Bernoulli **25**(4B), 2019, 3527–3554
https://doi.org/10.3150/18-BEJ1101

Bernstein-type exponential inequalities in survey sampling: Conditional Poisson sampling schemes

PATRICE BERTAIL[1] and STEPHAN CLÉMENÇON[2]

[1]*Modal'X, UPL, Université Paris-Nanterre, 200 avenue de la République, Nanterre 92100, France. E-mail: patrice.bertail@u-paris10.fr*
[2]*LTCI, Telecom ParisTech, Université Paris-Saclay, 46 rue Barrault, 75013, Paris, France. E-mail: stephan.clemencon@telecom-paristech.fr*

This paper is devoted to establishing exponential bounds for the probabilities of deviation of a sample sum from its expectation, when the variables involved in the summation are obtained by sampling in a finite population according to a rejective scheme, generalizing simple random sampling without replacement, and by using an appropriate normalization. In contrast to Poisson sampling, classical deviation inequalities in the i.i.d. setting do not straightforwardly apply to sample sums related to rejective schemes, due to the inherent dependence structure of the sampled points. We show here how to overcome this difficulty, by combining the formulation of rejective sampling as Poisson sampling conditioned upon the sample size with the Esscher transformation. In particular, the Bennett/Bernstein type bounds thus established highlight the effect of the asymptotic variance of the (properly standardized) sample weighted sum and are shown to be much more accurate than those based on the negative association property shared by the terms involved in the summation. Beyond its interest in itself, such a result for rejective sampling is crucial, insofar as it permit to obtain tail bounds for many other sampling schemes, namely those that can be accurately approximated by rejective plans in the sense of the total variation distance.

Keywords: coupling; Esscher transformation; exponential inequality; Poisson survey scheme; rejective sampling; survey sampling

1. Introduction

Whereas many upper bounds for the probability that a sum of independent real-valued (integrable) random variables exceeds its expectation by a specified threshold value $t \in \mathbb{R}$ are documented in the literature (see, *e.g.*, [12] and the references therein), very few results are available when the $n \geq 1$ random variables involved in the summation are sampled from a population of finite cardinality $N \geq n$ according to a given survey scheme and next appropriately normalized in order to obtain an unbiased estimator of a total (referred to as the Horvitz–Thompson estimator, using the related survey weights as originally proposed in [25]). The sole situation where results in the independent setting straightforwardly carry over to survey samples (without replacement) corresponds to the case where the variables are sampled independently with possibly unequal weights, *that is*, Poisson sampling. For more complex sampling plans, the dependence structure between the sampled variables makes the study of the fluctuations of the resulting weighted

1350-7265

sum estimating the total very challenging. The case of *simple random sampling without replacement* has been first considered in [24], and refined in [31] and [3]. In contrast, the asymptotic behavior of the Horvitz–Thompson estimator as N and n simultaneously tend to infinity is well-documented in the literature. Following in the footsteps of the seminal contribution [23], a variety of limit results (*e.g.* consistency, asymptotic normality) have been established for Poisson sampling and next extended to rejective sampling viewed as conditional Poisson sampling given the sample size and to sampling schemes that are close to the latter in a *coupling* sense in [29] and [5]. Although the nature of the results established in this paper are nonasymptotic, very similar arguments are involved in their proofs, essentially based on conditioning upon the sampling size and coupling.

It is the major purpose of this article to extend tail bounds proved for simple random sampling without replacement to the case of rejective sampling, a widely studied fixed size sampling scheme generalizing it (see, *e.g.*, [19,21] or [9]). The approach we develop is thus based on viewing rejective sampling as conditional Poisson sampling given the sample size and writing then the deviation probability as a ratio of two quantities. The numerator is the joint probability that a Poisson sampling-based total estimator exceeds the threshold t and the size of the cardinality of the Poisson sample equals the (deterministic) size n of the rejective plan considered, while the denominator is the probability that the Poisson sample size is equal to n. Whereas a sharp lower bound for the denominator can be straightforwardly derived from a local Berry–Esseen bound proved in [17] for sums of independent, possibly non-identically distributed, Bernoulli variables, an accurate upper bound for the numerator can be established by means of an appropriate exponential change of measure (*i.e.*, Esscher transformation), following in the footsteps of the method proposed in [34], a refinement of the classical argument of Bahadur–Rao's theorem in order to improve exponential bounds in the independent setting. The tail bounds (of Bennett/Bernstein type) established by means of this method are shown to be sharp in the sense that they explicitly involve the "small" asymptotic variance of the Horvitz–Thompson total estimator based on rejective sampling, in contrast to those proved by using the *negative association* property of the sampling scheme.

The article is organized as follows. A few key concepts pertaining to survey theory are recalled in Section 2, as well as specific properties of Poisson and rejective sampling schemes. For comparison purpose, preliminary tail bounds in the conditional Poisson case are stated in Section 3. The main results of the paper, sharper exponential bounds for conditional Poisson sampling namely, are proved in Section 4, while Section 5 explains how they can be used to establish tail bounds for other sampling schemes, sufficiently close to rejective sampling in the sense of the total variation norm. A few remarks are finally collected in Section 6 and some technical details are deferred to the Appendix section.

2. Background and preliminaries

As a first go, we start with briefly recalling basic notions in survey theory, together with key properties of (conditional) Poisson sampling schemes (one may refer to [30] for instance). Here and throughout, the indicator function of any event $\mathcal{E}$ is denoted by $\mathbb{I}\{\mathcal{E}\}$, the power set of any set E by $\mathcal{P}(E)$, the covariance matrix between square integrable random vectors Y and Z of same dimensionality by $\mathrm{Cov}(Z, Y)$, the cardinality of any finite set E by $\#E$ and the Dirac mass at any point a by δ_a. For any real number x, we set $x^+ = \max\{x, 0\}$, $x^- = \max\{-x, 0\}$, $\lceil x \rceil = \inf\{k \in \mathbb{Z} : x \le k\}$ and $\lfloor x \rfloor = \sup\{k \in \mathbb{Z} : k \le x\}$.

2.1. Sampling schemes and Horvitz–Thompson estimation

Consider a finite population of $N \ge 1$ distinct units, say $\mathcal{I}_N = \{1, \dots, N\}$. A survey sample of (possibly random) size $n \le N$ is any subset $s = \{i_1, \dots, i_{n(s)}\} \in \mathcal{P}(\mathcal{I}_N)$ of size $n(s) = n$. A sampling design without replacement is defined as a probability distribution R_N on the set of all possible samples $s \in \mathcal{P}(\mathcal{I}_N)$. For all $i \in \mathcal{I}_N$, the probability that the unit i belongs to a random sample S defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and drawn from distribution R_N is denoted by $\pi_i = \mathbb{P}\{i \in S\} = R_N(\{i\})$. The π_i's are referred to as *first order inclusion probabilities*. The *second order inclusion probability* related to any pair $(i, j) \in \mathcal{I}_N^2$ is denoted by $\pi_{i,j} = \mathbb{P}\{(i, j) \in S^2\} = R_N(\{i, j\})$ (observe that $\pi_{i,i} = \pi_i$). Here and throughout, we denote by $\mathbb{E}[\cdot]$ the $\mathbb{P}$-expectation and by $\mathrm{Var}(Z)$ the variance of any $\mathbb{P}$-square integrable r.v. $Z : \Omega \to \mathbb{R}$. The random vector $\boldsymbol{\epsilon}_N = (\epsilon_1, \dots, \epsilon_N)$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$, where $\epsilon_i = \mathbb{I}\{i \in S\}$ fully characterizes the random sample $S \in \mathcal{P}(\mathcal{I}_N)$. In particular, the sample size is given by $n(S) = \sum_{i=1}^N \epsilon_i$, its expectation and variance by $\mathbb{E}[n(S)] = \sum_{i=1}^N \pi_i$ and $\mathrm{Var}(n(S)) = \sum_{1 \le i,j \le N} \{\pi_{i,j} - \pi_i \pi_j\}$ respectively. The 1-dimensional marginal distributions of the random vector $\boldsymbol{\epsilon}_N$ are the Bernoulli distributions $\mathrm{Ber}(\pi_i) = \pi_i \delta_1 + (1 - \pi_i)\delta_0$, $1 \le i \le N$ and its covariance matrix is $\Gamma_N = (\pi_{i,j} - \pi_i \pi_j)_{1 \le i,j \le N}$.

We place ourselves here in the *fixed-population* or *design-based* sampling framework, meaning that we suppose that a fixed (unknown) real value x_i is assigned to each unit $i \in \mathcal{I}_N$. As originally proposed in the seminal contribution [25], the Horvitz–Thompson estimator of the population total $S_N = \sum_{i=1}^N x_i$ is given by

$$\widehat{S}_{\boldsymbol{\pi}_N}^{\boldsymbol{\epsilon}_N} = \sum_{i=1}^{N} \frac{\epsilon_i}{\pi_i} x_i = \sum_{i \in S} \frac{1}{\pi_i} x_i, \tag{1}$$

with $0/0 = 0$ by convention. Throughout the article, we assume that the π_i's are all strictly positive. Hence, the expectation of (1) is $\mathbb{E}[\widehat{S}_{\boldsymbol{\pi}_N}^{\boldsymbol{\epsilon}_N}] = S_N$ and, in the case where the size of the random sample is deterministic, its variance is

$$\mathrm{Var}\big(\widehat{S}_{\boldsymbol{\pi}}^{\boldsymbol{\epsilon}_N}\big) = \sum_{i<j} \left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2 \times (\pi_i \pi_j - \pi_{i,j}). \tag{2}$$

The goal of this paper is to establish accurate bounds for tail probabilities

$$\mathbb{P}\big\{\widehat{S}_{\boldsymbol{\pi}_N}^{\boldsymbol{\epsilon}_N} - S_N > t\big\}, \tag{3}$$

where $t \in \mathbb{R}$, when the sampling scheme $\boldsymbol{\epsilon}_N$ is *rejective*, a very popular sampling plan that generalizes simple random sampling without replacement and can be expressed as a conditional Poisson scheme, as recalled in the following subsection for clarity. One may refer to [18] for instance, for an excellent account of survey theory, including many more examples of sampling designs.

