's are uniformly distributed on $[0,1]$.

To extend this to a functional CLT with i.i.d. random variables $Y_1, Y_2, \dots$ with a general c.d.f. F, we can follow the argument in the proof of Theorem 14.3 from [Bil99]. First define the generalized inverse of F:

$$\varphi(s) = \inf\{t : s \le F(t)\},$$

that satisfies $s \le F(t)$ if and only if $\varphi(s) \le t$. This means that if $U_1, U_2, \dots$ are i.i.d. uniformly distributed on $[0,1]$, $\varphi(U_i)$ has the same distribution as Y_i, so that $\mathbb{1}_{\{Y_i\le t\}} \overset{d}{=} \mathbb{1}_{\{\varphi(U_i)\le t\}} = \mathbb{1}_{\{U_i\le F(t)\}}$. It follows that

$$\mathbb{X}_N(t) = \sqrt{n}\left\{\frac{1}{N}\sum_{i=1}^N\frac{\xi_i\mathbb{1}_{\{Y_i\le t\}}}{\pi_i} - \frac{1}{N}\sum_{i=1}^N\mathbb{1}_{\{Y_i\le t\}}\right\} \overset{d}{=} Z_N(F(t)), \quad t\in\mathbb{R},$$

where

$$Z_N(t) = \frac{\sqrt{n}}{N} \sum_{i=1}^{N} \left(\frac{\xi_i}{\pi_i} - 1 \right) \mathbb{1}_{\{U_i \leq t\}}, \quad t \in [0,1], \tag{8.11}$$

Hence, the general HT empirical process $\mathbb{X}_N$ is the image of the HT uniform empirical process Z_N under the mapping $\psi : D[0,1] \mapsto D(\mathbb{R})$ given by $[\psi x](t) = x(F(t))$. Note that, if $x_N \to x$ in $D[0,1]$ in the Skorohod topology and x has continuous sample paths, then the convergence is uniform. But then also ψx_N converges to ψx uniformly in $D(\mathbb{R})$. This implies that ψx_N converges to ψx in the Skorohod topology. We have established that $Z_N \Rightarrow Z$ weakly in $D[0,1]$ in the Skorohod topology, where Z has continuous sample paths. Therefore, according to the continuous mapping theorem, e.g., Theorem 2.7 in [Bil99], it follows that $\psi(Z_N) \Rightarrow \psi(Z)$ weakly. This proves the theorem for Y_i's with a general c.d.f. F. □

Proof of Proposition 3.1 The proof is similar to that of Theorem 3.1. First consider the case of uniform Y_i's with $F(t) = t$. We only have to verify the weak convergence of the finite dimensional projections of the process $\mathbb{X}_N = \sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - \mathbb{F}_N)$. Consider (8.7) represented as in (8.8). From (HT1) and Lemma 9.1(ii) in [BLRG15] we conclude that (8.7) converges in distribution to a mean zero normal random variable with variance

$$\begin{aligned} \sigma_{\mathrm{HT}}^2 &= \mu_{\pi 1} \mathbb{E}_m \left[V_{1k}^2 \right] + \mu_{\pi 2} \left(\mathbb{E}_m \left[V_{1k} \right] \right)^2 \\ &= \mu_{\pi 1} \mathbf{a}_k^t \mathbb{E}_m \left[\mathbf{Y}_{1k} \mathbf{Y}_{1k}^t \right] \mathbf{a}_k + \mu_{\pi 2} \mathbf{a}_k^t \left(\mathbb{E}_m \mathbf{Y}_{1k} \right) \left(\mathbb{E}_m \mathbf{Y}_{1k} \right)^t \mathbf{a}_k = \mathbf{a}_k^t \mathbf{\Sigma}_k \mathbf{a}_k, \end{aligned}$$

where $\mathbf{\Sigma}_k$ is the $k \times k$-matrix with (q,r)-element equal to $\mu_{\pi_1}(t_q \wedge t_r) + \mu_{\pi 2} t_q t_r$. We conclude that (8.7) converges in distribution to $a_1 N_1 + \cdots + a_k N_k$, where $(N_1, \ldots, N_k)$ has a k-variate mean zero normal distribution with covariance matrix $\mathbf{\Sigma}_k$. As in the proof of Lemma 8.2, by means of the Cramér-Wold device this establishes the limit distribution of $(\mathbb{X}_N(t_1), \ldots, \mathbb{X}_N(t_k))$, which is the same that of the vector $(\mathbb{G}^{\mathrm{HT}}(t_1), \ldots, \mathbb{G}^{\mathrm{HT}}(t_k))$, where $\mathbb{G}^{\mathrm{HT}}$ is a mean zero Gaussian process with covariance function $\mathbb{E}_{d,m} \mathbb{G}^{\mathrm{HT}}(s) \mathbb{G}^{\mathrm{HT}}(t) = \mu_{\pi 1}(s \wedge t) + \mu_{\pi 2} st$. From here on, the proof is completely the same as that of Theorem 3.1. □

To establish tightness for the process $\sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - F)$ we use the following decomposition

$$\sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - F) = \sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - \mathbb{F}_N) + \frac{\sqrt{n}}{\sqrt{N}} \cdot \sqrt{N}(\mathbb{F}_N - F). \tag{8.12}$$

The first process on the right hand side converges weakly to Gaussian process, according to Theorem 3.1. The process $\sqrt{N}(\mathbb{F}_N - F)$ also converges weakly to a Gaussian process, due to the classical Donsker theorem. In particular both processes on the right hand side are tight in $D(\mathbb{R})$ with the Skorohod metric. In general the sum of two tight processes in $D(\mathbb{R})$ is not necessarily tight. However, this will be the case if both processes converge weakly to continuous processes (see Lemma 9.2 in [BLRG15]).

Lemma 8.3. *Let $V_1, V_2, \ldots$ be a sequence of bounded i.i.d. random variables on $(\Omega, \mathfrak{F}, \mathbb{P}_m)$ with mean μ_V and variance σ_V^2, and let S_N^2 be defined by (3.2). Suppose (HT1) and (HT3) hold and $nS_N^2 \to \sigma_{\mathrm{HT}}^2 > 0$ in $\mathbb{P}_m$-probability. Then,*

$$\sqrt{n} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i V_i}{\pi_i} - \mu_V \right), \tag{8.13}$$

converges in distribution under $\mathbb{P}_{d,m}$ to a mean zero normal random variable with variance $\sigma_{\mathrm{HT}}^2 + \lambda \sigma_V^2$.

Note that, in view of the expression for σ^2_{HT} obtained in Lemma 9.1, for simple random sampling without replacement, the condition $\sigma^2_{\text{HT}} > 0$ implies that λ must differ from 1.

Proof. We decompose as follows

$$\begin{aligned}\frac{1}{S_N}\left(\frac{1}{N}\sum_{i=1}^N \frac{\xi_i V_i}{\pi_i} - \mu_V\right) =& \frac{1}{S_N}\left(\frac{1}{N}\sum_{i=1}^N \frac{\xi_i V_i}{\pi_i} - \frac{1}{N}\sum_{i=1}^N V_i\right) \\ &+ \frac{1}{\sqrt{n}S_N} \times \frac{\sqrt{n}}{\sqrt{N}} \times \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^N V_i - \mu_V\right).\end{aligned}$$

According to (HT3), the central limit theorem, Slutsky's theorem, and the fact that $nS_N^2 \to \sigma^2_{\text{HT}} > 0$ in probability,

$$\frac{1}{\sqrt{n}S_N} \times \frac{\sqrt{n}}{\sqrt{N}} \times \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^N V_i - \mu_V\right) \to N(0, \lambda\sigma_V^2/\sigma^2_{\text{HT}}), \tag{8.14}$$

in distribution under $\mathbb{P}_m$, whereas, thanks to (HT1),

$$\frac{1}{S_N}\left(\frac{1}{N}\sum_{i=1}^N \frac{\xi_i V_i}{\pi_i} - \frac{1}{N}\sum_{i=1}^N V_i\right) \to N(0,1), \quad \omega - \text{a.s.}, \tag{8.15}$$

in distribution under $\mathbb{P}_d$. Since the latter limit distribution does not depend on ω, we can apply Theorem 5.1(iii) from [RBSK05]. It follows that

$$\frac{1}{S_N}\left(\frac{1}{N}\sum_{i=1}^N \frac{\xi_i V_i}{\pi_i} - \mu_V\right) \to N\left(0, 1 + \lambda\sigma_V^2/\sigma^2_{\text{HT}}\right),$$

in distribution under $\mathbb{P}_{d,m}$. Together with $nS_N^2 \to \sigma^2_{HT}$ in probability, this implies that the random variable in (8.13) converges to a mean zero normal random variable with variance $\sigma^2_{\text{HT}} + \lambda\sigma_V^2$. ☐

