

# Uniform convergence of the empirical cumulative distribution function under informative selection from a finite population

DANIEL BONNÉRY<sup>1,\*,\*\*</sup>, F. JAY BREIDT<sup>2</sup> and FRANÇOIS COQUET<sup>1,†,‡</sup>

<sup>1</sup>Ensaï, Campus de Ker-Lann, Rue Blaise Pascal – BP 37203, 35172 Bruz – cedex, France.

E-mail: \*daniel.bonnery@ensai.fr; †fcoquet@ensai.fr, url: \*\*www.ensai.com/daniel-bonnery-rub,

‡www.ensai.com/francois-coquet-rub

<sup>2</sup>Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877, USA.

E-mail: jbreidt@stat.colostate.edu, url: www.stat.colostate.edu/~jbreidt

Consider informative selection of a sample from a finite population. Responses are realized as independent and identically distributed (i.i.d.) random variables with a probability density function (p.d.f.)  $f$ , referred to as the superpopulation model. The selection is informative in the sense that the sample responses, given that they were selected, are not i.i.d.  $f$ . In general, the informative selection mechanism may induce dependence among the selected observations. The impact of such dependence on the empirical cumulative distribution function (c.d.f.) is studied. An asymptotic framework and weak conditions on the informative selection mechanism are developed under which the (unweighted) empirical c.d.f. converges uniformly, in  $L_2$  and almost surely, to a weighted version of the superpopulation c.d.f. This yields an analogue of the Glivenko–Cantelli theorem. A series of examples, motivated by real problems in surveys and other observational studies, shows that the conditions are verifiable for specified designs.

*Keywords:* complex survey; cut-off sampling; endogenous stratification; Glivenko–Cantelli; length-biased sampling; superpopulation

## 1. Introduction

Consider informative selection of a sample from a finite population, with responses  $Y$  realized as independent and identically distributed (i.i.d.) random variables with probability density function (p.d.f.)  $f$ , referred to as the superpopulation model. (Regression problems, in which observations are conditionally independent given covariates, are also of interest, but the following discussion readily generalizes to that setting and we restrict attention to the i.i.d. case for simplicity of exposition.) In non-informative selection (e.g., Cassel *et al.* [6], Section 1.4, or Särndal *et al.* [34], Remark 2.4.4), the probability of drawing the sample does not depend explicitly on the responses  $Y$ . We consider informative selection in the sense that the sample responses, given that they were selected, are not i.i.d.  $f$ . A specification of informative selection that includes the i.i.d. case described here is given in Pfeiffermann and Sverchkov [30], Remark 1.2. We study the implications of this informative selection for estimation of the superpopulation model.

In general, the informative selection mechanism may induce dependence among the selected observations. Nevertheless, a large body of current methodological literature treats the observations as if they were independently distributed according to the *sample p.d.f.*, defined as the conditional distribution of the random variable  $Y$ , given that it was selected. Under informative selection, the sample p.d.f. differs from  $f$ . In particular, Pfeffermann *et al.* [26] (see some motivating work in Skinner [37]) have developed a *sample likelihood* approach to estimation and inference for the superpopulation model, which maximizes the criterion function formed by taking the product of the sample p.d.f.'s, as if the responses were i.i.d. This methodology has been extended in a number of directions Eideh and Nathan [11–13], Pfeffermann *et al.* [27], Pfeffermann and Sverchkov [28,29,31]. An extensive review of these and other approaches to inference under informative selection is given by Pfeffermann and Sverchkov [30].

Under a strong set of assumptions (in particular, sample size remains fixed as population size goes to infinity), Pfeffermann *et al.* [26] have established the pointwise convergence of the joint distribution of the responses to the product of the sample p.d.f.'s. This is taken as partial justification of the sample likelihood approach. Alternatively, the full likelihood of the data (selection indicators for the finite population and response variables and inclusion probabilities for the sample) can be maximized (Breckling *et al.* [3], Chambers *et al.* [7]), or the *pseudo-likelihood* can be obtained by plugging in Horvitz–Thompson estimators for unknown quantities in the log-likelihood for the entire finite population (e.g., Binder [2], Chambers and Skinner [8], Kish and Frankel [19], Section 2.4). Obviously, each of these likelihood-based approaches requires a model specification.

Rather than starting at the point of likelihood-based inferential methods for the superpopulation model, we take a step back and consider the problem of identifying a suitable model using observed data. In an ordinary inference problem with i.i.d. observations, we often begin not by constructing a likelihood and conducting inference, but by using basic sample statistics to help identify a suitable model. In particular, under i.i.d. sampling the empirical cumulative distribution function (c.d.f.) converges uniformly almost surely to the population c.d.f., by the Glivenko–Cantelli theorem (e.g., van der Vaart [39], Theorem 19.1). What is the behavior of the empirical c.d.f. under informative selection from a finite population? In this paper, we develop an asymptotic framework and weak conditions on the informative selection mechanism under which the (unweighted) empirical c.d.f. converges uniformly, in  $L_2$  and almost surely, to a weighted version of the superpopulation c.d.f. The corresponding quantiles also converge uniformly on compact sets. Our almost sure results rely on an embedding argument. Importantly, our construction preserves the original response vector for the finite population, not some independent replicate.

The conditions we propose are verifiable for specified designs, and involve computing conditional versions of first and second-order inclusion probabilities. Motivated by real problems in surveys and other observational studies, we give examples of where these conditions hold and where they fail. Where the conditions hold, the convergence results we obtain may be useful in making inference about the superpopulation model. For example, the results may be used in identifying a suitable parametric family for the weighted c.d.f., from which a selection mechanism and a superpopulation p.d.f. may be postulated using results in Pfeffermann *et al.* [26].

## 2. Results

### 2.1. Asymptotic framework and assumptions

In what follows, all random variables are defined on a common probability space  $(\Omega, \mathcal{A}, P)$ . Let  $\mathcal{B}(\mathbb{R})$  denote the  $\sigma$ -field of Borel sets. Assume that for  $k \in \mathbb{N}$ ,  $Y_k : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  are i.i.d. real random variables with a density  $f$  with respect to  $\lambda$ , the Lebesgue measure. Consider  $\{N_\gamma\}_{\gamma \in \mathbb{N}}$ , an increasing sequence of positive integers representing a sequence of population sizes, with  $\lim_{\gamma \rightarrow \infty} N_\gamma = \infty$ .

We consider a sequence of finite populations and samples. The  $\gamma$ th finite population is the set of elements indexed by  $U_\gamma = (1, \dots, N_\gamma)$ . In the sampling literature (e.g., Särndal *et al.* [34]),  $U_\gamma$  is often an unordered set, but it is convenient for us to order it and to write, for example,  $\sum_{k \in U_\gamma} = \sum_{k=1}^{N_\gamma}$ . The vector of responses for the population is  $\mathcal{Y}_\gamma = (Y_k)_{k \in U_\gamma}$  and the sample is indexed by the random vector  $\mathcal{I}_\gamma = (I_{\gamma k})_{k \in U_\gamma}$ , where the  $k$ th coordinate  $I_{\gamma k}$  indicates the number of times element  $k$  is selected: 0 or 1 under without-replacement sampling, or a non-negative integer under with-replacement sampling. Define the distribution of  $\mathcal{I}_\gamma$  conditional on  $\mathcal{Y}_\gamma$ :

$$g_\gamma(i_1, \dots, i_{N_\gamma}, y_1, \dots, y_{N_\gamma}) = P(\mathcal{I}_\gamma = (i_1, \dots, i_{N_\gamma}) | \mathcal{Y}_\gamma = (y_1, \dots, y_{N_\gamma})).$$

We assume that the index of the element  $k$  of the population plays no role in the way elements are selected. Specifically, let  $\sigma$  denote a permutation of a vector of length  $N_\gamma$ . Then, for all  $\gamma \in \mathbb{N}$ ,  $(\mathcal{I}_\gamma | \mathcal{Y}_\gamma)$  and  $(\sigma \cdot \mathcal{I}_\gamma | \sigma \cdot \mathcal{Y}_\gamma)$  are identically distributed, or equivalently

$$g_\gamma(i_1, \dots, i_{N_\gamma}, y_1, \dots, y_{N_\gamma}) = g_\gamma(\sigma \cdot (i_1, \dots, i_{N_\gamma}), \sigma \cdot (y_1, \dots, y_{N_\gamma})). \quad (1)$$

We refer to (1) as the exchangeability assumption. It corresponds to the condition of weakly exchangeable arrays (Eagleson and Weber [10]) applied to  $(I_{\gamma k}, Y_k)_{\gamma \in \mathbb{N}, k \in U_\gamma}$ .

**Definition 1.** For  $\gamma \in \mathbb{N}$ , the empirical c.d.f. is the random process  $F_\gamma : \mathbb{R} \rightarrow [0, 1]$  via

$$F_\gamma(\alpha) = \frac{\sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(Y_k) I_{\gamma k}}{\mathbb{1}_{\mathcal{I}_\gamma=0} + \sum_{k \in U_\gamma} I_{\gamma k}}.$$

**Definition 2.** Given  $\gamma$ , let  $k, \ell \in U_\gamma$  with  $k \neq \ell$ . Assume exchangeability as in (1) and let

$$\begin{aligned} m_\gamma(y) &= E[I_{\gamma k} | Y_k = y], \\ v_\gamma(y) &= \text{Var}(I_{\gamma k} | Y_k = y), \\ m'_\gamma(y_1, y_2) &= E[I_{\gamma k} | Y_k = y_1, Y_\ell = y_2], \\ c_\gamma(y_1, y_2) &= \text{Cov}(I_{\gamma k}, I_{\gamma \ell} | Y_k = y_1, Y_\ell = y_2). \end{aligned}$$

(These definitions do not depend on the choice of  $k, \ell$  under the exchangeability assumption).

The following conditions on  $m_\gamma$  are used in defining the limit c.d.f.:

**A0.** There exist  $M : \mathbb{R} \rightarrow \mathbb{R}^+$  and  $m : \mathbb{R} \rightarrow \mathbb{R}^+$ , both  $\lambda$ -measurable, such that

$$\begin{cases} \forall \gamma \in \mathbb{N}, & m_\gamma < M, \\ \int Mf \, d\lambda < \infty, \end{cases} \tag{0a}$$

$$\begin{cases} m_\gamma \rightarrow m & \text{pointwise as } \gamma \rightarrow \infty, \\ \int mf \, d\lambda > 0. \end{cases} \tag{0b}$$

**Definition 3.** Under A0, the limit c.d.f.  $F_s : \mathbb{R} \rightarrow [0, 1]$  is

$$F_s(\alpha) = \frac{\int \mathbb{1}_{(-\infty, \alpha]} mf \, d\lambda}{\int mf \, d\lambda}.$$

**Remark (relation to sample p.d.f.).** Because of informative selection, the empirical c.d.f. does not converge to the superpopulation c.d.f. Under some conditions to be specified below, it converges to  $F_s$ , a weighted integral of the superpopulation p.d.f. To see this, consider the case of without-replacement sampling and a single element,  $k$ . The sample p.d.f. defined in Krieger and Pfeffermann [20] is the conditional density of  $Y_k$  given  $I_{\gamma k} = 1$ . By Bayes' rule,

$$\begin{aligned} f_{s\gamma}(y) &= f(y|I_{\gamma k} = 1) = \frac{P(I_{\gamma k} = 1|Y_k = y)f(y)}{\int P(I_{\gamma k} = 1|Y_k = y)f \, d\lambda} \\ &= \frac{m_\gamma(y)}{\int m_\gamma f \, d\lambda} f(y) = w_\gamma(y) f(y). \end{aligned}$$

Define  $w = \lim_{\gamma \rightarrow \infty} w_\gamma$  and consider  $\alpha \in \mathbb{R}$ . Then

$$\lim_{\gamma \rightarrow \infty} \int \mathbb{1}_{(-\infty, \alpha]} f_{s\gamma} \, d\lambda = \lim_{\gamma \rightarrow \infty} \int \mathbb{1}_{(-\infty, \alpha]} w_\gamma f \, d\lambda = \int \mathbb{1}_{(-\infty, \alpha]} wf \, d\lambda = F_s(\alpha).$$

Thus, if observations were i.i.d. from the sample p.d.f.,  $F_s$  would be the natural limiting c.d.f. A related argument can be used to show that the same weighted c.d.f. is obtained under with-replacement sampling and a fixed number of draws, when considering the distribution of any observation in the sample.