2.2. Poisson and conditional Poisson samplings

Undoubtedly, one of the simplest sampling plan is the *Poisson survey scheme* (without replacement), a generalization of *Bernoulli sampling* originally proposed in [22] for the case of unequal weights: the ϵ_i's are independent and the sampling distribution P_N is thus entirely determined by the first order inclusion probabilities $\mathbf{p}_N = (p_1, \dots, p_N) \in]0,1[^N$:

$$\forall s \in \mathcal{P}(\mathcal{I}_N), \quad P_N(s) = \prod_{i \in s} p_i \prod_{i \notin s} (1 - p_i). \tag{4}$$

We point out that the (first order) inclusion probabilities of a Poisson scheme are systematically denoted by p_i rather than π_i throughout the article. Indeed, two sampling schemes are simultaneously considered in the subsequent analysis, a rejective scheme (defined below) with first order inclusion probabilities π_i and a Poisson scheme, which the rejective scheme appears to be a conditional version of, with different inclusion probabilities, therefore denoted differently in order to avoid confusion. Observe in addition that the behavior of the Horvitz–Thompson estimator (1) can be investigated by means of results established for sums of independent random variables. However, the major drawback of this sampling plan lies in the random nature of the corresponding sample size, impacting significantly the variability of (1). The variance of the Poisson sample size is given by $d_N = \sum_{i=1}^N p_i(1 - p_i)$, while the variance of (1) in this case is:

$$\mathrm{Var}\big(\widehat{S}^{\epsilon_N}_{\mathbf{p}_N}\big) = \sum_{i=1}^{N} \frac{1 - p_i}{p_i} x_i^2.$$

Because of the variance reduction it achieves, *rejective sampling*, a sampling design R_N of fixed size $n \le N$, is often preferred in practice. It generalizes the *simple random sampling without replacement*, where all samples with cardinality n are equally likely to be chosen, with probability $n!(N-n)!/N!$, all the corresponding first and second order probabilities being thus equal to n/N and $n(n-1)/(N(N-1))$, respectively. Denoting by $\boldsymbol{\pi}^R_N = (\pi^R_1, \dots, \pi^R_N)$ its first order inclusion probabilities and by $\mathcal{S}_n = \{s \in \mathcal{P}(\mathcal{I}_N) : \#s = n\}$ the subset of all possible samples of size n, it is defined by:

$$\forall s \in \mathcal{S}_n, \quad R_N(s) = C \prod_{i \in s} p^R_i \prod_{i \notin s} \big(1 - p^R_i\big), \tag{5}$$

where $C = 1/(\sum_{s \in \mathcal{S}_n} \prod_{i \in s} p^R_i \prod_{i \notin s}(1 - p^R_i))$ and the parameters $\mathbf{p}^R_N = (p^R_1, \dots, p^R_N) \in]0,1[^N$ yield first order inclusion probabilities equal to the π^R_i's and are such that $\sum_{i=1}^N p^R_i = n$. Under this latter additional condition, such a vector $\mathbf{p}^R_N$ exists and is unique (see [19]) and the related representation (5) is then said to be *canonical*. Notice incidentally that any vector $\mathbf{p}'_N \in]0,1[^N$ such that $p^R_i/(1 - p^R_i) = c p'_i/(1 - p'_i)$ for all $i \in \{1, \dots, n\}$ for some constant $c > 0$ can be used to write a representation of R_N of the same type as (5). Comparing (5) and (4) reveals that rejective sampling of fixed size n can be viewed as Poisson sampling given that the sample size is equal to n. It is for this reason that rejective sampling is usually referred to as *conditional Poisson sampling*. For simplicity's sake, the superscript R is omitted in the sequel. One must pay attention not to get the π_i's and the p_i's mixed up (except in the simple random sampling without replacement case, where these quantities are all equal to n/N): the latter are the first order inclusion probabilities of P_N, whereas the former are those of its conditional version R_N. However they can be related by means of the results stated in [23] (see Theorem 5.1 therein, as well as Lemma 7 in Section 4 and [9]): $\forall i \in \{1, \dots, N\}$,

$$\pi_i(1-p_i) = p_i(1-\pi_i) \times \big(1 - (\tilde{\pi} - \pi_i)/d_N^* + o\big(1/d_N^*\big)\big), \tag{6}$$

$$p_i(1-\pi_i) = \pi_i(1-p_i) \times \big(1 - (\tilde{p} - p_i)/d_N + o(1/d_N)\big), \tag{7}$$

where $d_N^* = \sum_{i=1}^N \pi_i(1-\pi_i)$, $d_N = \sum_{i=1}^N p_i(1-p_i)$, $\tilde{\pi} = (1/d_N^*)\sum_{i=1}^N \pi_i^2(1-\pi_i)$ and $\tilde{p} = (1/d_N)\sum_{i=1}^N p_i^2(1-p_i)$.

Since the major advantage of conditional Poisson sampling lies in its reduced variance property (compared to Poisson sampling in particular, see the discussion in Section 4), focus is next on exponential inequalities involving a variance term, of Bennett/Bernstein type namely.

3. Preliminary results

As a first go, we establish tail bounds for the Horvitz–Thompson estimator in the case where the variables are sampled according to a Poisson scheme. We next show how to exploit the *negative association* property satisfied by rejective sampling in order to extend the latter to conditional Poisson sampling. Of course, this approach does not account for the reduced variance property of Horvitz–Thompson estimators based on rejective sampling, it is the purpose of the next section to improve these first exponential bounds.

3.1. Tails bounds for Poisson sampling

As previously observed, bounding the tail probability (3) is easy in the Poisson situation insofar as the variables summed up in (1) are independent though possibly non identically distributed (since the inclusion probabilities are not assumed to be all equal). The following theorem thus directly follows from well-known results related to tail bounds for sums of independent random variables.

Theorem 1 (Poisson sampling). *Assume that the survey scheme $\boldsymbol{\epsilon}_N$ defines a Poisson sampling plan with first order inclusion probabilities $p_i > 0$, with $1 \le i \le N$. Then, we have almost-surely*: $\forall t > 0$, $\forall N \ge 1$,

$$\mathbb{P}\big\{\widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N} - S_N > t\big\} \le \exp\left(-\frac{\sum_{i=1}^N \frac{1-p_i}{p_i} x_i^2}{(\max_{1\le i\le N} \frac{x_i}{p_i})^2} h\left(\frac{\max_{1\le i\le N}\frac{|x_i|}{p_i} t}{\sum_{i=1}^N \frac{1-p_i}{p_i} x_i^2}\right)\right) \tag{8}$$

$$\le \exp\left(\frac{-t^2}{\frac{2}{3}\max_{1\le i\le N}\frac{|x_i|}{p_i} t + 2\sum_{i=1}^N \frac{1-p_i}{p_i} x_i^2}\right), \tag{9}$$

where $h(x) = (1+x)\log(1+x) - x$ *for* $x \ge 0$.

Bounds (8) and (9) straightforwardly result from Bennett inequality [4] and Bernstein exponential inequality [6] respectively, when applied to the independent random variables $(\epsilon_i/p_i)x_i$, $1 \le i \le N$. By applying these results to the variables $-(\epsilon_i/p_i)x_i$'s, the same bounds naturally hold for the deviation probability $\mathbb{P}\{\widehat{S}^{\boldsymbol{\epsilon}_N}_{\boldsymbol{p}_N} - S_N < -t\}$ (and, incidentally, for $\mathbb{P}\{|\widehat{S}^{\boldsymbol{\epsilon}_N}_{\boldsymbol{p}_N} - S_N| > t\}$ up to a factor 2). Details, as well as extensions to other deviation inequalities (see, *e.g.*, [20]), are left to the reader.

3.2. Exponential inequalities for sums of negatively associated random variables

For clarity, we first recall the definition of *negatively associated random variables*, see [27].

Definition 1. Let $Z_1, \dots, Z_n$ be random variables defined on the same probability space, valued in a measurable space $(E, \mathcal{E})$. They are said to be negatively associated iff for any pair of disjoint subsets A_1 and A_2 of the index set $\{1, \dots, n\}$

$$\mathrm{Cov}\Big(f\big((Z_i)_{i\in A_1}\big), g\big((Z_j)_{j\in A_2}\big)\Big) \le 0, \tag{10}$$

for any real valued measurable functions $f : E^{\#A_1} \to \mathbb{R}$ and $g : E^{\#A_2} \to \mathbb{R}$ that are both increasing in each variable.

The following result provides tail bounds for sums of negatively associated random variables, which extends the usual Bennett/Bernstein inequalities in the i.i.d. setting, see [4] and [6].

Theorem 2. *Let $Z_1, \dots, Z_N$ be square integrable negatively associated real valued random variables such that $|Z_i| \le c$ a.s. and $\mathbb{E}[Z_i] = 0$ for $1 \le i \le N$. Let $a_1, \dots, a_N$ be non negative constants and set $\sigma^2 = (1/N)\sum_{i=1}^N a_i^2 \,\mathrm{Var}(Z_i)$. Then, for all $t > 0$, we have: $\forall N \ge 1$,*

$$\mathbb{P}\left\{\sum_{i=1}^N a_i Z_i \ge t\right\} \le \exp\left(-\frac{N\sigma^2}{c^2} h\left(\frac{c\max_{1\le i\le N}|a_i| t}{N\sigma^2}\right)\right) \tag{11}$$

$$\le \exp\left(-\frac{t^2}{2N\sigma^2 + \frac{2c\max_{1\le i\le N}|a_i| t}{3}}\right), \tag{12}$$

where $h(x) = (1+x)\log(1+x) - x$ for $x \ge 0$.

Before detailing the proof, observe that the same bounds hold true for the tail probability $\mathbb{P}\{\sum_{i=1}^N a_i Z_i \le -t\}$ (and for $\mathbb{P}\{|\sum_{i=1}^N a_i Z_i| \ge t\}$ as well, up to a multiplicative factor 2). Refer also to Theorem 4 in [26] for a similar result in a more restrictive setting (*i.e.*, tail bounds for sums of *negatively related* r.v.'s) and to [32] as well.

Proof. The proof starts off with the usual Chernoff method: for all $\lambda > 0$,

$$\mathbb{P}\left\{\sum_{i=1}^{N} a_i Z_i \geq t\right\} \leq \exp\bigl(-t\lambda + \log \mathbb{E}\bigl[e^{\lambda \sum_{i=1}^{N} a_i Z_i}\bigr]\bigr). \tag{13}$$

Next, observe that, for all $\lambda > 0$, we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{n} a_i Z_i\right)\right] &= \mathbb{E}\left[\exp(\lambda a_n Z_n)\exp\left(\lambda \sum_{i=1}^{n-1} a_i Z_i\right)\right] \\ &\leq \mathbb{E}\bigl[\exp(\lambda a_n Z_n)\bigr]\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{n-1} a_i Z_i\right)\right] \\ &\leq \prod_{i=1}^{n} \mathbb{E}\bigl[\exp(\lambda a_i Z_i)\bigr], \end{aligned} \tag{14}$$

using (10) (with $f(Z_n) = \exp(\lambda a_n Z_n)$ and $g(Z_1, \ldots, Z_{n-1}) = \exp(\lambda \sum_{i\leq n} a_i Z_i)$) combined with a descending recurrence on i. The proof is finished by plugging (14) into (13), using the bound

$$\log \mathbb{E}\bigl[\exp(\lambda a_i Z_i)\bigr] \leq \lambda^2 a_i^2 \operatorname{Var}(Z_i)\Bigl(e^{\lambda c \max_{1\leq i\leq N}|a_i|} - 1 - \lambda c \max_{1\leq i\leq N}|a_i|\Bigr)\Big/\Bigl(c \max_{1\leq i\leq N}|a_i|\Bigr)^2$$

and optimizing finally the resulting bound w.r.t. $\lambda > 0$, just like in the proof of the classic Bennett/Bernstein inequalities, see [4] and [6]. □

The first assertion of the theorem stated below reveals that any rejective scheme $\boldsymbol{\epsilon}_N^*$ forms a collection of negatively associated r.v.'s, the second one appearing then as a direct consequence of Theorem 2. We underline that many sampling schemes (*e.g.*, Rao–Sampford sampling, Pareto sampling, Srinivasan sampling) of fixed size are actually described by random vectors $\boldsymbol{\epsilon}_N$ with negatively associated components, see [13] or [14], so that exponential bounds similar to that stated below can be proved for such sampling plans.

Theorem 3. *Let $N \geq 1$ and $\boldsymbol{\epsilon}_N^* = (\epsilon_1^*, \ldots, \epsilon_N^*)$ be the vector of indicator variables related to a rejective scheme on $\mathcal{I}_N$ with first order inclusion probabilities $(\pi_1, \ldots, \pi_N) \in]0, 1]^N$. Set $X'_N = \max_{i\leq N}|x_i|/\pi_i$ and $s_N^2 = \sum_{i\leq N}(1-\pi_i)x_i^2/\pi_i$. Then, the following assertions hold true.*

(i) *The Bernoulli random variables $\epsilon_1^*, \ldots, \epsilon_N^*$ are negatively associated.*
(ii) *For any $t \geq 0$ and $N \geq 1$, we have*:

$$\mathbb{P}\bigl\{\widehat{S}_{\boldsymbol{\pi}}^{\boldsymbol{\epsilon}_N^*} - S_N \geq t\bigr\} \leq 2\exp\left(-\frac{s_N^2}{X_N'^2} h\left(\frac{tX'_N/2}{s_N^2}\right)\right) \leq 2\exp\left(\frac{-t^2/4}{\frac{2}{3}X'_N t + 2s_N^2}\right),$$

where $h(x) = (1+x)\log(1+x) - x$ for $x \geq 0$.