Lemma 8.4. *Let $\mathbb{X}_N^F = \sqrt{n}(\mathbb{F}_N^{\text{HT}} - F)$ and suppose that (C1)-(C2),(HT1)-(HT4) hold. Then for any $k \in \{1, 2, \dots\}$, and $t_1, t_2, \dots, t_k \in \mathbb{R}$, the sequence $\left(\mathbb{X}_N^F(t_1), \dots, \mathbb{X}_N^F(t_k)\right)$ converges in distribution under $\mathbb{P}_{d,m}$ to a k-variate mean zero normal random vector with covariance matrix $\boldsymbol{\Sigma}^F_{\text{HT}} = \boldsymbol{\Sigma}_k^{\text{HT}} + \lambda\boldsymbol{\Sigma}_F$, where $\boldsymbol{\Sigma}_k^{\text{HT}}$ is given in (3.4) and $\boldsymbol{\Sigma}_F$ is the $k\times k$ matrix with (q,r)-entry $F(t_q \wedge t_r) - F(t_q)F(t_r)$, for $q, r = 1, 2, \dots, k$.*

Proof. The proof is similar to the proof of Lemma 8.2. The details can be found in [BLRG15]. ☐

Proof of Theorem 3.2 The proof is completely similar to that of Theorem 3.1. We first consider the process $\mathbb{X}_N^F = \sqrt{n}(\mathbb{F}_N^{\text{HT}} - F)$ for the case that the Y_i's follow a uniform distribution with $F(t) = t$. Decompose $\mathbb{X}_N^F$ as in (8.12). By Theorem 3.1, the first process on the right hand side of (8.12) converges weakly to a process in $C[0,1]$. Due to the classical Donsker theorem and (HT3), the second process on the right hand side of (8.12) also converges weakly to a process in $C[0,1]$. Tightness of $\mathbb{X}_N^F$ then follows from Lemma 9.2 in [BLRG15]. Convergence of the finite dimensional distributions is provided by Lemma 8.4. The theorem now follows from Theorem 13.5 in [Bil99] for the case that the Y_i's are uniformly distributed on $[0,1]$. Next, this is extended to Y_i's with a general c.d.f. F in the same way as in the proof of Theorem 3.1. ☐

To establish convergence in distribution of the finite dimensional distributions of $\sqrt{n}(\mathbb{F}_N^{\text{HT}} - F)$ under the conditions of Proposition 3.2, as in the proof of Lemma 8.4, we will use the Cramér-Wold device. To ensure that the limit in (9.2) is still strictly positive without imposing (HT4), we will need the following lemma. Its proof can befound in [BLRG15].

Lemma 8.5. *Let F be the c.d.f. of the i.i.d. $Y_1, \dots, Y_N$. For any k-tuple $(t_1, \dots, t_k) \in \mathbb{R}^k$, suppose that the values $F(t_1), \dots, F(t_k)$ are all distinct and such that $0 < F(t_i) < 1$. Let $a, b \in \mathbb{R}$, such that $a \geq b$. If $a > 0$, then the $k \times k$ matrix $\mathbf{M}$ with (i,j)-th element $M_{ij} = aF(t_i \wedge t_j) - bF(t_i)F(t_j)$ is positive definite.*

Lemma 8.6. *Let $\mathbb{X}_N^F = \sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - F)$ and suppose that n and π_i, π_{ij}, for $i, j = 1, 2, \dots, N$, are deterministic. Suppose that (C1)-(C2), (HT1) and (HT3) hold, as well as conditions (i)-(ii) of Proposition 3.2. Then, for any $k \in \{1, 2, \dots\}$, and $t_1, \dots, t_k \in \mathbb{R}$, $\big(\mathbb{X}_N^F(t_1), \dots, \mathbb{X}_N^F(t_k)\big)$ converges in distribution under $\mathbb{P}_{d,m}$ to a k-variate mean zero normal random vector with covariance matrix $\mathbf{\Sigma}_{\mathrm{HT}}^F$, with (q,r)-entry $(\mu_{\pi 1} + \lambda)F(t_q \wedge t_r) + (\mu_{\pi 2} - \lambda)F(t_q)F(t_r)$, for $q, r, = 1, 2, \dots, k$.*

Proof. The proof follows the same ideas as the proof of Lemma 8.4, but is a bit more technical. It can be found in [BLRG15]. □

Proof of Proposition 3.2 The proof is similar to that of Theorem 3.2. Tightness is obtained in the same way and the convergence of finite dimensional projections is provided by Lemma 8.6. The theorem now follows from Theorem 13.5 in [Bil99] for the case that the Y_i's are uniformly distributed on $[0, 1]$. Next, this is extended to Y_i's with a general c.d.f. F in the same way as in the proof of Theorem 3.1. □

Proof of Theorem 4.1 For part (i), note that with S_N^2 defined in (3.2) with $V_i = 1$, from (HT1) together with condition (4.6), it follows that

$$\sqrt{n} S_N \times \frac{1}{S_N} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} - 1 \right) \to N(0, \sigma_\pi^2), \quad \omega - \text{a.s.},$$

in distribution under $\mathbb{P}_d$. This implies

$$\sqrt{n} \left(\frac{\widehat{N}}{N} - 1 \right) = \sqrt{n} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} - 1 \right) \to N(0, \sigma_\pi^2), \tag{8.16}$$

in distribution under $\mathbb{P}_{d,m}$. In particular, since $n \to \infty$, this proves part (i).

The proof of part(ii) is along the same lines as the proof of Theorems 3.1 and 3.2. First consider the case, where the Y_i's are uniform, with $F(t) = t$ on $[0, 1]$. Then, with $\mathbb{F}_N^{\mathrm{HT}}$ defined in (2.1) and $\mathbb{X}_N^F = \sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - F)$, we can write $\mathbb{G}_N^\pi(t) = \mathbb{X}_N^F(t) - (\mathbb{X}_N^F(t) - \mathbb{G}_N^\pi(t))$. According to Theorem 3.2, the process $\mathbb{X}_N^F$ converges weakly to a continuous process. As a consequence of (8.16), the process

$$\mathbb{X}_N^F(t) - \mathbb{G}_N^\pi(t) = t\sqrt{n} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} - 1 \right),$$

also converges weakly to a continuous process. Hence, similar to the argument in the proof of Theorem 3.2, we conclude that the process $\mathbb{G}_N^\pi$ is tight. Next, we establish weak convergence of the finite dimensional projections

$$\big(\mathbb{G}_N^\pi(t_1), \dots, \mathbb{G}_N^\pi(t_k)\big). \tag{8.17}$$

To this end we apply the Cramér-Wold device and consider linear combinations

$$a_1 \mathbb{G}_N^\pi(t_1) + \cdots + a_k \mathbb{G}_N^\pi(t_k) = \frac{\sqrt{n}}{N} \sum_{i=1}^{N} \frac{\xi_i}{\pi_i} V_{ik}. \tag{8.18}$$

