Submitted to the Bernoulli
arXiv: math.PR/0000000

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

DANIEL BONNÉRY [1,*], F. JAY BREIDT [2] and FRANÇOIS COQUET [1,†]

[1] *Ensai, 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,65.html; www.ensai.com/francois-coquet-rub,30.html

[2] *Department of Statistics, Colorado State University, Fort Collins, Colorado 80523-1877,*
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 (iid) random variables with a probability density function (pdf) f, referred to as the superpopulation model. The selection is informative in the sense that the joint distribution of the sample responses, given that they were selected, is not iid 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 (cdf) is studied. An asymptotic framework and weak conditions on the informative selection mechanism are developed under which the (unweighted) empirical cdf converges uniformly, in L_2 and almost surely, to a weighted version of the superpopulation cdf. 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.

AMS 2000 subject classifications: Primary 62D05; secondary 62E20..
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 realized as independent and identically distributed (iid) random variables with probability density function (pdf) 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 iid case for simplicity of exposition.) The selection is informative in the sense that the joint distribution of the sample responses, *given that they were selected*, is not iid f.

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

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 pdf*, defined as the conditional distribution of the random variable Y, given that it was selected. Under informative selection, the sample pdf differs from f. In particular, Pfeffermann et al. (1998) (see some motivating work in Skinner, 1994) have developed a *sample likelihood* approach to estimation and inference, which maximizes the criterion function formed by taking the product of the sample pdf's, as if the responses were iid. This methodology has been extended in a number of directions (Pfeffermann and Sverchkov, 1999, 2003, 2007; Pfeffermann et al., 2006; Eideh and Nathan, 2006, 2007, 2009).

Under a strong set of assumptions (in particular, sample size remains fixed as population size goes to infinity), Pfeffermann et al. (1998) have established the pointwise convergence of the joint distribution of the responses to the product of the sample pdf'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., 1994; Chambers et al., 1998), 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. Kish and Frankel, 1974; Binder, 1983; Chambers and Skinner, 2003, §2.4). Obviously, each of these likelihood-based approaches requires a model specification.

Rather than starting at the point of likelihood-based inferential methods, we take a step back and consider the problem of identifying a suitable model using observed data. In an ordinary inference problem with iid 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 iid sampling the empirical cumulative distribution function (cdf) converges uniformly almost surely to the population cdf, by the Glivenko-Cantelli theorem (e.g., Van der Vaart, 1998, Theorem 19.1). What is the behavior of the empirical cdf 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 cdf converges uniformly, in L_2 and almost surely, to a weighted version of the superpopulation cdf. 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 convergence results we obtain may be useful, for example, in identifying a suitable parametric family for the weighted cdf, from which a selection mechanism and a superpopulation pdf may be postulated using results in Pfeffermann et al. (1998). 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.

2. Results

2.1. Asymptotic framework and assumptions

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

We consider a sequence of finite populations and samples. The γth finite population is the set of elements indexed by $U_\gamma = (1, \ldots, N_\gamma)$. In the sampling literature (e.g., Särndal et al., 1992), U_γ 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 kth 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 $\mathcal{Y}_\gamma$:

$$g_\gamma(i_1 \ldots, i_{N_\gamma}, y_1 \ldots y_{N_\gamma}) = \mathrm{P}\left(\mathcal{I}_\gamma = (i_1, \ldots, i_{N_\gamma}) \mid \mathcal{Y}_\gamma = (y_1, \ldots, y_{N\gamma})\right).$$

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

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

We refer to (1) as the exchangeability assumption. It corresponds to the condition of weakly exchangeable arrays (Eagleson and Weber, 1978) applied to $(I_{\gamma k}, Y_k)_{\gamma \in \mathbb{N}, k \in U_\gamma}$.