Proof. Considering the usual representation of the distribution of $(\epsilon_1, \dots, \epsilon_N)$ as the conditional distribution of a sample of independent Bernoulli variables $(\epsilon_1^*, \dots, \epsilon_N^*)$ conditioned upon the event $\sum_{i=1}^N \epsilon_i^* = n$ (see Section 2.2), Assertion (i) is a straightforward consequence from Theorem 2.6 in [27], see also [2]. Assertion (i) shows in particular that Theorem 2 can be applied to the random variables $\{(\epsilon_i^*/\pi_i - 1)x_i^+ : 1 \le i \le N\}$ and to the random variables $\{(\epsilon_i^*/\pi_i - 1)x_i^- : 1 \le i \le N\}$ as well. Using the union bound, we obtain that

$$\mathbb{P}\big\{\widehat{S}_{\boldsymbol{\pi}}^{\boldsymbol{\epsilon}_N^*} - S_N \ge t\big\} \le \mathbb{P}\left\{\sum_{i=1}^N \left(\frac{\epsilon_i^*}{\pi_i} - 1\right)x_i^+ \ge t/2\right\} + \mathbb{P}\left\{\sum_{i=1}^N \left(\frac{\epsilon_i^*}{\pi_i} - 1\right)x_i^- \le -t/2\right\},$$

and a direct application of Theorem 2 to each of the terms involved in this bound straightforwardly proves Assertion (ii). Of course, if the x_i's are all nonnegative (respectively, all negative), the upper bounds are simpler and given by $\exp(-(s_N^2/X_N'^2)h(tX_N'/s_N^2)) \le \exp(-t^2/(2X_N't/3 + 2s_N^2))$. □

The negative association property permits to handle the dependence of the terms involved in the summation. However, it may lead to rather loose probability bounds. Indeed, except the factor 2, the bounds of Assertion (ii) exactly correspond to those stated in Theorem 1, as if the ϵ_i^*'s were independent, whereas the asymptotic variance σ_N^2 of $\widehat{S}_{\boldsymbol{\pi}}^{\boldsymbol{\epsilon}_N^*}$ can be much smaller than $\sum_{i=1}^N (1-\pi_i)x_i^2/\pi_i$. It is the goal of the subsequent analysis to improve these preliminary results and establish exponential bounds involving the asymptotic variance σ_N^2.

Remark 1. We point out that in the specific case of simple random sampling without replacement, *that is*, when $\pi_i = n/N$ for all $i \in \{1, \dots, N\}$, the inequality stated in Assertion (ii) is quite comparable (except the factor 2) to that which can be derived from the Chernoff bound given in [24], see Proposition 1.4 in [3].

4. Main results – exponential inequalities for rejective sampling

More accurate deviation probabilities related to the total estimator (1) based on a rejective sampling scheme $\boldsymbol{\epsilon}_N^*$ of (fixed) sample size $n \le N$ with first order inclusion probabilities $\boldsymbol{\pi}_N = (\pi_1, \dots, \pi_N)$ and canonical representation $\mathbf{p}_N = (p_1, \dots, p_N)$ are now investigated. Consider $\boldsymbol{\epsilon}_N$ a Poisson scheme with $\mathbf{p}_N$ as vector of first order inclusion probabilities. As previously recalled, the distribution of $\boldsymbol{\epsilon}_N^*$ is equal to the conditional distribution of $\boldsymbol{\epsilon}_N$ given $\sum_{i=1}^N \varepsilon_i = n$:

$$\big(\varepsilon_1^*, \varepsilon_2^*, \dots, \varepsilon_N^*\big) \stackrel{d}{=} (\varepsilon_1, \dots, \varepsilon_N)\Big| \sum_{i=1}^N \varepsilon_i = n. \tag{15}$$

Hence, we have almost-surely: $\forall t > 0$, $\forall N \ge 1$,

$$\mathbb{P}\big\{\widehat{S}_{\boldsymbol{\pi}_N}^{\boldsymbol{\epsilon}_N^*} - S_N > t\big\} = \mathbb{P}\left\{\sum_{i=1}^N \frac{\epsilon_i}{\pi_i}x_i - S_N > t \,\Big|\, \sum_{i=1}^N \epsilon_i = n\right\}. \tag{16}$$

As a first go, we shall prove tail bounds for the quantity

$$\widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*} \stackrel{\text{def}}{=} \sum_{i=1}^{N} \frac{\epsilon_i^*}{p_i} x_i. \tag{17}$$

Observe that this corresponds to the Horvitz–Thompson (HT in abbreviated form) estimator of the total $\sum_{i=1}^{N}(\pi_i/p_i)x_i$. Refinements of relationships (6) and (7) between the p_i's and the π_i's shall next allow us to obtain an upper bound for (16). Notice that, though slightly biased (see Assertion (i) of Theorem 5 for a control of the bias), the statistic (17) is commonly used as an estimator of S_N (see, *e.g.*, [23]), insofar as the parameters p_i's are readily available from the canonical representation of $\boldsymbol{\epsilon}_N^*$, whereas the computation of the π_i's is much more complicated. One may refer to [15] for practical algorithms dedicated to this task. Hence, Theorem 4 is of practical interest to build non-asymptotic confidence intervals for the total S_N.

Asymptotic variance. Recall that $d_N = \sum_{i=1}^{N} p_i(1-p_i)$ is the variance $\mathrm{Var}(\sum_{i=1}^{N}\epsilon_i)$ of the size of the Poisson plan $\boldsymbol{\epsilon}_N$ and set

$$\theta_N = \frac{\sum_{i=1}^{N} x_i(1-p_i)}{d_N}.$$

As explained in [7], the quantity θ_N is the coefficient of the linear regression relating $\sum_{i=1}^{N}\frac{\epsilon_i}{p_i}x_i - S_N$ to the sample size $\sum_{i=1}^{N}\epsilon_i$. We may thus write

$$\sum_{i=1}^{N}\frac{\epsilon_i}{p_i}x_i - S_N = \theta_N \times \sum_{i=1}^{N}\epsilon_i + r_N,$$

where the residual r_N is orthogonal to $\sum_{i=1}^{N}\epsilon_i$. Hence, we have the following decomposition

$$\mathrm{Var}\left(\sum_{i=1}^{N}\frac{\epsilon_i}{p_i}x_i\right) = \sigma_N^2 + \theta_N^2 d_N, \tag{18}$$

where

$$\sigma_N^2 = \mathrm{Var}\left(\sum_{i=1}^{N}(\epsilon_i - p_i)\left(\frac{x_i}{p_i} - \theta_N\right)\right) \tag{19}$$

is the asymptotic variance of the statistic $\widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*}$, see [23]. In other words, the variance reduction resulting from the use of a rejective sampling plan instead of a Poisson plan is equal to $\theta_N^2 d_N$, and can be arbitrarily large in practice, depending on the values taken by the (x_i, p_i)'s. A Bernstein type probability inequality for $\widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*}$ should thus involve σ_N^2 rather than the Poisson variance $\mathrm{Var}(\sum_{i=1}^{N}(\epsilon_i/p_i)x_i)$. Using the fact that $\sum_{i=1}^{N}(\epsilon_i - p_i) = 0$ on the event $\{\sum_{i=1}^{N}\epsilon_i = n\}$, we may now write

$$\mathbb{P}\left\{\widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*} - S_N > t\right\} = \mathbb{P}\left\{\sum_{i=1}^{N}\frac{\epsilon_i}{p_i}x_i - S_N > t \,\middle|\, \sum_{i=1}^{N}\epsilon_i = n\right\}$$

$$= \frac{\mathbb{P}\{\sum_{i=1}^N (\epsilon_i - p_i)\frac{x_i}{p_i} > t, \sum_{i=1}^N \epsilon_i = n\}}{\mathbb{P}\{\sum_{i=1}^N \epsilon_i = n\}}$$

$$= \frac{\mathbb{P}\{\sum_{i=1}^N (\epsilon_i - p_i)(\frac{x_i}{p_i} - \theta_N) > t, \sum_{i=1}^N \epsilon_i = n\}}{\mathbb{P}\{\sum_{i=1}^N \epsilon_i = n\}}, \tag{20}$$

by virtue of (16) and the definition of $\boldsymbol{\epsilon}_N^*$. Based on the observation that the random variables $\sum_{i=1}^N (\epsilon_i - p_i)(x_i/p_i - \theta_N)$ and $\sum_{i=1}^N (\epsilon_i - p_i)$ are uncorrelated, Eq. (20) thus permits to establish directly the CLT $\sigma_N^{-1}(\widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*} - S_N) \Rightarrow \mathcal{N}(0,1)$, provided that $d_N \to +\infty$, as $N \to +\infty$, simplifyingg asymptotically the ratio, see [23]. Hence, the asymptotic variance of $\widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*} - S_N$ is the variance σ_N^2 of the quantity $\sum_{i=1}^N (\epsilon_i - p_i)(x_i/p_i - \theta_N)$, which is less than that of the Poisson HT estimator (18), since it eliminates the variability due to the sample size. We also point out that Lemma 7 proved in the Appendix section straightforwardly shows that the "variance term" $\sum_{i=1}^N x_i^2(1-\pi_i)/\pi_i$ involved in the bound stated in Theorem 2 is always larger than $(1+6/d_N)^{-1}\sum_{i=1}^N x_i^2(1-p_i)/p_i$.

The desired result here is non asymptotic and accurate exponential bounds are required for both the numerator and the denominator of (20). It is proved in [23] (see Lemma 3.1 therein) that, as $N \to +\infty$:

$$\mathbb{P}\left\{\sum_{i=1}^N \epsilon_i = n\right\} = (2\pi d_N)^{-1/2}\big(1 + o(1)\big). \tag{21}$$

As shall be seen in the proof of the theorem stated below, the approximation (21) can be refined by using the results in [17] (see Lemma 1) and we thus essentially need to establish an exponential bound for the numerator with a constant of order $d_N^{-1/2}$, sharp enough so as to simplify the resulting ratio bound and cancel off the denominator. We shall prove that this can be achieved by using a similar argument to that considered in [8] for establishing an accurate exponential bound for i.i.d. 1-lattice random vectors, based on a device introduced in [34] for refining Hoeffding's inequality.

Theorem 4. *Let $N \geq 1$. Suppose that $\boldsymbol{\epsilon}_N^*$ is a rejective scheme of size $n \leq N$ with canonical parameter $\boldsymbol{p}_N = (p_1, \ldots, p_N) \in]0,1[^N$. Set $X_N = 2\max_{1\leq j\leq N}|x_j|/p_j$. Then, there exist a universal constant $D_0 > 0$ such that we have, as soon as $\min\{d_N, d_N^*\} \geq 1$ and $d_N \geq D$ where D denotes any constant strictly larger than D_0, for all $t > 0$ and for all $N \geq 1$,*

$$\mathbb{P}\big\{\widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*} - S_N > t\big\} \leq C\exp\left(-\frac{\sigma_N^2}{X_N^2} h\left(\frac{tX_N}{\sigma_N^2}\right)\right)$$
$$\leq C\exp\left(-\frac{t^2}{\frac{2}{3}X_N t + 2\sigma_N^2}\right),$$

where $C > 0$ is a constant depending only on D, $d_N = \sum_{i=1}^N p_i(1-p_i)$, $d_N^* = \sum_{i=1}^N \pi_i(1-\pi_i)$ and $h(x) = (1+x)\log(1+x) - x$ for $x \geq 0$.

The form of the constant C as a function of $D > D_0$ as well as an overestimated value of the constant D_0 can be deduced by a careful examination of the proof given below, see the discussion in the Appendix section. Before we detail it, we point out that the exponential bound in Theorem 4 involves the asymptotic variance of (17), in contrast to bounds obtained by exploiting the *negative association* property of the ϵ_i^*'s.