Because informative selection from a finite population may induce dependence among the selected observations, observations are not i.i.d., and we next specify asymptotic weak dependence conditions among  $\mathcal{I}_\gamma$  coordinates.

For a sequence  $\{b_\gamma\}$ , let  $o_\gamma(b_\gamma)$  denote  $\lim_{\gamma \rightarrow \infty} o_\gamma(b_\gamma)b_\gamma^{-1} = 0$ . In the next two assumptions, we define sufficient conditions for uniform  $L_2$  convergence and uniform a.s. convergence of the empirical c.d.f.

**A1 (Uniform  $L_2$  convergence conditions).**

$$\int c_\gamma(y_1, y_2) f(y_1) f(y_2) dy_1 dy_2 = o_\gamma(1), \tag{1a}$$

$$\int (m'_\gamma(y_1, y_2) m'_\gamma(y_2, y_1) - m_\gamma(y_1) m_\gamma(y_2)) f(y_1) f(y_2) dy_1 dy_2 = o_\gamma(1), \tag{1b}$$

$$\int (v_\gamma + m_\gamma^2) f d\lambda = o_\gamma(N_\gamma), \tag{1c}$$

$$P(\mathcal{I}_\gamma = (0, \dots, 0)) = o_\gamma(1). \tag{1d}$$

**A2 (Uniform almost sure convergence conditions).** Let  $y \in \mathbb{R}^N$  satisfy

$$\sup_{\alpha' \in \mathbb{R}} \left| \frac{\sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha']}(y_k)}{N_\gamma} - \int \mathbb{1}_{(-\infty, \alpha']} f d\lambda \right| = o_\gamma(1).$$

Then for all  $\alpha \in \mathbb{R}$ ,

$$\text{Var} \left( \sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(y_k) I_{\gamma k} \mid \mathcal{Y}_\gamma = (y_1, \dots, y_{N_\gamma}) \right) = o_\gamma(N_\gamma^2), \tag{2a}$$

$$\sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(y_k) (E[I_{\gamma k} \mid \mathcal{Y}_\gamma = (y_1, \dots, y_{N_\gamma})] - m_\gamma(y_k)) = o_\gamma(N_\gamma), \tag{2b}$$

$$g_\gamma((0, \dots, 0), y) = o_\gamma(1). \tag{2c}$$

*Properties of sampling without replacement*

In the case of sampling without replacement,  $\mathcal{I}_\gamma : \Omega \rightarrow \{0, 1\}^{N_\gamma}$ , A0 and A1 can be replaced by a simpler set of sufficient conditions for uniform  $L_2$  convergence.

**A3 (Uniform  $L_2$  convergence conditions under sampling without replacement).**

$$\exists m : \mathbb{R} \rightarrow \mathbb{R}^+ \text{ } \lambda\text{-measurable s.t.} \quad \begin{cases} m_\gamma \rightarrow m & \text{pointwise as } \gamma \rightarrow \infty, \\ \int m f d\lambda > 0, \end{cases} \tag{3a}$$

$$\forall y_1, y_2, \quad c_\gamma(y_1, y_2) = o_\gamma(1), \tag{3b}$$

$$\forall y_1, y_2, \quad m'_\gamma(y_1, y_2) - m_\gamma(y_2) = o_\gamma(1), \tag{3c}$$

$$P(\mathcal{I}_\gamma = (0, \dots, 0)) = o_\gamma(1). \tag{3d}$$

These conditions imply A0 and A1.

**Proof.** Since  $I_{\gamma k} \in \{0, 1\}$ , (0a) and (1c) always hold. By applying the Lebesgue dominated convergence theorem, we obtain that (1a) is verified when  $\forall y_1, y_2, c_\gamma(y_1, y_2) = o_\gamma(1)$  and (1b) is verified when  $\forall y_1, y_2, m'_\gamma(y_1, y_2) - m_\gamma(y_2) = o_\gamma(1)$ . □

An important special case of sampling without replacement is non-informative selection, with  $\mathcal{I}_\gamma$  independent of  $\mathcal{Y}_\gamma$  for all  $\gamma \in \mathbb{N}$ . In this case, the sample obtained is an i.i.d. sample of size  $n_\gamma = \sum_{k \in U_\gamma} I_{\gamma k}$  (Fuller [14], Theorem 1.3.1), and the classic Glivenko–Cantelli theorem can be applied as soon as  $n_\gamma \xrightarrow{\text{a.s.}} \infty$  as  $\gamma \rightarrow \infty$ . The assumptions of Theorem 1 and Theorem 2 will then just ensure that the expectation of the sample size will grow to infinity, and that its variations are small enough to avoid very small samples. We can thus replace A0–A2 by a simpler set of sufficient conditions.

**A4 (Uniform  $L_2$  and a.s. convergence conditions under independent sampling without replacement).**

$$\begin{cases} N_\gamma^{-1} \mathbb{E}[n_\gamma] \rightarrow m \neq 0 & \text{as } \gamma \rightarrow \infty, \\ \text{Var}(n_\gamma) = o_\gamma(N_\gamma^2). \end{cases} \tag{4}$$

*These conditions imply A0–A2.*

**Proof.** We first show that A4 implies A3. Because  $\mathcal{I}_\gamma$  and  $\mathcal{Y}_\gamma$  are independent, the exchangeability assumption implies  $m_\gamma(y) = \mathbb{E}[I_{\gamma 1}] = N_\gamma^{-1} \mathbb{E}[n_\gamma]$  and  $N_\gamma^{-1} \mathbb{E}[n_\gamma] \rightarrow m$  by A4, so (3a) holds. Exchangeability also implies

$$\mathbb{E}[I_{\gamma 1} I_{\gamma 2}] = \frac{\sum_{k, \ell \in U_\gamma : k \neq \ell} \mathbb{E}[I_{\gamma k} I_{\gamma \ell}]}{N_\gamma(N_\gamma - 1)} = \mathbb{E}\left[\frac{\sum_{k, \ell \in U_\gamma : k \neq \ell} I_{\gamma k} I_{\gamma \ell}}{N_\gamma(N_\gamma - 1)}\right] = \mathbb{E}\left[\frac{n_\gamma(n_\gamma - 1)}{N_\gamma(N_\gamma - 1)}\right]$$

so

$$c_\gamma(y_1, y_2) = \text{Cov}(I_{\gamma 1}, I_{\gamma 2}) = \mathbb{E}\left[\frac{n_\gamma(n_\gamma - N_\gamma)}{N_\gamma^2(N_\gamma - 1)}\right] + \text{Var}\left(\frac{n_\gamma}{N_\gamma}\right) = o_\gamma(1) \tag{5}$$

by A4, so (3b) is obtained, and (3c) holds by independence. Finally,

$$\begin{aligned} \mathbb{P}(n_\gamma = 0) &= \mathbb{P}(n_\gamma < 1) = \mathbb{P}(n_\gamma - \mathbb{E}[n_\gamma] < 1 - \mathbb{E}[n_\gamma]) \\ &\leq \mathbb{P}(|n_\gamma - \mathbb{E}[n_\gamma]| > \mathbb{E}[n_\gamma] - 1) \leq \frac{\text{Var}(n_\gamma)}{(\mathbb{E}[n_\gamma] - 1)^2} = o_\gamma(1), \end{aligned} \tag{6}$$

establishing (3d).

We next show that (4) implies A2. For all  $\alpha \in \mathbb{R}$ ,

$$\begin{aligned} &\text{Var}\left(\sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(Y_k) I_{\gamma k} \mid \mathcal{Y}_\gamma = (y_1, \dots, y_{N_\gamma})\right) \\ &= \sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(y_k) \text{Var}(I_{\gamma k}) \\ &\quad + \sum_{k, \ell \in U_\gamma : k \neq \ell} \mathbb{1}_{(-\infty, \alpha]}(y_k) \mathbb{1}_{(-\infty, \alpha]}(y_\ell) \text{Cov}(I_{\gamma k}, I_{\gamma \ell}) \\ &\leq N_\gamma + N_\gamma(N_\gamma - 1) o_\gamma(1) = o_\gamma(N_\gamma^2) \end{aligned}$$

by equation (5), so (2a) holds. By independence,

$$E[I_{\gamma k} | \mathcal{Y}_\gamma = (y_1, \dots, y_{N_\gamma})] = E[I_{\gamma k} | Y_k = y_k] = m_\gamma(y_k),$$

so (2b) holds. Finally,

$$g_\gamma((0, \dots, 0), y) = P(n_\gamma = 0) = o_\gamma(1)$$

by independence and (6), so (2c) holds. □

**Remark.** In conventional finite population asymptotics (Breidt and Opsomer [4,5], Isaki and Fuller [17], Robinson and Särndal [33]), conditions on design covariances  $\text{Cov}(I_{\gamma k}, I_{\gamma \ell})$  are imposed to guarantee that the Horvitz–Thompson estimator  $\sum_{k \in U_\gamma} y_k I_{\gamma k} (E[I_{\gamma k}])^{-1}$  is consistent. Typically, these conditions imply that the variance of the Horvitz–Thompson estimator is  $O_\gamma(N_\gamma^2 / (N_\gamma \pi_{*\gamma}))$ , where  $N_\gamma \pi_{*\gamma} \rightarrow \infty$  is a sequence of lower bounds on the expected sample size,  $E[n_\gamma]$ . These same conditions can be used to show that  $\text{Var}(n_\gamma) = O_\gamma(N_\gamma^2 / (N_\gamma \pi_{*\gamma})) = o_\gamma(N_\gamma^2)$ , agreeing with A4.

## 2.2. Uniform convergence of the empirical c.d.f.

In this section, we state the main results of the paper: uniform  $L_2$  convergence of the empirical c.d.f. and uniform almost sure convergence of the empirical c.d.f. Important corollaries yield uniform convergence of sample quantiles on compact sets. Proofs are given in the [Appendix](#).

