# TAIL INDEX ESTIMATION BASED ON SURVEY DATA

PATRICE BERTAIL<sup>1,2</sup>, EMILIE CHAUTRU<sup>3</sup> AND STÉPHAN CLÉMENÇON<sup>4</sup>

**Abstract.** This paper is devoted to tail index estimation in the context of survey data. Assuming that the population of interest is described by a heavy-tailed statistical model, we prove that the survey scheme plays a crucial role in the design of consistent inference methods for extremes. As can be revealed by simulation experiments, ignoring the sampling plan generally induces a significant bias, jeopardizing the accuracy of the extreme value statistics thus computed. Focus is here on the celebrated Hill method for tail index estimation, it is shown how to modify it in order to take into account the survey design. Precisely, under specific conditions on the inclusion probabilities of first and second orders, we establish the consistency of the variant of the Hill estimator we propose. Additionally, its asymptotic normality is proved in a specific situation. Application of this limit result for building Gaussian confidence intervals is thoroughly discussed and illustrated by numerical results.

Mathematics Subject Classification. 62D05, 62F12, 62G32.

Received August 12, 2013. Revised February 5, 2014.

## 1. INTRODUCTION

It is the main purpose of this paper to study the impact of a survey sampling scheme on tail index estimation. Indeed, in many situations, statisticians have at their disposal not only data but also weights arising from some survey sampling stratification. These weights correspond either to true inclusion probabilities, as is often the case for institutional data, or to some calibrated or post-stratification weights (minimizing some discrepancy with the inclusion probabilities subject to some margin constraints). In most cases, the survey design is ignored; such an omission can induce a significant sampling bias on the computed statistics [5]. When considering statistics of extremes in particular, this may cause severe drawbacks and completely jeopardize the estimation, as can be revealed by simulation experiments. Whereas asymptotic analysis of the Horvitz–Thompson estimator [27], in the context of mean estimation and regression in particular, has been the subject of much attention (see [2, 16, 23, 24, 31, 32] for instance) and the last few years have witnessed significant progress towards a comprehensive functional limit theory for distribution function estimation (refer to [6–8, 18, 33]), no result on tail estimation has been documented in the survey sampling literature yet.

Keywords and phrases. Survey sampling, tail index estimation, Hill estimator, Poisson survey scheme, rejective sampling.

<sup>&</sup>lt;sup>1</sup> MODAL'X - Université Paris Ouest, 92001 Nanterre, France

<sup>&</sup>lt;sup>2</sup> Laboratoire de Statistique, CREST, France

<sup>&</sup>lt;sup>3</sup> Laboratoire AGM - Université de Cergy-Pontoise, 95000 Cergy-Pontoise, France

 $<sup>^4\,</sup>$ Institut Mines-Télécom - LTCI UMR Télécom Paris<br/>Tech/CNRS No. 5141, 75634 Paris, France.

 $<sup>{\</sup>tt stephan.clemencon@telecom-paristech.fr}$ 

Our goal here is to show how to incorporate the survey scheme into extreme value statistical techniques, in order to guarantee their asymptotic validity. Our approach is illustrated through the tail index estimation problem in the context of heavy-tailed survey data. We propose a specific modification of the Hill estimator, accounting for the survey plan by means of which data have been collected, and establish its consistency under adequate assumptions on the probabilities of inclusion of first and second orders. Following in the footsteps of [24], its asymptotic normality is also investigated in the particular situations of Poisson and rejective schemes. Based on this limit result, the issue of building Gaussian confidence intervals for the tail index is next considered.

The rest of the paper is structured as follows. Basics about survey sampling and tail index estimation in the standard i.i.d. setup are briefly recalled in Section 2. Section 3 describes at length the modification of the Hill estimator we propose in the context of a general sampling plan and proves its consistency. In Section 4, its asymptotic normality is established and the main asymptotic results of the paper are stated. Possible heuristic methods for selecting the number of observations that should be ideally involved in the computation of the modified Hill estimator are discussed in Section 5, together with illustrative numerical experiments. Technical proofs are deferred to the Appendix section.

## 2. Background and preliminaries

We first recall some crucial notions in survey sampling, which shall be extensively used in the subsequent analysis. Basic concepts of heavy-tail modeling, including Pareto-type distributions and standard strategies for statistical estimation of the related parameters are next briefly described for the sake of clarity. Throughout the article, the Dirac mass at  $x \in \mathbb{R}$  is denoted by  $\delta_x$ , the indicator function of any event  $\mathcal{E}$  by  $\mathbb{I}{\mathcal{E}}$ . The (left-continuous) inverse of any nondecreasing function  $H : (a, b) \to \mathbb{R}$ , where  $-\infty \leq a < b \leq +\infty$ , is denoted by  $H^{\leftarrow}(t) = \inf\{s \in (a, b) : H(s) \geq t\}, t \in \mathbb{R}$ , with the convention that the infimum over an empty set is  $-\infty$ . The minimum of two real numbers x and y is denoted by  $x \wedge y$  and the maximum by  $x \vee y$ . The expectation of an integrable random variable Z is denoted by  $\mathbb{E}(Z)$  and, when it is square integrable, its variance is denoted by  $\mathbb{V}(Z)$ .

# 2.1. Survey sampling

Here and throughout, we consider a finite population of size  $N \geq 1$  referred to as  $\mathcal{U}_N := \{1, \ldots, N\}$ and denote by  $\mathcal{P}(\mathcal{U}_N)$  its power set. We call a *sample* of (possibly random) size  $n \leq N$ , any subset  $s := \{i_1, \ldots, i_{n(s)}\} \in \mathcal{P}(\mathcal{U}_N)$  with cardinality n =: n(s) less than N. A sampling scheme (design/plan) without replacement is determined by a probability distribution  $R_N$  on the set of all possible samples  $s \in \mathcal{P}(\mathcal{U}_N)$ . For any  $i \in \{1, \ldots, N\}$ , the following quantity, generally called (first order) *inclusion probability*,

$$\pi_i(R_N) := \mathbb{P}_{R_N} \left( i \in S \right),$$

is the probability that the unit *i* belongs to a random sample *S* drawn from distribution  $R_N$ . In vectorial form, we shall write  $\pi(R_N) := (\pi_1(R_N), \ldots, \pi_N(R_N))$ . First order inclusion probabilities are assumed to be strictly positive in the subsequent analysis:  $\forall i \in \{1, \ldots, n\}, \pi_i(R_N) > 0$ . Additionally, the second order inclusion probabilities are denoted by

$$\pi_{i,j}(R_N) := \mathbb{P}_{R_N}\left((i,j) \in S^2\right),\,$$

for any  $i \neq j$  in  $\{1, \ldots, N\}^2$ . When no confusion is possible, we shall fail to mention the dependence in  $R_N$  when writing the first/second order probabilities of inclusion. The information related to the observed sample  $S \subset \{1, \ldots, N\}$  is encapsulated by the random vector  $\boldsymbol{\epsilon} := (\epsilon_1, \ldots, \epsilon_N)$ , where

$$\epsilon_i = \begin{cases} 1 \text{ if } i \in S, \\ 0 \text{ otherwise.} \end{cases}$$

The distribution of the sampling scheme  $\epsilon$  has 1-d marginals that correspond to the Bernoulli distributions  $\mathcal{B}(\pi_i)$ ,  $1 \leq i \leq N$ , and covariance matrix given by  $\{\pi_{i,j} - \pi_i \pi_j\}_{1 \leq i,j \leq N}$ . Notice incidentally that, equipped with these notations, we have  $\sum_{i=1}^{N} \epsilon_i = n(S)$ .

**Example 1** (Poisson survey sampling). Though of extreme simplicity, the Poisson scheme (without replacement) plays a crucial role in sampling theory, insofar as it can be used to approximate a wide range of survey plans. This is indeed a key observation to establish general asymptotic results in the survey context, see [24] and Section 4 of the present paper. For such a plan  $T_N$ , the  $\epsilon_i$ 's are independent Bernoulli random variables with parameters  $p_1, \ldots, p_N$  in (0, 1). The first order inclusion probabilities thus characterize fully such a plan. Observe in addition that the size  $n(\mathcal{S})$  of a sample generated this way is random and goes to infinity as  $N \to +\infty$  with probability one, provided that  $\min_{1 \le i \le N} p_i$  remains bounded away from zero.

The superpopulation model we consider here stipulates that a real-valued random variable X with distribution  $\mathbb{P}$  and cdf F is observable on the population  $\mathcal{U}_N$ , *i.e.*  $X_1, \ldots, X_N$  are i.i.d. realizations drawn from  $\mathbb{P}$ . In practice, it is customary to determine the first order inclusion probabilities as a function of an *auxiliary variable*, which is observed on the entire population. Here, it is denoted by  $\mathbf{W}$  with distribution  $\mathbb{P}_{\mathbf{W}}$ . Hence, for all  $i \in \{1, \ldots, N\}$  we can write  $\pi_i = \pi(\mathbf{W}_i)$  for some link function  $\pi(.)$ . When  $\mathbf{W}$  and X are strongly correlated, thus proceeding helps select more informative samples and subsequently reduce the variance of estimators (see Sect. 4 for a more detailed discussion on the use of auxiliary information in survey sampling). One may refer to [9, 15, 22] for accounts of survey sampling techniques.

We recall that the Horvitz–Thompson estimator of the empirical measure  $\mathbb{P}_N := N^{-1} \sum_{i=1}^N \delta_{X_i}$  based on the survey data described above is defined as follows [27]:

$$\mathbb{P}_{R_N}^{\boldsymbol{\pi}(R_N)} := \frac{1}{N} \sum_{i=1}^N \frac{\epsilon_i}{\pi_i} \,\delta_{X_i} = \frac{1}{N} \sum_{i \in S} \frac{1}{\pi_i} \,\delta_{X_i},$$

where the subscript  $R_N$  stipulates that the vector  $\boldsymbol{\epsilon} := (\epsilon_1, \ldots, \epsilon_N)$  is in correspondence with a sample S drawn at random from distribution  $R_N$ , and the superscript  $\boldsymbol{\pi}(R_N)$  indicates that the inclusion probabilities used in the formula are those of the design  $R_N$ . When there is no ambiguity, we shall simplify notations and write  $\mathbb{P}_N^{\boldsymbol{\pi}}$  instead of  $\mathbb{P}_{R_N}^{\boldsymbol{\pi}(R_N)}$ . We highlight the fact that, conditionally upon all vectors  $\{(X_i, \mathbf{W}_i), 1 \leq i \leq N\}$ , the latter is an unbiased estimator of  $\mathbb{P}_N$ , although it is not a probability measure. Its (pointwise) consistency and asymptotic normality are established in [2, 31]. Limit results of functional nature are established in [18] for specific biased sampling models (see also [3, 7, 8, 33]). The weighted quantity  $F_N^{\boldsymbol{\pi}}(x) := \mathbb{P}_N^{\boldsymbol{\pi}}(-\infty, x]$  is naturally different from the empirical cumulative distribution function of the observations,  $F_n(x) := n^{-1} \sum_{i \in S} \mathbb{I}\{X_i \leq x\}$ namely, whose asymptotic behavior is investigated in [5]. It is then straightforward to deduce the following (unbiased) estimate of the probability of exceedance  $\overline{F}(x) := \mathbb{P}(X > x), x \in \mathbb{R}$ , given by:

$$\overline{F}_{N}^{\pi}(x) := \frac{1}{N} \sum_{i=1}^{N} \frac{\epsilon_{i}}{\pi_{i}} \mathbb{I}\left\{X_{i} > x\right\} = \frac{1}{N} \sum_{i \in S} \frac{1}{\pi_{i}} \mathbb{I}\left\{X_{i} > x\right\}.$$
(2.1)

### 2.2. Tail index inference – the Hill estimator

In a wide variety of situations, it is appropriate to assume that a statistical population is described by a heavy-tailed probability distribution (the field of heavy-tail analysis is well depicted in [30]). A distribution with Pareto-like right tail is any probability measure  $\mathbb{P}$  on  $\mathbb{R}$  with cdf F such that for all  $x \in \mathbb{R}$ ,

$$1 - F(x) = \overline{F}(x) = x^{-1/\gamma} L(x), \qquad (2.2)$$

where  $\gamma > 0$  is the extreme value index (EVI) of distribution  $\mathbb{P}$  and L(x) is a slowly varying function, i.e. a function such that  $L(tx)/L(x) \to 1$  as  $x \to +\infty$  for all t > 0. Notice that instead of the EVI, focus is often on  $\alpha := 1/\gamma$ , the tail index of the distribution  $\mathbb{P}$ . Functions of the form introduced in equation (2.2) are said to be regularly varying with index  $-1/\gamma$ ; the set of such functions is denoted by  $\mathcal{R}_{-1/\gamma}$ . One may refer to [4] for an account of the theory of regularly varying functions. The Hill estimator [26] provides a popular way of estimating the EVI  $\gamma$ . Its asymptotic behavior and the practical issues related to its computation are well-documented in the literature, see for instance ([30], Chap. 4) and the references therein. Given an i.i.d. population  $X_1, \ldots, X_N$  of size  $N \ge 1$  drawn from  $\mathbb{P}$  and a number  $K \in \{1, \ldots, N-1\}$  of largest observations, it is written as

$$H_{K,N} := \frac{1}{K} \sum_{i=1}^{K} \log\left(\frac{X_{N-i+1,N}}{X_{N-K,N}}\right),$$
(2.3)

where  $X_{1,N} \leq \cdots \leq X_{N,N}$  denote the order statistics related to the population. Whereas the theory has been extensively developed in the case where the observations are independent and identically distributed, including issues related to the choice of K in equation (2.3), to the best of our knowledge the Hill procedure has received no attention when data arise from a general survey. We point out that there exist alternative methods for tail index or EVI estimation, refer for instance to Chapter 4 of [1], for further details. The argument of the subsequent analysis paves the way for studying extensions of such techniques in the context of survey sampling models.

### 3. The Hill estimator in survey sampling

Placing ourselves in the framework described in Section 2.1, we shall denote by  $X_{1,n} \leq \cdots \leq X_{n,n}$  the order statistics related to the survey sample  $(X_{i_1}, \ldots, X_{i_n})$ , where n = n(S) may be random. When unit j is such that  $X_j = X_{i,N}$ , the *i*-th largest observation in the population,  $1 \leq i, j \leq N$ , its inclusion indicator and probability are denoted by  $\epsilon_{i,N} = \epsilon_j$  and  $\pi_{i,N} = \pi_j$  respectively. Similarly, we write  $\pi_{i,n} := \pi_j$  when  $X_{i,n} = X_j, 1 \leq i, j \leq n$ . As a general rule, indexes in uppercase shall designate the full population, as opposed to those in lowercase, which shall refer to the sample. We assume in addition that the distribution function F has the semi-parametric form set out in equation (2.2) with unknown parameter  $\gamma > 0$  and, for the sake of simplicity, that its support is included in  $(0, +\infty]$ . Because it is destined to be extensively used in the sequel, we also introduce the *tail quantile function*, written for all  $x \in [1, +\infty]$  as

$$U(x) := F^{\leftarrow} \left(1 - \frac{1}{x}\right) \cdot$$

Its empirical and Horvitz-Thompson equivalents are respectively denoted by

$$U_N(x) := F_N^{\leftarrow} \left(1 - \frac{1}{x}\right)$$
 and  $U_N^{\pi}(x) := (F_N^{\pi})^{\leftarrow} \left(1 - \frac{1}{x}\right)$ .

Notice that when  $F \in \mathcal{R}_{-1/\gamma}$ , the corresponding tail quantile function U is also regularly varying with index  $\gamma$  ([1], Sects. 2.3.2 and 2.9.3). The goal pursued here is to estimate the tail parameter  $\gamma$  based on the survey data  $X_{i_1}, \ldots, X_{i_n}$  and the sampling plan  $R_N$ .