Remark 2. We underline that, in the particular case of sampling without replacement (*i.e.*, when $p_i = \pi_i = n/N$ for $1 \le i \le N$), the Bernstein type exponential inequality stated above provides a control of the tail similar to that obtained in [3], see Theorem 2 therein, with $k = n$. In this specific situation, we have $d_N = n(1 - n/N)$ and $\theta_N = S_N/n$, so that formula (19) then becomes

$$\sigma_N^2 = \left(1 - \frac{n}{N}\right)\frac{N^2}{n}\left\{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2\right\}.$$

The proof technique we consider here is very different from the (forward/backward) martingale methods used in [3], which consists in conditioning sequentially upon $\epsilon_1^*, \dots, \epsilon_k^*$ (respectively, upon $\epsilon_k^*, \dots, \epsilon_N^*$), k increasing from 1 to n (respectively, decreasing from n to 1), and in contrast applies to more general sampling schemes, with non-uniform weights. The control induced by Theorem 4 obtained this way is actually slightly better than that given by Theorem 3.5 in [3] when $n \le N/2$ (which situation is of particular interest in the context of survey sampling), insofar as the factor $(1 - n/N)$ is involved in the variance term, rather than $(1 - (n-1)/N)$. However, whereas the bound obtained in [3] holds true for any $n \le N$, that stated in Theorem 4 is valid only when $d_N = n(1 - n/N) \ge D$ (and thus, as soon as $d_N > 2 \times 923.1^2 \exp(1/4)/\pi$ using the results of the computations related to the evaluation of the constant C made in the Appendix section).

Proof. We first introduce additional notations. Set $Z_i = (\epsilon_i - p_i)(x_i/p_i - \theta_N)$ and $m_i = \epsilon_i - p_i$ for $1 \le i \le N$ and, for convenience, consider the standardized variables given by

$$\mathcal{Z}_N = n^{1/2}\frac{1}{N}\sum_{1 \le i \le N} Z_i \quad \text{and} \quad \mathcal{M}_N = d_N^{-1/2}\sum_{1 \le i \le N} m_i. \tag{22}$$

As previously announced, the proof technique is based on (20). Equipped with the notations introduced above, one may indeed write

$$\mathbb{P}\left\{\widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*} - S_N > t\right\} = \frac{\mathbb{P}\{\mathcal{Z}_N \ge t\sqrt{n}/N, \mathcal{M}_N = 0\}}{\mathbb{P}\{\mathcal{M}_N = 0\}}.$$

The proof then relies on the combination of two intermediary results: Lemma 1 which provides a lower bound for the denominator $\mathbb{P}\{\mathcal{M}_N = 0\}$ and Lemma 2 that gives an upper bound for the numerator $\mathbb{P}\{\mathcal{Z}_N \ge t\sqrt{n}/N, \mathcal{M}_N = 0\}$. They are established in the Appendix section.

Lemma 1. *Suppose that Theorem 4's assumptions are fulfilled. Then, there exists a universal constant D_0 such that, for all $N \geq 1$, we have as soon as $d_N \geq D$:*

$$\mathbb{P}\{\mathcal{M}_N = 0\} \geq C_1 \frac{1}{\sqrt{d_N}}, \tag{23}$$

where D is any constant strictly larger than D_0 and $C_1 = (1 - \sqrt{D_0/D}) \exp(-1/8)/\sqrt{2\pi}$.

As shown in the proof given in the Appendix section, it can be obtained by applying the local Berry–Esseen bound established in [17] for sums of independent (and possibly non identically) Bernoulli random variables. For completeness, an alternative lower bound (see Lemma 4), based on a binomial approximation result in [16] and involving a constant that can be evaluated much more easily, is also proved.

Lemma 2. *Suppose that Theorem 4's assumptions are fulfilled. Set $d_N^* = \sum_{1 \leq i \leq N} \pi_i(1-\pi_i)$. Then, we have for all $x \geq 0$, and for all $N \geq 1$ such that $\min\{d_N, d_N^*\} \geq 1$:*

$$\begin{aligned}\mathbb{P}\{\mathcal{Z}_N \geq x, \mathcal{M}_N = 0\} &\leq C_2 \frac{1}{\sqrt{d_N}} \exp\left(-\frac{\operatorname{Var}(\sum_{i=1}^N Z_i)}{X_N^2} h\left(\frac{N}{\sqrt{n}} \frac{x X_N}{\operatorname{Var}(\sum_{i=1}^N Z_i)}\right)\right) \\ &\leq C_2 \frac{1}{\sqrt{d_N}} \exp\left(-\frac{N^2 x^2/n}{2(\operatorname{Var}(\sum_{i=1}^N Z_i) + \frac{1}{3}\frac{N}{\sqrt{n}} x X_N)}\right),\end{aligned}$$

where $C_2 < +\infty$ is a universal constant and $h(x) = (1+x)\log(1+x) - x$ for $x \geq 0$.

The proof, which can be found in the Appendix section, relies on an appropriate exponential change of probability measure (originally used in [34] to refine Hoeffding's inequality) in order to make appear the factor $1/\sqrt{d_N}$. The bound stated in Theorem 4 now directly results from Eq. (20) combined with Lemmas 1 and 2, with $x = t\sqrt{n}/N$, by simplifying the factor $1/\sqrt{d_N}$. Due to this proof technique, the impact of an overestimation of the constant C_1 involved in Lemma 1 can be considerable, the constant in Theorem 4 being equal to C_2/C_1. □

Even if the computation of the biased statistic (17) is much more tractable from a practical perspective, we now come back to the study of the HT total estimator (1). The first part of the result stated below provides an estimation of the bias that replacement of (1) by (17) induces, whereas its second part finally gives a tail bound for (1).

Theorem 5. *Suppose that the assumptions of Theorem 4 are fulfilled and set $M_N = (4/d_N) \times \sum_{i=1}^N |x_i|/\pi_i$ and $X_N = 2\max_{1 \leq j \leq N} |x_j|/p_j$. The following assertions hold true.*

(i) *For all $N \geq 1$, we have almost-surely:*

$$\left|\widehat{S}_{\boldsymbol{\pi}_N}^{\boldsymbol{\epsilon}_N^*} - \widehat{S}_{\mathbf{p}_N}^{\boldsymbol{\epsilon}_N^*}\right| \leq M_N.$$

(ii) *There exists a universal constant $D_0 > 0$ such that, for all $t > M_N$ and for all $N \geq 1$, we have, as soon as $d_N^* \geq 1$ and $d_N > \max\{4, D\}$ where D is any constant strictly larger than D_0,*

$$\mathbb{P}\{\widehat{S}_{\boldsymbol{\pi}_N}^{\boldsymbol{\epsilon}_N^*} - S_N > t\} \leq C\exp\biggl(-\frac{\sigma_N^2}{X_N^2}h\biggl(\frac{N}{\sqrt{n}}\frac{(t-M_N)X_N}{\sigma_N^2}\biggr)\biggr)$$

$$\leq C\exp\biggl(-\frac{N^2(t-M_N)^2/n}{2(\sigma_N^2 + \frac{1}{3}\frac{N}{\sqrt{n}}(t-M_N)X_N)}\biggr),$$

where C is a constant depending only on D and $h(x) = (1+x)\log(1+x) - x$ for $x \geq 0$.

The proof is given in the Appendix section. We point out that, when $c_1 n/N \leq \pi_i \leq c_2 n/N$ for all $i \in \{1, \dots, N\}$ with $0 < c_1 \leq c_2 < +\infty$, if there exists $K < +\infty$ such that $\max_{1\leq i\leq N}|x_i| \leq K$ for all $N \geq 1$, then the bias term M_N is of order $o(N)$, provided that $\sqrt{N}/n \to 0$ as $N \to +\infty$.

On practical application. We point out that, as the (unknown) asymptotic variance involved in the bounds stated in Theorems 4 and 5 (essentially responsible for the refinement of the probability bounds obtained, compared with the results established in Section 3.2) can be expressed as

$$\sigma_N^2 = \frac{1}{2}\sum_{i,j}\frac{p_j(1-p_j)p_i(1-p_i)}{d_N}\biggl(\frac{X_i}{p_i} - \frac{X_j}{p_j}\biggr)^2, \tag{24}$$

which quantity can be straightforwardly estimated with a controlled error, as suggested in [3]. One may thus obtain confidence bounds for the total S_N by replacing it in the theoretical bounds deriving from Theorems 4 and 5 with the estimate thus computed, like in Theorem 4.3 of [3]. Indeed, we have

$$\sigma_N^2 \leq \frac{1}{32 d_N}\sum_{i,j}\biggl(\frac{X_i}{p_i} - \frac{X_j}{p_j}\biggr)^2. \tag{25}$$

Considering a sample uniformly drawn from the population without replacement of size n_1 (possibly different from n) with inclusion variables $\eta_1, \dots, \eta_N$ such that $\sum_{i=1}^N \eta_i = n_1$, an estimator of the bound on the right hand side of (25) is given by the empirical variance

$$b_{n_1}^2 = \frac{N^2}{16 d_N}\biggl(\frac{1}{n_1}\sum_{i=1}^N \eta_i\biggl(\frac{X_i}{p_i}\biggr)^2 - \biggl(\frac{1}{n_1}\sum_{j=1}\eta_j\frac{X_j}{p_j}\biggr)^2\biggr).$$

Now applying the results of [3] (Lemma 4.1 namely) to the X_i/p_i's, one obtains that, for any $0 < \delta < 1$, we have with probability larger than $1 - \delta$:

$$\sigma_N^2 \leq b_{n_1}^2 + \frac{N^2}{16 d_N}\max_{1\leq i\leq N}\biggl(\frac{|X_i|}{p_i}\biggr)(1 + \sqrt{1+\rho_{n_1}})\sqrt{\frac{\ln(3/\delta)}{2n_1}} = V_{N,n_1}^2(\delta),$$

where $\rho_{n_1} = (1 - n_1/N)$ when $n_1 < N/2$ (which is typically the relevant situation in survey sampling). Hence, if we replace σ_N^2 by V_{N,n_1}^2, one gets an empirical Bernstein version of the Bernstein bound given in Theorem 4: for any $0 < \delta < 1$,

$$\mathbb{P}\{\widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*} - S_N > t\} \leq C \exp\left(-\frac{t^2}{2(V_{N,n_1}^2(\delta) + \frac{1}{3} t X_N)}\right) + \delta. \tag{26}$$

Of course, replacing σ_N^2 by a surrogate bound deteriorates the probability inequality. Depending on the application considered, one may however try to calibrate δ and the size n_1 in order to build reasonable confidence intervals.

A numerical illustration. We now present numerical results related to the bounds obtained in Theorems 3, 4 and the bound (26) in a specific case. They illustrate that using the true variance rather than the Poisson like variance yield significant improvements. They also show that, although it deteriorates the theoretical bounds given in Theorems 4–5, the empirical Bernstein bound can yield much more accurate estimations than the bound stated in Theorem 3. Here, we consider an informative sampling plan based on inclusion probabilities $p_i = nW_i / \sum_{i=1}^N W_i$ that are proportional to the random variables $W_i = 1 + \gamma_i$, where the γ_i's are i.i.d. exponential variables with mean 1. The lower bound in Lemma 4 is close to $0.5n/N$ for the corresponding scheme, as can be shown by an immediate application of the law of large numbers. It follows that the constant C that can be computed explicitly given the p_i''s is of order 8.356. The X_i''s and the W_i's are linked through a linear model

$$X_i = \beta W_i + \sigma \varepsilon_i, \varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1),$$

the ε_i's being in addition independent from the W_i's. The parameters β and σ permit to tune the linear correlation between the variables X_i and W_i. The true tails of the Horvitz–Thompson estimator (17) are obtained by simulating 1999 samples according to respectively a Poisson scheme, a negatively associated sampling plan with fixed size (a pivotal sampling namely), a rejective sampling plan and a Rao–Sampford sampling plan. For this purpose, we used the R package "sampling" available at https://CRAN.R-project.org/package=sampling, see [36]. A simple box plot reveals that sampling with a an informative scheme and a fixed sample size considerably improves the variance of the HT-estimator. Figure 1 depicts such a boxplot for a correlation equal to 0.9. The results obtained are very similar when $n < N/2$, whatever the choice of large sizes N and n such that $n/N > 5$ (our bounds are not evaluated for N close to n, a situation that is not relevant in survey sampling). For a correlation of order 0.4, the Poisson plan still yields a very large variance and, when it is less than 0.2, the rejective sampling plan gives results very close the simple random sampling without replacement plan and the Poisson sampling plan.