Convergence of (8.18), is obtained completely similar to that of (9.1) in Lemma 8.4, but this time with

$$V_{ik} = a_1\left(\mathbb{1}_{\{Y_i \le t_1\}} - t_1\right) + \cdots + a_k(\mathbb{1}_{\{Y_i \le t_k\}}) - t_k\right),$$

and $\mu_k = 0$. Using the fact that (HJ4) allows the use of Lemma 8.3, one can deduce that (8.18) converges in distribution under $\mathbb{P}_{d,m}$ to $a_1N_1+\cdots+a_kN_k$, where $(N_1,\dots,N_k)$ has a k-variate normal distribution with covariance matrix $\mathbf{\Sigma}^\pi = \mathbf{\Sigma}_k^{\mathrm{HJ}} + \lambda\mathbf{\Sigma}_F$, where $\mathbf{\Sigma}_k^{\mathrm{HJ}}$ and $\mathbf{\Sigma}_F$ are given in (4.5) and Lemma 8.4, respectively. By means of the Cramér-Wold device, this proves that (8.17) converges in distribution under $\mathbb{P}_{d,m}$ to a mean zero k-variate normal random vector with covariance matrix $\mathbf{\Sigma}^\pi$. This distribution is the same as that of $\big(\mathbb{G}^\pi(t_1),\dots,\mathbb{G}^\pi(t_k)\big)$, where $\mathbb{G}^\pi$ is a mean zero Gaussian process with covariance function

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \mathbb{E}_m\left[n\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\left(\mathbb{1}_{\{Y_i\le s\}}-s\right)\left(\mathbb{1}_{\{Y_i\le t\}}-t\right)\right] + \lambda\left(s\wedge t - st\right), \quad s,t\in\mathbb{R}.$$

Since $\mathbb{G}^\pi$ is continuous at 1, the theorem then follows from Theorem 13.5 in [Bil99] for the case of uniform Y_i's. Extension to Y_i's with a general c.d.f. F is completely similar to the proof of Theorem 3.1. □

Proof of Theorem 4.2 We use (4.2). From the proof of Theorem 4.1, we know that $\mathbb{G}_N^\pi$ is tight. Together with Theorem 4.1(i), it then follows that the limit behavior of $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}} - \mathbb{F}_N)$ is the same as that of the process $\mathbb{Y}_N$ defined in (4.3). This process can be written as

$$\mathbb{Y}_N(t) = \frac{\sqrt{n}}{N}\sum_{i=1}^N\left(\frac{\xi_i}{\pi_i}-1\right)\mathbb{1}_{\{Y_i\le t\}} - F(t)\frac{\sqrt{n}}{N}\sum_{i=1}^N\left(\frac{\xi_i}{\pi_i}-1\right).$$

As in the proofs of Theorems 3.1, 3.2, and 4.1, we first consider the case of uniform Y_i's. The first process on the right hand side is $\sqrt{n}(\mathbb{F}_N^{\mathrm{HT}} - \mathbb{F}_N)$, which converges weakly to a continuous process, according to Theorem 3.1, whereas the second process also converges to a continuous process due to (8.16). As in the proof of Theorem 3.2, one can then argue that $\mathbb{Y}_N$, being the difference of these processes, is tight. Next, we prove weak convergence of the finite dimensional projections

$$\big(\mathbb{Y}_N(t_1),\dots,\mathbb{Y}_N(t_k)\big). \tag{8.19}$$

As before, we apply the Cramér-Wold device and consider

$$a_1\mathbb{Y}_N(t_1)+\cdots+a_k\mathbb{Y}_N(t_k) = \sqrt{n}\left\{\frac{1}{N}\sum_{i=1}^N\frac{\xi_i}{\pi_i}V_{ik} - \frac{1}{N}\sum_{i=1}^N V_{ik}\right\}, \tag{8.20}$$

with

$$V_{ik} = a_1\left(\mathbb{1}_{\{Y_i \le t_1\}} - t_1\right) + \cdots + a_k(\mathbb{1}_{\{Y_i \le t_k\}}) - t_k\right).$$

Convergence of (8.20) is obtained completely similar to that of (8.8) in the proof of Lemma 8.2. From (HT1) and (HJ2), it follows that (8.20) converges in distribution under $\mathbb{P}_{d,m}$ to $a_1N_1+\cdots+a_kN_k$, where $(N_1,\dots,N_k)$ has a k-variate normal distribution with covariance matrix $\mathbf{\Sigma}_k^{\mathrm{HJ}}$ given in (4.5). By means of the Cramér-Wold device, this proves that (8.19) converges in distribution under $\mathbb{P}_{d,m}$ to a mean zero k-variate normal random vector with covariance matrix $\mathbf{\Sigma}_k^{\mathrm{HJ}}$. This distribution is the same as that of $\big(\mathbb{G}^{\mathrm{HJ}}(t_1),\dots,\mathbb{G}^{\mathrm{HJ}}(t_k)\big)$, where $\mathbb{G}^{\mathrm{HJ}}$ is a mean zero Gaussian process with covariance function

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \mathbb{E}_m\left[n\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\left(\mathbb{1}_{\{Y_i\le s\}}-s\right)\left(\mathbb{1}_{\{Y_i\le t\}}-t\right)\right],$$

for $s, t \in \mathbb{R}$. As before, the theorem now follows from Theorem 13.5 in [Bil99] for the case of uniform Y_i's, and is then extended to Y_i's with a general c.d.f. F. □

Proof of Theorem 4.3 The theorem follows directly from relation (4.7) and Theorem 4.1. □

The proofs of Propositions 4.1 and 4.2 are similar to those of Theorems 4.2 and 4.1, respectively, and can be found in [BLRG15]. The proofs for Corollaries 5.1 and 5.2 are fairly straightforward and can be found in [BLRG15].

References

[BCC14] Patrice Bertail, Emilie Chautru, and Stéphan Clémençon. Empirical processes in survey sampling. *Submitted, See also* *https://hal.archives-ouvertes.fr/hal-00989585*, 2014.

[BD09] Garry F. Barrett and Stephen G. Donald. Statistical inference with generalized Gini indices of inequality, poverty, and welfare. *J. Bus. Econom. Statist.*, 27(1):1–17, 2009.

[Bha07] Debopam Bhattacharya. Inference on inequality from household survey data. *J. Econometrics*, 137(2):674–707, 2007.

[Bil99] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[BLRG12] Helène Boistard, Hendrik P. Lopuhaä, and Anne Ruiz-Gazen. Approximation of rejective sampling inclusion probabilities and application to high order correlations. *Electron. J. Stat.*, 6:1967–1983, 2012.

[BLRG15] Helène Boistard, Hendrik P. Lopuhaä, and Anne Ruiz-Gazen. Supplement to "functional central limit theorems in survey sampling". 2015.

[BM11] Debopam Bhattacharya and Bhaskhar Mazumder. A nonparametric analysis of black and white differences in intergenerational income mobility in the united states. *Quant. Econ.*, 2(3):335–379, 2011.

[BO00] F. Jay Breidt and Jean D. Opsomer. Local polynomial regresssion estimators in survey sampling. *Ann. Statist.*, 28(4):1026–1053, 2000.

[BR09] David Binder and Georgia Roberts. *Handbook of Statistics 29B: Sample Surveys: Design, Methods and Applications.*, chapter Chapter 24: Design- and Model-Based Inference for Model Parameters, pages 33–54. Elsevier, Amsterdam, 2009.

[BS03] Yves G. Berger and Chris J. Skinner. Variance estimation for a low income proportion. *J. Roy. Statist. Soc. Ser. C*, 52(4):457–468, 2003.

[BW07] Norman E. Breslow and Jon A. Wellner. Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scand. J. Statist.*, 34(1):86–102, 2007.

[CCGL10] Hervé Cardot, Mohamed Chaouch, Camelia Goga, and Catherine Labruère. Properties of design-based functional principal components analysis. *J. Statist. Plann. Inference*, 140(1):75–91, 2010.

[Dav09] Russell Davidson. Reliable inference for the Gini index. *J. Econometrics*, 150(1):30–40, 2009.

[Dd08] Fabien Dell and Xavier d'Haultfœuille. Measuring the evolution of complex indicators: Theory and application to the poverty rate in France. *Ann. Économ. Statist.*, (90):259–290, 2008.

[DS92] Jean-Claude Deville and Carl-Erik Särndal. Calibration estimators in survey sampling. *Journal of the American statistical Association*, 87(418):376–382, 1992.

[Dud02] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[FF91] Carol A. Francisco and Wayne A. Fuller. Quantile estimation with a complex survey design. *Ann. Statist.*, 19(1):454–469, 1991.

[Ful09] W.A. Fuller. *Sampling Statistics*. Wiley Series in Survey Methodology. Wiley, New York, 2009.