Definition 1. *For $\gamma \in \mathbb{N}$, the empirical cdf is the random process $F_\gamma : \mathbb{R} \to [0,1]$ 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 γ, let $k, \ell \in U_\gamma$ with $k \neq \ell$. Assume exchangeability as in (1) and let*

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

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

The following conditions on m_γ are used in defining the limit cdf:

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

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

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

Definition 3. *Under A0, the limit cdf $F_s : \mathbb{R} \to [0, 1]$ is*

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

Remark: relation to sample pdf

Because of informative selection, the empirical cdf does not converge to the superpopulation cdf. Under some conditions to be specified below, it converges to F_s, a weighted integral of the superpopulation pdf. To see this, consider the case of without-replacement sampling and a single element, k. The sample pdf defined in Krieger and Pfeffermann (1992) is the conditional density of Y_k given $I_{\gamma k} = 1$. By Bayes' rule,

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

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

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

Thus, if observations were iid from the sample pdf, F_s would be the natural limiting cdf. A related argument can be used to show that the same weighted cdf 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 iid, and we next specify asymptotic weak dependence conditions among $\mathcal{I}_\gamma$ coordinates.

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

•A1. Uniform L_2 convergence conditions:

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

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

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

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

•A2. Uniform almost sure convergence conditions: *Let $y \in \mathbb{R}^{\mathbb{N}}$ 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 $\alpha \in \mathbb{R}$,

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

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

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

Properties of sampling without replacement

In the case of sampling without replacement, $\mathcal{I}_\gamma : \Omega \to \{0, 1\}^{N\gamma}$, 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:

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

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

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

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

These conditions imply A0 and A1.

Proof. Since $I_{\gamma k} \in \{0,1\}$, A0a and A1c always hold. By applying the Lebesgue dominated convergence theorem, we obtain that A1a is verified when $\forall y_1, y_2, c_\gamma(y_1,y_2) = o_\gamma(1)$ and A1b 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 iid sample of size $n_\gamma = \sum_{k\in U_\gamma} I_{\gamma k}$ (Fuller, 2009, Thm. 1.3.1), and the classic Glivenko-Cantelli theorem can be applied as soon as $n_\gamma \overset{\text{a.s.}}{\to} \infty$ as $\gamma \to \infty$. 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:

$$\left\{ \begin{array}{l} {N_\gamma}^{-1} E[n_\gamma] \to m \neq 0 \text{ as } \gamma \to \infty \\ Var(n_\gamma) = o_\gamma(N_\gamma^2). \end{array} \right. \tag{4}$$

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 $m_\gamma(y) = \mathrm{E}[I_{\gamma 1}] = N_\gamma^{-1}\mathrm{E}\left[n_\gamma\right]$ and ${N_\gamma}^{-1}\mathrm{E}\left[n_\gamma\right] \to m$ by A4, so A3a holds. Exchangeability also implies

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

so

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

by A4, so A3b is obtained, and A3c holds by independence. Finally,

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

establishing A3d.

We next show that A4 implies A2. For all $\alpha \in \mathbb{R}$,

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

by equation (5), so A2a holds. By independence,

$$
\mathrm{E}\left[I_{\gamma k} \mid \mathcal{Y}_\gamma = (y_1, \ldots, y_{N_\gamma})\right] = \mathrm{E}\left[I_{\gamma k} \mid Y_k = y_k\right] = m_\gamma(y_k),
$$

so A2b holds. Finally,

$$
g_\gamma((0, \ldots, 0), y) = \mathrm{P}\left(n_\gamma = 0\right) = o_\gamma(1)
$$

by independence and (6), so A2c holds.

□

Remark

In conventional finite population asymptotics (Isaki and Fuller, 1982; Robinson and Särndal, 1983; Breidt and Opsomer, 2000, 2008), conditions on design covariances $\mathrm{Cov}(I_{\gamma k}, I_{\gamma \ell})$ are imposed to guarantee that the Horvitz-Thompson estimator $\sum_{k \in U_\gamma} y_k I_{\gamma k} (\mathrm{E}\left[I_{\gamma k}\right])^{-1}$ 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 $N_\gamma \pi_{*\gamma} \to \infty$ is a sequence of lower bounds on the expected sample size, $\mathrm{E}\left[n_\gamma\right]$. These same conditions can be used to show that $\mathrm{Var}\left(n_\gamma\right) = O_\gamma(N_\gamma^2 / (N_\gamma \pi_{*\gamma})) = o_\gamma(N_\gamma^2)$, agreeing with A4.