Since the results are very similar for the three other sampling plans, only the tail of the HT estimator under the rejective sampling plan is plotted below and compared to the bounds we obtained. The bounds for the optimal constant C (equal to 1 here) are plotted as well for completeness. The graphic in Figure 2 displays the comparisons for a moderate sampling size $N = 300$ and $n = 30$ (with a true sum of order 600) and that in Figure 3 depicts the results for a large sampling size $N = 10^4$ and $n = 10^3$ (with a true sum equal to 20,000). All the simulations we

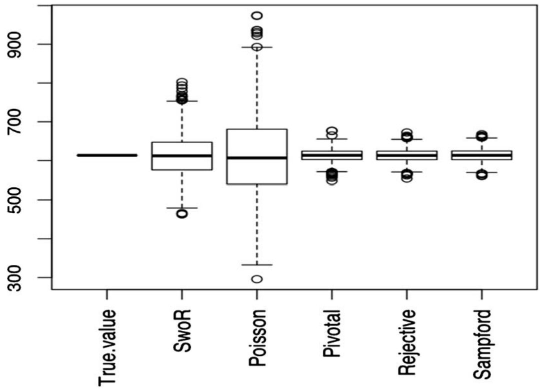

Figure 1. Box plots of the HT estimates for several sampling plans.

performed are in accordance with those presented here. The green curve on the left corner correspond to the true tail. The Bernstein bounds with $C = 1$ (in pink) are of course the best ones, then come the "rejective" bounds in blue (here the constant in our theorem is of order 8.6), the empirical Bernstein bounds (in black) with $n_1 = 2n$ are less accurate but still slightly better than the Poisson type bounds (in red).

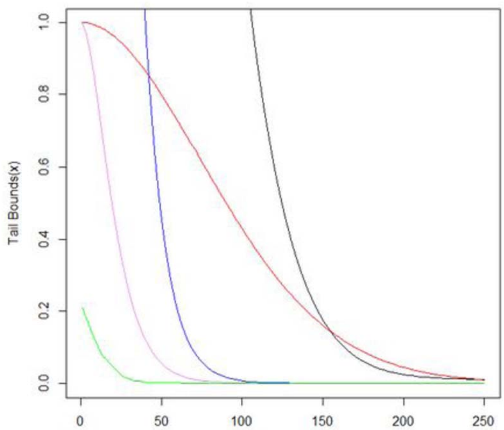

Figure 2. Comparisons between the true tail if the HT-estimator of the sum with the Bernstein bounds (resp. The optimal bound (green), the Rejective type bound (blue), the empirical Bernstein bound (black), the Poisson or NA bound (red), $N = 300$, $n = 30$).

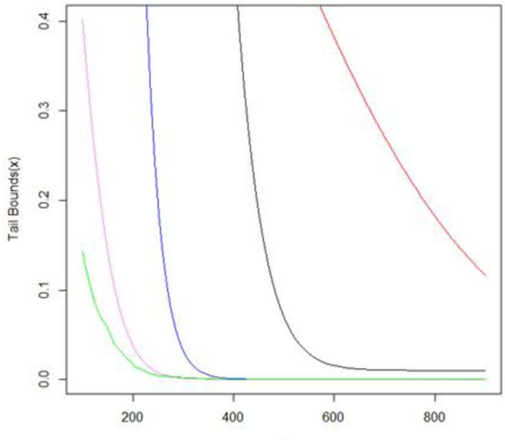

Figure 3. Comparisons between the true tail if the HT-estimator of the sum with the Bernstein bounds (resp. The optimal bound (green), the Rejective type bound (blue), the empirical Bernstein bound (black), the Poisson or NA bound (red), $N = 10^4$, $n = 10^3$).

5. Extensions to more general sampling schemes

We finally explain how the results established in the previous section for rejective sampling may permit to control tail probabilities for more general sampling plans. A similar argument is used in [5] to derive CLT's for HT estimators based on complex sampling schemes that can be approximated by more simple sampling plans, see also [7] and [10]. Let $\widetilde{R}_N$ and R_N be two sampling plans on the population $\mathcal{I}_N$ and consider the *total variation metric*

$$\|\widetilde{R}_N - R_N\|_1 \stackrel{\text{def}}{=} \sum_{s \in \mathcal{P}(\mathcal{I}_N)} \big|\widetilde{R}_N(s) - R_N(s)\big|,$$

as well as the *Kullback–Leibler divergence*

$$D_{\mathrm{KL}}\big(R_N || \widetilde{R}_N\big) \stackrel{\text{def}}{=} \sum_{s \in \mathcal{P}(\mathcal{I}_N)} R_N(s) \log\left(\frac{R_N(s)}{\widetilde{R}_N(s)}\right).$$

Equipped with these notations, we can state the following result.

Lemma 3. *Let $\boldsymbol{\epsilon}_N$ and $\widetilde{\boldsymbol{\epsilon}}_N$ be two schemes defined on the same probability space and drawn from plans R_N and $\widetilde{R}_N$ respectively and let $\mathbf{w}_N \in]0,1]^N$. Then, we have*: $\forall N \geq 1$, $\forall t \in \mathbb{R}$,

$$\left|\mathbb{P}\big\{\widehat{S}^{\boldsymbol{\epsilon}_N}_{\mathbf{w}_N} - S_N > t\big\} - \mathbb{P}\big\{\widehat{S}^{\widetilde{\boldsymbol{\epsilon}}_N}_{\mathbf{w}_N} - S_N > t\big\}\right| \leq \|\widetilde{R}_N - R_N\|_1 \leq \sqrt{2D_{\mathrm{KL}}\big(R_N || \widetilde{R}_N\big)}.$$

Proof. The first bound immediately results from the following elementary observation:

$$\mathbb{P}\big\{\widehat{S}^{\boldsymbol{\epsilon}_N}_{\mathbf{w}_N} - S_N > t\big\} - \mathbb{P}\big\{\widehat{S}^{\widetilde{\boldsymbol{\epsilon}}_N}_{\mathbf{w}_N} - S_N > t\big\}$$

$$= \sum_{s \in \mathcal{P}(\mathcal{I}_N)} \mathbb{I}\left\{\sum_{i \in s} x_i/w_i - S_N > t\right\} \times \left(R_N(s) - \widetilde{R}_N(s)\right),$$

while the second bound is the classical Pinsker's inequality. □

In practice, R_N is typically the rejective sampling plan investigated in the previous subsection (or eventually the Poisson sampling scheme), $\mathbf{w}_N$ corresponds to its first order inclusion probabilities $\boldsymbol{\pi}_N$ or to the Poisson weights $\mathbf{p}_N$ and $\widetilde{R}_N$ is a sampling plan for which the Kullback–Leibler divergence to R_N asymptotically vanishes as $N \to \infty$, *e.g.* the rate at which $D_{\mathrm{KL}}(R_N || \widetilde{R}_N)$ decays to zero has been investigated in [5] when $\widetilde{R}_N$ corresponds to Rao–Sampford, successive sampling or Pareto sampling under appropriate regular conditions (see also [11]). Lemma 3 combined with Theorem 5 permits then to obtain the upper bound

$$\mathbb{P}\left\{\widehat{S}_{\boldsymbol{\pi}_N}^{\widetilde{\boldsymbol{\epsilon}}_N} - S_N > t\right\} \le D_{\mathrm{KL}}\left(R_N || \widetilde{R}_N\right) + C \exp\left(-\frac{N^2(t - M_N)^2/n}{2(\sigma_N^2 + \frac{1}{3}\frac{N}{\sqrt{n}}(t - M_N)X_N)}\right).$$

A similar bound can be obtained for $\mathbb{P}\{\widehat{S}_{\mathbf{p}_N}^{\widetilde{\boldsymbol{\epsilon}}_N} - S_N > t\}$ using Theorem 4. As the first term on the right hand side is independent from t, it is essentially useful in situations where N is large, which is the typical framework for survey sampling. Denoting by $\widetilde{\boldsymbol{\pi}}_N$ the first order inclusion probabilities of $\widetilde{R}_N$, one may straightforwardly deduce an upper bound for $\mathbb{P}\{\widehat{S}_{\widetilde{\boldsymbol{\pi}}_N}^{\widetilde{\boldsymbol{\epsilon}}_N} - S_N > t\}$, as soon as $|\widehat{S}_{\widetilde{\boldsymbol{\pi}}_N}^{\widetilde{\boldsymbol{\epsilon}}_N} - \widehat{S}_{\boldsymbol{\pi}_N}^{\widetilde{\boldsymbol{\epsilon}}_N}|$ can be controlled like in Assertion (i) of Theorem 5 by bounding the deviations $|1/\pi_i - 1/\widetilde{\pi}_i|$ as in Lemma 7. One may refer to [5] for bounds of this type in the case of Rao–Sampford sampling or successive sampling.

6. Conclusion

In this article, we proved Bernstein-type tail bounds to quantify the deviation between a total and its Horvitz–Thompson estimator when based on conditional Poisson sampling, extending (and even slightly improving) results proved in the case of basic sampling without replacement. The original technique used to establish these inequalities is not based on coupling but relies on expressing the deviation probabilities related to a conditional Poisson scheme as conditional probabilities related to a Poisson plan. This permits to recover tight exponential bounds, involving the exact asymptotic variance of the Horvitz–Thompson estimator. Beyond the fact that rejective sampling is of prime importance in the practice of survey sampling (see, *e.g.*, [1,21] or [35]), this result may also yield tail bounds for sampling schemes that can be accurately approximated by rejective sampling in the total variation sense.

Appendix: Technical proofs

Proof of Lemma 1

For clarity, we first recall the following result. An estimation of the constant C involved in the bound it provides is given at the end of the appendix.

Theorem 6 ([17], Theorem 1.3). *Let $(Y_{j,n})_{1\le j\le n}$ be a triangular array of independent Bernoulli random variables with means $q_{1,n},\dots,q_{n,n}$ in $(0,1)$, respectively. Denote by $\sigma_n^2=\sum_{i=1}^n q_{i,n}(1-q_{i,n})$ the variance of the sum $\Sigma_n=\sum_{i=1}^n Y_{i,n}$ and by $\nu_n=\sum_{i=1}^n q_{i,n}$ its mean. Considering the cumulative distribution function (cdf) $F_n(x)=\mathbb{P}\{\sigma_n^{-1}(\Sigma_n-\nu_n)\le x\}$, we have: $\forall n\ge 1$,*

$$\sup_{k\in\mathbb{Z}}\left|F_n(x_{n,k})-\Phi(x_{n,k})-\frac{1-x_{n,k}^2}{6\sigma_n}\phi(x_{n,k})\left\{1-\frac{2\sum_{i=1}^n q_{i,n}^2(1-q_{i,n})}{\sigma_n^2}\right\}\right|\le\frac{C_0}{\sigma_n^2},$$

where $x_{n,k}=\sigma_n^{-1}(k-\nu_n+1/2)$ for any $k\in\mathbb{Z}$, $\Phi(x)=(2\pi)^{-1/2}\int_{-\infty}^x\exp(-z^2/2)\,dz$ is the cdf of the standard normal distribution $\mathcal{N}(0,1)$, $\phi(x)=\Phi'(x)$ and $C_0<+\infty$ is a universal constant.