### 2.2.1. Uniform $L_2$ convergence of the empirical c.d.f.

**Theorem 1.** Under A0 and A1, the empirical c.d.f. converges uniformly in  $L_2$  in the sense that

$$\sup_{\alpha \in \mathbb{R}} |F_\gamma(\alpha) - F_s(\alpha)| = \|F_\gamma - F_s\|_\infty \xrightarrow[\gamma \rightarrow \infty]{L_2} 0.$$

**Definition 4.** The limit quantiles  $\xi_s : (0, 1) \rightarrow \mathbb{R}$  are given by

$$\xi_s(p) = \inf\{y \in \mathbb{R} : F_s(y) \geq p\}$$

and the empirical quantiles  $\xi_\gamma : (0, 1) \rightarrow \mathbb{R}$  are given by

$$\xi_\gamma(p) = \inf\{y \in \mathbb{R} : F_\gamma(y) \geq p\}.$$

With this definition, we have the following corollary.

**Corollary 1.** Suppose that  $F_s$  is continuous on  $\mathbb{R}$  and  $0 < F_s(y_1) = F_s(y_2) < 1 \Rightarrow y_1 = y_2$ . Then, under A0 and A1, the empirical quantiles converge uniformly in probability to the limit quantiles,

$$\sup_{p \in K} |\xi_\gamma(p) - \xi_s(p)| \xrightarrow[\gamma \rightarrow \infty]{P} 0$$

for all  $K$  a compact subset of  $(0, 1)$ . Under the further hypothesis that  $f$  has compact support, the convergence is uniform in  $L_2$ :

$$\sup_{p \in K} |\xi_\gamma(p) - \xi_s(p)| \xrightarrow[\gamma \rightarrow \infty]{L_2} 0.$$

2.2.2. Uniform almost sure convergence of the empirical c.d.f.

The Glivenko–Cantelli theorem gives uniform almost sure convergence of the empirical c.d.f. under i.i.d. sampling. We now consider uniform almost sure convergence under dependent sampling satisfying the second-order conditions of A2.

Asymptotic arguments in survey sampling consist first in embedding a specific sample scheme in a sequence of sample schemes. In the proof of the following representation theorem, we link the elements of the sequence of sample schemes in a way that ensures uniform almost sure convergence of the empirical c.d.f. We stress that in our result the vector of responses for the population remains the original  $\mathcal{Y}_\gamma = (Y_k)_{k \in U_\gamma}$ , and not another set of identically distributed random variables.

**Theorem 2.** Under A0 and A2, there exist sequences of random variables  $(I'_{\gamma k})_{\gamma \in \mathbb{N}, k \in U_\gamma}$ ,  $(Y'_k)_{k \in \mathbb{N}}$  defined on the probability space  $(\Omega \times [0, 1], \mathcal{A} \otimes \mathcal{B}_{[0,1]}, \mathbf{P}' = \mathbf{P} \otimes \lambda_{[0,1]})$  such that

- $\|F'_\gamma - F_s\|_\infty$  converges  $\mathbf{P}'$ -a.s. to 0
- $\forall \gamma \in \mathbb{N}$ ,  $(\mathcal{I}'_\gamma, \mathcal{Y}'_\gamma)$  and  $(\mathcal{I}_\gamma, \mathcal{Y}_\gamma)$  have the same law
- $\forall \gamma \in \mathbb{N}, \omega \in \Omega, x \in [0, 1], \mathcal{Y}'_\gamma(\omega, x) = \mathcal{Y}_\gamma(\omega)$ ,

where  $\mathcal{B}_{[0,1]}$  is the  $\sigma$ -field of Borel sets,  $\lambda_{[0,1]}$  is the Lebesgue measure on  $[0, 1]$ ,  $\mathcal{I}'_\gamma = (I'_{\gamma 1}, \dots, I'_{\gamma N_\gamma})$ ,  $\mathcal{Y}'_\gamma = (Y'_1, \dots, Y'_{N_\gamma})$  and  $F'_\gamma : \mathbb{R} \rightarrow [0, 1]$  via

$$F'_\gamma(\alpha) = \frac{\sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(Y'_{\gamma k}) I'_{\gamma k}}{\sum_{k \in U_\gamma} I'_{\gamma k} + \mathbb{1}_{I'_\gamma = 0}}. \tag{7}$$

**Corollary 2.** Suppose that  $F_s$  is continuous and  $0 < F_s(y_1) = F_s(y_2) < 1 \Rightarrow y_1 = y_2$ . If A0 and A2 hold, then for  $(I'_{\gamma k})_{\gamma \in \mathbb{N}, k \in U_\gamma}$  and  $(Y'_k)_{k \in \mathbb{N}}$  that satisfy the conditions of Theorem 2, the empirical quantiles

$$\xi'_\gamma(p) = \inf\{y \in \mathbb{R}: F'_\gamma(y) \geq p\}$$

converge uniformly almost surely,

$$\sup_{p \in K} |\xi'_\gamma(p) - \xi_s(p)| \xrightarrow[\gamma \rightarrow \infty]{a.s.} 0$$

for all  $K$  a compact subset of  $(0, 1)$ .

### 3. Examples

We now consider a series of examples of selection mechanisms, motivated by real problems in surveys and other observational studies. We give examples where conditions A0, A1, A2 hold and where they fail.

#### 3.1. Non-informative selection without replacement

- For any sequence of *fixed-size without-replacement designs* with  $\mathcal{I}_\gamma$  independent of  $\mathcal{Y}_\gamma$  (e.g., simple random sampling, stratified sampling with stratification variables independent of  $\mathcal{Y}_\gamma$ , rejective sampling (Hájek [15]) with inclusion probabilities independent of  $\mathcal{Y}_\gamma$ , etc.), the condition A4 holds provided that  $n_\gamma N_\gamma^{-1}$  converges to a strictly positive sampling rate.
- For a sequence of *Bernoulli samples* with parameter  $p \in (0, 1)$ , the  $\{I_{\gamma k}\}$  are i.i.d. Bernoulli( $p$ ) random variables, so  $\text{Var}(n_\gamma) = N_\gamma p(1 - p)$  and condition A4 holds.
- *Poisson sampling* corresponds to a design in which, given a random vector  $(\Pi_{\gamma 1}, \dots, \Pi_{\gamma N_\gamma}) : \Omega \rightarrow (0, 1]^{N_\gamma}$ , the  $\{I_{\gamma k}\}$  are a sequence of independent Bernoulli( $\Pi_{\gamma k}$ ) random variables (Poisson [32]). In this case, the variance of  $n_\gamma$  is given by

$$\text{Var}(n_\gamma) = \sum_{k \in U_\gamma} \text{E}[\Pi_{\gamma k}(1 - \Pi_{\gamma k})] + \text{Var}\left(\sum_{k \in U_\gamma} \Pi_{\gamma k}\right).$$

Note that the first term in this expression is always  $o_\gamma(N_\gamma^2)$ , so it suffices to consider the second.

- In the case where the vector  $[\Pi_{\gamma k}]_{k \in U_\gamma}$  is just a random permutation of a non-random vector  $[\pi_{\gamma k}]_{k \in U_\gamma}$ , then  $\text{Var}(\sum_{k \in U_\gamma} \Pi_{\gamma k}) = \text{Var}(\sum_{k \in U_\gamma} \pi_{\gamma k}) = 0$  and A4 is satisfied when  $N_\gamma^{-1} \sum_{k \in U_\gamma} \pi_{\gamma k}$  converges to a non-zero constant.
- Suppose that  $Z_\gamma$  is a random positive real vector of size  $N_\gamma$ , and suppose that the law of  $(Z_\gamma, \mathcal{Y}_\gamma)$  is invariant under any permutation of the coordinates. For  $n_\gamma^*$  fixed, consider the design in which  $\Pi_{\gamma k} = n_\gamma^* Z_{\gamma k} \{\sum_{k \in U_\gamma} Z_{\gamma k}\}^{-1}$ . Then

$$\text{Var}\left(\sum_{k \in U_\gamma} \Pi_{\gamma k}\right) = \text{Var}(n_\gamma^*) = 0$$

and A4 is satisfied when  $Z_\gamma$  and  $\mathcal{Y}_\gamma$  are independent and  $N_\gamma^{-1} n_\gamma^*$  converges to a non-zero constant.

- Let  $a_\gamma, b_\gamma \in (0, 1]$  with  $a_\gamma \neq b_\gamma$ . If

$$(\Pi_{\gamma 1}, \dots, \Pi_{\gamma N_\gamma}) \equiv \begin{cases} (a_\gamma, \dots, a_\gamma), & \text{with probability } 1/2, \\ (b_\gamma, \dots, b_\gamma), & \text{with probability } 1/2, \end{cases}$$

then

$$\text{Var}\left(\sum_{k \in U_\gamma} \Pi_{\gamma k}\right) = N_\gamma^2 \frac{(a_\gamma - b_\gamma)^2}{4} \neq o_\gamma(N_\gamma^2).$$

Then A4 is not verified and in fact if  $N_\gamma a_\gamma = o_\gamma(1)$  we do not have uniform convergence of the empirical c.d.f.

### 3.2. Length-biased sampling

Length-biased sampling, in which  $P(I_{\gamma k} = 1|Y_k = y_k) = m_\gamma(y_k) \propto y_k$ , is pervasive in real surveys and observational studies. Cox [9] gives a now-classic example of sampling fibers in textile manufacture, in which  $m_\gamma(y_k) \propto y_k = \text{fiber length}$ . In surveys of wildlife abundance, “visibility bias” means that larger individuals or groups are more noticeable (e.g., Patil and Rao [25]), so  $m_\gamma(y_k) \propto y_k = \text{size of individual or group}$ . “On-site surveys” are sometimes used to study people engaged in some activity like shopping in a mall (Nowell and Stanley [24]) or fishing at the seashore (Sullivan *et al.* [38]); the longer they spend doing the activity, the more likely the field staff are to intercept and interview them, so  $m_\gamma(y_k) \propto y_k = \text{activity time}$ . In mark-recapture surveys of wildlife populations, individuals that live longer are more likely to be recaptured, so  $m_\gamma(y_k) \propto y_k = \text{lifetime}$  (e.g., Leigh [22]). Similarly, in epidemiological studies of latent diseases, individuals who become symptomatic seek treatment and drop out of eligibility for sampling, while those with long latency periods are more likely to be sampled:  $m_\gamma(y_k) \propto y_k = \text{latency period}$ . Finally, propensity to respond to a survey is often related to a variable of interest; for example, higher response rates from higher-income individuals.