### 3.1. The Horvitz–Thompson variant of the Hill estimator

Notice first that, under the heavy-tail assumption above, we have:

$$\gamma = \lim_{x \to \infty} \int_{x}^{+\infty} \frac{\overline{F}(u)}{\overline{F}(x)} \frac{\mathrm{d}u}{u},\tag{3.1}$$

see Section 2.6 of [1], for instance. In the case of the iid population  $X_1, \ldots, X_N$  drawn from  $\mathbb{P}$ , one classically recovers the celebrated Hill estimator by substituting F with the empirical cdf  $F_N$  in equation (3.1) and taking  $x = U_N(N/K) = X_{N-K,N}$  for some number  $1 \le K \le N - 1$  of largest observations, supposedly representative of the tail of the distribution. Indeed, we have:

$$\int_{X_{N-K,N}}^{+\infty} \frac{\overline{F}_N(u)}{\overline{F}_N(X_{N-K,N})} \frac{\mathrm{d}u}{u} = \sum_{i=1}^K \int_{X_{N-i+1,N}}^{X_{N-i+1,N}} \frac{\overline{F}_N(u)}{\overline{F}_N(X_{N-K,N})} \frac{\mathrm{d}u}{u}$$
$$= \sum_{i=1}^K \frac{i}{K} \left(\log X_{N-i+1,N} - \log X_{N-i,N}\right)$$
$$= \frac{1}{K} \sum_{i=1}^K \log\left(\frac{X_{N-i+1,N}}{X_{N-K,N}}\right)$$
$$= H_{K,N}.$$

Consistency of this estimator is classically guaranteed as soon as  $K \to +\infty$  when  $N \to +\infty$  so that K = o(N), see [28]. In this case, the empirical threshold  $X_{N-K,N}$  (equivalent in probability to U(N/K)) goes to infinity (again in probability) as  $N \to +\infty$ .

Going back to the survey data situation, one may naturally replace U(N/K) by  $U_N^{\pi}(N/K)$  and build a plug-in estimate of the EVI  $\gamma$  based on the Horvitz–Thompson estimator given in equation (2.1) of the tail probability  $\overline{F}(x)$ . Observe that by definition  $U_N^{\pi}(N/K)$  corresponds to one of the observations in the sample, say  $X_{i^*}$  with rank  $\ell \in \{1, \ldots, n\}$ . To this  $\ell$  obviously corresponds an index  $k \in \{0, \ldots, n-1\}$  such that  $\ell = n - k$ , implying  $X_{n-k,n} = U_N^{\pi}(N/K)$ , the Horvitz–Thompson estimator of the quantile of order 1 - K/N. We denote by  $\kappa_N^{\pi}$ the map linking k to K in  $\mathcal{U}_N$  under the sampling scheme  $R_N$ :

$$\kappa_N^{\boldsymbol{\pi}} : \begin{pmatrix} \{1, \dots, N-1\} \longrightarrow \{1, \dots, n-1\} \\ K \longmapsto k := \kappa_N^{\boldsymbol{\pi}}(K) \end{pmatrix},$$

where

$$\kappa_N^{\boldsymbol{\pi}}(K) := n - \inf\left\{ i \in \{1, \dots, n-1\} : \sum_{j=1}^i \frac{1}{\pi_{j,n}} \ge N - K \right\}.$$
(3.2)

This leads to the quantity:

$$\begin{split} \widehat{\gamma} &= \int_{X_{n-k,n}}^{+\infty} \frac{\overline{F}_{R_N}^{\pi(R_N)}(u)}{\overline{F}_{R_N}^{\pi(R_N)}(X_{n-k,n})} \frac{\mathrm{d}u}{u} = \sum_{i=1}^k \int_{X_{n-i,n}}^{X_{n-i+1,n}} \frac{\overline{F}_{R_N}^{\pi(R_N)}(u)}{\overline{F}_{R_N}^{\pi(R_N)}(X_{n-k,n})} \frac{\mathrm{d}u}{u} \\ &= \sum_{i=1}^k \left( \frac{\sum_{j=1}^i \frac{1}{\pi_{n-j+1,n}}}{\sum_{j=1}^k \frac{1}{\pi_{n-j+1,n}}} \right) \times \left( \log(X_{n-i+1,n}) - \log(X_{n-i,n}) \right) \\ &= \left( \sum_{j=1}^k \frac{1}{\pi_{n-j+1,n}} \right)^{-1} \sum_{i=1}^k \frac{1}{\pi_{n-i+1,n}} \log\left( \frac{X_{n-i+1,n}}{X_{n-k,n}} \right). \\ &=: H_{k,n}^{\pi}. \end{split}$$

Hence, k is to the sample what K is to the population: the number of upper values on which the estimation should rely. Notice that we may also write

$$H_{k,n}^{\pi} = \left(\sum_{j=1}^{K} \frac{\epsilon_{N-j+1,N}}{\pi_{N-j+1,N}}\right)^{-1} \sum_{i=1}^{K} \frac{\epsilon_{N-i+1,N}}{\pi_{N-i+1,N}} \log\left(\frac{X_{N-i+1,N}}{X_{N-K,N}}\right) =: H_{K,N}^{\pi},$$
(3.3)

where K is the chosen number of largest observations in the population from which k was constructed. Observe that  $\kappa_N^{\pi}$  in equation (3.2) is a surjective, non-injective random map, which suggests that the subsequent asymptotic analysis better rely on some appropriately chosen K and its random image  $k := \kappa_N^{\pi}(K)$  rather than the contrary. Since in practice only k can be computed from the  $X_i$ 's and  $\pi_i$ 's,  $i \in S$ , the total population being partly unobserved, considerations about the choice of an appropriate k are discussed in detail in Subsection 5.1. From this point forward, the Horvitz–Thompson Hill estimator shall be written  $H_{K,N}^{\pi}$  with  $K \in \{1, \ldots, N-1\}$ held fixed.

## 3.2. Consistency result

Here we investigate the limit properties of the estimator  $H_{K,N}^{\pi}$  as N and n simultaneously go to infinity, with  $n \leq N$ . The following assumptions, related to the sample design, shall be involved in the asymptotic analysis.

Assumption 1. There exist  $\pi_* > 0$  and  $N_0 \in \mathbb{N} \setminus \{0\}$  such that for all  $N \ge N_0$  and  $i \in \mathcal{U}_N$ ,

$$\pi_i > \pi_\star$$

Assumption 2. There exists  $\ell < +\infty$  such that we have  $\forall N \ge 1$ ,

$$\max_{1 \le i, \ j \le N} |\pi_{i,j} - \pi_i \pi_j| \le \frac{\ell}{n}$$

Hypothesis 1 guarantees that first order inclusion probabilities do not vanish asymptotically, while hypotesis 2 corresponds to the situation where the second order inclusion probabilities are not too different from those in the case of independent sampling (it is thus fulfilled by the Poisson design, see Sect. 4.1).

**Remark 3.1** (On Assumptions 1 and 2). The assumptions introduced herein-before are rather mild and are fulfilled in a wide variety of situations. Indeed, Assumption 2 is standard in asymptotic analysis of sampling techniques, see [24,25] for instance. As for Assumption 1, recall that  $\pi_i = \pi(\mathbf{W}_i)$  with  $\mathbf{W}$  an auxiliary variable and  $\pi(.)$  a link function. Then, it is fulfilled as soon as  $\pi(.)$  is continuous and the support of  $\mathbb{P}_{\mathbf{W}}$  is a compact subset of  $(\mathbb{R}^*_+)^d$ , where d denotes the dimension of the random vector  $\mathbf{W}$  and  $\mathbb{R}^*_+$  the set of positive real numbers.

Given this framework, following in the footsteps of Section 4.4.1 from [30], the consistency of  $H_{K,N}^{\pi}$  can be handled by exploiting the properties of regularly varying distributions. Indeed, under the heavy-tail assumption in equation (2.2), provided that  $K \to +\infty$  and  $K/N \to 0$  as  $N \to +\infty$ , we have

$$\frac{N}{K} \mathbb{P}\left(\frac{X}{U(N/K)} \in .\right) \xrightarrow[N \to \infty]{v} \nu_{-1/\gamma}(.)$$

in the space of Radon measures on  $(0, +\infty]$ . There, " $\stackrel{"v}{\rightarrow}$ " stands for the vague convergence of measures<sup>5</sup> and  $\nu_{-1/\gamma}(.)$  is such that  $\nu_{-1/\gamma}(x, \infty] = x^{-1/\gamma}$  for all x > 0 (see [30], Thm. 3.6 for instance). Its empirical counterpart in the population, usually called the *tail empirical measure*, is defined as follows:

$$\nu_N := \frac{1}{K} \sum_{i=1}^N \delta_{X_i/U(N/K)}.$$

When replacing U(N/K) by its estimate  $U_N(N/K)$  in the expression above and assuming that  $K = K(N) \rightarrow +\infty$  where  $K/N \rightarrow 0$  as  $N \rightarrow +\infty$ , it can be shown that it converges to  $\nu_{-1/\gamma}$  in probability ([30], Eq. (4.21)). Since we have

$$H_{K,N} = \int_{1}^{\infty} \nu_N \left( \frac{X_{N-K,N}}{U(N/K)} \left( x, +\infty \right] \right) \frac{\mathrm{d}x}{x}$$

<sup>&</sup>lt;sup>5</sup>Recall that, in the space of non-negative Radon measures on  $(0, +\infty]$ , a sequence  $(\mu_m)_{m\geq 1}$  is said to converge vaguely to  $\mu$  iff for any compactly supported continuous function  $h: (0, +\infty] \to \mathbb{R}$ , we have:  $\int_0^{+\infty} h(x)\mu_m(\mathrm{d}x) \to \int_0^{+\infty} h(x)\mu(\mathrm{d}x)$  as  $m \to +\infty$ .

and

$$\gamma = \int_1^\infty \nu_{-1/\gamma}(x,+\infty] \, \frac{\mathrm{d}x}{x},$$

the asymptotic properties of  $\nu_N$  naturally convey the consistency of the Hill estimator. Generalizing this result to the *Horvitz–Thompson tail empirical measure* 

$$\nu_N^{\pi} = \frac{1}{K} \sum_{i=1}^N \frac{\epsilon_i}{\pi_i} \,\delta_{X_i/U_N^{\pi}(N/K)} = \frac{1}{K} \sum_{i \in S} \frac{1}{\pi_i} \,\delta_{X_i/U_N^{\pi}(N/K)} \tag{3.4}$$

would then yield the theorem below (see the proof in the Appendix section). It reveals that, in regard to the asymptotic statistical estimation of the EVI  $\gamma$ , the Horvitz–Thompson variant of the Hill estimator  $H_{K,N}^{\pi}$  is consistent.

**Theorem 3.2** (Consistency). Let K = K(N) be a sequence of integers such that  $K \to +\infty$  and  $K/N \to 0$  as  $N, n \to +\infty$ . Provided that Assumptions 1 and 2 are fulfilled, we then have, as N and n tend to  $+\infty$ :

$$H_{K,N}^{\pi} \xrightarrow{\mathbb{P}} \gamma. \tag{3.5}$$

# 4. Asymptotic normality

Whereas the consistency of the standard Hill estimator in equation (2.3) can be proved for any sequence K going to infinity at a reasonable rate, asymptotic normality cannot be guaranteed at such a level of generality. Higher-order regular variation properties of the heavy-tail model in equation (2.2) are required [12, 14]. More specifically, consider the hypothesis below, referred to as the *Von Mises condition* [20].

Assumption 3. The regularly varying tail quantile function  $U \in \mathcal{R}_{\gamma}$  with  $\gamma > 0$  is such that there is a real parameter  $\rho < 0$ , referred to as the *second order parameter*, and a positive or negative function A with  $\lim_{x \to +\infty} A(x) = 0$  such that for any t > 0,

$$\frac{1}{A(x)} \left( \frac{U(tx)}{U(x)} - t^{\gamma} \right) \underset{x \to \infty}{\longrightarrow} t^{\gamma} \frac{t^{\rho} - 1}{\rho},$$

or equivalently

$$\frac{1}{A\left(\frac{1}{\overline{F}(x)}\right)} \left(\frac{\overline{F}(t\,x)}{\overline{F}(x)} - t^{-1/\gamma}\right) \xrightarrow[x \to \infty]{} t^{-1/\gamma} \frac{t^{\rho/\gamma} - 1}{\gamma \,\rho} \cdot$$

This condition simply establishes some constraints about the slowly varying function L(.) in equation (2.2) to ensure its influence vanishes quickly enough not to interfere with the Pareto form  $x^{-1/\gamma}$  of  $\overline{F}$ .

The limit distribution of the standard Hill estimator in equation (2.3) has been investigated by means of Rényi's exponential representation of log-spacings under Assumption 3. Of course, this condition can hardly be checked in practice and the choice of the number of extremal observations is generally selected so as to minimize an estimate of the asymptotic mean squared error (MSE), see Section 5.1 and the references therein.

**Remark 4.1** (On the Hill estimator and the tail empirical process). Other approaches than the Rényi decomposition in log-spacings have been developed to prove the asymptotic normality of the Hill estimator  $H_{K,N}$ (see [30], Chaps. 4 and 9 and the references therein). The version presented at length in [30] involves the preliminary study of the *tail empirical process* 

$$\mathbb{T}_N := \sqrt{K} \left( \nu_N - \nu_{-1/\gamma} \right) (x^{-1/\gamma}, \infty], \quad x \ge 0,$$

34

from which the asymptotic properties of the Hill estimator are later deduced. We refer to Theorem 9.1 and Section 9.1.2 in [30] for more details on this seemingly simple but actually quite intricate procedure.

Unfortunately, contrary to the classical empirical process, the asymptotic properties of  $\mathbb{T}_N$  cannot be extended to its Horvitz-Thompson equivalent. This is essentially due to the fact that the population  $\mathcal{U}_N$  contains a *finite* number of observations. Hence, there is always a maximum  $X_{N,N} < \infty$  bounding  $\mathbb{P}_N$  and sampling fails to distinguish between a distribution with finite support such as those in the Weibull domain of attraction and a heavy-tailed distribution with no endpoint. Actually, these arguments are exactly the same as those introduced when discussing the applicability of bootstrap in extreme value analysis. Indeed, survey sampling under a superpopulation model can be viewed as a generalization of weighted bootstrap. The interested reader may refer to Section 6.4 in [30], for a brief introduction to the difficulty of bootstrapping heavy-tailed phenomena.

In this section, we shall aim at proving first that under Poisson survey schemes, the Horvitz–Thompson version of the Hill estimator computed on the sample is asymptotically close to its standard version calculated over the entire population. This result is next extended to rejective sampling plans, which can be viewed as Poisson plans conditioned upon a fixed sample size.

### 4.1. The case of the Poisson survey scheme

In this subsection we assume that the vector  $\boldsymbol{\epsilon}$  corresponds to that of a Poisson survey scheme, such as depicted in Example 1. The distribution of this design is denoted by  $T_N$  and the first order inclusion probabilities by  $p_1, \ldots, p_N$ . Under this setting, the Horvitz–Thompson variant of the Hill estimator is naturally denoted by  $H_{K,N}^{\mathbf{p}}$  and the  $\pi_i$ 's in equation (3.3) are to be replaced by the corresponding  $p_i$ 's. In addition, just like we previously set  $\pi_i = \pi(\mathbf{W}_i)$ , we write  $p_i = p(\mathbf{W}_i)$  for all  $i \in \mathcal{U}_N$  and p the Poisson link function. In order to prove the asymptotic normality of  $H_{K,N}^{\mathbf{p}}$ , we shall require that the  $p_i$ 's fulfill Assumption 1 with lower bound  $p_{\star}$  and the auxiliary variable from which they are built to satisfy the condition below.