Recall first that the ϵ_i's denote the independent Bernoulli variables with parameters $p_1,\dots,p_N$ related to the Poisson scheme and, given the definition of the quantity $\mathcal{M}_N$ in Eq. (22), observe that we can write:

$$\begin{aligned}\mathbb{P}\{\mathcal{M}_N=0\}&=\mathbb{P}\left\{\sum_{i=1}^N(\epsilon_i-p_i)\in\left]-1/2,1/2\right]\right\}\\&=\mathbb{P}\left\{d_N^{-1/2}\sum_{i=1}^N m_i\le\frac{1}{2}d_N^{-1/2}\right\}-\mathbb{P}\left\{d_N^{-1/2}\sum_{i=1}^N m_i\le-\frac{1}{2}d_N^{-1/2}\right\}.\end{aligned}$$

Applying Theorem 6 to bound the first term of this decomposition (with $k=\nu_n$ and $x_{n,k}=1/(2\sqrt{d_N})$) directly yields that

$$\begin{aligned}\mathbb{P}\left\{\frac{\sum_{i=1}^N m_i}{\sqrt{d_N}}\le\frac{1}{2\sqrt{d_N}}\right\}&\ge\Phi\left(\frac{1}{2\sqrt{d_N}}\right)\\&\quad+\frac{1-\frac{1}{4d_N}}{6\sqrt{d_N}}\phi\left(\frac{1}{2\sqrt{d_N}}\right)\left\{1-\frac{2\sum_{i=1}^n p_i^2(1-p_i)}{d_N}\right\}-\frac{C_0}{d_N}.\end{aligned}$$

For the second term, its application with $k=\nu_n-1$ entails that:

$$\begin{aligned}-\mathbb{P}\left\{\frac{1}{\sqrt{d_N}}\sum_{i=1}^N m_i\ge-\frac{1}{2\sqrt{d_N}}\right\}&\ge-\Phi\left(-\frac{1}{2\sqrt{d_N}}\right)\\&\quad-\frac{1-\frac{1}{4d_N}}{6\sqrt{d_N}}\phi\left(-\frac{1}{2\sqrt{d_N}}\right)\left\{1-\frac{2\sum_{i=1}^n p_i^2(1-p_i)}{d_N}\right\}-\frac{C_0}{d_N}.\end{aligned}$$

If $d_N\ge 1$, it follows that

$$\begin{aligned}\mathbb{P}\{\mathcal{M}_N=0\}&\ge\Phi\left(\frac{1}{2\sqrt{d_N}}\right)-\Phi\left(-\frac{1}{2\sqrt{d_N}}\right)-\frac{2C_0}{d_N}\\&=2\int_0^{\frac{1}{2\sqrt{d_N}}}\phi(t)\,dt-\frac{2C_0}{d_N}\ge\left(\phi(1/2)-\frac{2C_0}{\sqrt{d_N}}\right)\frac{1}{\sqrt{d_N}}.\end{aligned}$$

We thus obtain the desired result for $d_N \geq D$, where D is any constant strictly larger than $D_0 = 4C_0^2/\phi^2(1/2)$, and $C_1 = \phi(1/2) - 2C_0/\sqrt{D} = \phi(1/2)(1 - \sqrt{D_0/D})$.

For completeness, we state an alternative lower bound result, based on [16]. As claimed in the lemma below, the constant involved can be straightforwardly evaluated.

Lemma 4. *Suppose that Theorem 4's assumptions are fulfilled and that there exists $c > 0$ such that $p_i \geq cn/N$ for all $i \in \{1, \dots, N\}$. Then, we have*: $\forall N \geq 1$, $\forall n < N$,

$$\mathbb{P}\{\mathcal{M}_N = 0\} \geq \frac{e^{-1/6}\sqrt{c}}{\sqrt{2\pi d_N}}. \tag{27}$$

Proof. The argument relies on the binomial approximation of the distribution of the number of successes in independent and possibly non identically distributed Bernoulli trials. Denote by $\mathcal{B}_N$ a binomial r.v. of size N and parameter $\sum_{i=1}^N p_i/N = n/N$. Applying Theorem A in [16] twice and next Stirling's formula (*i.e.*, for all $k \geq 1$, $\sqrt{2\pi k}(k/e)^k \leq k! \leq \sqrt{2\pi k}(k/e)^k \exp(1/(12k))$), one obtains that: $\forall n < N$,

$$\begin{aligned}\mathbb{P}\{\mathcal{M}_N = 0\} &= \mathbb{P}\left\{\sum_{i=1}^N \epsilon_i \leq n\right\} - \mathbb{P}\left\{\sum_{i=1}^N \epsilon_i \leq n-1\right\} \geq \mathbb{P}\{\mathcal{B}_N = n\} \\ &\geq \binom{N}{n} \times \left(\frac{n}{N}\right)^n \times \left(1 - \frac{n}{N}\right)^n \geq \frac{\exp(-\frac{1}{12n} - \frac{1}{12(N-n)})}{\sqrt{2\pi n(1-n/N)}} \\ &\geq \frac{e^{-1/6}}{\sqrt{2\pi d_N}} \times \sqrt{\frac{d_N}{n(1-n/N)}} \geq \frac{e^{-1/6}\sqrt{c}}{\sqrt{2\pi d_N}}.\end{aligned}$$

□

Remark 3. The constant C_1 obtained above is valid for any sampling distribution but is clearly not optimal in specific cases. In the case when all the inclusion probabilities are equal to n/N, an estimate of $\mathbb{P}_N\{\mathcal{M}_N = 0\}$ can be obtained by Stirling formula

$$\frac{1}{\sqrt{2\pi n}} \frac{1}{\sqrt{1 - \frac{n}{N}}} \exp\left(-\frac{1}{12n} - \frac{1}{12(N-n)}\right) \leq \mathbb{P}_N\{\mathcal{M}_N = 0\} \leq \frac{1}{\sqrt{2\pi n}} \frac{1}{\sqrt{1 - \frac{n}{N}}} \exp\left(\frac{1}{12N}\right).$$

The constant C appearing in the bounds of Theorems 4 and 5 is precisely equal to the ratio between the bounds due to the technic of the proof. In this case it is equal to $\exp(1/(12N) + 1/(12n) + 1/(12(N-n)))$ which is less than 1.24 (for $N = 2$ and $n = 1$) and its is very close to 1.001 for large N and $n \geq 10$. Under the more general assumption that there exists constants $0 < c < 1 < c' < \infty$ such that $cn/N \leq p_i \leq c'n/N < 1$, then for all $i \in \{1, \dots, N\}$ a bound for $\mathbb{P}_N\{\mathcal{M}_N = 0\}$ can be obtained by means of results of [28] and our estimate in Lemma 6

$$\frac{e^{-1/6}\sqrt{c}}{\sqrt{2\pi d_N}} \leq \mathbb{P}_N\{\mathcal{M}_N = 0\} \leq \min\left(\binom{N}{n} \Pi_{i=1}^N (1-p_i)\left(\frac{1}{N}\sum_{i=1}^N p_i/(1-p_i)\right)^n, \frac{4.838}{\sqrt{2\pi d_N}}\right).$$

A.1. Proof of Lemma 2

Recall first that we set $Z_i = (\epsilon_i - p_i)(x_i/p_i - \theta_N)$ for $1 \le i \le N$ with $\theta_N = \sum_{i=1}^N x_i(1-p_i)/d_N$ and observe that

$$\mathrm{Var}\left(\sum_{i=1}^N Z_i\right) = \sum_{i=1}^N \mathrm{Var}(Z_i) = \mathrm{Var}\left(\sum_{i=1}^N \epsilon_i^* \frac{x_i}{p_i}\right) = \mathrm{Var}\big(\widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*}\big). \tag{28}$$

Let $\psi_N(u) = \log \mathbb{E}[\exp(\langle u, (\mathcal{Z}_N, \mathcal{M}_N)\rangle)]$, $u = (u_1, u_2) \in \mathbb{R}^+ \times \mathbb{R}$, be the log-Laplace of the 1-lattice random vector $(\mathcal{Z}_N, \mathcal{M}_N)$, where $\langle \cdot, \cdot \rangle$ is the usual scalar product on $\mathbb{R}^2$. Recall that, as the logarithm of a moment-generating function, ψ_N is convex. Denote by $\psi_N^{(1)}(u)$ and $\psi_N^{(2)}(u)$ its gradient and its Hessian matrix respectively. Consider now the probability measure $\mathbb{P}_{u,N}$ defined by the Esscher transform

$$d\mathbb{P}_{u,N} = \exp\big(\big\langle u, (\mathcal{Z}_N, \mathcal{M}_N)\big\rangle - \psi_N(u)\big)\, d\mathbb{P}. \tag{29}$$

The $\mathbb{P}_{u,N}$-expectation is denoted by $\mathbb{E}_{u,N}[\cdot]$, the covariance matrix of a $\mathbb{P}_{u,N}$-square integrable random vector Y under $\mathbb{P}_{u,N}$ by $\mathrm{Var}_{u^*,N}(Y)$. With $x = t\sqrt{n}/N$, by exponential change of probability measure, we can rewrite the numerator of (20) as

$$\begin{aligned}\mathbb{P}\{\mathcal{Z}_N \ge x, \mathcal{M}_N = 0\} &= \mathbb{E}_{u,N}\big[e^{\psi_N(u) - \langle u, (\mathcal{Z}_N, \mathcal{M}_N)\rangle} \mathbb{I}\{\mathcal{Z}_N \ge x, \mathcal{M}_N = 0\}\big] \\ &= H(u)\mathbb{E}_{u,N}\big[e^{-\langle u, (\mathcal{Z}_N - x, \mathcal{M}_N)\rangle} \mathbb{I}\{\mathcal{Z}_N \ge x, \mathcal{M}_N = 0\}\big],\end{aligned}$$

where we set $H(u) = \exp(-\langle u, (x, 0)\rangle + \psi_N(u))$. Now, as ψ_N is convex, the point defined by

$$u^* = \big(u_1^*, 0\big) = \arg \sup_{u \in \mathbb{R}_+ \times \{0\}} \big\{\big\langle u, (x, 0)\big\rangle - \psi_N(u)\big\}$$

is such that $\psi_N^{(1)}(u^*) = (x, 0)$. Since $\mathbb{E}[\exp(< u, (\mathcal{Z}_N, \mathcal{M}_N) >)] = \exp(\psi_N(u))$, by differentiating w.r.t. u one gets

$$\mathbb{E}\big[e^{\langle u, (\mathcal{Z}_N, \mathcal{M}_N)\rangle} (\mathcal{Z}_N, \mathcal{M}_N)\big] = \psi_N^{(1)}(u) e^{\psi_N(u)}.$$

Taking next $u = u^*$ yields

$$\mathbb{E}\big[e^{\langle u^*, (\mathcal{Z}_N, \mathcal{M}_N)\rangle} (\mathcal{Z}_N, \mathcal{M}_N)\big] = (x, 0) e^{\psi_N(u^*)}. \tag{30}$$

Since $u^* = (u_1^*, 0)$ with $u_1^* \ge 0$, we have, under the condition $\mathcal{Z}_N \ge x$, $e^{-\langle u^*, (\mathcal{Z}_N - x, \mathcal{M}_N\rangle} \le 1$ and the straightforward bound

$$\mathbb{E}_{u^*,N}\big[e^{-\langle u^*, (\mathcal{Z}_N - x, \mathcal{M}_N)\rangle} \mathbb{I}\{\mathcal{Z}_N \ge x, \mathcal{M}_N = 0\}\big] \le \mathbb{P}_{u^*,N}\{\mathcal{M}_N = 0\}.$$

Hence, we have the bound:

$$\mathbb{P}\{\mathcal{Z}_N \ge x, \mathcal{M}_N = 0\} \le H\big(u^*\big) \times \mathbb{P}_{u^*,N}\{\mathcal{M}_N = 0\}. \tag{31}$$

We shall bound each factor involved in (31) separately. We start with bounding $H(u^*)$, which essentially boils down to bounding $\mathbb{E}[e^{\langle u^*, (\mathcal{Z}_N, \mathcal{M}_N)\rangle}]$.