Suppose that  $f$  has compact, positive support:  $\int \mathbb{1}_{[\epsilon, M]} f \, d\lambda = 1$  for some  $0 < \epsilon < M < \infty$ . For the  $\gamma$ th finite population, consider Poisson sampling with inclusion probability proportional to  $Y$ , in the sense that  $\{I_{\gamma k}\}_{k \in U_\gamma}$  are independent binary random variables, with

$$P(I_{\gamma k} = 1|Y_k = y_k) = 1 - P(I_{\gamma k} = 0|Y_k = y_k) = m_\gamma(y_k) \propto y_k.$$

Let  $\tau_\gamma = y_k^{-1}P(I_{\gamma k} = 1|Y_k = y_k)$  be the common proportionality constant (independent of  $k$ ), and assume that  $\tau_\gamma \rightarrow \tau \in (0, M^{-1}]$  as  $\gamma \rightarrow \infty$ . Then

$$\begin{aligned} m_\gamma(y) &= \tau_\gamma y \quad \rightarrow \quad \tau y = m(y), \\ c_\gamma(y_k, y_\ell) &= 0, \quad m'_\gamma(y_k, y_\ell) - m_\gamma(y_k) = 0, \end{aligned}$$

$$\begin{aligned} P(\mathcal{I}_\gamma = (0, \dots, 0)) &= E \left[ \prod_{k \in U_\gamma} (1 - \tau_\gamma y_k) \right] \\ &\leq (1 - \tau_\gamma \epsilon)^{N_\gamma} = \exp(N_\gamma \ln(1 - \tau_\gamma \epsilon)) = o_\gamma(1), \end{aligned}$$

so that A3 is verified. It then follows that the limiting c.d.f. is given by

$$F_s(\alpha) = \int \mathbb{1}_{(-\infty, \alpha]} \frac{y}{E[Y_1]} f \, d\lambda. \tag{8}$$

### 3.3. Cluster sampling

Let  $F$  denote the superpopulation c.d.f.:  $F(\tau) = \int \mathbb{1}_{(-\infty, \tau]} f \, d\lambda$ . Let  $\tau \in \mathbb{R}$  be such that  $F(\tau) > 0$ . Define  $i_{1\gamma} = (\mathbb{1}_{(-\infty, \tau]}(Y_k))_{k \in U_\gamma}$  and  $i_{2\gamma} = (\mathbb{1}_{(\tau, \infty)}(Y_k))_{k \in U_\gamma}$ . The selection mecha-

nism is  $\mathcal{I}_\gamma = i_{1\gamma}$  or  $i_{2\gamma}$ , each with probability 1/2, so uniform convergence of the empirical c.d.f. is not possible. Note that

$$\begin{aligned} \text{Cov}(I_{\gamma k}, I_{\gamma \ell} | Y_k = y_1, Y_\ell = y_2) &= \frac{1}{2} \mathbb{1}_{(-\infty, \tau]}(y_1) \mathbb{1}_{(-\infty, \tau]}(y_2) \\ &\quad + \frac{1}{2} \mathbb{1}_{(\tau, \infty)}(y_1) \mathbb{1}_{(\tau, \infty)}(y_2) - \frac{1}{4} \end{aligned}$$

so that

$$\int c_\gamma(y_1, y_2) f(y_1) f(y_2) dy_1 dy_2 = \frac{1}{2} F^2(\tau) + \frac{1}{2} (1 - F(\tau))^2 - \frac{1}{4} \neq o_\gamma(1),$$

and (1a) fails to hold. This example can be regarded as a “worst-case” cluster sample: the sample consists of many elements but only one cluster, and the population is made up of a small number of large clusters, none of which is fully representative of the population.

### 3.4. Cut-off sampling and take-all strata

In cut-off sampling a part of the population is excluded from sampling, so that  $I_{\gamma k} = 0$  with probability one for some subset of  $U_\gamma$ . This may be due to physical limitations of the sampling apparatus, like a net that lets small animals escape, or may be due to a deliberate design decision. For example, a statistical agency may be willing to accept the bias inherent in cutting off small  $y$ -values if the  $y$ -distribution is highly skewed, as is often the case in establishment surveys (e.g., Särndal *et al.* [34], Section 14.4).

Consider cut-off sampling with  $I_{\gamma k} = 0$  for  $\{k \in U_\gamma: y_k \leq \tau\}$ , and simple random sampling without replacement of size  $\min\{n_\gamma, N_\gamma - \sum_{j \in U_\gamma} \mathbb{1}_{(-\infty, \tau]}(y_j)\}$  from the remaining population,  $\{j \in U_\gamma: y_j > \tau\}$ .

Define  $Z_k = \mathbb{1}_{(-\infty, \tau]}(Y_k)$  with corresponding realization  $z_k = \mathbb{1}_{(-\infty, \tau]}(y_k)$ . Let  $\rho_\gamma = N_\gamma^{-1} n_\gamma$  and assume that  $\lim_{\gamma \rightarrow \infty} \rho_\gamma = \rho$ . We now verify A3.

Define  $S_{\gamma A} = \sum_{j \in U_\gamma: j \notin A} Z_j$ . By the weak law of large numbers,  $N_\gamma^{-1} S_{\gamma A} \xrightarrow{P} F(\tau)$  as  $\gamma \rightarrow \infty$  for  $A = \{k\}$  or  $A = \{k, \ell\}$ , and so for those sets  $A$  we have

$$\lim_{\gamma \rightarrow \infty} \mathbb{E} \left[ \frac{\rho_\gamma - N_\gamma^{-1} S_{\gamma A}}{1 - N_\gamma^{-1} S_{\gamma A}} \mathbb{1}_{\{\rho_\gamma > N_\gamma^{-1} S_{\gamma A}\}} \right] = \frac{(\rho - F(\tau)) \mathbb{1}_{\{\rho > F(\tau)\}}}{1 - F(\tau)}$$

by the uniform integrability of the integrand. With the same argument,

$$\begin{aligned} \lim_{\gamma \rightarrow \infty} \mathbb{E} \left[ \frac{(n_\gamma - S_{\gamma\{k, \ell\}})(n_\gamma - 1 - S_{\gamma\{k, \ell\}})}{(N_\gamma - S_{\gamma\{k, \ell\}})(N_\gamma - 1 - S_{\gamma\{k, \ell\}})} \mathbb{1}_{\{n_\gamma > S_{\gamma\{k, \ell\}}\}} \right] \\ = \left( \frac{\rho - F(\tau)}{1 - F(\tau)} \right)^2 \mathbb{1}_{\{\rho > F(\tau)\}}. \end{aligned}$$

Using conditional first and second-order inclusion probabilities under simple random sampling, we have

$$\begin{aligned}
 m_\gamma(y_k) &= z_k + (1 - z_k)E\left[\frac{n_\gamma - S_\gamma\{k\}}{N_\gamma - S_\gamma\{k\}}\mathbb{1}_{\{n_\gamma > S_\gamma\{k\}\}}\right] \\
 &\rightarrow z_k + (1 - z_k)\frac{(\rho - F(\tau))\mathbb{1}_{\{\rho > F(\tau)\}}}{1 - F(\tau)}, \\
 m'_\gamma(y_\ell, y_k) &= z_k + (1 - z_\ell)(1 - z_k)E\left[\frac{n_\gamma - S_\gamma\{k, \ell\}}{N_\gamma - S_\gamma\{k, \ell\}}\mathbb{1}_{\{n_\gamma > S_\gamma\{k, \ell\}\}}\right] \\
 &\quad + z_\ell(1 - z_k)E\left[\frac{n_\gamma - 1 - S_\gamma\{k, \ell\}}{N_\gamma - 1 - S_\gamma\{k, \ell\}}\mathbb{1}_{\{n_\gamma - 1 > S_\gamma\{k, \ell\}\}}\mathbb{1}_{\{N_\gamma - 1 > S_\gamma\{k, \ell\}\}}\right] \\
 &\rightarrow z_k + (1 - z_k)\frac{(\rho - F(\tau))\mathbb{1}_{\{\rho > F(\tau)\}}}{1 - F(\tau)}, \\
 d_\gamma(y_k, y_\ell) &= E[I_{\gamma k}I_{\gamma \ell}|Y_k = y_k, Y_\ell = y_\ell] \\
 &= z_k z_\ell + \{z_k(1 - z_\ell) + (1 - z_k)z_\ell\}E\left[\frac{n_\gamma - 1 - S_\gamma\{k, \ell\}}{N_\gamma - 1 - S_\gamma\{k, \ell\}}\mathbb{1}_{\{n_\gamma - 1 > S_\gamma\{k, \ell\}\}}\right] \\
 &\quad + (1 - z_k)(1 - z_\ell)E\left[\frac{(n_\gamma - S_\gamma\{k, \ell\})(n_\gamma - 1 - S_\gamma\{k, \ell\})}{(N_\gamma - S_\gamma\{k, \ell\})(N_\gamma - 1 - S_\gamma\{k, \ell\})}\mathbb{1}_{\{n_\gamma > S_\gamma\{k, \ell\}\}}\right] \\
 &\rightarrow z_k z_\ell + (1 - z_k)(1 - z_\ell)\left(\frac{\rho - F(\tau)}{1 - F(\tau)}\right)^2\mathbb{1}_{\{\rho > F(\tau)\}} \\
 &\quad + \{z_k(1 - z_\ell) + (1 - z_k)z_\ell\}\frac{(\rho - F(\tau))\mathbb{1}_{\{\rho > F(\tau)\}}}{1 - F(\tau)}, \\
 c_\gamma(y_k, y_\ell) &= d_\gamma(y_k, y_\ell) - m'_\gamma(y_k, y_\ell)m'_\gamma(y_\ell, y_k) = o_\gamma(1),
 \end{aligned}$$

and A3 is verified.

Cut-off sampling for  $y_k \leq \tau$  is essentially the complement of stratified sampling with a ‘‘take-all stratum’’:  $I_{\gamma k} = 1$  for the set  $\{k \in U_\gamma : z_k = 1\}$ . Take-all strata are common in practice, particularly for the highly-skewed populations in which cut-off sampling is attractive. Arguments nearly identical to those above can be used to establish A3 in the take-all case. This take-all stratified design is analogous to the well-known class of *case-control studies* in epidemiology. We specifically consider prospective case-control studies (e.g., Mantel [23], Langholz and Goldstein [21], Arratia *et al.* [1]), in which the finite population of all disease cases and controls is stratified, disease cases ( $z_k = 1$ ) are selected with probability one, and controls ( $z_k = 0$ ) are selected with probability less than one.

### 3.5. With-replacement sampling with probability proportional to size

Let  $\{n_\gamma\}$  be a non-random sequence of positive integers with  $n_\gamma < N_\gamma$  and suppose that  $f$  has strictly positive support:  $\int \mathbb{1}_{(-\infty, 0]} f \, d\lambda = 0$ . Consider the case of with-replacement sam-

pling of  $n_\gamma$  draws, with probability of selecting element  $k$  on the  $h$ th draw equal  $p_{\gamma k} \in [0, 1]$ ,  $\sum_{k \in U_\gamma} p_{\gamma k} = 1$ . While  $p_{\gamma k}$  could be constructed in many ways, a case of particular interest is  $p_{\gamma k} \propto Y_k$ . This design is usually not feasible in practice, but statistical agencies often attempt to draw samples with probability proportional to a size measure (p.p.s.) that is highly correlated with  $Y$ . Such a design will be highly efficient for estimation of the  $Y$ -total (indeed, a fixed-size p.p.s. design with probabilities proportional to  $Y_k$  would exactly reproduce the  $Y$ -total).