Assumption 4. The random vectors  $\mathbf{W}_1, \ldots, \mathbf{W}_N$  are iid with continuous distribution  $\mathbb{P}_{\mathbf{W}}$  on  $\mathcal{W} \subset \mathbb{R}^d$ , *d*-variate cdf  $F_{\mathbf{W}}$  and marginals  $F_{W_1}, \ldots, F_{W_d}$ . The joint distribution of the entailed iid sequence  $\{(X_i, \mathbf{W}_i), 1 \leq i \leq N\}$  is denoted by  $\mathbb{P}_{X,\mathbf{W}}$  with corresponding cdf  $F_{X,\mathbf{W}}$ .

The following result reveals that under the Poisson survey scheme, when based on the K largest values among the whole population  $X_1, \ldots, X_N$ ,  $H_{K,N}^{\mathbf{p}}$  converges at the same rate  $(1/\sqrt{K} \text{ namely})$  to the same limit distribution as  $H_{K,N}$ , up to a multiplicative term in the asymptotic variance induced by the sampling scheme. Further details about the convergence of the classical Hill estimator can be found *e.g.* in ([12], Thm. 1) and ([30], Sect. 9).

**Theorem 4.2** (Limit distribution in the Poisson survey case). Suppose that Assumption 3 is fulfilled by the underlying heavy-tailed model and that Assumption 1 is satisfied by the considered sequence of Poisson inclusion probabilities  $p_1, \ldots, p_N, N \ge 1$ , constructed from some set of auxiliary variables as in Assumption 4. In addition, assume that as  $N, K \to \infty, K/N \to 0$ ,

$$\mathbb{E}\left(\frac{1}{p(\mathbf{W})} \mid X > U(N/K)\right) \underset{N \to \infty}{\longrightarrow} \sigma_p^2 < \infty.$$

Then, provided that  $K \to +\infty$  as  $N \to +\infty$  so that  $\sqrt{K}A(N/K) \to \lambda$  for some constant  $\lambda \in \mathbb{R}$ , we have the convergence in distribution as  $N \to +\infty$ :

$$\sqrt{K} \left( H_{K,N}^{\mathbf{p}} - \gamma \right) \Rightarrow \mathcal{N} \left( \frac{\lambda}{1 - \rho}, \gamma^2 \sigma_p^2 \right).$$
(4.1)

#### P. BERTAIL ET AL.

As can be seen by examining the proof of this theorem in the Appendix section, the limit result in equation (4.1) can be obtained using the following decomposition:

$$\sqrt{K} \left( H_{K,N}^{\mathbf{p}} - \gamma \right) = \underbrace{\frac{\sqrt{K}}{r_{K,N}} \left( r_{K,N} H_{K,N}^{\mathbf{p}} - H_{K,N} \right)}_{Q_N^{(1)}} + \underbrace{\frac{\sqrt{K}}{r_{K,N}} \left( H_{K,N} - \gamma \right)}_{Q_N^{(2)}} + \underbrace{\gamma \sqrt{K} \left( \frac{1}{r_{K,N}} - 1 \right)}_{Q_N^{(3)}},$$
(4.2)

where

$$r_{K,N} := r_{K,N}(T_N, \mathbf{p}) := \frac{1}{K} \sum_{i=1}^{K} \frac{\epsilon_{N-i+1,N}}{p_{N-i+1,N}}.$$
(4.3)

These three quantities are studied independently under the hypotheses required in Theorem 4.2. First, we show that  $r_{K,N}$  converges to 1 in probability as  $N \to +\infty$ . Combined with Rényi's decomposition in log-spacings of the Hill estimator (refer for instance to [1], Sect. 4.4), this establishes the asymptotic convergence, in probability, of  $Q_N^{(1)}$  to 0. It also implies that  $Q_N^{(2)}$  is equivalent to  $\sqrt{K}$   $(H_{K,N} - \gamma)$ , a well-known quantity which tends to a Gaussian distribution with expectation  $\lambda/(1 - \rho)$  and variance  $\gamma^2$  under the second order condition stipulated in Assumption 3 ([11], Thm. 3.2.5). As for  $Q_N^{(3)}$ , we calculate its expectation and variance conditionally on the full vector of observations  $\{(X_i, \mathbf{W}_i), 1 \leq i \leq N\}$ , yielding expressions where the randomness induced by the survey scheme has been controlled. Then, following in the lines of a Lindeberg–Feller theorem for independent and non-identically distributed variables ([17], Thm. 3, p. 262), we analyze the conditions under which the conditional variance has a finite limit in probability relative to  $\{(X_i, \mathbf{W}_i), 1 \leq i \leq N\}$ . Provided that they are fulfilled,  $Q_N^{(3)}$  converges weakly to a centered Gaussian distribution with variance  $\gamma^2 (\sigma_p^2 - 1)$ . Thus proceeding enables to consider  $Q_N^{(2)}$  and  $Q_N^{(3)}$  as independent random variables (one depends on the data and the other on the survey scheme). Thus, the limit distribution of their sum is simply the sum of their limit distributions, thereby yielding equation (4.1).

In the remark below, we exhibit some sufficient conditions that guarantee the existence of  $\sigma_p^2$ , the limit of the conditional expectation appearing in Theorem 4.2.

**Remark 4.3** (Sufficient conditions for Theorem 4.2). It is possible to give a more explicit expression of  $\sigma_p^2$  under some additional conditions. Let  $C_{X,\mathbf{W}} : [0,1]^{d+1} \to [0,1]$  be the copula of cdf  $F_{X,\mathbf{W}}$  such that for all  $x \in \mathbb{R}^*_+$ ,  $\mathbf{w} = (w_1, \ldots, w_d) \in \mathcal{W}$ ,

$$F_{X,\mathbf{W}}(x,\mathbf{w}) = C_{X,\mathbf{W}}(F(x), F_{W_1}(w_1), \dots, F_{W_d}(w_d))$$

In addition to the set of assumptions required in Theorem 4.2, assume that the conditions below hold.

- $(\mathcal{A}_1)$  The cdf F of X is absolutely continuous with respect to the Lebesgue measure with density f.
- $(\mathcal{A}_2)$  The one-dimensional marginal cdfs  $F_{W_1}, \ldots, F_{W_d}$  of **W** are absolutely continuous with respect to the Lebesgue measure with respective densities  $f_{W_1}, \ldots, f_{W_d}$ .
- $(\mathcal{A}_3)$  The copula  $C_{X,\mathbf{W}}$  is absolutely continuous with Lebesgue-integrable density  $c_{X,\mathbf{W}}$  bounded in a neighborhood of  $\{1\} \times [0,1]^d$ .

Then,

$$\sigma_p^2 = \int_{\mathcal{W}} \frac{1}{p(\mathbf{w})} c_{X,\mathbf{W}}(1, F_{W_1}(w_1), \dots, F_{W_d}(w_d)) \prod_{j=1}^d f_{W_j}(w_j) \, \mathrm{d}w_1 \dots \mathrm{d}w_d$$

36

In order to understand the behavior of  $\sigma_p^2$  as a function of the copula density and of the choice of the inclusion probabilities, we illustrate our result by means of a simple example.

**Example 2** (Clayton copula). Consider the simple situation where W is a one-dimensional random variable with cdf  $F_W$  and density  $f_W$ , and where the copula  $C_{X,W}$  belongs to the Clayton class of copulas, *i.e.* there exists some  $\theta > 0$  such that for all  $(u, v) \in [0, 1]^2 \setminus \{(0, 0)\}$ ,

$$C_{X,W}(u,v) = C_{\theta}(u,v) := (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$$

Written this way, the function  $C_{X,W}$  models the positive dependence of rare events. It is known that the related tail dependence coefficient is given by  $2^{-1/\theta}$  ([29], Example 5.22 p. 215, Family (4.2.1)) and the Kendall's tau by  $\theta/(\theta+2)$  ([29], Example 5.4 p. 163), yielding a wide range of possible correlations. Straightforward computations permit to show that the density is given by: for all  $(u, v) \in [0, 1]^2 \setminus \{(0, 0)\}$ ,

$$c_{\theta}(u,v) = (\theta+1) u^{-1-\theta} v^{-1-\theta} (u^{-\theta} + v^{-\theta} - 1)^{-2-1/\theta}$$

Notice that the density function  $c_{\theta}$  is not bounded near (0,0). Nevertheless it has no singularity near (1,1) and is integrable at (1,0). Hence, conditions  $\mathcal{A}_1 - \mathcal{A}_3$  in Remark 4.3 hold true. Thus, for all  $v \in [0,1]$ , we have the simpler expression

$$c_{\theta}(1,v) = (\theta+1) v^{\theta}$$

Now take

$$p(w) = \begin{cases} F_W(w) & \text{if } F_W(w) \in (p_\star, 1], \\ p_\star & \text{if } F_W(w) \in [0, p_\star]. \end{cases}$$

When i is not too small, these inclusion probabilities are simply proportional to the rank of  $W_i$  divided by N. Then we have

$$\sigma_p^2 = \frac{1}{p_\star} \int_0^{p_\star} (\theta + 1) v^\theta \, \mathrm{d}v + \int_{p_\star}^1 (\theta + 1) v^{\theta - 1} \, \mathrm{d}v = 1 + \frac{1}{\theta} (1 - p_\star^\theta).$$

It follows that, as the parameter  $\theta$  grows to  $\infty$  (*i.e.* the greater the correlation in the extremes),  $\sigma_p^2$  tends to 1. In other words, sampling in this case does not degrade at all the asymptotic variance of the Hill estimator.

Before turning to the extension of the results above to more general sampling schemes, a few remarks are in order.

**Remark 4.4** (A singular situation). A situation of particular interest is when W is one-dimensional and there exists a continuous increasing function  $g: \mathcal{W} \to \mathbb{R}^*_+$  such that W = g(X). In that case, the copula  $C_{X,W}$  linking both cdfs F of X and  $F_W$  of W is such that for all  $(u, v) \in [0, 1]^2$ ,  $C_{X,W}(u, v) = u \wedge v$ , which is not differentiable. Nevertheless, assuming that F is absolutely continuous with respect to the Lebesgue measure with density f and that p is increasing, the result still holds with  $\sigma_p^2 = 1/p(w^*)$ , where  $w^* := \inf\{w: F_W(w) = 1\}$  is the right endpoint of the distribution of W (potentially  $+\infty$ ). Indeed, we have

$$\mathbb{E}\left(\frac{1}{p \circ g(X)} \mid X > U(N/K)\right) = \frac{N}{K} \int_{U(N/K)}^{\infty} \frac{1}{p \circ g(x)} f(x) \, \mathrm{d}x.$$

Setting  $u = \frac{N}{K} (1 - F(x))$  entails

$$\mathbb{E}\left(\frac{1}{p \circ g(X)} \mid X > U(N/K)\right) = \int_0^1 \frac{1}{p \circ g(U(N/(Ku)))} \,\mathrm{d}u,$$

and since by virtue of Assumption 1, the Poisson link function p is bounded, as  $N \to \infty$  this last integral tends to  $1/(p \circ g(\infty)) = 1/p(w^*)$  for g increasing. For such a  $\sigma_p^2$  to be close to 1, it suffices to give a probability of inclusion of 1 to the most extreme observations of W (and a fortiori of X).

#### P. BERTAIL ET AL.

**Remark 4.5** (On the asymptotic variance). Looking at the variance term in equation (4.1) of Theorem 4.2, we see that the influence of the survey scheme is encapsulated by the multiplicative term  $\sigma_p^2$ , which is by definition greater than or equal to 1. Ideally, we would like to have at our disposal inclusion probabilities for which  $\sigma_p^2$  is as close to 1 as possible. This aspect is discussed in Section 5.2. In any case, it is possible to use the Horvitz–Thompson empirical version of  $\sigma_p^2$  given by

$$(\tilde{S}_N^{\mathbf{p}})^2 := \frac{1}{K} \sum_{i=1}^N \frac{\epsilon_i}{p(W_i)} \frac{1}{p(W_i)} \mathbb{I}\{X_i > U(N/K)\}$$

to estimate this variance term. Indeed, conditionally on the full vector of observations  $(\mathbf{W}_i, X_i)_{1 \le i \le N}$ , we have that

$$\mathbb{E}\left((\widetilde{S}_N^{\mathbf{p}})^2\right) = \frac{1}{K} \sum_{i=1}^N \frac{1}{p(W_i)} \mathbb{I}\left\{X_i > U(N/K)\right\} \underset{N \to \infty}{\longrightarrow} \sigma_p^2 \text{ a.s.}$$

(see Lem. 19). It follows that  $(\widetilde{S}_N^{\mathbf{p}})^2$  converges in  $L_1$  and thus in probability to  $\sigma_p^2$ .

## 4.2. Extension – Rejective sampling schemes

We now show how the result stated in Theorem 4.2 can be extended to an important class of survey plans, namely *rejective sampling schemes*. For the sake of clarity, we first provide a brief description of the latter, refer to [2, 24] for further details.

Fix  $n \leq N$  and consider a vector  $(\pi_1, \ldots, \pi_N)$  of first order inclusion probability. The rejective sampling, sometimes referred to as *conditional Poisson sampling* (CPS in short), exponential design without replacement or maximum entropy design [34], is the sampling plan  $R_N$  which picks samples of fixed size n(S) := n in order to maximize the entropy measure

$$H(R_N) = -\sum_{\{s \in \mathcal{P}(\mathcal{U}_N): \#s=n\}} R_N(s) \log R_N(s)$$

subject to the constraint stipulating that its vector of first order inclusion probabilities coincides with  $(\pi_1, \ldots, \pi_N)$ . It can be implemented in two steps, as follows.

- 1. Draw a sample S with a Poisson sampling plan (without replacement), with properly chosen first order inclusion probabilities  $(p_1, \ldots, p_N)$ . The representation is called canonical if  $\sum_{i=1}^{N} p_i = n$ . In that case the relationships between  $p_i$  and  $\pi_i$ ,  $1 \le i \le N$ , are established in [24].
- 2. If  $n(\mathcal{S}) \neq n$ , then reject it and go back to step one, otherwise stop.

The vector  $(p_1, \ldots, p_N)$  must be chosen in a way that the resulting first order inclusion probabilities coincide with  $\pi_1, \ldots, \pi_N$ , by means of a dedicated optimization algorithm, see [34]. The corresponding probability distribution is given by:  $\forall s \in \mathcal{P}(\mathcal{U}_N)$ ,

$$R_N(s) = \frac{T_N^{\mathbf{p}}(s)\mathbb{I}\{\#s=n\}}{\sum_{\{s'\in\mathcal{P}(\mathcal{U}_N): \ \#s'=n\}} T_N^{\mathbf{p}}(s')} \propto \prod_{i\in s} p_i \prod_{i\notin s} (1-p_i) \times \mathbb{I}\{\#s=n\}.$$

Refer to ([24], p. 1496), for more details on the  $p_i$ 's.

Turning now to the extension of the result stated in Theorem 4.2 for the Poisson survey scheme to the case of rejective sampling, we proceed in two steps. For clarity, we shall write  $H_{K,N}^{\pi}(R_N)$  when the Horvitz–Thompson version of the Hill estimator involves the inclusion variables  $\epsilon_1, \ldots, \epsilon_N$  drawn under the sampling plan  $R_N$  and the probabilities of inclusion  $\pi_1, \ldots, \pi_N$ . First, we show that the asymptotic distribution of  $\sqrt{K}(H_{K,N}^{\mathbf{p}}(R_N) - \gamma)$  is the same as that of  $\sqrt{K}(H_{K,N}^{\mathbf{p}}(T_N) - \gamma)$ , where  $R_N$  (resp.  $T_N$ ) denotes a rejective (resp. Poisson) sampling plan

with inclusion probabilities  $\pi_1, \ldots, \pi_N$  (resp.  $p_1, \ldots, p_N$ ). Then, we control the difference between  $H_{K,N}^{\mathbf{p}}(R_N)$ and  $H_{K,N}^{\boldsymbol{\pi}}(R_N)$ . For this, we introduce the following quantities:  $\forall K \leq N$ ,

$$D_{K,N}(R_N, T_N) := r_{K,N}(R_N, \pi) H_{K,N}^{\pi}(R_N) - r_{K,N}(R_N, \mathbf{p}) H_{K,N}^{\mathbf{p}}(R_N)$$

and

$$r_{K,N}(R_N,T_N) := r_{K,N}(R_N,\boldsymbol{\pi}) - r_{K,N}(R_N,\mathbf{p}).$$

The ensuing approach follows in the footsteps of [24] and relies more specifically on the results displayed in Theorem 5.1 (p. 508). Let us start by defining the quantities

$$d_N = \sum_{i=1}^N p_i(1-p_i)$$
 and  $\bar{p}_N = \frac{1}{d_N} \sum_{i=1}^N p_i^2(1-p_i).$ 

We assume that both  $R_N$  and  $T_N$  fulfill Assumption 1 for minoring constants  $\pi_{\star}$  and  $p_{\star}$  respectively and that the Poisson inclusion probabilities further satisfy the following condition.