2.2. Uniform convergence of the empirical cdf

In this section, we state the main results of the paper: uniform L_2 convergence of the empirical cdf and uniform almost sure convergence of the empirical cdf. 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 cdf

Theorem 1. *Under A0 and A1, the empirical cdf converges uniformly in L_2 in the sense that*

$$
\sup_{\alpha \in \mathbb{R}} \left| F_\gamma(\alpha) - F_s(\alpha) \right| = \left\| F_\gamma - F_s \right\|_\infty \underset{\gamma \to \infty}{\overset{L_2}{\to}} 0.
$$

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

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

and the empirical quantiles $\xi_\gamma : (0,1) \to \mathbb{R}$ 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,*

$$\sup_{p \in K} |\xi_\gamma(p) - \xi_s(p)| \underset{\gamma \to \infty}{\overset{P}{\to}} 0$$

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:

$$\sup_{p \in K} |\xi_\gamma(p) - \xi_s(p)| \underset{\gamma \to \infty}{\overset{L_2}{\to}} 0.$$

2.2.2. Uniform almost sure convergence of the empirical cdf

The Glivenko-Cantelli theorem gives uniform almost sure convergence of the empirical cdf under iid 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 cdf. 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 $\left(\Omega \times [0,1], \mathscr{A} \otimes \mathscr{B}_{[0,1]}, P' = P \otimes \lambda_{[0,1]}\right)$ such that*

- *$\|F'_\gamma - F_s\|_\infty$ converges 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 $\mathscr{B}_{[0,1]}$ is the σ-field of Borel sets, $\lambda_{[0,1]}$ is the Lebesgue measure on $[0,1]$, $\mathcal{I}'_\gamma = \left(I'_{\gamma 1}, \ldots, I'_{\gamma N_\gamma}\right)$, $\mathcal{Y}'_\gamma = \left(Y'_1, \ldots, Y'_{N_\gamma}\right)$ and $F'_\gamma : \mathbb{R} \to [0,1]$ 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,

$$\sup_{p \in K} |\xi_\gamma(p) - \xi_s(p)| \overset{a.s.}{\underset{\gamma \to \infty}{\longrightarrow}} 0$$

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, 1981) 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 iid Bernoulli(p) random variables, so $\mathrm{Var}\,(n_\gamma) = N_\gamma p(1-p)$ and condition A4 holds.
- **Poisson sampling** corresponds to a design in which, given a random vector $\left(\Pi_{\gamma 1}, \ldots, \Pi_{\gamma N_\gamma}\right) : \Omega \to (0,1]^{N_\gamma}$, the $\{I_{\gamma k}\}$ are a sequence of independent Bernoulli($\Pi_{\gamma k}$) random variables (Poisson, 1837). In this case, the variance of n_γ is given by

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

  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 $\mathrm{Var}\left(\sum_{k \in U_\gamma} \Pi_{\gamma k}\right) = \mathrm{Var}\left(\sum_{k \in U_\gamma} \pi_{\gamma k}\right) = 0$ 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_γ is a random positive real vector of size N_γ, and suppose that the law of $(Z_\gamma, \mathcal{Y}_\gamma)$ is invariant under any permutation of the coordinates. For n^*_γ fixed, consider the design in which $\Pi_{\gamma k} = n^*_\gamma Z_{\gamma k}\{\sum_{k\in U_\gamma} Z_{\gamma k}\}^{-1}$. Then

$$\operatorname{Var}\left(\sum_{k\in U_\gamma} \Pi_{\gamma k}\right) = \operatorname{Var}\left(n^*_\gamma\right) = 0$$

and A4 is satisfied when Z_γ 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

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

then

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

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 cdf.