For  $h = 1, \dots, n_\gamma$ , let  $R_{\gamma h}$  be i.i.d. random variables with

$$P(R_{\gamma h} = k | \mathcal{Y}_\gamma) = \frac{Y_k}{\sum_{j \in U_\gamma} Y_j}.$$

Then  $I_{\gamma k} = \sum_{h=1}^{n_\gamma} \mathbb{1}_{\{R_{\gamma h}=k\}}$  counts the number of draws for which element  $k$  is selected. Define  $W_{\gamma A} = N_\gamma^{-1} \sum_{j \in U_\gamma: j \notin A} Y_j$ . Then

$$\begin{aligned} m_\gamma(y_k) &= \frac{n_\gamma}{N_\gamma} y_k E \left[ \frac{1}{N_\gamma^{-1} y_k + W_{\gamma\{k\}}} \right], \\ m'_\gamma(y_k, y_\ell) &= \frac{n_\gamma}{N_\gamma} y_k E \left[ \frac{1}{N_\gamma^{-1} (y_k + y_\ell) + W_{\gamma\{k,\ell\}}} \right], \\ v_\gamma(y_k) &= \left( \frac{n_\gamma}{N_\gamma} y_k \right)^2 \text{Var} \left( \frac{1}{N_\gamma^{-1} y_k + W_{\gamma\{k\}}} \right) + \frac{n_\gamma}{N_\gamma} \frac{y_k}{N_\gamma} E \left[ \frac{W_{\gamma\{k\}}}{(N_\gamma^{-1} y_k + W_{\gamma\{k\}})^2} \right], \\ c_\gamma(y_k, y_\ell) &= \left( \frac{n_\gamma}{N_\gamma} \right)^2 y_k y_\ell \left\{ -\frac{1}{N_\gamma} E \left[ \frac{1}{(N_\gamma^{-1} (y_k + y_\ell) + W_{\gamma\{k,\ell\}})^2} \right] \right. \\ &\quad \left. + n_\gamma \text{Var} \left( \frac{1}{N_\gamma^{-1} (y_k + y_\ell) + W_{\gamma\{k,\ell\}}} \right) \right\}. \end{aligned}$$

Under mild additional conditions, A1 and A2 can be established using straightforward bounding and limiting arguments. A sufficient set of conditions for either A1 or A2 is  $n_\gamma N_\gamma^{-1} \rightarrow \tau \in [0, 1]$  as  $\gamma \rightarrow \infty$  and  $E[Y_1^6] < \infty$ . Under these conditions,  $m_\gamma(y) = \tau y (E[Y_1])^{-1} + o_\gamma(1)$ , and the limiting c.d.f. is the same as in length-biased sampling, as given by equation (8).

### 3.6. Endogenous stratification

Endogenous stratification, in which the sample is effectively stratified on the value of the dependent variable, is common in the health and social sciences (e.g., Hausman and Wise [16], Jewell [18], Shaw [36]). Often, it arises by design when a screening sample is selected, the dependent variable is observed, and then covariates are measured for a sub-sample that is stratified on the dependent variable: for example, undersampling the high-income stratum (Hausman and Wise [16]). It can also arise through uncontrolled selection effects, in much the same way as length-biased sampling. One such example comes from fisheries surveys, in which a field interviewer is stationed at a dock for a fixed length of time, and intercepts recreational fishing boats

as they return to the dock. The interviewer tends to select high-catch boats and, while busy measuring the fish caught on those boats, misses more of the low-catch boats. Thus, sampling effort is endogenously stratified on catch (Sullivan *et al.* [38]).

We consider a sample endogenously stratified on the order statistics of  $Y$ . Let  $\{H_\gamma\}$  be a non-random sequence of positive integers, which may remain bounded or go to infinity. For each  $\gamma$ , let  $\{N_{\gamma h}\}_{h=1}^{H_\gamma}$  be a set of non-random positive integers with  $\sum_{h=1}^{H_\gamma} N_{\gamma h} = N_\gamma$ , and let  $\{n_{\gamma h}\}_{h=1}^{H_\gamma}$  be a set of non-random positive integers with  $n_{\gamma h} \leq N_{\gamma h}$ . Let

$$Y_{(1)} < Y_{(2)} < \dots < Y_{(N_\gamma)}$$

denote the order statistics for the  $\gamma$ th population, which is stratified by taking the first  $N_{\gamma 1}$  values as stratum 1, the next  $N_{\gamma 2}$  as stratum 2, etc., with the last  $N_{\gamma H_\gamma}$  values constituting stratum  $H_\gamma$ . The  $\gamma$ th sample is then a stratified simple random sample without replacement of size  $n_{\gamma h}$  from the  $N_{\gamma h}$  elements in stratum  $h$ .

Define  $M_{\gamma 0} = 0$  and  $M_{\gamma h} = \sum_{g=1}^h N_{\gamma g}$ . Because  $H_\gamma$ ,  $N_\gamma$  and  $n_\gamma$  are not random, we then have

$$\begin{aligned} m_\gamma(y) &= \sum_{h=1}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} \mathbb{P}(Y_{(M_{\gamma, h-1})} < Y_k \leq Y_{(M_{\gamma h})} | Y_k = y) \\ &= \sum_{h=1}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} \mathbb{P}\left(\frac{M_{\gamma, h-1}}{N_\gamma - 1} < F_{N_\gamma - 1}(y) \leq \frac{M_{\gamma, h}}{N_\gamma - 1}\right), \end{aligned}$$

where  $F_{N_\gamma - 1}(\cdot)$  is the empirical cumulative distribution function for  $\{Y_j\}_{j \in U_\gamma : j \neq k}$ . From the classical Glivenko–Cantelli theorem,  $F_{N_\gamma - 1}(y)$  converges uniformly almost surely to  $F$ . Similar computations can be used to derive  $m'_\gamma(y_1, y_2)$  and  $c_\gamma(y_1, y_2)$  and their limits. With such derivations, it is possible to establish the following result, the proof of which is omitted.

**Result 1.** *If  $G(\alpha) = \lim_{\gamma \rightarrow \infty} \sum_{h=1}^{H_\gamma} n_{\gamma h} N_{\gamma h}^{-1} \mathbb{1}_{(N_\gamma^{-1} M_{\gamma, h-1}, N_\gamma^{-1} M_{\gamma h})}(\alpha)$  exists except for a finite number of points and is a piecewise continuous nonnull function, and the convergence is uniform in  $\alpha$  then A3 and A2 hold.*

## 4. Conclusion

We have given assumptions on the selection mechanism and the superpopulation model under which the unweighted empirical c.d.f. converges uniformly to a weighted version of the superpopulation c.d.f. Because the conditions we specify on the informative selection mechanism are closely tied to first and second-order inclusion probabilities in a standard design-based survey sampling setting, the conditions are verifiable. Our examples illustrate the computations for selection mechanisms encountered in real surveys and observational studies. We expect these conditions to be useful in studying the properties of other basic sample statistics under informative selection, which will be the subject of further research.

## Appendix A: Proofs of Theorems 1 and 2

The first subsection contains the proof of Theorem 1. The proof consists in showing the uniform  $L_2$  convergence of the empirical c.d.f., seen as a ratio of two random variables. First, we show that from A1 we can deduce the  $L_2$  convergence of both the numerator and denominator, then the classical proof of Glivenko–Cantelli is adapted to obtain a uniform  $L_2$  convergence.

The second subsection contains the proof of Theorem 2. We first construct two sequences of random variables  $(\mathcal{I}'_\gamma)$  and  $Y'$  such that  $\forall \gamma$ ,  $(\mathcal{I}'_\gamma, \mathcal{Y}'_\gamma)$  and  $(\mathcal{I}_\gamma, \mathcal{Y}_\gamma)$  have the same distribution. We then prove uniform  $L_2$  convergence of the empirical c.d.f. defined from  $(\mathcal{I}'_\gamma)$  and  $Y'$ , almost surely in  $Y'$ . The result is “design-based” in the sense that it is conditional on  $Y'$ , and is of independent interest. We conclude by showing the almost sure convergence.

### A.1. Proof of Theorem 1: Uniform $L_2$ convergence of the empirical c.d.f.

**Lemma 1.** *Given a bounded measurable function  $b: \mathbb{R} \rightarrow \mathbb{R}$ , A0 and A1 imply that*

$$\frac{\sum_{k \in U_\gamma} b(Y_k) I_{\gamma k}}{N_\gamma} \xrightarrow[\gamma \rightarrow \infty]{L_2} \int b m f \, d\lambda.$$

**Proof.** Assume A0 and A1. The exchangeability property (1) implies

$$\mathbb{E} \left[ \frac{\sum_{k \in U_\gamma} b(Y_k) I_{\gamma k}}{N_\gamma} \right] = \frac{\sum_{k \in U_\gamma} \mathbb{E}[b(Y_k) I_{\gamma k}]}{N_\gamma} = \int b m_\gamma f \, d\lambda \xrightarrow[\gamma \rightarrow \infty]{} \int b m f \, d\lambda$$

by (0a), (0b) and the dominated convergence theorem. Further, (1) implies

$$\begin{aligned} & \text{Var} \left( \frac{\sum_{k \in U_\gamma} b(Y_k) I_{\gamma k}}{N_\gamma} \right) \\ &= \frac{1}{N_\gamma^2} \sum_{k, \ell \in U_\gamma} \{ \text{Cov}(b(Y_k) \mathbb{E}[I_{\gamma k} | Y_k, Y_\ell], b(Y_\ell) \mathbb{E}[I_{\gamma \ell} | Y_k, Y_\ell]) \\ & \quad + \mathbb{E}[b(Y_k) b(Y_\ell) \text{Cov}(I_{\gamma k}, I_{\gamma \ell} | Y_k, Y_\ell)] \} \\ &= \left( 1 - \frac{1}{N_\gamma} \right) \left( \int b(y_1) b(y_2) m'_\gamma(y_1, y_2) m'_\gamma(y_2, y_1) f(y_1) f(y_2) \, dy_1 \, dy_2 \right. \\ & \quad \left. - \left( \int b(y_1) m'_\gamma(y_1, y_2) f(y_1) f(y_2) \, dy_1 \, dy_2 \right)^2 \right. \\ & \quad \left. + \int b(y_1) b(y_2) c_\gamma(y_1, y_2) f(y_1) f(y_2) \, dy_1 \, dy_2 \right) \\ & \quad + \frac{1}{N_\gamma} \left( \int b^2 v_\gamma f \, d\lambda + \int b^2 m_\gamma^2 f \, d\lambda - \left( \int b m_\gamma f \, d\lambda \right)^2 \right) \end{aligned}$$

$$\begin{aligned}
 &= \left(1 - \frac{1}{N_\gamma}\right) \left( \int b(y_1)b(y_2)(m'_\gamma(y_1, y_2)m'_\gamma(y_2, y_1) \right. \\
 &\quad \left. - m_\gamma(y_1)m_\gamma(y_2)) f(y_1) f(y_2) dy_1 dy_2 \right. \\
 &\quad \left. + \int b(y_1)b(y_2)c_\gamma(y_1, y_2) f(y_1) f(y_2) dy_1 dy_2 \right) \\
 &\quad + \frac{1}{N_\gamma} \left( \int b^2(v_\gamma + m_\gamma^2) f d\lambda - \left( \int bm_\gamma f d\lambda \right)^2 \right) \\
 &= o_\gamma(1)
 \end{aligned}$$

by (1a), (1b), and (1c), and the result is proved. □

**Lemma 2.** *Under A0 and A1, the numerator of the empirical c.d.f. converges uniformly in  $L_2$ :*

$$\lim_{\gamma \rightarrow \infty} E \left[ \left( \sup_{\alpha \in \mathbb{R}} \left| \frac{\sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(Y_k) I_{\gamma k}}{N_\gamma} - \int \mathbb{1}_{(-\infty, \alpha]} m_\gamma f d\lambda \right| \right)^2 \right] = 0.$$

**Proof.** We first define  $G_\gamma : \mathbb{R} \rightarrow \mathbb{R}^+$  and  $G_s : \mathbb{R} \rightarrow \mathbb{R}^+$  as

$$G_\gamma(\alpha) = \frac{1}{N_\gamma} \sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(Y_k) I_{\gamma k} \quad \text{and} \quad G_s(\alpha) = \int \mathbb{1}_{(-\infty, \alpha]} m f d\lambda.$$