#### Assumption 5.

$$\limsup_{N \to +\infty} \frac{1}{N} \sum_{i=1}^{N} p_i(T_N) < 1.$$

Notice that, in this situation,  $d_N = o(1/K)$  and  $\bar{p}_N$  is bounded. In addition, as shown in [24] (see p. 1510 therein), the decomposition below holds for all  $i \in \{1, \ldots, N\}$ :

$$p_i - \pi_i = \left(\frac{\bar{p}_N - p_i}{d_N} + o(1/d_N)\right) p_i(1 - \pi_i).$$

This roughly means that the inclusion probabilities of the rejective sampling scheme are very close to those of the underlying Poisson design from which it was built. So close in fact that  $D_{K,N}(R_N, T_N)$  and  $r_{K,N}(R_N, T_N)$ asymptotically vanish. Therefore, as revealed by the following result, Theorem 4.2 also holds when the sample is constructed with a rejective plan.

**Theorem 4.6** (Limit distribution in the rejective survey case). Suppose that all the conditions required in Theorem 4.2 hold together with Assumption 5. Then, for  $\sigma_p^2$  as in Theorem 4.2 and provided that  $K \to +\infty$  as  $N \to +\infty$  so that  $\sqrt{K}A(N/K) \to \lambda$  for some constant  $\lambda \in \mathbb{R}$ , we have the convergence in distribution as  $N \to +\infty$ :

$$\sqrt{K} \left( H_{K,N}^{\pi} - \gamma \right) \Rightarrow \mathcal{N} \left( \frac{\lambda}{1 - \rho} \,, \, \gamma^2 \, \sigma_p^2 \right). \tag{4.4}$$

The proof of this theorem is available in the Appendix section. Notice that the limit variance does not depend on the inclusion probabilities  $\pi_1, \ldots, \pi_N$ , but on those of the underlying Poisson design,  $p_1, \ldots, p_N$  namely. The asymptotic properties of the rejective sampling scheme are intricately linked to the Poisson plan which it is associated with.

# 5. PRACTICAL ISSUES AND ILLUSTRATIVE EXPERIMENTS

### 5.1. On the choice of an optimal k

All results presented in the previous section depend on some appropriately number K of largest observations in the population  $X_1, \ldots, X_N$ . Unfortunately, the estimated tail quantile  $X_{N-K,N}$  from which  $H_{K,N}^{\pi}$  is computed may not be included in the sample. Hence, we need to choose a number k of largest values in the sample to which we may associate some K that respects the necessary conditions for consistency and asymptotic normality to

39

#### P. BERTAIL ET AL.

hold  $(K = K(N) \to +\infty \text{ and } \sqrt{K}A(N/K) \to \lambda < \infty \text{ as } N \to +\infty)$ . Recall that we defined  $\kappa_N^{\pi}$  in equation (3.2), a non-injective random map that assigns an index k in the sample to any index K in the population so that  $X_{n-k,n} = U_N^{\pi}(N/K)$ . Setting

$$\widehat{K}(k) := (\kappa_N^{\boldsymbol{\pi}})^{\leftarrow}(k) := \left\lceil N - \sum_{i=1}^{n-k} \frac{1}{\pi_{i,n}} \right\rceil,$$

where  $\lceil .\rceil$  is the ceiling function, it is straightforward to show that the limit results stated in the sections above remains true for  $\sqrt{\hat{K}(k)} \left(H_{k,n}^{\pi} - \gamma\right)$  as  $N, k \to +\infty$  so that  $\sqrt{\hat{K}(k)} A(N/\hat{K}(k)) \to \lambda$  almost-surely. This result can be naturally used to ground the construction of asymptotic Gaussian confidence intervals. Indeed, by virtue of Slutsky's Lemma combined with Theorem 3.2, the quantity  $\sqrt{\hat{K}(k)} \left(H_{k,n}^{\pi} - \gamma\right) / H_{k,n}^{\pi}$  is then asymptotically pivotal, distributed as a standard Gaussian random variable as  $N, k \to +\infty$ .

In practice, choosing an optimal threshold  $X_{N-K,N}$  is already complicated in the i.i.d. case. Many techniques have been proposed in the literature, often based on the minimization of the MSE (see [10, 19, 21] and the references therein). Since they involve in general the estimation of the second order parameter  $\rho$ , which goes beyond the scope of our analysis, we leave such considerations for future research. In the meantime, we propose to simply rely on heuristics such as the stability of the Horvitz–Thompson version of the Hill estimator around the appropriate k.

## 5.2. On the choice of the sampling weights

A question of interest is the design of a survey sampling scheme yielding a weighted Hill estimator of minimum asymptotic variance. In order to give some hints of the corresponding plan, assume for simplicity that W is a onedimensional random variable with distribution  $F_W$  and density  $f_W$  from which we observe N i.i.d. realizations, denoted by  $W_{i,N}$  once ordered. Also consider the conditions of Theorem 4.2 to hold together with those of Remark 4.3. Then, minimizing the variance of the Horvitz-Thompson variant of the Hill estimator boils down to minimizing

$$\sigma_p^2 := \int_{\mathcal{W}} \frac{1}{p(w)} c_{X,W}(1, F_W(w)) f_W(w) \,\mathrm{d}w$$

subject to the constraints

$$\begin{cases} \int_{\mathcal{W}} p(w) f_W(w) \, \mathrm{d}w =: \eta = \lim_{N \to \infty} \frac{\mathbb{E}(n)}{N} < 1, \\ \forall w \in \mathcal{W}, \ p_\star \le p(w) \le 1. \end{cases}$$

This variational problem may be translated in terms of the simpler convex optimisation problem (for chosing the weights) given by

$$\min \frac{1}{N} \sum_{i=1}^{N} \frac{c_{X,W}\left(1, \frac{i}{N}\right)}{p\left(W_{i,N}\right)} \text{ subject to } \begin{cases} \sum_{i=1}^{N} p\left(W_{i,N}\right) = \mathbb{E}\left(n\right), \\ p_{\star} < p\left(W_{i,N}\right) < 1. \end{cases}$$
(5.1)

If for all  $i \in \{1, ..., N\}$  we set  $p_i := p(W_{i,N})$  and denote by  $\lambda_0, ..., \lambda_{2N}$  the KKT multipliers, then the Lagrangian of this problem is simply given by

$$\mathcal{L}(\mathbf{p}) := \frac{1}{N} \sum_{i=1}^{N} \frac{1}{p_i} c_{X,W} \left( 1, \frac{i}{N} \right)$$
$$+ \lambda_0 \left( \sum_{i=1}^{N} p_i - \mathbb{E}\left( n \right) \right) + \sum_{i=1}^{N} \lambda_i (p_i - 1) + \sum_{i=1}^{N} \lambda_{N+i} (p_\star - p_i).$$

Provided that the set of linear constraint is non-empty (it may be infeasible for instance if we chose  $p_{\star}$  bigger than  $\eta$ ), the Karush–Kuhn–Tucker theorem yields that the solution is given by the conditions

$$(\mathcal{C}_1) \quad \widehat{p}_i = \frac{\sqrt{c_{X,W}\left(1, \frac{i}{N}\right)}}{\sqrt{N}\sqrt{\lambda_0 + \lambda_i - \lambda_{N+i}}},$$
  

$$(\mathcal{C}_2) \quad \lambda_i(\widehat{p}_i - 1) = 0, \ \lambda_i \ge 0,$$
  

$$(\mathcal{C}_3) \quad \lambda_{N+i}(p_\star - \widehat{p}_i) = 0, \ \lambda_{N+i} \ge 0,$$

for all  $i \in \{1, \ldots, N\}$ , plus the initial conditions in equation (5.1). This means that the weights are either proportional to  $\sqrt{c_{X,W}(1,i/n)}$  or saturated respectively to 1 for too large values of  $c_{X,W}(1,.)$  and to  $p_{\star}$  for too small ones. When  $c_{X,W}(1,v)$  is an increasing function of  $v \in [0,1]$ , it is possible to give a more explicit solution. Denote by  $n_0$  (resp.  $n_1$ ) the resulting number of inclusion probabilities equal to  $p_{\star}$  (resp. 1). Then it is easy to see that the equality

$$n_0 p_{\star} + \frac{1}{\sqrt{N \lambda_0}} \sum_{i=n_0+1}^{N-n_1} \sqrt{c_{X,W}\left(1, \frac{i}{N}\right)} + n_1 = n,$$

yields for  $p_{\star} < \hat{p}_i < 1$ :

$$\widehat{p}_{i} = \frac{\left(n - n_{1} - n_{0} p_{\star}\right) \sqrt{c_{X,W}\left(1, \frac{i}{n}\right)}}{\sum_{j=n_{0}+1}^{N-n_{1}} \sqrt{c_{X,W}\left(1, \frac{j}{N}\right)}}$$

Further assume that  $n_0/N \to \beta_0$  and  $n_1/N \to \beta_1$  as  $N \to \infty$ , then for large N we may use

$$p(w) = \frac{(\eta - \beta_1 - \beta_0 p_\star) \sqrt{c(1, F_W(w))}}{\int_{\beta_0}^{1 - \beta_1} \sqrt{c(1, v)} \,\mathrm{d}v}$$

and obtain

$$\sigma_p^2 = \frac{1}{p_\star} \int_0^{\beta_0} c_{X,W}(1,v) \, \mathrm{d}v + \frac{1}{(\eta - \beta_1 - \beta_0 \, p_\star)} \left( \int_{\beta_0}^{\beta_1} \sqrt{c_{X,W}(1,v)} \, \mathrm{d}v \right)^2 + \int_{1-\beta_1}^1 c_{X,W}(1,v) \, \mathrm{d}v.$$

Notice that the weights are never saturated when for all  $v \in [0,1]$  and some fixed  $p_{\star} \in (0,\eta)$ ,

$$\frac{n\sqrt{c_{X,W}(1,v)}}{\int_{\beta_0}^{\beta_1}\sqrt{c_{X,W}(1,v)}\,\mathrm{d}v} \in [p_\star, 1].$$

Then the asymptotic variance boils down to the simpler expression

$$\sigma_p^2 = \frac{\left(\int_0^1 \sqrt{c_{X,W}(1,v)} \,\mathrm{d}v\right)^2}{\eta}$$

From a practical viewpoint, to design an optimal survey sampling plan we need to estimate first the copula density by means of a simple sampling plan (*e.g.* sampling with equal probability without replacement) so as to build a (kernel or wavelet) estimator of  $c_{X,W}(1, v)$ ,  $\hat{c}_{X,W}(v)$  say, and to solve next the discrete optimization

#### P. BERTAIL ET AL.

problem in equation (5.1) (this may be done easily using Mathematica with the function Minimize for instance) with  $c_{X,W}(1, v)$  replaced by  $\hat{c}_{X,W}(v)$ . The performance of this procedure will be studied in a future work.

The following example illustrates the point discussed above.

**Example 3** (Frank copula). Consider the case where  $C_{X,W}$  is a Frank copula, *i.e.* there exists some  $\theta > 0$  such that, for all  $(u, v) \in [0, 1]^2$ ,

$$C_{X,W}(u,v) = C_{\theta}(u,v) := -\frac{1}{\theta} \log \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)} \right) \cdot$$

It is known that the associated Kendall's tau is given by

$$1 - \frac{4}{\theta} + \frac{4}{\theta^2} \int_0^\theta \frac{t}{e^t - 1} \,\mathrm{d}t,$$

[29] (Exercise 5.9 p. 171), which is an increasing function of  $\theta > 0$  (from 0 to 1). By some tedious but easy calculations, one can show that the density is bounded and that, for all  $v \in [0, 1]$ , we have the simple expression

$$c_{\theta}(1,v) = \frac{\theta}{e^{\theta}-1} e^{\theta v}.$$

Notice that this is itself a density on [0, 1], lower bounded by  $\theta/(e^{\theta} - 1)$ . We may thus apply our results and choose the optimal weights given by

$$p(w) = \frac{\theta \eta e^{\theta F_W(w)/2}}{2(e^{\theta/2} - 1)}$$

provided that they all lie in [0, 1]. This forces us to choose  $\eta \leq 2(e^{\theta/2}-1)/(\theta e^{\theta/2})$  (*i.e.* the greater the dependence, the smaller the survey sampling size). Then, we have

$$\sigma_p^2 = \frac{4 \left(e^{\theta/2} - 1\right)^2}{\eta \left(e^{\theta} - 1\right)\theta} \cdot$$

Notice that when  $\theta$  is close to 0 (*i.e.* in the non-correlated case), all optimal inclusion probabilities are close to  $\eta$  and the loss due to the sampling stage is only  $\sigma_p^2 = 1/\eta$ . Conversely, for a very large  $\theta$ , if we choose  $\eta$  of order  $2/\theta$  for instance,  $\sigma_p^2$  will be then of order 2. This means that a drastic reduction of the sample size may result from a very small variability of the Hill estimator, with a factor of order 2 for these optimal weights.

For a given survey sampling size, we may also solve the program in equation (5.1), but the calculations are more complicated in this case.

We would like to mention that such an approach fails for the family of extreme value copulas ([29], Sect. 3.3.4), which, except in the independent case, have a singularity and a discontinuity at the point (1, 1). The study of this specific case requires a much more sophisticated approach; this will be the subject of a future research.

### 5.3. Numerical experiments

As a complement to the theoretical results established in the previous section, we provide here some illustrations based on simulations. In particular, we consider a model that does not fulfill condition (ii) in Theorem 4.2, which requires the absolute continuity of  $F_{X,\mathbf{W}}$ . The encouraging empirical results we obtain nonetheless give hope that this assumption may be relaxed. Such desirable extensions are left for future research.

		$K^{\star}(N)$			
Scenario	$\gamma$	$N = 10^{3}$	$N = 5 \times 10^3$	$N = 10^4$	$N = 5 \times 10^4$
$S_1$	1/2	11	26	37	83
$S_2$	1	125	368	584	1709
$S_3$	2	514	1863	3245	11760

TABLE 1. List of scenarios depending on  $\gamma$  and corresponding optimal  $K^{\star}(N)$ .