3.2. Length-biased sampling

Length-biased sampling, in which $\mathrm{P}(I_{\gamma k} = 1 \mid Y_k = y_k) = m_\gamma(y_k) \propto y_k$, is pervasive in real surveys and observational studies. Cox (Cox, 1969) gives a now-classic example of sampling fibers in textile manufacture, in which $m_\gamma(y_k) \propto y_k =$ fiber length. In surveys of wildlife abundance, "visibility bias" means that larger individuals or groups are more noticeable (e.g., Patil and Rao, 1978), so $m_\gamma(y_k) \propto y_k =$ 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, 1991) or fishing at the seashore (Sullivan et al., 2006); the longer they spend doing the activity, the more likely the field staff are to intercept and interview them, so $m_\gamma(y_k) \propto y_k =$ activity time. In mark-recapture surveys of wildlife populations, individuals that live longer are more likely to be recaptured, so $m_\gamma(y_k) \propto y_k =$ lifetime (e.g., Leigh, 1988). 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: $m_\gamma(y_k) \propto y_k =$ latency period. Finally, propensity to respond to a survey is often related to a variable of interest; e.g., 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 γth 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

$$\mathrm{P}\left(I_{\gamma k} = 1 \mid Y_k = y_k\right) = 1 - \mathrm{P}\left(I_{\gamma k} = 0 \mid Y_k = y_k\right) = m_\gamma(y_k) \propto y_k.$$

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

$$\begin{aligned} m_\gamma(y) = \tau_\gamma y &\to \tau y = m(y) \\ c_\gamma(y_k, y_\ell) = 0, &\quad m'_\gamma(y_k, y_\ell) - m_\gamma(y_k) = 0 \\ \mathrm{P}\left(\mathcal{I}_\gamma = (0, \ldots, 0)\right) &= \mathrm{E}\left[\prod_{k \in U_\gamma} (1 - \tau_\gamma y_k)\right] \\ &\leq (1 - \tau_\gamma \epsilon)^{N_\gamma} = \exp(N_\gamma \ln(1 - \tau_\gamma \epsilon)) = o_\gamma(1), \end{aligned}$$

so that A3 is verified. It then follows that the limiting cdf is given by

$$F_s(\alpha) = \int \mathbb{1}_{(-\infty,\alpha]} \frac{y}{\mathrm{E}\left[\mathrm{Y}_1\right]} f \, d\lambda. \tag{8}$$

3.3. Cluster sampling

Let F denote the superpopulation cdf: $F(\tau) = \int \mathbb{1}_{(-\infty,\tau]} f d\lambda$. 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 mechanism is $\mathcal{I}_\gamma = i_{1\gamma}$ or $i_{2\gamma}$, each with probability $1/2$, so uniform convergence of the empirical cdf is not possible. Note that

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

so that

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

and A1a 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_γ. 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., 1992, §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 $\lim_{\gamma\to\infty} \rho_\gamma = \rho$. 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 $\gamma \to \infty$ for $A = \{k\}$ or $A = \{k, \ell\}$, and so for those sets A we have

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

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

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

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

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

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 (e.g., Prentice and Pyke, 1979; Scott and Wild, 1986), in which 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 sampling of n_γ draws, with probability of selecting element k on the hth 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 (pps) that is highly correlated with Y. Such a design will be highly efficient for estimation of the Y-total (indeed, a fixed-size pps design with probabilities proportional to Y_k would exactly reproduce the Y-total).

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

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

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

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

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 $n_\gamma N_\gamma^{-1} \to \tau \in [0,1]$ as $\gamma \to \infty$ and $\mathrm{E}\left[Y_1^6\right] < \infty$. Under these conditions, $m_\gamma(y) = \tau y (\mathrm{E}[Y_1])^{-1} + o_\gamma(1)$, and the limiting cdf 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, 1981; Jewell, 1985; Shaw, 1988). 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, 1981). 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., 2006).

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 γ, 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

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

denote the order statistics for the γth 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_γ. The γth 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 $M_{\gamma h} = \sum_{g=1}^{h} N_{\gamma h}$. Because $H_\gamma,\ N_\gamma$ and n_γ are not random, we then have

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

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

4. Conclusion

We have given assumptions on the selection mechanism and the superpopulation model under which the unweighted empirical cdf converges uniformly to a weighted version of the superpopulation cdf. 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 cdf, 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 cdf 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 cdf

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

$$\frac{\sum_{k\in U_\gamma} b(Y_k) I_{\gamma k}}{N_\gamma} \underset{\gamma\to\infty}{\overset{L_2}{\to}} \int bmf\, d\lambda.$$