With these definitions,

$$\sup_{\alpha \in \mathbb{R}} \left| \frac{\sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(Y_k) I_{\gamma k}}{N_\gamma} - \int \mathbb{1}_{(-\infty, \alpha]} m_\gamma f d\lambda \right| = \|G_\gamma - G_s\|_\infty.$$

Let  $\eta \in \mathbb{N}^*$  index the positive integers and define a sequence of subdivisions  $\{\alpha_{\eta, q}\}_{q=0}^{\eta+1}$  of  $\mathbb{R}$  via  $\alpha_{\eta, 0} = -\infty$ ,  $\alpha_{\eta, \eta+1} = \infty$ , and for  $q = 1, \dots, \eta$ ,

$$\alpha_{\eta, q} = \inf\{\alpha \in \mathbb{R} \mid G_s(\alpha) \geq \eta^{-1} q G_s(\infty)\}.$$

We first show that for all positive integers  $\eta$ ,

$$\sup_{\alpha \in \mathbb{R}} \{|G_\gamma(\alpha) - G_s(\alpha)|\} \leq \max_{0 \leq q \leq \eta+1} \{|G_\gamma(\alpha_{\eta, q}) - G_s(\alpha_{\eta, q})|\} + \frac{G_s(\infty)}{\eta}.$$

Let  $\eta \in \mathbb{N}$  and  $\alpha \in \mathbb{R}$ . Then  $\alpha \in [\alpha_{\eta, q}, \alpha_{\eta, q+1}]$  for some  $0 \leq q \leq \eta$ , and

$$\begin{aligned}
 G_\gamma(\alpha_{\eta, q}) &\leq G_\gamma(\alpha) \leq G_\gamma(\alpha_{\eta, q+1}), \\
 G_s(\alpha_{\eta, q}) &\leq G_s(\alpha) \leq G_s(\alpha_{\eta, q+1}), \\
 G_s(\alpha_{\eta, q+1}) - \frac{G_s(\infty)}{\eta} &\leq G_s(\alpha) \leq G_s(\alpha_{\eta, q}) + \frac{G_s(\infty)}{\eta}.
 \end{aligned}$$

Combining these inequalities, we have

$$\begin{aligned} G_\gamma(\alpha_{\eta,q}) - G_s(\alpha_{\eta,q}) - \frac{G_s(\infty)}{\eta} &\leq G_\gamma(\alpha) - G_s(\alpha) \\ &\leq G_\gamma(\alpha_{\eta,q+1}) - G_s(\alpha_{\eta,q+1}) + \frac{G_s(\infty)}{\eta}, \end{aligned}$$

so that

$$\begin{aligned} |G_\gamma(\alpha) - G_s(\alpha)| &\leq \max\{|G_\gamma(\alpha_{\eta,q}) - G_s(\alpha_{\eta,q})|, |G_\gamma(\alpha_{\eta,q+1}) - G_s(\alpha_{\eta,q+1})|\} + \frac{G_s(\infty)}{\eta} \\ &\leq \max_{0 \leq q' \leq \eta+1} \{|G_\gamma(\alpha_{\eta,q'}) - G_s(\alpha_{\eta,q'})|\} + \frac{G_s(\infty)}{\eta}. \end{aligned}$$

Thus, for all  $\alpha \in \mathbb{R}$ ,

$$|G_\gamma(\alpha) - G_s(\alpha)|^2 \leq 2 \left( \max_{0 \leq q' \leq \eta+1} \{|G_\gamma(\alpha_{\eta,q'}) - G_s(\alpha_{\eta,q'})|^2\} + \frac{G_s(\infty)^2}{\eta^2} \right),$$

so that

$$E[\|G_\gamma - G_s\|_\infty^2] \leq 2E \left[ \max_{0 \leq q \leq \eta+1} \{|G_\gamma(\alpha_{\eta,q}) - G_s(\alpha_{\eta,q})|^2\} \right] + \frac{2G_s(\infty)^2}{\eta^2}. \tag{9}$$

Let  $\varepsilon > 0$  be given. Choose  $\eta \in \mathbb{N}$  so large that  $2G_s(\infty)^2\eta^{-2} < \varepsilon/2$ , then use Lemma 1 to choose  $\Gamma$  so that  $\gamma \geq \Gamma$  implies

$$2E \left[ \max_{0 \leq q \leq \eta+1} \{|G_\gamma(\alpha_{\eta,q}) - G_s(\alpha_{\eta,q})|^2\} \right] < \frac{\varepsilon}{2}.$$

Hence, for all  $\gamma \geq \Gamma$ , the right-hand side of (9) is bounded by  $\varepsilon$ , which was arbitrary, so  $\lim_{\gamma \rightarrow \infty} E[(\|G_\gamma - G_s\|_\infty)^2] = 0$ . □

**Proof of Theorem 1.** By Definitions 1 and 3 and A0, for all  $\alpha \in \mathbb{R}$ ,

$$F_\gamma(\alpha) = \frac{G_\gamma(\alpha)}{G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0}}, \quad F_s(\alpha) = \frac{G_s(\alpha)}{G_s(\infty)},$$

so

$$\begin{aligned} \|F_\gamma - F_s\|_\infty &= \left\| \frac{G_\gamma}{G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0}} - \frac{G_s}{G_s(\infty)} \right\|_\infty \\ &= \left\| \frac{G_\gamma - G_s}{G_s(\infty)} + G_\gamma \frac{G_s(\infty) - (G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0})}{G_s(\infty)(G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0})} \right\|_\infty \end{aligned}$$

$$\begin{aligned} &\leq \frac{\|G_\gamma - G_s\|_\infty}{G_s(\infty)} + \frac{\|G_\gamma\|_\infty}{G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0}} \frac{|G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0} - G_s(\infty)|}{G_s(\infty)} \\ &\leq \frac{\|G_\gamma - G_s\|_\infty}{G_s(\infty)} + \frac{|G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0} - G_s(\infty)|}{G_s(\infty)} \\ &\leq \frac{\|G_\gamma - G_s\|_\infty}{G_s(\infty)} + \frac{|G_s(\infty) - G_\gamma(\infty)|}{G_s(\infty)} + \frac{\mathbb{1}_{G_\gamma(\infty)=0}}{G_s(\infty)}. \end{aligned}$$

From Lemma 2, the first two summands converge to 0 in  $L_2$ . From (1d), so does the third summand. □

### A.2. Proof of Theorem 2: Uniform almost sure convergence of the empirical c.d.f.

*Construction of  $\mathcal{I}'_\gamma, Y'$*

We define  $Y'$  and  $\mathcal{I}'_\gamma$  on the probability space  $(\Omega \times [0, 1], \mathcal{A} \otimes \mathcal{B}_{[0,1]}, \mathbf{P}' = \mathbf{P} \otimes \lambda_{[0,1]})$ . First, define  $Y' : \Omega \times [0, 1] \rightarrow \mathbb{R}^{\mathbb{N}}$  via

$$Y'(\omega, x) = Y(\omega).$$

Let  $\mathcal{Y}'_\gamma$  be the vector of random variables  $(Y'_1, \dots, Y'_{N_\gamma})$  and note that  $\mathcal{Y}'_\gamma(\omega, x) = \mathcal{Y}_\gamma(\omega)$ . Let  $S_{\gamma y} = \{i \in \mathbb{N}^{N_\gamma} : g_\gamma(i, y) \neq 0\}$  and note that for a given  $y \in \mathbb{R}^{N_\gamma}$ ,  $\sum_{i \in S_{\gamma y}} g_\gamma(i, y) = 1$ . Define  $h_\gamma : \mathbb{R}^{N_\gamma} \times \mathbb{N}^{N_\gamma} \rightarrow \mathbb{R}$  via

$$h_\gamma(y, i) = \sup_{\alpha \in \mathbb{R}} \left| \frac{\sum_{k \in U_\gamma} i_k \mathbb{1}_{(-\infty, \alpha]}(y_k)}{\mathbb{1}_{i=0} + \sum_{k \in U_\gamma} (i_k)} - G_s(\alpha) \right|.$$

We now impose an order on the  $M_{\gamma y}$  vectors in  $S_{\gamma y}$  by requiring  $h_\gamma$  to be non-increasing; that is, for vectors  $i^{(t)}, i^{(u)} \in S_{\gamma y}$ ,  $t < u$  if and only if  $h_\gamma(y, i^{(t)}) \geq h_\gamma(y, i^{(u)})$ . Any ties can be resolved, for example, by randomization. For  $\omega \in \Omega$  and  $x \in [0, 1]$ , we then define  $\mathcal{I}'_\gamma(\omega, 0) = i^{(1)}$  and for  $x > 0$

$$\mathcal{I}'_\gamma(\omega, x) = \sum_{u=1}^{M_{\gamma y}} i^{(u)} \mathbb{1}_{(\sum_{t < u} g_\gamma(i^{(t)}, \mathcal{Y}_\gamma(\omega)), \sum_{t \leq u} g_\gamma(i^{(t)}, \mathcal{Y}_\gamma(\omega))]}(x).$$

Because we use uniform measure on  $\mathcal{B}_{[0,1]}$ , the vector  $i^{(u)}$  is sampled from  $S_{\gamma \mathcal{Y}_\gamma(\omega)}$  with probability  $g_\gamma(i^{(u)}, \mathcal{Y}_\gamma(\omega))$ . Thus, by construction we have for all  $\gamma$ ,

$$\mathbf{P}'[\mathcal{I}'_\gamma = i | \mathcal{Y}'_\gamma = y] = g_\gamma(i, y) = \mathbf{P}[\mathcal{I}_\gamma = i | \mathcal{Y}_\gamma = y]$$

and  $\mathbf{P}'[\mathcal{Y}'_\gamma = y] = \mathbf{P}[\mathcal{Y}_\gamma = y]$ , so that

$$\mathbf{P}'[\mathcal{I}'_\gamma = i, \mathcal{Y}'_\gamma = y] = \mathbf{P}[\mathcal{I}_\gamma = i, \mathcal{Y}_\gamma = y].$$

This yields the following property.

**Property 1.** For all  $\gamma$ ,

$$h_\gamma(\mathcal{J}'_\gamma, \mathcal{I}'_\gamma) = \sup_{\alpha \in \mathbb{R}} |F'_\gamma(\alpha) - F_s(\alpha)| = \|F'_\gamma - F_s\|_\infty$$

has the same law as  $\|F_\gamma - F_s\|_\infty$ , where  $F'_\gamma$  is defined in (7).

Define  $G'_\gamma : \mathbb{R} \rightarrow \mathbb{R}^+$  via

$$G'_\gamma(\alpha) = \frac{\sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(Y'_k) I'_{\gamma k}}{N_\gamma},$$

noting that  $F'_\gamma = G'_\gamma (G'_\gamma(\infty) + \mathbb{1}_{G'_\gamma(\infty)=0})^{-1}$ . We then have the following lemma.