### 5.3.1. Experiment setting

Simulations were based on the following model, chosen for its simplicity in terms of both computation and interpretation. Let W, the auxiliary information, have truncated Normal distribution  $F_W$  over  $[w_\star, w^\star] \subset \mathbb{R}$  with expectation  $\mu$  and variance  $\sigma_W^2$ , and define X, the variable of interest, as follows:

$$X = \frac{(1 - F_W(W))^{\gamma} - 1}{\gamma}, \ \gamma > 0.$$

Under such a representation, the distribution of X is a General Pareto with scale parameter 1 and EVI  $\gamma$ , *i.e.*  $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$ . This is a well-known family of distributions, the second order properties of which are easily derived ([11], Sect. 3.2). In particular, we have  $\rho = -\gamma$  and  $A(x) = x^{-\gamma}/\gamma$ . Following ([11], p. 80), the optimal number of largest observations in the population is

$$K^{\star} = K^{\star}(N) \sim \left[ \left( \frac{N^{2\gamma}}{2} \gamma^{3} (1+\gamma)^{2} \right)^{1/(2\gamma+1)} \right],$$

where  $\lfloor x \rfloor$  is the integer part of x. It follows that  $\sqrt{K^*}A(N/K^*) \to 0$  as  $N \to +\infty$ . Concerning the joint distribution of X and W, it is straightforward to see that

$$F_{X,W}(x,w) = F(x) \wedge F_W(w),$$

which means that the copula linking both marginals is the well-known singular copula  $M(u, v) := u \wedge v$ ,  $(u, v) \in [0, 1]^2$ . As stipulated in Remark 4.4, the multiplicative term in the asymptotic variance of the Horvitz–Thompson variant of the Hill estimator is then  $\sigma_p^2 = 1/p(w^*)$ .

For a given population  $\mathcal{U}_N$  of size  $\dot{N}$ , where it is assumed that  $\{W_i, i \in \mathcal{U}_N\}$  are independent realizations of W, inclusion probabilities of the Poisson sampling scheme are defined as

$$p_{i} = p(W_{i}) = n \, \frac{W_{i}}{\sum_{j=1}^{N} W_{j}},\tag{5.2}$$

with  $n = \eta N$ ,  $\eta \in (0, 1)$ , the desired expected sample size ([24], Sect. 6, p. 1512). Thus defined,  $p(W) \in [p_{\star}, p^{\star}]$ , where  $p_{\star} = \eta w_{\star}/\mu$  and  $p^{\star} = \eta w^{\star}/\mu$ , which offers an easy way of ensuring Assumption 1 is fulfilled. Furthermore, given the formula used to compute X as a function of W, the more extreme the observations, the greater the probabilities of inclusion.

Numerical experiments were conducted on a set of populations with increasing sizes  $N = 10^3$ ,  $5 \times 10^3$ ,  $10^4$ and  $5 \times 10^4$ . Several scenarios were investigated depending on the EVI  $\gamma$ ; they are summarized in Table 1. For each scenario, two sample sizes were considered: one small with  $n = 0.1 \times N$  and one relatively large with  $n = 0.5 \times N$ . Parameters of the distribution of W were chosen to ensure that for all  $i \in \mathcal{U}_N$ ,  $p_i \in [0.01, 1]$ . Specifically, we set  $\mu = 1$ ,  $\sigma_W^2 = 0.09$ ,  $w_* = 0.1$  and  $w^* = 2$ , thereby implying that  $(p_*, p^*) = (0.01, 0.02)$  when  $n = 0.1 \times N$  and  $(p_*, p^*) = (0.05, 1)$  when  $n = 0.5 \times N$ . Notice that in this last situation we obtain  $\sigma_p^2 = 1$ , *i.e.* sampling does not deteriorate the quality of estimation.



FIGURE 1. Average values of  $H_{K,N}^{\pi}$  (red line) and empirical 95% confidence band (pink area) computed on the 1000 simulated samples under scenario S<sub>2</sub> for  $n = 0.1 \times N$  (left hand plots) and  $n = 0.5 \times N$  (right hand plots), then compared to  $H_{K,N}$  (black dotted line) for  $N = 10^3$  (upper plots) and  $N = 5 \times 10^4$  (lower plots). (In color online)

For each scenario, we drew 1000 samples according to a rejective sampling scheme, following Algorithm 5.9 in [34]. The true inclusion probabilities, denoted by  $\pi_i$ ,  $1 \le i \le N$ , were deduced from their Poisson equivalents defined in equation (5.2) using a Monte-Carlo approximation technique, based on the repetition (10<sup>5</sup> times) of the basic algorithm stated in Section 4.2. Notice that since rejective sampling is a Poisson sampling conditioned upon its size, we have  $(p_i = 1) \Rightarrow (\pi_i = 1)$ .

The Horvitz–Thompson version of the Hill estimator was calculated using equation (3.3) on each of the 1000 simulated samples. The ensuing results are presented herein-after.

### 5.3.2. Experiment results

Illustrations of the behavior of  $H_{K,N}^{\pi}$  in a neighborhood of  $K^*(N)$  as N grows are presented in Figure 1 for each scenario. They can be compared to those of  $H_{k,n}$ , the Hill estimator in the sample that ignores the survey weights, in Figure 2. As a complement, we display in Figure 3 the empirical estimator of  $\eta \gamma^2 \sigma_p^2$  in a



FIGURE 2. Average values of  $H_{k,n}$  (red line) and empirical 95% confidence band (pink area) computed on the 1000 simulated samples under scenario S<sub>2</sub> for  $n = 0.1 \times N$  (left hand plots) and  $n = 0.5 \times N$  (right hand plots), then compared to  $H_{K,N}$  (black dotted line) for  $N = 10^3$  (upper plots) and  $N = 5 \times 10^4$  (lower plots). (In color online)

neighborhood of  $K^{\star}(N)$  for each scenario and each sample size as N increases; for large populations, this gives some indication as to the form of the variance of  $\sqrt{K} (H_{K,N}^{\pi} - \gamma)$ .

Since we only considered one fixed population, these results should be interpreted with caution: they only illustrate the behavior of  $\sqrt{K} \left( H_{K,N}^{\pi} - H_{K,N} \right)$  given the full vector  $(X_1, \mathbf{W}_1), \ldots, (X_N, \mathbf{W}_N)$ . We can see on Figure 1 that the Horvitz–Thompson version of the Hill estimator behaves perfectly well, even if the conditions in Remark 4.3 are not all satisfied; this corroborates the comments in Remark 4.4. In particular, both its mean and variance decrease with N, more quickly when  $n = 0.5 \times N$  (which is the ideal case with  $\sigma_p^2 = 1$ ) than when  $n = 0.1 \times N$ , and the distribution of the estimator appears to be symmetric around its classical version, which advocates normality. Observe that on the contrary,  $H_{k,n}$ , which ignores the survey weights, is clearly biased. Scrutinizing Figure 3, we can see that the asymptotic variance of  $\sqrt{K} \left( H_{K,N}^{\pi} - H_{K,N} \right)$  seems indeed to be finite,



FIGURE 3. Estimation of  $\eta \gamma^2 \sigma_p^2$  based on the 1000 simulated samples for  $\eta = 0.1$  (dotted lines) and  $\eta = 0.5$  (plain lines) under scenarios S<sub>1</sub> (grey lines), S<sub>2</sub> (black lines) and S<sub>3</sub> (red lines) compared to the theoretical value  $\gamma^2/2$  ("+" symbols on the ordinate axis). (In color online)

close to the theoretical values deduced from Remark 4.4. It depends on both the sample size and  $\gamma^2$ : the smaller the EVI, the smaller the variance.

# 6. CONCLUSION

In an attempt to start adapting classical extreme value analysis to the case of survey data, we introduced a Horvitz–Thompson version of the widely celebrated Hill estimator of the extreme value index. After exhibiting some sufficient hypotheses on both the superpopulation model and the sampling scheme for the consistency of this novel statistic to hold, we proved its asymptotic convergence to a Gaussian distribution when the survey design is of Poisson type. The exhibited rate of convergence appeared to be the same as the standard Hill estimator, namely  $\sqrt{K}$ , and the asymptotic variance was simply perturbed by a multiplicative constant depending solely on the sampling plan. In view of Remark 4.4 and of the empirical results presented, hope is

that the existence of a density copula linking those of the variable of interest and of the auxiliary information is not necessary for the asymptotic normality to be true. This encourages further research to try and relax this assumption. Other improvements may be brought to these first results, for instance situations where the true inclusion probabilities are not available and replaced by an estimated version issued from post-calibration methods could be inspected. Sampling designs of other nature than the Poisson type may also be considered.

# APPENDIX A. TECHNICAL PROOFS

# Proof of Theorem 3.2

We start by establishing the following intermediate results, in order to describe next the limit behavior of the Horvitz–Thompson tail empirical process.

First, we introduce the point measure:

$$\tilde{\nu}_N^{\boldsymbol{\pi}} := \frac{1}{K} \sum_{i=1}^N \frac{\epsilon_i}{\pi_i} \,\delta_{X_i/U(N/K)}.$$

Notice that the point measure  $\nu_N^{\pi}$  can be obtained from the latter by replacing the threshold U(N/K) by the empirical counterpart  $U_N^{\pi}(N/K)$ .

**Lemma A.1.** Under the assumptions of Theorem 3.2, as N, n and K tend to infinity so that  $K/N \rightarrow 0$ , we have:

$$\tilde{\nu}_N^{\pi} \Rightarrow \nu_{-1/\gamma},$$
(A.1)

where " $\Rightarrow$ " denotes weak convergence in the space of positive Radon measures on  $(0, +\infty]$ .

Proof. Consider first the tail empirical process

$$\nu_N := \frac{1}{K} \sum_{i=1}^N \delta_{X_i/U(N/K)}$$

We shall prove that for any t > 0, as N, n and K tend to  $+\infty$ , provided K/N converges to 0,

$$\mathcal{D}_N^{\boldsymbol{\pi}}(t) := \tilde{\nu}_N^{\boldsymbol{\pi}}(t, +\infty] - \nu_N(t, +\infty] \to 0 \text{ in } L_2.$$
(A.2)

Indeed,  $\forall t > 0$ , provided Assumptions 1 and 2 hold, we have

$$\mathbb{E}\left(\mathcal{D}_{N}^{\boldsymbol{\pi}}(t)\right) = \frac{1}{K} \sum_{i=1}^{N} \mathbb{E}\left(\left(\frac{\mathbb{E}\left(\epsilon_{i} \mid \left\{(X_{i}, \mathbf{W}_{i})_{1 \leq i \leq N}\right\}\right)}{\pi_{i}} - 1\right) \mathbb{I}\left\{X_{i} > t U(N/K)\right\}\right)$$
$$= 0,$$

together with

$$\begin{split} \mathbb{E} \left( \mathcal{D}_{N}^{\pi}(t)^{2} \right) &= \frac{1}{K^{2}} \sum_{i=1}^{N} \mathbb{E} \left( \mathbb{E} \left( \left( \frac{\epsilon_{i}}{\pi_{i}} - 1 \right)^{2} \middle| \{(X_{i}, \mathbf{W}_{i})_{1 \leq i \leq N}\} \right) \mathbb{I} \{X_{i} > t U(N/K)\} \right) \\ &+ \frac{2}{K^{2}} \sum_{1 \leq i \neq j \leq N} \mathbb{E} \left( \mathbb{E} \left( \left( \frac{\epsilon_{i}}{\pi_{i}} - 1 \right) \left( \frac{\epsilon_{j}}{\pi_{j}} - 1 \right) \middle| \{(X_{i}, \mathbf{W}_{i})_{1 \leq i \leq N}\} \right) \right) \\ &\times \mathbb{I} \{X_{i} \wedge X_{j} > t U(N/K)\} \right) \\ &= \frac{1}{K^{2}} \sum_{i=1}^{N} \mathbb{E} \left( \left( \frac{1}{\pi_{i}} - 1 \right) \mathbb{I} \{X_{i} > t U(N/K)\} \right) \\ &+ \frac{2}{K^{2}} \sum_{1 \leq i \neq j \leq N} \mathbb{E} \left( \left( \frac{\pi_{i,j} - \pi_{i} \pi_{j}}{\pi_{i} \pi_{j}} \mathbb{I} \{X_{i} \wedge X_{j} > t U(N/K)\} \right) \right) \\ &\leq \left( \frac{1}{\pi_{\star}} - 1 \right) \frac{1}{K^{2}} \sum_{i=1}^{N} \mathbb{P} \left( X_{i} > t U(N/K) \right) \\ &+ \left( \frac{\ell/n}{\pi_{\star}^{2}} \right) \frac{2}{K^{2}} \sum_{1 \leq i \neq j \leq N} \mathbb{P} \left( X_{i} \wedge X_{j} > t U(N/K) \right) =: \ell_{N,K}(t). \end{split}$$

Since  $\overline{F}$  is supposed to be regularly varying with index  $-1/\gamma$  and  $K/N \to 0$ , we have  $\mathbb{P}(X_i > t U(N/K)) \sim t^{-1/\gamma} K/N$  for all  $i \in \{1, \ldots, N\}$  as N and K go to infinity. It follows that as  $N, n, K \to +\infty$ ,

$$\ell_{N,K}(t) \sim \left(\frac{1}{\pi_{\star}} - 1\right) t^{-1/\gamma} \frac{1}{K} + \left(\frac{\ell/n}{\pi_{\star}^2}\right) t^{-2/\gamma} \left(1 - \frac{1}{N}\right) \to 0.$$

Hence, the convergence in equation (A.2) is proved and the desired convergence will then result from the fact that  $\nu_N \Rightarrow \nu_{-1/\gamma}$ , (see [30], Thm. 4.1).

We next prove the lemma below, claiming that the threshold  $U_N^{\pi}(N/K)$  and U(N/K) are asymptotically equivalent in probability.

Lemma A.2. Under the assumptions of Theorem 3.2, we have: as N, n and K tend to infinity,

$$\frac{U_N^{\pi}(N/K)}{U(N/K)} \to 1 \text{ in probability.}$$
(A.3)

*Proof.* This is a straightforward consequence of Lemma A.1. Indeed, for all  $\varepsilon > 0$ , we have:

$$\mathbb{P}\left(\left|\frac{U_{N}^{\pi}(N/K)}{U(N/K)} - 1\right| > \varepsilon\right) = \mathbb{P}\left(U_{N}^{\pi}(N/K) > (1 + \varepsilon) U(N/K)\right) \\ + \mathbb{P}\left(U_{N}^{\pi}(N/K) < (1 - \varepsilon) U(N/K)\right) \\ \leq \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\epsilon_{i}}{\pi_{i}}\delta_{X_{i}}\left((1 + \varepsilon) U(N/K), +\infty\right] \ge \frac{K}{N}\right) \\ + \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\epsilon_{i}}{\pi_{i}}\delta_{X_{i}}\left((1 - \varepsilon) U(N/K), +\infty\right] < \frac{K}{N}\right) \\ \leq \mathbb{P}\left(\tilde{\nu}_{N}^{\pi}\left(1 + \varepsilon, +\infty\right] \ge 1\right) + \mathbb{P}\left(\tilde{\nu}_{N}^{\pi}\left(1 - \varepsilon, +\infty\right] \le 1\right).$$

48

Therefore, by virtue of the lemma previously established, we asymptotically have:  $\tilde{\nu}_N^{\pi}(1 + \varepsilon, +\infty] \rightarrow 1/(1 + \varepsilon)^{1/\gamma} < 1$  and  $\tilde{\nu}_N^{\pi}(1 + \varepsilon, +\infty] \rightarrow 1/(1 - \varepsilon)^{1/\gamma} > 1$  in probability. Combined with the bound above, this proves the lemma.

Equipped with these preliminary results, we may now tackle the proof of Theorem 3.2. The consistency result can be established by following line by line the proof for the consistency of the Hill estimator in the i.i.d. situation given in [30]: by a continuous mapping theorem argument, one derives from Lemmas A.1 and A.2 that the Horvitz–Thompson tail empirical process  $\nu_N^{\pi}$  converges in probability to  $\nu_{-1/\gamma}$  in the space of positive Radon measures on  $(0, +\infty]$ . Then, it classically suffices to integrate the tail measures against dt/t (cf. Eqs. (3.1) and (3.3)) and apply the convergence previously mentioned. See ([30], Sect. 4.4.1) for further details.

## Proof of Theorem 4.2

As a first go, we start by establishing three intermediate results, which are introduced herein-after. The first lemma claims that the quantity  $r_{K,N}$  defined in equation (4.3) converges to 1 in probability.