Lemma 5. *Under Theorem 4's assumptions, we have:*

$$H(u^*) \le \exp\left(-\frac{\mathrm{Var}(\sum_{i=1}^N Z_i)}{X_N^2} h\left(\frac{N}{\sqrt{n}} \frac{x X_N}{\mathrm{Var}(\sum_{i=1}^N Z_i)}\right)\right) \tag{32}$$

$$\le \exp\left(-\frac{N^2 x^2/n}{2(\mathrm{Var}(\sum_{i=1}^N Z_i) + \frac{1}{3}\frac{N}{\sqrt{n}} x X_N)}\right), \tag{33}$$

where $h(x) = (1+x)\log(1+x) - x$ *for* $x \ge 0$.

Proof. Recall first that we set $\mathcal{Z}_N = (n^{1/2}/N)\sum_{1\le i\le N} Z_i$. Using the standard argument leading to the Bennett–Bernstein bound, observe that: $\forall i \in \{1,\dots,N\}$, $\forall u_1 > 0$,

$$\mathbb{E}\left[e^{u_1 Z_i}\right] \le \exp\left(\mathrm{Var}(Z_i)\frac{\exp(u_1 X_N) - 1 - u_1 X_N}{X_N^2}\right).$$

since we almost-surely have $|Z_i| \le 2\max_{1\le j\le N}|x_j|/p_j = X_N$ for all $i \in \{1,\dots,N\}$. Using the independence of the Z_i's, we obtain that: $\forall u_1 > 0$,

$$\mathbb{E}\left[e^{u_1 \mathcal{Z}_N}\right] \le \exp\left(\mathrm{Var}\left(\sum_{i=1}^N Z_i\right)\frac{\exp(\frac{\sqrt{n}}{N}u_1 X_N) - 1 - \frac{\sqrt{n}}{N}u_1 X_N}{X_N^2}\right).$$

The resulting upper bound for $H((u_1,0))$ being minimized for

$$u_1 = \frac{N}{\sqrt{n}} \frac{\log(1 + \frac{N}{\sqrt{n}}\frac{x X_N}{\mathrm{Var}(\sum_{i=1}^N Z_i)})}{X_N},$$

this yields

$$H(u^*) \le \exp\left(-\frac{\mathrm{Var}(\sum_{i=1}^N Z_i)}{X_N^2} h\left(\frac{N}{\sqrt{n}} \frac{x X_N}{\mathrm{Var}(\sum_{i=1}^N Z_i)}\right)\right). \tag{34}$$

Using the classical inequality

$$h(x) \ge \frac{x^2}{2(1+x/3)} \quad \text{for } x \ge 0,$$

we also get that

$$H(u^*) \le \exp\left(-\frac{N^2 x^2/n}{2(\mathrm{Var}(\sum_{i=1}^N Z_i) + \frac{1}{3}\frac{N}{\sqrt{n}} x X_N)}\right). \qquad \square$$

We now prove the lemma stated below, which provides an upper bound for $\mathbb{P}_{u^*,N}\{\mathcal{M}_N = 0\}$.

Lemma 6. *Under Theorem 4's assumptions, there exists a universal constant* $C' \leq 4.838$ *such that*: $\forall N \geq 1$,

$$\mathbb{P}_{u^*,N}\{\mathcal{M}_N = 0\} \leq C' \frac{1}{\sqrt{2\pi d_N}}. \tag{35}$$

Proof. Under the probability measure $\mathbb{P}_{u^*,N}$, the ε_i's are still independent Bernoulli variables, with means given by

$$\pi_i^* \stackrel{\text{def}}{=} \sum_{s\in\mathcal{P}(\mathcal{I}_N)} e^{\langle u^*,(\mathcal{Z}_N(s),\mathcal{M}_N(s))\rangle-\psi_N(u^*)}\mathbb{I}\{i \in s\}R_N(s) > 0,$$

for $i \in \{1, \dots, N\}$. Since $\mathbb{E}_{u^*,N}[\mathcal{M}_N] = 0$, we have $\sum_{i=1}^N \pi_i^* = n$ and thus

$$d_{N,u^*} \stackrel{\text{def}}{=} \operatorname{Var}_{u^*,N}\left(\sum_{i=1}^N \varepsilon_i\right) = \sum_{i=1}^N \pi_i^*\left(1-\pi_i^*\right) \leq n.$$

Now, applying twice the Berry–Esseen bound to the sum of centered independent (and possibly non-identically) Bernoulli random variables $\sum_{i=1}^N m_i$ yields

$$\begin{aligned}\mathbb{P}_{u^*,N}\{\mathcal{M}_N = 0\} &= \mathbb{P}_{u^*,N}\left\{d_{N,u^*}^{-1/2}\sum_{i=1}^N m_i \leq 0\right\} - \mathbb{P}_{u^*,N}\left\{d_{N,u^*}^{-1/2}\sum_{i=1}^N m_i \leq -d_{N,u^*}^{-1/2}\right\} \\ &\leq \frac{2C''\sum_{i=1}^N \mathbb{E}_{u^*,N}[|\epsilon_i - \pi_i^*|^3]}{d_{N,u^*}^{3/2}} + \frac{1}{\sqrt{2\pi d_{N,u^*}}} \leq \left(\frac{1}{\sqrt{2\pi}} + 2C''\right)\frac{1}{\sqrt{d_{N,u^*}}},\end{aligned}$$

where $C'' \leq 0.7655$, see [33]. Finally, observe that

$$d_{N,u^*} = \mathbb{E}_{u^*,N}\left[\left(\sum_{i=1}^N m_i\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^N m_i\right)^2 \Big/ H\left(u^*\right)\right] \geq \mathbb{E}\left[\left(\sum_{i=1}^N m_i\right)^2\right] = d_N,$$

since we proved that $H(u^*) \leq 1$. Combined with the previous bound and the fact that we assumed $d_N \geq 1$, this yields the desired result. □

Lemmas 5 and 6 combined with Eq. (31) leads to the bound stated in Lemma 2.

Proof of Theorem 5

We start with proving the preliminary result below.

Lemma 7. *Let* $\pi_1, \dots, \pi_N$ *be the first order inclusion probabilities of a rejective sampling of size* n *with canonical representation characterized by the Poisson weights* $p_1, \dots, p_N$. *Provided* *that* $d_N = \sum_{i=1}^{N} p_i(1-p_i) > 4$, *we have*: $\forall i \in \{1,\dots,N\}$,

$$\left|\frac{1}{\pi_i} - \frac{1}{p_i}\right| \leq \frac{4}{d_N} \times \frac{1-\pi_i}{\pi_i}.$$

Proof. The proof is based on the representation (5.14) on page 1509 of [23]. For all $i \in \{1,\dots,N\}$, we have:

$$\frac{\pi_i}{p_i}\frac{1-p_i}{1-\pi_i} = \frac{\sum_{s\in\mathcal{P}(\mathcal{I}_N):i\in\mathcal{I}_N\setminus\{s\}} P_N(s)\sum_{h\in s}\frac{1-p_h}{\sum_{j\in s}(1-p_j)+(p_h-p_i)}}{\sum_{s\in\mathcal{P}(\mathcal{I}_N):i\in\mathcal{I}_N\setminus\{s\}} P_N(s)}$$

$$= \frac{\sum_{s:i\in\mathcal{I}_N\setminus\{s\}} P_N(s)\sum_{h\in s}\frac{1-p_h}{\sum_{j\in s}(1-p_j)(1+\frac{(p_h-p_i)}{\sum_{j\in s}(1-p_j)})}}{\sum_{s:i\in\mathcal{I}_N\setminus\{s\}} P_N(s)}.$$

Now recall that for any $x \in]-1/2, 1[$, we have:

$$1 - x \leq \frac{1}{1+x} \leq 1 - x + 2x^2.$$

Applying this to $x = (p_h - p_i)/\sum_{j\in s}(1-p_j)$ for all $h \in s$, which can be seen to belong to $]-1/2,+1[$ by noticing that $\sum_{j\in s}(1-p_j) \geq d_N/2 > 2$ as soon as $d_N > 4$ by virtue of Lemma 2.2 in [23], we obtain that

$$\frac{\pi_i}{p_i}\frac{1-p_i}{1-\pi_i} \leq 1 - \left(\sum_{s:i\in\mathcal{I}_N\setminus\{s\}} P_N(s)\right)^{-1} \sum_{s:i\in\mathcal{I}_N\setminus\{s\}} P_N(s)\sum_{h\in s}\frac{(1-p_h)(p_h-p_i)}{(\sum_{j\in s}(1-p_j))^2}$$
$$+ 2\left(\sum_{s:i\in\mathcal{I}_N\setminus\{s\}} P_N(s)\right)^{-1} \sum_{s:i\in\mathcal{I}_N\setminus\{s\}} P_N(s)\sum_{h\in s}\frac{(1-p_h)(p_h-p_i)^2}{(\sum_{j\in s}(1-p_j))^3}.$$

Following now line by line the proof on p. 1510 in [23] and using the fact that $\sum_{j\in s}(1-p_j) \geq d_N/2$, we get

$$\left|\sum_{h\in s}\frac{(1-p_h)(p_h-p_i)}{(\sum_{j\in s}(1-p_j))^2}\right| \leq \frac{1}{(\sum_{j\in s}(1-p_j))} \leq \frac{2}{d_N},$$

and similarly

$$\sum_{h\in s}\frac{(1-p_h)(p_h-p_i)^2}{(\sum_{j\in s}(1-p_j))^3} \leq \frac{1}{(\sum_{j\in s}(1-p_j))^2} \leq \frac{4}{d_N^2}.$$

This yields: $\forall i \in \{1,\dots,N\}$,

$$1 - \frac{2}{d_N} \leq \frac{\pi_i}{p_i}\frac{1-p_i}{1-\pi_i} \leq 1 + \frac{2}{d_N} + \frac{8}{d_N^2}$$

and

$$p_i(1-\pi_i)\left(1-\frac{2}{d_N}\right) \le \pi_i(1-p_i) \le p_i(1-\pi_i)\left(1+\frac{2}{d_N}+\frac{8}{d_N^2}\right),$$

leading then to

$$-\frac{2}{d_N}(1-\pi_i)p_i \le \pi_i - p_i \le p_i(1-\pi_i)\left(\frac{2}{d_N}+\frac{8}{d_N^2}\right)$$

and finally to

$$-\frac{(1-\pi_i)}{\pi_i}\frac{2}{d_N} \le \frac{1}{p_i}-\frac{1}{\pi_i} \le \frac{(1-\pi_i)}{\pi_i}\left(\frac{2}{d_N}+\frac{8}{d_N^2}\right).$$

Since $1/d_N^2 \le 4/d_N$ as soon as $d_N \ge 4$, the lemma is proved. □

By virtue of Lemma 7, we obtain that:

$$\left|\widehat{S}_{\boldsymbol{\pi}_N}^{\boldsymbol{\epsilon}_N^*} - \widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*}\right| \le \frac{4}{d_N}\sum_{i=1}^{N}\frac{1}{\pi_i}|x_i| = M_N.$$

It follows that

$$\mathbb{P}\left\{\widehat{S}_{\boldsymbol{\pi}_N}^{\boldsymbol{\epsilon}_N^*} - S_N > x\right\} \le \mathbb{P}\left\{\left|\widehat{S}_{\boldsymbol{\pi}_N}^{\boldsymbol{\epsilon}_N^*} - \widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*}\right| + \widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*} - S_N > x\right\} \le \mathbb{P}\left\{M_N + \widehat{S}_{\boldsymbol{p}_N}^{\boldsymbol{\epsilon}_N^*} - S_N > x\right\}.$$

A direct application of Theorem 4 finally gives the desired result.