**Lemma 3.** Under A0 and A2, for all  $\alpha \in \mathbb{R}$ ,

$$\lim_{\gamma \rightarrow \infty} \int_{[0,1]} (G'_\gamma(\alpha)(\omega, x) - G_s(\alpha))^2 d\lambda(x) = 0 \quad P\text{-a.s. } (\omega).$$

**Proof.** Let

$$\Omega_{GC} = \left\{ \omega \in \Omega : \lim_{\gamma \rightarrow \infty} \sup_{\alpha \in \mathbb{R}} \left| N_\gamma^{-1} \sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(Y_k)(\omega) - \int \mathbb{1}_{(-\infty, \alpha]} f d\lambda \right| = 0 \right\}.$$

From the Glivenko–Cantelli theorem,  $P(\Omega_{GC}) = 1$ . We will show that for all  $\omega \in \Omega_{GC}$ ,

$$\int_{[0,1]} (G'_\gamma(\alpha)(\omega, x) - G_s(\alpha))^2 d\lambda(x) = o_\gamma(1).$$

Let  $\omega \in \Omega_{GC}$ . We then have

$$\begin{aligned} & \sqrt{\int_{[0,1]} (G'_\gamma(\alpha)(\omega, x) - G_s(\alpha))^2 d\lambda(x)} \\ & \leq \sqrt{\int_{[0,1]} \left( G'_\gamma(\alpha)(\omega, x) - \frac{\sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(Y_k(\omega)) \int_{[0,1]} I'_{\gamma k}(\omega, u) d\lambda(u)}{N_\gamma} \right)^2 d\lambda(x)} \\ & \quad + \left| \frac{\sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(Y_k(\omega)) \int_{[0,1]} I'_{\gamma k}(\omega, u) d\lambda(u)}{N_\gamma} - \frac{\sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(Y_k(\omega)) m_\gamma(Y_k(\omega))}{N_\gamma} \right| \\ & \quad + \left| \frac{\sum_{k \in U_\gamma} \mathbb{1}_{(-\infty, \alpha]}(Y_k(\omega)) m_\gamma(Y_k(\omega))}{N_\gamma} - \int \mathbb{1}_{(-\infty, \alpha]} m_\gamma f d\lambda \right| \\ & \quad + \left| \int \mathbb{1}_{(-\infty, \alpha]} m_\gamma f d\lambda - \int \mathbb{1}_{(-\infty, \alpha]} m f d\lambda \right|. \end{aligned}$$

The first term is the square root of

$$\text{Var}(G'_\gamma(\alpha) | \mathcal{Y}'_\gamma = (Y_1(\omega), \dots, Y_{N_\gamma}(\omega))) = N_\gamma^{-2} \mathfrak{o}_\gamma(N_\gamma^2) = \mathfrak{o}_\gamma(1)$$

by (2a). The second term is

$$\left| \sum_{k \in U_\gamma} \frac{\mathbb{1}_{(-\infty, \alpha]}(Y_k(\omega))}{N_\gamma} (\mathbb{E}[I'_{\gamma k} | \mathcal{Y}'_\gamma = (Y_1(\omega), \dots, Y_{N_\gamma}(\omega))] - m_\gamma(Y_k(\omega))) \right| = \mathfrak{o}_\gamma(1)$$

by (2b). The third term is  $\mathfrak{o}_\gamma(1)$  because the convergence of the empirical measure given by A2 implies the convergence of the integral for all bounded random variables. Finally, the fourth term is  $\mathfrak{o}_\gamma(1)$  by A0 and the dominated convergence theorem.  $\square$

The following lemma has its own interest, yielding design-based uniform  $L_2$  convergence of the empirical c.d.f.

**Lemma 4.** *Under A0 and A2,*

$$\int (h_\gamma(\mathcal{Y}'_\gamma(\omega, x), \mathcal{I}'_\gamma(\omega, x)))^2 d\lambda(x) = \mathfrak{o}_\gamma(1) \quad P\text{-a.s. } (\omega).$$

**Proof.** Starting from Lemma 3 and adapting the proof of Lemma 2, we have that:  $A2 \Rightarrow \int (\|G_\gamma(\mathcal{Y}'_\gamma(\omega, x), \mathcal{I}'_\gamma(\omega, x)) - G_s\|_\infty)^2 d\lambda(x) = \mathfrak{o}_\gamma(1)$  P-a.s.  $(\omega)$ . We then adapt the end of the proof of Theorem 1 and get the result.  $\square$

**Definition 5.** *For  $\omega \in \Omega$ ,  $\gamma \in \mathbb{N}$  and all  $\varepsilon > 0$ ,  $a_{\varepsilon, \gamma, \omega} \in [0, 1]$  is defined as*

$$a_{\varepsilon, \gamma, \omega} = \int_{[0, 1]} \mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, x) \geq \varepsilon\}} d\lambda(x) = \lambda_{[0, 1]}(\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, \cdot) \geq \varepsilon\}).$$

**Property 2.** *For all  $\varepsilon > 0$ ,*

$$\limsup_{\gamma \rightarrow \infty} \mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, x) > \varepsilon\}} = \mathbb{1}_{\{0\}} \quad P\text{-a.s. } (\omega).$$

**Proof.** First note that  $\forall x \in [0, 1]$ ,  $\mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, x) > \varepsilon\}} = \mathbb{1}_{(0, a_{\varepsilon, \gamma, \omega}]}(x)$ , because by construction of  $\mathcal{I}'_\gamma, \mathcal{Y}'_\gamma$ ,  $\{x \in [0, 1]: h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, x) > \varepsilon\}$  is a subinterval of  $[0, 1]$  containing 0 of measure  $a_{\varepsilon, \gamma, \omega}$ . Further,  $\forall x \in [0, 1]$ ,

$$\limsup_{\gamma \rightarrow \infty} \mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, x) > \varepsilon\}} = \mathbb{1}_{[0, \limsup_{\gamma \rightarrow \infty} a_{\varepsilon, \gamma, \omega}]}(x). \tag{10}$$

By Lemma 4, the random variable

$$h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, \cdot) : ([0, 1], \mathcal{B}_{[0, 1]}, \lambda_{[0, 1]}) \rightarrow \mathbb{R}$$

converges in  $L_2(\lambda)$  to 0, P-a.s. ( $\omega$ ), hence it also converges in probability to 0, and so  $\lim_{\gamma \rightarrow \infty} a_{\varepsilon, \gamma, \omega} = 0$ . The result then follows from equation (10).  $\square$

**Proof of Theorem 2.** We want to show that

$$A0, A2 \Rightarrow \|F'_\gamma - F_s\|_\infty \xrightarrow{\text{a.s.}} 0 \quad \text{as } \gamma \rightarrow \infty,$$

which is equivalent to showing that

$$A0, A2 \Rightarrow P'\left(\left\{\lim_{\gamma \rightarrow \infty} h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma) = 0\right\}\right) = 1.$$

Assume A0 and A2. We calculate:

$$\begin{aligned} & P'\left(\left\{\lim_{\gamma \rightarrow \infty} h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma) = 0\right\}\right) \\ &= P'\left(\bigcap_{\varepsilon > 0} \bigcup_{\Gamma} \bigcap_{\gamma > \Gamma} \{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma) < \varepsilon\}\right) \\ &= \lim_{\varepsilon \rightarrow 0} P'\left(\bigcup_{\Gamma} \bigcap_{\gamma > \Gamma} \{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma) < \varepsilon\}\right) \\ &= \lim_{\varepsilon \rightarrow 0} 1 - P'\left(\bigcap_{\Gamma} \bigcup_{\gamma > \Gamma} \{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma) \geq \varepsilon\}\right) \\ &= 1 - \lim_{\varepsilon \rightarrow 0} \int \limsup_{\gamma \rightarrow \infty} \mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, x) \geq \varepsilon\}} dP'(\omega, x). \end{aligned}$$

Let  $\varepsilon > 0$ . Applying Fubini's theorem,

$$\begin{aligned} & \int \limsup_{\gamma \rightarrow \infty} \mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, x) \geq \varepsilon\}} dP'(\omega, x) \\ &= \int \left( \int \limsup_{\gamma} \mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, x) \geq \varepsilon\}} d\lambda_{[0,1]}(x) \right) dP(\omega). \end{aligned}$$

Since we have  $\limsup_{\gamma \rightarrow \infty} \mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, x) \geq \varepsilon\}} = \mathbb{1}_{\{0\}}(x)$  P-a.s. ( $\omega$ ), we also have for all  $\varepsilon > 0$  that

$$\int \limsup_{\gamma \rightarrow \infty} \mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, x) \geq \varepsilon\}} d\lambda_{[0,1]}(x) = \int_{[0,1]} \mathbb{1}_{\{0\}}(x) d\lambda_{[0,1]}(x) = 0$$

P-a.s. ( $\omega$ ). Thus,

$$P'\left(\left\{\lim_{\gamma \rightarrow \infty} h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma) = 0\right\}\right) = 1. \quad \square$$

## Appendix B: Proof of Corollaries 1, 2

We state the following lemma which is a consequence of a theorem due to P olya (e.g., Serfling [35], page 18). The proof is omitted.

**Lemma 5.** *Let  $\{u_\gamma(\cdot)\}_{\gamma \in \mathbb{N}}$  be a sequence of increasing step functions,  $u_\gamma: \mathbb{R} \rightarrow [0, 1]$ , that converges pointwise to a continuous increasing function  $u: \mathbb{R} \rightarrow [0, 1]$  with  $\lim_{y \rightarrow -\infty} u(y) = 0$ ,  $\lim_{y \rightarrow \infty} u(y) = 1$  and  $0 < u(y_1) = u(y_2) < 1 \Rightarrow y_1 = y_2$ . Define  $q_\gamma(p) = \inf\{y \in \mathbb{R}: u_\gamma(y) \geq p\}$ ,  $q(p) = \inf\{y \in \mathbb{R}: u(y) \geq p\}$ . Then for all  $K$  a compact subset of  $(0, 1)$ ,  $\lim_{\gamma \rightarrow \infty} \sup_{p \in K} \{q_\gamma(p) - q(p)\} = 0$ .*

### B.1. Proof of Corollary 1

**Proof.** As  $m_\gamma f$  and  $mf$  may have different supports, we extend the definition of  $\xi_s$  by

$$\forall p \in \mathbb{R}, \quad \xi_s(p) = \inf\{y \in \mathbb{R}: F_s(y) \geq p\}.$$

Let  $K$  be a compact subset of  $(0, 1)$ . Then

$$\sup_{p \in K} |\xi_\gamma(p) - \xi_s(p)| \xrightarrow[\gamma \rightarrow \infty]{P} 0$$

if from all subsequences one can extract a subsequence that converges a.s. to 0. Let  $\tau: \mathbb{N} \rightarrow \mathbb{N}$  be a strictly increasing function. If  $\|F_\gamma - F_s\|_\infty \rightarrow^{L_2} 0$  then  $\|F_{\tau(\gamma)} - F_s\|_\infty \rightarrow^{L_2} 0$  and  $\|F_{\tau(\gamma)} - F_s\|_\infty \rightarrow^P 0$ . Then there exists  $\rho: \mathbb{N} \rightarrow \mathbb{N}$  strictly increasing such that  $\|F_{\tau(\rho(\gamma))} - F_s\|_\infty \rightarrow^{\text{a.s.}} 0$  and by Lemma 5,  $P(\lim_{\gamma \rightarrow \infty} \sup_{p \in K} |\xi_{\tau(\rho(\gamma))}(p) - \xi_s(p)| = 0) = 1$ .