**Lemma A.3.** Let  $\epsilon_1, \ldots, \epsilon_N$  and  $p_1, \ldots, p_N$  be respectively the inclusion indicators and probabilities of a Poisson survey plan in some population  $\mathcal{U}_N := \{1, \ldots, N\}$ . Then, provided Assumption 1 holds, for any  $K \in \{1, \ldots, N\}$  such that  $K := K(N) \xrightarrow[N \to \infty]{} +\infty$  we have

$$r_{K,N} := \frac{1}{K} \sum_{i=1}^{K} \frac{\epsilon_{N-i+1,N}}{p_{N-i+1,N}} \xrightarrow{\mathbb{P}} 1$$

*Proof.* Recall that under a Poisson sampling plan, all  $\epsilon_1, \ldots, \epsilon_N$  are independent. We had set  $\mathbb{E}(\epsilon_i \mid (X_i, \mathbf{W}_i)) := p_i$  for all  $i \in \{1, \ldots, N\}$ , hence

$$\mathbb{E}(r_{K,N}) = \frac{1}{K} \sum_{i=1}^{K} \mathbb{E}\left(\frac{\mathbb{E}\left(\epsilon_{N-i+1,N} \mid (X_1, \mathbf{W}_1), \dots, (X_N, \mathbf{W}_N)\right)}{p_{N-i+1,N}}\right) = 1.$$

In addition,  $\mathbb{V}(\epsilon_i \mid (X_i, \mathbf{W}_i)) := p_i(1 - p_i)$  for all  $i \in \{1, \ldots, N\}$ , therefore

$$\mathbb{V}(r_{K,N}) = \frac{1}{K^2} \sum_{i=1}^{K} \mathbb{E}\left(\frac{\mathbb{V}\left(\epsilon_{N-i+1,N} \mid (X_1, \mathbf{W}_1), \dots, (X_N, \mathbf{W}_N)\right)}{p_{N-i+1,N}^2}\right)$$
$$+ \frac{1}{K^2} \sum_{i=1}^{K} \mathbb{V}\left(\frac{\mathbb{E}\left(\epsilon_{N-i+1,N} \mid (X_1, \mathbf{W}_1), \dots, (X_N, \mathbf{W}_N)\right)}{p_{N-i+1,N}}\right)$$
$$= \frac{1}{K^2} \sum_{i=1}^{K} \mathbb{E}\left(\frac{1}{p_{N-i+1,N}}\right) - \frac{1}{K}$$

Under Assumption 1 it is clear that  $\mathbb{V}(r_{K,N}) = O(\frac{1}{K})$ . Since  $K \xrightarrow[N \to \infty]{} +\infty$ , this concludes the proof.  $\Box$ 

**Remark A.4.** The result stated in Lemma A.3 is also true when  $(\epsilon_1, \ldots, \epsilon_N)$  are distributed according to a rejective sampling plan  $R_N$ . Indeed, it is easy to check that its expectation equals 1 and that its variance tends to 0 by noticing that the inclusion probabilities of the rejective design satisfy both Assumption 1 and Assumption 2 (see [24], Eq. (5.27)).

We now move to the quantity  $\sqrt{K} \left( r_{K,N} H_{K,N}^{\mathbf{p}} - H_{K,N} \right)$  appearing in equation (4.2). The result below reveals that is vanishes asymptotically.

**Lemma A.5.** Suppose that Assumption 3 is fulfilled by the underlying heavy-tailed model and that Assumption 1 is satisfied by the considered sequence of Poisson inclusion probabilities  $p_1, \ldots, p_N$ ,  $N \ge 1$ . Assume also that  $K \to +\infty$  as  $N \to +\infty$  so that  $\sqrt{KA(N/K)} \to \lambda$  for some constant  $\lambda \in \mathbb{R}$ . Then we have

$$\sqrt{K}\left(r_{K,N} H_{K,N}^{\mathbf{p}} - H_{K,N}\right) \xrightarrow[N \to \infty]{\mathbb{P}} 0.$$

*Proof.* Let us start by introducing the weighted versions of *log-spacings* in the population, given by

$$\forall i \in \{1, \dots, N\}, \ \Delta_i := i \ (\log X_{N-i+1,N} - \log X_{N-i,N}).$$

Let  $K \in \{1, ..., N\}$ , these random variables are intrinsically linked to both  $H_{K,N}$  and  $H_{K,N}^{\mathbf{p}}$ , given by equations (2.3) and (3.3), respectively. Indeed, they can be expressed as

$$H_{K,N} = \frac{1}{K} \sum_{i=1}^{K} \Delta_i$$

and

$$H_{K,N}^{\mathbf{p}} = \left(\frac{1}{K} \sum_{j=1}^{K} \frac{\epsilon_{N-j+1,N}}{p_{N-j+1,N}}\right)^{-1} \frac{1}{K} \sum_{i=1}^{K} \left(\frac{1}{i} \sum_{j=1}^{i} \frac{\epsilon_{N-j+1,N}}{p_{N-j+1,N}}\right) \Delta_{i}$$

Combining the two we immediately obtain

$$r_{K,N} H_{K,N}^{\mathbf{p}} - H_{K,N} = \frac{1}{K} \sum_{i=1}^{K} \left( \frac{1}{i} \sum_{j=1}^{i} \frac{\epsilon_{N-j+1,N}}{p_{N-j+1,N}} - 1 \right) \Delta_i.$$
(A.4)

When Assumption (3) is fulfilled, it is possible to approximate the distribution of the  $\Delta_i$ 's corresponding to the K + 1 largest values. Denoting by  $E_1, \ldots, E_K$  a collection of independent random variables with standard exponential distribution, the random variables  $\Delta_i$  are approximately distributed as

$$\left(\gamma + \left(\frac{i}{K+1}\right)^{-\rho} A\left(\frac{N+1}{K+1}\right)\right) E_i, \quad 1 \le i \le K.$$
(A.5)

This property is at the basis of most of the asymptotic analyses that were lead concerning  $H_{K,N}$ , see [13] for more details. As mentioned in Remark 4.1, alternative approaches taking advantage of the Glivenko–Cantelli and Donsker theorems in the formulation of the Hill estimator in equation (2.3) were also developed, see [30] for instance.

Given the decomposition in equation (A.4), just like  $r_{K,N}$  in Lemma A.3 the expectation and variance of  $r_{K,N} H_{K,N}^{\mathbf{p}} - H_{K,N}$  are easily derived by conditioning upon the full vector of observations  $(X_1, \mathbf{W}_1), \ldots, (X_N, \mathbf{W}_N)$ . In particular, it is straightforward to see that

$$\mathbb{E}\left(r_{K,N}H_{K,N}^{\mathbf{p}}-H_{K,N}\right) = \mathbb{E}\left(\frac{1}{K}\sum_{i=1}^{K}\left(\frac{1}{i}\sum_{j=1}^{i}\frac{\epsilon_{N-j+1,N}}{p_{N-j+1,N}}-1\right)\Delta_{i}\right) = 0.$$

Turning now to the variance, under Assumption 1, we have

$$\mathbb{V}\left(r_{K,N} H_{K,N}^{\mathbf{p}} - H_{K,N}\right) = \frac{1}{K^2} \sum_{i=1}^{K} \mathbb{E}\left(\Delta_i^2 \frac{1}{i^2} \sum_{j=1}^i \left(\frac{1}{p_{N-j+1,N}} - 1\right)\right)$$
$$\leq \frac{1}{K^2} \sum_{i=1}^K \mathbb{E}\left(\Delta_i^2\right) \frac{p_\star^{-1} - 1}{i}.$$

To simplify notations, set  $\zeta = p_{\star}^{-1} - 1$  and consider the variable  $A_{K,N} = A\left(\frac{N+1}{K+1}\right)$  involved in equation (A.5). Using this particular equation, we are able to establish asymptotic properties of the right hand side in the inequality herein above: as  $N, K \to +\infty$  we have

$$\frac{1}{K^2} \sum_{i=1}^K \mathbb{E} \left( \Delta_i^2 \right) \frac{\zeta}{i} \sim \frac{1}{K^2} \sum_{i=1}^K \left( \gamma + \left( \frac{i}{K+1} \right)^{-\rho} A_{K,N} \right)^2 \mathbb{E} \left( E_i^2 \right) \frac{\zeta}{i} \\ \sim \frac{2\zeta\gamma^2}{K^2} \sum_{i=1}^K \frac{1}{i} + \frac{2\zeta}{K^2(K+1)^{-2\rho}} A_{K,N}^2 \sum_{i=1}^K \frac{1}{i^{2\rho+1}} \\ + \frac{4\zeta\gamma}{K^2(K+1)^{-\rho}} A_{K,N} \sum_{i=1}^K \frac{1}{i^{\rho+1}} \\ \sim \frac{\log K}{K^2} \left( 2\zeta\gamma^2 \right) - \frac{K^{-2\rho} - 1}{K^2(K+1)^{-2\rho}} \frac{\zeta}{\rho} A_{K,N}^2 \\ - \frac{K^{-\rho} - 1}{K^2(K+1)^{-\rho}} \frac{4\zeta\gamma}{\rho} A_{K,N} \\ \sim \frac{\log K}{K^2} \left( 2\zeta\gamma^2 \right) - \frac{1}{K^2} \frac{\zeta}{\rho} A_{K,N}^2 - \frac{1}{K^2} \frac{4\zeta\gamma}{\rho} A_{K,N}.$$

Because  $A \in \mathcal{R}_{\rho}$  with  $\rho < 0$ , then  $A_{K,N} \rightarrow 0$  as  $N, K \rightarrow +\infty$ , and we can conclude that  $\mathbb{V}\left(r_{K,N} H_{K,N}^{\mathbf{p}} - H_{K,N}\right) = o(1/K).$ 

The last intermediate result concerns the quantity  $Q_N^{(3)}$  in equation (4.2). It claims that it converges weakly to a centered Normal distribution.

Lemma A.6. Suppose that Assumption 1 is satisfied by the considered sequence of Poisson inclusion probabilities  $p_1, \ldots, p_N, N \geq 1$ , constructed from some set of auxiliary variables as in Assumption 4. Further assume that as  $N, K \to \infty, K/N \to 0$ ,

$$\mathbb{E}\left(\frac{1}{p(\mathbf{W})} \mid X > U(N/K)\right) \underset{N \to \infty}{\longrightarrow} \sigma_p^2 < \infty.$$

Then, provided that  $K \to +\infty$  as  $N \to +\infty$  so that K = o(N), we have the convergence in distribution as  $N \to +\infty$ :

$$\sqrt{K}(1-r_{K,N}) \Rightarrow \mathcal{N}(0, \sigma_p^2-1).$$

*Proof.* This proof is based on the application of ([17], Thm. 3, p. 262) to the collection of random variables  $\{Z_{i,N}(\boldsymbol{\epsilon}), 1 \leq i \leq N\}$  defined for all  $i \in \{1, \ldots, N\}$  as follows:

$$Z_{i,N}(\boldsymbol{\epsilon}) := \frac{1}{\sqrt{K}} \left( 1 - \frac{\epsilon_i}{p_i} \right) \mathbb{I} \left\{ X_i > X_{N-K,N} \right\},$$

with distribution  $\mathbb{P}_i$  only depending on the survey scheme (they are conditioned upon the vectors  $\{(X_i, \mathbf{W}_i), 1 \leq i \leq n \}$  $i \leq N$ }). Indeed, notice that we have  $\sqrt{K} (1 - r_{K,N}) = \sum_{i=1}^{N} Z_{i,N}(\epsilon)$ . In order to apply this theorem, we first have to check the three conditions below.

 $(\mathcal{C}_1)$  For all  $i \in \{1, \ldots, N\}$ , we have  $\mathbb{E}(Z_{i,N}(\boldsymbol{\epsilon})) = 0$  and  $\mathbb{V}(Z_{i,N}(\boldsymbol{\epsilon})) = \sigma_i^2 < \infty$ .

- ( $C_2$ ) There exists some real constant  $\sigma^2 < \infty$  such that  $S_N^2 := \sum_{i=1}^N \sigma_i^2 \xrightarrow[N \to \infty]{\mathbb{P}} \sigma^2$ . ( $C_3$ ) For each t > 0, we have  $\sum_{i=1}^N \int_{|z| \ge t S_N} z^2 \mathbb{P}_i(\mathrm{d}z) \xrightarrow[N \to \infty]{\mathbb{P}} 0$ .

#### P. BERTAIL ET AL.

**Condition** ( $C_1$ ). Let us start by calculating the expectation of  $Z_{i,N}(\epsilon)$  for all  $i \in \{1, \ldots, N\}$ . Because the  $\epsilon_1, \ldots, \epsilon_N$  are independent random variables with respective Bernoulli distributions  $\mathcal{B}(p_1), \ldots, \mathcal{B}(p_N)$ , it is straightforward to see that

$$\mathbb{E}\left(Z_{i,N}(\boldsymbol{\epsilon})\right) = \frac{1}{\sqrt{K}} \left(1 - \frac{\mathbb{E}\left(\epsilon_i \mid \{(X_i, \mathbf{W}_i)_{1 \le i \le N}\}\right)}{p_i}\right) \mathbb{I}\left\{X_i > X_{N-K,N}\right\} = 0.$$

As for the variance, we have

$$\mathbb{V}\left(Z_{i,N}(\boldsymbol{\epsilon})\right) = \mathbb{E}\left(Z_{i,N}^{2}(\boldsymbol{\epsilon})\right)$$
$$= \frac{1}{K} \mathbb{E}\left(\left(1 - \frac{\epsilon_{i}}{p_{i}}\right)^{2} \middle| \left\{(X_{i}, \mathbf{W}_{i})_{1 \leq i \leq N}\right\}\right) \mathbb{I}\left\{X_{i} > X_{N-K,N}\right\}$$
$$= \frac{1}{K} \frac{1 - p_{i}}{p_{i}} \mathbb{I}\left\{X_{i} > X_{N-K,N}\right\} =: \sigma_{i}^{2} < \infty.$$

Therefore, condition  $(\mathcal{C}_1)$  is fulfilled.

**Condition** ( $C_2$ ) We now have to prove that  $S_N^2$  converges in probability to a finite constant  $\sigma^2$  as N tends to infinity and exhibit the required conditions for this property to hold. First observe that

$$S_N^2 := \sum_{i=1}^N \sigma_i^2 = \frac{1}{K} \sum_{i=1}^N \frac{1}{p_i} \mathbb{I}\{X_i > X_{N-K,N}\} - 1,$$

where  $X_{N-K,N}$  is a consistent estimator of U(N/K) ([30], Sect. 4.4.1, p. 81). With this remark in mind, we will proceed in two steps and successively prove that there exists a real constant  $\sigma^2 > 0$  such that

$$\tilde{S}_N^2 := \frac{1}{K} \sum_{i=1}^N \frac{1}{p_i} \mathbb{I}\left\{X_i > U(N/K)\right\} - 1 \xrightarrow[N \to \infty]{\mathbb{P}} \sigma^2,$$

then that  $S_N^2 - \tilde{S}_N^2 \xrightarrow[N \to \infty]{\mathbb{P}} 0$ . Since  $S_N^2 = S_N^2 - \tilde{S}_N^2 + \tilde{S}_N^2$ , this will yield the desired result.

Let us start with  $\tilde{S}_N^2$ . Since  $\mathbb{P}\left(X > U(N/K)\right) = K/N$ , its expectation can be written

$$\mathbb{E}\left(\tilde{S}_{N}^{2}\right) = \mathbb{E}\left(\frac{1}{p(\mathbf{W})} \mid X > U(N/K)\right) - 1,$$

and under Assumption 1 its variance asymptotically vanishes. Hence, as  $N \to +\infty$  we have

$$\tilde{S}_N^2 + 1 = \mathbb{E}\left(\frac{1}{p(\mathbf{W})} \mid X > U(N/K)\right) + o_{\mathbb{P}}(1),$$

provided that this expectation exists, and when it is possible, we write  $\sigma_p^2$  its limit.