[GT14] Eric Graf and Yves Tillé. Variance estimation using linearization for poverty and social exclusion indicators. *Survey Methodology*, 40(1):61–79, 2014.

[Háj64] Jaroslav Hájek. Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Statist.*, 35:1491–1523, 1964.

[KG98] Edward L. Korn and Barry I. Graubard. Variance estimation for superpopulation parameters. *Statist. Sinica*, 8(4):1131–1151, 1998.

[KR81] D. Krewski and J. N. K. Rao. Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods. *Ann. Statist.*, 9(5):1010–1019, 1981.

[OAB15] M. Oguz-Alper and Y. G. Berger. Variance estimation of change of poverty based upon the turkish eu-silc survey. *Journal of Official Statistics*, 31(2):155–175, 2015.

[PW93] Jens Præstgaard and Jon A. Wellner. Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.*, 21(4):2053–2086, 1993.

[RBSK05] Susana Rubin-Bleuer and Ioana Schiopu Kratina. On the two-phase framework for joint model and design-based inference. *Ann. Statist.*, 33(6):2789–2810, 2005.

[Sil86] B. W. Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.

[SW13] Takumi Saegusa and Jon A. Wellner. Weighted likelihood estimation under two-phase sampling. *Ann. Statist.*, 41(1):269–295, 2013.

[Tho97] M. E. Thompson. *Theory of sample surveys*, volume 74 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1997.

[vdV98] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

[vdVW96] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.

[Wan12] Jianqiang C. Wang. Sample distribution function based goodness-of-fit test for complex surveys. *Comput. Statist. Data Anal.*, 56(3):664–679, 2012.

Hélène Boistard
Toulouse School of Economics
21 allée de Brienne
31000 Toulouse, France
e-mail: helene@boistard.fr

Hendrik P. Lopuhaä
Delft Institute of Applied Mathematics
Delft University of Technology
Delft, The Netherlands
e-mail: h.p.lopuhaa@tudelft.nl

Anne Ruiz-Gazen
Toulouse School of Economics
21 allée de Brienne
31000 Toulouse, France
e-mail: anne.ruiz-gazen@tse-fr.eu

9 Supplemental Material

9.1 Proofs of Lemmas, Propositions and Corollaries in the main text

Proof of Lemma 8.4 We will use the Cramér-Wold device. To this end, we determine the limit distribution of $a_1\mathbb{X}_N^F(t_1)+\cdots+a_k\mathbb{X}_N^F(t_k)$, for $a_1,\ldots,a_k\in\mathbb{R}$ fixed and $\mathbf{a}_k^t=(a_1,\ldots,a_k)\neq(0,\ldots,0)$. As in the proof of Lemma 8.2, we consider

$$a_1\mathbb{X}_N^F(t_1)+\cdots+a_k\mathbb{X}_N^F(t_k)=\sqrt{n}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{\pi_i}V_{ik}-\mu_k\right),\tag{9.1}$$

where V_{ik} is defined in (8.9). We want to apply Lemma 8.3. As in (8.10),

$$nS_N^2\to\mathbf{a}_k^t\mathbf{\Sigma}_k^{\mathrm{HT}}\mathbf{a}_k,\quad \omega-\text{a.s.},\tag{9.2}$$

where $\mathbf{a}_k^t\mathbf{\Sigma}_k^{\mathrm{HT}}\mathbf{a}_k>0$, thanks to (HT4). This means that, according to Lemma 8.3, the right hand side of (9.1) converges in distribution under $\mathbb{P}_{d,m}$ to a mean zero normal random variable with variance

$$\mathbf{a}_k^t\mathbf{\Sigma}_k^{\mathrm{HT}}\mathbf{a}_k+\lambda\left\{\mathbb{E}_m[V_{1k}^2]-(\mathbb{E}_m[V_{1k}])^2\right\}=\mathbf{a}_k^t\mathbf{\Sigma}_{\mathrm{HT}}^F\mathbf{a}_k,$$

where

$$\mathbf{\Sigma}_{\mathrm{HT}}^F=\mathbf{\Sigma}_k^{\mathrm{HT}}+\lambda\mathbf{\Sigma}_F.\tag{9.3}$$

We conclude that (9.1) converges in distribution to $a_1N_1+\cdots+a_kN_k$, where $(N_1,\ldots,N_k)$ has a mean zero k-variate normal distribution with covariance matrix $\mathbf{\Sigma}_{\mathrm{HT}}^F$. By the Cramér-Wold device, this proves the lemma. □

Proof of Lemma 8.5 Without loss of generality we may assume $0<F(t_1)<\cdots<F(t_k)<1$, since we can permute the rows and columns of $\mathbf{M}$ without changing the determinant. For the entries of $\mathbf{M}$ we can distinguish three situations:

1. if $1\leq j<i\leq k$, then $M_{ij}=aF(t_j)-bF(t_i)F(t_j)$
2. if $1\leq i=j\leq k$, then $M_{ij}=aF(t_i)-bF(t_i)^2$
3. if $1\leq i<j\leq k$, then $M_{ij}=aF(t_i)-bF(t_i)F(t_j)$.

Now, for $2\leq i\leq k$, multiply the i-th row by $F(t_1)/F(t_i)$. This changes the determinant with a factor $F(t_1)^{k-1}/F(t_2)\cdots F(t_k)>0$, and as a result, all entries in column j, at positions $1\leq i\leq j\leq k$, are the same: $aF(t_1)-bF(t_1)F(t_j)$. Hence, if we subtract row-2 from row-1, then row-3 from row-2, ..., and then row-k from row-$(k-1)$, we get a new matrix $\mathbf{M}'$ with a right-upper triangle consisting of zero's and a main diagonal with elements $M'_{ii}=aF(t_1)-aF(t_1)F(t_i)/F(t_{i+1})$, if $1\leq i\leq k-1$, and $M'_{kk}=aF(t_1)-bF(t_1)F(t_k)$. It follows that

$$\begin{aligned}\det(\mathbf{M})&=\frac{F(t_2)\cdots F(t_k)}{F(t_1)^{k-1}}\det(\mathbf{M}')\\&=a^{k-1}F(t_1)(F(t_2)-F(t_1))\cdots(F(t_k)-F(t_{k-1}))(a-bF(t_k))>0,\end{aligned}$$

since $a>0$, $0<F(t_1)<\cdots<F(t_k)<1$, and $a-bF(t_k)>a-b\geq 0$. □

Proof of Lemma 8.6 The proof is similar to that of Lemma 8.4. We determine the limit distribution of (9.1). Note that without loss of generality we can assume that $0 \leq F(t_1) \leq \cdots \leq F(t_k) \leq 1$. In contrast with the proof of Lemma 8.4, we now have to distinguish between several cases.

We first consider the situation where all $F(t_i)$'s are distinct and such that $0 < F(t_i) < 1$. From (HT1) and Lemma 9.1(ii) we conclude that

$$nS_N^2 \to \sigma_{\mathrm{HT}}^2 = \mu_{\pi 1}\mathbb{E}_m[V_{1k}^2] + \mu_{\pi 2}\left(\mathbb{E}_m[V_{1k}]\right)^2 = \mathbf{a}_k^t\mathbf{\Sigma}_k\mathbf{a}_k,$$

where

$$\mathbf{\Sigma}_k = \Big(\mu_{\pi 1}F(t_q \wedge t_r) + \mu_{\pi 2}F(t_q)F(t_r)\Big)_{q,r=1}^k. \tag{9.4}$$

First note that

$$\mu_{\pi 1} + \mu_{\pi 2} = \lim_{N\to\infty}\frac{n}{N^2}\sum_{i=1}^N\sum_{j=1}^N\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j} = \lim_{N\to\infty}\frac{n}{N^2}\mathrm{Var}\left(\sum_{i=1}^N\frac{\xi_i}{\pi_i}\right) \geq 0.$$