For the uniform  $L_2$  convergence, let  $p \in (0, 1)$  and  $\alpha \in \mathbb{R}$ . Then  $|F_\gamma(\alpha) - F_s(\alpha)| \leq \|F_\gamma - F_s\|_\infty$ , so that

$$\begin{aligned} \{\alpha \in \mathbb{R}: F_s(\alpha) \geq p + \|F_\gamma - F_s\|_\infty\} &\subset \{\alpha \in \mathbb{R}: F_\gamma(\alpha) \geq p\} \\ &\subset \{\alpha \in \mathbb{R}: F_s(\alpha) \geq p - \|F_\gamma - F_s\|_\infty\}, \end{aligned}$$

and

$$\begin{aligned} \inf\{\alpha \in \mathbb{R}: F_s(\alpha) \geq p + \|F_\gamma - F_s\|_\infty\} &\geq \inf\{\alpha \in \mathbb{R}: F_\gamma(\alpha) \geq p\} \\ &\geq \inf\{\alpha \in \mathbb{R}: F_s(\alpha) \geq p - \|F_\gamma - F_s\|_\infty\}. \end{aligned}$$

Hence,  $\forall p \in (0, 1)$ ,  $\xi_s(p + \|F_\gamma - F_s\|_\infty) \geq \xi_\gamma(p) \geq \xi_s(p - \|F_\gamma - F_s\|_\infty)$ .

Further,  $f$  has compact support by hypothesis, so there exists  $b > 0$  such that the supports of  $(m_\gamma f)_{\gamma \in \mathbb{N}}$  and  $mf$  are included in  $[-b, b]$ . So  $\forall p \in (0, 1)$ ,  $\gamma \in \mathbb{N}$ ,  $-b \leq \xi_\gamma(p) \leq b$ ,  $-b \leq \xi_s(p) \leq b$ . By combining these three inequalities, we have,  $\forall p \in (0, 1)$ :

$$|\xi_s(p) - \xi_\gamma(p)| \leq \min\{b, \xi_s(p + \|F_\gamma - F_s\|_\infty)\} - \max\{-b, \xi_s(p - \|F_\gamma - F_s\|_\infty)\}. \quad (11)$$

Since  $K \subset (0, 1)$  is compact, there exists  $a \in (0, 1)$  such that  $K \subset [a, 1 - a]$ . With the assumed continuity of  $F_s$ , we have that  $\xi_s$  is uniformly continuous on any subinterval of  $[0, 1]$  that does not contain zero. Thus, for  $\varepsilon > 0$ , there exists  $\eta \in (0, a/2)$  such that  $p \in K$  implies  $|\xi_s(p + \eta) - \xi_s(p - \eta)| \leq \varepsilon$ . If  $\|F_\gamma - F_s\|_\infty \leq \eta$ , then  $p + \|F_\gamma - F_s\|_\infty \leq p + \eta < 1 - a/2$ , and  $\xi_s(p + \|F_\gamma - F_s\|_\infty) < b$ ,  $p - \|F_\gamma - F_s\|_\infty \geq p - \eta > a/2$  and  $\xi_s(p - \|F_\gamma - F_s\|_\infty) > -b$ , so equation (11) is bounded by  $\varepsilon$ . If  $\|F_\gamma - F_s\|_\infty > \eta$ , then (11) is bounded by  $(2b) \mathbb{1}_{(\|F_\gamma - F_s\|_\infty > \eta)}$ . Thus,

$$\mathbb{E} \left[ \left( \sup_{p \in K} |\xi_\gamma(p) - \xi_s(p)| \right)^2 \right] \leq \varepsilon^2 + 4b^2 \mathbb{P}(\|F_\gamma - F_s\|_\infty > \eta).$$

Since  $\varepsilon$  was arbitrary and  $\mathbb{P}(\|F_\gamma - F_s\|_\infty > \eta) \rightarrow 0$  as  $\gamma \rightarrow \infty$ , the result follows.  $\square$

## B.2. Proof of Corollary 2

**Proof.** If  $\|F'_\gamma - F_s\|_\infty \xrightarrow{\text{a.s.}} 0$ , then for all  $K$  a compact subset of  $(0, 1)$ , and all  $(\omega, x) \in \{(\omega, x): \|F'_\gamma - F_s\|_\infty \rightarrow 0\}$ , we apply Lemma 5 with  $u_\gamma = F'_\gamma(\omega, x)$ ,  $u = F_s$ , and obtain that  $\mathbb{P}'(\lim_{\gamma \rightarrow \infty} \sup_{p \in K} |\xi'_\gamma(p) - \xi'_s(p)| = 0) = 1$ .  $\square$

## Acknowledgement

This research was supported in part by the US National Science Foundation (SES-0922142).

## References

- [1] Arratia, R., Goldstein, L. and Langholz, B. (2005). Local central limit theorems, the high-order correlations of rejective sampling and logistic likelihood asymptotics. *Ann. Statist.* **33** 871–914. [MR2163162](#)
- [2] Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Internat. Statist. Rev.* **51** 279–292. [MR0731144](#)
- [3] Breckling, J.U., Chambers, R.L., Dorfman, A.H., Tam, S.M. and Welsh, A.H. (1994). Maximum likelihood inference from sample survey data. *Internat. Statist. Rev.* **62** 349–363.
- [4] Breidt, F.J. and Opsomer, J.D. (2000). Local polynomial regression estimators in survey sampling. *Ann. Statist.* **28** 1026–1053. [MR1810918](#)
- [5] Breidt, F.J. and Opsomer, J.D. (2008). Endogenous post-stratification in surveys: Classifying with a sample-fitted model. *Ann. Statist.* **36** 403–427. [MR2387977](#)
- [6] Cassel, C.M., Särndal, C.E. and Wretman, J.H. (1977). *Foundations of Inference in Survey Sampling*. Wiley Series in Probability and Mathematical Statistics. New York: Wiley-Interscience. [MR0652527](#)
- [7] Chambers, R.L., Dorfman, A.H. and Wang, S. (1998). Limited information likelihood analysis of survey data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 397–411. [MR1616053](#)
- [8] Chambers, R.L. and Skinner, C. J. (2003). *Analysis of Survey Data*. Chichester: Wiley. [MR1978840](#)
- [9] Cox, D. (1969). Some sampling problems in technology. In *New Developments in Survey Sampling* (U. Johnson and H. Smith, eds.) 506–527. New York: Wiley Interscience.
- [10] Eagleson, G.K. and Weber, N.C. (1978). Limit theorems for weakly exchangeable arrays. *Math. Proc. Cambridge Philos. Soc.* **84** 123–130. [MR0501275](#)

- [11] Eideh, A. and Nathan, G. (2009). Two-stage informative cluster sampling-estimation and prediction with applications for small-area models. *J. Statist. Plann. Inference* **139** 3088–3101. [MR2535185](#)
- [12] Eideh, A.A.H. and Nathan, G. (2006). Fitting time series models for longitudinal survey data under informative sampling. *J. Statist. Plann. Inference* **136** 3052–3069. [MR2256216](#)
- [13] Eideh, A.A.H. and Nathan, G. (2007). Corrigendum to “Fitting time series models for longitudinal survey data under informative sampling” [*J. Statist. Plann. Inference* **136** 3052–3069]. *J. Statist. Plann. Inference* **137** 628.
- [14] Fuller, W. (2009). *Sampling Statistics*. New York: Wiley.
- [15] Hájek, J. (1981). *Sampling from a Finite Population. Statistics: Textbooks and Monographs* **37**. New York: Dekker. Edited by Václav Dupač, With a foreword by P. K. Sen. [MR0627744](#)
- [16] Hausman, J. and Wise, D. (1981). Stratification on endogenous variables and estimation: The Gary income maintenance experiment. In *Structural Analysis of Discrete Data*. Cambridge, MA: MIT Press.
- [17] Isaki, C.T. and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.* **77** 89–96. [MR0648029](#)
- [18] Jewell, N.P. (1985). Least squares regression with data arising from stratified samples of the dependent variable. *Biometrika* **72** 11–21. [MR0790196](#)
- [19] Kish, L. and Frankel, M.R. (1974). Inference from complex samples. *J. Roy. Statist. Soc. Ser. B* **36** 1–37. [MR0365812](#)
- [20] Krieger, A.M. and Pfeffermann, D. (1992). Maximum likelihood estimation for complex sample surveys. *Survey Methodology* **18** 225–239.
- [21] Langholz, B. and Goldstein, L. (2001). Conditional logistic analysis of case–control studies with complex sampling. *Biostatistics* **2** 63–84.
- [22] Leigh, G.M. (1988). A comparison of estimates of natural mortality from fish tagging experiments. *Biometrika* **75** 347–353.
- [23] Mantel, N. (1973). Synthetic retrospective studies and related topics. *Biometrics* **29** 479–486.
- [24] Nowell, C. and Stanley, L.R. (1991). Length-biased sampling in mall intercept surveys. *Journal of Marketing Research* **28** 475–479.
- [25] Patil, G.P. and Rao, C.R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics* **34** 179–189. [MR0507202](#)
- [26] Pfeffermann, D., Krieger, A.M. and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statist. Sinica* **8** 1087–1114. [MR1666233](#)
- [27] Pfeffermann, D., Moura, F.A.D.S. and Silva, P.L.d.N. (2006). Multi-level modelling under informative sampling. *Biometrika* **93** 943–959. [MR2285081](#)
- [28] Pfeffermann, D. and Sverchkov, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhyā Ser. B* **61** 166–186. [MR1720710](#)
- [29] Pfeffermann, D. and Sverchkov, M. (2007). Small-area estimation under informative probability sampling of areas and within the selected areas. *J. Amer. Statist. Assoc.* **102** 1427–1439. [MR2412558](#)
- [30] Pfeffermann, D. and Sverchkov, M. (2009). Inference under informative sampling. In *Handbook of Statistics* (D. Pfefferman and C.R. Rao, eds.) **29B** 455–487. Amsterdam: North-Holland.
- [31] Pfeffermann, D. and Sverchkov, M.Y. (2003). Fitting generalized linear models under informative sampling. In *Analysis of Survey Data. Wiley Ser. Surv. Methodol.* 175–195. Chichester: Wiley. [MR1978851](#)
- [32] Poisson, S.D. (1837). Recherches sur la probabilité des jugements en matière criminelle et en matière civile. In *Procédés des Règles Générales du Calcul des Probabilités*. Paris: Bachelier, Imprimeur-Libraire pour les Mathématiques.
- [33] Robinson, P.M. and Särndal, C.E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhyā Ser. B* **45** 240–248. [MR0748468](#)

- [34] Särndal, C.E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer. [MR1140409](#)
- [35] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. [MR0595165](#)
- [36] Shaw, D. (1988). On-site samples' regression: Problems of nonnegative integers, truncation, and endogenous stratification. *J. Econometrics* **37** 211–223. [MR0932141](#)
- [37] Skinner, C. (1994). Sample models and weights. In *Proceedings of the Section on Survey Research Methods* 133–142. Washington, DC: American Statistical Association.
- [38] Sullivan, P., Breidt, F., Ditton, R., Knuth, B., Leaman, B., O'Connell, V., Parsons, G., Pollock, K., Smith, S. and Stokes, S. (2006). *Review of Recreational Fisheries Survey Methods*. Washington, DC: National Academies Press.
- [39] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. [MR1652247](#)

*Received October 2010 and revised March 2011*