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

$$\mathrm{E}\left[\frac{\sum_{k\in U_\gamma} b(Y_k) I_{\gamma k}}{N_\gamma}\right] = \frac{\sum_{k\in U_\gamma} \mathrm{E}\left[b(Y_k) I_{\gamma k}\right]}{N_\gamma} = \int bm_\gamma f\, d\lambda \underset{\gamma\to\infty}{\to} \int bmf\, d\lambda$$

by A0a, A0b and the dominated convergence theorem. Further, (1) implies

$$
\begin{aligned}
&\mathrm{Var}\left(\frac{\sum_{k\in U_\gamma} b(Y_k)I_{\gamma k}}{N_\gamma}\right)\\
&= \frac{1}{N_\gamma^2}\sum_{k,\ell\in U_\gamma}\left\{\mathrm{Cov}\left(b(Y_k)\mathrm{E}\left[I_{\gamma k}\mid Y_k,Y_\ell\right], b(Y_\ell)\mathrm{E}\left[I_{\gamma l}\mid Y_k,Y_\ell\right]\right)\right.\\
&\qquad \left.+\,\mathrm{E}\left[b(Y_k)b(Y_\ell)\mathrm{Cov}\left(I_{\gamma k}, I_{\gamma \ell}\mid Y_k,Y_\ell\right)\right]\right\}\\
&= \left(1-\frac{1}{N_\gamma}\right)\left(\int b(y_1)b(y_2)m'_\gamma(y_1,y_2)m'_\gamma(y_2,y_1)f(y_1)f(y_2)dy_1dy_2\right.\\
&\qquad -\left(\int b(y_1)m'_\gamma(y_1,y_2)f(y_1)f(y_2)dy_1dy_2\right)^2\\
&\qquad \left.+\int b(y_1)b(y_2)c_\gamma(y_1,y_2)f(y_1)f(y_2)dy_1dy_2\right)\\
&\quad +\frac{1}{N_\gamma}\left(\int b^2 v_\gamma f\,d\lambda+\int b^2 m_\gamma^2 f\,d\lambda-\left(\int b m_\gamma f\,d\lambda\right)^2\right)\\
&= \left(1-\frac{1}{N_\gamma}\right)\left(\int b(y_1)b(y_2)\left(m'_\gamma(y_1,y_2)m'_\gamma(y_2,y_1)\right.\right.\\
&\qquad \left.-m_\gamma(y_1)m_\gamma(y_2)\right)f(y_1)f(y_2)dy_1dy_2\\
&\qquad \left.+\int b(y_1)b(y_2)c_\gamma(y_1,y_2)f(y_1)f(y_2)dy_1dy_2\right)\\
&\quad +\frac{1}{N_\gamma}\left(\int b^2(v_\gamma+m_\gamma^2)f\,d\lambda-\left(\int b m_\gamma f\,d\lambda\right)^2\right)\\
&= o_\gamma(1)
\end{aligned}
$$

by A1a, A1b, and A1c, and the result is proved. □

Lemma 2. *Under A0 and A1, the numerator of the empirical cdf converges uniformly in L_2:*

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

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

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

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 $q=1,\ldots,\eta$,

$$\alpha_{\eta,q}=\inf\left\{\alpha\in\mathbb{R}\,\middle|\,G_s(\alpha)\geq\eta^{-1}qG_s(\infty)\right\}.$$

We first show that for all positive integers η,

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

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

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

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 $\alpha\in\mathbb{R}$,

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

so that

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

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 Γ so that $\gamma \geq \Gamma$ implies

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

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

Proof of Theorem 1

By Definitions 1 and 3 and A0, for all $\alpha \in \mathbb{R}$,

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

so

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

From Lemma 2, the first two summands converge to 0 in L_2. From A1d, so does the third summand.