Therefore, together with condition (i) we can apply Lemma 8.5 with $a = \mu_{\pi 1}$ and $b = -\mu_{\pi 2}$. It follows that $\mathbf{\Sigma}_k$ is positive definite, so that $\sigma_{\mathrm{HT}}^2 > 0$. This means that, according to Lemma 8.3, the right hand side of (9.1) converges in distribution under $\mathbb{P}_{d,m}$ to a mean zero normal random variable with variance $(\mu_{\pi 1}+\lambda)\mathbb{E}_m[V_{1k}^2] + (\mu_{\pi 2}-\lambda)\left(\mathbb{E}_m[V_{1k}]\right)^2 = \mathbf{a}_k^t\mathbf{\Sigma}_{\mathrm{HT}}^F\mathbf{a}_k$, where

$$\mathbf{\Sigma}_{\mathrm{HT}}^F = \Big((\mu_{\pi 1}+\lambda)F(t_q \wedge t_r) + (\mu_{\pi 2}-\lambda)F(t_q)F(t_r)\Big)_{q,r=1}^k. \tag{9.5}$$

We conclude that (9.1) converges in distribution to $a_1N_1 + \cdots + a_kN_k$, where $(N_1, \ldots, N_k)$ has a mean zero k-variate normal distribution with covariance matrix $\mathbf{\Sigma}_{\mathrm{HT}}^F$. By means of the Cramér-Wold device, this proves the lemma for the case that $0 < F(t_1) < \cdots < F(t_k) < 1$.

The case that the $F(t_i)$'s are not all distinct, but still satisfy $0 < F(t_i) < 1$, can be reduced to the case where all $F(t_i)$'s are distinct. This can be seen as follows. For simplicity, suppose $F(t_1) = \cdots = F(t_m) = F(t_0)$, with $0 < F(t_0) < F(t_{m+1}) < \cdots < F(t_k) < 1$. Then we can write (9.1) as

$$a_0\mathbb{X}_N^F(t_0) + a_{m+1}\mathbb{X}_N^F(t_{m+1}) + \cdots + a_k\mathbb{X}_N^F(t_k), \tag{9.6}$$

where $a_0 = a_1 + \cdots + a_m$. As before, with (HT4) and Lemma 8.5, it follows from Lemma 8.3 that (9.6) converges in distribution to a mean zero normal random variable with variance $\mathbf{a}_0^t\mathbf{\Sigma}_0^F\mathbf{a}_0$, where $\mathbf{a}_0 = (a_0, a_{m+1}, \ldots, a_k)^t$ and

$$\mathbf{\Sigma}_0^F = \gamma_{\pi 1}\mathbb{E}_m[\mathbf{Y}_0\mathbf{Y}_0^t] + (\gamma_{\pi 2}-\lambda)\left(\mathbb{E}_m[\mathbf{Y}_0]\right)\left(\mathbb{E}_m[\mathbf{Y}_0]\right)^t,$$

with $\mathbf{Y}_0 = (\mathbb{1}_{\{Y_i \leq t_0\}}, \mathbb{1}_{\{Y_i \leq t_{m+1}\}}, \ldots, \mathbb{1}_{\{Y_i \leq t_k\}})^t$. However, note that

$$\begin{aligned}\mathbf{a}_0^t\mathbf{Y}_0 &= (a_1 + \cdots + a_m)\mathbb{1}_{\{Y_i \leq t_0\}} + a_{m+1}\mathbb{1}_{\{Y_i \leq t_{m+1}\}} + \cdots + a_k\mathbb{1}_{\{Y_i \leq t_k\}} \\ &= a_1\mathbb{1}_{\{Y_i \leq t_1\}} + \cdots + a_k\mathbb{1}_{\{Y_i \leq t_k\}} = \mathbf{a}_k^t\mathbf{Y}_{1k},\end{aligned},$$

where $\mathbf{a}_k = (a_1, \ldots, a_k)^t$ and $\mathbf{Y}_{1k} = (\mathbb{1}_{\{Y_i \leq t_1\}}, \ldots, \mathbb{1}_{\{Y_i \leq t_k\}})^t$, as before. This means that $\mathbf{a}_0^t\mathbf{\Sigma}_0^F\mathbf{a}_0 = \mathbf{a}_k^t\mathbf{\Sigma}_{\mathrm{HT}}^F\mathbf{a}_k$, with $\mathbf{\Sigma}_{\mathrm{HT}}^F$ from (9.3). It follows that (9.1) converges in distribution to $a_1N_1 + \cdots + a_kN_k$, where $(N_1, \ldots, N_k)$ has a mean zero k-variate normal distribution with covariance matrix $\mathbf{\Sigma}_{\mathrm{HT}}^F$. By means of the Cramér-Wold device, this proves the lemma for the case $F(t_1) = \cdots = F(t_m) = F(t_0) < F(t_{m+1}) < \cdots < F(t_k) < 1$. The argument is the same for other cases with multiple $F(t_i) \in (0,1)$ being equal to each other.

Next, consider the case $F(t_1) = 0$. In this case, $\mathbb{1}_{\{Y_i \leq t_1\}} = 0$ with probability one. This means that the summation on the left hand side of (9.1) reduces to $a_2\mathbb{X}_N^F(t_2) + \cdots + a_k\mathbb{X}_N^F(t_k)$ and

$$\mathbf{\Sigma}_{\text{HT}} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{\Sigma}_{\text{HT},k-1} & \\ 0 & & & \end{pmatrix}, \tag{9.7}$$

where $\mathbf{\Sigma}_{\text{HT},k-1}$ is the matrix in (9.4) based on $0 < F(t_2) < \cdots < F(t_k) < 1$. When $\mathbf{a}_{k-1}^t = (a_2, \ldots, a_k) \neq (0, \ldots, 0)$, then

$$\sigma_{\text{HT}}^2 = \mathbf{a}_k^t \mathbf{\Sigma}_{\text{HT}}^F \mathbf{a}_k = \mathbf{a}_{k-1}^t \mathbf{\Sigma}_{\text{HT},k-1} \mathbf{a}_{k-1} > 0,$$

because $\mathbf{\Sigma}_{\text{HT},k-1}$ is positive definite, due to (HT4) and Lemma 8.5. This allows application of Lemma 8.3 to (9.1). As in the previous cases, we conclude that (9.1) converges in distribution to $a_1N_1 + \cdots + a_kN_k$, where $(N_1, \ldots, N_k)$ has a mean zero k-variate normal distribution with covariance matrix $\mathbf{\Sigma}_{\text{HT}}^F$ given by (9.3). When $\mathbf{a}_k^t = (a_1, 0, \ldots, 0)$, with $a_1 \neq 0$, then both (9.1) and $a_1N_1 + \cdots + a_kN_k$ are equal to zero. According to the Cramér-Wold device, this proves the lemma for the case $F(t_k) = 0$.

It remains to consider the case $F(t_k) = 1$. In this case, the (k, k)-th element of the matrix $\mathbf{\Sigma}_{\text{HT}}$ in (9.4) is equal to $\mu_{\pi 1} + \mu_{\pi 2}$. We distinguish between $\mu_{\pi 1} + \mu_{\pi 2} = 0$ and $\mu_{\pi 1} + \mu_{\pi 2} > 0$. In the latter case, from the proof of Lemma 8.5 we find that $\mathbf{\Sigma}_{\text{HT}}$ has determinant

$$\mu_{\pi 1}^{k-1} F(t_1) \prod_{i=2}^{k} (F(t_i) - F(t_{i-1}))(\mu_{\pi 1} + \mu_{\pi 2}) > 0,$$

using (HT4) and $0 < F(t_1) < \cdots < F(t_{k-1}) < F(t_k) = 1$. This allows application of Lemma 8.3 to (9.1). As before, we conclude that (9.1) converges in distribution to $a_1N_1 + \cdots + a_kN_k$, where $(N_1, \ldots, N_k)$ has a k-variate mean zero normal distribution with covariance matrix $\mathbf{\Sigma}_{\text{HT}}^F$ from (9.3). According to the Cramér-Wold device, this proves the lemma for the case $F(t_k) = 1$ and $\mu_{\pi 1} + \mu_{\pi 2} > 0$.