On the constant C_0 in Theorem 6

As an examination of the proof of Theorem 4 shows, an evaluation of the constants appearing in the tail bounds we establish here can be deduced from that of the constant C_0 involved in Theorem 6 (Theorem 1.3 in [17]). Precisely, we prove below that one may choose the value

$$C_0 = 923.12.$$

For clarity, we use exactly the same notations and numbering as those in [17]. Observe first that one may choose $\alpha = 0.61$ in Lemma 3.2 of [17] and that, in Lemma 3.3, the crude asymptotic expansion can be replaced by the more accurate bound

$$\sup_x\left|\Phi(\gamma x) - \Phi(x)\right| \le |\gamma - 1|.$$

In order to see this, differentiate w.r.t. γ, so as to establish that: $\forall\gamma \in [1/2, 1]$, $\sup_x|\Phi(\gamma x) - \Phi(x)| \le 2|\gamma - 1|/\sqrt{2\pi e}$. Similarly, differentiating w.r.t. x yields

$$\sup_x \frac{\gamma\phi(x\gamma) - \phi(x)}{\gamma - 1} \le \frac{2}{\sqrt{2\pi}} \quad \text{and} \quad \sup_x \frac{x^2|\gamma^3\phi(x\gamma) - \phi(x)|}{\gamma - 1} \le \frac{2}{\sqrt{2\pi}}.$$

These inequalities imply that one may choose $M_1 = 1$ and $M_2 = 2/(6\sqrt{2\pi})$ in Lemma 3.4 of [17], so that taking $c = 0.61$ and $d = 1$ as constants in the beginning of the proof of Corollary 1.3 is fair.

Considering now the constants in Theorem 1.3 (see the proof on p. 290 of [17]), notice that a straightforward computation, *that is*, computing the derivatives of $\Gamma_n(x) = \Phi(x) + ((1 - x^2)/(6\sigma_n^3))\phi(x)\sum_{j=1}^n \mathbb{E}[(Y_j, n - q_{j,n})^3]$, shows that, In Lemma 4.1, we can take

$$\rho = \frac{1}{\sqrt{2\pi e}} \quad \text{and} \quad E = \frac{\rho}{24},$$

since we have for all $n \geq 2$:

$$\sup_x \Gamma_n''(x) \leq \frac{1}{\sqrt{2\pi e}}(1 + 1/\sigma_n).$$

In Lemma 4.2, by computing the second order derivative, we obtain that

$$\sup_x \Gamma_n''(x) \leq \frac{1}{\sqrt{2\pi}} + \frac{1}{6\sigma_n} \times k_1$$

with

$$\exp\big((-3 + \sqrt{6})/2\big)\sqrt{\frac{3(3 - \sqrt{6})}{\pi}} \leq 0.051.$$

Hence, one can take $K = 1/\sqrt{2\pi}$ in order to control the term I_2 on p. 292 in [17]. We prove now that a crude evaluation of L in Lemma 4.3 is given by

$$L \leq 698.786 < 699.$$

In order to see this, take $q = 1/2$ in Lemma 4.4 (notice incidentally that this value is by no means optimal but permits to carry out all the calculations explicitly). Following the proof line by line, one can take $M = e^4/(\sqrt{2\pi}(1 - q)) \leq 43.57$. In this case, on p. 294 we can choose any δ such that

$$\delta \leq 0.0305637 \simeq \min\left\{\frac{e^{-1}}{2}, \frac{1}{4(1/6 + e^4 e^{-1}/\sqrt{2\pi})}\right\},$$

yielding a crude evaluation of $R = 2141.1$ and the bound

$$A_n \leq 2141.1/\sigma_n.$$

A straightforward integral computation shows that

$$\alpha_n \int_{\alpha_n}^{\infty} t^2 e^{-t^2/2} \leq 1.05,$$

so that we can take $R = 1.05$ in (4.20), which leads to

$$B_n \leq \frac{1.05}{0.0305}\sigma_n^{-2} \leq 34.355\sigma_n^{-2}.$$

We can now evaluate the terms in (4.22) by noticing that

$$\delta^{-1} + \frac{1}{4\pi}\sum_{i=1}^{\infty} l^{-2} \leq 1070{,}505 + \frac{\pi}{24} \leq 1070.64.$$

Taking now $M = \frac{e^4}{\sqrt{2\pi}}\frac{1}{1-q}$ in (4.25) and choosing $d = \delta = 0.0305637$ yields

$$E_n \leq \frac{2.74}{\sigma_n^2}$$

in (4.28), by observing that

$$\int_0^{2\pi} \sin(y/2)\,dy = 4 \quad \text{and} \quad \int_0^{\infty} v \exp\bigl(-v^2/4\bigr)\,dv = 2.$$

Next, by combining (4.21)–(4.29) with the constants above, we get $Q = 2181$ in Lemma 4.5. In their turn, these estimates permit to get

$$I_2 \leq 9.58\sigma_n^{-2}(1 + 1/\sigma_n)$$

and, by combining (4.14)–(4.17) with (4.13),

$$I_{11} \leq \sigma_n^{-2} 2(8/9 + 8M) \leq 699\sigma_n^{-2}.$$

Using Lemma 4.5, we then obtain

$$I_1 \leq I_{11} + I_{12} \leq (2181 + 699)\sigma_n^{-2} = 2880\sigma_n^{-2},$$

that gives

$$I_1 + I_2 \leq 2900\sigma_n^{-2}.$$

From these estimates, we next deduce that, in (4.3),

$$\bigl|\widetilde{F}_n(x) - \widetilde{\Gamma}_n(x)\bigr| \leq \frac{2900}{\pi}\sigma_n^{-2}$$

and finally that

$$\bigl|\widetilde{F}_n(x) - \Gamma_n(x)\bigr| \leq \frac{2900}{\pi}\sigma_n^{-2} + \frac{1}{24\sqrt{2\pi e}}\sigma_n^{-2}\bigl(1 + \sigma_n^{-1}\bigr) \leq 923.12\sigma_n^{-2},$$

which is the desired result.

Acknowledgements

We would like to thank the referees and the Associate Editor for their insightful comments, that greatly improved the presentation of the results. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01) for P. Bertail.

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] Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson Approximation. Oxford Studies in Probability* **2**. New York: Oxford University Press. MR1163825

[3] Bardenet, R. and Maillard, O.-A. (2015). Concentration inequalities for sampling without replacement. *Bernoulli* **21** 1361–1385. MR3352047

[4] Bennett, G. (1962). Probability inequalities for a sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

[5] Berger, Y.G. (1998). Rate of convergence to normal distribution for the Horvitz–Thompson estimator. *J. Statist. Plann. Inference* **67** 209–226. MR1624693

[6] Bernstein, S.N. (1964). On a modification of Chebyshev's inequality and on the error in Laplace formula. *Collected Works*, *Izd-vo 'Nauka'*, *Moscow* (*in Russian*) **4** 71–80.

[7] Bertail, P., Chautru, E. and Clémençon, S. (2017). Empirical processes in survey sampling with (conditional) Poisson designs. *Scand. J. Stat.* **44** 97–111. MR3619696

[8] Bertail, P. and Clémençon, S. (2010). Sharp bounds for the tails of functionals of Markov chains. *Theory Probab. Appl.* **54** 505–515.

[9] Boistard, H., Lopuhaä, H.P. and Ruiz-Gazen, A. (2012). Approximation of rejective sampling inclusion probabilities and application to high order correlations. *Electron. J. Stat.* **6** 1967–1983. MR3020253

[10] Boistard, H., Lopuhaä, H.P. and Ruiz-Gazen, A. (2017). Functional central limit theorems for single-stage sampling designs. *Ann. Statist.* **45** 1728–1758. MR3670194

[11] Bondesson, L., Traat, I. and Lundqvist, A. (2006). Pareto sampling versus Sampford and conditional Poisson sampling. *Scand. J. Stat.* **33** 699–720. MR2300911

[12] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities*: *A Nonasymptotic Theory of Independence*. Oxford: Oxford Univ. Press. MR3185193

[13] Brändén, P. and Jonasson, J. (2012). Negative dependence in sampling. *Scand. J. Stat.* **39** 830–838. MR3000852

[14] Kramer, J.B., Cutler, J. and Radcliffe, A.J. (2011). Negative dependence and Srinivasan's sampling process. *Combin. Probab. Comput.* **20** 347–361. MR2784632

[15] Chen, X.-H., Dempster, A.P. and Liu, J.S. (1994). Weighted finite population sampling to maximize entropy. *Biometrika* **81** 457–469. MR1311090

[16] Choi, K.P. and Xia, A. (2002). Approximating the number of successes in independent trials: Binomial versus Poisson. *Ann. Appl. Probab.* **12** 1139–1148. MR1936586

[17] Deheuvels, P., Puri, M.L. and Ralescu, S.S. (1989). Asymptotic expansions for sums of nonidentically distributed Bernoulli random variables. *J. Multivariate Anal.* **28** 282–303. MR0991952

[18] Deville, J.C. (1987). Réplications d'échantillons, demi-échantillons, Jackknife, bootstrap dans les sondages. In *Les Sondages*. Economica, Ed. Droesbeke, Tassi, Fichet.

[19] Dupačová, J. (1979). A note on rejective sampling. In *Contributions to Statistics* 71–78. Dordrecht-Boston, Mass.-London: Reidel. MR0561260
[20] Fuk, D.H. and Nagaev, S.V. (1971). Probabilistic inequalities for sums of independent random variables. *Theory Probab. Appl.* **16** 643–660.
[21] Fuller, W.A. (2009). Some design properties of a rejective sampling procedure. *Biometrika* **96** 933–944. MR2564501
[22] Goodman, L.A. (1949). On the estimation of the number of classes in a population. *Ann. Math. Stat.* **20** 572–579. MR0032165
[23] Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Stat.* **35** 1491–1523. MR0178555
[24] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363
[25] Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460
[26] Janson, S. (1994). Large deviation inequalities for sums of indicator variables. Unpublished manuscript. Available at www2.math.uu.se/~svante/papers/sj107.ps.
[27] Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables, with applications. *Ann. Statist.* **11** 286–295. MR0684886
[28] Percus, O.E. and Percus, J.K. (1985). Probability bounds on the sum of independent nonidentically distributed binomial random variables. *SIAM J. Appl. Math.* **45** 621–640. MR0796099
[29] Robinson, P.M. (1982). On the convergence of the Horvitz–Thompson estimator. *Aust. J. Stat.* **24** 234–238. MR0678263
[30] Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. *Springer Series in Statistics*. New York: Springer. MR1140409
[31] Serfling, R.J. (1974). Probability inequalities for the sum in sampling without replacement. *Ann. Statist.* **2** 39–48. MR0420967
[32] Shao, Q.-M. (2000). A comparison theorem on moment inequalities between negatively associated and independent random variables. *J. Theoret. Probab.* **13** 343–356. MR1777538
[33] Shiganov, I.S. (1986). Refinement of the upper bound of the constant in the central limit theorem. *J. Sov. Math.* **35** 2545–2550.
[34] Talagrand, M. (1995). The missing factor in Hoeffding's inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.* **31** 689–702. MR1355613
[35] Tan, Z. (2013). Simple design-efficient calibration estimators for rejective and high-entropy sampling. *Biometrika* **100** 399–415. MR3068442
[36] Tillé, Y. (2006). *Sampling Algorithms*. *Springer Series in Statistics*. New York: Springer. MR2225036

Received July 2017 and revised September 2018