□

A.2. Proof of Theorem 2: Uniform almost sure convergence of the empirical cdf

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

We define Y' and $\mathcal{I}'_\gamma$ on the probability space $\left(\Omega \times [0,1], \mathscr{A} \otimes \mathscr{B}_{[0,1]}, \mathrm{P}' = \mathrm{P} \otimes \lambda_{[0,1]}\right)$. First, define $Y' : \Omega \times [0,1] \to \mathbb{R}^{\mathbb{N}}$ 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 $h_\gamma : \mathbb{R}^{N_\gamma} \times \mathbb{N}^{N_\gamma} \to \mathbb{R}$ 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_γ 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, e.g., by randomization. For $\omega \in \Omega$ and $x \in [0, 1]$, we then define $\mathcal{I}'_\gamma(\omega, 0) = i^{(1)}$ and for $x > 0$

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

Because we use uniform measure on $\mathscr{B}_{[0,1]}$, 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 γ,

$$\mathrm{P}'\left[\mathcal{I}'_\gamma = i \mid \mathcal{Y}'_\gamma = y\right] = g_\gamma(i, y) = \mathrm{P}\left[\mathcal{I}_\gamma = i \mid \mathcal{Y}_\gamma = y\right]$$

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

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

This yields the following property:

Property 1. *For all γ,*

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

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

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

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

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

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

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

Proof. Let

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

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

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

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

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

The first term is the square root of

$$\mathrm{Var}\left(G'_\gamma(\alpha) \mid \mathcal{Y}'_\gamma = (Y_1(\omega), \ldots, Y_{N_\gamma}(\omega)) \right) = N_\gamma^{-2} o_\gamma(N_\gamma^2) = o_\gamma(1)$$

by A2a. The second term is

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

by A2b. The third term is $o_\gamma(1)$ 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 $o_\gamma(1)$ by A0 and the dominated convergence theorem. □

The following lemma has its own interest, yielding design-based uniform L_2 convergence of the empirical cdf.

Lemma 4. *Under A0 and A2,*

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

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

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

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

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

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

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 $\mathcal{I}'_\gamma, \mathcal{Y}'_\gamma$, $\left\{x \in [0,1] : h_\gamma(\mathcal{Y}'_\gamma,\mathcal{I}'_\gamma)(\omega,x) > \varepsilon\right\}$ is a subinterval of $[0,1]$ containing 0 of measure $a_{\varepsilon,\gamma,\omega}$. Further, $\forall x \in [0,1]$,

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

By Lemma 4, the random variable

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

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

Proof of Theorem 2

Proof. We want to show that

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

which is equivalent to showing that

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

Assume A0 and A2. We calculate:

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

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

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

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

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

P-a.s.(ω). Thus,

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

□

Appendix B: Proof of Corollaries 1, 2

We state the following lemma which is a consequence of a theorem due to Pólya (e.g., Serfling, 1980, p. 18). The proof is omitted.

Lemma 5. *Let* $\{u_\gamma(\cdot)\}_{\gamma\in\mathbb{N}}$ *be a sequence of increasing step functions,* $u_\gamma:\mathbb{R}\to[0,1]$, *that converges pointwise to a continuous increasing function* $u:\mathbb{R}\to[0,1]$ *with* $\lim_{y\to-\infty}u(y)=0$, $\lim_{y\to\infty}u(y)=1$ *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* $(0,1)$, $\lim_{\gamma\to\infty}\sup_{p\in K}\{q_\gamma(p)-q(p)\}=0$.

B.1. Proof of Corollary 1

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

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

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

$$\sup_{p\in K} |\xi_\gamma(p) - \xi_s(p)| \underset{\gamma\to\infty}{\overset{P}{\to}} 0$$

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

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

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, $\forall p \in (0,1)$:

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

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 ξ_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 ε. If $\|F_\gamma - F_s\|_\infty > \eta$, then (11) is bounded by $(2b)\mathbb{1}_{\{\|F_\gamma - F_s\|_\infty > \eta\}}$. Thus

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

Since ε was arbitrary and $\mathrm{P}\left(\|F_\gamma - F_s\|_\infty > \eta\right) \to 0$ as $\gamma \to \infty$, the result follows.

□

B.2. Proof of Corollary 2

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

References

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review / Revue Internationale de Statistique*, 51(3):279–292.

Breckling, J. U., Chambers, R. L., Dorfman, A. H., Tam, S. M., and Welsh, A. H. (1994). Maximum likelihood inference from sample survey data. *International Statistical Review / Revue Internationale de Statistique*, 62(3):349–363.

Breidt, F. J. and Opsomer, J. D. (2000). Local polynomial regression estimators in survey sampling. *Annals of Statistics*, 28:1026–1053.