Next, consider the case $F(t_k) = 1$ and $\mu_{\pi 1} + \mu_{\pi 2} = 0$. This means

$$\mathbf{\Sigma}_{\text{HT}} = \begin{pmatrix} & & & 0 \\ & \mathbf{\Sigma}_{\text{HT},k-1} & & \vdots \\ & & & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}, \tag{9.8}$$

where $\mathbf{\Sigma}_{\text{HT},k-1}$ is the matrix in (9.4) corresponding to $0 < F(t_1) < \cdots < F(t_{k-1}) < 1$. When $\mathbf{a}_{k-1}^t = (a_1, \ldots, a_{k-1}) \neq (0, \ldots, 0)$, then

$$\sigma_{\text{HT}}^2 = \mathbf{a}_k^t \mathbf{\Sigma}_{\text{HT}} \mathbf{a}_k = \mathbf{a}_{k-1}^t \mathbf{\Sigma}_{\text{HT},k-1} \mathbf{a}_{k-1} > 0,$$

because $\mathbf{\Sigma}_{\text{HT},k-1}$ is positive definite, due to (HT4) and Lemma 8.5. This allows application of Lemma 8.3 to (9.1). As in the previous cases, we conclude that (9.1) converges in distribution to $a_1N_1 + \cdots + a_kN_k$, where $(N_1, \ldots, N_k)$ has a k-variate mean zero normal distribution with covariance matrix $\mathbf{\Sigma}_{\text{HT}}^F$ given by (9.3). When $\mathbf{a}_k^t = (0, \ldots, 0, a_k)$, with $a_k \neq 0$, then $a_1N_1 + \cdots + a_kN_k = 0$ and

$$a_1\mathbb{X}_N^F(t_1) + \cdots + a_k\mathbb{X}_N^F(t_k) = a_k\sqrt{n}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{\pi_i} - 1\right).$$

converges to zero in probability. The latter follows from the fact that according to (HT1) and Lemma 9.1, we have that

$$\sqrt{n}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{\pi_i}-1\right)\to N(0,\mu_{\pi1}+\mu_{\pi2}), \tag{9.9}$$

in distribution under $\mathbb{P}_{d,m}$. According to the Cramér-Wold device, this proves the lemma for the case $F(t_k)=1$ and $\mu_{\pi1}+\mu_{\pi2}=0$. Finally, the argument for the case that $F(t_1)=0$ and $F(t_k)=1$ simultaneously, either with or without repeated among the $F(t_i)$'s, is completely similar. This finishes the proof. □

Proof of Proposition 4.1 The proof is similar to that of Theorem 4.2. We find that the limit behavior of $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}}-\mathbb{F}_N)$ is the same as that of the process $\mathbb{Y}_N$ defined in (4.3). When we first consider the case of uniform Y_i's with $F(t)=t$, tightness of the process $\mathbb{Y}_N$ follows in the same way as in the proof of Theorem 4.2. It remains to establish weak convergence of the finite dimensional projections (8.19). This can be done in the same way as in the proof of Proposition 3.1, but this time with

$$V_{ik}=a_1\left(\mathbb{1}_{\{Y_i\leq t_1\}}-t_1\right)+\cdots+a_k(\mathbb{1}_{\{Y_i\leq t_k\}})-t_k\big).$$

From (HT1) and Lemma 9.1(i) we conclude that (8.20) converges in distribution to a mean zero normal random variable with variance

$$\sigma^2_{\mathrm{HT}}=\mu_{\pi1}\mathbb{E}_m\left[V_{1k}^2\right]=\mathbf{a}_k^t\widetilde{\mathbf{\Sigma}}_k\mathbf{a}_k,$$

where $\widetilde{\mathbf{\Sigma}}_k$ is the $k\times k$-matrix with (q,r)-element equal to $\mu_{\pi_1}(t_q\wedge t_r-t_qt_r)$. We conclude that (8.20) converges in distribution to $a_1N_1+\cdots+a_kN_k$, where $(N_1,\ldots,N_k)$ has a k-variate mean zero normal distribution with covariance matrix $\widetilde{\mathbf{\Sigma}}_k$. By means of the Cramér-Wold device this establishes the limit distribution of (8.19), which is the same as that of the vector $(\mathbb{G}^{\mathrm{HJ}}(t_1),\ldots,\mathbb{G}^{\mathrm{HJ}}(t_k))$, where $\mathbb{G}^{\mathrm{HJ}}$ is a mean zero Gaussian process with covariance function

$$\mathbb{E}_{d,m}\mathbb{G}^{\mathrm{HJ}}(s)\mathbb{G}^{\mathrm{HJ}}(t)=\mu_{\pi1}\left(s\wedge t-st\right).$$

From here on, the proof is completely the same as that of Theorem 4.2. □

Proof of Proposition 4.2 From relation (4.7) and Theorem 4.1 we know that the limit behavior of $\sqrt{n}(\mathbb{F}_N^{\mathrm{HJ}}-F)$ is the same as that of $\mathbb{G}_N^{\pi}$. Tightness of $\mathbb{G}_N^{\pi}$ has been obtained in the proof of Theorem 4.1. It remains to establish weak convergence of (8.17). This can be done in the same way as in the proof of Lemma 8.6, but this time with

$$V_{ik}=a_1\left(\mathbb{1}_{\{Y_i\leq t_1\}}-F(t_1)\right)+\cdots+a_k\left(\mathbb{1}_{\{Y_i\leq t_k\}}-F(t_k)\right)$$

and $\mu_k=0$. When $0<F(t_1)<\cdots<F(t_k)<1$, from (HT1) and Lemma 9.1 we find that $nS_N^2\to\mu_{\pi1}\mathbb{E}_m[V_{1k}^2]=\mathbf{a}_k^t\mathbf{\Sigma}_k\mathbf{a}_k$, where

$$\mathbf{\Sigma}_k=\mu_{\pi1}\Big(F(t_q\wedge t_r)-F(t_q)F(t_r)\Big)_{q,r=1}^{k}. \tag{9.10}$$

From condition (i) of Proposition 3.2 and Lemma 8.5, it follows that $\mathbf{\Sigma}_k$ is positive definite, so that $\mathbf{a}_k^t\mathbf{\Sigma}_k\mathbf{a}_k>0$. Hence, according to Lemma 8.3, the right hand side of (8.18) converges in distribution under $\mathbb{P}_{d,m}$ to a mean zero normal random variable with variance $(\mu_{\pi1}+\lambda)\mathbb{E}_m[V_{1k}^2]=\mathbf{a}_k^t\mathbf{\Sigma}_{\mathrm{HJ}}^F\mathbf{a}_k$, where

$$\mathbf{\Sigma}_{\mathrm{HJ}}^F=\Big((\mu_{\pi1}+\lambda)F(t_q\wedge t_r)\Big)_{q,r=1}^{k}. \tag{9.11}$$

We conclude that the right hand side of (8.18) converges in distribution to $a_1N_1+\cdots+a_kN_k$, where $(N_1,\ldots,N_k)$ has a mean zero k-variate normal distribution with covariance matrix $\boldsymbol{\Sigma}^F_{\mathrm{HJ}}$. By means of the Cramér-Wold device, this proves weak convergence of $\big(\mathbb{G}^\pi_N(t_1),\ldots,\mathbb{G}^\pi_N(t_k)\big)$ for the case that $0<F(t_1)<\cdots<F(t_k)<1$. As in the proof of Lemma 8.6, the case where the $F(t_i)$'s are not all distinct, but satisfy $0<F(t_i)<1$, the case $F(t_1)=0$, and the case $F(t_k)=1$, can be reduced to the previous case. From here on, the proof is completely the same as that of Theorem 4.1. □

Proof of (5.2) Following [Dd08], one can write $\phi=\psi_2\circ\psi_1$, where

$$\begin{aligned}\psi_1(F)&=\big(F,\beta F^{-1}(\alpha)\big)\\ \psi_2(F,x)&=F(x).\end{aligned}$$

The Hadamard-derivative of ϕ can then be obtained from the chain rule, e.g., see Lemma 3.9.3 in [vdVW96]. According to Lemma 3.9.20 in [vdVW96], for $0<\alpha<1$ and $F\in\mathbb{D}_\phi$ that have a positive derivative at $F^{-1}(\alpha)$, the map ψ_1 is Hadamard-differentiable at F tangentially to the set of functions $h\in D(\mathbb{R})$ that are continuous at $F^{-1}(\alpha)$ with derivative

$$\psi'_{1,F}(h)=\left(h,-\beta\frac{h(F^{-1}(\alpha))}{f(F^{-1}(\alpha))}\right).$$