Under assumptions  $\mathcal{A}_1$  to  $\mathcal{A}_3$  in Remark 4.2, we can write

$$\mathbb{E}\left(\frac{1}{p(\mathbf{W})} \mid X > U(N/K)\right) = \frac{N}{K} \int_{\mathcal{W}} \int_{U(N/K)}^{\infty} \frac{1}{p(\mathbf{w})} c_{X,\mathbf{W}}\left(F(x), F_{W_1}(w_1), \dots, F_{W_d}(w_d)\right)$$
$$\times f(x) \prod_{j=1}^d f_{W_j}(w_j) \, \mathrm{d}x \, \mathrm{d}w_1, \dots, \mathrm{d}w_d.$$

Further set  $u := \frac{N}{K}\overline{F}(x)$ , then for  $K := K(N) \to +\infty$  as  $N \to +\infty$  and K = o(N) we have:

$$\mathbb{E}\left(\frac{1}{p(\mathbf{W})} \mid X > U(N/K)\right) = \int_{\mathcal{W}} \int_{0}^{1} \frac{1}{p(\mathbf{w})} c_{X,\mathbf{W}}\left(1 - \frac{K}{N}u, F_{W_{1}}(w_{1}), \dots, F_{W_{d}}(w_{d})\right)$$
$$\times \prod_{j=1}^{d} f_{W_{j}}(w_{j}) \, mdu \, \mathrm{d}w_{1}, \dots, \mathrm{d}w_{d},$$

which tends to

$$\sigma_p^2 = \int_{\mathcal{W}} \frac{1}{p(\mathbf{w})} c_{X,\mathbf{W}}(1, F_{W_1}(w_1), \dots, F_{W_d}(w_d)) \prod_{j=1}^d f_{W_j}(w_j) \, \mathrm{d}w_1 \dots \mathrm{d}w_d$$

as  $N \to \infty$ .

There remains to control the quantity  $|S_N^2 - \tilde{S}_N^2|$ , which we denote by  $S_N$  for simplicity. For any fixed  $N \in \mathbb{N}^*$  and  $\eta > 0$ , it can be decomposed as follows:

$$\mathcal{S}_N := \frac{1}{K} \sum_{i=1}^N \frac{1}{p_i} \left( \mathbb{I}\left\{ X_i > U(N/K) \right\} - \mathbb{I}\left\{ X_i > X_{N-K,N} \right\} \right) = \mathcal{S}_N^{(1)}(\eta) + \mathcal{S}_N^{(2)}(\eta),$$

where

$$\mathcal{S}_{N}^{(1)}(\eta) := \mathcal{S}_{N} \mathbb{I}\left\{ \left| \frac{X_{N-K,N}}{U(N/K)} - 1 \right| > \eta \right\}$$

and

$$S_N^{(2)}(\eta) = S_N \mathbb{I}\left\{ \left| \frac{X_{N-K,N}}{U(N/K)} - 1 \right| \le \eta \right\}.$$

We shall use this decomposition to prove that  $S_N \xrightarrow[N \to \infty]{\mathbb{P}} 0$ . Referring to Assumption 1, it is easy to see that  $\mathcal{S}_N^{(1)}(\eta) \xrightarrow[N \to \infty]{\mathbb{P}} 0.$  Indeed,

$$\begin{split} \left| \mathcal{S}_{N}^{(1)}(\eta) \right| &\leq \frac{1}{p_{\star}} \left| \frac{1}{K} \sum_{i=1}^{N} \left( \mathbb{I} \left\{ X_{i} > U(N/K) \right\} - \mathbb{I} \left\{ X_{i} > X_{N-K,N} \right\} \right) \right| \\ & \times \mathbb{I} \left\{ \left| \frac{X_{N-K,N}}{U(N/K)} - 1 \right| > \eta \right\} \\ & \leq \frac{1}{p_{\star}} \left| \frac{1}{K} \sum_{i=1}^{N} \mathbb{I} \left\{ X_{i} > U(N/K) \right\} - 1 \right| \times \mathbb{I} \left\{ \left| \frac{X_{N-K,N}}{U(N/K)} - 1 \right| > \eta \right\} \end{split}$$

The law of large numbers ensures that as  $N \to \infty$ ,

$$\frac{1}{K}\sum_{i=1}^{N} \mathbb{I}\left\{X_i > U(N/K)\right\} - 1 = \frac{N}{K} \mathbb{P}\left(X_i > U(N/K)\right) - 1 + o_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} 0.$$

In addition, we know that for all  $\eta > 0$ ,

$$\mathbb{I}\left\{\left|\frac{X_{N-K,N}}{U(N/K)} - 1\right| > \eta\right\} \xrightarrow[N \to \infty]{\mathbb{P}} 0$$

 $[\textbf{30}] \text{ (Sect. 4.4.1, p. 81), hence } \mathcal{S}_N^{(1)}(\eta) \xrightarrow[N \to \infty]{\mathbb{P}} 0.$ 

As for  $\mathcal{S}_N^{(2)}(\eta)$ , combining Assumption 1 with triangular inequalities provides an absolute bound:

$$\begin{split} \mathcal{S}_{N}^{(2)}(\eta) \bigg| &\leq \frac{1}{p_{\star}} \frac{1}{K} \sum_{i=1}^{N} \left| \mathbb{I} \left\{ X_{i} > U(N/K) \right\} - \mathbb{I} \left\{ X_{i} > X_{N-K,N} \right\} \right| \\ &\times \mathbb{I} \left\{ \left| \frac{X_{N-K,N}}{U(N/K)} - 1 \right| \leq \eta \right\} \\ &\leq \frac{1}{p_{\star}} \frac{1}{K} \sum_{i=1}^{N} \mathbb{I} \left\{ (1-\eta) U(N/K) < X_{i} \leq (1+\eta) U(N/K) \right\} \\ &\leq \frac{1}{p_{\star}} \frac{1}{K} \sum_{i=1}^{N} \mathbb{I} \left\{ X_{i} > (1-\eta) U(N/K) \right\} - \mathbb{I} \left\{ X_{i} > (1+\eta) U(N/K) \right\} \end{split}$$

Denote by  $c_N(\eta)$  the quantity in the right hand part of the last inequality, *i.e.* 

$$c_N(\eta) := \frac{1}{p_\star} \frac{1}{K} \sum_{i=1}^N \mathbb{I} \left\{ X_i > (1-\eta) U(N/K) \right\} - \mathbb{I} \left\{ X_i > (1+\eta) U(N/K) \right\}.$$

Applying the law of large numbers and recalling that  $\overline{F} \in \mathcal{R}_{-1/\gamma}$ , we get that as  $N \to +\infty$ ,

$$c_N(\eta) = \frac{1}{p_\star} \frac{N}{K} \left( (1-\eta)^{-1/\gamma} - (1+\eta)^{-1/\gamma} \right) \overline{F}(U(N/K)) + o_{\mathbb{P}}(1)$$
$$\xrightarrow{\mathbb{P}} \frac{1}{p_\star} \left( (1-\eta)^{-1/\gamma} - (1+\eta)^{-1/\gamma} \right) =: c_\eta.$$

Furthermore, notice that  $c_\eta \xrightarrow[\eta \to 0]{} 0$ , since

$$\begin{aligned} c_{\eta} &| = \frac{1}{p_{\star}} \left| (1-\eta)^{-1/\gamma} - (1+\eta)^{-1/\gamma} \right| \\ &\leq \frac{1}{p_{\star}} \left| (1+\frac{\eta}{\gamma} + o(\eta)) - (1-\frac{\eta}{\gamma} + o(\eta)) \right| \\ &\leq \frac{2\eta}{p_{\star} \gamma} + o(\eta) \underset{\eta \to 0}{\longrightarrow} 0. \end{aligned}$$

We may now prove that  $\mathcal{S}_N \xrightarrow[N \to \infty]{\mathbb{P}} 0$  by going back to the definition.

Formally, we have shown that  $|\mathcal{S}_N| \leq |\mathcal{S}_N^{(1)}(\eta)| + c_N(\eta) =: \beta_N(\eta)$ , for all  $\eta > 0$ . We also know that  $\beta_N(\eta) \xrightarrow[N \to \infty]{N \to \infty} c_\eta$  and  $c_\eta \xrightarrow[\eta \to 0]{} 0$ . Now for any fixed  $\delta > 0$ , we have to verify that  $\mathbb{P}(|\mathcal{S}_N| > 2\delta) \xrightarrow[N \to \infty]{N \to \infty} 0$ . First choose  $\eta_0 > 0$  such that for all  $0 < \eta \leq \eta_0$ ,  $|c_\eta| \leq \delta$ , then take some  $\xi > 0$ . We need to prove that there exists  $N_0 \in \mathbb{N}$  such that for all  $N \geq N_0$ ,  $\mathbb{P}(|\mathcal{S}_N| > 2\delta) \leq \xi$ . In order to construct this  $N_0$ , first fix any  $\eta > 0$  such that  $\eta \leq \eta_0$ . Since  $\beta_N(\eta) \xrightarrow[N \to \infty]{\mathbb{P}} c_\eta$ , there exists  $N_0 \in \mathbb{N}$  such that for all  $N \geq N_0$ , we have  $\mathbb{P}(|\beta_N(\eta) - c_\eta| > \delta) \leq \xi$ . In parallel, since  $\eta \leq \eta_0$ , we have for all  $N \in \mathbb{N}$ :

$$|\mathcal{S}_N| \le |\beta_N(\eta)| \le |\beta_N(\eta) - c_\eta| + |c_\eta| \le |\beta_N - c_\eta| + \delta.$$

This implies that for all  $N \ge N_0$ ,

$$\mathbb{P}\left(|\mathcal{S}_N| > 2\delta\right) \le \mathbb{P}\left(|\beta_N(\eta) - c_\eta| + \delta > 2\delta\right) = \mathbb{P}\left(|\beta_N(\eta) - c_\eta| > \delta\right) \le \xi.$$

Since this is true for any  $\delta > 0$ , this means that  $\mathcal{S}_N \xrightarrow[N \to \infty]{\mathbb{P}} 0$ .

In fine, we can conclude that under all the hypotheses stated in Lemma A.6,

$$S_N^2 \xrightarrow[N \to \infty]{\mathbb{P}} \sigma_p^2 - 1.$$

**Condition** ( $C_3$ ). Our last task consists in verifying that for any t > 0, the quantity below, denoted for simplicity by  $\mathcal{Z}_N(t)$ , converges to 0 as  $N \to +\infty$ . By definition, we have

$$\begin{aligned} \mathcal{Z}_{N}(t) &:= \sum_{i=1}^{N} \int_{|z| \ge t \, S_{N}} z^{2} \, \mathbb{P}_{i}(\mathrm{d}z) \\ &= \sum_{i=1}^{N} \mathbb{E} \left( Z_{i,N}^{2}(\epsilon) \, \mathbb{I} \left\{ |Z_{i,N}(\epsilon)| \ge t \, S_{N} \right\} \mid \left\{ (X_{i}, \mathbf{W}_{i})_{1 \le i \le N} \right\} \right) \\ &= \frac{1}{K} \sum_{i=1}^{N} \mathbb{I} \left\{ X_{i} > X_{N-K,N} \right\} \\ &\times \mathbb{E} \left( \left( 1 - \frac{\epsilon_{i}}{p_{i}} \right)^{2} \, \mathbb{I} \left\{ \left| 1 - \frac{\epsilon_{i}}{p_{i}} \right| \ge \frac{t \, S_{N}}{K \, \mathbb{I} \left\{ X_{i} > X_{N-K,N} \right\}} \right\} \mid \left\{ (X_{i}, \mathbf{W}_{i})_{1 \le i \le N} \right\} \right). \end{aligned}$$

Using Hölder's inequality, we obtain

$$\mathcal{Z}_{N}(t) \leq \frac{1}{K} \sum_{i=1}^{N} \mathbb{I}\left\{X_{i} > X_{N-K,N}\right\} \mathbb{E}\left(\left(1 - \frac{\epsilon_{i}}{p_{i}}\right)^{3} \middle| \left\{(X_{i}, \mathbf{W}_{i})_{1 \leq i \leq N}\right\}\right)^{2/3} \times \mathbb{E}\left(\mathbb{I}\left\{\left|1 - \frac{\epsilon_{i}}{p_{i}}\right| \geq \frac{t S_{N}}{K \mathbb{I}\left\{X_{i} > X_{N-K,N}\right\}}\right\} \middle| \left\{(X_{i}, \mathbf{W}_{i})_{1 \leq i \leq N}\right\}\right)^{1/3}.$$

Observe that under Assumption 1, we have

$$\left(\mathbb{E}\left(\left(1-\frac{\epsilon_i}{p_i}\right)^3 \mid \{(X_i, \mathbf{W}_i)_{1 \le i \le N}\}\right)\right)^{2/3} = \left(\frac{1}{p_i}\left(3-\frac{1}{p_i}\right)-2\right)^{2/3}$$
$$\leq \left(3\left(\frac{1}{p_\star}-1\right)\right)^{2/3}.$$

Moreover, conditionally upon the vectors  $\{(X_i, \mathbf{W}_i), 1 \leq i \leq N\}$ , the random variable  $\left|1 - \frac{\epsilon_i}{p_i}\right|$  equals either  $(p_i - 1)/p_i$  with probability  $p_i$  or 1 with probability  $1 - p_i$ . Therefore, by virtue of Markov's inequality, we can further bound  $\mathcal{Z}_N(t)$  from above:

$$\begin{aligned} \mathcal{Z}_{N}(t) &\leq \frac{1}{K} \sum_{i=1}^{N} \mathbb{I} \left\{ X_{i} > X_{N-K,N} \right\} \left( 3 \left( \frac{1}{p_{\star}} - 1 \right) \right)^{2/3} \\ &\times \left( \frac{1}{K} \mathbb{I} \left\{ X_{i} > X_{N-K,N} \right\} \frac{2 \left( 1 - p_{i} \right)}{t \, S_{N}} \right)^{1/3} \\ &\leq \frac{3^{2/3}}{K^{4/3}} \left( \frac{1}{p_{\star}} - 1 \right)^{2/3} \sum_{i=1}^{N} \mathbb{I} \left\{ X_{i} > X_{N-K,N} \right\} \left( \frac{2 \left( 1 - p_{i} \right)}{t \, S_{N}} \right)^{1/3}. \end{aligned}$$

Using again Assumption 1, this yields

$$\begin{aligned} \mathcal{Z}_N(t) &\leq \frac{2^{1/3} \times 3^{2/3}}{K^{4/3}} \left(\frac{1}{p_\star} - 1\right)^{2/3} \left(\frac{1 - p_\star}{t \, S_N}\right)^{1/3} \sum_{i=1}^N \mathbb{I}\left\{X_i > X_{N-K,N}\right\} \\ &\leq \frac{2^{1/3} \times 3^{2/3}}{K^{1/3}} \left(\frac{1}{p_\star} - 1\right)^{2/3} \left(\frac{1 - p_\star}{t \, S_N}\right)^{1/3}, \end{aligned}$$

where we have shown that  $S_N \xrightarrow[N \to \infty]{\mathbb{P}} \sigma$ . Consequently, the right hand part of this last inequality tends to 0 in probability as N tends to infinity for any t > 0. Hence, condition ( $\mathcal{C}_3$ ) is fulfilled.