Breidt, F. J. and Opsomer, J. D. (2008). Endogenous post-stratification in surveys: classifying with a sample-fitted model. *Annals of Statistics*, 36:403–427.

Chambers, R. L., Dorfman, A. H., and Wang, S. (1998). Limited information likelihood analysis of survey data. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 60(2):397–411.

Chambers, R. L. and Skinner, C. (2003). *Analysis of Survey Data*. Wiley Ser. Surv. Methodol. Wiley, Chichester.

Cox, D. (1969). Some sampling problems in technology. In Johnson, U. and Smith, H., editors, *New Developments in Survey Sampling*, pages 506–527. Wiley Interscience.

Eagleson, G. K. and Weber, N. C. (1978). Limit theorems for weakly exchangeable arrays. *Mathematical Proceedings of the Cambridge Philosophical Society*, 84(1):123–130.

Eideh, A. and Nathan, G. (2009). Two-stage informative cluster sampling-estimation and prediction with applications for small-area models. *Journal of Statistical Planning and Inference*, 139(9):3088–3101.

Eideh, A. A. H. and Nathan, G. (2006). Fitting time series models for longitudinal survey data under informative sampling. *Journal of Statistical Planning and Inference*, 136(9):3052–3069.

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** (2006), no. 9, 3052–3069]. *Journal of Statistical Planning and Inference*, 137(2):628.

Fuller, W. (2009). *Sampling Statistics*. Wiley, New York.

Hájek, J. (1981). *Sampling from a Finite Population*, volume 37 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York. Edited by Václav Dupač, With a foreword by P. K. Sen.

Hausman, J. and Wise, D. (1981). Stratification on endogenous variables and estimation: The Gary income maintenance experiment. In *Structural Analysis of Discrete Data*. MIT press.

Isaki, C. T. and Fuller, W. A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77(377):89–96.

Jewell, N. P. (1985). Least squares regression with data arising from stratified samples of the dependent variable. *Biometrika*, 72(1):11–21.

Kish, L. and Frankel, M. (1974). Inference from complex samples. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 36(1):1–37.

Krieger, A. M. and Pfeffermann, D. (1992). Maximum likelihood estimation for complex sample surveys. *Survey Methodology*, 18(2):225–239.

Leigh, G. M. (1988). A comparison of estimates of natural mortality from fish tagging experiments. *Biometrika*, 75(2):347–353.

Nowell, C. and Stanley, L. R. (1991). Length-biased sampling in mall intercept surveys. *Journal of Marketing Research*, 28(4):475–479.

Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, 34(2):179–189.

Pfeffermann, D., Krieger, A. M., and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, 8(4):1087–1114.

Pfeffermann, D., Moura, F. A. D. S., and Silva, P. L. d. N. (2006). Multi-level modelling under informative sampling. *Biometrika*, 93(4):943–959.

Pfeffermann, D. and Sverchkov, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhya: The Indian Journal of Statistics, Series B*, 61(1):166–186.

Pfeffermann, D. and Sverchkov, M. (2007). Small-area estimation under informative probability sampling of areas and within the selected areas. *Journal of the American Statistical Association*, 102(480):1427–1439.

Pfeffermann, D. and Sverchkov, M. Y. (2003). Fitting generalized linear models under informative sampling. In *Analysis of Survey Data*, Wiley Ser. Surv. Methodol., pages 175–195. Wiley, Chichester.

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 Probabilitiés*, Paris. Bachelier, Imprimeur-Libraire pour les Mathématiques.

Prentice, R. L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, 66(3):403–411.

Robinson, P. M. and Särndal, C.-E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhya: The Indian Journal of Statistics, Series B*, 45:240–248.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling.* Springer Series in Statistics. Springer-Verlag, New York.

Scott, A. J. and Wild, C. J. (1986). Fitting logistic models under case-control or choice based sampling. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 48(2):170–182.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Shaw, D. (1988). On-site samples' regression: problems of nonnegative integers, truncation, and endogenous stratification. *Journal of Econometrics*, 37(2):211–223.

Skinner, C. (1994). Sample models and weights. In *Proceedings of the Section on Survey Research Methods*, pages 133–142, Washington, DC. American Statistical Association.

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*. National Academies Press, Washington, DC.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York.