It is fairly straightforward to show that for F that are differentiable at x, the mapping ψ_2 is Hadamard-differentiable at (F,x) tangentially to the set of pairs (h,ϵ), such that h is continuous at x and $\epsilon\in\mathbb{R}$, with derivative

$$\psi'_{2,(F,x)}(h,\epsilon)=\epsilon f(x)+h(x).$$

Then for $F\in\mathbb{D}_\phi$ that are differentiable at $\beta F^{-1}(\alpha)$, the mapping ψ_2 is Hadamard-differentiable at $\psi_1(F)=\big(F,\beta F^{-1}(\alpha)\big)$. It follows from the chain rule that $\phi(F)=F\left(\beta F^{-1}(\alpha)\right)=\psi_2\circ\psi_1(F)$ is Hadamard-differentiable at F tangentially to the set $\mathbb{D}_0$ consisting of functions $h\in D(\mathbb{R})$ that are continuous at $F^{-1}(\alpha)$ with derivative

$$\phi'_F(h)=-\beta\frac{f(\beta F^{-1}(\alpha))}{f(F^{-1}(\alpha))}h(F^{-1}(\alpha))+h(\beta F^{-1}(\alpha)).$$

□

Proof of Corollary 5.1 The mapping $\phi:\mathbb{D}_\phi\subset D(\mathbb{R})\mapsto\mathbb{R}$ is Hadamard-differentiable at F tangentially to the set $\mathbb{D}_0$ consisting of functions $h\in D(\mathbb{R})$ that are continuous at $F^{-1}(\alpha)$. According to Theorem 3.2, the sequence $\sqrt{n}(\mathbb{F}^{\mathrm{HT}}_N-F)$ converges weakly to a mean zero Gaussian process $\mathbb{G}^{\mathrm{HT}}_F$ with covariance structure

$$\mathbb{E}_{d,m}\mathbb{G}^{\mathrm{HT}}_F(s)\mathbb{G}^{\mathrm{HT}}_F(t)=(\mu_{\pi1}+\lambda)F(s\wedge t)+(\mu_{\pi2}-\lambda)F(s)F(t),\tag{9.12}$$

for $s,t\in\mathbb{R}$. It then follows from Theorem 3.9.4 in [vdVW96], that the random variable $\sqrt{n}(\phi(\mathbb{F}^{\mathrm{HT}}_N)-\phi(F))$ converges weakly to

$$-\beta\frac{f(\beta F^{-1}(\alpha))}{f(F^{-1}(\alpha))}\mathbb{G}^{\mathrm{HT}}_F(F^{-1}(\alpha))+\mathbb{G}^{\mathrm{HT}}_F(\beta F^{-1}(\alpha)),$$

which has a normal distribution with mean zero and variance

$$\begin{aligned}\sigma^2_{\mathrm{HT},\alpha,\beta}=\beta^2&\frac{f(\beta F^{-1}(\alpha))^2}{f(F^{-1}(\alpha))^2}\mathbb{E}\left[\mathbb{G}^{\mathrm{HT}}_F(F^{-1}(\alpha))^2\right]\\ &+\mathbb{E}\left[\mathbb{G}^{\mathrm{HT}}_F(\beta F^{-1}(\alpha))^2\right]\\ &-2\beta\frac{f(\beta F^{-1}(\alpha))}{f(F^{-1}(\alpha))}\mathbb{E}\left[\mathbb{G}^{\mathrm{HT}}_F(F^{-1}(\alpha))\mathbb{G}^{\mathrm{HT}}_F(\beta F^{-1}(\alpha))\right].\end{aligned}$$

The precise expression can then be derived from (9.12), which proves part one. For part two, write

$$\sqrt{n}\left(\phi(\mathbb{F}_N^{\mathrm{HT}}) - \phi(\mathbb{F}_N)\right) = \sqrt{n}\left(\phi(\mathbb{F}_N^{\mathrm{HT}}) - \phi(F)\right) + \frac{\sqrt{n}}{\sqrt{N}}\sqrt{N}\left(\phi(\mathbb{F}_N) - \phi(F)\right).$$

The process $\sqrt{N}(\mathbb{F}_N - F)$ converges weakly to a mean zero Gaussian process $\mathbb{G}_F$. Then, Hadamard-differentiability of ϕ together with Theorem 3.9.4 in [vdVW96] yields that the sequence $\sqrt{N}(\phi(\mathbb{F}_N) - \phi(F))$ converges weakly to $\phi'_F(\mathbb{G}_F)$. As $n/N \to 0$, the theorem follows from part one. □

Proof of Corollary 5.2 The proof is completely the same as that of Corollary 5.1, with the only difference that the covariance structure of the limiting process $\sqrt{n}(\phi(\mathbb{F}_N^{\mathrm{HJ}}) - \phi(F))$ is now given in Theorem 4.3. □

9.2 Additional Lemmas

Lemma 9.1. *Let S_N^2 be defined by (3.2), where $V_1, V_2, \dots$ is a sequence of i.i.d. random variables on $(\Omega, \mathfrak{F}, \mathbb{P}_m)$ with $\mathbb{E}_m[V_1^4] < \infty$. Suppose that n and π_i, π_{ij}, for $i, j = 1, 2, \dots, N$ are deterministic and let $\mathbb{V}_m(S_N^2)$ denote the variance of S_N^2. If (C1)-(C2) hold, then $n^2\mathbb{V}_m[S_N^2] = O(1/N)$. Then,*

(i) if $\mathbb{E}_m[V_1] = 0$ and condition (i) in Proposition 3.1 holds,

$$nS_N^2 \to \sigma_{\mathrm{HT}}^2 = \mu_{\pi 1}\mathbb{E}_m[V_1^2], \quad \textit{in } \mathbb{P}_m\textit{-probability.}$$

(ii) if $\mathbb{E}_m[V_1] \neq 0$ and conditions (i)-(ii) in Proposition 3.1 hold,

$$nS_N^2 \to \sigma_{\mathrm{HT}}^2 = \mu_{\pi 1}\mathbb{E}_m[V_1^2] + \mu_{\pi 2}\left(\mathbb{E}_m[V_1]\right)^2, \quad \textit{in } \mathbb{P}_m\textit{-probability.}$$

Proof. For any $\epsilon > 0$, by Markov inequality we have

$$\mathbb{P}_m\left\{|nS_N^2 - \mathbb{E}_m[nS_N^2]| > \epsilon\right\} < \frac{n^2\mathbb{V}_m[S_N^2]}{\epsilon^2}, \tag{9.13}$$

where $\mathbb{V}_m$ denotes the variance of S_N^2 under the super-population model. In order to compute $\mathbb{V}_m[S_N^2]$, we first have

$$\begin{aligned}\mathbb{E}_m[S_N^2] &= \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}\mathbb{E}_m(V_iV_j) \\ &= \frac{\mathbb{E}_m[V_1^2]}{N^2}\sum_{i=1}^N \frac{1-\pi_i}{\pi_i} + \frac{(\mathbb{E}_m[V_1])^2}{N^2}\sum\sum_{i\neq j}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}.\end{aligned} \tag{9.14}$$