With all three conditions  $(\mathcal{C}_1)$ ,  $(\mathcal{C}_2)$  and  $(\mathcal{C}_3)$  satisfied, by virtue of [17] (Thm. 3, p. 262) we finally have

$$\sqrt{K} (1 - r_{K,N}) \underset{N \to +\infty}{\Rightarrow} \mathcal{N} (0, \sigma_p^2 - 1).$$

We are now fully equipped to prove Theorem 4.2. Recall the decomposition in equation (4.2):

$$\sqrt{K} \left( H_{K,N}^{\mathbf{p}} - \gamma \right) = \underbrace{\frac{\sqrt{K}}{r_{K,N}} \left( r_{K,N} H_{K,N}^{\mathbf{p}} - H_{K,N} \right)}_{Q_N^{(1)}} + \underbrace{\frac{\sqrt{K}}{r_{K,N}} \left( H_{K,N} - \gamma \right)}_{Q_N^{(2)}} + \underbrace{\gamma \sqrt{K} \left( \frac{1}{r_{K,N}} - 1 \right)}_{Q_N^{(3)}}.$$

Combining Lemmas A.3 and A.5, provided that Assumptions 1 and 3 hold and that  $K = K(N) \to +\infty$ , K = o(N) and  $\sqrt{K}A(N/K) \to \lambda$  for some constant  $\lambda \in \mathbb{R}$ , we have

$$Q_N^{(1)} \xrightarrow[N \to \infty]{\mathbb{P}} 0.$$

Lemma A.3 also ensures that under Assumption 1,  $Q_N^{(2)}$  is equivalent to  $\sqrt{K} (H_{K,N} - \gamma)$ . Referring for instance to [11] (Thm. 3.2.5), this entails that provided Assumption 3 holds and that  $K = K(N) \to +\infty$ , K = o(N) and  $\sqrt{K}A(N/K) \to \lambda$  for some constant  $\lambda \in \mathbb{R}$ , we have the convergence in distribution as  $N \to +\infty$ :

$$Q_N^{(2)} \Rightarrow \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

Finally, by virtue of Lemmas A.3 and A.6, if Assumptions 1, 4 and conditions stipulated in Theorem 4.2 hold together with  $K = K(N) \rightarrow +\infty$  and K = o(N), we have the convergence in distribution as  $N \rightarrow +\infty$ :

$$Q_N^{(3)} \Rightarrow \mathcal{N}\left(0, \gamma^2 \left(\sigma_p^2 - 1\right)\right).$$

Because the limit distribution of  $Q_N^{(3)}$  was established conditionally upon the random vector  $\{(X_i, \mathbf{W}_i), 1 \leq i \leq N\}$  and in probability relative to this full vector of observations, we can consider  $Q_N^{(2)}$  and  $Q_N^{(3)}$  as independent random variables (one depends on the data and the other on the survey scheme). The limit distribution of their sum is then the sum of their limit distributions. This concludes the proof.

# Proof of Theorem 4.6

Consider a Poisson scheme  $T_N$  with probabilities  $p_1, \ldots, p_N$  and a rejective scheme  $R_N$  with probabilities  $\pi_1, \ldots, \pi_N$ . We shall prove the asymptotic normality of  $\sqrt{K} \left(H_{K,N}^{\pi}(R_N) - \gamma\right)$  using the following decomposition:

$$\sqrt{K} \left( H_{K,N}^{\pi}(R_N) - \gamma \right) = \underbrace{\sqrt{K} \left( H_{K,N}^{p}(R_N) - \gamma \right)}_{Q_N^{(4)}} + \underbrace{\sqrt{K} \left( H_{K,N}^{\pi}(R_N) - H_{K,N}^{\mathbf{p}}(R_N) \right)}_{Q_N^{(5)}}.$$

We shall successively prove that as  $N \to \infty$ ,  $Q_N^{(4)}$  is normally distributed and  $|Q_N^{(5)}|$  tends to 0 in probability. Let us start with  $Q_N^{(4)}$ . This expression may be treated as in [3] by recalling that a rejective survey plan is a Poisson scheme conditioned upon a fixed sample size  $\sum_{i=1}^{N} \epsilon_i = n$ :

$$\mathbb{P}_{R_N}\left(\sqrt{K}\left(H_{K,N}^{\mathbf{p}}(R_N) - \gamma\right) \le x\right) = \mathbb{P}_{T_N}\left(\sqrt{K}\left(H_{K,N}^{\mathbf{p}}(T_N) - \gamma\right) \le x \left|\sum_{i=1}^N \epsilon_i = n\right)\right)$$
$$= \frac{\mathbb{P}_{T_N}\left(\sqrt{K}\left(H_{K,N}^{\mathbf{p}}(T_N) - \gamma\right) \le x, \sum_{i=1}^N \epsilon_i = n\right)}{\mathbb{P}\left(\sum_{i=1}^N \epsilon_i = n\right)}.$$
(A.6)

Now consider the following regression:

$$\sqrt{K} \left( H_{K,N}^{\mathbf{p}}(T_N) - \gamma \right) = \frac{\operatorname{cov} \left( H_{K,N}^{\mathbf{p}}(T_N), \sum_{i=1}^N \epsilon_i \right)}{N^{-1} d_N} \times \frac{\sqrt{K}}{N} \left( \sum_{i=1}^N \epsilon_i - n \right) + \sqrt{K} Z_{K,N},$$

where  $Z_{K,N}$  is a residual term orthogonal to  $\frac{\sqrt{K}}{N} \left( \sum_{i=1}^{N} \epsilon_i - n \right)$ . Using the same type of decomposition as in equation (4.2) under the assumptions of Theorem 4.2, it is easy to see that the regression coefficient  $cov \left( H_{K,N}^{\mathbf{p}}(T_N), \sum_{i=1}^{N} \epsilon_i \right) N d_N^{-1}$  is bounded and that  $\frac{\sqrt{K}}{N} \left( \sum_{i=1}^{N} \epsilon_i - n \right) \to 0$  in probability as  $N, K \to \infty$ . Therefore,  $\sqrt{K} \left( H_{K,N}^{\mathbf{p}}(T_N) - \gamma \right)$  is asymptotically uncorrelated with the sample size, so that after applying the bivariate Central Limit Theorem to equation (A.6), we obtain that the asymptotic distribution of  $Q_N^{(4)}$  is the same as that of  $\sqrt{K} \left( H_{K,N}^{\mathbf{p}}(T_N) - \gamma \right)$ .

Let us now turn to  $Q_N^{(5)}$ . We have:

$$\begin{aligned} |Q_N^{(5)}| &= \sqrt{K} \left| H_{K,N}^{\pi}(R_N) - H_{K,N}^{\mathbf{p}}(R_N) \right| \\ &= \left| \frac{\sqrt{K} D_{K,N}(R_N, T_N) - H_{K,N}^{\mathbf{p}}(R_N) \sqrt{K} r_{K,N}(R_N, T_N)}{r_{K,N}(R_N, T_N) + r_{K,N}(R_N, \mathbf{p})} \right| \\ &\leq \frac{\sqrt{K} \left| D_{K,N}(R_N, T_N) \right| + H_{K,N}^{\mathbf{p}}(R_N) \sqrt{K} \left| r_{K,N}(R_N, T_N) \right|}{r_{K,N}(R_N, T_N) + r_{K,N}(R_N, \mathbf{p})}. \end{aligned}$$

#### P. BERTAIL ET AL.

We start by analyzing  $D_{K,N}(R_N, T_N)$ . Observe that it can be written as a function of  $d_N$  and  $\bar{p}$ , which have nice asymptotic properties:

$$D_{K,N}(R_N, T_N) = \frac{1}{K} \sum_{i=1}^{K} \epsilon_{N-i+1,N} \left( \frac{1}{\pi_{N-i+1,N}} - \frac{1}{p_{N-i+1,N}} \right) \log \left( \frac{X_{N-i+1,N}}{X_{N-K,N}} \right)$$
$$= \frac{1}{K} \sum_{i=1}^{K} \frac{\epsilon_{N-i+1,N}}{p_{N-i+1,N}} \frac{p_{N-i+1,N} - \pi_{N-i+1,N}}{\pi_{N-i+1,N}} \log \left( \frac{X_{N-i+1,N}}{X_{N-K,N}} \right)$$
$$= \frac{1}{K} \sum_{i=1}^{K} \frac{\epsilon_{N-i+1,N}}{p_{N-i+1,N}} \left( \frac{\bar{p}_N - p_{N-i+1,N}}{d_N} + o(1/d_N) \right)$$
$$\times p_{N-i+1,N} \left( \frac{1}{\pi_{N-i+1,N}} - 1 \right) \log \left( \frac{X_{N-i+1,N}}{X_{N-K,N}} \right).$$

Set  $\xi_{\star} := (1 - p_{\star}) \left(\frac{1}{\pi_{\star}} - 1\right)$ , then under Assumption 1 we have  $\forall K \leq N$ ,

$$\begin{aligned} |D_{K,N}(R_N, T_N)| &\leq \frac{1}{K} \sum_{i=1}^{K} \frac{\epsilon_{N-i+1,N}}{p_{N-i+1,N}} \left| \frac{\xi_{\star}}{d_N} + o(1/d_N) \right| \log\left(\frac{X_{N-i+1,N}}{X_{N-K,N}}\right) \\ &\leq r_{K,N}(R_N, \mathbf{p}) H_{K,N}^{\mathbf{p}}(R_N) \left| \xi_{\star} + o(1) \right| \frac{1}{d_N} \\ &\leq \left( \left( r_{K,N}(R_N, \mathbf{p}) H_{K,N}^{\mathbf{p}}(R_N) - r_{K,N}(T_N, \mathbf{p}) H_{K,N}^{\mathbf{p}}(T_N) \right) \\ &+ \left( r_{K,N}(T_N, \mathbf{p}) H_{K,N}^{\mathbf{p}}(T_N) - H_{K,N} \right) \right) \left| \xi_{\star} + o(1) \right| \frac{1}{d_N} \\ &+ \frac{H_{K,N}}{d_N} \left| \xi_{\star} + o(1) \right|. \end{aligned}$$

Recall that under Assumption 5 we have  $d_N = o(1/K)$  and that  $H_{K,N}$  is a consistent estimator of  $\gamma$  [28]. Therefore, combining Lemma A.5, Remark A.4, and the result for  $Q_N^{(4)}$ , we conclude that  $D_{K,N}(R_N, T_N) = o_{\mathbb{P}}(1/\sqrt{K})$ . Mimicking exactly this procedure and considering the same set of assumptions, we also obtain:

$$|r_{K,N}(R_N,T_N)| \le r_{K,N}(R_N,\mathbf{p}) |\xi_{\star} + o(1)| \frac{1}{d_N}$$

leading to  $r_{K,N}(R_N, T_N) = o_{\mathbb{P}}(1/\sqrt{K})$  by virtue of Remark A.4. Combining these two results with Theorem 3.2 yields

$$\sqrt{K} \left( H_{K,N}^{\pi}(R_N) - H_{K,N}^{\mathbf{p}}(R_N) \right) \xrightarrow{\mathbb{P}}_{N \to \infty} 0,$$

provided that  $K = K(N) \to +\infty$  as  $N \to +\infty$  and that K = o(N). This concludes the proof.

58

Acknowledgements. We would like to thank Pr. Johan Segers for his helpful remarks on this work. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01) and was partly funded by Inra, Anses and Crest-Ensai.

#### TAIL INDEX ESTIMATION BASED ON SURVEY DATA

### References

- J. Beirlant, Y. Goegebeur, J. Segers and J. Teugels, Statistics of extremes: theory and applications. John Wiley & Sons Inc (2004).
- [2] Y.G. Berger, Rate of convergence to normal distribution for the Horvitz-Thompson estimator. J. Stat. Plann. Inference 67 (1998) 209–226.
- [3] P. Bertail, E. Chautru and S. Clémençon, Empirical processes in survey sampling. Submitted (2013).
- [4] N.H. Bingham, C.M. Goldie and J.L. Teugels, Regular variation. Encycl. Math. Appl. Cambridge Univ Press, Cambridge (1987).
- [5] D. Bonnéry, J. Breidt and F. Coquet, Propriétés asymptotiques de l'échantillon dans le cas d'un plan de sondage informatif. Submitted (2011).
- [6] N.E. Breslow, T. Lumley, C. Ballantyne, L. Chambless and M. Kulich, Improved Horvitz-Thompson estimation of model parameters from two-phase stratified samples: applications in epidemiology. *Stat. Biosci.* 1 (2009) 32–49.
- [7] N.E. Breslow and J.A. Wellner, Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. Scand. J. Stat. 35 (2007) 186–192.
- [8] N.E. Breslow and J.A. Wellner, A Z-theorem with estimated nuisance parameters and correction note for "Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression". Scand. J. Stat. 35 (2008) 186–192.
- [9] W.G. Cochran, Sampling techniques. Wiley, New York (1977).
- [10] J. Danielsson, L. De Haan, L. Peng and C.G. De Vries, Using a bootstrap method to choose the sample fraction in tail index estimation. J. Multivariate Anal. 76 (2001) 226–248.
- [11] L. De Haan and A. Ferreira, Extreme value theory: an introduction. Springer Verlag (2006).
- [12] L. de Haan and L. Peng. Comparison of tail index estimators. *Stat. Neerl.* **52** (1998) 60–70.
- [13] L. de Haan and S. Resnick, On asymptotic normality of the Hill estimator. Stoch. Models 14 (1998) 849–867.
- [14] L. de Haan and S. Stadtmüller, Generalized regular variation of second order. J. Austral. Math. Soc. Ser. A 61 (1996) 381–295.
- [15] J.C. Deville, Réplications d'échantillons, demi-échantillons, Jackknife, bootstrap dans les sondages. Economica, Ed. Droesbeke, Tassi, Fichet (1987).
- [16] J.C. Deville and C.E. Särndal, Calibration estimators in survey sampling. J. Am. Stat. Assoc. 87 (1992) 376–382.
- [17] W. Feller, An introduction to probability theory and its applications, 2nd edition. John Wiley & Sons Inc., New York (1971).
- [18] R.D. Gill, Y. Vardi and J.A. Wellner, Large sample theory of empirical distributions in biased sampling models. Ann. Stat. 16 (1988) 1069–1112.
- [19] Y. Goegebeur, J. Beirlant and T. de Wet, Linking Pareto-tail kernel goodness-of-fit statistics with tail index at optimal threshold and second order estimation. *Revstat* 6 (2008) 51–69.
- [20] C.M. Goldie and R.L. Smith, Slow variation with remainder: theory and applications. Quart. J. Math. Oxford 38 (1987) 45–71.
- [21] M.I. Gomes and O. Oliveira, The bootstrap methodology in statistics of extremes choice of the optimal sample fraction. Extremes 4 (2001) 331–358.
- [22] C. Gourieroux, Théorie des sondages. Economica (1981).
- [23] C. Gourieroux, Effets d'un sondage: cas du  $\chi^2$  et de la régression. Economica, Ed. Droesbeke, Tassi, Fichet (1987).
- [24] J. Hajek, Asymptotic theory of rejective sampling with varying probabilities from a finite population. Ann. Math. Stat. 35 (1964) 1491–1523.
- [25] H.O. Hartley and J.N.K. Rao, Sampling with unequal probabilities and without replacement. Ann. Math. Stat. 33 (1962) 350–374.
- [26] B.M. Hill, A simple general approach to inference about the tail of a distribution. Ann. Stat. 3 (1975) 1163–1174.
- [27] D.G. Horvitz and D.J. Thompson, A generalization of sampling without replacement from a finite universe. J. Am. Stat. Assoc. 47 (1951) 663–685.
- [28] D.M. Mason, Laws of large numbers for sums of extreme values. Ann. Probab. 10 (1982) 756–764.
- [29] R.B. Nelsen, An introduction to copulas. Springer (1999).
- [30] S.I. Resnick, Heavy-tail phenomena: probabilistic and statistical modeling. Springer Verlag (2007).
- [31] P.M. Robinson, On the convergence of the Horvitz-Thompson estimator. Austral. J. Stat. 24 (1982) 234-238.
- [32] P. Rosen, Asymptotic theory for successive sampling. J. Am. Math. Soc. 43 (1972) 373-397.
- [33] T. Saegusa and J.A. Wellner, Weighted likelihood estimation under two-phase sampling. Preprint available at http://arxiv.org/abs/1112.4951v1 (2011).
- [34] Y. Tillé, Sampling algorithms. Springer Ser. Stat. (2006).