From this, tedious but straightforward calculus leads to the expression for $(\mathbb{E}_m[S_N^2])^2$ and $\mathbb{E}_m[S_N^4]$. One finds

$$N^4\left(\mathbb{E}_m\left[S_N^2\right]\right)^2 = a_1\left(\mathbb{E}_m[V_1]\right)^4 + a_2\mathbb{E}_m\left[V_1^2\right]\left(\mathbb{E}_m\left[V_1\right]\right)^2 + a_3\left(\mathbb{E}_m\left[V_1^2\right]\right)^2,$$

where, according to (C1)-(C2):

$$
\begin{aligned}
a_1 &= \sum\sum\sum\sum_{(i,j,k,l)\in D_{4,N}} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l} \\
&\quad + 4\sum\sum\sum_{(i,j,l)\in D_{3,N}} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\frac{\pi_{il}-\pi_i\pi_l}{\pi_i\pi_l} + 2\sum\sum_{(i,j)\in D_{2,N}} \left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)^2 \\
&= \sum\sum\sum\sum_{(i,j,k,l)\in D_{4,N}} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l} + O(N^3/n^2) + O(N^2/n^2) \\
a_2 &= 2\sum\sum\sum_{(i,k,l)\in D_{3,N}} \frac{1-\pi_i}{\pi_i}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l} + 4\sum\sum_{(i,k)\in D_{2,N}} \frac{1-\pi_i}{\pi_i}\frac{\pi_{ik}-\pi_i\pi_k}{\pi_i\pi_k} \\
&= 2\sum\sum\sum_{(i,k,l)\in D_{3,N}} \frac{1-\pi_i}{\pi_i}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l} + O(N^3/n^2) \\
a_3 &= \sum\sum_{(i,j)\in D_{2,N}} \frac{1-\pi_i}{\pi_i}\frac{1-\pi_j}{\pi_j} + \sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i}\right)^2 \\
&= \sum\sum_{(i,j)\in D_{2,N}} \frac{1-\pi_i}{\pi_i}\frac{1-\pi_j}{\pi_j} + O(N^3/n^2).
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
N^4\mathbb{E}_m\left[S_N^4\right] &= b_1\left(\mathbb{E}_m[V_1]\right)^4 + b_2\mathbb{E}_m\left[V_1^2\right]\left(\mathbb{E}_m[V_1]\right)^2 \\
&\quad + b_3\left(\mathbb{E}_m\left[V_1^2\right]\right)^2 + b_4\mathbb{E}_m[V_1]\mathbb{E}_m\left[V_1^3\right]
\end{aligned}
$$

where

$$
\begin{aligned}
b_1 &= \sum\sum\sum\sum_{(i,j,k,l)\in D_{4,N}} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l} + \sum_{i=1}^{N}\left(\frac{1-\pi_i}{\pi_i}\right)^2 \\
&= \sum\sum\sum\sum_{(i,j,k,l)\in D_{4,N}} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l} + O(N^3/n^2) \\
b_2 &= 2\sum\sum\sum_{(i,k,l)\in D_{3,N}} \frac{1-\pi_i}{\pi_i}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l} + 4\sum\sum\sum_{(i,j,l)\in D_{3,N}} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\frac{\pi_{il}-\pi_i\pi_l}{\pi_i\pi_l} \\
&= 2\sum\sum\sum_{(i,k,l)\in D_{3,N}} \frac{1-\pi_i}{\pi_i}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l} + O(N^3/n^2)
\end{aligned}
$$

$$
\begin{aligned}
b_3 &= \sum\sum_{(i,k)\in D_{2,N}} \frac{1-\pi_i}{\pi_i}\frac{1-\pi_k}{\pi_k} + 2\sum\sum_{(i,j)\in D_{2,N}} \left(\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\right)^2 \\
&= \sum\sum_{(i,k)\in D_{2,N}} \frac{1-\pi_i}{\pi_i}\frac{1-\pi_k}{\pi_k} + O(N^2/n^2) \\
b_4 &= 4\sum\sum_{(i,j)\in D_{2,N}} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\frac{1-\pi_j}{\pi_j} = O(N^3/n^2).
\end{aligned}
$$

The variance expression for S_N^2 is deduced easily from the previous computations. From the expression derived in [BLRG15], we find that $a_i - b_i = O(N^3/n^2)$, for $i = 1, 2, 3$, and $b_4 = O(N^3/n^2)$, so that

$$n^2\mathbb{V}_m[S_N^2] = n^2\mathbb{E}_m[S_N^4] - n^2\left(\mathbb{E}_m[S_N^2]\right)^2 = O(1/N). \tag{9.15}$$

From (9.13) we conclude that $nS_N^2 - \mathbb{E}_m[nS_N^2]$ tends to zero in $\mathbb{P}_m$-probability. As a consequence, statements (i) and (ii) follow from (9.14). □

Lemma 9.2. *If $x_N \rightsquigarrow x$ and $y_N \rightsquigarrow y$ in $D[0,1]$ with the Skorohod metric, and $x, y \in C[0,1]$, then the sequence $\{x_N + y_N\}$ is also tight in $D[0,1]$.*

Proof. We can use Theorem 13.2 from [Bil99]. The first condition follows easily since

$$\sup_{t\in[0,1]} |x_N(t) + y_N(t)| \leq \sup_{t\in[0,1]} |x_N(t)| + \sup_{t\in[0,1]} |y_N(t)|.$$

Because $x_N \rightsquigarrow x$ and $y_N \rightsquigarrow y$ both sequences $\{x_N\}$ and $\{y_N\}$ are tight, so that they satisfy the first condition of Theorem 13.2 individually. For condition (ii) of Theorem 13.2 in [Bil99], choose $\epsilon > 0$. According to (12.7) in [Bil99], for any $0 < \delta < 1/2$,

$$w'_x(\delta) \leq w_x(2\delta).$$

This means that

$$\begin{aligned}\mathbb{P}\left\{w'_{x_N+y_N}(\delta) \geq \epsilon\right\} &\leq \mathbb{P}\left\{w_{x_N+y_N}(2\delta) \geq \epsilon\right\} \\ &\leq \mathbb{P}\left\{w_{x_N}(2\delta) \geq \epsilon/2\right\} + \mathbb{P}\left\{w_{y_N}(2\delta) \geq \epsilon/2\right\}.\end{aligned}$$

Consider the first probability. Since $x_N \rightsquigarrow x$ in $D[0,1]$ with the Skorohod metric, according to the almost sure representation theorem (see, e.g., Theorem 11.7.2 in [Dud02]), there exist $\widetilde{x}_n$ and $\widetilde{x}$, having the same distribution as x_N and x, respectively, such that $\widetilde{x}_N \to \widetilde{x}$, with probability one, in the Skorohod metric. Because $\widetilde{x} \stackrel{d}{=} x$ and $x \in C[0,1]$, also $\widetilde{x} \in C[0,1]$. Hence, since $\widetilde{x}$ is continuous, it follows that

$$\sup_{t\in[0,1]} |\widetilde{x}_N(t) - \widetilde{x}(t)| \to 0, \qquad \text{with probability one.} \tag{9.16}$$

We then find that

$$\begin{aligned}&\mathbb{P}\left\{w_{x_N}(2\delta) \geq \epsilon/2\right\} = \mathbb{P}\left\{\sup_{|s-t|<2\delta} |x_N(s) - x_N(t)| \geq \epsilon/2\right\} \\ &= \mathbb{P}\left\{\sup_{|s-t|<2\delta} |\widetilde{x}_N(s) - \widetilde{x}_N(t)| \geq \epsilon/2\right\} \\ &\leq \mathbb{P}\left\{\sup_{|s-t|<2\delta} |\widetilde{x}(s) - \widetilde{x}(t)| \geq \epsilon/4\right\} \\ &\quad + \mathbb{P}\left\{\sup_{s\in[0,1]} |\widetilde{x}_N(s) - \widetilde{x}(s)| \geq \epsilon/8\right\} + \mathbb{P}\left\{\sup_{t\in[0,1]} |\widetilde{x}_N(t) - \widetilde{x}(t)| \geq \epsilon/8\right\}.\end{aligned}$$

The latter two probabilities tend to zero due to to (9.16). For the first probability on the right hand side, note that $C[0,1]$ is separable and complete. This means that each random element in $C[0,1]$ is tight. Hence, $\widetilde{x} \in C[0,1]$ is tight, so that according to Theorem 7.3 in [Bil99], there exists a $0 < \delta < 1/2$, such that

$$\mathbb{P}\left\{\sup_{|s-t|<2\delta} |x(s) - x(t)| \geq \epsilon/4\right\} = \mathbb{P}\left\{w_x(2\delta) \geq \epsilon/4\right\} \leq \eta.$$

We conclude that $\mathbb{P}\left\{w_{x_N}(2\delta) \geq \epsilon/2\right\} \to 0$, and the same result for y_N can be obtained similarly. This proves the lemma. □